

Pilot Study Design for the SOII Employee Study

Authors

Lou Rizzo, Ph.D.
Cynthia Helba, Ph.D.
J. Michael Brick, Ph.D.



September 1, 2015

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

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

Table of Contents

<u>Chapter</u>		<u>Page</u>
1	Introduction	1-1
2	Summary of Cost-Benefit Analysis	2-1
3	Face-to-Face Sample Design.....	3-1
4	Telephone Survey Sample Design.....	4-1
5	Integration of the Employee Study with Larger National Studies	5-1
6	Questionnaire Development and Proxies	6-1
7	Weighting and Estimation.....	7-1
8	Information to Collect from the Pilot Study and Preliminary Activities.....	8-1

Tables

3-1	Illustration of design and design effect calculations.....	3-3
3-2	Possible face-to-face design with twelve-month recall period.....	3-4
4-1	Possible telephone design with twelve-month recall period	4-3

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. 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. Therefore, BLS has decided to pursue a supplementary data collection of information directly from a nationally representative sample of employees (called the “SOII Employee Study” below). 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 (this is called the ‘Literature Review’ below). A cost-benefit analysis of the various possible options is provided in “Survey of Occupational Injuries and Illnesses Employee Survey cost Benefit Analysis Criteria” by Rizzo, Helba, Brick, Bernstein, and Leonard at Westat (dated May 12, 2015), also under this contract (this is called the ‘Cost-Benefit Analysis’ below).

The starting point for this document is the Cost-Benefit Analysis, which discusses the relative costs and benefits of a wide range of possible designs for the SOII Employee Study. Section 2 of this report summarizes these results, and discusses some choices made by BLS based on the Cost-Benefit Analysis with regard to moving forward. The primary substance of *this* document is the development of two full-scale sample designs for the pilot study for the SOII Employee Study.

The goal of the pilot study is to provide a full-scale dress rehearsal of an employee survey. This pilot study will be designed to provide nationally representative estimates of prevalence of occupational injuries and illnesses over a year period (to compare to the SOII employer study). The pilot study is intended primarily to differ from the main version of this study only in terms of sample size— the pilot will not be large enough to give sufficient precision for domains (see Section 2 of the Cost-Benefit Report for a listing of these domains). The main version of this study will be a ‘scale-up’ of the pilot study. Another difference is that we recommend that the pilot study have embedded experiments to compare possible detailed methods, and that more extensive information might be collected to inform the main study.

Section 2 of this document summarizes the Cost-Benefit Analysis, and the decisions made based on that document and other considerations. Section 3 presents a face-to-face design for the SOII Employee Study. Section 4 provides a telephone design. Section 5 discusses issues of the integration of the SOII Employee Study with larger national surveys. Section 6 discusses questionnaire development. Section 7 discusses estimation and variance estimation. Section 8 discusses information that should be collected from the pilot study to facilitate future studies.

Summary of Cost-Benefit Analysis

2

The Cost-Benefit Analysis set an effective sample size benchmark of 5,100 employee/years (5,100 employees and a one-year time window per employee). This is the sample size necessary, assuming simple random sampling, to detect a prevalence difference of 20 percent with 80 percent power (e.g., a difference between 3.5 injuries per 100 employees per year and 4.2 injuries per 100 employees per year). We will continue to utilize this benchmark in this report.

The Cost-Benefit Analysis provided two main branches: a household-based study and an employer-based study. The employer-based study is thoroughly explored in Section 6 of the Cost-Benefit Analysis. BLS has chosen not to move forward with the employer-based design due to fears of low response rates (both on the employer part and on the employee part), and because it has many of the same features as the current SOII. The employer-based design is not discussed in this report. The two designs developed in this report are household-based studies.

Three different designs were studied in the Cost-Benefit Analysis for the household-based survey: stand-alone, module, and follow-on. The stand-alone study is a direct national sample of households for this SOII employee study. The module option puts the relevant questions as a module in a larger study such as the National Health Interview Survey (NHIS). The follow-on option also piggy-backs on a larger household study, but in this case a subsample of cooperative households from the larger study is targeted for followup interviews for the SOII.

The mode of data collection is a critical tradeoff between cost and quality. Data collection can be face-to-face (in-person), by telephone, or by mail. There are mixtures of these as well: for example mail followed by telephoning of nonrespondents. Face-to-face interviews are likely to have the highest response rates and allow for more complex instruments, probing, and other attributes that are possible in this setting. Of course, face-to-face interviews are very expensive per interview. Telephone interviewing allows for interviewer administration with a trained interviewer and is much less expensive than face-to-face interviews. Response rates have been in decline for years and may not be acceptable to all stakeholders. Mail interviewing is relatively inexpensive and has response rates that are generally higher than telephone interviewing, but the interviews must be self-administered and that places constraints on the complexity of the instrument and features such as within household sampling and probing are not available. BLS decided against mail interviews as an

option as they do not include the careful probing from direct interaction with a trained interviewer that questions about occupational injuries and illnesses require. The idea of using mail only to determine eligibility of the household was entertained, but given the relatively high eligibility rate for the SOII Westat does not see many benefits to this type of screening approach using mail.

The distinction between prospective and retrospective interviews is another important branching for both the household-based study and the employer-based study. This is discussed in detail in Section 4.1 in the Cost-Benefit Analysis. In that report, it was argued that the prospective panel approach helped reduce the correlation between incidence and response (as in the prospective approach injuries will be in the future, and cannot influence response at the time of the survey contact). This is a much larger issue in the employer-based study than the household-based study. Since we are only considering a household-based study, the response propensity tradeoff between the prospective and retrospective studies reduces in importance, and the nonresponse associated with a followup interview under the prospective option pushes the advantage towards retrospective studies.

The time window is an important determinant of both quality and cost. Current research on recall error as discussed in the Section 3.1 of the Cost-Benefit Analysis shows that having a time window wider than three months will likely incur measurement error, especially for more minor injuries and illnesses. But having a short time window will mean a large number of required interviews (e.g., for a three-month time interval, four times as many interviews are required as for a twelve-month time interval). Based on these findings, BLS tentatively decided that a window shorter than three months (e.g., one month, two months) may not be economically feasible. The range of possible time windows for the SOII Employee Study that is deemed feasible is three months to twelve months.

Face-to-Face Sample Design

3

This section proposes a face-to-face design option. Given the expense of recruiting and fielding the households under this mode, a retrospective sample design is the most cost-effective (i.e., asking about incidences in the most recent time window up to the interview date). If a twelve-month window is used (all events in the past year), then the effective sample size of working adults should be 5,100. If a three-month window is used (all events in the past three months), then an effective sample size of working adults will be 20,400.

A face-to-face national probability design requires a multi-stage design. We recommend as a benchmark the design for the National Health Interview Survey, which is a three-stage design.¹ The goals of this study are similar to that of NHIS, making NHIS a good exemplar (if not in fact the parent survey for a modular or follow-on design), although some features may differ. The Primary Sample Units (PSUs) are counties or groups of counties. The sample of PSUs should be a stratified probability proportionate to size (PPS) sample, with the estimated number of workers as the measure of size for each PSU. The larger metropolitan PSUs will be self-representing (SR) and will be selected with certainty. The smaller PSUs will be placed into strata, and will be selected PPS without replacement. A systematic selection is common, but other methodologies can be used. The stratification structure for the first stage of selection should be selected carefully. The strata should be homogeneous in the primary characteristic being measured within strata, and heterogeneous in this characteristic across strata. In this case, the primary characteristic of interest is prevalence of occupational injuries and illnesses. An analysis of prevalence by geography using the SOII employer survey data may provide the right stratification structure. It may or may not be the case that prevalence varies across county-level PSUs in some kind of systematic way. It may even be possible to draw from the NHIS PSU sample (something that can be considered to save on development costs). The number of sampled noncertainty PSUs (NSRs) for the employee study should be large enough to guarantee adequate degrees of freedom for variance estimation. Sixty noncertainty PSUs

¹ The source for the NHIS design is Parsons, V. L., C. Moriarity, K. Jonas, et al. “Design and Estimation for the National Health Interview Survey, 2006-2015, National Center for Health Statistics. Vital Health Stat 2(165), 2014. Parsons, V.L. (2014), “Designing Flexibility for State Samples into the 2016 NHIS”, Proceedings of the Survey Research Methods Section, American Statistical Association, 3037-3043, provides an overview of the still unfinished sample design for the next decade NHIS cycle 2016-2025.

sampled from 30 strata is probably a minimum for the NSR PSU sample size (providing 30 degrees of freedom in variance estimation for this component of variance).²

An alternative ‘exemplar’ or even a source of final sample units is the Current Population Survey. The questionnaire for the SOII Employee Study will certainly deviate from the CPS questionnaire, but the target population for this study is very similar to that of the CPS (CPS targets the working-age population: employed and unemployed workers). The PSU structure is similar to that of NHIS (counties and groups of counties). CPS and the new cycle of NHIS have separate samples by states. This would also need to be ‘undone’ by subsampling if CPS or NHIS PSUs were used for the SOII Employee Study, as the SOII Employee Study should not have oversampling of small states. Of course, there are administrative details associated with using the CPS that BLS knows very well.

The next stage of selection is of second-stage units (SSUs) which are intermediate between the county-level PSUs and the households which are the final stage of selection. SSUs can be block groups, tracts, blocks, or modifications thereof. The choice of SSU should be determined by a trade-off between intra-SSU correlation and cost of fielding the sample. Smaller SSUs tend to have higher intra-SSU correlations, but lower costs. A full-scale optimality analysis could measure intra-SSU correlations and relative costs. NHIS and CPS both use Census blocks. In this case, NHIS or CPS can probably provide a sufficiently efficient exemplar that can be used for the pilot study. The information gathered from the pilot study about intra-stage correlations and cost ratios would then inform future employee studies. Alternatively, relevant intra-stage correlation coefficients can be computed at the NHIS PSU level by using data on prevalence of occupational injury and illness by PSU from a recent NHIS survey. This will require securing the files with all the needed data on geography (there may be confidentiality restrictions on getting PSU and SSU identifiers) and carrying through the appropriate analysis.

Westat would recommend exploring the possibility of using block groups rather than the traditional blocks. Block groups are less homogeneous, reducing the intra-cluster correlation. The larger listing cost can be offset by using address lists from the Postal Service (an ABS procedure) as the entire frame or with supplementation for missing data. For the urban block groups, the address lists from the post office will cover the block group quite well. The block groups will also increase data collection costs due to travel within the PSU for interviewers, but our experience is that this increase is not great in many surveys.

² Note that there needs to be certainty PSUs as well in addition to the minimum sixty noncertainty PSUs.

Table 3-1 illustrates a possible design based on CPS and NHIS designs. It assumes that 65 percent of the population falls into self-representing PSUs (the actual percentage varies based on the exact design and cannot be determined until frame development). A total of 1,348 SSUs are assumed, with 78 NSR PSUs. The NSR within-PSU correlation is assumed to be low, 0.005. This is likely realistic of occupational injury and illness prevalence, but an empirically based estimate should be computed eventually. The mean number of sampled SSUs per NSR PSU is assumed to be 6, with the mean number of interviewed households in both SR and NSR SSUs assumed to be 4. The within-SSU correlation is assumed to be 0.05. The cluster design effect for the NSR SSUs is approximately given by³ $1 + ab\rho_1 + (b - 1)\rho_2$, where ρ_1 is the intra-PSU correlation, ρ_2 is the intra-SSU correlation, b is the average number of responding households per SSU, and a is the average number of SSUs per PSU. For the SR PSUs it is simply $1 + (b - 1)\rho_2$, as the first stage of selection is eliminated. With all of these parameters defined, the overall design effect from clustering is 1.2. This is probably realistic for this study for the effect of clustering.

Table 3-1. Illustration of design and design effect calculations

	SR PSUs	NSR PSUs	Total
Percent of population	65%	35%	100%
Number of PSUs	NA	78	
Within-PSU correlation	NA	0.5%	
Number of SSUs	880	468	1,348
Number of SSUs per PSU	NA	6	
Number of intvd households	3,520	1,872	5,392
Number of intvd households per SSU	4	4	4
Within SSU correlation	5%	5%	5%
Design effect	1.15	1.27	1.192
Effective HH sample size	3,061	1,474	4,522

Table 3-2 illustrates a possible design which achieves the goal of an effective sample size of at least 5,100 (note this assumes a twelve-month recall period: a three-month recall will require a 4 times larger sample size). We adjust for the assumed design effect from clustering from Table 3-1, and also include an assumed design effect of 1.2 from weighting adjustments (e.g., nonresponse adjustments, adjustments from variable cluster sizes). The sample design assumed here includes interviewing up to two working adults per household, resulting in a mean number of 1.6 adults per household (see Section 5.1.1 in the Cost-Benefit Report). The adult response rate (conditional on the screener being completed) is assumed to be 85 percent, and the screener response rate 75 percent. The

³ Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953), *Sample Survey Methods and Theory, Volume I*, Equation (3.1), p. 370. New York: Wiley Classics Library.

household eligibility rate (households with at least one working adult) is assumed to be 80 percent, as in Section 5-1-1 of the Cost-Benefit Report. A total of 9,000 initially sampled households will lead to the desired effective sample size of 5,100.

Table 3-2. Possible face-to-face design with twelve-month recall period

Household sample size	8,987
Household response rate	75%
Household screener size	6,740
Household eligibility rate	80%
Eligible screener households	5,392
Mean sampled adults/eligible HHs	1.6
Sampled adults	8,627
Adult response rate	85%
Adult prevalence interviews	7,333
Design effect from weighting	1.2
Design effect from clustering	1.192
Effective sample size	5,125

Telephone Survey Sample Design

4

A telephone survey design at the national level can be a one-stage design (no PSUs or SSUs). Landline numbers should be sampled from an RDD list-assisted frame, and cellphone numbers from a cellphone exchange frame. There are relatively inexpensive commercial vendors who can provide access to the most recent versions of these frames, and allow for the draw of valid samples. Each cellphone interview is more expensive to collect, as each call has to be by a live interviewer. With the landlines, a ‘half-ring’ call can be made by automated dialers to establish that the number is a working number, and then a live interviewer only calls numbers which have been initially screened by the automatic dialer. The optimal design would not draw cellphone numbers in proportion to the population, but draw somewhat fewer cellphone numbers because of their higher cost. There is no clustering from PSUs and SSUs, but the sample sizes must be larger to offset the low response rates that are typical (20-25%).

Lohr and Brick (2014)⁴ present two types of dual-frame designs: a ‘screener design’ and an ‘overlap design’. The screener design screens out any cellphone numbers where the screener respondent indicates that the household has a landline number. The overlap design takes these numbers and interviews a sampled adult (usually there is only one adult associated with a cellphone number). Under the overlap design, there needs to be an adjustment for the multiple chances of selection of the household. The optimal design depends on the following factors:

- The percentages of households which are cellphone-only, landline-only, and both cellphone and landline (these numbers are in the series of papers published by Blumberg and Luke⁵ based on NHIS);
- The relative OII prevalence levels for cellphone-only, landline-only, and both cellphone and landline households;

⁴ Lohr, S., and J. M. Brick (2014). “Allocation for dual frame telephone surveys with nonresponse”. *Journal of Survey Statistics and Methodology* 2 (4), 388-409.

⁵ The most recent reference in this series is Blumberg, S.J., and J. V. Luke (2014). “Wireless substitution: early release of estimates from the National Health Interview Survey, January-June 2014”. National Center for Health Statistics. Available at <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201412.pdf>.

- The relative costs of cellphone and landline screeners and cellphone and landline completed interviews; and
- The relative response rates for cellphone and landline screeners and cellphone and landline completed interviews.

Based on the values of these key parameters and the relative prevalence levels in the three domains, Lohr and Brick (2014) can provide a methodology for choosing an optimal design. In the pilot study, rough guesses can be made for the relative prevalence levels, relative costs, and relative response rates.

We recommend a retrospective design here with a three-month or twelve-month window. A prospective design has a strong rationale in the context of an employer-based study as it decouples the response from prevalence (very important). In a household study, the nonresponse will not likely be correlated to prevalence if the questionnaire is worded carefully and the potential respondents can be convinced that their confidentiality will be protected. Any followup interviews in the context of telephone interviewing will likely simply be a source of considerable panel nonresponse on top of the considerable initial nonresponse. Thus we recommend the retrospective design over the prospective design for the telephone survey as well as the face-to-face survey.

Table 4-1 illustrates a telephone design for a retrospective design with a twelve-month window. A three-month window would require a four times larger sample size. The design effect from weighting is assumed to be the relatively large value of 1.3 as the nonresponse adjustments will need to be considerable in this case. We assume that one sampled adult will be sampled for prevalence interviews, in-person or by proxy. Unlike the face-to-face design, we take only one sampled adult because of the difficulty of reaching other household adults through cellphone numbers, which tend to be personal. A screening interview will be required to specify the working adults within the household, and to draw a sample of those adults. Another option is to sample only one working adult within the household. This will simplify the questionnaire and protocol for each sampled household, but will then require a larger sample size of households.

Table 4-1. Possible telephone design with twelve-month recall period

Household sample size	39,184
Household response rate	25%
Household screener size	9,796
Household eligibility rate	80%
Eligible screener households	7,837
Mean sampled adults/eligible HHs	1
Sampled adults	7,837
Adult response rate	85%
Adult prevalence interviews	6,661
Design effect from weighting	1.3
Effective sample size	5,124

We recommend a vigorous adaptive design approach for minimizing nonresponse bias under the telephone mode, to the extent this is possible. Incentives and extensive mailings preceding the initial telephone contact, and mailings following an unsuccessful initial telephone contact, for households in which there is an address linked to the (largely landline) telephone number.⁶ Noncontacts and initial nonrespondents should be followed up intensively. We also recommend an approach in which sampled households or persons who are nonrespondents after a given period of time are subsampled at a 50 percent rate for more intensive followup, dropping the other sample numbers from further followup. The concentration of interviewer efforts on a subsample of the initial nonrespondents and the other added field procedures discussed elsewhere should increase the response rate, appropriately weighted, over the response rate that would have been obtained with a non-adaptive design. We have used all of these methods but response rates are still relatively low as indicated earlier. The approach also results in an additional design effect from doubling the weights of the subsampled converted initial nonrespondents, but this can be limited by not subsampling until reasonable efforts are made on all cases. The technique of drawing followup subsamples from a set of initial nonrespondents dates back to Hansen and Hurwitz (1946).⁷

⁶ An excellent up-to-date, comprehensive meta-analysis of this is given in Mercer, A., Caporaso, Cantor, D. and Townsend, R. (2015), How much gets you how much? Monetary incentives and response rates in household surveys. *Public Opinion Quarterly* 79, 105-129. This is also discussed in Section 3.3 of the Cost-Benefit Report.

⁷ Hansen, M. H., and Hurwitz, W. N. (1946). The problem of nonresponse in sample surveys. *Journal of the American Statistical Association*, 41, 517-529.

Integration of the Employee Study with Larger National Studies

5

There are three plans which the Cost-Benefit Report had put forward for the SOII Employee household survey option: a stand-alone study, a module option, and a follow-on option. In the case of the module and follow-on options, a larger study needs to be the ‘host’ for the SOII Employee Study. As discussed in the Cost-Benefit Report, the best option for this appears to be the National Health Interview Survey (NHIS). NHIS is a face-to-face survey, so a face-to-face option for the SOII employee study is a good fit. It should be noted that NHIS for recent cycles had about 37,000 interviewed households.⁸ Our best information about the next decade’s NHIS cycle (2016-2025) is that there will be a ‘core’ national sample of about 25,000 interviewed households with no oversampling that can be used as a host for the SOII Employee Study. This provides enough NHIS households in one cycle for a national study to cover SOII Employee Study household sample of about 9,000 for a twelve-month recall period (see Table 3-2), but not quite enough for a three-month recall period (which requires 36,000 households).

If the telephone survey option is chosen for the SOII employee study, then only the stand-alone or follow-on options are available for the linking. The follow-on option becomes a telephone followup to the CPS or NHIS face-to-face interview. Table 4-1 indicates a household sample size of about 40,000, which would likely be beyond the available sample size from NHIS for one cycle. But this presupposes a response rate of 25 percent, which is what would be expected currently for stand-alone telephone surveys. But in this case, the household pool would be NHIS respondents, so the likely follow-on response rate might be higher. If the follow-on response rate can be edged up to 50 percent, then a household sample size base of 20,000 is sufficient, and this could be covered by NHIS. The telephone numbers of the respondents in the CPS or NHIS are collected but there will be some loss due to the delay in processing the data from that survey and not all the telephone numbers are captured in the NHIS.

All of the sample sizes in Sections 1 through 4 presuppose a simple national estimate of overall prevalence of occupational injuries and illnesses over a particular year period. If particular industries, occupational groups, or types of injuries/illnesses are specified to receive smaller CVs, then all of

⁸ See for example Parsons, V. L, C. Moriarity, K. Jonas, et al. “Design and Estimation for the National Health Interview Survey, 2006-2015, National Center for Health Statistics. Vital Health Stat 2(165), 2014, page 9.

the sample sizes will increase to meet these requirements. For example, if ‘Case 1’ and ‘Case 2’ injuries are designated to allow for 80 percent power for 20 percent differences in prevalence between the employer and the employee studies, then the sample sizes will need to be twice as large (effective sample sizes of 10,730), as given in Table 2-4 of the Cost-Benefit Report. A study on this scale will likely be too much for NHIS as a host survey (depending on the specified recall period), but CPS can still provide the necessary households. If there are enough small subgroups which are designated with high power requirements it may be too large even for CPS as a host survey.

One option is to combine samples across cycles for subgroups (adding together the samples for the subgroup from two, three, four cycles). This achieves the power goal, but the estimate now refers to a mean value of prevalence over several years, rather than a single year’s estimate.

While the mechanics of setting up and operating with a study like the NHIS is an administrative detail outside of our area, we expect that the pilot study will be stand-alone, even if the final SOII employee household survey may be a module or a follow-on option. This provides some complications in the development of the questionnaire and protocols. The pilot study will have differences with the later main household survey that will impact the questionnaire and protocols. This seems unavoidable.

Questionnaire Development and Proxies

6

Primary issues in questionnaire development are the need to specify the type of injury or illness, which will require a complex branching questionnaire structure, with probes and clarifications required to lead the sampled person (or proxy) to the appropriate answers. Also required is a specification of both the industry and the occupational subgroups. Specifying industry and occupational subgroup is necessary to allow for post-survey weighting adjustments in addition to being needed by analysts.

Another important issue is recall. If a twelve-month window is utilized, prompting recall becomes very important. There is an extensive literature regarding the prompting of recall of calendar-based events that should be explored. It is important to maximize recall on the part of the respondent as to their occupational injuries and illnesses, and to make sure the events do in fact fall within the time window. This is even of greater concern when proxies need to be utilized, as proxies may forget quickly about less serious occupational injuries and illnesses, or more readily get the time of the occurrence wrong. One possible option would be to randomize the sample into two subsamples: one to receive a twelve-month recall time-window and one to receive a three-month recall time-window. Ideally, each branch should have an effective sample size of 5,100 on its own to support a national estimate on its own.

It is important to allow for ‘family respondents’ (proxy response) to allow for those persons who are in hospitals or rehabilitation facilities due to injuries and illnesses suffered in the time window to be covered in the retrospective interview.

The Cost-Benefit Analysis Report discusses the ‘prevalence’ and ‘incidence’ interviews. The prevalence interview specifies that an occupational injury or illness has likely occurred to a sampled adult in the time window. The incidence interview then confirms this and asks for details about it. In some cases, the incidence interview can follow directly as another module in the face-to-face or telephone contact. In other cases, the incidence interview will need to be scheduled as a separate interview. The key design decision is whether to accept household respondent reports as proxies for reporting of incidents of other members. We encourage this approach in the pilot test, but would propose a large enough subsample of adults who have no reported incidents as reported by a proxy to be sampled for estimating the error rates associated with the proxies. Given the low prevalence

rate we can expect among the persons reported for by proxies, this subsample unfortunately needs to be quite large to provide any kind of power to distinguish prevalence. This needs to be worked out carefully in the questionnaire and protocol development.

Sampling weights should be developed for unbiased estimation from the pilot study. These sampling weights should begin with base weights: reciprocals of the probability of selection of each sampled adult. For the face-to-face mode, this includes PSU, SSU, household, and adult sampling probabilities. For the telephone mode, this includes household and adult sampling probabilities, and adjustments for multiple household telephone numbers. Since the telephone survey is a dual frame survey a composite estimation scheme is needed to deal with the overlap (if screening for cell phones is not done). The composite estimation scheme we would recommend is described in detail in Brick et al, (2011).⁹

Nonresponse adjustments will be necessary for both the face-to-face and the telephone modes, though the stakes will be higher for the telephone mode with its lower response rates. Nonresponse weighting adjustment cells should be defined based on characteristics known at the sample level which are both related to response propensity and to prevalence. Calibration adjustments should be considered so that the weights match control totals for total employees by industry and occupational group. This will minimize biases from getting too many or too few employees in particular industries or occupational groups. Control totals can come from other BLS employment studies. The questionnaire needs to accurately assign sampled working adults to their industry and occupational group for this adjustment to be accurate.

We prefer replicate weights be generated using balanced repeated replication (BRR) or the jackknife. The replicate structure is considerably different between face-to-face and telephone modes, given the very different sample designs. Taylor series method variance estimation should also be accommodated even though it may not capture all the components of variance associated with nonresponse and calibration.

⁹ Brick, J. M., I. F. Cervantes, S. Lee, & G. Norman (2011). Nonsampling errors in dual frame telephone surveys. *Survey Methodology*, 37(1), 1-12.

Information to Collect from the Pilot Study and Preliminary Activities

8

The key information that will be collected from the pilot study is the national occupational injury and illness prevalence rate for the time window covered by the study, and this can then be compared to the SOII Employer Survey prevalence. This is the primary purpose of the SOII Employee Study, and its success or failure will be determined by how well it achieves this primary purpose.

Other information of importance that should be collected from the pilot includes:

- Computation of response rates and response rate differences across subgroups (“R-indicators”);
- Differences between recall periods: whether three-month and twelve-month recall periods give systematically different prevalence rates, and how these differentials differ by injury type and by other subgroups;
- Correlation coefficients for prevalence: within household;
- For the face-to-face option, within-PSU and within-SSU correlations for prevalence;
- For the face-to-face option, relative costs for PSUs, SSUs, and households;
- For the telephone option, relative costs for landline and cellphone numbers;
- For the telephone option, relative prevalence rates for landline and cellphone numbers;
- For the telephone option, relative response rates for landline and cellphone numbers; and
- Effects of proxy interviews: how well proxies report on prevalence as compared to the sampled adult themselves.

In some cases, information can be collected from preliminary activities preceding the pilot study (pre-pilots, focus studies, etc.). It is likely that the sample sizes will be small for these preliminary activities, but they can shed some light on some of the issues. The gold standard for evaluating competing options is a fully randomized approach. Under full randomization, a randomly selected set of sample units receive one branch, and the complement set receives another branch. (Note that here can be more than two branches.). Randomization can be used for example to determine effects of differing time windows for recall periods (three months vs. twelve months), and to determine

differences from proxies. Randomization can also be utilized to decide upon the working adult sample size within each sampled household (one, two, or more working adults as a maximum sample size). The sample sizes need to be large enough however to provide for sufficient power. The pilot study itself will have these large sample sizes.