Westat

Improving lives through research™

Survey of Occupational Injuries and Illnesses Employee Survey Cost Benefit Analysis Criteria

Authors

Lou Rizzo, Ph.D.
Cynthia Helba, Ph.D.
J. Michael Brick, Ph.D.
Matthew Bernstein
Mariel Leonard

May 12, 2015

Prepared for:
U.S. Bureau of Labor Statistics
2 Massachusetts Avenue
NE Washington, DC
20212-0001

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Executive Summary

This report provides an analysis of potential survey designs for the projected Employee Survey Component of the Survey of Occupational Injuries and Illnesses.

The key purpose of this study is to measure potential bias in the current employer-based SOII study. Researchers over the years have criticized the employer-based study as being biased downwards as a true measure of the prevalence of occupational injuries and illnesses. A preliminary power calculation shows that an effective sample size of 5,100 persons-years must be covered to achieve 80% power for detecting a 20% difference from the employer-based SOII study at its current prevalence level. (One person-year means one year of occupational injury and illness history is covered for one person.) This is only for an overall prevalence estimate. Once subgroups such as occupational categories, industries, and types of injuries and illnesses are considered, the sample size needs increase rapidly. A careful analysis must be done as to what power is needed for which subgroups, as measured against the resources available to carry out the study. The tables in Section 2 provide a starting point for this analysis.

The two major branches for carrying out the study which are presented in this report are a household-based survey and an employer-based survey. The household-based survey would be a nationally representative sample of households, and would ask for adults who have had occupational injuries or illnesses in the recent past. The prevalence of adults who would fall into this category is sufficiently high to make this a reasonable alternative. The household-based survey would be a stand-alone study, a component of a current national household study, or a follow-on to a current national household study. An employer-based survey would sample employers first, ask for a roster of current employees from those employers, and sample from these employee lists.

Major issues that are discussed include the problem of some employees (especially more vulnerable ones) who are expected to attempt to conceal injuries and illnesses from their employer, as is well-documented in the literature. This requires considerable effort to re-assure employees that the information they provide BLS would be completely confidential. This will be harder (but not impossible) in an employer-based study, but will still be necessary in a household-based study.

A very important issue as is documented in the literature is the issue of recall of injuries and illnesses. The longer the recall period, the more measurement error will be incurred as persons do not remember, or do not place correctly in time, their incidents. As the primary goal of this

employee-based study is to provide a fully unbiased estimator to supplement the employer-based SOII study currently done by BLS, this issue is a real concern. The time window asked about should be ideally shorter than three months. The gold standard would be a single month. But having interviews which have short recall periods requires then a large number of interviews. For example, the 5,100 person-year interviews specified as satisfying sufficient power for overall estimates at a national level will become 61,200 interviews, if each interview covers only one single month. There is a key cost-benefit tradeoff between the need for a short recall period and the cost of the multiple interviews (from the same or from different employees) that needs to be well thought-out.

Another major branch for household and employer studies is retrospective vs. prospective studies. Retrospective studies consist of identifying respondents, and then asking them about the recent past. If the interview focuses on a short recall period (e.g., one month), then a sufficient number of persons need to be sampled (e.g., 61,200). Retrospective studies are not recommended for the employer-based branch, as it is likely that recently injured employees will be left off of the employee lists provided by the employer. Prospective studies (or panel studies) consist of identifying respondents and recruiting them in an initial screening interview, but then following them into the future and asking them about their experiences in future months. This is the best solution for the employer-based branch, as these followups can be done outside of the worksite, and the initial self-selection into the study (by employer and employee) will not be dependent on incidence, as incidence is in the future at the time of the screening interview. In general, we would prefer prospective studies, as this guarantees that the initial response is not in any way correlated to prevalence. This reduces the potential for nonresponse bias in the prevalence estimates, which is strongly exacerbated by any dependence of response on prevalence.

For the household study, a stand-alone study could be retrospective or prospective or both. For a household study which is a component of a current national study, only a retrospective study is possible. For a household study which is a follow-on for a current national study, a prospective study is a natural fit.

Table of Contents

Contents Continued

Contents Continued

Contents Continued

Table **Page**

Figure **Page**

1 Introduction

The U.S. Bureau of Labor Statistics (BLS) Survey of Occupational Injuries and Illnesses (SOII) is the primary source of information on injuries and illnesses that take place in the workplace in the United States (excepting fatal injuries, which is covered in a census). The SOII is collected yearly from a sample of employers who report information from their OSHA logs and other documentation. There has been concern on the part of BLS researchers and outside researchers that the SOII is systematically undercounting the number of these injuries and illnesses. A literature review is provided in "SOII Research on Data Collection from Employees Literature Review" by Helba, Leonard and Bernstein at Westat (dated December 29, 2014) under this contract. The best way to evaluate this undercount in a scientifically sound way is to draw a nationally representative sample of employees who will themselves report on their injuries and illnesses during a specified time window.

This document takes as a starting point the literature review and elaborates a number of different proposals for carrying out this national sample of employees, based on the research presented in the papers in the literature review, based on Westat's experience in carrying out these kinds of studies, and based on feedback elicited from BLS staff as this document was being drafted. The relative costs and benefits of the various proposals are presented.

The key estimate of interest then is the prevalence of occupational injuries and illnesses per worker per year, nationally, and by occupation, by industry, and by injury/illness types. Section 2 provides a base power calculation for the sample sizes necessary to adequately discriminate and measure a bias in the SOII prevalence estimate, for a national estimate of all nonfatal injuries/illnesses, and for subgroups defined by occupation, industry, and injury type. Sections 3 and 4 then discuss issues of concern in developing a methodology and general methodological issues respectively, before launching into a systematic development of the alternatives. Sections 5 and 6 then articulate the detailed alternatives for household surveys and employer-based surveys respectively, which are the two major branches for achieving a nationally representative sample. Also included are a breakdown of relative costs and benefits of the possible alternatives. All of this is developed in-depth. Section 7 then provides conclusions and a summary of the alternatives and their cost/benefit tradeoffs.

2 Prevalence Differences

Important goals of the employee survey are coverage of the full population of employees of U.S. private industry and state and local government establishments, excepting specialized populations such as miners who are covered elsewhere. It would be an added benefit if the overlap with SOII sampled employers (employers providing logs to BLS) was high. This will improve analyses that link the two data sources. But this is not completely necessary. The important thing is that the employee study provides an independent benchmark on the prevalence estimates provided by the SOII employer logs, covering essentially the same population of employees in both cases.

The primary goal is to estimate the prevalence of employee workplace injury and illness: the percentage of employees in the universe who were injured or became ill in the workplace during a designated benchmark period. SOII reported that in 2013 there were 3.5 occupational injuries and/or illnesses per 100 employees in all industries including state and local government[1]. The current literature (see for example Wiatrowski 2014) indicates that this is likely an undercount of the truth, but it is not known by how much. For an auxiliary employee study to be useful in indicating the level of undercount bias, the precision of the auxiliary study must have enough power to successfully measure a difference from the SOII employer study estimate. For example, suppose in a future year the SOII employer study indicates 3.5 again, and the auxiliary employee study generates an estimate of 4.5 occupational injuries per 100 employees. If a 95% confidence interval is for example [2.5, 6.5], then we are really not sure that the difference between the employer study prevalence of 3.5 and the employee study prevalence of 4.5 is just a matter of sampling error. The 4.5 estimate is not measured with a great enough degree of precision to allow us to say with confidence that there is in fact a bias in the employer study prevalence.

The real key is being able to measure the difference between the SOII employer study national overall occupational injuries and illness prevalence estimate and the auxiliary employee study estimate with a sufficient level of precision so that the confidence interval around that difference is tight enough to identify meaningful differences. There are further issues of bias and measurement error that are difficult, and it is likely there will be interest in important subgroups (types of injury and/or illness; industries), but this simple criterion is a starting point.

[1] BLS news Release (USDL-14-2183) December 4, 2014, Table 1.

The relative standard error of the SOII employer study 2013 prevalence estimate is 0.5%. This is a standard error of 0.0175%, which is quite low. An auxiliary employee study is unlikely to be close to this, unless it is quite large. Table 1 below presents seven scenarios. In each scenario, the null hypothesis that the employer study prevalence and the auxiliary employee study prevalence are equal is tested, against the alternative that the auxiliary employee study prevalence is larger (i.e., a one-sided test). The null hypothesis is consistent with the employer study prevalence estimate not being biased downwards. The test is a 95% one-sided test using the difference between the two surveys, with a standard error of the difference computed based on assuming independence between the studies[2]. The effective sample size[3] for each scenario is the effective sample size that will provide 80% power for the alternative auxiliary employee survey prevalence as given in the third column. For example, for scenario 1, the designated alternative is a prevalence of 5.0% for the auxiliary employee study. This is a difference of 1.5% from the employer prevalence of 3.5%. An effective sample size of 1,310 will provide 80% power under the alternative. Also provided in Table 1-1 is the standard error of the auxiliary employee study prevalence (assuming the alternative), and a 95% confidence interval for the auxiliary employee study prevalence, assuming that the prevalence estimate is equal to the alternative mean.

Table 2-1 Scenarios with 80% power for measuring a difference between the employer study (with prevalence 3.5%) and an auxiliary employee survey[4]

Scenario	Effective sample size employee study	Auxiliary employee survey estimate	Standard error em-ployee estimate	Lower bound 95% CI	Upper bound 95% CI
1	1,310	5.00%	0.60%	3.82%	6.18%
2	1,675	4.80%	0.52%	3.78%	5.82%
3	2,250	4.60%	0.44%	3.73%	5.47%
4	3,220	4.40%	0.36%	3.69%	5.11%
5	5,100	4.20%	0.28%	3.65%	4.75%
6	9,570	4.00%	0.20%	3.61%	4.39%
7	25,660	3.80%	0.12%	3.57%	4.03%

As can be seen in Table 2-1, the effective sample size varies enormously given the alternative specified. The sample size of 1,310 will allow the distinguishing (with 80% probability) of a difference of 1.5 percentage points between the employer study prevalence estimate and the

[2] Some of the potential plans include having the employee survey nested within the employer study. This will induce correlation between the estimates. But the very small standard error of the employer study makes the effect of correlations negligible, and thus this effect can be disregarded.

[3] The effective sample size is the sample size of a simple random sample from the population that will achieve the indicated precision level.

[4] All of these calculations assume an underlying dichotomous variable: equal to 0 or 1 with probability 3.5-4.25%. This is only an approximation, as there are cases where individuals have multiple injuries or illnesses in a given year.

employee study prevalence estimate. But if the employee mean is closer to 3.5% (but still larger), then it is unlikely that the null hypothesis of no difference can be rejected, and the 95% confidence interval around the employee prevalence estimate will contain 3.5%.

The scenario that is of the greatest interest is the one for which the difference is 20% of the prevalence estimate. This is Scenario 5: the 4.2% employee survey percentage is equal to 1.2 times the 3.5% employer estimate (20% higher). The national employee effective sample size necessary to detect this difference with 80% power is 5,100. We will use this as a benchmark.

Estimating prevalence itself is one major goal of the study. Another goal is to generate improved overall estimates for all employees who have had an occupational injury/illness (an 'incident') in the past year. These are estimates for questionnaire items pertaining only to those who have had an incident in the past year (e.g., what kind of injury/illness did the employee have). These goals are certainly connected, but different, as is pointed out in Kalton (1993), Section 3. Table 2-2 provides expected confidence intervals for a 50% sample percentage within the set of employees with occupational injuries/illnesses (e.g., the sample percentage of employees with occupational injuries/illnesses in the past year who had a more serious injury (by some definition)). The confidence intervals are 95% confidence intervals, assuming no design effects. The assumed prevalence is 4% (a little higher than the 3.5% measured by the SOII employer study in 2013).

Table 2-2 Scenarios with 80% power for measuring a difference between the employer study (with prevalence 3.5%) and an auxiliary employee survey[5]

Scenario	Effective sample size auxiliary employee study	Assumed annual prevalence	Expected sampled employees with incident this year	Lower bound CI 50% sample percentage	Upper bound CI 50% sample percentage
1	1,310	4.00%	52	36.5%	63.5%
2	1,675	4.00%	67	38.0%	62.0%
3	2,250	4.00%	90	39.7%	60.3%
4	3,220	4.00%	129	41.4%	58.6%
5	5,100	4.00%	204	43.1%	56.9%
6	9,570	4.00%	383	45.0%	55.0%
7	25,660	4.00%	1,026	46.9%	53.1%

[5] All of these calculations assume an underlying dichotomous variable: equal to 0 or 1. This is only an approximation, as there are cases where individuals have multiple injuries or illnesses of the same type in a given year.

2.1 Prevalence Differences for Occupation Injury and Illness Cases

The total cases of occupational injuries and illnesses among all US eligible workers can be divided into four subgroups, as follows:

- Case 1: occupational injury/illness cases that cause at least one day away from work, or job transfer or restrictions;
  - Case 1.1: occupational injury/illness cases that cause at least one day away from work;
  - Case 1.2: occupational injury/illness cases that cause job transfers or restrictions;
- Case 2: other recordable occupational injury.

In 2013, the SOII estimate for the prevalence of these four cases per 100 person-years was 1.8, 1.1, 0.7, and 1.7 respectively[6]. Table 2-3 below presents the power for a one-sided test of the null hypothesis that the employer-based study and the employee-based study have the same prevalence, assuming the benchmark effective sample size of 5,100. The alternative hypothesis in each case is 20% higher than the null prevalences of 1.8%, 1.1%, 0.7% and 1.7% (1.8% and 2.16%, 1.1% and 1.32%, 0.7% and 0.84%, and 1.7% and 2.04% respectively). As can be seen the power is low for all four cases. The lower the prevalence percentages, the harder it is to achieve sufficient power.

Table 2-3 Power for five prevalence cases with an effective sample size of 5,100

Case	Effective sample size auxiliary employee study	SOII employer-based estimate	Standard error SOII employer-based estimate	Auxiliary employee survey estimate	Standard error em-ployee estimate	Lower bound 95% CI	Upper bound 95% CI	Power
Total	5,100	3.50%	0.0175%	4.20%	0.28%	3.65%	4.75%	80.02%
Case 1	5,100	1.80%	0.0090%	2.16%	0.20%	1.76%	2.56%	54.85%
Case 1.1	5,100	1.10%	0.0055%	1.32%	0.16%	1.01%	1.63%	39.39%
Case 1.2	5,100	0.70%	0.0035%	0.84%	0.13%	0.59%	1.09%	29.12%
Case 2	5,100	1.70%	0.0085%	2.04%	0.20%	1.65%	2.43%	52.83%

Table 2-4 presents the effective sample sizes necessary for achieving 80% power for each of the four cases.

[6] BLS News Release (USDL-14-2183) December 4, 2014, Table 1.

Table 2-4 Scenarios with 80% power for measuring a difference between the employer study and an auxiliary employee survey for five cases

	Effective sample size auxiliary employee study	SOII employer-based estimate	Standard error SOII employer-based estimate	Auxiliary employee survey estimate	Standard error employee estimate	Lower bound 95% CI	Upper bound 95% CI	Power
Total	5,100	3.50%	0.0175%	4.20%	0.28%	3.65%	4.75%	80.02%
Case 1	10,125	1.80%	0.0090%	2.16%	0.14%	1.88%	2.44%	80.01%
Case 1.1	16,700	1.10%	0.0055%	1.32%	0.09%	1.15%	1.49%	79.99%
Case 1.2	26,370	0.70%	0.0035%	0.84%	0.06%	0.73%	0.95%	79.99%
Case 2	10,730	1.70%	0.0085%	2.04%	0.14%	1.77%	2.31%	80.00%

The necessary sample sizes are much larger than 5,100. The goal of 80% power to distinguish a 20% difference in prevalence can be achieved for Case 1 and Case 2 with 10,730 as the effective sample size. It can only be achieved for Case 1.1 and 1.2 as well with 26,370 as the effective sample size.

Table 2-5 presents a breakdown for individual types of injuries and illness with the largest prevalence within the general class of Case 1. The estimated prevalence for each injury and illness type is the percentage of those injuries multiplied by the overall prevalence for Case 1. The coefficients of variations (the standard errors divided by prevalence) are provided rather than power calculations.

Table 2-5 Scenarios with 80% power for measuring a difference between the employer study and an auxiliary employee survey for five cases

Injury and illness code	Injury and illness type	Total injuries and illnesses 2013 (in 1000s)	Percent of injuries and illnesses	Effective sample size auxiliary employee study	Prevalence estimate	StdEr	CV
Total	*Total case 1*	*913.5*	*100.0%*	*5,100*	*1.80%*	*0.19%*	*10.3%*
123	Sprains, strains, tears	327.1	35.8%	5,100	0.64%	0.11%	17.4%
197	Nonspecified injuries	184.8	20.2%	5,100	0.36%	0.08%	23.2%
111	Fractures	78.0	8.5%	5,100	0.15%	0.05%	35.7%
143	Bruises, contusions	72.7	8.0%	5,100	0.14%	0.05%	37.0%
132	Cuts, lacerations	71.2	7.8%	5,100	0.14%	0.05%	37.4%
133	Puncture wounds	17.1	1.9%	5,100	0.03%	0.03%	76.3%
152	Heat burns	13.8	1.5%	5,100	0.03%	0.02%	85.0%
124	Hernias due to injury	13.5	1.5%	5,100	0.03%	0.02%	85.8%
162	concussions	10.7	1.2%	5,100	0.02%	0.02%	96.5%
182	Sprains and other injuries	10.1	1.1%	5,100	0.02%	0.02%	99.4%
Others	All others	114.7	12.6%	5,100	0.23%	0.07%	29.4%

2.2 Prevalence Differences by NAICS Three-digit Industry Subgroup

There is interest in each particular NAICS three-digit industry classification. In this section, expected effective sample sizes are presented for the case in which the benchmark effective sample size of 5,100 is allocated proportionately across the three-digit classifications. This would occur in any of the sample designs described in this paper if there is no explicit effort to oversample, assuming no differential response rates across industries to the employee study. If there are in fact differential response rates, then this needs to be taken into account in the sample design (in particular, those industries with systematically lower employee response rates may need to be oversampled to offset the lower response rates).

Table 2-6 presents the estimated employment and percent employment in the NAICS subgroup for 2013. These employment figures by NAICS subgroup are from the Quarterly Census of Employment and Wages - Bureau of Labor Statistics, 2013. The subgroups are ordered by employment size, with the largest subgroup at the top. Only the top 30 subgroups in terms of

employment are included in the table, with the remaining subgroups collapsed into an 'All Other Subgroups' category.

Also included in Table 2-6 are the incidence rates for occupational injuries and illnesses of all types for 2013 for the Table 2-6 subgroups. These are from the BLS Employer-Reported Workplace Injuries and Illnesses News Release for 2013 (December 4, 2014). The expected effective sample sizes are the employment percentages multiplied by 5,100, the benchmark effective sample size. This assumes completely proportional allocation of final effective sample sizes. The coefficients of variation (CVs) are the expected standard errors computed from these effective sample sizes, divided by the prevalence rate. Note that the CV is a function both of the effective sample size and the prevalence rate. Industries with lower prevalence rates (e.g., the financial industry) have higher CVs than industries with higher prevalence rates for the same sample sizes. It is harder to measure a smaller prevalence accurately.

The only three industries with CVs less than 30% are food services and drinking places, hospitals, and nursing care and residential facilities. These industries combine high employment and high prevalence rates.

It may be possible to improve the CVs for particular industries by oversampling these industries. This is especially easy for an employer-based sample design (see Section 5 below). This will degrade the precision of the national estimates however. If it is desired to have better CVs for a wide range of industries, the overall effective sample size needs to be increased. An n-times increase in the effective sample size will decrease the CVs listed in Table 1-5 by a factor of $\sqrt{n}$. For example, a four times increase in the effective sample size from 5,100 to 20,400 will reduce all of the CVs by a factor of 2 (e.g., from 25.1% to 12.5% for food services and drinking places).

Table 2-6 Expected Coefficients of variation for NAICS subgroups with the largest employment in 2013

NAICS Subgroup	Estimated employment 2013	Percent employment 2013	Incidence rate	Expected effective sample size	Expected CV
NAICS 722 Food services and drinking places	10,316,259	9.13%	3.3%	465.8	25.1%
NAICS 541 Professional and technical services	8,122,350	7.19%	1.0%	366.7	52.0%
NAICS 561 Administrative and support services	7,893,439	6.99%	2.5%	356.4	33.1%
NAICS 621 Ambulatory health care services	6,462,954	5.72%	2.7%	291.8	35.1%
NAICS 622 Hospitals	4,748,092	4.20%	6.4%	214.4	26.1%
NAICS 238 Specialty trade contractors	3,668,939	3.25%	4.2%	165.7	37.1%
NAICS 623 Nursing and residential care facilities	3,219,398	2.85%	7.3%	145.4	29.6%
NAICS 624 Social assistance	3,144,449	2.78%	3.4%	142.0	44.7%
NAICS 452 General merchandise stores	3,086,621	2.73%	4.8%	139.4	37.7%
NAICS 445 Food and beverage stores	2,934,003	2.60%	4.7%	132.5	39.1%
NAICS 423 Merchant wholesalers durable goods	2,867,522	2.54%	2.9%	129.5	50.9%
NAICS 611 Educational services	2,629,459	2.33%	2.0%	118.7	64.2%
NAICS 522 Credit intermediation	2,608,712	2.31%	0.8%	117.8	102.6%
NAICS 524 Insurance carriers	2,130,569	1.89%	0.8%	96.2	113.5%
NAICS 551 Management of companies	2,087,081	1.85%	1.2%	94.2	93.5%
NAICS 424 Merchant wholesalers, nondurable	1,980,354	1.75%	3.9%	89.4	52.5%
NAICS 721 Accommodation	1,849,249	1.64%	5.3%	83.5	46.3%
NAICS 441 Motor vehicle and parts dealers	1,789,864	1.58%	3.6%	80.8	57.6%
NAICS 336 Transportation equipment	1,513,893	1.34%	4.9%	68.4	53.3%
NAICS 311 Food manufacturing	1,467,238	1.30%	5.0%	66.2	53.6%
NAICS 713 Amusements, gambling, recreation	1,461,948	1.29%	4.1%	66.0	59.5%
NAICS 531 Real estate	1,449,906	1.28%	2.6%	65.5	75.6%
NAICS 332 Metal product manufacturing	1,421,941	1.26%	5.2%	64.2	53.3%
NAICS 448 Clothing and clothing accessories	1,388,530	1.23%	2.6%	62.7	77.3%
NAICS 484 Truck transportation	1,371,104	1.21%	4.6%	61.9	57.9%
NAICS 812 Personal and laundry services	1,339,327	1.19%	2.3%	60.5	83.8%
NAICS 813 Membership associations	1,331,923	1.18%	2.2%	60.1	86.0%
NAICS 236 Construction of buildings	1,281,387	1.13%	3.3%	57.9	71.2%
NAICS 811 Repair and maintenance	1,207,796	1.07%	2.8%	54.5	79.8%
NAICS 444 Building material & garden supplies	1,199,903	1.06%	4.8%	54.2	60.5%
All Other Subgroups	24,984,125	22.12%	4.1%	1,128.0	14.4%
TOTAL	112,958,335	100.00%	3.5%	5,100.0	7.4%

2.3 Prevalence Differences by Occupational Subgroups

There is interest in each particular occupational classification as well as the industry subgroups presented in Section 2.1. As in Section 2.1, expected effective sample sizes are presented for the case in which the benchmark effective sample size of 5,100 is allocated proportionately across the occupational classifications. This would occur in any of the sample designs described in this paper if there is no explicit effort to oversample, assuming no differential response rates across industries to the employee study.

Table 2-7 presents the estimated employment and percent employment in the occupational subgroup for 2013. These employment figures by occupational subgroup are from the 2013 United States National Occupational Employment and Wage Estimates (Bureau of Labor Statistics). The subgroups are ordered by employment size, with the largest subgroup at the top. Only the top 30 subgroups in terms of employment are included in the table, with the remaining subgroups collapsed into an 'All Other Subgroups' category.

Also included in Table 2-7 are the incidence rates for occupational injuries and illnesses that require days away from work or work transfers for 2013 for the Table 2-7 subgroups (from SOII and OES). The expected effective sample sizes are the employment percentages multiplied by 5,100, the benchmark effective sample size. This assumes completely proportional allocation of final effective sample sizes. The coefficients of variation (CVs) are the expected standard errors computed from these effective sample sizes, divided by the prevalence rate. Note that the CV is a function both of the effective sample size and the prevalence rate. Occupational subgroups with lower prevalence rates (e.g., the financial industry) have higher CVs than industries with higher prevalence rates for the same sample sizes. It is harder to measure a smaller prevalence accurately.

All of the CVs are rather large, but we are dealing with occupational injuries and illnesses that require days away from work, which is a smaller subgroup than all occupational injuries and illnesses.

Table 2-7 Expected Coefficients of variation for occupational subgroups with the largest employment in 2013

Occupation Code	Description	Employment count 2013 (in 1000s)	Pct employment	Private Industry count injuries 2013	Injury prevalence rate	Expected effective sample size	Expected CV
41-2000	Retail Sales Workers	8,501	6.41%	38,940	0.46%	327	81.52%
35-3000	Food and Beverage Serving Work	6,679	5.04%	22,370	0.33%	257	107.62%
43-4000	Information and Record Clerks	5,345	4.03%	12,050	0.23%	206	146.72%
29-1000	Health Diagnosing and Treating	4,761	3.59%	26,970	0.57%	183	97.90%
53-7000	Material Moving Workers	4,201	3.17%	69,930	1.66%	162	60.46%
13-1000	Business Operations Specialist	4,138	3.12%	5,690	0.14%	159	213.61%
25-2000	Teachers	4,024	3.04%	3,430	0.09%	155	275.19%
43-9000	Other Administrative Support W	3,879	2.93%	7,480	0.19%	149	186.25%
43-5000	Material Recording, Scheduling	3,799	2.87%	36,490	0.96%	146	84.00%
47-2000	Construction Trades Workers	3,761	2.84%	62,760	1.67%	145	63.82%
53-3000	Motor Vehicle Operators	3,664	2.76%	87,280	2.38%	141	53.92%
43-6000	Secretaries and Administrative	3,648	2.75%	3,770	0.10%	140	262.47%
15-1100	Computer Occupations	3,573	2.69%	1,640	0.05%	137	398.06%
43-3000	Financial Clerks	3,275	2.47%	5,170	0.16%	126	224.07%
37-2000	Building Cleaning and Pest Con	3,100	2.34%	41,470	1.34%	119	78.65%
35-2000	Cooks and Food Preparation Wor	2,979	2.25%	31,700	1.06%	115	90.08%
29-2000	Health Technologists and Techn	2,849	2.15%	23,010	0.81%	110	105.86%
49-9000	Other Installation and Repair	2,676	2.02%	47,520	1.78%	103	73.31%
13-2000	Financial Specialists	2,521	1.90%	1,750	0.07%	97	385.30%
39-9000	Other Personal Care and Servic	2,439	1.84%	16,140	0.66%	94	126.50%
51-9000	Other Production Occupations	2,430	1.83%	40,120	1.65%	93	79.83%
31-1000	Nursing, Psychiatric, and Home	2,362	1.78%	49,480	2.09%	91	71.72%
11-1000	Top Executives	2,278	1.72%	4,560	0.20%	88	238.53%
11-9000	Other Management Occupations	2,074	1.56%	12,350	0.60%	80	144.66%
51-4000	Metal Workers and Plastic Work	1,888	1.42%	26,280	1.39%	73	98.77%
21-1000	Community and Social Service S	1,830	1.38%	7,660	0.42%	70	183.84%
51-2000	Assemblers and Fabricators	1,758	1.33%	15,450	0.88%	68	129.15%
41-4000	Sales Representatives, Wholesa	1,757	1.32%	2,290	0.13%	68	336.72%
41-3000	Sales Representatives, Service	1,654	1.25%	2,290	0.14%	64	336.71%
11-3000	Operations Specialties Manager	1,582	1.19%	4,400	0.28%	61	242.74%
17-2000	Engineers	1,548	1.17%	1,310	0.08%	60	445.30%

Table 2-7 Expected Coefficients of variation for occupational subgroups with the largest employment in 2013 (continued)

Occupation Code	Description	Employment count 2013 (in 1000s)	Pct employment	Private Industry count injuries 2013	Injury prevalence rate	Expected effective sample size	Expected CV
25-1000	Postsecondary Teachers	1,511	1.14%	480	0.03%	58	735.83%
49-3000	Vehicle and Mobile Equipment M	1,464	1.10%	23,930	1.63%	56	103.38%
41-1000	Supervisors of Sales Workers	1,459	1.10%	15,350	1.05%	56	129.45%
33-9000	Other Protective Service Worke	1,457	1.10%	9,390	0.64%	56	165.86%
25-9000	Other Education Occupations	1,448	1.09%	2,240	0.15%	56	340.41%
31-9000	Other Healthcare Support Occup	1,402	1.06%	7,590	0.54%	54	184.57%
43-1000	Supervisors of Office Workers	1,367	1.03%	1,800	0.13%	53	379.79%
All Oth	All Others	21,510	16.2%	141,850	0.66%	827	42.67%
Total	Total	132,589	100%	914,380	0.69%	5,100	16.80%

Issues of Concern 3

The calculations in Section 1 above are based on a rather ideal world. There is no undercoverage, no nonresponse, and no measurement error. In this series of sections, we explore basic issues in a general way which have been identified as important in the literature search as important sources of bias. Section 2.1 discusses recall and telescoping issues, Section 2.2 discusses concealment issues, and Sections 2.3 and 2.4 discusses response rate issues.

3.1 Measurement Error from Recall Problems

An important species of measurement error in this case is the issue of recall. Warner et al. (2005) found considerable memory decay and telescoping issues in recalling injuries in the National Health Interview Survey (NHIS). Table 3-1 presents results from Table 3 in the Warner et al. paper. These are estimates from the NHIS 1997-1999 of annualized injury episodes per week. The window for the question items was a three month period, and information was elicited about all thirteen of these weeks going backwards in time[7]. There is no reason why weeks 6 through 13 before the questionnaire date should be any different from weeks 1 through 5 before the questionnaire date as the date of the questionnaire was arbitrary, except that recall was different about the earlier period.

Table 3-1 Average annual weighted number of injury episodes reported per week for first six weeks of NHIS questionnaire window and last seven weeks of window

	Weeks 1-5	Weeks 6-13	Percent difference
All episodes	658	605	-8.1%
Sprains/strains	179	167	-6.7%
Open wounds	141	121	-14.2%
Fractures	106	108	1.9%
Contusions/superficial injuries	100	76	-24.0%
Hospitalized for injury–Yes	41	43	4.9%
Not hospitalized for injury	615	559	-9.1%
Time lost from school or work	231	230	-0.4%
No time lost from school or work	303	266	-12.2%

[7] The NHIS questionnaire item was "During the past three months, that is since [91 days before today's date], [were/was] [you/anyone in the family] [injured/poisoned] seriously enough that [you/they] got medical advice or treatment?"

For fractures and for injuries requiring hospitalization, there was almost no fall-off, indicating recall for the earlier period was not different from recall for the later period. But for contusion/superficial injuries, the difference was considerable (24% less in the later period).

Warner et al. (2005) based on their NHIS research recommended a very short recall period (one month or a little longer) for these kinds of injuries. Especially smaller injuries are recalled with less accuracy when there is a longer lag time between the event and the time of questionnaire completion.

For the purposes of this cost-benefit analysis we will posit the following recall model. We define the parameter RL as the percentage of interview respondents who will fail to recall an incident if the window is extended for one month longer. If m is the number of months between the event point in time and the recall point in time, then the percentage of interview respondents who will respond under this model is $P(m) = (1 - RL)^m$. This kind of attrition model is broadly consistent with the Warner et al. (2005) data. We chose an exponential model rather than a linear model as a linear model projects too much loss of recall for longer periods (after a while, the linear model gives zero and negative recall percentages, which isn't realistic). The exponential model posits that each fixed time period loses a fixed percentage of respondents recalling an event accurately.

The particular value of RL that is consistent with an 8.1% decrease in recall between Weeks 1 and 6 and Weeks 7 and 13 for the Warner NHIS data is 6.0%[8]. Figure 3-1 below presents the percent recall one could expect for month window lengths from 1 through 12. A one-month window has a recall rate of 97.0%; a twelve-month window a recall rate of 49.2%.

[8] If the monthly exponential decay rate is 6.0%, then a weekly exponential decay rate will be (7/30)*6.0%, or 1.4%. This translates to a six-week decay of (1-.0.014)^6, which is 0.919, or an 8.1% decay for a six-week period.

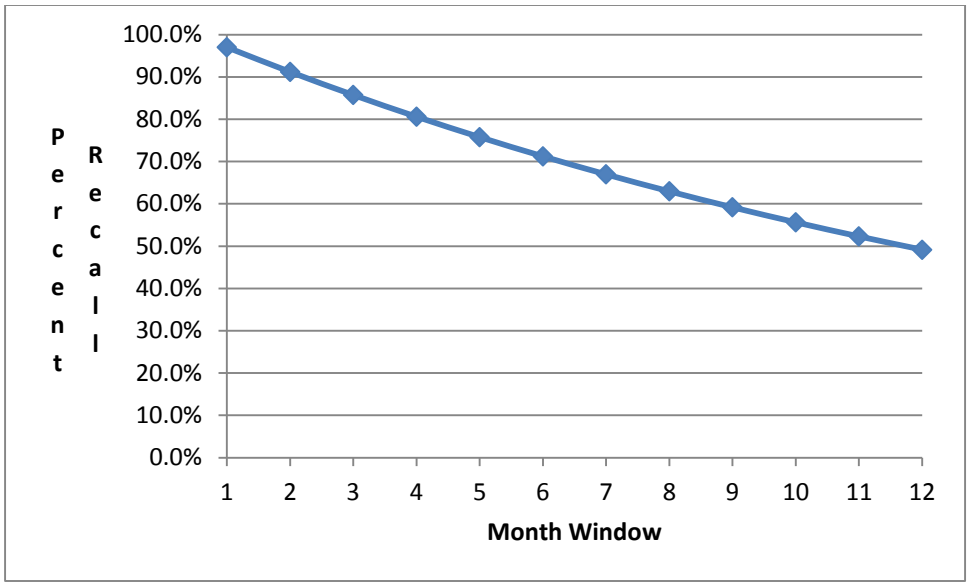

Figure 3-1 Percent expected recall by month window length assuming exponential decay model with monthly decay percentage 6.0%

The value of *RL* depends on the actual type of injury/illness being measured. The *RL* value of 6.0% is a mean value: serious injuries requiring hospitalization for example will have much lower *RL* values. Minor injuries which require no medical care and no lost days will have higher *RL* values. In addition, the NHIS question item on which the Warner et al. (2006) research was based allows for proxy answers (family members answering for other family members). The use of proxy information likely increases the *RL* value. If no proxy is utilized, the *RL* value may be in fact lower than 6.0%.

Figure 3-2 below assumes a exponential decay model with monthly decay percentage 4.5%: 25% smaller than the 6.0% decay rate. This is the assumed reduction in error which would come about from not allowing for proxies. We are certain that proxies will fail to remember about adults they are reporting for more often than the adults themselves, but have no empirical evidence on how much. We assume a 25% smaller rate in this cost-benefit analysis, but further research needs to be done to develop sound estimates. An experimental study is one strong way of developing such estimates (randomly sampling to experimental arms that allow proxies and do not allow proxies).

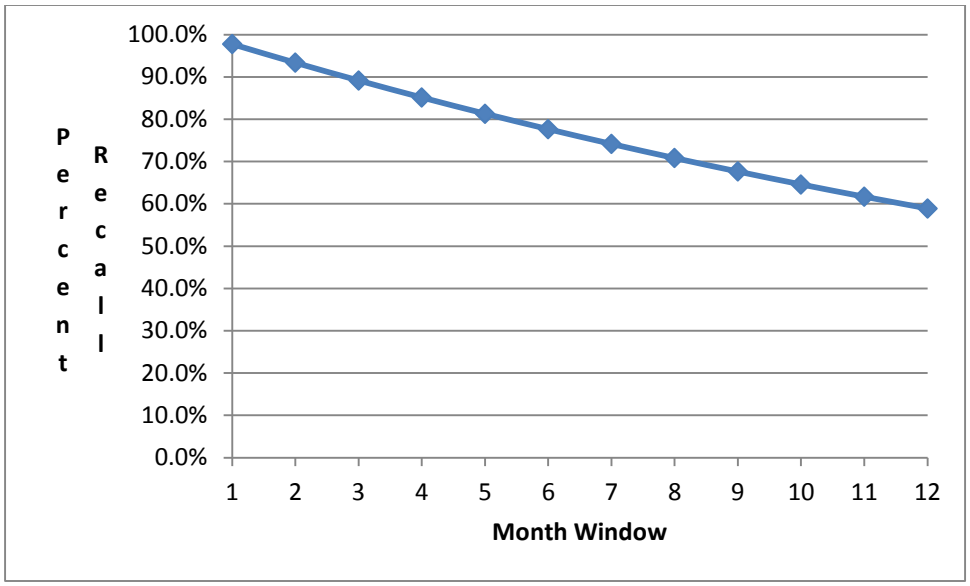

Figure 3-2 **Percent expected recall by month window length assuming exponential decay model with monthly decay percentage 4.5%**

3.2 Measurement Error from Concealment

Occupational injuries are something that employees tend in many cases to hide from their employers. Azaroff et. al (2002) discusses at great length the many reasons that employees might have for trying to conceal their injuries in some cases from their employers. These reasons include a fear of losing their job on the part of the employee. These fears are mediated by the presence of a union (which reduces the degree of concealment: for example Morse et al. (2003) and Hirsch et al. (1997)), the vulnerability of the workers (limited-English-proficiency workers, foreign-born workers: for example Forst et al. (2013) and Pransky et al. (2002)), and the policies of the employer which discourage injury reports (safety awards, etc.).

It should be noted that this is primarily an issue with the two-stage employer sampling options, which sample from employee lists provided by the employer, and begin the interview process from employer worksites. This should not be an issue with the household survey option if precautions are taken to ensure that the sampled household members know there is no chance that their volunteered data will be shared with their employer.

3.3 Survey Design Determinants of Response Rates

The response rate is deeply related to the degree of effort devoted to achieving high response rate. A carefully designed experiment was reported in Keeter et al.(2000). This study was a Pew Center study of political viewpoints. Sampled households were randomized into a group receiving a 'standard' followup and a group receiving a 'rigorous' followup. The standard followup group received one week of calls to reach the households, with a maximum one followup call for initial nonresponse. The rigorous followup received up to eight weeks of calls with exhaustive attempts at recontact and to follow up initial nonrespondents, and a small monetary incentive. The standard followup group had a response rate of 36.0 percent. The rigorous followup group had a response rate of 60.6 percent. The huge differences in response rates based on intensity with all other things being equal (the questionnaire, the sponsor, the population) indicates the enormous effect of intensity of followup on response outcomes. A wide range of literature in the past decades (e.g., Church (1993) for mail surveys; Singer et al. (1999) for telephone and face-to-face surveys) has indicated the positive effects on response rates from monetary incentives. Some of this literature reports on designed experiments, where the only difference is the presence or absence of an incentive, the type of incentive, and the magnitude of the cost. Again all other things are equal (questionnaire, sponsor, population, followup regime), simple differences in just the presence of and level of monetary incentives can make a sizeable difference in response outcomes.

The length of the questionnaire can also have an influence on response. A short interview length will achieve a higher response rate in general than a long interview, all other things being equal. The form of the questionnaire will also have some effect. A pleasing form will gain higher response than a confusing, unpleasant form. (See for example Dillman et al. (1993), Sahlqvist et al. (2011)). There are many influences on response that all contribute their effects.

The Methodology Group at Westat has published a new meta-analysis of the effect of monetary incentives on surveys, including also all of these other mediating effects (Mercer et al. (2015)). This meta-analysis includes 178 designed experiments in the effect of monetary incentives between 1992 and 2013. The new meta-analysis confirms the older literature in that incentives do increase response rates, and each increment of incentive increases response rate. A linear model between logarithm of incentive and response rate fits very well across the many designated experiments. The new meta-analysis also confirms the importance of mediating factors as discussed above. Among these important mediating factors are:

- Year of survey: later years have lower response rates all other things being equal;

- The mode (face-to-face, telephone, and mail):
  - Face-to-face has significantly higher response rates than mail and telephone among the 178 studies;
  - The leverage of incentives is stronger for mail surveys than the other two;
- Prepaid incentives vs. promised incentives:
  - Prepaid incentives (e.g., a $2 bill sent in the introductory letter) are more effective at each dollar value than promised incentives, all other things being equal;
  - This prepaid-promised difference disappears in face-to-face studies, and is strongest in mail studies;
- Government and university sponsors have higher response rates in these 178 studies than private sponsors, but with no interaction effects with the other characteristics (i.e., government surveys don't have a stronger or weaker incentive effect);
- High-burden studies (those with long questionnaires) have lower response rates than low-burden studies (short questionnaires), all other things being equal.

3.4 Response Propensity and R-Indicators

Nonresponse is a serious issue. There are methods for adjusting for nonresponse by weighting and analysis to avoid nonresponse bias, but these methods are imperfect. The general consensus still is that response rates should be as high as possible. Recent literature however has put forward a revised consensus that lower response rates are not as bad if there is not a great deal of differential nonresponse (Keeter et al. (2000), Groves (2006), Peytchev et al. (2009), Särndal and Lundquist (2014)). In principle, we can specify a response propensity for each sampled employee (Bethlehem and Kersten (1985), Oh and Scheuren (1983)). This response propensity is the underlying probability that the sampled employee responds if sampled. The mean value of these response propensities is then the response rate. In an extreme case, if the response propensity is exactly equal to 40% for each sampled person (and we somehow know that for certain), then nonresponse in this case will generate no bias, even though the response rate is only 40%. On the other extreme, we could have a situation with 80% of the sample having a response propensity of 100% and 20% of the population having a response propensity of 0%. The response rate will be 80%, but the nonresponse bias could be extreme. In the context of SOII, the 20% who are certain nonrespondents (0% propensity) may include most of the employees with occupational injuries and illnesses. In that case, the prevalence estimate will be badly biased even though the response rate is 80%. A high response rate is no

guarantee of small nonresponse bias, and a low response rate is not necessarily an unmitigated disaster. It is very important to know about and control the variation in response propensity. Another reference for this basic strategy is Schouten et al. (2009), who defines the concept of the 'R-indicator', which is basically the variation in response propensity. Minimizing the 'R-indicator' is more important than maximizing the response rate.

For the sake of prevalence estimates, it is important to assure ourselves that the response rate differential between employees who do have occupational injuries/illnesses and those who do not in the relevant time period is not too large. The design needs to be such that we can have this confidence. This will not be easy to achieve. For estimates within the population of employees with occupational injuries/illnesses, we need to assure ourselves that subgroups such as those with very serious injuries/illnesses and those with minor injuries/illnesses have similar response propensities. Otherwise there will be serious biases within the occupational injury/illness incident sample. Again, achieving this will require careful design.

General Methodological Issues 4

In this section, we discuss general methodological issues which cut across the basic design choice. Section 4.1 discusses the prospective vs. retrospective issues. Section 4.2 discusses the time window issue. Section 4.3 discusses benchmarking.

4.1 Prospective vs. Retrospective Studies

All but one of the studies which were described in the literature review are retrospective in nature: the respondents are questioned about the near past (three months up to the interview date; one year up to the interview date, etc.). This has the advantage of requiring only one interview: the respondent is interviewed about the past, and no further followup is necessary. An alternative which is sometimes used is the prospective (or panel) study. Huang et al. (2012) was a prospective study in the occupational safety and health field: this was a 12-week prospective study on injury in which restaurant workers reported weekly by telephone, internet, or writing about workplace injuries and hours worked in the past week.

Halbesleben (2010) is another example of a well-designed panel prospective study in a study of the relationship between exhaustion and burnout and occupational injuries among health care professionals. This study included two separate samples of health professionals (one of all full-time employees from a large teaching hospital and the other of registered nurses from a small community hospital). Surveys were done in 3 time intervals six months apart. Both health-care facilities were cooperative. Final response rates were 44% and 67% respectively, but the panel retention rates were quite good. Only 20% dropped out between Time 1 and Time 2, and only 1-2% between Time 2 and Time 3.

A third example is the New Immigrant Survey, which tracked new legal immigrants forward (for example Jasso et al. (2006)). Kalton et al. (1989) is an older general reference for panel surveys. The respondent is entered in the study at a particular point in time, and then followed into the future, and interviewed at particular future dates. This has the drawback of requiring multiple interviews, and also requiring followup, both of which can lead to nonresponse. Nonresponse can be reduced with extensive followup but is expensive. It has the advantage that at the point of initial recruitment,

the future is not known by anyone, so nonresponse at either the employer or employee level cannot be correlated to future injuries and illnesses. In practice, a study can be both retrospective and prospective in that questions can be asked about the near past, and then the respondent is followed into the future. Panel studies would allow this approach.

4.2 Time Window

The SOII employer study uses an annual time period as its time window. The prevalence rate is the number of incidents per year per 100 workers. Using this long a time period for the critical occupational injury/illness prevalence questions however will lead to recall and telescoping issues (see Section 3.1). Using a much shorter time window for the prevalence questions such as three months or less will lead to much more reliable and accurate answers, especially with regard to relatively limited injuries and illnesses, as has been found by Warner et al. (2005). But in order to cover the annual time window, much larger sample sizes will be needed. For example, if the time interval of the questions is two months, then the sample size will need to be six times larger to provide the same coverage as a corresponding survey which has a full year as a time interval. Table 4-1 presents the necessary interviews for with four different time windows (annual, three-month, two-month, one-month). The greater accuracy by making the time window short comes at a high price.

Table 4-1 Numbers of interviews to provide precision levels for each scenario with four different time windows

Scenario	Effective sample size auxiliary employee study	Necessary person sample size annual window	Necessary person sample size three-month window	Necessary person sample size two-month window	Necessary person sample size one-month window
1	1,310	1,310	5,240	7,860	15,720
2	1,675	1,675	6,700	10,050	20,100
3	2,250	2,250	9,000	13,500	27,000
4	3,220	3,220	12,880	19,320	38,640
5	5,100	5,100	20,400	30,600	61,200
6	9,570	9,570	38,280	57,420	114,840
7	25,660	25,660	102,640	153,960	307,920

Suppose for example we have specified a three-month window. Then to achieve the precision as specified for Scenario 5 (see Tables 2-1 and 2-2), a total of 20,400 interviews are needed. This can either be four interviews with 5,100 persons, or one interview with 20,400 persons. The former can be implemented through a prospective panel study with an initial interview and then three followup

interviews at three followup time points, and the latter can be implemented through a retrospective panel study (only a one-and-out interview with a single three-month window).

The issue of the length of the time period is intimately tied to the sample size, as 'person-years' have to be covered: not just persons. But this issue is relatively independent of many of the other issues such as household survey vs. employer-based survey, prospective vs. retrospective, etc. The time window is an issue with all of these options, and the concerns are similar under each of these options.

Table 4-2 presents a total of 21 scenarios which correspond to 61,200 person-months covered (5,100 person years). This is Scenario 5 from Table 4-1. The 'A' scenarios cover these person-months with interviews that cover only one month. This minimizes the measurement error. The 'B' scenarios cover these person-months with interviews that cover a two-month time window and so forth (the 'C' scenarios have a three-month time window, the 'D' scenarios a four-month time window, the 'E' scenarios a six-month time window, and the 'F' scenarios a twelve-month time window).

The sub-scenarios within 'A' through 'F' vary the number of sampled persons and the number of interviews. The '1' subscenarios have only one interview per person (a retrospective interview, or a single prospective interview). The '2' subscenarios have two interviews per person (two prospective interviews), and so forth. In the case of the A-6 scenario, there are 12 interviews of a single person.

Table 4-2 includes for each scenario the persons interviewed, the interviews per person, the total interviews, and the measurement error bias assuming the exponential decay model with a monthly decay rate of 4.5% and 6% (see Section 2.1). The measurement error bias is 1 minus the expected recall rate from Figure 3-1, and represents the aggregate percentage of 'lost' injury/illness reports. The 4.5% rate represents a reasonable decay rate assuming no proxies, and 6.0% a reasonable decay rate with proxies. (The 6.0% rate is derived from the Warner et al. (2005) results, which do include a proxy: a family member is asked to answer for all family members. So the expected decay rate with no proxy can be expected to be a bit lower.)

Table 4-2 Numbers of interviews to provide precision levels for each scenario with four different time windows

Scenario	Months covered for interview	Persons interviewed	Interviews per person	Person-months covered	Total interviews	Measurement error bias monthly decay 4.5%	Measurement error bias monthly decay 6%
A-1	1	61,200	1	61,200	61,200	2.3%	3.0%
A-2	1	30,600	2	61,200	61,200	2.3%	3.0%
A-3	1	20,400	3	61,200	61,200	2.3%	3.0%
A-4	1	15,300	4	61,200	61,200	2.3%	3.0%
A-5	1	10,200	6	61,200	61,200	2.3%	3.0%
A-6	1	5,100	12	61,200	61,200	2.3%	3.0%
B-1	2	30,600	1	61,200	30,600	6.6%	8.8%
B-2	2	15,300	2	61,200	30,600	6.6%	8.8%
B-3	2	10,200	3	61,200	30,600	6.6%	8.8%
B-4	2	7,650	4	61,200	30,600	6.6%	8.8%
B-5	2	5,100	6	61,200	30,600	6.6%	8.8%
C-1	3	20,400	1	61,200	20,400	10.8%	14.3%
C-2	3	10,200	2	61,200	20,400	10.8%	14.3%
C-3	3	6,800	3	61,200	20,400	10.8%	14.3%
C-4	3	5,100	4	61,200	20,400	10.8%	14.3%
D-1	4	15,300	1	61,200	15,300	14.9%	19.4%
D-2	4	7,650	2	61,200	15,300	14.9%	19.4%
D-3	4	5,100	3	61,200	15,300	14.9%	19.4%
E-1	6	10,200	1	61,200	10,200	22.4%	28.8%
E-2	6	5,100	2	61,200	10,200	22.4%	28.8%
F-1	12	5,100	1	61,200	5,100	41.1%	50.9%

Multiple interviews with a single person require followup. As is well-known, response can drop away with each followup. In most cases, response in panel studies tends to be monotone in nature (response up to a certain point, and then non-response from then on out: if they drop out for any followup they do not return), but there can also be 'swiss-cheese' patterns of response-nonresponse-response. This tends to be rarer. See for example Kalton et al.(1989).

We posit for now 100% response for the first followup (or the single interview, for one-interview scenarios). Nonresponse to the first interview is discussed in the later sections that separate out by household and employer studies, so here we condition on first interview cooperation. When there is no proxy, we assume a 10% dropoff for the first follow-on prevalence interview (the second interview) and further 5% drop-offs for each further follow-on prevalence interview (an exponential decay). When there is a proxy allowed (a family member who can answer for any family member), we assume a 7.5% dropoff for the first follow-on prevalence interview (the second interview) and further 3.75% drop-offs for each further follow-on prevalence interview (an exponential decay). The attrition rates when proxies are allowed for are assumed to be 25% lower. This is entirely a

presupposition at this point: we don't have any hard evidence from other related surveys. Figure 4-1 presents the expected response rates in graphical form.

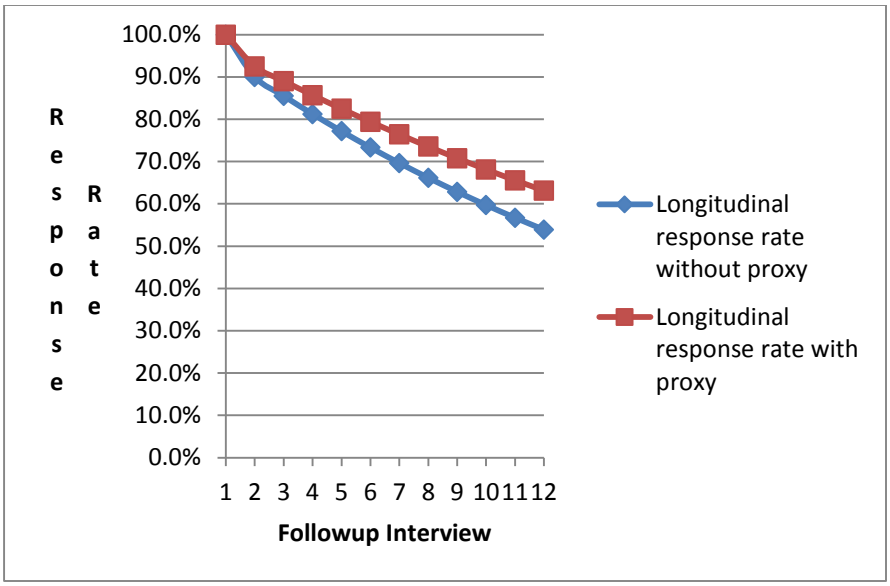

Figure 4-1 **Percent expected response rate for followup interviews by month assuming exponential decay model with 5.0% decay per interview (except for the second at 10%).**

In practice in panel studies, one does not discard a sample person because they do not cooperate with all of the followups assigned, and in this study one would not do that as well. The interviews collected are of value even if the sample person falls out and does not respond to later followup interviews. The critical thing is the monotone response pattern: that the interviews are coherent and correct up to the point of breakoff. Swiss-cheese patterns are more problematic as the interviews after a missing interview will need to be processed more carefully, but even here with sufficient analysis the later interviews after a missed interview(s) should be usable. Weighting adjustments need to be made to adjust each followup interview to the particular response rate for that followup interview.

Table 4-3 presents expected numbers of collected interviews assuming a completely monotone response pattern with the response rates as given in Figure 4-1. For example, for two-interview followup scenarios (A-2, ….,E-2), 90% are expected for the second followup without proxies (92.5% with proxies). For the cooperating sample without proxies (cooperating with at least the first interview), we have 100% first interviews and 90% second interviews. There are two possible interviews, and the expected yield is 1.90 total interviews, for a 95% cooperation rate. With proxies, the expected yield is 1.93, for a 96.25% cooperation rate. For the three-interview followup scenarios

(A-3,…,D-3) without proxies, 91.83% are expected to complete all three interviews, with 95% completing two, and 100% completing one. This is an aggregate rate of 2.76 interviews per sampled person (without proxies). For the twelve-interview scenario A-6 without proxies, the expectation from the exponential decay model is that an aggregate rate of 8.76 interviews will be completed out of the potential total of 12 (a 73.0% completion rate). With proxies, an aggregate rate of 9.47 interviews is expected (a 78.9% completion rate).

Table 4-3 Expected yield of interviews by followup scenario

Scenarios	Total potential interviews per sample person	Expected interviews per sample person w/o proxy	Expected interviews per sample person with proxy	Expected interview aggregate cooper-ation rate w/o proxy	Expected interview aggregate cooper-ation rate with proxy
A-1, B-1, C-1, D-1, E-1, F-1	1	1.00	1.00	100.00%	100.00%
A-2, B-2, C-2, D-2, E-2	2	1.90	1.93	95.00%	96.25%
A-3, B-3, C-3, D-3	3	2.76	2.82	91.83%	93.84%
A-4, B-4, C-4	4	3.57	3.67	89.18%	91.81%
A-5, B-5	6	5.07	5.29	84.53%	88.18%
A-6	12	8.76	9.47	73.01%	78.89%

The correct approach would be to also adjust the person-level sample sizes to allow for this expected attrition (and possibly also for the design effects from nonresponse weighting adjustments as well, but we will not do that here). For example, if longitudinal nonresponse is ignored, Scenario A-2 without proxies requires 30,600 sample persons to complete 61,200 interviews. If we adjust for loss from 5% longitudinal nonresponse (reducing the number of interviews by 5%), then the sample person count should be 30,600/1.95, or 32,211.

Table 4-4-1 presents the number of persons required who complete at least one interview, as well as expected interviews and total interviews, not allowing proxies. The measurement error bias with a monthly decay rate of 4.5% in recall is also presented. The numbers are such that the total interviews is equal to 61,200 divided by the month window. Table 4-4-2 presents the same expected interviews, but this time allowing for proxies. The measurement error bias with a monthly decay rate of 6.0% in recall is also presented.

Table 4-4-1 Numbers of interviews to provide precision levels for each scenario allowing for longitudinal nonresponse (without proxies)

Scenario	Months covered for interview	Target persons interviewed	Target interviews per person	Expected interview aggregate cooperation rate	Expected interviews per sampled person	Total interviews	Measurement error bias monthly decay rate 4.5%
A-1	1	61,200	1	100.00%	1.00	61,200	2.3%
A-2	1	32,211	2	95.00%	1.90	61,200	2.3%
A-3	1	22,214	3	91.83%	2.76	61,200	2.3%
A-4	1	17,156	4	89.18%	3.57	61,200	2.3%
A-5	1	12,066	6	84.53%	5.07	61,200	2.3%
A-6	1	6,985	12	73.01%	8.76	61,200	2.3%
B-1	2	30,600	1	100.00%	1.00	30,600	6.6%
B-2	2	16,105	2	95.00%	1.90	30,600	6.6%
B-3	2	11,107	3	91.83%	2.76	30,600	6.6%
B-4	2	8,578	4	89.18%	3.57	30,600	6.6%
B-5	2	6,033	6	84.53%	5.07	30,600	6.6%
C-1	3	20,400	1	100.00%	1.00	20,400	10.8%
C-2	3	10,737	2	95.00%	1.90	20,400	10.8%
C-3	3	7,405	3	91.83%	2.76	20,400	10.8%
C-4	3	5,719	4	89.18%	3.57	20,400	10.8%
D-1	4	15,300	1	100.00%	1.00	15,300	14.9%
D-2	4	8,053	2	95.00%	1.90	15,300	14.9%
D-3	4	5,554	3	91.83%	2.76	15,300	14.9%
E-1	6	10,200	1	100.00%	1.00	10,200	22.4%
E-2	6	5,368	2	95.00%	1.90	10,200	22.4%
F-1	12	5,100	1	100.00%	1.00	5,100	41.1%

Table 4-4-2 Numbers of interviews to provide precision levels for each scenario allowing for longitudinal nonresponse (with proxies)

Scenario	Months covered for interview	Target persons interviewed	Target interviews per person	Expected interview aggregate cooperation rate	Expected interviews per sampled person	Total interviews	Measurement error bias monthly decay rate 4.5%
A-1	1	61,200	1	100.00%	1.00	61,200	3.0%
A-2	1	31,792	2	96.25%	1.93	61,200	3.0%
A-3	1	21,738	3	93.84%	2.82	61,200	3.0%
A-4	1	16,666	4	91.81%	3.67	61,200	3.0%
A-5	1	11,567	6	88.18%	5.29	61,200	3.0%
A-6	1	6,465	12	78.89%	9.47	61,200	3.0%
B-1	2	30,600	1	100.00%	1.00	30,600	8.8%
B-2	2	15,896	2	96.25%	1.93	30,600	8.8%
B-3	2	10,869	3	93.84%	2.82	30,600	8.8%
B-4	2	8,333	4	91.81%	3.67	30,600	8.8%
B-5	2	5,784	6	88.18%	5.29	30,600	8.8%
C-1	3	20,400	1	100.00%	1.00	20,400	14.3%
C-2	3	10,597	2	96.25%	1.93	20,400	14.3%
C-3	3	7,246	3	93.84%	2.82	20,400	14.3%
C-4	3	5,555	4	91.81%	3.67	20,400	14.3%
D-1	4	15,300	1	100.00%	1.00	15,300	19.4%
D-2	4	7,948	2	96.25%	1.93	15,300	19.4%
D-3	4	5,435	3	93.84%	2.82	15,300	19.4%
E-1	6	10,200	1	100.00%	1.00	10,200	28.8%
E-2	6	5,299	2	96.25%	1.93	10,200	28.8%
F-1	12	5,100	1	100.00%	1.00	5,100	50.9%

Figure 4-2-1 and 4-2-2 present a summarization of the 21 scenarios in terms of their cost and bias properties. These figures are derived from the numbers in Table 4-4-1, and show how the different scenarios measure up as far as cost (interviews and interviewed persons), and bias (recall measurement error bias and longitudinal nonresponse). In both tables 'low is good' and 'high is bad': the scenarios in the lower left hand corner have the best cost properties in Figure 4-2-1 and the scenarios in the lower left hand corner of Figure 4-2-2 have the best bias properties.

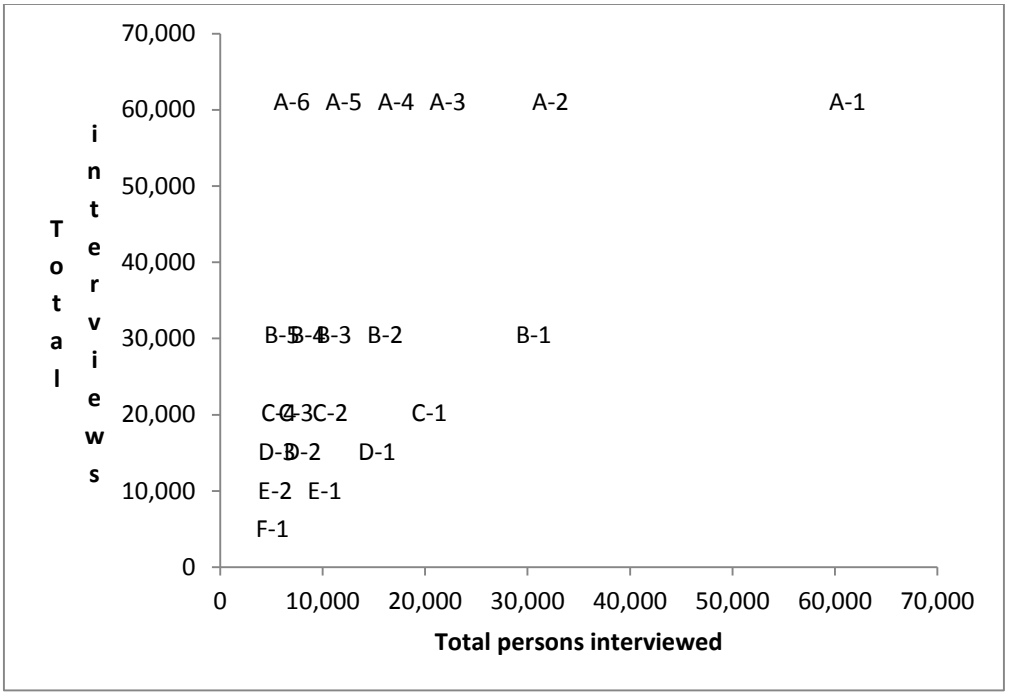

Figure 4-2-1 Interviewed persons vs. interviews for the 21 scenarios (with no proxy)

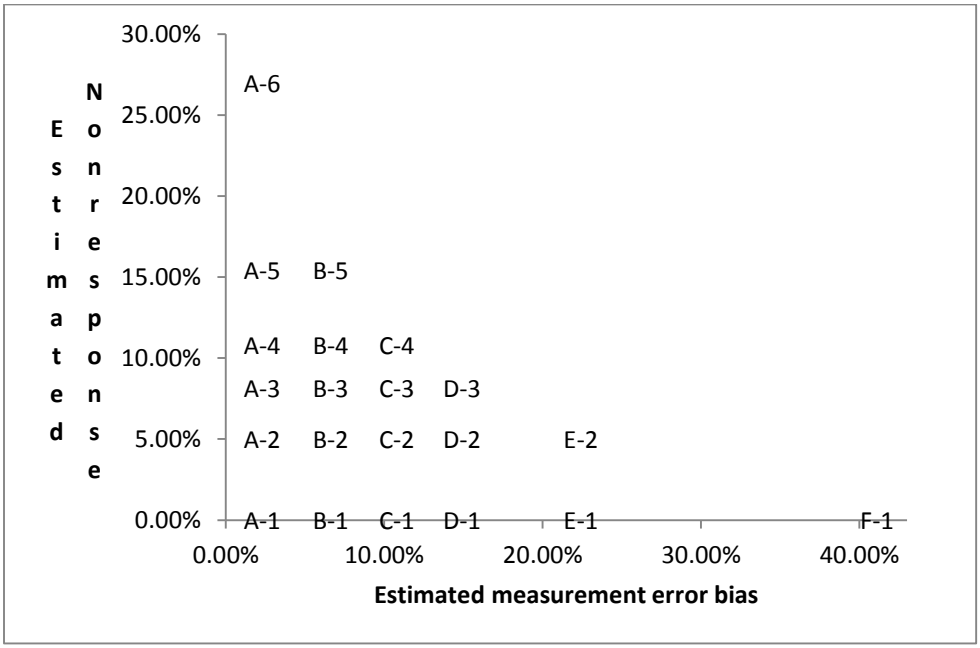

Figure 4-2-2 Estimated recall measurement error bias vs. longitudinal interview nonresponse (with no proxy)

As Tables 4-4-1 and 4-4-2 and Figures 4-2-1 and 4-2-2 show, there is a complex trade-off with these scenarios. The 'A' scenarios have the lowest measurement error, but they either have a very large number of interviews (A-1), or they have a high nonresponse rate with the followup interviews (or a high cost to minimize the longitudinal nonresponse). The 'F' scenario on the other hand has a low number of interviews and a no longitudinal nonresponse, but it has a terrible measurement error bias potential. There are no obvious winners, only tradeoffs.

There is also a tradeoff between allowing proxies and not allowing proxies. Allowing proxies the longitudinal response rate is better (reducing sample size needs as well), but the potential for measurement error bias is worse. Not allowing proxies, measurement error bias is improved, but the longitudinal response rate is degraded, increasing the necessary sample sizes. The actual magnitude of these effects is speculative at this point, but these are the right qualitative effects.

4.3 Clustering

The two basic approaches discussed below are both two-stage samples. In the household survey case, households are selected first (households are the primary sample units, or PSUs), and then adults within the household. In the employer-based study case, establishments are selected as the PSU, and then employees within each establishment. The two stage approach means that the effective sample size will be less than the nominal sample size in general. Only cases where the characteristics of interest have intra-PSU correlations of zero will have an effective sample size equal to the nominal sample size.

If we have a simple random sample size of m primary sampling units (PSUs) and a mean value of $\bar{n}$ second-stage sampling units (SSUs) per PSU (the total number of sampled SSUs is $m\bar{n}$), then an approximate variance for the sample mean $\bar{y}$ is

$$Var(\bar{y}) = \frac{\sigma_y^2}{m\bar{n}}(1 + \rho\{\bar{n} - 1\})$$

ρ is the intra-PSU correlation coefficient. See for example Hansen, Hurwitz, and Madow (1953), Chapter 6, Eq (8.1). This ignores finite population corrections, and weighting effects. The design effect (the ratio of the sampling variance of the sample mean under the design to the sample mean of a simple random sample design with the same ultimate sample size) is equal to $1 + \rho\{\bar{n} - 1\}$.

If the y-statistic estimand is actually prevalence of an occupational injury or illness in the past year, then a household survey will show minimal intra-PSU correlation. Many households will have only one eligible worker, and for those with multiple eligible workers, it will be highly unlikely for more than one of these workers to have an incident in the past year. The correlation coefficient ρ may not be zero for these households, but it will tend to be very close to zero. If only one eligible adult is sampled in each sampled household, then $\bar{n}$ will be 1 and the design effect will be 1 (no difference with a simple random sample of eligible adults from the general population). If more than one eligible adult is sampled within households with multiple eligible workers, then the design effect will be greater than 1, but $\bar{n}$ will still be close to 1 and ρ will be close to 0, so the design effect will not be much greater than 1.

An employer-based survey is likely to show a larger design effect. In this case, the PSUs are the employers, and the SSUs are employees. The intra-PSU correlation coefficient ρ is likely to be bigger than 1 in general, though if employers are stratified by industry groups, with relative prevalence of occupational injuries/illnesses determining the industry groups (very high prevalence industries, etc.), then within each of these strata the intra-PSU correlation coefficient ρ will be much lower (the variation among employers in prevalence within a strata is less than that over all establishments in all industries). But it will still be greater than zero: individual employers will differ in prevalence levels even in the same industry (based on variations in their workplace environments).

The second aspect of clustering is the cost aspect. Most cost analyses of clustering (for example Valliant et al. (2013), Section 9.3.1) utilize a linear function for cost as follows:

$$C = C_0 + mc_1 + m\bar{n}c_2$$

where C_0 is an overall overhead cost, c_1 is the cost of fielding a PSU, and c_2 is the cost of recruiting, interviewing, and following up single SSUs. In most two-stage studies, the cost c_1 is many times the cost of c_2. For example, in school achievement testing studies, the cost of recruiting a school and sending a field person to carry out the test is high, with the follow-on cost of giving the test to each student not very large.

Under these assumptions, the optimal sample mean SSU sample size per PSU $\bar{n}$ is

$$\bar{n}_{opt} = \sqrt{\frac{c_1}{c_2}\frac{1-\rho}{\rho}}$$

(Equation 9.24 in Valliant et al. (2013).) Table 4-5 below presents the optimal $\bar{n}_{opt}$ values for intra-PSU correlations ρ ranging from 1.0% to 10.0%, and cost ratios c_1/c_2 ranging from 1.0 to 100. This covers a wide range of possible correlations and cost ratios. Correlation coefficients will vary across questionnaire items (for example, in households race/ethnicity will be fairly constant within the household and have a high correlation but gender may actually have a negative correlation). Relevant correlation coefficients for household surveys can be computed from a survey such as NHIS. The survey should be national, have multiple sample persons within a household, and provide a household indicator on a public-use file. For employer surveys, the BLS SOII employer survey can be used to generate employer correlation coefficients for incidence (by computing incidence percentages for individual employers, and finding the variance component). This should be done within industries separately, as employer prevalence rates should vary across industries.

For cost ratios, the household survey cost structure will vary by the type of survey. Telephone surveys will have considerably lower cost ratios than in-person surveys. For a telephone survey, the cost ratio should be fairly small, as contacting the household and interviewing the adult will be roughly comparable costs (e.g., one person-hour for each). The actual cost ratio depends on many factors, such as the intensity of the followup (average person-hours to reach a household and do the screener) and the length of the screening, prevalence, and incidence interviews.

Table 4-5 Optimal SSU per PSU mean sample size values for varying cost ratios and intra-PSU correlation coefficients

Intra-PSU correlation	Cost ratio										
	1.0	1.5	2	4	6	8	10	15	20	50	100
1.0%	9.9	12.2	14.1	19.9	22.2	24.4	31.5	38.5	44.5	70.4	99.5
2.5%	6.2	7.6	8.8	12.5	14.0	15.3	19.7	24.2	27.9	44.2	62.4
5.0%	4.4	5.3	6.2	8.7	9.7	10.7	13.8	16.9	19.5	30.8	43.6
7.5%	3.5	4.3	5.0	7.0	7.9	8.6	11.1	13.6	15.7	24.8	35.1
10.0%	3.0	3.7	4.2	6.0	6.7	7.3	9.5	11.6	13.4	21.2	30.0

4.4 Interviewing for Prevalence vs. Incidence

In most cases, the worker will not have had an occupational injury/illness in the time window in question. For these workers, a very short 'prevalence' interview will generally be adequate, just collecting information about how much they were off work, how much on reduced schedules due to occupational injuries or illnesses, etc., to decide or not whether there was an incident or not. If there was an incident, then the interview is much longer to collect information about the nature of the

injury or illness. It may be necessary to have a followup 'incident' interview by another interviewer who is trained to carry out the more extensive interview. The 'prevalence interview' should be relatively short. The more extensive 'incident' interview is only carried out on the small percentage of employees who do have incidents in the time window.

The distinction between prevalence and incidence is also of relevance in allowing for proxy interviews (interviews by persons who know well the information about the sampled persons, such as a spouse). We speculate that the proxy interviews may be realistic for the prevalence interview, but not as much for the incidence interview, as the prevalence interview requires only knowledge about whether there was an occupational injury/illness incident or not for an adult in the household, whereas the incidence interview requires details about the incident. The injury/poisoning question for NHIS which was studied in Warner (2005) allows for proxies. The Warner (2005) results includes the proxy responses allowed for by the NHIS questionnaire on this item. We don't know what would happen under the counterfactual that the injury/poisoning question is asked in the adult questionnaire and is directed only at the sampled adult. That research would be greatly beneficial if it existed. To get a clear answer on this issue, we would recommend a designed experiment to check the effects of allowing for proxies or not in the pilot study or in the early years of a main study. This would give a clear answer to this very important question.

The Option of Household Surveys 5

A nationally representative household survey in which households with at least one worker in the targeted industries is one option. Screening would be necessary to subset to the households with at least one worker in private industry (excluding mining and other out-of-scope industries) or state and local governments. From the American Community Survey, we know that 85% of US families have at least one worker[9]. There is some extra screening cost to needing to exclude households with no eligible workers, but this is limited (we are assuming below that only 20% or so of households will be screened as being ineligible).

We discuss three variants of household surveys below: a stand-alone targeted nationally representative household study for occupational injuries and illnesses, having the study as a module in a larger household study such as the National Health Interview Study, and the closely related option of having the study as a follow on to a larger household study.

5.1 A Targeted Nationally Representative Household Study

Under this plan, a specialized national sample of households is drawn using a technique such as ABS (Address Based Sampling). There are a variety of data collection methods possible, with even mixed modes. The gold standard would be a face-to-face interview in the sampled households. The response rates and quality will be the highest under this mode. The data collection costs per interview will be many times higher than that for alternative modes (see for example Brick and Kalton (2007)). In the 1990s, the major alternative to face-to-face interviews would have been a Random Digit Dialing (RDD) Survey: telephone data collection. But due to the onset of cellphones with their special issues and the collapsing contact and response rates, telephone surveys are becoming less and less common as the method of choice. One can cite for example Dillman et al. (2014), p. 11 or Brick and Williams (2013).

[9] This includes workers in federal government and mining, but these should be relatively small percentages (no more than an extra 5%). This is from Table S2302 from American Fact Finder from the 2013 ACS 1-Year Estimates.

The best approach in the current environment for this plan, given the relative prevalence expected and the budget we believe will be available, is a mail survey. Any followups will also be by mail, though the household can be invited to respond by internet if they are inclined in that direction. See for example Brick et al. (2015) for an argument for a single-stage mail design. It is also possible to carry out a two-stage process (e.g., Montaquila et al. (2013)), separating out screening surveys from prevalence and incidence surveys, but we believe a single-stage mail collection for each prevalence check might work well in this instance. Mail surveys now have higher response rates than telephone surveys if the mail surveys are done well, and are much less costly by a factor of between 2 and 5 per respondent (Brick and Kalton (2007)). The only potential advantage of a telephone mode as opposed to a mail mode is dealing with languages other than English and Spanish in areas where there are many spoken languages, so that interviewers in those languages can be utilized. The best that can generally be done with a mail survey is to send the forms in English and Spanish (e.g., Montaquila et al. (2013)).

The full eligible household set can only be used to actually measure prevalence: the percentage of workers who have had an occupational injury/illness during the specified time window (e.g., the past one month, three months, one year). Specific questions for those who indicate an occupational injury/illness in the specified time window can only be addressed to those respondents. This will be a much smaller percentage of households who have at least one person with an occupational injury/illness. Suppose we assume the incidence rate of 3.5 occupational injuries/illnesses per 100 workers per year (see above), and disregard as an approximation the issues of persons who have multiple incidents and households with more than one worker. If the time window is a full year (which is subject to fairly sizeable recall and telescoping measurement error issues), then we can expect 3.5% of households will yield persons who can answer these questions. If the time window is only three months (much better in terms of measurement error), then we can expect only 0.875% of households will yield persons who had incidents in the time window.

If the survey and the questions are clearly about occupational injury and illness in particular then there will be some chance that the respondents may be on their guard, and some measurement error from concealment will result. The questionnaire design can address this at least partially by including more content related to other occupational issues rather than only injury and illness, but care must be taken to keep the length within reason. Furthermore, it is clear that it will be easier to reassure the respondent that the information that they provide will not be shared with the present employer or any potential future employer if the study is a household study rather than an employer-based study. One variant of this option is to allow one member of the household to speak for occupational injuries/illnesses for all working adults in the household. This would allow for a greater adult-level

sample size for a given number of sampled households. This option would allow for taking all eligible adults within a household (eliminating within household sampling for a mail survey as discussed below). But there would be sometimes some inaccuracy in the proxy report, especially for more limited injuries/illnesses. This is a tradeoff.

5.1.1 Household Study Option: Within-Household Sampling

The next step in the stand-alone study is to sample eligible adults. In some cases, the eligible household will have only one eligible worker. The American Community Survey show that 15% of families have had no eligible workers in the past 12 months, 34% have exactly one eligible worker, and 40% have exactly 2 eligible workers, and 11% have three or more eligible workers[10]. In this case, this worker is automatically the sampled adult. If there are multiple eligible workers in the household, then there are several options. Stand-alone Household Option 1 is to sample up to one eligible worker. Stand-alone Household Option 2 is to sample up to two eligible workers[11]. Sub-options of both options would allow proxy interviews for the prevalence interview. These proxy interviews would be collected from other adults in the household (eligible or not) about the sampled eligible adults. This will increase the resultant response rate as there are more options for getting a completed interview. It will also lead to some measurement error, as a proxy will not know or remember as well as the eligible adult themselves. However, if the proxy option is restricted to the prevalence interview we speculate that then the measurement error may not be that great, as the proxy is only required to have accurate information about whether or not an incident has occurred for the subject adult, and not to be knowledgeable about details of it.

Our recommendation would be for Option 2, but with a mail survey it might be possible to take a slight extension of sampling all eligible adults in the household if proxies are allowed as discussed above. Generally Option 1 is a good option if the intra-household correlation is expected to be relatively high, so that getting multiple interviews within the household does not add that much information. But in this case, the prevalence incidence itself (whether or not the eligible worker has had an occupational injury/illness in the specified time window, and how many incidences) should have very low within-household correlation. The fact that one adult in the household has had an incidence shouldn't increase or decrease the likelihood other adults have had incidences, unless they

[10] From Table S2302 from American FactFinder: 2013 American Community Survey 1-Year Estimates, and from a downloaded table from Data Ferret from the 2013 Public Use Microdata Set for Workers in Families.

[11] It is possible to have options to take more eligible workers than two, but the number of households this will apply to will be limited, so the properties will not be much different from Option 2.

happen to be working for the same company in the same occupational area. This very low expected correlation makes it a good option to interview as many eligible adults as possible once the household is recruited. Thus Option 1 is not the optimal approach in this particular context.

Table 5-1 below presents results for Option 1 and Option 2 with a 1% correlation. Both options are presented with and without proxies. The table presents necessary household sample sizes under these options to achieve effective adult interview sample sizes of 5,100, making the following assumptions:

- Household response rates equal to 50%[12];
- Household eligibility rates equal to 80% (eligible in terms of having at least one eligible working adult);
- Adult first-interview[13] response rates of 75%, with response rates with proxies allowed equal to 85%;
- Mean sampled adults of 1.6 for Option 2 (one sampled adult in households with one eligible adult; two sampled adults in households with two or more eligible adults[14]);
- For Option 2, a within-household correlation of 1% is assumed.

The design effect for Option 2 is computed using the formula $1 + \rho\{\bar{n} - 1\}$ with the ρ of 1% and an $\bar{n}$ of 1.6.

[12] This assumes a fairly rigorous followup of nonrespondents (a field period of several months, with refusal conversion efforts), and a relatively short prevalence interview (10-15 minutes).

[13] I.e.,, the response rate for the first interview before longitudinal attrition.

[14] This assumes that 40% of households with at least one eligible adult will have exactly one eligible adult, and the others will have more than one eligible adult.

Table 5-1 Household sample sizes to achieve effective adult interview sample sizes of 5,100

	Option 1 no proxies	Option 1 with proxies	Option 2 1% ρ no proxies	Option 2 1% ρ with proxies
Household sample size	17,000	15,000	10,689	9,431
Household response rate	50%	50%	50%	50%
Household screener size	8,500	7,500	5,345	4,716
Household eligibility rate	80%	80%	80%	80%
Eligible households	6,800	6,000	4,276	3,773
Mean sampled adults/eligible HHs	1	1	1.60	1.60
Sampled adults	6,800	6,000	6,841	6,036
Adult first-interview response rate	75%	85%	75%	85%
Adult prevalence interviews	5,100	5,100	5,131	5,131
Within household correlation	NA	NA	0.01	0.01
Design effect	1	1	1.006	1.006
Effective sample size	5,100	5,100	5,100	5,100

Tables 5-2-1 through 5-2-4 present the required household sample sizes under the four options (Option 1 no Proxy; Option 1With Proxy; Option 2 No Proxy 1% rho; Option 2 With Proxy 1% rho) for the 21 scenarios as given in Section 4.2, utilizing the assumptions about response rates, etc., as given in Table 5-1. Note that the Table 5-1 figures correspond to the Scenario F-1 row of Tables 5-2-1 through 5-2-4.

Table 5-2-1 Household, person, and interview sample sizes for Household Option 1 No Proxy

Scenario	Household sample size	HH response and eligibility rate	Responding households	Sampled adults	Adult first interview response rate	Target persons interviewed	Expected interviews per person	Total interviews
A-1	204,000	40%	81,600	81,600	75%	61,200	1.00	61,200
A-2	107,368	40%	42,947	42,947	75%	32,211	1.90	61,200
A-3	74,047	40%	29,619	29,619	75%	22,214	2.76	61,200
A-4	57,187	40%	22,875	22,875	75%	17,156	3.57	61,200
A-5	40,221	40%	16,089	16,089	75%	12,066	5.07	61,200
A-6	23,283	40%	9,313	9,313	75%	6,985	8.76	61,200
B-1	102,000	40%	40,800	40,800	75%	30,600	1.00	30,600
B-2	53,684	40%	21,474	21,474	75%	16,105	1.90	30,600
B-3	37,024	40%	14,809	14,809	75%	11,107	2.76	30,600
B-4	28,593	40%	11,437	11,437	75%	8,578	3.57	30,600
B-5	20,111	40%	8,044	8,044	75%	6,033	5.07	30,600
C-1	68,000	40%	27,200	27,200	75%	20,400	1.00	20,400
C-2	35,789	40%	14,316	14,316	75%	10,737	1.90	20,400
C-3	24,682	40%	9,873	9,873	75%	7,405	2.76	20,400
C-4	19,062	40%	7,625	7,625	75%	5,719	3.57	20,400
D-1	51,000	40%	20,400	20,400	75%	15,300	1.00	15,300
D-2	26,842	40%	10,737	10,737	75%	8,053	1.90	15,300
D-3	18,512	40%	7,405	7,405	75%	5,554	2.76	15,300
E-1	34,000	40%	13,600	13,600	75%	10,200	1.00	10,200
E-2	17,895	40%	7,158	7,158	75%	5,368	1.90	10,200
F-1	17,000	40%	6,800	6,800	75%	5,100	1.00	5,100

Table 5-2-2 Household, person, and interview sample sizes for Household Option 1 With Proxy

Scenario	Household sample size	HH response and eligibility rate	Responding households	Sampled adults	Adult first interview response rate	Target persons interviewed	Expected interviews per person	Total interviews
A-1	204,000	40%	81,600	81,600	75%	61,200	1.00	61,200
A-2	105,974	40%	42,390	42,390	75%	31,792	1.93	61,200
A-3	72,461	40%	28,984	28,984	75%	21,738	2.82	61,200
A-4	55,552	40%	22,221	22,221	75%	16,666	3.67	61,200
A-5	38,557	40%	15,423	15,423	75%	11,567	5.29	61,200
A-6	21,550	40%	8,620	8,620	75%	6,465	9.47	61,200
B-1	102,000	40%	40,800	40,800	75%	30,600	1.00	30,600
B-2	52,987	40%	21,195	21,195	75%	15,896	1.93	30,600
B-3	36,230	40%	14,492	14,492	75%	10,869	2.82	30,600
B-4	27,776	40%	11,110	11,110	75%	8,333	3.67	30,600
B-5	19,278	40%	7,711	7,711	75%	5,784	5.29	30,600
C-1	68,000	40%	27,200	27,200	75%	20,400	1.00	20,400
C-2	35,325	40%	14,130	14,130	75%	10,597	1.93	20,400
C-3	24,154	40%	9,661	9,661	75%	7,246	2.82	20,400
C-4	18,517	40%	7,407	7,407	75%	5,555	3.67	20,400
D-1	51,000	40%	20,400	20,400	75%	15,300	1.00	15,300
D-2	26,494	40%	10,597	10,597	75%	7,948	1.93	15,300
D-3	18,115	40%	7,246	7,246	75%	5,435	2.82	15,300
E-1	34,000	40%	13,600	13,600	75%	10,200	1.00	10,200
E-2	17,662	40%	7,065	7,065	75%	5,299	1.93	10,200
F-1	17,000	40%	6,800	6,800	75%	5,100	1.00	5,100

Table 5-2-3 Household, person, and interview sample sizes for Household Option 2 No Proxy

Scenario	Household sample size	HH response and eligibility rate	Responding households	Sampled adults	Adult first interview response rate	Target persons interviewed	Expected interviews per person	Total interviews
A-1	128,265	40%	51,306	82,090	75%	61,567	1.00	61,567
A-2	67,508	40%	27,003	43,205	75%	32,404	1.90	61,567
A-3	46,557	40%	18,623	29,797	75%	22,347	2.76	61,567
A-4	35,956	40%	14,383	23,012	75%	17,259	3.57	61,567
A-5	25,289	40%	10,116	16,185	75%	12,139	5.07	61,567
A-6	14,639	40%	5,856	9,369	75%	7,027	8.76	61,567
B-1	64,133	40%	25,653	41,045	75%	30,784	1.00	30,784
B-2	33,754	40%	13,502	21,603	75%	16,202	1.90	30,784
B-3	23,279	40%	9,311	14,898	75%	11,174	2.76	30,784
B-4	17,978	40%	7,191	11,506	75%	8,630	3.57	30,784
B-5	12,645	40%	5,058	8,093	75%	6,069	5.07	30,784
C-1	42,755	40%	17,102	27,363	75%	20,522	1.00	20,522
C-2	22,503	40%	9,001	14,402	75%	10,801	1.90	20,522
C-3	15,519	40%	6,208	9,932	75%	7,449	2.76	20,522
C-4	11,985	40%	4,794	7,671	75%	5,753	3.57	20,522
D-1	32,066	40%	12,827	20,522	75%	15,392	1.00	15,392
D-2	16,877	40%	6,751	10,801	75%	8,101	1.90	15,392
D-3	11,639	40%	4,656	7,449	75%	5,587	2.76	15,392
E-1	21,378	40%	8,551	13,682	75%	10,261	1.00	10,261
E-2	11,251	40%	4,501	7,201	75%	5,401	1.90	10,261
F-1	10,689	40%	4,276	6,841	75%	5,131	1.00	5,131

Table 5-2-4 Household, person, and interview sample sizes for Household Option 2 With Proxy

Scenario	Household sample size	HH response and eligibility rate	Responding households	Sampled adults	Adult first interview response rate	Target persons interviewed	Expected interviews per person	Total interviews
A-1	113,175	40%	45,270	72,432	85%	61,567	1.00	61,567
A-2	58,792	40%	23,517	37,627	85%	31,983	1.93	61,567
A-3	40,200	40%	16,080	25,728	85%	21,869	2.82	61,567
A-4	30,819	40%	12,328	19,724	85%	16,766	3.67	61,567
A-5	21,391	40%	8,556	13,690	85%	11,636	5.29	61,567
A-6	11,955	40%	4,782	7,651	85%	6,504	9.47	61,567
B-1	56,588	40%	22,635	36,216	85%	30,784	1.00	30,784
B-2	29,396	40%	11,758	18,814	85%	15,991	1.93	30,784
B-3	20,100	40%	8,040	12,864	85%	10,934	2.82	30,784
B-4	15,410	40%	6,164	9,862	85%	8,383	3.67	30,784
B-5	10,695	40%	4,278	6,845	85%	5,818	5.29	30,784
C-1	37,725	40%	15,090	24,144	85%	20,522	1.00	20,522
C-2	19,597	40%	7,839	12,542	85%	10,661	1.93	20,522
C-3	13,400	40%	5,360	8,576	85%	7,290	2.82	20,522
C-4	10,273	40%	4,109	6,575	85%	5,589	3.67	20,522
D-1	28,294	40%	11,318	18,108	85%	15,392	1.00	15,392
D-2	14,698	40%	5,879	9,407	85%	7,996	1.93	15,392
D-3	10,050	40%	4,020	6,432	85%	5,467	2.82	15,392
E-1	18,863	40%	7,545	12,072	85%	10,261	1.00	10,261
E-2	9,799	40%	3,919	6,271	85%	5,330	1.93	10,261
F-1	9,431	40%	3,773	6,036	85%	5,131	1.00	5,131

A total cost is computed using a simple linear cost function (as is standard):

$$C = H * c_h + A * 1 + F * c_f$$

where H is the number of sampled households, c_h is the marginal cost of one more sampled household, A is the total number of interviewed adults, F is the total number of followup interviews, and c_f is the marginal cost of one more followup interview. Note we are assuming the marginal cost of one more interviewed adult is equal to 1.

We will present the costs using this cost function, and the following six pairs of cost factors (c_h, c_f). This is a fairly broad range of possibilities, but it may not cover the real cost ratios. Cost depends on many details as to how the households are sampled and recruited, and how followup interviews are recruited. A more careful cost analysis should be done after the range of possible options is narrowed down.

Table 5-3 Six cost-scenarios for household sampling

Cost factor	Cost Scen 1	Cost Scen 2	Cost Scen 3	Cost Scen 4	Cost Scen 5	Cost Scen 6
c_h	2	3.5	5	3.5	5	5
c_f	2	3.5	5	2	2	3.5

Tables 5-4-1 through 5-4-4 then present the calculated costs for each of the 21 scenarios and each of the 4 Household Sample Options, for the six cost scenarios presented in Table 5-3.

Table 5-4-1 Estimated relative costs for the Scenarios for Household Option 1 No Proxy

Scenario	Responding households	Target persons interviewed	Expected interviews per sampled person	Cost scenario (in 1000s of cost units)					
				1	2	3	4	5	6
A-1	81,600	61,200	1.00	224.4	346.8	469.2	346.8	469.2	469.2
A-2	42,947	32,211	1.90	176.1	284.0	391.9	240.5	304.9	348.4
A-3	29,619	22,214	2.76	159.4	262.3	365.2	203.9	248.3	306.8
A-4	22,875	17,156	3.57	151.0	251.4	351.7	185.3	219.6	285.7
A-5	16,089	12,066	5.07	142.5	240.3	338.2	166.6	190.8	264.5
A-6	9,313	6,985	8.76	134.0	229.3	324.6	148.0	162.0	243.3
B-1	40,800	30,600	1.00	112.2	173.4	234.6	173.4	234.6	234.6
B-2	21,474	16,105	1.90	88.0	142.0	195.9	120.3	152.5	174.2
B-3	14,809	11,107	2.76	79.7	131.2	182.6	101.9	124.1	153.4
B-4	11,437	8,578	3.57	75.5	125.7	175.9	92.7	109.8	142.8
B-5	8,044	6,033	5.07	71.3	120.2	169.1	83.3	95.4	132.2
C-1	27,200	20,400	1.00	74.8	115.6	156.4	115.6	156.4	156.4
C-2	14,316	10,737	1.90	58.7	94.7	130.6	80.2	101.6	116.1
C-3	9,873	7,405	2.76	53.1	87.4	121.7	68.0	82.8	102.3
C-4	7,625	5,719	3.57	50.3	83.8	117.2	61.8	73.2	95.2
D-1	20,400	15,300	1.00	56.1	86.7	117.3	86.7	117.3	117.3
D-2	10,737	8,053	1.90	44.0	71.0	98.0	60.1	76.2	87.1
D-3	7,405	5,554	2.76	39.9	65.6	91.3	51.0	62.1	76.7
E-1	13,600	10,200	1.00	37.4	57.8	78.2	57.8	78.2	78.2
E-2	7,158	5,368	1.90	29.3	47.3	65.3	40.1	50.8	58.1
F-1	6,800	5,100	1.00	18.7	28.9	39.1	28.9	39.1	39.1

Table 5-4-2 Estimated relative costs for the scenarios for Household Option 1 with proxy

Scenario	Responding households	Target persons interviewed	Expected interviews per sampled person	Cost scenario (in 1000s of cost units)					
				1	2	3	4	5	6
A-1	72,000	61,200	1.00	205.2	313.2	421.2	313.2	421.2	421.2
A-2	37,403	31,792	1.93	165.4	265.6	365.8	221.5	277.6	321.7
A-3	25,574	21,738	2.82	151.8	249.4	346.9	190.2	228.5	287.7
A-4	19,607	16,666	3.67	144.9	241.2	337.4	174.4	203.8	270.6
A-5	13,608	11,567	5.29	138.0	232.9	327.8	158.5	178.9	253.3
A-6	7,606	6,465	9.47	131.1	224.7	318.2	142.6	154.0	236.1
B-1	36,000	30,600	1.00	102.6	156.6	210.6	156.6	210.6	210.6
B-2	18,701	15,896	1.93	82.7	132.8	182.9	110.8	138.8	160.9
B-3	12,787	10,869	2.82	75.9	124.7	173.5	95.1	114.3	143.9
B-4	9,803	8,333	3.67	72.5	120.6	168.7	87.2	101.9	135.3
B-5	6,804	5,784	5.29	69.0	116.5	163.9	79.2	89.4	126.7
C-1	24,000	20,400	1.00	68.4	104.4	140.4	104.4	140.4	140.4
C-2	12,468	10,597	1.93	55.1	88.5	121.9	73.8	92.5	107.2
C-3	8,525	7,246	2.82	50.6	83.1	115.6	63.4	76.2	95.9
C-4	6,536	5,555	3.67	48.3	80.4	112.5	58.1	67.9	90.2
D-1	18,000	15,300	1.00	51.3	78.3	105.3	78.3	105.3	105.3
D-2	9,351	7,948	1.93	41.4	66.4	91.5	55.4	69.4	80.4
D-3	6,394	5,435	2.82	38.0	62.3	86.7	47.5	57.1	71.9
E-1	12,000	10,200	1.00	34.2	52.2	70.2	52.2	70.2	70.2
E-2	6,234	5,299	1.93	27.6	44.3	61.0	36.9	46.3	53.6
F-1	6,000	5,100	1.00	17.1	26.1	35.1	26.1	35.1	35.1

Table 5-4-3 Estimated relative costs for the scenarios for Household Option 2 no proxy

Scenario	Responding households	Target persons interviewed	Expected interviews per sampled person	Cost scenario (in 1000s of cost units)					
				1	2	3	4	5	6
A-1	51,306	61,567	1.00	164.2	241.1	318.1	241.1	318.1	318.1
A-2	27,003	32,404	1.90	144.7	229.0	313.2	185.2	225.7	269.5
A-3	18,623	22,347	2.76	138.0	224.8	311.6	166.0	193.9	252.7
A-4	14,383	17,259	3.57	134.6	222.7	310.7	156.2	177.8	244.3
A-5	10,116	12,139	5.07	131.2	220.5	309.9	146.4	161.6	235.7
A-6	5,856	7,027	8.76	127.8	218.4	309.0	136.6	145.4	227.2
B-1	25,653	30,784	1.00	82.1	120.6	159.0	120.6	159.0	159.0
B-2	13,502	16,202	1.90	72.4	114.5	156.6	92.6	112.9	134.7
B-3	9,311	11,174	2.76	69.0	112.4	155.8	83.0	97.0	126.4
B-4	7,191	8,630	3.57	67.3	111.3	155.4	78.1	88.9	122.1
B-5	5,058	6,069	5.07	65.6	110.3	154.9	73.2	80.8	117.9
C-1	17,102	20,522	1.00	54.7	80.4	106.0	80.4	106.0	106.0
C-2	9,001	10,801	1.90	48.2	76.3	104.4	61.7	75.2	89.8
C-3	6,208	7,449	2.76	46.0	74.9	103.9	55.3	64.6	84.2
C-4	4,794	5,753	3.57	44.9	74.2	103.6	52.1	59.3	81.4
D-1	12,827	15,392	1.00	41.0	60.3	79.5	60.3	79.5	79.5
D-2	6,751	8,101	1.90	36.2	57.2	78.3	46.3	56.4	67.4
D-3	4,656	5,587	2.76	34.5	56.2	77.9	41.5	48.5	63.2
E-1	8,551	10,261	1.00	27.4	40.2	53.0	40.2	53.0	53.0
E-2	4,501	5,401	1.90	24.1	38.2	52.2	30.9	37.6	44.9
F-1	4,276	5,131	1.00	13.7	20.1	26.5	20.1	26.5	26.5

Table 5-4-4 Estimated relative costs for the scenarios for Household Option 2 with proxy

Scenario	Responding households	Target persons interviewed	Expected interviews per sampled person	Cost scenario (in 1000s of cost units)					
				1	2	3	4	5	6
A-1	45,270	61,567	1.00	152.1	220.0	287.9	220.0	287.9	287.9
A-2	23,517	31,983	1.93	138.2	217.8	297.5	173.5	208.7	253.1
A-3	16,080	21,869	2.82	133.4	217.1	300.8	157.5	181.7	241.2
A-4	12,328	16,766	3.67	131.0	216.7	302.4	149.5	168.0	235.2
A-5	8,556	11,636	5.29	128.6	216.3	304.1	141.4	154.3	229.2
A-6	4,782	6,504	9.47	126.2	216.0	305.7	133.4	140.5	223.1
B-1	22,635	30,784	1.00	76.1	110.0	144.0	110.0	144.0	144.0
B-2	11,758	15,991	1.93	69.1	108.9	148.7	86.7	104.4	126.6
B-3	8,040	10,934	2.82	66.7	108.5	150.4	78.8	90.8	120.6
B-4	6,164	8,383	3.67	65.5	108.4	151.2	74.8	84.0	117.6
B-5	4,278	5,818	5.29	64.3	108.2	152.0	70.7	77.1	114.6
C-1	15,090	20,522	1.00	50.7	73.3	96.0	73.3	96.0	96.0
C-2	7,839	10,661	1.93	46.1	72.6	99.2	57.8	69.6	84.4
C-3	5,360	7,290	2.82	44.5	72.4	100.3	52.5	60.6	80.4
C-4	4,109	5,589	3.67	43.7	72.2	100.8	49.8	56.0	78.4
D-1	11,318	15,392	1.00	38.0	55.0	72.0	55.0	72.0	72.0
D-2	5,879	7,996	1.93	34.5	54.5	74.4	43.4	52.2	63.3
D-3	4,020	5,467	2.82	33.4	54.3	75.2	39.4	45.4	60.3
E-1	7,545	10,261	1.00	25.4	36.7	48.0	36.7	48.0	48.0
E-2	3,919	5,330	1.93	23.0	36.3	49.6	28.9	34.8	42.2
F-1	3,773	5,131	1.00	12.7	18.3	24.0	18.3	24.0	24.0

The costs given above are really relevant once the mode of data collection has been fixed and efforts to fine tune the sample design are being considered. This follows due to the very large differences in cost between face-to-face administered surveys and those conducted in other modes without sending interviewers into the field. For example, a mail survey might have a total cost in the range of $85 to $175 per completed household for a standard type of survey. Additional features could raise or lower the costs, but this range gives a reasonable order of magnitude. A telephone survey would be in the same range and a mixed mode survey (mail and Web) would also fail roughly in the same range. On the other hand, a face-to-face survey would typically have a cost in the range of $1000-$1500 per completed household. Again, the range would vary depending on the specifics of the interview and the intensity of the effort to attain high response rates, but the ratio of the face-to-face cost per complete and the mail/mixed mode interview cost per complete would probably be in a range of 8 to 12.

The cost ratios for the completed interviews by mode imply that the first decision in the design is on the mode of data collection. Any face-to-face interviewing (even nonresponse followup) drives the cost ratios much higher compared to a design without face-to-face interviewing. The implication is

that the effective sample sizes for a design with a face-to-face component would probably be less than 25% of the effective sample size for a design without face-to-face interviewing. Given the incidence rates, this has a profound effect on the ability to make estimates with adequate precision. We suspect that the potential budget for the survey will play a very major role in the design decisions.

5.2 A Module of a Nationally Representative Household Study Used for Other Purposes

Under this plan, the employee study becomes a module of an already existing nationally representative household study (a set of specialized questions added onto the questionnaire). It would be necessary to subset to the households with at least one worker in private industry (excluding mining and other out-of-scope industries) or state and local governments. The prevalence question is asked of this subset of households, and the followup questions are asked of those respondents who indicate an occupational injury/illness during the specified time window. Potential candidates for this are the NHIS, and the Current Population Survey as carried out by BLS. There are many complexities that need to be overcome, including the reality of 'context effects': getting systematically different answers to the same questions based on the larger context of the questionnaire. The answers one gets from a prevalence item nested in the NHIS or nested in the CPS are likely to not be identical to each other, or to those from a stand-alone study or an employer-based study, even setting aside coverage differences, nonresponse, etc., as is well-attested in the literature. Context alone can register a mode effect. If at all possible, these context effects should be tested for through experimentation, though the cost of this in this instance might be considerable (it requires administering the survey through several entirely different modes).

The advantages of this approach as opposed to a targeted study is that there is very little issue about concealment from a desire to hide information from their employer. The occupational injury and illness questions are nested within a much larger questionnaire, and it will not be clear to the respondent that there is any particular importance attached to these questions. It will still be necessary to reassure the respondent that the information collected will not go outside the survey organization. But the measurement error from concealment under this approach should be the lowest of any of the options considered in this report.

The same two within-household sampling options as indicated for the stand-alone study are possible here. The same sample size calculations as given in Table 5 are still relevant here, except there needs to be an allowance for a further design effects generated from the larger study (which is not done in these calculations). This option does not allow realistically for followup interviews. Only one retrospective interview is possible. The only possible scenarios are then A-1 through F-1. Table 5-5 presents sample sizes for the six scenarios under two household options as in Section 5.1 (one eligible adult interviewed per household and all eligible adults per household).

Table 5-5 Household, person, and interview sample sizes for Module Option

Household Option	Scenario	Household sample size	Adults per sampled household	Sampled adults	Interviews
1	A-1	61,200	1	61,200	61,200
1	B-1	30,600	1	30,600	30,600
1	C-1	20,400	1	20,400	20,400
1	D-1	15,300	1	15,300	15,300
1	E-1	10,200	1	10,200	10,200
1	F-1	5,100	1	5,100	5,100
2	A-1	38,480	1.6	61,567	61,567
2	B-1	19,240	1.6	30,784	30,784
2	C-1	12,827	1.6	20,522	20,522
2	D-1	9,620	1.6	15,392	15,392
2	E-1	6,413	1.6	10,261	10,261
2	F-1	3,207	1.6	5,131	5,131

It should be noted that this very inexpensive compared to the stand-alone household study if the cost can exclude the cost for the main study of which this is simply a module, and only reflects the costs of completing the prevalence interview. Also, nonresponse from the larger nesting household study is not accounted for here.

5.3 A Follow-On Study from a Nationally Representative Household Study Used for Other Purposes

This is a variant of the module plan discussed in the previous section. Under this plan, the employee study becomes a followup study from an already existing nationally representative household study. The respondent from the larger study is checked for eligibility, and then eligible respondents are subsampled for the employee study. The subsampled persons are asked whether they would be willing to participate in a followup study, and if they respond favorably are contacted at a later date to fill out the SOII questionnaire.

This approach has the advantage over the modular approach in that very few questions are actually asked in the instrument in the main study. This reduces the burden in the main study, and may make it more palatable as a regular (or semi-regular) part of the larger study. The disadvantage then is that there will certainly be attrition from inability to contact the person later and refusals to participate in the SOII study when contacted.

Table 5-6 below is a revised version of Table 5-1 for the Follow-On Study Option.

Table 5-6 Household sample sizes to achieve effective adult interview sample sizes of 5,100 for Follow-On Study Option

	Option 1 no proxies	Option 1 with proxies	Option 2 1% correlation no proxies	Option 2 1% correlation with proxies
Parent study household sample size	9,067	8,000	5,701	5,030
Household recontact rate	75%	75%	75%	75%
Eligible households	6,800	6,000	4,276	3,773
Mean sampled adults/eligible HHs	1	1	1.60	1.60
Sampled adults	6,800	6,000	6,841	6,036
Adult response rate	75%	85%	75%	85%
Adult prevalence interviews	5,100	5,100	5,131	5,131
Within household correlation	NA	NA	0.01	0.01
Design effect	1	1	1.006	1.006
Effective sample size	5,100	5,100	5,100	5,100

The only difference between the stand-alone study and the follow-on study is the method of initially recruiting households. We leave out the costs of the lead-in study, as well as the response rates from the lead-in study. One benefit of the follow-on study as opposed to the stand-alone study is that all of the households are initially known to be eligible households. There is no need to screen out ineligible households or non-households.

The Employer-Based Study Option 6

The other major option (contrasting to a nationally representative household survey) is an employer-based survey. Under this option, a carefully stratified subsample of the SOII employer sample would be drawn. From this subsample, a list of their current employees would be requested. A sample of employees would be taken from this list, and a final interview done with the employees. These interviews could possibly be at the worksites, or if contact information can be gained, in their homes or by a telephone or mobile phone. If this is planned, it is important to make sure that employees who are on sick leave or on short-term disability are not excluded.

This plan will take different forms depending on whether a retrospective option, prospective or panel option, or both a retrospective and prospective option is utilized.

Under the retrospective option, the sampled employee is contacted and questioned about the recent past (e.g., the past month, or past three months, or past year). This is a one time and out interview. The fact that no followup is necessary into the future is a definite benefit of this approach. The retrospective study does not require multiple interviews. The response rate will be highest of the options discussed here. But, the single interview will only cover the designated recall window.

Under the prospective approach, the sampling process might be similar but only an initial interview with the employees would be needed at the worksites. Contact information would be collected from the employees so that they could be contacted outside of the worksite. One or more followup interviews then would be conducted off site. The employees would be followed up even if they were not working at the time of the followup interview for any reason including injury or illness. They would also be followed up even if they had left the company and gone to another job or been placed on disability, for example.

One great advantage of the prospective study is that the initial employee list is complete, at least with regard to future injuries and illnesses. There is no loss due to persons who are on sick leave or disability at the current time as there might be with a retrospective study. In some cases the company may leave some or all of these persons off the list provided to BLS. Careful followup can be done to make sure persons are not lost because they are sick or on disability in the future (i.e., they are not present at the company site). Also, future interviews can be done outside of the context

of the employer: by internet, telephone, or in-person. In-person interviews could take place at workers' homes, in the hospital, or at their new jobs, for example.

The disadvantage to the prospective study is the cost per employee. Followup interviews will require the usual extensive fielding activities to bring make contact with individuals and interview them. Panel studies tend to have a declining response rate over time as contact is lost with some persons and other persons become weary of the interviews. But then each interview can cover a window of time well, avoiding recall biases. This improves the quality of the study.

A third option is a combined prospective-retrospective study. The initial interview can be extensive and cover a designated recall window, and then there can be future interviews. This will provide the greatest amount of information per employee, but will obviously incur the highest cost, and the two components will need to be combined together in analyses.

6.1 The Employer-Based Study: Two-Stage Design

The employer-based study is a two-stage sample design, with employers the primary sample units (PSUs) and employees the secondary sample units (SSUs). The recommendation is to carry out a subsample of the SOII employer sample for the first stage PSU sample, though this is not completely necessary. An independent sample can be drawn from the original SOII employer frame for the purposes of the employee survey, but it seems like it would be less costly to contact the same companies as carry out the SOII employer survey. If SOII employer sample establishments are in fact resistant to cooperating to the employee survey because they are in SOII employer sample (added burden?), then it may be necessary to draw a minimally overlapping sample of employers from the SOII employer frame (a sample which minimizes the overlap with the SOII employer sample). This will lose the possible benefits of having the employer-level estimates matching the employee sample from the sampled employer (of possible value in benchmarking). But if it helps with response it ought to be done.

6.2 The Employer-Based Study: Prospective Option

Under the prospective option, a list of current employees as of a given date (including temporarily disabled workers) is obtained from the company. A simple random sample is drawn from this list of

employees. We understand that the company may not be willing to provide contact information about the employees, so the sampled employees are interviewed by BLS field agents at the worksite after being called in by the employer. This screening interview is short and is meant primarily to collect contact information from the employee and to receive consent from the employee for the followup interviews, which will cover occupational injuries/illnesses in the time window (which starts from the point of the screening interview and extends into at least a year into the future). The employee is re-assured that later interviews will not be at the worksite but instead over the telephone, by email, or at their homes, and any information they will provide in those followup interviews will not be passed on to the employer (in as strong and legally binding language as possible). The prospective aspect of this option is very helpful, as the employee's response will not be correlated to any occupational injuries or illnesses in the time window, which is forward in time from this initial screening interview (though it certainly may be correlated to health issues from the past). This means that the response propensity will not differ between those who will be injured or ill in the time window and those who will not be. This will strongly reduce biases from potential differential response propensities between the injured and the non-injured.

The sampled employee is encouraged to keep an informal diary of their experience in the time window, with multiple interviews by their choice of mode (in-person, by telephone, by web). One option is to have four followup interviews at three month intervals, covering a time window of one year. The followup prevalence interviews should be generally relatively short, just collecting information about how much they were off work, how much on reduced schedules due to occupational injuries or illnesses, etc., to decide or not whether there was an incident or not. If there was an incident, then the interview is much longer to collect information about the nature of the injury or illness. It may be necessary to have the followup incident interview by another interviewer who is trained to carry out the more extensive interview.

The followup regime should be relatively intense to avoid differential response propensities from arising based on incidence. Employees who do have incidents may become more difficult to reach, and they may be shy about revealing their statuses. Followup contacts should be much more intense than the initial contact: once they respond to the screener and move into the time window all effort should be made not to allow them to attrit in a differential way (in particular, we don't want those who have incidents in the time window to attrit at a higher rate than those who don't have incidents in the time window). The employees should be tracked in the time window wherever they end up (in the hospital, in new jobs, at home on disability, etc.). This will require considerable effort, but this effort is central to minimizing response biases in the prevalence estimates.

The cost ratio c_1/c_2 mentioned in Section 4.3 is between the cost c_1 of recruiting and fielding a PSU (an employer) and the cost c_2 of recruiting and fielding an SSU (an employee), and is important in deciding the relative employer and employee sample sizes. The employer needs to be recruited for the employee survey and a list of employees needs to be collected from the employer, and an initial visit needs to be made in person to the employer's worksite for a workday or two (or more, depending on the employee sample size) to recruit the sampled employees and to carry out the screening interviews. The c_1 cost includes these employer contacts and followups, and the cost of sending field personnel to the employer worksite for the screening interviews. The c_2 cost includes the differential cost of the particular screening interview (relatively small), and then the follow-on costs of prevalence interviews for all of the sampled employees, and incidence interviews for the small proportion of employees who have an occupational injury/illness incident during the time window.

The c_1 cost may be reduced somewhat if the employer sample comes from the already fielded SOII employer sample, as these employers are already cooperating with BLS, and only need to be recruited for this additional task, and if the field visits are coordinated geographically. If necessary, geographic clustering can be included in the sample design to reduce these travel costs further. The c_2 cost should be relatively high, as intensity in the followup interviewing is going to be critical to minimizing response bias in the prevalence estimates. Thus we may expect a relatively low c_1/c_2 ratio (5 to 20). BLS field staff can make these ratios more precise, especially after a pilot study. The intra-PSU correlations as discussed in Section 3.3 are likely to be in the 0% to 10% range. It may be possible to use the SOII data base to generate an estimate of intra-employer correlation at least for SOII estimates (which we assume are somewhat biased as estimates for employee incidence) which can be used for planning purposes. At this point, we will assume a 5 to 20 c_1/c_2 ratio and a 1%-10% intra-PSU correlation. Table 4-5 shows a wide range of optimal SSU per PSU $\bar{n}$ mean sample sizes: from 6.0 to 44.5. We will work with mean SSU sample sizes $\bar{n}$ of 6, 8, 10, 15, 20, 30. Tables 6-1-1 through 6-1-5 below present employer sample sizes necessary to achieve an effective employee sample size of 5,100 with an intra-PSU correlation of 5% for each of these $\bar{n}$ values. The effective sample sizes vary by the value of the intra-PSU correlation. The 5,100 benchmark is from Table 2-1 (Scenario 5): the effective sample size necessary to achieve 80% power to distinguish a difference of 20% in prevalence rates between employer and employee surveys.

Table 6-1-1 Employer and employee sample sizes and design effects: mean SSU per PSU sample size of 6

Intra-PSU correlation	Employer sample size	Employee sample size	Total employee sample size	Design effect	Effective employee sample size
1.0%	1,064	6	6,384	1.05	6,080
2.5%	1,064	6	6,384	1.125	5,675
5.0%	1,064	6	6,384	1.25	5,107
7.5%	1,064	6	6,384	1.375	4,643
10.0%	1,064	6	6,384	1.5	4,256

Table 6-1-2 Employer and employee sample sizes and design effects: mean SSU per PSU sample size of 8

Intra-PSU correlation	Employer sample size	Employee sample size	Total employee sample size	Design effect	Effective employee sample size
1.0%	862	8	6,896	1.07	6,445
2.5%	862	8	6,896	1.175	5,869
5.0%	862	8	6,896	1.35	5,108
7.5%	862	8	6,896	1.525	4,522
10.0%	862	8	6,896	1.7	4,056

Table 6-1-3 Employer and employee sample sizes and design effects: mean SSU per PSU sample size of 10

Intra-PSU correlation	Employer sample size	Employee sample size	Total employee sample size	Design effect	Effective employee sample size
1.0%	740	10	7,400	1.09	6,789
2.5%	740	10	7,400	1.225	6,041
5.0%	740	10	7,400	1.45	5,103
7.5%	740	10	7,400	1.675	4,418
10.0%	740	10	7,400	1.9	3,895

Table 6-1-4 Employer and employee sample sizes and design effects: mean SSU per PSU sample size of 15

Intra-PSU correlation	Employer sample size	Employee sample size	Total employee sample size	Design effect	Effective employee sample size
1.0%	578	15	8,670	1.14	7,605
2.5%	578	15	8,670	1.35	6,422
5.0%	578	15	8,670	1.7	5,100
7.5%	578	15	8,670	2.05	4,229
10.0%	578	15	8,670	2.4	3,613

Table 6-1-5 Employer and employee sample sizes and design effects: mean SSU per PSU sample size of 20

Intra-PSU correlation	Employer sample size	Employee sample size	Total employee sample size	Design effect	Effective employee sample size
1.0%	498	20	9,960	1.19	8,370
2.5%	498	20	9,960	1.475	6,753
5.0%	498	20	9,960	1.95	5,108
7.5%	498	20	9,960	2.425	4,107
10.0%	498	20	9,960	2.9	3,434

Table 6-2 presents the 21 scenarios from Section 4.2 as they would be applied to the employer prospective option. All of these scenarios are 'in play' for the employer prospective option: any number of followups is possible once the necessary contact information has been collected from the employees in the initial employer sample. Table 6-2 assumes the two-stage design from Table 6-1-3: a total of 10 cooperative employees per employer. Note that 'cooperation' in this case means the employee agrees to provide contact information, and also participates in the first prevalence interview. All of the employee and interview sample sizes from Section 4.2 are inflated here to allow for the expected design effect from employer-based sampling with 10 sampled employees per sampled employer (1.45). The 'adjusted persons interviewed' is the "persons interviewed total" (assuming no nonresponse or design effects) from Table 4-2 for the scenario multiplied by the design effect and divided by the expected longitudinal interview cooperation rate (from Table 4-3). Note that the total interviews after longitudinal attrition is 1.45 times the corresponding totals from Table 4-4. The two-stage design with its assumed correlation forces everything to be proportionately higher as per the assumed design effect.

Table 6-2 Prospective interview scenarios for Employer-Based Option: Employee Sample Size 10 (Part 1)

Scenario	Months covered for interview	Target interviews per person	Persons interviewed	Design effect	Expected interview aggregate cooperation rate	Adjusted persons interviewed	Expected interviews per sampled person	Total interviews
A-1	1	1	61,200	1.45	100.0%	88,740	1.00	88,740
A-2	1	2	30,600	1.45	95.0%	46,705	1.90	88,740
A-3	1	3	20,400	1.45	91.8%	32,211	2.76	88,740
A-4	1	4	15,300	1.45	89.2%	24,876	3.57	88,740
A-5	1	6	10,200	1.45	84.5%	17,496	5.07	88,740
A-6	1	12	5,100	1.45	73.0%	10,128	8.76	88,740
B-1	2	1	30,600	1.45	100.0%	44,370	1.00	44,370
B-2	2	2	15,300	1.45	95.0%	23,353	1.90	44,370
B-3	2	3	10,200	1.45	91.8%	16,105	2.76	44,370
B-4	2	4	7,650	1.45	89.2%	12,438	3.57	44,370
B-5	2	6	5,100	1.45	84.5%	8,748	5.07	44,370
C-1	3	1	20,400	1.45	100.0%	29,580	1.00	29,580
C-2	3	2	10,200	1.45	95.0%	15,568	1.90	29,580
C-3	3	3	6,800	1.45	91.8%	10,737	2.76	29,580
C-4	3	4	5,100	1.45	89.2%	8,292	3.57	29,580
D-1	4	1	15,300	1.45	100.0%	22,185	1.00	22,185
D-2	4	2	7,650	1.45	95.0%	11,676	1.90	22,185
D-3	4	3	5,100	1.45	91.8%	8,053	2.76	22,185
E-1	6	1	10,200	1.45	100.0%	14,790	1.00	14,790
E-2	6	2	5,100	1.45	95.0%	7,784	1.90	14,790
F-1	12	1	5,100	1.45	100.0%	7,395	1.00	7,395

Table 6-3 presents the employer-employee sample design that will achieve the adjusted persons interviewed, assuming a 50% screener response rate for employees (50% of sampled employees agree to provide contact information at the employer work-site) and a 74.1% first-interview response rate (74.1% of the screener employees are found and interviewed for the first prevalence interview).

Table 6-3 Prospective interview scenarios for Employer-Based Option: Employee Sample Size 10 (Part 2)

Scenario	Months covered for interview	Target interviews per person	Employer sample size	Employees sampled per employer	Total employees sampled	Expected screener response rate	Expected first interview response rate	Employees interviewed per employer	Adjusted persons interviewed
A-1	1	1	8,874	27	239,598	50%	74.1%	10	88,740
A-2	1	2	4,671	27	126,104	50%	74.1%	10	46,705
A-3	1	3	3,221	27	86,968	50%	74.1%	10	32,211
A-4	1	4	2,488	27	67,166	50%	74.1%	10	24,876
A-5	1	6	1,750	27	47,240	50%	74.1%	10	17,496
A-6	1	12	1,013	27	27,346	50%	74.1%	10	10,128
B-1	2	1	4,437	27	119,799	50%	74.1%	10	44,370
B-2	2	2	2,335	27	63,052	50%	74.1%	10	23,353
B-3	2	3	1,611	27	43,484	50%	74.1%	10	16,105
B-4	2	4	1,244	27	33,583	50%	74.1%	10	12,438
B-5	2	6	875	27	23,620	50%	74.1%	10	8,748
C-1	3	1	2,958	27	79,866	50%	74.1%	10	29,580
C-2	3	2	1,557	27	42,035	50%	74.1%	10	15,568
C-3	3	3	1,074	27	28,989	50%	74.1%	10	10,737
C-4	3	4	829	27	22,389	50%	74.1%	10	8,292
D-1	4	1	2,219	27	59,900	50%	74.1%	10	22,185
D-2	4	2	1,168	27	31,526	50%	74.1%	10	11,676
D-3	4	3	805	27	21,742	50%	74.1%	10	8,053
E-1	6	1	1,479	27	39,933	50%	74.1%	10	14,790
E-2	6	2	778	27	21,017	50%	74.1%	10	7,784
F-1	12	1	740	27	19,967	50%	74.1%	10	7,395

A total cost is computed using a simple linear cost function (as is standard):

$$C = EMPR * c_{EMP} + EMPY * 1 + F * c_f$$

where $EMPR$ is the number of sampled employers visited, c_{EMP} is the marginal cost of one more visited employer site, A is the total number of interviewed employees at the worksite, F is the total number of followup interviews, and $c_f$is the marginal cost of one more followup employee interview. Note we are assuming the marginal cost of one more interviewed employee is equal to 1.

We will present the costs using this cost function, and the following three pairs of cost factors (c_{EMP}, c_f). This is a fairly broad range of possibilities, but it may not cover the real cost ratios. Cost depends on many details as to how the employers are sampled and recruited, and how followup interviews are recruited. Note that we are assuming a high cost per sampled employer, as a lot of effort will be necessary including a possible multiple day onsite visit.

Table 6-4 Three cost-scenarios for employer-based prospective sampling

Cost factor	Cost Scen 1	Cost Scen 2	Cost Scen 3
c_{EMP}	60	60	60
c_f	2	3.5	5

Table 6-5 then presents the calculated costs for each of the 21 scenarios, for the three cost scenarios presented in Table 6-4.

Table 6-5 Estimated relative costs for the scenarios for employer-based sampling

Scenario	Employer sample size	Adjusted persons interviewed	Expected interviews per sampled person	Cost Scenario (in 1000s of cost units)		
				1	2	3
A-1	8,874	88,740	1.00	621.2	621.2	621.2
A-2	4,671	46,705	1.90	411.0	474.1	537.1
A-3	3,221	32,211	2.76	338.5	423.3	508.1
A-4	2,488	24,876	3.57	301.9	397.7	493.5
A-5	1,750	17,496	5.07	265.0	371.8	478.7
A-6	1,013	10,128	8.76	228.1	346.0	464.0
B-1	4,437	44,370	1.00	310.6	310.6	310.6
B-2	2,335	23,353	1.90	205.5	237.0	268.6
B-3	1,611	16,105	2.76	169.3	211.7	254.1
B-4	1,244	12,438	3.57	150.9	198.8	246.7
B-5	875	8,748	5.07	132.5	185.9	239.3
C-1	2,958	29,580	1.00	207.1	207.1	207.1
C-2	1,557	15,568	1.90	137.0	158.0	179.0
C-3	1,074	10,737	2.76	112.8	141.1	169.4
C-4	829	8,292	3.57	100.6	132.6	164.5
D-1	2,219	22,185	1.00	155.3	155.3	155.3
D-2	1,168	11,676	1.90	102.8	118.5	134.3
D-3	805	8,053	2.76	84.6	105.8	127.0
E-1	1,479	14,790	1.00	103.5	103.5	103.5
E-2	778	7,784	1.90	68.5	79.0	89.5
F-1	740	7,395	1.00	51.8	51.8	51.8

6.3 The Employer-Based Study: Retrospective Option

Under the retrospective option, the sampling of employees is the carried through in the same way as for the prospective option: a list of current employees as of a given date (including temporarily disabled workers) is obtained from the company, and a sample is drawn from it. A simple random sample is drawn from this list of employees. The sampled employees are called in at the worksite, and the full interview is carried out for the time window, which in this case is the year before the sample interview date. The employee in this case has to be re-assured that the information that is provided will not be given to the employer (in as strong and legally binding language as possible). This will be difficult to make credible given that the interview is at the worksite itself. Every effort will need to be made to include employees who are on sick leave or disability, possibly with a later 'make-up' interview. It will be very difficult to avoid differential response propensities between employees with incidents over the time window and those without under this option. All efforts need to be made in this direction to avoid biases in the prevalence estimates.

Table 6-6 below presents the sample sizes necessary to achieve an effective employee sample size of 5,100 with an intra-PSU correlation of 5% for each an $\bar{n}$ value of 10, for the six scenarios from Section 4.2 which correspond to only one interview (A-1, B-1, C-1, D-1, E-1, F-1). A 34.5% (10 divided by 29) response rate is assumed at the employee level: 34.5% of employees are assumed to complete the prevalence interview. It will require considerable effort to achieve this response rate under the circumstances. The design effect of 1.45 is assumed. Also included is a cost, assuming a cost function with 60 times the number of employers visited, plus 1 times the number of worksite interviews.

Table 6-6 Retrospective option employer and employee sample sizes and design effects: six scenarios

Scenario	Employer sample size	Employees sampled per employer	Total employees sampled	Expected interview response rate	Employees interviewed per employer	Total interviews	Design effect	Effective interviews	Cost (in 1000 units)
A-1	8,874	29	257,346	34.5%	10	88,740	1.45	61,200	621.2
B-1	4,437	29	128,673	34.5%	10	44,370	1.45	30,600	310.6
C-1	2,958	29	85,782	34.5%	10	29,580	1.45	20,400	207.1
D-1	2,219	29	64,337	34.5%	10	22,185	1.45	15,300	155.3
E-1	1,479	29	42,891	34.5%	10	14,790	1.45	10,200	103.5
F-1	740	29	21,446	34.5%	10	7,395	1.45	5,100	51.8

6.4 The Employer-Based Study: Retrospective-Prospective Option

There is the possibility of having both a retrospective and a prospective option. One variant of this is to have a retrospective interview with a three-month window, and then a prospective option with three followup interviews with three-month windows. There are further variants on this (e.g., a retrospective interview with a six-month window with two followup interviews). It seems however that this option has some of the weaknesses of both the retrospective and prospective approaches without having some of the strengths of either. The RP options will suffer the weakness of the retrospective options in that the first prevalence interview will take place immediately during the screening interview at the worksite. The RP option then suffers the extra expense and difficulty of the prospective option in requiring a followup of the sampled employees. We don't recommend this approach and will not develop its properties.

7 Summary

There are several major issues in the development of a sample design for an Employee Survey for SOII. A large enough sample is needed to assure that there is an 80% chance of perceiving a 20% difference between the SOII employer prevalence of occupational injuries and illnesses in the past year and the SOII employee prevalence at the national level. The effective sample size needed for this is 5,100 (the equivalent of a simple random sample of size 5,100 from the employee population). If it is also necessary to achieve this precision standard for subgroups such as occupational injuries and illnesses that cause a loss of worktime, or cause a need for temporary restrictions or job transfers, (or other injuries/illnesses) then the effective sample sizes increase accordingly (by a factor of up to 4 to achieve the standard for all three subgroups). If it is necessary to achieve this precision standard for particular industries and/or occupations, then either oversampling is required, or a much larger effective sample size. These are decision branches that need to be chosen.

The population of interest is not simply eligible workers, but a conceptual combination of eligible workers and time intervals. The object is to measure all occupational injuries and illnesses among eligible workers during the year corresponding to the survey. Research has shown that simply asking workers about their past years' experience may lead to recall error. The least expensive approach is to ask the worker about an entire year's time in a single interview. But this will have the highest measurement error, with bias levels dependent on the type of injury/illness. Another approach is to have a shorter time window (three months for example), and have a larger sample size. The annual time period can be covered with 1 interview per employee covering m months, with a $12/m$ times larger employee sample size. Or alternatively, the time period of the interview can be m months and the sample size remains constant, but there are d total interviews of the sampled person (with $d * m = 12$). This can be an initial prevalence interview with $d - 1$ followup interviews (a retrospective/prospective approach), or d followup interviews after an initial screening interview (a fully prospective approach). We have defined 21 possible scenarios which cover all reasonable permutations of time window, number of interviews, and number of persons interviewed. The pros and cons of these scenarios are generally independent of other aspects of the design. Figures 4-2-1 and 4-2-2 present these aspects. The option of having followup interviews is not possible with the retrospective sample design options. For these designs, only six of the 21 scenarios are possible (A-1, B-1, C-1, D-1, E-1, F-1).

Our general sense is that a one-month time interval, though the best in terms of measurement error, may be too expensive for the benefit thereof. On the other hand, a twelve-month time interval would probably defeat the purpose of getting relatively unbiased prevalence estimates. The trade-offs are probably between two, three, four, and six-month time intervals. It may be best to test this explicitly through designed experiments in the pilot study and in early years of the main study. The allowance of proxies is also a critical dimension in this. A longer time-interval is compatible with not using proxies for prevalence. The allowance for proxies will change the calculus towards shorter time windows. These are options that offer themselves well to designed experiments.

A major decision branch is retrospective vs. prospective. The only option that ***has*** to be retrospective is the household survey module option. All other options can be either retrospective, or prospective, or both. The prospective option is argued to be better as the initial response is unrelated to prevalence (very important in reducing nonresponse bias, possibly drastically), but it is more expensive as it requires extensive followup interviews.

The household survey vs. employer-based option is of course a major decision branch. The household survey will be more expensive per sampled employee, as reaching the households will be more difficult than reaching employers, as the SOII employer sample is already available to be drawn from. The household survey can be a stand-alone survey, a module of an already existing national household survey such as NHIS, or a follow-on survey to an already existing national household survey. Tying into another national survey will require cooperation with BLS from another agency. If this can be done, then the expense per sampled employee will be much less, as considerable screening can be done by the larger survey. But there will be little flexibility: the study will have to be single-interview retrospective. A follow-on study from a larger national survey will provide considerable flexibility (a prospective option is possible, with multiple interviews), but the cost of bringing in these followup interviews successfully may defeat much of the cost savings from having the larger survey as a lead-in. The module option has the added benefit that the prevalence questions are one among many questions, and the respondent will not feel that there is any special focus on occupational injuries and illnesses, reducing to a minimum concealment bias. The module option has to be retrospective, but given the modularity this particular retrospective option is unlikely to suffer differential nonresponse propensities: nonresponse will not be related to occupational injuries/illnesses unlike 'stand-alone' retrospective options.

The employer-based study will be less expensive per interviewed employee, as the employers are easier to recruit (the SOII employer sample can be subsampled). But a larger number of employees need to be sampled in order to offset the design effect from intra-employer prevalence correlation.

One must expect that employers will have differing prevalence rates, and this will reduce the efficiency of the two-stage design. Another drawback of the employer-based study as compared to the household-level study is the risk of concealment error: employees may not wish to reveal information to their employers about their injuries that they may be hiding from their employers. A prospective design can mitigate this risk, as well as careful reassurances. The prospective/retrospective branching is more strongly weighted towards the prospective option in an employer-based study as opposed to a household-based study.

Table 7-1 below presents a summary of the options and some of the considerations in choosing these options.

Table 7-1 Summary of options with costs and benefits

Design Option	Compatible scenarios	Clustering design effects	Potential for prevalence response rate differentials	Cost in bringing in first-stage sample	Cost in follow-up interviews	Extra costs
Stand-alone Household Study–Retrospective	A-1, B-1, C-1, D-1, E-1, F-1	Lower[1]	Lowest[3]	Higher[6]	None possible	None
Stand-alone Household Study–Prospective	All 21	Lower[1]	Lowest[3]	Higher[6]	Depends on scenario	None
Module of Existing Household Study–Retrospective	A-1, B-1, C-1, D-1, E-1, F-1	Lower[1]	Lowest[3]	Lower[7]	None possible	Need to accommodate existing study
Follow-on to Existing Household Study–Retrospective	A-1, B-1, C-1, D-1, E-1, F-1	Lower[1]	Lowest[3]	Lower[7]	None possible	Need to accommodate existing study
Follow-on to Existing Household Study–Prospective	All 21	Lower[1]	Lowest[3]	Lower[7]	Depends on scenario	Need to accommodate existing study
Employer Based Study–Retrospective	A-1, B-1, C-1, D-1, E-1, F-1	Higher[2]	Higher[4]	Much higher[8]	None possible	None
Employer Based Study–Prospective	All 21	Higher[2]	Lower[5]	Much higher[8]	Depends on scenario	None
Employer Based Study-Retrospective and Prospective	All 21	Higher[2]	Higher[4]	Much higher[8]	Depends on scenario	None

[1]Design effect not expected to be large as correlations in prevalence incidence within household not expected to be large.

[2]Design effects expected to be larger if employers are the first stage, but this can be mitigated by stratifying the employer frame carefully (by relative prevalence levels).

[3]Prevalence response rate differentials likely to be low as contact point will be independent of employer.

[4]Prevalence response rate differential for retrospective employer-based study likely to be high without careful safeguards as interview is done on the worksite.

[5]Prevalence response rate differential for prospective employer-based study is better than retrospective if followup interviews done outside worksite.

[6]Stand-alone household study requires recruitment of households for screener through mail, telephone, or in-person.

[7]Households in module or follow-on option are already recruited from the parent survey.

[8]Recruitment at employer worksites has to be in-person, with need to gain employer buy-in in advance as well.

References R

Azaroff, L. S., C. Levenstein, D. H. Wegman (2002). Occupational Injury and Illness Surveillance: Conceptual Filters Explain Underreporting. *American Journal of Public Health, 92(9)*, 1421-1429.

Bethlehem, J. G., and H. M. P. Kersten (1985). On the Treatment of Nonresponse in Sample Surveys. *Journal of Official Statistics, 3*, 141-153.

Brick, J. M., W. R. Andrews, and N. A. Mathiowetz (2015). Single-Phase Mail Survey Design for Rare Population Subgroups. *Field Methods* (upcoming). http://fmx.sagepub.com.

Brick, J. M., and G. Kalton (2007). Survey Mode Choices: Data Quality and Costs. *Proceedings of the Ninth Conference on Health Survey Research Methods.*

Brick, J. M., and D. Williams (2013). Explaining Rising Nonresposne Rates in cross-Sectional Surveys. *Annals of the American Academy of Political and Social Science, 645(1)*, 36-59.

Church, A. H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly, 57*, 62-79.

Dillman, D. A., M. D. Sinclair, and J. R. Clark (1993). Effects of Questionnaire Length, Respondent-Friendly Design, and a Diffficult Question on Response Rates for Occupant-Addressed Census Mail Surveys. *Public Opinion Quarterly 57 (3)*, 289-304.

Dillman, D. A., J. D. Smyth, L. M. Christian (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method. 4th ed.* Hoboken, NJ: John Wiley & Sons.

Forst, L., E. Ahonen, J. Zanoni, A. Holloway-Beth, M. Oschner, L. Kimmel, C. Martino, E. Rodriguez, A. Kader, E. Ringholm, and R. Sokas. More Than Training: Community-based Participatory Research to Reduce Injuries Among Hispanic Construction Workers. *American Journal of Industrial Medicine 56*, 827-837.

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly 70 (5),* 646-675.

Halbesleben, J. R. B. (2010). The Role of Exhaustion and Workarounds in Predicting Occupational Injuries: A Cross-Laged Panel Study of Health Care Professionals. *Journal of Occupational Health Psychology 13*, 1-16.

Hansen, M. H., W. N. Hurwitz, W. G. Madow (1953). *Sample Survey Methods and Theory: Volume I--Methods and Applications.* New York: John Wiley & Sons.

Hirsch, B., D. Macpherson, and J. M. Dumond (1997). Workers' Compensation Recipiency in Union and Nonunion Workplaces. *Industrial and Labor Relations Review. 50*, 213-236.

Huang, Y.-H., S. K. Verma, W.-R. Chang, T. K. Courtney, D. A. Lombardi, M. J. Brennan, and M. J. Perry (2012). Supervisor vs. Employee Safety Perceptions and Association with Future Injury in U.S. Limited-Service Restaurant Workers. *Accident Analysis and Prevention 47*, 45-51.

Jasso, G., D. Massey, M. Rosenzweig, and J. Smith (2006). The New Immigrant Survey 2003 Round 1 (NIS-2003-1) Public Release Data. http://nis.princeton.edu.

Kalton, G. (1993). *Sampling Rare and Elusive Populations.* United Nations: Department of Economic and Social Information and Policy Analysis, Statisticaal Division, National Household Survey Capability Programme.

Kalton, G., D. Kasprzyk, and D. B. McMillen (1989). Nonsampling Errors in Panel Surveys. In *Panel Surveys* (eds. D. Kasprzyk, G. Duncan, G. Kalton, M. P. Singh). New York: John Wiley & Sons.

Keeter, S., C. Miller, A. Kohut, R. M. Groves, S. Presser (2000). Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly 64*, 125-148.

Mercer, A., A. Caporaso, D. Cantor, R. Townsend (2015). How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys. *Public Opinion Quarterly 79*, 105-129.

Montaquila, J. M., J. M. Brick, D. Williams, K. Kim, D. Han (2013). A Study of Two-Phase Mail Survey Data Collection Methods. *Journal of Survey Statistics and Methodology 1*, 66-87.

Morse, T., L. Punnett, N. Warren, C. Dillon, and A. Warren (2003). The Relationship of Unions to Prevalence and Claim Filing for Work-Related Upper-Extremity Musculoskeletal Disorders. *American Journal of Industrial Medicine 44*, 83-93.

Oh, H. L., and F. J. Scheuren (1983). Weighting Adjustment for Unit Nonresponse. *Incomplete Data in Sample Surveys, Vol. 2* (eds. W. G. Madow, I. Olkin, D. B. Rubin). New York: Academic Press, 143-184.

Peytchev, A., R. K. Baxter., L. R. Carley-Baxter (2009). Not All Survey Effort is Equal: Reduction of Nonresponse Bias and Nonresponse Error. *Public Opinion Quarterly 73*, 785-806.

Pransky, G., D. Moshenberg, K. Benjamin, S. Portillo, J. L. Thackrey, and C. Hill-Fotouhi (2002). Occupational Risks and Injuries in Non-Agricultural Immigrant Latino Workers. *American Journal of Industrial Medicine 42*, 117-123.

Sahlqvist, S., Y. Song, F. Bull, E. Adams, J. Preston, and D. Ogilvie (2011). Effect of Questionnaire Length, Personalisation and Reminder Type on Response Rate to a Complex Postal Survey: Randomised Controlled Trial. BMC Medical Research Methodology 11:62, http://www.biomedcentral.com/1471-2288/11/62.

Särndal, C.-E., and P. Lundquist (2014). Accuracy in Estimation with Nonresponse: A Function of Degree of Imbalance and Degree of Explanation. *Journal of Survey Statistics and Methodology 2*, 361-387.

Schouten, B., F. Cobben, and J. Bethlehem (2009). Indicators for the Representativeness of Survey Response. *Survey Methodology 35 (1)*, 101-113.

Singer, E., R. M. Groves, and A. D. Corning (1999). Differential Incentives: Beliefs About Practices, Perceptions of Equity, and Effects on Survey Participation, *Public Opinion Quarterly, 63*, 251-260.

Valliant, R., J. A. Dever, F. Kreuter (2013). *Practical Tools for Designing and Weighting Survey Samples.* New York: Springer.

Warner, M., N. Schenker, M. A. Heinen, and L. A. Fingerhut (2005). The Effect of Recall on Reporting Injury and Poisoning Episodes in the National Health Interview Survey. *Injury Prevention, 11*, 282-287.

Wiatrowski, W. J. (2014). Examining the Completeness of Occupational Injury and Illness Data: An Update on Current Research. Monthly Labor Review June 2014. http://www.bls.gov/opub/mlr/2014/article/examining-the-completeness-of-occupational-injury-and-illness-data-an-update-on-current research.htm.