Underreporting of Occupational Injuries and Illnesses by Workers: A NEISS-Work Telephone Interview Survey

Request for Office of Management and Budget Review and

Approval for Federally Sponsored Data Collection

Section B

Project officer: Larry L. Jackson, Chief, Injury Surveillance Team
National Institute for Occupational Safety and Health
Division of Safety Research
1095 Willowdale Road, MS H1808
Morgantown, WV 26505

Phone: 304-285-5980
Fax: 304-285-5774
E-mail: LLJackson@cdc.gov

January 26, 2012

Table of Contents

B. Collection of Information Employing Statistical Methods

B.1   Respondent Universe and Sampling Methods

The telephone interviews will be a new data collection effort. Respondents will be identified through an ongoing emergency department surveillance system: the occupational injury supplement of the National Electronic Injury Surveillance System (NEISS-Work). NEISS is used to capture and report product-related injuries. Respondents will be civilian, non-institutionalized workers treated in a NEISS-Work hospital for an apparent occupational injury or illness. Selection of cases will be restricted to NEISS-Work hospitals among the small, medium, large, and very large hospital stratum (cases treated in a Children's Hospital will be excluded). Because of variations in the age of majority across states and the added complication of obtaining parental or guardian consent for a very small number of cases, respondents will be aged 20-64 years. Additionally, respondents must have been working for a wage or salary or be self-employed at the time of treatment. Volunteers and day laborers will be excluded (see Appendix J for the rationale for excluding day laborers). Respondents must be conversant in English or Spanish to be included.

Background on NEISS and NEISS-Work

In 1972, as authorized by statute (the Consumer Product Safety Act Sec. 5. [15 U.S.C. § 2054]), the Consumer Product Safety Commission (CPSC) initiated the collection of consumer product-related injury and illness information through a surveillance system that uses a national probability-based sample of hospital emergency departments—the National Electronic Injury and Illness System (NEISS). The NEISS data are abstractions of selected information from emergency department medical records as collected by paid records abstractors at hospitals contracted to collect data for CPSC. CPSC uses information obtained through NEISS to conduct in-depth follow-up investigations by telephone. CPSC collects information through the follow-up investigations with the approval of OMB (OMB Control No: 3041-0029) based on information collection extension requests every three years (e.g., Federal Register: Vol. 75, No. 65; Tuesday, April 6, 2010; 17391-17393).

NIOSH conducts research using an occupational supplement to NEISS as authorized by the Occupational Safety and Health Act, Section 20, "Research and Related Activities" and Section 22(d), "Authority of Director, National Institute for Occupational Safety and Health" (29 U.S.C. 669, 671 (d)). The work-related data obtained by NIOSH from CPSC through what is referred to as NEISS-Work do not contain direct personal identifiers such as name or contact information. Personal identifiers will not be provided to NIOSH in the proposed study, but will be obtained from the participating hospitals and retained by CPSC only for the purposes of conducting the intended follow-up interviews.

In the 1990's NIOSH conducted several follow-up telephone interview studies in collaboration with CPSC under their approval to collect information. OMB was regularly informed of such. Beginning in approximately 2002, CPSC requested that all federal agencies using NEISS for follow-up investigations seek their own OMB approval for the specific follow-up study (as is being done herein).

NEISS-Work

NEISS-Work data are collected from a national stratified probability sample created in 1997 of 67 of the approximately 5,400 rural and urban hospitals in the U.S. and its territories that have a minimum of six beds and that operate a 24-hour emergency department. General, specialty care, and military hospitals are included in the sample population; however, prison, psychiatric, rehabilitation and long-term care facilities, and Veterans Administration hospitals are excluded. Selection of the current hospital sample was based on a 1995 census of U.S. hospitals. Hospitals were stratified on the basis of both geographical location and hospital size (as determined by the number of annual ED visits). Data collection using the current hospital sample began in 1997 at 101 CPSC NEISS hospitals for consumer product-related cases and at 67 hospitals (2/3 of the CPSC sample) for work-related cases. Any eligible NEISS-Work sample hospital that stops reporting or refuses to continue participating in NEISS-Work is replaced with another hospital. These replacement hospitals were pre-designated as part of the original NEISS-Work sample design. If the hospital participates in the NEISS-Work data collection, but no ED patient record abstraction for NEISS-Work is done in a particular month, the hospital is retained in the sample and a nonresponse adjustment is made to the hospital weight.

Since selection of the NEISS-Work hospital sample in 1997, four small hospitals of the original 67 hospitals have permanently closed and thus were not replaced in the sample. The remaining 63 hospitals in the sample are distributed among the five sample strata with 28 small, 9 medium, 6 large, 15 very large sized hospitals, and 5 children's hospitals. Currently, one of the nine medium size hospitals is not reporting and a replacement hospital is being sought. Case weights are adjusted to account for this nonresponse until the hospital is replaced.

Sample Design

For the interview survey we are using a statistical sample design optimized for the NEISS-Work hospital sample. The resulting sample will be representative of a national workforce and maximize our ability to detect significant differences among the study subpopulations. Approximately 1,500 to 3,000 completed interviews are anticipated pending funding limitations and other operational constraints in place at the time of

interview. The sample design was developed based on an assumption of 2,000 completed interviews.

The sample design for selecting interview participants takes into account the underlying stratified hospital sample design for the NEISS-Work surveillance. It maintains the nationally representative aspects of the hospital sample design while maximizing our ability to detect significant differences among the study population (e.g., minimize variance issues). The design incorporates methodology to apply an appropriate statistical weight to each interview. This weight accounts for potential respondent biases when respondent characteristics are compared to NEISS-Work case characteristics as a whole. Finally, the sample design lends itself to a robust data analysis plan.

The sample design for the project was developed by Westat, a research services company that provides services to the United States government, among other entities. Westat has extensive experience in all aspects of survey design and analysis and its staff includes internationally recognized experts in research methodology, sample design, and estimation. For this project, NIOSH requested that the sample design focus on self-employed workers while simultaneously collecting representative information on chronic injuries and illnesses from all workers. NIOSH further requested that the sample design use methodology to (1) minimize the variance within hospital strata by using balanced designs in lieu of simple random sampling; (2) apply an appropriate statistical weight to each interview, taking into account potential respondent biases when respondent characteristics are compared to NEISS-Work case characteristics as a whole; and (3) optimize the ability for this project to attain reportable, stable, and valid data results that meet NIOSH privacy requirements. NIOSH has three criteria for determining reportability of NEISS-Work data results that are intended to ensure reasonable and reliable data quality and appropriate interpretation and use of these data[1]:

1. Number of cases treated within the hospital sample must exceed a specified value;

2. The extrapolated national estimates must exceed a specified value; and

3. The coefficient of variation must be less than or equal to 33%.

Based on these recommendations, Westat provided a report detailing the recommended sample design and the assumptions used in constructing it. The recommended sampling plan was based on a target of 2,000 completed interviews spread over the course of a year. The sampling frame excludes patients who are day laborers or volunteers. It consists of all work-related injury- or illness-related ED patient records for

---

[1] Because of privacy restrictions, NIOSH does not publicly release the minimum sample case or national estimate requirements. Variance requirements are released.

individuals aged 20 to 64 years who were reported at the NEISS-Work sample hospitals (excluding five Children's hospitals) over the course of the year. In the 2011 NEISS-Work files, this consists of 58 hospitals.

The intrinsic aspects of underreporting and the exclusion of certain populations in other surveys served as a driver for determining priority research domains for our study. These domains are self-employed persons not working in agricultural settings, farm workers, Hispanic persons, government workers, and persons anticipating worker's compensation to cover their expenses. Self-employed farm workers will only be combined with self-employed persons not working in agricultural settings if a sufficient sample cannot be collected to analyze these populations separately; wage earners on farms will be kept separate from non-wage self-employed farm workers who work on a farm that they or their family own. These characteristics can be identified by the NEISS-Work worker status, race, and payer variables. The business and occupation type variables can also be scanned to identify additional self-employed cases. For sample design purposes, the prevalence of each of these groups was estimated using NEISS-Work 2009 second and third quarter data with the hospital weights. The prevalence rates give an indication of how many of each group to expect for a given overall sample size if there were no oversampling.

When setting sampling rates, the goals were to minimize variation in final patient weights, obtain the required total initial sample size, and to acquire enough cases to make subgroup estimates. For this study, given the importance of the self-employed and farm workers and their very low prevalence in the sampling frame of each of these populations, all such persons will be taken into the sample with certainty to provide enough cases for producing estimates that meet the NIOSH minimum reporting requirements.

Next, the sampling rate for the remaining eligible patients at each hospital ED was calculated by first solving for the rate that gives an overall constant weight for the remaining patients across all 58 hospitals (the overall patient base weight is the product of the hospital weight x 1/(within-hospital sampling rate)). The rates were then "rounded" to integer rates for ease of sampling, and adjusted slightly if necessary to produce the total desired initial sample size (Table B.1). The initial within-hospital sampling rates do not vary within a stratum since the hospital weights are the same within each stratum (with the exception of one hospital in the medium stratum). The hospital sampling rates can be adjusted periodically to account for variation in response rates across hospitals to prevent a sample size shortfall. The rates have been set to produce 2,000 completed interviews per year. The design effect (deff)[2] that results from variation in the overall patient base weights

---

[2] The design effect is the ratio of the variance under the actual sample design to the variance for a simple random sample of the same size. It measures the effect on the variance of stratification, clustering of patients within hospitals, and weighting. Since the design effect reduces the effective sample size ($n_{eff} = n$/deff), and hence

for each set of sampling rates is also given in Table B.1 If the individual hospital sampling rates are modified over time to increase the yields, the overall patient weights may become more variable and this design effect will increase. The final patient weights will also become more variable when they are adjusted for interview nonresponse.

Table B.1 Patient sampling rates

Stratum	Hospital Weight	Sampling Rate per year		Patient Base Weight	
		All Others	Self-Employed, Farm Workers	All Others	Self-Employed, Farm Workers
Small	120.34	1 in 6	1 in 1	722.02	120.34
Medium	126.51	1 in 6	1 in 1	759.05	126.51
Large	88.78	1 in 9	1 in 1	799.03	88.78
Very Large	22.45	1 in 36	1 in 1	808.04	22.45

n=2,000 completed interviews over 1 year, deff(weights)=1.34

Each hospital is assigned a within-hospital sampling rate based on its stratum. The same sampling rate will initially be assigned to all hospitals in the stratum, based on the rate needed to minimize variation in the final patient weights and obtain the total required sample size. The initial total sample size will be inflated to allow for loss due to noncontact and nonresponse. An overall completion rate of 40 percent for sampled ED patients has been assumed. In reality, response rates will differ by patient characteristics and by hospital. Thus, a sample tracking system will be implemented to review the sample yields as the study progresses. The rates for some hospitals may need to be adjusted periodically to keep the sample yields on target should the response rate and contact assumptions prove to be inaccurate. The sample tracking system will also store the sampling rate used for each batch of sampled cases for use in calculating patient weights for analysis.

The expected number of completed interviews and expected precision for several subgroups of interest are given in Tables B.2 and B.3 The expected number of completed interviews was calculated based on prevalence estimates from 2009 second and third quarter NEISS-Work data, the proposed within-Hospital Sampling Rates, an overall patient

---

precision, we would prefer a sample design with the lowest design effect possible. The design effect due to weight variability is calculated as $1 + cv^2$ (where cv is the coefficient of variation of the weights); see Kish, 1992.

interview completion rate of 40 percent, and the total number of completed interviews. Table B.2 gives the coefficient of variation (*cv*) and 95 percent confidence interval half-widths for estimated proportions based on the entire sample and the expected number of completed interviews for each subgroup. Table B.3 gives the *cv* for the estimated total annual number of work-related injuries or illnesses for subgroups defined by employment status, race/ethnicity, sex, and expected payer. The tables show that for estimates of proportions and total number of work-related injuries or illnesses in each of these subgroups, the NIOSH precision requirement would be met.[3]

The design effects are important because they reflect the effects of stratification, weighting, and clustering of patients within hospitals. The effects of clustering are most harmful for characteristics with a high intra-class correlation coefficient (ICC) and when the cluster sizes are large. The ICC is a measure of how similar patients are within a hospital. It is the proportion of the total variance due to variability across hospitals – the more homogeneous patients are within hospitals, the larger the between-hospital component of the total variance. The design effect due to clustering within hospitals can be approximated by $1 + \rho*(m\text{-}1)$, where ρ is the ICC and m is the average number of patients per hospital for the sample design (Kish, 1992). This relationship shows the harmful effect of large cluster sizes on the variance for even moderate ρ. The ICC's for several characteristics were estimated using the 2009 second and third quarter NEISS-Work ED patients data (without any sub-sampling of patients) and are given in Table B.4

Table B.2 Precision for estimated proportions in the total NEISS-work population

Characteristic (%)	n	P-value	Design Effects (DEFF)	Standard Error (P) (SE(P))	Coefficient of Variation (P) (CV(P))	95% Confidence Interval (95% CI) (Half-Width)

Self-Employed	4280.03	2.1	0.0055	0.184	0.011
Farm Workers	2150.02	3.0	0.0054	0.271	0.011
Govt. Employees	1640.12	6.4	0.0184	0.153	0.037
Worker's Comp	9270.64	30.4	0.0592	0.092	0.119
Hispanic	2150.11	10.6	0.0228	0.207	0.046
Female	5780.35	2.4	0.0165	0.047	0.033
Below 1st Quartile Age	4800.25	3.0	0.0168	0.067	0.034
Above Median Age	10.50	2.8	0.0187	0.037	0.038

n = 2,000 completed interviews

Table B.3 Precision for estimated total number of ED patients reporting work-related injuries by subgroup

Subgroup (Number)	2009 Q2Q3					Expected Annual		
	Sampled n	Est.Total	SE(total)	cv(total)	Deff	Completes	Deff	cv(total)
Self-Employed	535	40,471	7,346	0.182	18.1	428	18.1	0.203
Farm workers	269	27,388	5,640	0.206	11.6	215	11.6	0.230
Gov. Employees	205	163,511	33,309	0.204	9.6	164	9.6	0.228
Worker's Comp	1,159	870,525	155,148	0.178	101	927	101	0.199
Hispanic	269	155,325	33,246	0.214	13.0	215	13.0	0.239
Female	723	477,055	58,973	0.124	16.7	578	16.7	0.138

n=2,000 completed interviews

Table B.4 Intra-class correlation coefficients for NEISS-Work patient characteristic proportions

Characteristic (%)	ICC
Self-Employed	0.03
Farm Worker	0.04
Government Employee	0.09
Worker's Compensation	0.46
Hispanic	0.15
Female	0.01

The power analysis is focused on testing equality of proportions for subgroups using a two-tailed t-test with alpha=0.05. Power is the probability the test will correctly detect a significant difference between the subgroups when there truly is a difference. The power to detect significant differences between subgroups is based on the effective sample sizes and proportions for the subgroups. (The effective sample size is the actual sample size divided by the design effect.) Power is given in Table B.5.

The minimum detectable difference (MDD) between two proportions is also given for a power of 0.80 for each subgroup comparison. The characteristics chosen for the proportion estimates are only examples based on variables available in the 2009 second and third quarter NEISS-Work data files. Examples in the table are given for characteristics with high intra-class correlation (Hispanic, expected payer is worker's compensation) and low intra-class correlation (sex). In general, MDD's of 8 to 11 percentage points could be detected.

Table B.5 Power to detect differences in two proportions P1 and P2 for a two-tailed t-test at α=.05

Total n	Self-Employed			All Others			%Female		Power	MDD for Power =.8
	n_1	deff	$n_{effective}$	n_2	deff	$n_{effective}$	P_1	P_2		
2000	428	1	428	1,572	2.4	655	0.09	0.36	0.99	0.083

H_0: $P_{self-employed}$ (female)=$P_{all\ others}$ (female)

Total n	Self-Employed			All Others			%Hispanic		Power	MDD for Power= .80
	n_1	deff	$n_{effective}$	n_2	deff	$n_{effective}$	P_1	P_2		
2000	428	1	428	1,572	11	143	0.08	0.12	0.26	0.075

H_0: $P_{self-employed}$ (Hispanic)=$P_{all\ others}$ (Hispanic)

Table B.5 Continued

Total n	Farm Worker			All Others			%GE Median Age		Power	MDD for Power= .80
	n_1	deff	$n_{effective}$	n_2	deff	$n_{effective}$	P_1	P_2		
2000	215	1	215	1,785	2.8	638	0.63	0.49	0.93	0.113

H_0: $P_{farm\ worker}$ (GE median age)=$P_{all\ others}$ (GE median age)

Total n	Hispanic			All Others			%Female		Power	MDD for Power =.80
	n_1	deff	$n_{effective}$	n_2	deff	$n_{effective}$	P_1	P_2		
2000	215	1	215	1,785	2.4	744	0.3	0.37	0.43	0.105

H_0: $P_{Hispanic}$ (female)=$P_{all\ others}$ (female)

Total n	Hispanic			All Others			%Worker's Comp		Power	MDD for Power= .80
	n_1	deff	$n_{effective}$	n_2	deff	$n_{effective}$	P_1	P_2		
2000	215	8.2	26	1,785	27.8	64	0.61	0.65	0.03	0.345

H_0: $P_{Hispanic}$ (worker's comp)=$P_{all\ others}$ (worker's comp)

One aspect of the sample design not described above nor investigated by Westat involves sampling aspects of "illness" cases. NIOSH does not currently differentiate between "injuries" and "illnesses" in reporting NEISS-Work results. In part, this arises from issues of appropriately defining and identifying in the abstracted record information what an injury is versus what an illness is. In part, it arises from the NIOSH emphasis on using NEISS-Work for improving injury prevention. For example, for routine classification purposes oriented towards prevention, NIOSH classifies health effects such as secondary infections post trauma (e.g., an infection arising secondary to a nail puncture) as the originating condition, that is, an "injury." Prior unpublished work suggested that illnesses represented less than 5% and up to no more than 10% of NEISS-Work cases. To aide this particular study, NIOSH conducted a more rigorous assessment of illnesses captured by

NEISS-Work over one year. Among workers presenting to an ED, approximately 4% of workers were diagnosed with conditions related to dermal, respiratory, circulatory, and digestive systems, plus infections and general illness signs and symptoms. For this interview study, these results suggest that to obtain sufficient interviews with workers presenting with illnesses to meet minimum reporting guidelines and to adequately address illness-related concerns in general, illness cases must be sampled with certainty in a similar fashion to cases involving self-employed workers. ED diagnosis categories and keyword searches will be used to identify illness-related cases for selection with certainty.

Sampling illness cases with certainty is expected to have similar sample design effects to self-employed (non-farm) and farm workers illustrated above. However, the sampling rate of "all other" cases will be decreased with a commensurate increase in their sample weight. Nevertheless to meet the most critical goals of this interview study within the constraints of the NEISS-Work data characteristics and worker populations, sampling self-employed workers and illness-related cases with certainty is a necessary methodological trade off. Obviously a small portion of the self-employed and farm workers selected with certainty will have been treated in the ED for an illness. Similarly, some of the "all other" cases evaluated above will have included illnesses. Thus, compared to the design evaluations presented above, which do not include illness, the effect of sampling illnesses with certainty on other evaluation categories is likely to be slightly diminished compared to the effects of sampling self-employed workers with certainty, assuming that illness cases and self-employed worker cases exhibit similar clustering affects.

B.2   Procedures for the Collection of Information

Stratification and Sample Selection

NEISS-Work Sample Selection

The hospital population for NEISS-Work data is based on two-thirds of the CPSC NEISS sample. The NEISS sample design is based on a stratified simple random sample of hospitals with an emergency department (ED) in the U.S. and its territories. A hospital is defined as a general or specialty care facility with a minimum of six beds and a 24-hour ED. The requirement for a hospital to have at least six beds conforms to the American Hospital Association (AHA) registration requirements (AHA, 2006).

The sample is stratified by hospital size based on the number of emergency department visits annually. Two organizations have historically maintained data on U.S. hospitals and ED usage. Data from the American Hospital Association and the SMG Marketing Group (now doing business as Verispan) have been used at various times to create the NEISS sample frame (Marker & Lo, 1996). Since 1988, the SMG hospital lists and ED usage data have been used for all sample redesigns and annual hospital adjustments. The SMG data were used to construct the current NEISS hospital sample with four size-related strata and a children's hospital stratum. In addition to stratification by hospital size, the NEISS sample is stratified geographically. Within each size stratum, a systematic

Figure B.1.1 U.S. distribution of CPSC hospitals

hospital sample was drawn from a geographically-ordered SMG hospital list. The U.S. distribution of CPSC hospitals in the NEISS sample is shown in Figure B.1.1.
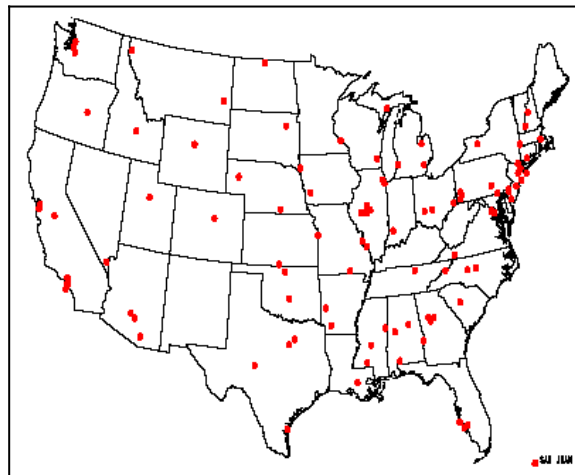
Since the initiation of the NEISS program in 1972, the CPSC hospital sample has been redesigned three times with implementation in 1978, 1990, and 1997. In addition to redesign changes, the number of hospitals in the sample has changed over time as CPSC has tried to enhance the data collection or reduce the system cost depending upon the vagaries of budgetary constraints. NIOSH has undergone similar expansions and contractions in its NEISS-Work data collection efforts. Currently, NIOSH collects data on all work-related injuries and illnesses treated in the ED at two thirds of the CPSC hospitals. The NEISS-Work data collection has been uniform and systematic since January 1, 1998, the last effective date of a break in series.

For the purposes of NEISS-Work methodology descriptions, the number of hospitals in the sample is defined as the number of hospitals in the sample at the time the current sample was initially selected. Currently, NEISS-Work uses the 1997 redesign hospital sample of 67 hospitals. When a hospital closes, the number of *in-scope* hospitals decreases because closures are not replaced in the sample. If a hospital simply withdraws from participating in NEISS, a new hospital is recruited and the original hospital is replaced, although there may be an extended lapse in reporting. The withdrawal of a hospital from NEISS or hospital non-response for a period of time does not result in a reduction of the number of in-scope hospitals (although it does influence the case weights for the period). The number of in-scope hospitals and reporting hospitals may change in any month of the year. At this time there are 63 in-scope reporting hospitals, including one non-reporting

hospital that will be replaced in the future. The 1997 CPSC sample redesign is based on a 1995 SMG sample frame. The full sample had 102 hospitals (1.9% of qualifying hospital EDs), but by the time the sample was implemented one hospital had closed resulting in 101 in-scope hospitals. CPSC used a Keyfitz procedure for resampling a stratified simple random sample that maximized the probability of retaining hospitals from the former sample (i.e., participating hospitals in 1996). As a result, 75 hospitals were retained and 26 new hospitals were recruited. As a part of this redesign, the children's hospital stratum became a probability based sample and no longer a simple convenience sample.

Although not used for the NEISS sample frame, the American Hospital Association annual surveys illustrate the decrease in emergency departments in community hospitals (nonfederal, short-term general and other special hospitals), while ED visits have increased in number and rate (Figure B.2.2) (AHA, 2006). Whereas the NEISS sample includes Federal and non-federal hospitals with 5,388 EDs in 1995, the AHA sample of community hospitals with 4,923 EDs in 1995 is generally representative of U.S. hospital trends as a whole.

Figure B.2.2 (a) Number of ED visits and number of EDs in community hospitals; (b) rate of ED visits per 1,000 persons; 1991-2004 (AHA, 2006).
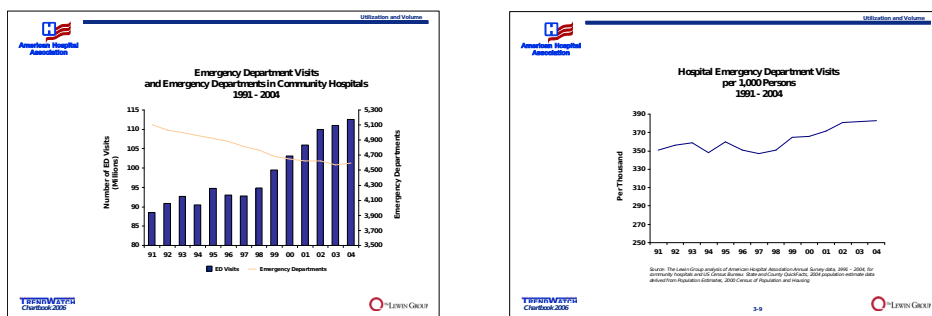
Source: The Lewin Group analysis of American Hospital Association Annual Survey data, 1991 – 2004, for community hospitals and US Census Bureau: State and County QuickFacts, 2004 population estimate data derived from Population Estimates, 2000 Census of Population and Housing

In October 1997, NIOSH implemented the 1997 CPSC sample design. However, budgetary constraints prohibited using the full 102 hospital sample. To continue with a sample of approximately the same size (i.e., ~65 hospitals) NIOSH obtained a new sample of 67 hospitals that was approximately two-thirds of the CPSC sample at that time. For the new NIOSH sample, 52 hospitals were retained from the prior sample and 15 new hospitals were added. Although adding a large number of new hospitals to the sample created some difficulties, work-related case reporting appeared stabilized by January 1998.

Each NEISS-Work case is assigned a statistical weight based on the inverse probability of selection. National estimates (i.e., the number of injuries and illnesses) are obtained by summing weights for all cases or particular cases of interest. The basic case weight is the inverse probability of selection for the hospitals in each stratum. The inverse probability of selection is the number of hospitals in the stratum universe divided by the number of hospitals in the NEISS sample for the stratum.

CPSC makes two types of routine weight adjustments to the basic case weights. First, weights are adjusted for non-participation if a hospital does not report fully during any given month or to account for hospital mergers, hospital closings or withdrawal from the NEISS-Work sample. Secondly, CPSC makes an annual ratio adjustment to the case weights by comparing the most recent U.S. hospital sample frame (i.e., for the prior year) with the 1995 sampling frame (used in 1997 for the latest NEISS sample). This adjustment is designed to account for changes in ED usage and the number of hospitals with EDs over time to provide the best opportunity for trend analysis and to minimize the expense of frequent sample redesigns. Thus, final case weights for each hospital stratum by month and year are calculated from the basic weight with adjustments for non-reporting and changes in the sampling frame over time.

Congressional Project Sample Selection

As stated on page six, the goals when setting sampling rates for this project were to minimize variation in final patient weights, obtain the required total initial sample size, and to acquire enough cases to make subgroup estimates. Thus, using the sample design described in section B.1 and the patient sampling rates shown in Table B.1 (p. 7), CPSC will select potential respondents weekly from incoming routine NEISS-Work case data. Prescreening using the basic NEISS-Work data elements will be used to restrict the potential respondents to those individuals most likely to meet the respondent definition (e.g., ages <20 and >64 and volunteers will be excluded). CPSC will then contact the participating hospital and request patient contact information. Individuals identified with potentially viable contact information will be sent one letter notifying them of the interview study and giving them the opportunity to "Opt Out." Contact information for individuals who do not opt out, or who fail to respond to the letter within ten days, will be provided to a third-party contractor who will conduct the interviews. Contact information will be provided by the CPSC approximately three weeks after the date of treatment. At no time will NIOSH have access to the individual identifiers or contact information for the potential respondents of the final interview survey.

The patient sampling will be done in batches on a flow basis. The frequency of sampling will depend on the volume of work-related injuries at the hospital ED. However, sampling

will occur throughout the entire 12 month period to avoid seasonal effects bias. Every eligible case will be given one (and only one) chance of selection. Prior to sampling cases, if possible, the list of patients will be sorted by demographic characteristics such as race/ethnicity, occupation type, sex, and age.

Each time sampling is done, the following information will be recorded in an electronic sample tracking sheet: hospital name and ID, date of sampling, total number reported, total number sampled, number of self-employed, number of farm workers, and the sampling rate used. Periodically, the total number sampled, number of self-employed, and number of farm workers will be tallied to check sample yields against the targets. If the total number sampled is below the expected number given how far the field period has progressed, the sampling rate in the hospital will be increased. New sampling rates will be calculated as follows: 1) update the prevalence rates of self-employed and farm workers in each hospital, 2) update the total number of eligible cases reported in each hospital over the first six months, 3) calculate the remainder sample size needed given what has been obtained so far, and 4) calculate new sampling rates to obtain the remainder needed. As the interviewing begins, interview response rates and number of completed interviews will also be monitored. If the response rates are lower than expected, the sampling rates will need to be increased.

Collection of Telephone Interview Data

Telephone interviewers are contracted through CPSC to complete the follow-back interviews. These interviewers are experienced interviewers and will receive additional training specific to the Underreporting questionnaire to be used for this study.

Prior to being contacted by telephone, potential participants will receive a letter describing the study and their protections as a participant should they choose to participate (Appendix E). This letter also provides them with the opportunity to opt out of participating in the study by calling a toll-free number. While the time for the telephone interview is not initially scheduled, participants do have the option at the time of contact to state that it is not a good time and schedule a better time to complete the interview. Also, if the potential participant initially declines to participate, text has been included in the telephone interview script to encourage them to reconsider.

Data Quality Control

Quality control of the data will not involve any additional contact with participants. Throughout data collection, a data cleaner will review the CATI database for appropriate values and skip pattern consistency. Analyses that will be used for this review include:

- One-way, labeled frequency distributions of database variables.
- Cross-tabulations of database variables to check skip patterns and other relationships.
- A query-by-identifier interactive report used to browse variable values by case.
- A query-by-value report to identify every record or record group matching a value, condition, or pattern.
- Interviewer comment file review – interviewers may enter comments about anything that was said or happened during an interview. The data cleaner will review this file and use it to resolve issues during data collection, such as an interviewer believing that the response did not fit any of the available categories or because it was outside a hard range.

Using all of these resources, the data cleaner may make changes to specific variables in specific interview records or a set of records. Any changes will be automatically captured in an edit log, which becomes part of the permanent documentation of the database. At this stage, the edit log contains any and all updates performed on a dataset during data collection, along with a brief note describing the reason for each edit. If an edit is performed, both the original coded value and the new updated value are documented in the log for each variable, for each affected case. As described below, this log is passed to the post-data collection data manager and maintained through all subsequent processing stages.

In addition to the CATI data cleaner's ongoing review of data during data collection, a second, independent review will be performed by the project data manager on the stable survey database immediately following data collection. The data manager will use the CATI instrument specifications and develop an independent SAS program that tests the integrity of the data collected. Any skip patterns/coding inconsistencies or violations of hard range values will be reviewed and any edits/updates will be documented. It should also be noted that before any edits or updates are performed, a back-up copy of the original dataset, as collected, will always be stored separately to allow for recourse in rare instances when there are problems with manipulated or processed datasets.

B.3   Methods to Maximize Response Rates and Deal with Nonresponse

We acknowledge that our projected response rate of 40%, based on the current CPSC reported 40-45% response rate, is low. However, it must be noted that this rate of overall response includes cases identified in NEISS-Work for which hospitals will not release contact information or correct contact information is unavailable. These insurmountable barriers drive the response rate down prior to us beginning to contact potential participants. In a recent study, this accounted for 35% of all potential cases.

Given a potentially low response rate, we plan to take several steps to help access potential participants and facilitate their willingness to participate. These steps include:

1. A letter describing the study will be sent to potential participants in advance of the initial phone call. This letter will alert and prepare potential participants for the phone call requesting their participation.

2. Telephone interviewers are required to make at least ten attempts to reach a potential respondent. The contact attempts are made at varying, but reasonable, hours of the day and on varying days of the week. When no personal contact is made after a number of attempts, the interview is set aside and contact attempts are made at a later date as time permits to maximize the response rate while minimizing recall bias issues. Interviewers are trained to be considerate of respondents and their families, leaving a minimal number of messages or speaking with the respondent or another individual of the residence to arrange a convenient interview time. Messages include a toll-free response number so that the respondent may call at their convenience. If personal contact is not made, a message system is not available, or there is an indication of an incorrect number, the interviewer will typically spread call attempts over a longer time period and commonly will make more than 10 contact attempts over the initial contact attempt period and the subsequent missed-interview follow-ups.

3. This project will use trained telephone interviewers who are experienced at conducting interviews. This will facilitate ease of survey participation for the respondent, increasing the likelihood that they will complete the survey in its entirety.

4. If the participant refuses the initial offer to participate in the study in a non-firm way, the interviewer will emphasize the importance of their participation and inquire as to whether they would be willing to participate at another time of their choosing. The training and experience of the telephone interviewers will be a key factor to understanding the reactions of potential participants and appropriately encouraging their participation in cases of refusal.

5. The questionnaire has been designed to be as easy and non-burdensome as possible. This includes ordering the questions in a logical sequence and asking only those questions that are needed for analysis purposes.

Despite a potentially low response rate, one of the benefits of this study is that we capture basic demographic and injury or illness information on all potential participants. Ultimately, we will compare the information we have on respondents and non-respondents using the NEISS-Work dataset to provide insight on any potential response bias. At a minimum, the case weights are adjusted for non-response within each stratum so that the

completed interviews within each stratum truly represent that stratum. If other factors are determined to influence answers, raking is performed so that the analysis weights for each variable of interest are equal to the corresponding national estimate.

B.4   Test of Procedures or Methods to be Undertaken

Interview Questionnaire

-

To achieve the aims set by Congress, we will use a telephone interview questionnaire that has been developed by NIOSH, based on applicable existing research and related questionnaires, and results of cognitive testing. This questionnaire will maximize our ability to identify an injured or ill worker's economic relationship to their job, confirm the work-relationship of the treated injury or illness, examine chronic injury or illness history, categorize the current injury or illness as chronic versus acute, and estimate the prevalence of chronic injury or illness in the sample population.

The interview, including the introductory materials, will be about 30 minutes or less in length. The interview will begin with an explanation of the study purpose and provide the information needed for informed consent. The subsequent questionnaire will begin with a brief series of qualifying questions, followed by an opportunity for the respondent to give a free form narrative statement of the recent injury or illness event. The remainder of the questionnaire will consist of separate modules that address specific worker or incident characteristics; issues related to reporting the injury or illness and the medical payer; and prior work-related injury/illness experience with a focus on chronic conditions. The specific modules included are: (1) initial introduction and screening questions; (2) classification of current injury or illness as acute or chronic; (3) current injury or illness characteristics; (4) type of employment; (5) employment characteristics; (6) ED reporting of the current injury or illness; (7) work reporting; (8) medical coverage and return to work; (9) history of chronic conditions; (10) demographic information; and (11) post-interview questions for the interviewer.

The initial draft questionnaire was developed by Westat, Rockville, MD, under contract to NIOSH. NIOSH staff revised the questionnaire extensively and harmonized the questionnaire with another underreporting survey being conducted by NIOSH. Additional revisions were completed based on reviewer comments and testing. Review comments were received from members of the NIOSH Division of Safety Research (DSR); the NIOSH Division of Surveillance, Hazard Evaluations, and Field Studies (DSHEFS); and the NIOSH Division of Respiratory Disease Studies (DRDS). In addition, the questionnaire was tested on a small number of employees at the NIOSH Morgantown branch who acted as

questionnaire respondents using constructed scenarios in order to test the skip pattern, flow, understandability, and comprehensiveness of the questions and their answer choices. Finally, survey experts from Research Triangle Institute (RTI), an independent, nonprofit research institution with more than 45 years of experience in survey methodology, reviewed and commented on the questionnaire, and conducted cognitive testing (explained in more detail below). Revisions based on the cognitive testing created the final questionnaire for interviewer administration.

The final English version questionnaire was translated to Spanish. The Spanish questionnaire was simplified in selected areas to minimize language, cultural, and conceptual differences among English-speaking and Spanish-speaking workers. The Spanish questionnaire version was tested using back translation, but did not undergo formal cognitive testing. Because the questionnaire changed somewhat when translated due to language and cultural differences, data from the Spanish-language interviews will be analyzed separately.

<u>Cognitive Testing</u>

RTI conducted cognitive testing of the questionnaire with nine potential respondents to insure clarity of questionnaire language and identify problems related to timing, skip patterns, and other complex conceptual issues that may not be readily obvious from simple reading of the questionnaire.

To identify the pool of participants for the cognitive interviews, CPSC selected 48 potential respondents from incoming routine NEISS-Work case data. Prescreening using the basic NEISS-Work data elements was used to restrict the potential respondents to those individuals who were most likely to meet the respondent definition (e.g., ages <20 and >64 and volunteers were excluded). CPSC then contacted the participating hospital and requested patient contact information. Individuals identified with potentially viable contact information were sent a letter (Appendix K) notifying them of the cognitive testing for the NIOSH interview study and giving them the opportunity to "Opt Out." Contact information for individuals who did not opt out were provided to CPSC by the hospital approximately three weeks after the date of treatment. CPSC provided the contact information for potential respondents to NIOSH for transmission to RTI following NIOSH privacy protocols. In compliance with Office of Management and Budget (OMB) requirements, RTI interviewed no more than 9 individuals as a part of this cognitive testing.

B.5 Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

Individuals who were consulted on statistical aspects
David A. Marker, Ph.D.
Senior Statistician
Westat, Inc.
1650 Research Blvd.
Rockville, MD 20850-3195
Phone: 301-251-1500

Pam Broene, Ph.D.
Senior Statistician
Westat, Inc.
1650 Research Blvd.
Rockville, MD 20850-3195
Phone: 301-251-1500

Tom Schroeder, MS
Statistician, Director
Division of Hazard and Injury Data Systems
U.S. Consumer Product Safety Commission
Phone: 301-504-0539 x1179
E-mail: TSchroeder@cpsc.gov

Individuals who will collect the data
CPSC staff and contracted interviewers under the direction of:
Tom Schroeder, MS
Statistician, Director
Division of Hazard and Injury Data Systems
U.S. Consumer Product Safety Commission
Phone: 301-504-0539 x1179
E-mail: TSchroeder@cpsc.gov

Individuals who will analyze the data

Larry Jackson, PhD
Chief, Injury Surveillance Team
Division of Safety Research, NIOSH
Phone: 304-285-5980
E-mail: LLJackson@cdc.gov

Susan Derk, MA
Epidemiologist, Injury Surveillance Team
Division of Safety Research, NIOSH
Phone: 304-285-6245
E-mail: SDerk@cdc.gov

Suzanne Marsh, MPA
Statistician, Injury Surveillance Team
Division of Safety Research, NIOSH
Phone: 304-285-6009
Email: SMMarsh@cdc.gov

Audrey Reichard, MPH, OTR
Epidemiologist, Injury Surveillance Team
Division of Safety Research, NIOSH
Phone: 304-285-6019
E-mail: AReichard@cdc.gov

Tom Schroeder, MS
Statistician, Director
Division of Hazard and Injury Data Systems
U.S. Consumer Product Safety Commission
Phone: 301-504-0539 x1179
E-mail: TSchroeder@cpsc.gov

Selected Citations

AHA. Hospital Statistics, 2006 ed: Health Forum, Chicago, IL. 2006.

Kish, L. "Weighting for unequal P(i)." *Journal of Official Statistics* 8, no. 2 (1992):183-200.

Marker, D.A., and Lo, A. (1996). *Update of the NEISS sampling frame and sample, final report*. Prepared by Westat for the Consumer Product Safety Commission, October 11, 1996.