Survey Design and Questionnaire for a Household Survey on Occupational Injuries and Injuries

Report on Suitability of Existing Surveys and Frames (Deliverable #5)

JANUARY 11, 2016

PRESENTED TO:
U.S. Bureau of Labor Statistics
2 Massachusetts Ave, NW
Washington, DC 20212

PRESENTED BY:
NORC at the University of Chicago
Kennon R. Copeland, PhD
Senior Vice President and Director
Statistics and Methodology
4350 East West Highway
Bethesda MD 20814
Phone (301) 634 9347

NORC at the UNIVERSITY of CHICAGO

Table of Contents

Executive Summary **1**

1. Introduction **3**

2. Background **4**

2.1 Design Options 4

2.2 Assessment Criteria 5

3. Review and Assessment Scope **6**

4. Assumptions **8**

4.1 Eligible Population 8

4.2 Within-HU Sampling 8

4.3 Minimum Sample Size 9

4.4 Potential Maximum Sample Size 9

4.4 Impact of I/O Sample Requirements on Possible Maximum Sample Size 11

5. Findings **13**

5.1 Population Representation 13

5.2 Survey Mode and Use of Proxy Respondents 19

5.3 Availability of Key Information 21

5.4 Timeliness of Calendar Year Estimates 25

5.5 Availability of Auxiliary Data for Sample Design Efficiency 27

6. Discussion **28**

6.1 Screening vs. Add-on Module 28

6.2 Survey Costs 31

7. Summary **37**

Appendix A. Detail About Design Options **39**

Appendix B. Links to Home Pages for Existing Surveys **46**

Appendix C. Definitions for Criteria Ratings **47**

List of Tables and Exhibits

Table 3.1: Existing surveys under consideration 6
Table 3.2: List of sampling frames used for existing surveys under consideration 7
Table 4.4.1: Population data, distributions, incidence rates for 7 industry group illustration 10
Table 4.4.2: Population data, distributions, incidence rates for 6 industry group illustration 11
Table 5.1.1: Profile of existing survey coverage..... 14
Table 5.1.2: Profile of existing survey sample size and response rates 16
Table 5.1.3: Summary ratings for sample frame coveage..... 17
Table 5.1.4: Summary ratings for existing survey coverage, sample size, and response rates 18
Table 5.2.1: Profile of existing survey data collection mode and rules 19
Table 5.2.2: Summary ratings for existing survey data collection rules 21
Table 5.3.1: Profile of existing survey employment information 22
Table 5.3.2: Profile of existing survey injury/illness information 23
Table 5.3.3: Summary ratings for existing survey employment and industry/occupation information 24
Table 5.4.1: Profile of existing survey frequency, respondent burden, and timing 25
Table 6.2.1: Preliminary assessment of relative costs for existing surveys and design options 36
Table 7.1: Summary assessment of use of existing surveys and sampling frames relative to design options 1-4 and single frame vs. multi-frame and/or multi-year design 37

Executive Summary

The Bureau of Labor Statistics is seeking to develop a household survey on nonfatal occupational injuries and illnesses (HSOII) to provide information on the level of underestimation of incidence rates from the Survey of Occupational Injuries and Illnesses and to provide information relative to the potential reasons for the underestimates. This report provides NORC's assessment of the suitability of existing surveys and sampling frames for potential use for a household occupational injury and illness module or follow-on survey. The key limiter in terms of what existing surveys could be used as a single sampling frame for the HSOII is sample size, given the need to provide sufficient sample to allow estimates by selected industry and occupation groups. Other factors such as respondent rules, availability of information on employment/industry/occupation/injuries/illnesses, and timing of data collection and release also affect the extent to which an existing survey is suitable for use as a sampling frame for the HSOII.

The least expensive option for implementing the HSOII would be to add a supplemental module to an existing survey (referred to as Option 1 in this report), as contact/response conversion efforts and associated costs are already covered through the existing survey. Only three existing surveys – the American Community Survey (ACS), the Current Population Survey (CPS), and the Behavioral Risk Factor Surveillance System (BRFSS) – offer sufficient sample sizes for such a design. Adding a supplemental HSOII module to the ACS raises substantial feasibility issues in obtaining approvals, while adding to the BRFSS raises substantial operational issues given the federal-state cooperative nature of the survey and states' determining which modules to include. Thus, CPS appears to be the only existing survey that can be considered for use as a single frame approach in which a supplemental HSOII module is added to an existing survey. CPS offers other advantages as well, such as currently collecting employment/industry/occupation data and having short data collection and release periods.

A second, but more expensive option would be to utilize data from an existing survey to screen and select sample for a HSOII conducted as a follow-up survey to the existing survey (referred to as Option 2 in this report). For this option, CPS, ACS, and BRFSS again offer sufficient sample sizes for consideration; however, employment data for BRFSS are collected only if a state-optional module is implemented and thus it would be more difficult to efficiently identify HSOII-eligible individuals from BRFSS. A refinement to this option, which would add the National Immunization Survey (NIS) as a candidate survey in light of its large sample size of selected housing units, would be to add HSOII-specific screener questions to an existing survey and utilize the expanded existing survey data to screen and select sample for a HSOII conducted as a follow-up survey to the existing survey (referred to as Option 3 in this report).

A third, but still more expensive, option would be to develop a sample design specific to HSOII utilizing an existing sampling frame, such as Census Master Address File, United States Postal Service Delivery Sequence File, or landline and telephone numbers (referred to as Option 5 in this report). This option would not have access to information on employment/industry/occupation for housing units and would thus result in a less efficient sample design than that possible under Options 2 or 3. Given the availability of existing surveys under Options 1-3 and the higher costs with developing a sample from a sampling frame rather than from an existing survey, this option does not appear to offer advantages to warrant its recommendation.

Refinements to the selected option, through use of multiple surveys (referred to as Option 4 in this report), implementation of multi-year surveys (referred to as Option 6 in this report), and use of panel surveys (referred to as Option 7 in this report), could be considered made to reduce costs and/or recall error. There appear to be only four other existing surveys sample sufficient to be considered for Option 4: National Health Interview Survey (NHIS); Survey of Income and Program Participation (SIPP); AmeriSpeak (AS); and KnowledgePanel (KP), although issues of coordinating implementation across multiple surveys make such an approach less desirable should other alternatives be available. Options 6 and 7 appear to be the most promising refinements for Options 1-3, offering the ability to target smaller annual sample sizes and thus lower annual costs with a modification to the concept of the time period represented by annual estimates (Option 6) or allowing multiple contacts with sample HUs to reduce data collection costs and reduce recall error.

1. Introduction

The Survey of Occupational Injuries and Illnesses (SOII) is the primary source of information on nonfatal workplace injuries and illnesses. Existing research points to an underestimate of injuries and illnesses in the SOII attributed to a variety of factors including an employer/employee incentive to underreport these occurrences.

One way to avoid the filtering effect of collecting establishment data on employee injuries and illnesses is to collect data directly from workers in a household survey. In such a survey, workers eligible for the survey would be identified outside the sphere of the employer. The Bureau of Labor Statistics (BLS) has contracted with NORC at the University of Chicago (NORC) to provide assistance in the development of a survey design and questionnaire for a household survey on nonfatal occupational injuries and illnesses.

This report provides NORC's assessment of the suitability of existing surveys and sampling frames for potential use for a household occupational injury and illness module or follow-on survey. The assessment includes such factors as frame coverage, subject matter overlap, data collection mode, timeline, and cost. This report is intended to provide BLS with a comprehensive review of the key aspects of the existing surveys and sampling frames as they relate to the target population for a household survey of occupational injuries and illnesses (abbreviated in this report as HSOII), along with summary comparisons of the alternatives. This report is not intended to provide recommendations to BLS.

The report begins with background for the problem at hand – key issues associated with the design of a HSOII, including likely design options and the key assessment criteria used in this report. The set of existing surveys and sampling frames considered within this report is then identified. Assumptions made while carrying out the review and assessment of existing surveys are then presented. Detailed and summary information for each criterion is then presented, along with a discussion of implications as they relate to the HSOII. A summary discusses the relationship between the design options, and use of existing surveys, as well as the relative costs associated with the design options and use of existing surveys and sampling frames.

2. Background

The sample design for the HSOII must result in a nationally representative household survey of workers that allows estimates by employment relationship (employee vs. self-employed), sector (private vs. government), industry, and occupation meeting publishability/quality criteria for the HSOII.

Total sample size requirements are not addressed in this document, other than a discussion of relative magnitudes that would likely be required. For discussion purposes in this report, the assumption will be made that each sampled person will provide a full year of data for the survey, although we recognize other designs may be considered. This aspect will be addressed later, along with development of recommended data collection procedures.

2.1 Design Options

Rather than developing a new survey design and sample for the HSOII, BLS would like to consider cost-efficient options that leverage existing surveys and/or sampling frames. Given this, several design options could be considered for the HSOII[1]. (See Appendix A for additional detail.)

1) Add a supplemental module to an existing survey. Under this option, BLS would work with the survey sponsor to add a module to the existing survey questionnaire to collect any additional data required for the HSOII.
2) Use data from an existing survey to screen for persons eligible for the HSOII, select a subsample of HSOII eligible persons, and conduct a follow-on survey to collect any additional data required for the HSOII.
3) A variant of Option 2 would be to add appropriate supplemental questions to the existing survey questionnaire to aid in screening for or stratifying persons eligible for the HSOII. As in Option 2, the enhanced data from the existing survey would be used to screen for persons eligible for the HSOII, select a subsample of HSOII eligible persons, and conduct a follow-on survey to collect any additional data required for the HSOII.
4) Utilize multiple surveys under Options 1, 2, and/or 3.
5) Utilize an existing sampling frame to develop an optimal sample design for the HSOII.
6) Implement any of Options 1-5 as a multi-year survey, analogous to the design utilized for the ACS, in which independent samples are interviewed each year and the data aggregated to create multi-year estimates. Such a design could support annual estimates at the national level and multi-year estimates for lower-level estimates, such as by industry and occupation.
7) Implement any of Options 2-6 as a panel survey, analogous to the rotation group design utilized for the CPS, in which sample persons are interviewed in more than one year. Under such an approach, persons sampled for the HSOII would be eligible for interview in multiple years. Note

[1] Options 1, 2, and 7 were identified in Survey of Occupational Injuries and Illnesses: Employee Survey Cost Benefit Analysis, May 12, 2015.

that this option cannot be used with Option 1 (Supplemental Module in Existing Survey) unless the existing survey utilizes a panel design.

2.2 Assessment Criteria

As stated in the RFP, assessment of HSOII sample frame options must take the following factors into consideration (reordered to reflect the sequence in which they are discussed within this document):

1. Does the survey frame or survey provide an adequate (i.e. unbiased) representation of the population of workers that are in scope for the proposed survey?
2. Is the mode conducive to asking questions that may involve recall bias?
3. Does the respondent selected to respond to the survey also serve as a proxy respondent for others in the household?
4. For existing surveys, is any of the information we wish to collect on occupation, industry, employment relationship, and/or occupational injuries and illnesses already collected?
5. Will the frame permit producing calendar year estimates in a time frame similar to that currently used by SOII?
6. Does the survey frame contain variables that would be useful (i.e. increase sampling efficiencies) for stratification?
7. Would the answers to questions 1-6 change if the survey is merely used to screen for potential respondents rather than including a full module of 15-20 questions on occupational injury and illness?

While information relevant to Factors 1-6 are presented in Section 5: Findings, Factor 7 is addressed, along with information on survey costs, in Section 6: Discussion.

3. Review and Assessment Scope

NORC has identified a set of fourteen existing surveys to consider for use in the HSOII using information provided by BLS[2] and NORC knowledge of government and non-governmental surveys. The final list includes the six surveys listed by BLS (CPS, NHIS, BRFSS, MEPS-HH, NHANES, and NLSY), five additional government surveys (ACS, NIS, SIPP, SCF, and NSFG), the General Social Survey (GSS), and two probability-based online panels (POPs) offered by NORC (AmeriSpeak, abbreviated here as AS) and GfK (KnowledgePanel, abbreviated here as KP). Table 3.1 lists the existing surveys, along with the organization conducting the survey, and the sampling frame used for the survey[3].

Table 3.1: Existing surveys under consideration

Survey	Organization	Sampling Frame
American Community Survey (ACS)	Census	Census Master Address File (MAF)
Current Population Survey (CPS)	BLS, Census	Census Master Address File (MAF)
National Health Interview Survey (NHIS)	NCHS, Census	Census data for selecting segments; Listing of housing units (HUs) within sampled segments
Behavioral Risk Factor Surveillance System (BRFSS)	CDC, States	RDD landline and cell phones
Medical Expenditure Panel Survey - Household Component (MEPS-HC)	AHRQ	NHIS Respondents
National Health and Nutrition Examination Survey (NHANES)	NCHS	Census data for selecting segments; Listing of dwelling units (HUs) within sampled segments
National Longitudinal Survey of Youth (NLSY)	BLS	Census data for selecting segments; Listing of housing units (HUs) within sampled segments
National Immunization Survey (NIS)	CDC, NORC	RDD landline and cell phones
Survey of Income and Program Participation (SIPP)	Census	Census Master Address File (MAF)
Survey of Consumer Finance (SCF)	FRB, NORC	US Postal Service Delivery Sequence File (USPS DSF); IRS tax return administrative listing
National Survey of Family Growth (NSFG)	NCHS, ISR	US Postal Service Delivery Sequence File (USPS DSF)
General Social Survey (GSS)	NORC	US Postal Service Delivery Sequence File (USPS DSF)
AmeriSpeak (AS)	NORC	US Postal Service Delivery Sequence File (USPS DSF)
KnowledgePanel (KP)	GfK	US Postal Service Delivery Sequence File (USPS DSF)

In addition, as requested by BLS during an early project status meeting, NORC has identified the set of sampling frames used for the existing surveys. Table 3.2 lists the six sampling frames, noting the existing surveys using the sampling frame.

[2] SOII Research on Data Collection from Employees Literature Review, January 15, 2015

[3] Links to home pages for each of the existing surveys listed in Table 3.1 are provided in Appendix B.

Table 3.2: List of sampling frames used for existing surveys under consideration

Sampling Frame	Surveys
Census Master Address File (MAF)	ACS, CPS, SIPP
Census Data plus Listing within Sampled Segments	NHIS, NHANES, NLSY
US Postal Service Delivery Sequence File (USPS DSF)	GSS, SCF, AS, KP
RDD Frame of both Landline and Cellular Numbers	BRFSS, NIS
IRS tax return administrative listing	SCF
NHIS	MEPS-HC

IRS tax return administrative data listing, which is utilized along with USPS DSF for SCF, is excluded for consideration as a sampling frame. NHIS, which serves as the sampling frame for MEPS-HC, is captured in consideration of the existing surveys. Thus the first four sampling frames listed are considered here.

4. Assumptions

We present here assumptions that were made for purposes of discussion in this document, along with discussion around derivation of these assumptions.

4.1 Eligible Population

For purposes of this report, it is assumed the target population for the HSOII is workers aged 16+ years, with a worker being anyone who worked in at least one of the prior 52 weeks. Sample counts for existing surveys represent the total number of persons for which a completed interview was obtained, regardless of employment status. Thus, these counts must be scaled to the HSOII target population of workers. It is assumed that 65.1% of the population aged 18+ is eligible for the HSOII, which was determined in the following manner.

Data from the 2014 ACS show that 75.1% of the U.S. population aged 16-64 worked at least one of the prior 52 weeks. This is as compared to November 2015 CPS data reporting 69.0% of the U.S. population aged 16-64 was employed the week containing November 12, 2015. While the difference between the two estimates is a function of the differing collection periods (calendar year 2014 vs Nov 2015), a more important factor is the reference period (employed in the 12 months vs. employed in the week containing Nov 12, 2015). The Nov 2015 CPS ratio of Civilian Labor Force (CLF) to Population (72.5%) for persons aged 16-64 is closer (difference of -2.6 percentage points) to the ACS result. As the Nov 2015 CPS ratio of CLF to Population for persons aged 16+ is 62.5%, adding the 2.6 percentage point difference yields the assumed worker eligibility rate of 65.1% for persons aged 16+. For existing surveys in which the sample population is other than persons aged 16+, similar calculations were used to determine the eligibility rate for the covered population.

4.2 Within-HU Sampling

As little or no correlation among persons within a HU would be expected for the HSOII, it would be most cost-efficient to sample all eligible persons within a HU. The assumption is made that all eligible persons within a HU would be selected. With the exception of designs that are limited by the information available for sampled persons within a sampled HU (such as Option 1 (Supplemental Module in Existing Survey) and possibly Option 2 (Follow-on to Existing Survey, no Change to Existing Survey) for some existing surveys), we assume all persons within a HU eligible for HSOII will be included in the sample.

4.3 Minimum Sample Size

As discussed in the background material provided by BLS[4], an effective sample size of 5,100 person-years should be targeted to achieve 80% power for detecting a 20% difference from the employer-based SOII (abbreviated here as ESOII) at the national level based on current overall prevalence levels. In addition to quantifying the under-reporting of injuries and illnesses in the ESOII, BLS is also interested in better understanding where under-reporting occurs. This will require a larger total sample size, with determination of specific population subgroups for which larger sample sizes are to be targeted than that expected from an overall effective sample size of 5,100 person-years.

Assuming a required effective sample size of 5,100 to provide adequately powered estimates at the national level for the overall incidence rate and recognizing possible design effects[5], the minimum number of completed interviews with workers is 6,630 for a CATI survey and 7,295 for a CAPI survey[6].

We might assume a maximum 75% yield rate of completed interviews from persons sampled from existing surveys, which takes into account person level response rates along with possible eligibility rates (which must be considered as either the information available from the existing survey does not provide the ability to identify all and only eligible workers, or the sampled persons' eligibility status may have changed). This results in minimum numbers of sampled persons of 8,840 for a CATI survey and 9,277 for a CAPI survey.

Thus, the minimum number of sampled persons is 1.73 (CATI) to 1.91 (CAPI) times the effective sample size. As these are likely minimums, for ease of conversion, we assume a ratio of 2.0. These values will be further examined and determined during the course of the project as survey design recommendations are developed.

4.4 Potential Maximum Sample Size

While not an objective of this project, there are several approaches which could be taken to identifying specific subgroups for oversampling and determining the target sample sizes for the subgroups. One approach, which would result in the largest overall effective sample size need, would be to target achievement of 80% power for either detecting a 20% difference from the employer-based SOII

[4] Survey of Occupational Injuries and Illnesses: Employee Survey Cost Benefit Analysis, May 12, 2015.

[5] Within the Pilot Study Design for the SOII Employee Study report (September 1, 2015), possible design effects were listed as 1.3 (due to weighting) for a telephone (CATI) survey and 1.43 (the product of 1.2 due to weighting and 1.192 due to clustering) for a face-to-face (CAPI) survey.

[6] Estimated minimum number of completed interviews calculated as the minimum effective sample size (5,100) as reported in the Survey of Occupational Injuries and Illnesses: Employee Survey Cost Benefit Analysis (May 12, 2015) and the possible design effects (1.3 for a CATI survey and 1.43 for a CAPI survey), as reported in the Pilot Study Design for the SOII Employee Study report (September 1, 2015).

(abbreviated here as ESOII) or a difference of 0.7 percentage points from the ESOII, based on current prevalence levels at the industry sector and occupational group level. A very rough sample size estimate can be made assuming effective sample size needs of 5,100 for each subgroup[7], which would yield a total effective sample size need of 102,000 to support the 20 industry sector levels and of 117,300 to support the 23 major occupational groups, implying a total effective sample size need well above 100,000 to support both industry and occupation (this sample size would also support employment relationship and sector), with total number of completed interviews above 130,000-145,000 and a total number of sampled persons in excess of 200,000.

An alternative to the most expansive approach would be to collapse industry and occupation to derive a smaller set of subgroups, say 5-10 each. This would yield total effective sample size needs on the order of 25,000-50,000 (this sample size would also support employment relationship and sector), and a total number of sampled persons on the order of 50,000-100,000.

As an illustration, provided in Table 4.4.1, one could consider defining the seven collapsed industry groups, based upon relative population size and 2014 incidence rates. This grouping would ensure each industry group has at least 10% of workers, and would yield five groupings with 2014 incidence rates about the overall incidence rate. Total effective sample size required would be on the order of 36,000, and a total number of sampled persons on the order of 72,000.

Table 4.4.1: Population data, distributions, incidence rates for 7 industry group illustration

	Nov 2015 Employment		2014 Incidence Rate		
Industry Group	**Count**	**Proportion**	**Group**	**2 digit Min**	**2 digit Max**
Agriculture & Goods-producing[1]	21,960	15.4%	3.8	3.6	5.5
Trade & Transportation[2]	26,507	18.6%	>3.6	2.9	4.8
Professional & Business Services	19,973	14.0%	1.5	0.9	2.6
Health Care and Social Assistance	18,811	13.2%	4.5	N/A	N/A
Leisure & Hospitality	15,330	10.8%	3.6	3.5	4.2
Other Service-producing[3]	20,598	14.5%	<2.6	0.7	2.6
State & Local Government	19,240	13.5%	5.0	4.1	5.4
Total Non-farm	142,418	100.0%	3.4		

[1]Agriculture & Related Industries (employment from CPS), Mining/Logging, Construction, Manufacturing,
[2]Wholesale Trade, Retail Trade, Transportation
[3]Utilities, Information, FIRE, Education, Other Services

A third alternative would be to identify selected industry and occupations which might most likely be associated with under-reporting that is of practical significance as well as statistical significance, and are

[7] Actual sample size needs would be derived to power detecting differences for the subgroup based upon the current prevalence level for the subgroup.

of sufficient size to allow for adequate sample. One could focus on those industries and occupations with high rates of injuries and illnesses, under the assumption that the magnitude of under-reporting of industries and occupations with low rates would be relatively small. For example, the 2014 incidence rate for Finance and Insurance was 0.7. Assuming a 20% under-reporting of injuries and illnesses for this industry, the actual rate would be 0.9, which could be considered not of practical significance. Conversely, the 2014 incidence rate for Health Care and Social Assistance was 4.5; a 20% under-reporting of injuries and illnesses would mean an actual rate of 5.6, a much more dramatic difference in level. Again, this sample size would also support employment relationship and sector.

As an illustration, provided in Table 4.4.2, one could identify the five industry groups with an incidence rate above the overall average of 3.4 and which contain at least 10% of total workers, based upon relative population size and 2014 incidence rates. A sixth grouping would consist of the remaining Service-providing industries. Total effective sample size required would be on the order of 31,000, and a total number of sampled persons on the order of 62,000.

Table 4.4.2: Population data, distributions, incidence rates for 6 industry group illustration

	Nov 2015 Employment		2014 Incidence Rate		
Industry	**Count**	**Proportion**	**Group**	**2 digit Min**	**2 digit Max**
Agriculture & Goods-producing[1]	21,960	15.4%	3.8	3.6	5.5
Retail Trade & Transportation[2]	20,570	14.4%	N/A	3.6	4.8
Health Care and Social Assistance	18,811	13.2%	4.5	N/A	N/A
Leisure & Hospitality	15,330	10.8%	3.6	3.5	4.2
Other Service-producing[3]	46,608	32.7%	N/A	0.7	2.9
State & Local Government	19,240	13.5%	5.0	4.1	5.4
Total Non-farm	142,518	100.0%	3.4		

[1]Agriculture & Related Industries (employment from CPS), Mining/Logging, Construction, Manufacturing,
[2]Retail Trade, Transportation
[3]Wholesale Trade, Utilities, Information, FIRE, Professional & Business Services, Education, Other Services

4.4 Impact of I/O Sample Requirements on Possible Maximum Sample Size

When considering the adequacy of the sample sizes for the surveys under consideration (this factor does not affect assessment of the sampling frames), it must be remembered that the sample distribution for the existing surveys is predetermined based upon the sample design of the survey and cannot be adjusted to fit the needs of the HSOII. In other words, given the sample designs for the surveys are focused on providing general population estimates (whether for the full population of persons aged 18+ years, or for some selected age subgroups), the sample distribution relative to industry and occupation can be reasonably expected to be similar to the distribution of the total population. There will be some

deviations based upon oversampling of selected populations, if utilized by the survey, but the oversampled populations tend to be minorities and the elderly, which should not dramatically affect the industry and occupation distributions. Thus, in assessing the adequacy of the survey to provide sufficient sample for selected or collapsed industries and occupations, one cannot compare the total sample size need for the HSOII with the total sample size of the survey.

To achieve the roughly 62,000 sampled persons from the immediately prior illustration, we make an assumption for discussion purposes within this document that an effective sample size of 5,100 is needed for each grouping, and a total number of sampled persons from each grouping on the order of 10,200. If sampling from an existing survey of the full population, the existing survey would need to have sufficient numbers of workers to fill sample size requirements for all industry groups. Given the proportion of workers within each industry group (with a minimum value of 10.9%), the existing survey would need to have ~100,000+ workers for an expectation of meeting the sample size requirements for the six industry groups.

5. Findings

In this section we address each of the assessment criteria, presenting detailed information on sample design and data collection for each existing survey and sampling frame from documentation available online, and discussing the implications of that information as it relates to an HSOII.

To aid in summarizing the information on the existing surveys, topline "ratings," ranging from 0 (zero, lowest rating) to +++++ (five, highest rating) were developed for the assessment criteria and are presented here. These topline "ratings" are intended to provide a quick view of the relative value of each survey for the criteria. Appendix C provides the definitions used in the ratings for all the criteria.

5.1 Population Representation

Population representation can be viewed not only in terms of potential for bias (i.e., extent and nature of undercoverage of the population of interest), as referenced in the RFP, but also in terms of adequacy of sample size for generating reliable estimates (i.e., of sufficiently small variance) for subpopulations of interest.

Sampling Frame and Survey Coverage Information

Undercoverage for a sampling frame refers to the extent of the population of interest which is missing from the frame. When considering use of an existing survey as the frame for the HSOII, there can also be undercoverage due to population exclusions made as part of the survey data collection process.

Table 5.1.1 provides information on coverage for the surveys under consideration – the sampling frame used by each survey, known coverage gaps in the frame, population exclusions that are made during data collection, and an estimate of the resultant coverage of the target population of workers (based upon the frame exclusions and population exclusions).

Table 5.1.1: Profile of existing survey coverage

Survey	Sampling Frame	Coverage		
		Frame Coverage Gaps	Survey Exclusions	Estimated Coverage of Workers in Eligible Sample
ACS	Census Master Address File (MAF)	None	None	100%
CPS	Census Master Address File (MAF)	None	Military, Institutionalized Population	100%
NHIS	Census Data for selecting segments; HU Listing for sampled segments	None	Military, Institutionalized Population	100%
BRFSS	RDD frame of both landline and cellular numbers	HU's without phones	Institutionalized Population	~97%
MEPS-HC	NHIS Respondents	None	Military, Institutionalized Population	100%
NHANES	Census Data for selecting segments; HU Listing for sampled segments	None	Military, Institutionalized Population	100%
NLSY	Census Data for selecting segments; HU Listing for sampled segments	None	Only includes NLSY-97: US residents born 1980-1984 NLSY-79: US residents born 1957-1964	~27%
NIS	RDD frame of both landline and cellular numbers	HU's without phones	Only includes HU's with children aged 6 mo - 17 years	~40%
SIPP	Census Master Address File (MAF)	None	None	100%
SCF	USPS DSF	None	None	100%
NSFG	USPS DSF	None	Only includes persons aged 15-44 years	~50%
GSS	USPS DSF	None	Non-English/Spanish	98%+
AS	USPS DSF	None	Non-English/Spanish	98%+
KP	USPS DSF	None	Non-English/Spanish	98%+

NOTES

Coverage is relative to the U.S. population of all workers, including those who work as independent contractors

Frames utilizing the Census Master Address File (MAF), Census data plus listing of selected segments, and the US Postal Service Delivery Sequence File (USPS DSF) offer essentially complete coverage of the population of interest, assuming listing is carried out in areas where the DSF addresses do not provide sufficient information to locate the address (~2% of all HUs). Surveys utilizing these sampling frames that have no exclusions from the population of interest, as is the case for ACS, CPS, NHIS, NHANES, SIPP, and SCF (military and institutionalized population are not in scope for the HSOII), would thus provide unbiased representation of the population of interest for the HSOII.

RDD landline and cell phone sampling frames lack coverage for the roughly 2.9% of adults living in households with no telephone[8]. Persons living in households with no telephone tend to be younger, minority, renters, lower income, and lower education[9].

As mentioned above, survey exclusions may also contribute to undercoverage of the target population of workers. GSS, AS, and KP include only English and Spanish speakers. Census data on languages spoken

[8] Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, July–December 2014. National Center for Health Statistics. June 2015. Available from: http://www.cdc.gov/nchs/nhis.htm.

[9] Khare, M. and Chowdhury, S. An Evaluation of Methods to Compensate for Noncoverage of Nontelephone Households using Information on Interruptions in Telephone Service and Presence of Wireless Phones. Proceedings of the Section on Survey Research Methods, Joint Statistical Meetings, American Statistical Association. 2006. 3221-3228.

and degree of English language proficiency from the ACS suggest such an exclusion would affect less than 2% of the population[10]. Eligibility for NSFG is limited to persons aged 15-44 years, and thus represent approximately 50% of the full HSOII target population of workers aged 18+. While the sample frame used for the NIS is nearly complete, the target population is households with children, resulting in a net coverage of the eligible sample on the order of 35% of the full HSOII target population of workers (2014 ACS data show 31.3% of households contain a related child; however, such households will have younger adults who are more likely to be workers than the full 18+ population; thus it was assumed to bump up the coverage of eligible workers). Finally, the target population for the NLSY is an eight year birth cohort (NLSY-79) and a five year birth cohort (NLSY-97), which combined represent roughly 27% of the full HSOII target population of workers aged 18+[11].

Availability of Sample Cases

Here we look at the extent to which the existing surveys offer sample sufficient to meet the needs for the HSOII (expected to be 100,000+ as discussed in Section 4) and at the frequency necessary to allow for annual estimates for the HSOII.

Table 5.1.2 provides the periodicity of the existing surveys under consideration along with information on annual sample sizes – approximate HU sample sizes (number selected for the survey, number screened, and number at which a completed interview was obtained) and an estimate of the number of completed interviews with workers, based upon number of completed interviews and the assumptions detailed in Section 4.

[10] Ryan, C. Language Use in the United States: 2011. U.S. Census Bureau. August 2013. Available from https://www.census.gov/prod/2013pubs/acs-22.pdf.

[11] 2014 ACS data show that adults aged 30-34 and 50-57 years (the 2014 age range for the two NLSY cohorts) make up ~23.5% of adults aged 18+ years; however, their higher employment rates relative to older age groups increases their proportion of the population of workers aged 16-64 years to 27%.

Table 5.1.2: Profile of existing survey sample size and response rates

Survey	Sample Selection Frequency	Sample Size (approximate) Selected HU's	Screened HU's	Completed HU's	Completed Workers	Response Rate	Comments
ACS	Monthly	3.5MM	3.5MM	2.3MM	~3.2MM	96.7%	
CPS	Monthly, 4-8-4 Rotation	225,000	180,000	165,000	~230,000	~90%	Half of annual sample is new in current year
NHIS	Quarterly	68,000	57,000	44,500	~30,000	58.9%	
BRFSS	Annual	6.4MM	850,000	460,000	~300,000	~50%	RoR varies from 34% to 64% across states
MEPS-HC	Annual	17,000	17,000	14,000	~19,000	52.8%	Half of annual sample is new current year
NHANES	Biennial	46,000	Not Found	10,000	~4,000	71.0%	Same sample cases across consecutive odd/even years
NLSY*	Biennial	150,000	135,000	Not Applicable	~11,000	78.7% (79) 83.9% (97)	Same sample cases across time
NIS	Quarterly	9.5MM (phone #'s)	1.1MM	50,000	0	62.6% (Landline) 33.5% (Cell)	State-level design
SIPP	Sample selected every 2.5 to 4 years	65,500	52,000	42,000 (Wave 1) 20,000 (Wave 16)	~58,000 (Wave 1) ~28,000 (Wave 16)	80.6% (Wave 1) 61.8% (Wave 16)	2014 panel to be interviewed once annually 2014-2017
SCF	Triennial	14,000	13,500	6,000	~4,000	51.7%: 68.7% (ABS) 34.7% (list)	Dual-frame design (ABS (4,500) + list sample (1,500) of high income HU's)
NSFG	Varies	12,000	11,500	10,400	~6,700	72.8%	Most recent round collected data from 2011 to 2013
GSS	Biennial	5,200	4,300	2,500	~1,700	71.2%	
AS	As Needed	59,000	15,000	7,500 Panel HUs	~10,000 Panel Members	36.6%	Panel members replaced for attrition; Panel size to triple in 2016
KP	As Needed	Not Found	Not Found	30,000 Panel HUs	~41,000 Panel Members	<10%	Panel members replaced for attrition

*** NLSY79 Response rates represent retention rates of 1991 sample respondents;**
NLSY97 Response rates represent retention rates of original respondents

The surveys which appear to have sufficient numbers of completed interviews with workers (100,000+, as discussed in Section 4) to be considered as a stand-alone sample frame for the HSOII are ACS, CPS, and BRFSS. The large number of HU's selected and screened for NIS make it a candidate for use under Option 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey).

While NHIS and MEPS-HC both have more than 15,000 annual completed interviews with workers and thus could support the sample needed for the HSOII at the national level, they could not support achieving target sample sizes for selected or collapsed industry and occupation. They could, however, be used in a multiple survey design (Option 4), although not together as MEPS-HC is a subset of NHIS.

SIPP has from ~58,000 to ~28,000 completed interviews with workers, depending upon Wave. However, the SIPP sample is selected every 2.5 to 4 years, suggesting SIPP should be considered only as part of a multiple survey (Option 4) and/or panel design (Option 7).

The POPs (AS, KP), while offering 30,000+ workers on their panels (with the AS expansion planned for 2016), are fairly constant panels, and thus offer little additional sample across time. As such, they should be considered for use only in a multiple survey (Option 4) and/or panel design (Option 7).

The remaining existing surveys (NHANES, NLSY, SCF, NSFG) offer very small sample sizes. In addition samples are not selected every year, which further adversely affects their viability for use in the HSOII, although they could be considered as part of a multiple frame and/or panel design.

Further information on survey and frame suitability for Options 1-4 and single frame vs. multi-frame/multi-year designs is provided in Section 7.

Implications

Table 5.1.3 provides the summary "rating" indicators of value relative to Population Representation (limited to frame coverage) for the four sampling frames under consideration.

Table 5.1.3: Summary ratings for sample frame coveage

Sampling Frame	Frame Coverage
Census MAF	+++++
Census Data	+++++
USPS DSF	+++++
RDD Telephone	++++

Three sampling frames offer complete coverage for the HSOII population of interest – Census MAF, Census data plus listing within sampled segments, and USPS DSF with listing in areas where the DSF addresses do not provide sufficient information to locate the address (~2% of all HUs). Each could thus be used as the sampling frame for HSOII. There is often reluctance on the part of federal agencies to utilize the Census MAF, as confidentiality of the addresses is protected and cannot be provided by Census under Title 13. Another issue for consideration is whether Census data review board (DRB) would allow release of a public use microdata sample (PUMS) file that would meet all BLS needs. RDD frames of landline and cellular phones offer an alternative choice with nearly complete coverage (~97%), although persons in non-telephone households do have different demographic and socio-economic characteristics than the full population, and thus appropriate ratio adjustment methods would be required to minimize any potential coverage bias.

Table 5.1.4 provides the summary "rating" indicators of value relative to Population Representation for the existing surveys under consideration.

Table 5.1.4: Summary ratings for existing survey coverage, sample size, and response rates

Survey	Coverage		Sample Selection Frequency	Sample Size			Response Rate
	Frame Coverage	Estimated Coverage of Workers in Eligible Sample		Screened HU's	Completed HU's	Completed Workers	
ACS	+++++	+++++	+++++	+++++	+++++	+++++	+++++
CPS	+++++	+++++	+++++	++++	++++	++++	+++++
NHIS	+++++	+++++	+++++	+++	+++	++	++
BRFSS	++++	++++	++++	+++++	+++++	++++	++
MEPS-HC	+++++	+++++	++++	++	++	++	++
NHANES	+++++	+++++	++	+	++	0	+++
NLSY	+++++	+	++	0	0	+	++++
NIS	++++	0	+++++	+++++	+++	0	++
SIPP	+++++	+++++	++	+++	+++	+++	+++
SCF	+++++	+++++	++	+	0	0	++
NSFG	+++++	++	++	+	++	+	+++
GSS	+++++	+++++	++	0	0	0	+++
AS	+++++	++++	+	0	++	++	+
KP	+++++	++++	+	0	++	++	0

All surveys with the exception of NLSY, NIS, and NSFG were rated 4 or 5 for Coverage. NLSY and NSFG eligible populations are subsets of the full 18+ population, while NIS eligible population is children <18. Six surveys (ACS, CPS, NHIS, BRFSS, MEPS-HC, NIS) were rated 4 or 5 for Sample Selection Frequency, being selected annually or more frequently. Three surveys (ACS, CPS, BRFSS) were rated 4 or 5 for Sample Size, with SIPP being the only other survey rated 3 or higher for Sample Size. Finally, ACS, CPS, and NLSY were rated 4 or 5 for Response Rate, with four others (NHANES, SIPP, NSFG, GSS) being rated 3.

ACS is rated 5 for all criteria, while CPS is rated 4 or 5 for all criteria, suggesting both could be candidates to serve as a single sampling frame for the HSOII. BRFSS is rated 4 or 5 for all criteria with the exception of Response Rate (rating of 2 for response rate ~50%, ranging from 34%-64% across states), suggesting it could likewise serve as a single sampling frame for the HSOII depending upon the potential magnitude of nonresponse bias not controlled through weighting. Finally, although the NIS interviewed sample does not include any workers (NIS target population is children), the selected sample size is of sufficient size to consider for use should additional screening questions be added to the for use in the HSOII design.

Two of the remaining surveys (NHIS, MEPS-HC) could be considered under Option 4 (Multiple Surveys); however, these two surveys could not be used together, given MEPS-HC is selected from NHIS. Although the remaining surveys could also be considered under Option 4 (Multiple Surveys), their

profiles relative to Sample Selection Frequency and Sample Size would likely make their use cost ineffective.

5.2 Survey Mode and Use of Proxy Respondents

Given Mode and Respondent Rules are components of the Data Collection process, both are discussed in this section. In addition, given the impact of within-HU person sampling, that factor is also discussed in this section. Table 5.2.1 provides information on data collection mode, within-HU person sampling, and respondent rules for the existing surveys.

Table 5.2.1: Profile of existing survey data collection mode and rules

Survey	Data Collection Mode	Sampled Persons	Respondent Rules
ACS	Three stage: 1) Mail/Internet (50%) 2) CATI (7%) 3) CAPI (44%)	All persons	Adult HH respondent
CPS	PV (MIS 1, 5) Phone (MIS 2-4, 6-8 if HH agrees; ~85% telphone)	All persons 16+	Self-report, if possible; Knowledgeable Adult as proxy (~50%)
NHIS	CAPI	One adult 18+ One child <18	Self-report (adult) Knowledgeable adult (child)
BRFSS	CATI	One adult 18+	Self-report
MEPS-HC	CAPI	All persons	Most knowledgeable adult
NHANES	CAPI	One, some, all, or none of persons in the HU	Self-report
NLSY	CATI (90%)/CAPI (10%)	All persons in HH within age cohort	Self-report
NIS	CATI	All children 19-35 mo (NIS) One child 13-17 years (NIS-Teen) One child 6-18 mo/3-12 years (NIS-Flu)	Knoweldgeable Adult
SIPP	CAPI (Waves 1,2,6) CATI (Waves 3,4,5,7-16)	All persons 15+	Self-report
SCF	CAPI	Economically dominant single individual or financially most knowledgeable member	Self-report
NSFG	CAPI	One eligible person aged 15-44 years	Self-report
GSS	CAPI	One adult 18+	Self-report
AS	Internet/CATI	All persons 18+	Self-report
KP	Internet	All persons 18+	Self-report

Survey Mode

Survey mode is of interest under Option 1 (Supplemental Module in Existing Survey). For options utilizing a follow-up survey, the mode utilized for HSOII would not have to be the same as that used in the existing survey.

Most surveys utilize an interviewer/respondent interaction in conducting the survey (CAPI and/or CATI). The exceptions are the POPs (AS, KP, which use Web self-reporting) and ACS, in which 50% of completes are self-reported via mail/internet.

Thus, the distinguishing factor among the surveys is whether interviews are conducted in person or over the phone, in either case using computer-assisted interviewing (CAI) technology. The relative advantages/disadvantages of mode relative to minimizing potential recall bias will be addressed in a separate report addressing questionnaire and data collection process.

Use of Proxy Respondents

Respondent rules and the use of proxy respondents are of interest under Option 1 (Supplemental Module in Existing Survey). For options utilizing a follow-up survey, respondent rules utilized for HSOII would not have to be the same as those used in the existing survey. All surveys with the exception of ACS and MEPS-HC instruct interviewers to obtain self-reports for sampled persons aged 18+ years. For CPS, although the interviewer seeks to obtain self-reports, proxy respondents are utilized for roughly half the completed interviews.

Sampled Persons within HU

A consideration for all options, but especially for Option 1 (Supplemental Module in Existing Survey), is whether all persons are sampled within a HU. As little to no correlation between persons within a HU would be expected for the HSOII, it would be most cost-efficient to sample all eligible persons within a HU. ACS, CPS, MEPS-HC, NLSY, and the POPs sample all persons within the HU. NHANES may select all persons, but they may also select one, some, or no persons within the HU.

Implications

Table 5.2.2 provides the mode(s) used and summary “rating” indicators of value relative to Data Collection Rules.

Table 5.2.2: Summary ratings for existing survey data collection rules

Survey	Mode	Data Collection	
		Sampled Persons	Respondent Rules
ACS	Mail/Web/CATI/CAPI	+++++	++
CPS	CAPI/CATI	+++++	+++
NHIS	CAPI	+++	+++++
BRFSS	CATI	+++	+++++
MEPS-HC	CAPI	+++++	++
NHANES	CAPI	+++	+++++
NLSY	CAPI/CATI	+++++	+++++
NIS	CATI	N/A	N/A
SIPP	CAPI/CATI	+++++	+++++
SCF	CAPI	+++	+++++
NSFG	CAPI	+++	+++++
GSS	CAPI	+++	+++++
AS	Web/CATI	+++++	+++++
KP	Web	+++++	+++++

Collection mode and respondent rules are of interest under Option 1 (Supplemental Module in Existing Survey), adding a supplemental module to an existing survey. With the exception of ACS (which utilized Mail, Web, CATI, and CAPI collection), all surveys use CAI, either CATI (BRFSS, NIS), CAPI (NHIS, MEPS-HC, NHANES, SCF, NSFG, GSS), Web (KP), a combination of CATI and CAPI (CPS, NLSY), or a combination of Web and CATI (AS). Advantages/ disadvantages of CATI vs. CAPI relative to recall bias will be discussed in a separate report addressing questionnaire and data collection process.

NLSY, SIPP, AS, and KP were each rated 5 for both Sampled Persons and Respondent Rules. ACS, CPS, and MEPS-HC sample all adults within a HU; however, ACS and MEPS-HC obtain data from one HH respondent, while roughly 50% of CPS person level completes are obtained from a proxy respondent. While NHIS, BRFSS, NHANES, SCF, NSFG, and GSS seek self-reports, they do not sample all adults within a HU.

5.3 Availability of Key Information

The two primary types of information of interest for the HSOII data collection are employment data and injury/illness data. The greater the concordance between data collected in the existing survey and data needed for the HSOII, the greater the value of the existing survey.

Employment Data

Table 5.3.1 describes availability of industry, occupation, employment relationship, and sector information from the current data collection for the surveys.

Table 5.3.1: Profile of existing survey employment information

Survey	Available Information					
	Employment Status	Multiple Jobholder	Occupation	Industry	Employment Relationship	Sector
ACS	Prior 12 months	Yes	Chief job activity/ business	Chief job activity/ business	Chief job activity/ business	Chief job activity/ business
CPS	Week containing 12th of month	Yes	Up to two jobs	Up to two jobs	Yes	Yes
NHIS	Prior 12 months	Yes	Main current job activity/ business, else for job worked longest	Main current job activity/ business, else for job worked longest	Main current job activity/ business, else for job worked longest	Main current job activity/ business, else for job worked longest
BRFSS	Current/Prior 12 months	No	Obtained in optional module (Industry/ Occupation)	Obtained in optional module (Industry/ Occupation)	Obtained in optional module (Industry/ Occupation)	No
MEPS-HC	Prior 12 months	Yes	Main current job activity/ business, else for most recent job	Main current job activity/ business, else for most recent job	Yes	Yes
NHANES	Current	No	Main job or business	Main job or business	Yes	Yes
NLSY	Prior 12 months	Yes	All jobs worked	Obtained for all jobs worked	Yes	Yes
NIS	No	No	No	No	No	No
SIPP	Prior 3 months	Yes	Up to two jobs	Type (Mfg, Whole Trade, Retail Trade, Service, Oth) for up to two jobs	Yes	Yes
SCF	Current (R); Current (R spouse/partner)	Yes	Up to two jobs	Up to two jobs	Yes	Yes
NSFG	Prior 12 months (R); Prior week (R spouse/partner)	Yes	No	No	No	No
GSS	Current	No	One current job	No	Yes	Yes
AS	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey	Yes	If asked as part of a given survey
KP	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey

Industry and occupation (I/O) information and employment relationship, is obtained in all surveys with the exception of NIS (as the survey target population is children) and BRFSS (for which employment data are collected in only if a state-optional module is implemented). In some surveys, I/O information is obtained only if currently employed (CPS, NHANES), while in other surveys I/O information is obtained on most recent job if not currently employed (ACS) or the job the individual worked at for the longest time period (NHIS).

Two other relevant pieces of information for the HSOII are whether the sample person has been employed at some time in the prior 12 months (obtained in ACS, NHIS, MEPS-HC, and NLSY), and whether the sample person is a multiple jobholder (obtained in CPS and NLSY). For multiple jobholders, CPS obtains I/O information for a second job, whereas other surveys collect I/O information for only one job. In collecting I/O information when persons have multiple jobs, some ask the person to report for the

"chief" job activity/business (ACS), and others collect I/O information for the "main" job activity/business (NHIS, MEPS-HC).

Injury and Illness Data

Table 5.3.2 describes availability of industry, occupation, employment relationship, and sector information from the current data collection for the existing surveys.

Table 5.3.2: Profile of existing survey injury/illness information

Survey	Injury/Illness Data		
	Injuries	Illnesses	Comments
ACS	No	No	
CPS	No	No	
NHIS	Number of work days missed in prior 12 months due to injury/illness; non-specific as to where injury/illness occurred	Number of work days missed in prior 12 months due to injury/illness; non-specific as to where injury/illness occurred	Obtains other information in injuries/illnesses which could be informative
BRFSS	Falls within prior 12 months; did fall result in injury	Chronic	
MEPS-HC	Identifies accidents/injuries that occurred while at work	Obtains information on ER/medical provider visits, hospital stays	
NHANES	Obtains information on ER/medical provider visits, hosptial stays	Obtains information on ER/medical provider visits, hospital stays	
NLSY	Extent to which pain interfered with work activities in prior 4 weeks	Selected chronic	
NIS	No	No	
SIPP	No	No	Questions on disability payments
SCF	No	No	Questions on disability payments
NSFG	No	No	
GSS	No	No	
AS	If asked as part of a given survey	Selected chronic	
KP	If asked as part of a given survey	If asked as part of a given survey	

As might be expected, NHIS, BRFSS, MEPS-HC, and NHANES are the surveys in which injury and illness information is collected. MEPS-HC is the one survey which collects identifies work-related accidents/injuries, and thus would be of value for screening under Option 2 (Follow-on to Existing Survey, no Change to Existing Survey), as well as providing useful background information for a supplemental module to MEPS-HC under Option 1 (Supplemental Module in Existing Survey). For the other surveys, use of injury and illness data could be of benefit when developing a supplemental module under Option 1 (Supplemental Module in Existing Survey).

Implications

Table 5.3.3 provides the summary "rating" indicators of value relative to Availability of Employment and Injury/Illness Information.

Table 5.3.3: Summary ratings for existing survey employment and industry/occupation information

Survey	Employment Information						Injury/Ilness Information	
	Employment Status	Multiple Jobholders	Occupation	Industry	Employment Relationship	Sector	Injury	Illness
ACS	+++++	+++++	+++	+++	+++++	+++++	0	0
CPS	+++	+++++	++++	++++	+++++	+++++	0	0
NHIS	+++++	+++++	+++	+++	+++++	+++++	++++	++++
BRFSS	+++++	0	++	++	+++++	0	++	+
MEPS-HC	+++++	+++++	+++	+++	+++++	+++++	+++++	++
NHANES	+++	0	+++	+++	+++++	+++++	++	++
NLSY	+++++	+++++	+++++	+++++	+++++	+++++	+++	+
NIS	0	0	0	0	0	0	0	0
SIPP	++++	+++++	++++	+++	+++++	+++++	0	0
SCF	+++	+++++	++++	++++	+++++	+++++	0	0
NSFG	+++++	+++++	0	0	0	0	0	0
GSS	+++	0	+++	0	+++++	+++++	0	0
AS	+	+	+	+	+++++	+	+	+
KP	+	+	+	+	+	+	+	+

Availability of employment and I/O information from an existing survey can allow for targeted interviewing under either a supplemental module (Option 1: Supplemental Module in Existing Survey) or use of collected data to screen for follow-up (Option 2: Follow-on to Existing Survey, no Change to Existing Survey) design. NLSY provides the most comprehensive information, being rated 5 for all employment information categories, identifying persons employed at any point within the prior 12 months (the target population for HSOII), and I/O information for all jobs. Other surveys with a rating of 3+ for all employment information categories are ACS, CPS, NHIS, MEPS-HC, SIPP, and SCF. For these surveys, employment status may not cover the prior 12 months and I/O information is often not collected for all jobs. The information from these surveys may be of value under Options 1 (Supplemental Module in Existing Survey), 2 (Follow-on to Existing Survey, no Change to Existing Survey), or 3 Follow-on to Existing Survey, Additional Screening Questions in Existing Survey ().

Availability of relevant injury and illness data is very limited. Only MEPS-HC (identifies accidents/injuries occurring at work) and NHIS (obtains number of work days missed due to injury/illness) obtain data that could potentially be useful for Options 1 (Supplemental Module in Existing Survey), 2 (Follow-on to Existing Survey, no Change to Existing Survey), or 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey).

5.4 Timeliness of Calendar Year Estimates

Data from the ESOII is released roughly eleven months following the end of the ESOII estimation period, which is the calendar year. Survey timeliness is primarily of consideration for Option 1 (Supplemental Module in Existing Survey), adding a supplemental module to an existing survey. Survey timeliness must also be considered for Options 2 (Follow-on to Existing Survey, no Change to Existing Survey) and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), wherein data from the survey is utilized in the sample design of the HSOII, to avoid deterioration over time of the accuracy of the information for use in the HSOII sample design and data collection process.

In addition to data timeliness, additional related considerations include the frequency with which sample cases are interviewed, the respondent burden associated with the existing survey (based on OMB clearance packages), and the frequency with which estimates are generated from the existing survey.

Table 5.4.1 provides information on these characteristics of the existing survey data collection process and estimate generation. Availability of the survey data is based on publication of the survey data. It may be possible that the information needed for use in a HSOII design may be available for use by BLS earlier.

Table 5.4.1: Profile of existing survey frequency, respondent burden, and timing

Survey	Collection Period	Collection Frequency for Sample Unit	Respondent Burden	Estimate Periodicity	Data Release Lag
ACS	Calendar Year	Once	30-40 min	Annual	~9 months
CPS	Two weeks	4 consecutive months, same 4 months 12 months later	~8 min	Monthly	~3 weeks
NHIS	Calendar Year	Once	~53 min	Annual	~6 months
BRFSS	Calendar Year	Once	~16 min	Annual	~9 months
MEPS-HC	5 Months	5 times over 2 years	~107 min	Annual	~7 months
NHANES	24 Month Period	Once	~150 min	Biennial	~10 months
NLSY	12 Months	Every 2 years	~61 min	Biennial	~15 months
NIS	Calendar Year	Once	~32 min	Annual	~9 months
SIPP	One Calendar Year	Once a year for 4 years (current panel)	~60 min	Annual	~9 months
SCF	One Calendar Year	Once	~75 min	Triennial	~9 months
NSFG	12 Month Period (for each of 4 samples)	Once	~70 min (M) ~100 min (F)	Varies (4 years, current)	~12 months
GSS	7 Months	Once	~90 min	Biennial	~9 months
AS	Varies	Multiple	Varies	Varies	~1 month
KP	Varies	Multiple	Varies	Varies	~1 month

Two surveys (NLSY, NSFG) do not publish results prior to current publication of ESOII data. Seven other surveys (ACS, BRFSS, NHANES, NIS, SIPP, SCF, GSS) publish data less than two months prior to

current publication of ESOII data. Only five surveys (CPS, NHIS, MEPS-HC, AS, KP) publish estimates within a timeframe that might be considered "safe" for HSOII.

Several surveys (CPS, MEPS-HC) collect data from a given sample unit multiple times within a year, Such a collection frequency could be advantageous under Option 1 (Supplemental Module in Existing Survey) implemented as a panel survey (Option 7), as it would allow for bounding interviews and recall periods of less than 12 months. These issues will be discussed further in a separate report addressing questionnaire and data collection process.

Seven surveys (MEPS-HC, NHANES, NLSY, SIPP, SCF, NSFG, GSS) have estimated respondent burden of one hour or more, while two (CPS, BRFSS) have estimated respondent burdens less than 20 minutes.

Implications

Table 5.4.2 provides the summary "rating" indicators of value relative to Timeliness of Calendar Year Estimates.

Table 5.4.2: Summary ratings for existing survey frequency, respondent burden, and timing

Survey	Collection Frequency for Sample Unit	Interview Length	Estimate Periodicity	Data Release Lag
ACS	++++	+++	+++++	++++
CPS	+++++	+++++	+++++	+++++
NHIS	++++	++	+++++	+++++
BRFSS	++++	++++	+++++	++++
MEPS-HC	+++++	+	+++++	+++++
NHANES	++++	+	++	++++
NLSY	++	+	++	+
NIS	++++	+++	+++++	++++
SIPP	++++	+	+++++	++++
SCF	++++	+	++	++++
NSFG	++++	+	++	+
GSS	++++	+	++	++++
AS	+++++	++++	+++++	+++++
KP	+++++	++++	+++++	+++++

For Option 1 (Supplemental Module in Existing Survey), only CPS has a short interview length in conjunction with a data release timeframe meeting likely timing necessary for BLS to obtain information from the existing survey and supplemental module to support the existing SOII timeframe for publication of results.

Utilizing screening information from an existing survey to select the HSOII sample (Options 2: Follow-on to Existing Survey, no Change to Existing Survey, 3: Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) should be possible from a timing perspective under all surveys with the exception of NLSY and NSFG.

5.5 Availability of Auxiliary Data for Sample Design Efficiency

Data collected in the existing surveys can be utilized to improve the efficiency of the HSOII design under Options 2 (Follow-on to Existing Survey, no Change to Existing Survey) and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey). Under Option 2 (Follow-on to Existing Survey, no Change to Existing Survey), the assumption is that HSOII design will be developed utilizing the information collected as part of the existing survey data collection instrument, whereas Option 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) assumes additional screening information may be collected in the existing survey to enhance the efficiency of the HSOII design.

The primary gains to be made will be derived from information discussed in Section 5.3: Availability of Key Information. Availability of employment data for persons within sample HUs from existing surveys will be valuable in stratification for the HSOII design under Options 2 (Follow-on to Existing Survey, no Change to Existing Survey) and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), especially in terms of obtaining sufficient sample sizes within industry and occupation groups.

While the design could consider screening out sample HUs with only elderly, retired individuals, it is possible that the residents at the HU have changed and thus eligible persons could be missed. A sample design could sample such HUs at a lower rate than other HUs so as to ensure representation but with lower sample sizes and costs.

Other information about the HU and persons within the HU, such as age, race, gender, and number of persons, may be of some value for stratification. Number of person in the HU is available from all existing surveys. Age, race, and gender is available for the sampled persons.

6. Discussion

Within this section, we discuss some two topics that are more global in nature:

1) From Section 2, assessment criterion 7 – Would the implications discussed for assessment criteria 1-6 change if the survey is merely used to screen for potential respondents (i.e., Options 2: Follow-on to Existing Survey, no Change to Existing Survey, or 3: Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) rather than including a full module of 15-20 questions on occupational injury and illness (i.e., Option 1: Supplemental Module in Existing Survey)?
2) What can be posited at this stage of the project as to potential costs associated with use of various existing surveys and design options?

6.1 Screening vs. Add-on Module

Add-on Module

Option 1 (Supplemental Module in Existing Survey) would result in the most cost-effective survey design, as contact/response conversion efforts and associated costs are carried out through the existing survey. The primary challenges in implementing Option 1 (Supplemental Module in Existing Survey) are existing survey: 1) sample size; 2) timing; 3) and respondent burden.

Table 6.1.1 presents counts of the summary ratings of 5, 4, or 3 assigned to existing surveys, across the assessment criteria.

Existing surveys with the largest number of assigned high summary ratings are CPS (12-5's, 17-5/4's, 19-5/4/3's) and ACS (13-5's, 15-5/4's, 18-5/4/3's). These two surveys each have sufficient sample size to be considered as single frames for the HSOII and have high assigned ratings for Employment and I/O Information. These two aspects make CPS and ACS potential candidates for adding a HSOII module to the existing survey. Timing of ACS data release (~9 months) and respondent burden (~30-40 min), result in ACS being less ideal for Option 1 (Supplemental Module in Existing Survey) than is CPS. It may also be argued a priori that ACS could not be utilized in an HSOII design as the likelihood of receiving approval to add a supplemental module to ACS would be near, if not at, zero.

Table 6.1.1: Counts of summary ratings of 5/4/3 for existing surveys, by assessment criteria group

Survey	Assessment Criteria Group				Total
	Coverage/ Sample Size/ Response Rate (7)	Data Collection Mode & Rules (2)	Employment & I/O Information (8)	Survey Frequency, Respondent Burden & Timing (4)	
ACS	7/0/0	1/0/0	4/0/2	1/2/1	13/2/3
CPS	4/3/0	1/0/1	3/2/1	4/0/0	12/5/2
NHIS	3/0/2	1/0/1	4/2/2	2/1/0	10/3/5
BRFSS	2/4/0	1/0/1	2/0/0	1/3/0	6/7/1
MEPS-HC	2/1/0	1/0/0	5/0/2	3/0/0	11/1/2
NHANES	2/0/1	1/0/1	2/0/3	0/2/0	5/2/5
NLSY	1/1/0	2/0/0	6/0/1	0/0/0	9/1/1
NIS	2/1/1	0/0/0	0/0/0	1/2/1	3/3/2
SIPP	2/0/4	2/0/0	3/2/1	1/2/0	8/4/5
SCF	2/0/0	1/0/1	3/2/1	0/2/0	6/4/2
NSFG	1/0/1	1/0/1	2/0/0	0/1/0	4/1/2
GSS	2/0/1	1/0/1	2/0/2	0/2/0	5/2/4
AS	1/1/0	2/0/0	1/0/0	3/1/0	7/2/0
KP	1/1/0	2/0/0	0/0/0	3/1/0	6/2/0

NHIS (18-5/4/3's) and SIPP (17-5/4/3's) are the two other existing surveys with similar counts of high ratings. While both have high assigned ratings for Employment and I/O Information, their sample sizes are not sufficient to be considered as single frames for adding a HSOII module to the existing survey, and their respondent burdens are each high. In addition, SIPP sample are only selected every 2.5 to 4 years.

A combination of Option 1 (Supplemental Module in Existing Survey) and Option 4 (Multiple Surveys) could be considered to open up the existing surveys which could be considered. However, coordinating approvals to add modules to multiple surveys and the field effort across multiple agencies make this appear undesirable.

A combination of Option 1 (Supplemental Module in Existing Survey) and Option 6 (Multi-year Survey) could also be considered to increase sample size. However, sample sizes for the other existing surveys would require use of a three-year estimation period, not a two-year estimation period.

CPS is currently used as a frame for a number of supplemental modules, as indicated in Table 6.1.2. Each supplement is conducted in one given month of the year on an annual, biennial, or triennial basis.

Table 6.1.2: Timing, frequency of repeated CPS supplements

Month	Supplement	Frequency
Jan	Displace Workers	Biennial
	Job Tenure/Occupational Mobility	Biennial
	Unbanked/Underbanked	Biennial
Feb	Public Participation in the Arts	Annual
Mar	Annual Social and Economic Supplement	Annual
Apr	Child Support	Biennial
May		
Jun	Fetility	Biennial
Jul	Tobacco Use	Triennial
Aug	Veterans	Annual
Sep	Volunteers	Annual
Oct	School Enrollment	Annual
Nov	Voting and Registration	Biennial
Dec	Food Security	Annual

There are a set of nine criteria considered before implementing a CPS supplement, which HSOII would appear to meet. Key among the criteria as they would relate to HSOI are respondent burden ("supplemental inquiry must not add more than 10 minutes of interview time per respondent or 25 minutes per household"), compatibility of the supplement with CPS ("must not introduce a concept that could affect the accuracy of responses to the basic CPS information"), and confidentiality ("information that could be used to indirectly identify an individual with a high probability of success will be suppressed"). Note that, to provide sufficient sample size in implementing HSOII as a CPS supplement, and to reduce the potential for seasonality affecting survey responses, it would be necessary to implement HSOII as a CPS supplement in multiple months

Screening

When considering implementing Options 2 (Follow-on to Existing Survey, no Change to Existing Survey) or 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), the primary challenges are existing survey: 1) sample size; 2) timing; 3) and availability of key information.

As discussed in Section 5.1, ACS, CPS, and BRFSS are the only surveys with sample sizes sufficient to serve as single frames for an HSOII survey design. As indicated in Table 6.1.1, the Employment and I/O Information available from BRFSS would likely be insufficient to develop a sufficiently efficient sample design for use with Option 2 (Follow-on to Existing Survey, no Change to Existing Survey). In addition,

as BRFSS is implemented by individual states, not all of which use currently existing available optional modules, it seems best not to consider as a single frame source.

CPS could be considered for both Options 2 (Follow-on to Existing Survey, no Change to Existing Survey) and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey). As with Option 1 (Supplemental Module in Existing Survey), it is unlikely ACS could be used for Option 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), although it could be considered for Option 2 (Follow-on to Existing Survey, no Change to Existing Survey). The discussion relative to other existing surveys and Option 1 (Supplemental Module in Existing Survey) would also argue against their use for Option 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), with one exception. NIS could be considered for use with Option 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), given its large sample size of sampled telephone numbers. This would require adding a few additional questions to the NIS screener, seeking to identify eligible workers with the sample HU, as well as eligible children.

Use of other existing surveys could be considered for Option 2 (Follow-on to Existing Survey, no Change to Existing Survey), although the smaller sample sizes would likely require combining Option 2 (Follow-on to Existing Survey, no Change to Existing Survey) with Options 4 (Multiple Surveys) and 6 (Multi-year Survey) to achieve the overall sample sizes needed for HSOII.

6.2 Survey Costs

Survey costs are also a critical factor in establishing an approach to an HSOII. The following factors will affect survey costs:

1) Data collection costs associated with Options 1 (Supplemental Module in Existing Survey), 2 (Follow-on to Existing Survey, no Change to Existing Survey), 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), and 5 (Utilize Existing Sampling Frame) is utilized.
2) The need for and complexity of the sample design and sample selection;
3) The extent of files required and complexity associated with pre-sample selection processing of data from existing surveys;
4) Use of CAPI vs CATI data collection;
5) Required sample size;
6) The extent of files required and complexity associated with processing the HSOII data files;
7) The complexity associated with survey weighting.
8) Use of Option 4 (Multiple Surveys);

9) Use of Option 6 (Multi-year Survey);
10) Use of Option 7 (Panel Survey);
11) Which existing survey(s) is utilized.

Each factor is discussed briefly below. More extensive and detailed discussion will be provided in the Survey Design report.

Data Collection

Option 1 (Supplemental Module in Existing Survey) will incur the lowest data collection survey costs, as it utilizes an existing survey sample and data collection process. Data collection costs would be incurred related to interviewer time to ask the additional questions in the HSOII and potentially additional required follow-ups to complete the HSOII module beyond those needed for the existing survey.

Options 2 (Follow-on to Existing Survey, no Change to Existing Survey) and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) would incur substantially higher data collection costs as all interviewer costs associated with contacting, obtaining participation, and screening would be borne by HSOII. Option 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) would also incur additional costs associated with asking the additional screening questions as part of the existing survey interview. This may be partially, but likely not fully, offset by a reduction in interview time for the HSOII under Option 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) given improved screening information.

Option 5 (Utilize Existing Sampling Frame) would incur the highest HSOII data collection costs, given additional efforts needed for locating, contacting, obtaining participation, and screening for sample that has not been contacted previously as part of an existing survey.

Sample Design

Option 1 (Supplemental Module in Existing Survey) will incur no costs for developing a sample design and selecting a sample as it utilizes that for the existing survey.

Options 2 (Follow-on to Existing Survey, no Change to Existing Survey) and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) will incur sample design and selection costs, for which the design component will likely be more involved than that associated with Option 5 (Utilize Existing Sampling Frame). This will be due to the additional design information available from the existing survey and the need to oversample selected industry and occupation groups to control total sample size.

Sample design and selection costs for Option 4 (Multiple Surveys) will be even higher than those for Options 2 (Follow-on to Existing Survey, no Change to Existing Survey), 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), and 5 (Utilize Existing Sampling Frame), due to the additional complexity of optimally designing a dual frame survey.

Pre-sample Selection Data Processing

Costs for this component would be expected to be lowest for Option 5 (Utilize Existing Sampling Frame). This is due to the need under Options 1 (Supplemental Module in Existing Survey), 2 (Follow-on to Existing Survey, no Change to Existing Survey), and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) to become familiar with and process data from an existing survey on an annual or perhaps more frequent basis. Depending upon how early in the existing survey data processing the HSOII relevant data files are created, the extent of data checking and cleaning needed before the files are ready for use in the HSOII sample design, selection, questionnaire data feeds may be significant. This issue would be compounded under Option 4 (Multiple Surveys).

CAPI vs. CATI

Data collection costs would be expected to be greater for CAPI than for CATI under Options 2 (Follow-on to Existing Survey, no Change to Existing Survey) and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), given higher costs associated with field travel, and assumed useful information from the existing survey for CATI to utilize in contacting the sample cases. Given HSOII would not incur most of the overhead interviewing costs (locating, contacting, screening, obtaining participation, etc.), the cost differential between CAPI and CATI should be much less for Option 1 (Supplemental Module in Existing Survey).

Under Option 5 (Utilize Existing Sampling Frame), there would likely be a need for a larger selected sample size for CATI, to account for ability to contact HUs though a telephone call, along with associated higher locating and contacting costs for CATI than for CAPI, relative to Options 1 (Supplemental Module in Existing Survey), 2 (Follow-on to Existing Survey, no Change to Existing Survey), and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey). However, CAPI costs would also be much higher given the need for locating and contacting sample HUS.

Sample Size

Costs will be greater with larger required sample sizes, due almost exclusively to increased data collection costs. The impact on costs for sample design and selection, data processing, and survey weighing associated with larger sample sizes can be assumed to be negligible.

Pre-weighting Data Processing

As with Pre-sample Selection Data Processing, costs for this component would be expected to be lowest for Option 5 (Utilize Existing Sampling Frame), again due to the need under Options 1 (Supplemental Module in Existing Survey), 2 (Follow-on to Existing Survey, no Change to Existing Survey), and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) to become familiar with and process data from an existing survey on an annual or perhaps more frequent basis. Depending upon how early in the existing survey data processing the HSOII relevant data files are created, the extent of data checking and cleaning needed before the files are ready for use in the HSOII sample design, selection, questionnaire data feeds may be significant. This issue would again be compounded under Option 4 (Multiple Surveys).

Survey Weighting

Cost differences for survey weighting would primarily be associated with the initial development of the weighting methodology and programming. Option 1 (Supplemental Module in Existing Survey) should again be expected to have the lowest costs, as it leverages the survey weights developed for the existing survey. While development of the weighting methodology and programming for Options 2 (Follow-on to Existing Survey, no Change to Existing Survey), 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), and 5 (Utilize Existing Sampling Frame) would be somewhat greater, they would be expected to have a minimal impact of the overall survey costs. Development of the weighting methodology and programming for Options 6 (Multi-year Survey) and 7 (Panel Survey) should likewise have minimal impact on the overall survey costs. Development of the weighting methodology and programming for Option 4 (Multiple Surveys) will require the greatest effort and costs; however, again should not noticeably affect the overall survey cost.

Option 4 (Multiple Surveys)

Costs for multiple survey approach (Option 4) would be greater than the costs for a corresponding single frame approach, as well as adding more complexity to the sample design creation and post-survey processing. These costs may, however, be partially offset through improved sample design efficiencies resulting in lower required sample sizes than a single survey approach.

Option 6 (Multi-Year Survey)

Costs for a multi-year survey (Option 6) would be lower than those for a corresponding single year survey, due to spreading a given sample size requirement across multiple years. Data collection costs could be as much as 50% lower (as number of interviews per year in a two year design would be roughly half that for the single year survey), although the actual reduction may be less.

Option 7 (Panel Survey)

Costs for a panel survey (Option 7) with one interview per year would be lower than those for a corresponding survey selecting independent sample each survey period, as subsequent interviews of sample persons previously interviewed generally will cost less than new contacts. Note that the HSOII in general should benefit from this situation as it will be recontacting sample persons from the existing surveys (with the exception of Option 5: Utilize Existing Sampling Frame). Cost reductions would be dependent upon the rotation scheme and number of interviews defined for the panel survey. Should a panel design be developed that requires more frequent (e.g., bi-annual or quarterly) interviews of sample cases, then the cost reduction mentioned above would not be realized.

Existing Survey(s)

While pending further investigation and consideration, the survey(s) utilized is believed to have little impact on costs, after accounting for the other factors previously listed.

Rough Cost Assessment

In an attempt to summarize the information presented above and to provide an initial point of discussion, Table 6.2.1 presents a preliminary and very rough assessment of the likely relative costs of the existing survey/sampling frame and design option combinations. Within Table 6.2.1, dollar signs ($) are used to represent ordering of costs associated with various categories/levels within a factor. The number of dollar signs, however, should not be interpreted as signifying the relative magnitude of costs among categories/levels, nor relative magnitude of costs across factors. Further work along this line will be carried out as part of development of the Survey Design report.

Table 6.2.1: Preliminary assessment of relative costs for existing surveys and design options

Factor	HSOII Survey Design Option						
	Option 1: Supplemental Module in Existing Survey	Option 2: Follow-on to Existing Survey, no Change to Existing Survey	Option 3: Follow-on to Existing Survey, Additional Screening Questions in Existing Survey	Option 5: Utilize Existing Sampling Frame	Option 4: Multiple Surveys	Option 6: Multi-year Survey	Option 7: Panel Survey
Data Collection	$	$$	$$-$$$	$$$	N/A		
Sample Design	$	$$-$$$	$$-$$$	$$	$$$$		
Pre-sample Selection Data Processing	$$-SSS	$$-$$$	$$-$$$	$	$$$$		
Mode							
CAPI	$-$$	$$$$	$$$$	$$$$$	N/A		
CATI	$	$$$	$$$	$$$$$	N/A		
Sample Size							
50,000	$	$	$	$	N/A		
100,000	$-$$	$$	$$	$$	N/A		
150,000	$$	$$$	$$$	$$$	N/A		
Pre-weighting Data Processing	$$-$$$	$$-$$$	$$-$$$	$	$$$$		
Survey Weighting	$	$-$$	$-$$	$-$$	$$$		
Option 4: Multiple Surveys							
Single survey					$$		
Two surveys					$$$-$$$$		
Option 6: Multi-year Surveys							
One year						$$$-$$$$	
Two years						$-$$	
Option 7: Panel Surveys+A10							
One interview per year							$
Two interview per year							$$
Four interviews per year							$$$-$$$$

7. Summary

Utilization of the existing surveys and sampling frames in the HSOII survey design is related to the design option of interest. Table 7.1 presents the existing surveys/sampling frames relative to their potential use in design Options 1-4 and single frame vs. multi-frame/multi-year designs. The assessment as to potential use is based upon sample size, which were assumed to be required to be 100,000+ for use as a single frame, and at least 15,000 to be considered for use in a multi-survey and/or multi-year design.

Table 7.1: Summary assessment of use of existing surveys and sampling frames relative to design options 1-4 and single frame vs. multi-frame and/or multi-year design

Survey/Sampling Frame	Design Option					
	Option 1: Supplemental Module in Existing Survey	Option 2: Follow-on to Existing Survey, no Change to Existing Survey		Option 3: Follow-on to Existing Survey, Additional Screening Questions in Existing Survey		Option 4: Utilize Existing Sampling Frame
	Single Frame	Single Frame	Multi-Frame/ Multi-year	Single Frame	Multi-Frame/ Multi-year	Single Frame
ACS	No	**Yes**	No	No	No	N/A
CPS	**Yes**	**Yes**	No	**Yes**	No	N/A
NHIS	No	No	**Yes**	No	**Yes**	N/A
BRFSS	No	**Yes**	No	No	No	N/A
MEPS-HC	No	No	**Yes**	No	**Yes**	N/A
NHANES	No	No	No	No	No	N/A
NLSY	No	No	No	No	No	N/A
NIS	No	No	No	**Yes**	No	N/A
SIPP	No	No	**Yes**	No	**Yes**	N/A
SCF	No	No	No	No	No	N/A
NSFG	No	No	No	No	No	N/A
GSS	No	No	No	No	No	N/A
AS	No	No	No	No	**Yes**	N/A
KP	No	No	No	No	**Yes**	N/A
Census MAF	N/A	N/A	N/A	N/A	N/A	Yes
Census Data	N/A	N/A	N/A	N/A	N/A	Yes
USPS DSF	N/A	N/A	N/A	N/A	N/A	Yes
RDD Telephone	N/A	N/A	N/A	N/A	N/A	Yes

When considering most likely approval constraints, CPS is the only existing survey that can be considered for use as a single frame approach under Option 1 (Supplemental Module in Existing Survey), as ACS raises substantial feasibility issues in obtaining approvals, while adding to the BRFSS raises substantial operational issues given the federal-state cooperative nature of the survey and states' determining which modules to include. CPS also appears to be the only existing survey that could be used as a single frame approach under both Options 2 (Follow-on to Existing Survey, no Change to Existing Survey) and 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey). ACS could also be considered for use as a single frame approach under Option 2 (Follow-on to

Existing Survey, no Change to Existing Survey) but not Option 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey), while NIS could also be considered for use as a single frame approach under Option 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey). Given states have the option of determining which modules to implement, BRFSS does not appear to be a desirable source for a sampling frame for HSOII.

NHIS and SIPP have sufficient sample to be considered for use under Option 1 (Supplemental Module in Existing Survey), 2 (Follow-on to Existing Survey, no Change to Existing Survey), or 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) in a multi-survey (Option 4) and/or multi-year (Option 6) approach. Although they do not routinely collect employment information, AS and KP have sufficient sample to be considered for use under Option 3 (Follow-on to Existing Survey, Additional Screening Questions in Existing Survey) in a multi-survey (Option 4) and/or multi-year (Option 6) approach. Implementation of a multi-survey (Option 4) approach does require consideration of operational aspects of implementation for two or more surveys, such as coordinating approval to add a supplemental module to more than one existing survey, as well as the need for different structures in the modules to account for differences among the multiple surveys.

Appendix A. Detail About Design Options

As listed in Section 2, seven design options have been identified. These options are not, however, mutually exclusive and some could be used in combination. Options 1, 2, and 3 refer to the survey instrument and data collection process that would be used for the HSOII. Option 4 (Multiple Surveys) was proposed to address potential sample size issues associated with some of the existing surveys, and could be used in conjunction with Options 1, 2, and/or 3. Option 5 (Utilize Existing Sampling Frame) (utilize an existing sample frame) is separate from considering use of an existing survey as the sample frame for the HSOII. Option 5 (Utilize Existing Sampling Frame) (multiple years) was proposed to address potential survey data collection costs associated with sufficient annual sample sizes, and could be used with any of Options 1-5. Finally, Option 7 (panel survey) was proposed to address potential recall error associated with a 12 month recall period, and could be utilized with any of Options 2-6 (this approach would only be utilized with Option 1 (Supplemental Module in Existing Survey) for an existing panel survey [CPS, MEPS-HC]).

Option 1: Add a supplemental module to an existing survey

Under this option, a module of questions would be added to the existing survey questionnaire, to be asked following the completion of the existing survey questionnaire. To the extent possible, data collected as part of the existing survey questionnaire would be used in screening for eligible respondents; however, it may be necessary to include additional screening questions to the HSOII module.

To avoid changes to the data collection process for the existing survey, the assumption is made that:

1. The same within-HU sample selection rules used for the existing survey would be used for the HSOII module (e.g.., if the existing survey selects one person age 18+ years for the survey, the HSOII would also treat that same person as sampled for HSOII and would not collect data for other eligible residents within the HU).
2. The same respondent rules used for the existing survey would be used for the HSOII module (i.e., if each existing survey eligible person in the HU is interviewed, then the HSOII module would be conducted with each specific HSOII eligible person; if the existing survey interviews one HU respondent who reports data for all existing survey eligible persons within the HU, then the HSOII module would obtain data for all HSOII eligible persons from the HU respondent).

Under this option, the HSOII would be collected under the same time schedule as for the existing survey. Upon completion of data collection and data processing for the existing survey, HSOII data processing could begin. Data files required to be transmitted from the existing survey would include cleaned data from the existing survey (for specified data elements such as HU characteristics and person demographics), the raw data from the HSOII supplemental module, sample disposition codes from the existing survey and HSOII supplemental module, and the existing survey weight file (to include weights and adjustment factors from the existing survey weighting calculation). Note that the existing survey weight file could be delivered after the other data files, to allow HSOII data cleaning and editing to begin as early as possible.

HSOII data collected as a supplement to an existing survey could be affected by seasonality if data collection is not evenly distributed across the calendar year; such a situation could occur if the collection period is less than 12 months or if data collection is clustered nonrandomly in certain months. Refinements to this option could be considered, such as, if the existing survey does not collect data for all HSOII eligible persons, should HSOII data collection be limited to those persons jointly eligible for the existing and HSOII surveys, or should an attempt be made to collect data under the HSOII module for all HSOII eligible person, regardless of their interview status for the existing survey?.

By collecting HSOII data concurrent with collection of existing survey data, Option 1 will result in the lowest HSOII data collection costs among Options 1, 2,and 3.

Option 2: Use data from an existing survey to screen for persons eligible for the HSOII, select a subsample of HSOII eligible persons, and conduct a follow-on survey to collect any additional data required for the HSOII

Under this option, the HSOII would utilize data from the existing survey to identify HSOII eligible persons within the HU, after data collection for the existing survey has been completed and a data file of sufficient completeness and quality from the existing survey is available. This option assumes the data currently collected in the existing survey, along with data available on the sample frame and usual auxiliary data will be the only data available for use in the sample design and data collection initiation.

The extent of relevant and correlated data collected in the existing survey will affect the efficiency of the HSOII sample design. For example, if no data on and individual's industry/occupation were available, the HSOII sample design would not be able to a priori oversample selected industry and occupation groups to ensure sufficient sample sizes for each group, such oversampling would need to be built into the data collection process. In addition, the extent of information available from the existing survey would drive the extent of additional screening questions needed to be included in the HSOII.

Regardless of the extent of data available from the existing survey, it would be necessary to also select sample, albeit it at a lower sampling rate, from HUs with no HSOII eligible persons. This is necessary to ensure representation of the target population as persons may change work status between the existing survey interview and the HSOII interview, and residents within a HU may change.

Given the need to collect HSOII data during a follow-up contact with the HU, Option 2 would incur higher data collection costs than Option 1. These higher costs would be more significant for a CAPI approach than for a CATI approach.

Option 3: Add appropriate supplemental questions to the existing survey questionnaire to aid in screening for or stratifying persons eligible for the HSOII

Under this option, it is assumed additional screening questionnaires could be added to the existing survey questionnaire. The purpose of the additional screening questions would be to improve the efficiency of the HSOII sample design and selection by providing information that would allow for stratification of the existing survey sample as well as for oversampling of selected subpopulations needed to meet sample size and publication requirements.

By ensuring collection of relevant and correlated data in the existing survey, this Option will result in the most efficient HSOII sample design among Options 1-3. As with Option 2, regardless of the extent of data available from the existing survey, it would be necessary to also select sample, albeit it at a lower sampling rate, from HUs with no HSOII eligible persons

Given the addition of questions to the existing survey, Option 2 would incur higher data collection costs for a given sample size and collection Mode than Option 1 and likely than Option 2. However the improved efficiency of the sample design may offset the per complete collection costs and result in survey costs of a similar magnitude to those for Option 1.

Option 4: Utilize multiple surveys under Options 1, 2, and/or 3

This option would be used primarily if it is determined there is no single existing survey which can be used or which meets the planned HSOII budget under Options 1, 2, or 3, and that there are two or more other existing surveys which can be used but which individually do not meet the HSOII sample size requirements. This approach is analogous to that used in dual-frame estimation for RDD surveys selecting sample from both a landline and cell telephone sample frame.

Under this option, the multiple existing surveys can be thought of as independent samples of the population of interest. From a weighting and estimation perspective, the data from each sample is independently weighted to derive survey weights representing the total population. It is at this point the multiple samples are combined, with weights adjusted so as to minimize the variance associated with estimates from the combined samples. The adjustment made to the individual sample weights is proportional to the inverse of the individual variances; thus, samples with lower variance receive larger relative adjustments.

Survey weighting for the multiple survey approach is initially carried out separately for each survey. The final stage of weighting entails adjusting survey weights to combine the sample so as to minimize variance. Essentially, the weights from each survey are adjusted inverse to their variance, with weights for sample cases from the lower variance survey receiving a larger adjustment, and weights for sample cases from the higher variance survey receiving a smaller adjustment. The weight adjustments sum to one, and are derived as in the formula below.

$$\lambda_A = \frac{var\left(\widehat{Y_B}\right)}{var\left(\widehat{Y_A}\right) + var\left(\widehat{Y_B}\right)}$$

where A and B refer to the surveys.

This approach can be utilized for integration of more than two surveys. If information about the bias of each estimate is available, then mean squared error (MSE) is used rather than variance to derive the adjustment factor.

This approach can also be used to combine surveys that only partially overlap in terms of population, such as would be the case if NSFG (for which the eligible adult population is 18-44) were utilized. This is also the situation for dual-frame RDD telephone surveys, as the landline and cell telephone sample frames only overlap for dual use landline/cell telephone HU's. The landline telephone frame uniquely covers the landline only HU population, while the cell telephone frame uniquely covers the cell only HU population. In the case of partially overlapping existing surveys, weighting is again initially carried out separately for each survey, with weights derived appropriate to the overlap population and the non-overlap population. The final stage of weighting entails adjusting survey weights for sample representing the overlap population. Weights for the non-overlap population are not affected.

$$\lambda_{A,P_1} = \frac{var\left(\widehat{Y_{B,P_1}}\right)}{var\left(\widehat{Y_{A,P_1}}\right) + var\left(\widehat{Y_{B,P_1}}\right)}$$

Where P1 refers to the overlap population between survey A and survey B.

Option 4 would incur sample design/selection, questionnaire design, and data processing costs somewhat greater than those for a single existing survey approach. However, the efficiency of the dual-frame design may yield sample sizes sufficiently less than those from a single survey design to result in total survey costs not substantially higher than those of a similarly efficient single survey design utilizing the same collection mode.

Option 5: Utilize an existing sampling frame to develop an optimal sample design for the HSOII

This option makes use of an existing sample frame rather than existing surveys. The advantage of Option 5 is the access to the full target population for sample design and selection, as opposed to being limited to that sample selected and interviewed for an existing sample. This advantage becomes more important the larger the HSOII sample requirement and the more publication levels required.

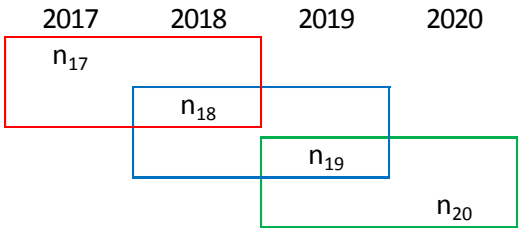
The HSOII sample design would only be able to utilize information available on the sample frame, which are limited to geographic information about the frame units, perhaps supplemented with data from auxiliary data sources. As a result, an HSOII sample design selected from one of the four sample frames would likely be less efficient than that of an HSOII sample design selected from an existing survey.

In addition, data collection costs can be expected to be noticeably greater for an HSOII sample design selected from one of the four sample frames, as all initial locating, sample clean up, etc., will be incurred by the HSOII rather than leveraging information from an existing survey. In addition, one might expect higher response rates when contacting sample cases interviewed previously than when contacting "virgin" sample cases.

Option 6: Implement any of Options 1-5 as a multi-year survey

Under this option, sample is combined across multiple years in creating an estimate for a given year (e.g., data collected in 2014 and 2015 would be combined to generate estimates for 2015, data collected in 2015 and 2016 would be combined to generate estimates for 2016, etc.). This approach is analogous to that utilized in the ACS for generating 3- and 5-year estimates.

As illustrated below, utilizing a two-year survey design, the sample (n17) from 2017 would be combined with the sample (n18) from 2018 to generate a two-year estimate for the period 2017-2018. The sample (n18) from 2018 would then be combined with the sample (n19) from 2019 to generate a two-year estimate for the period 2018-2019.



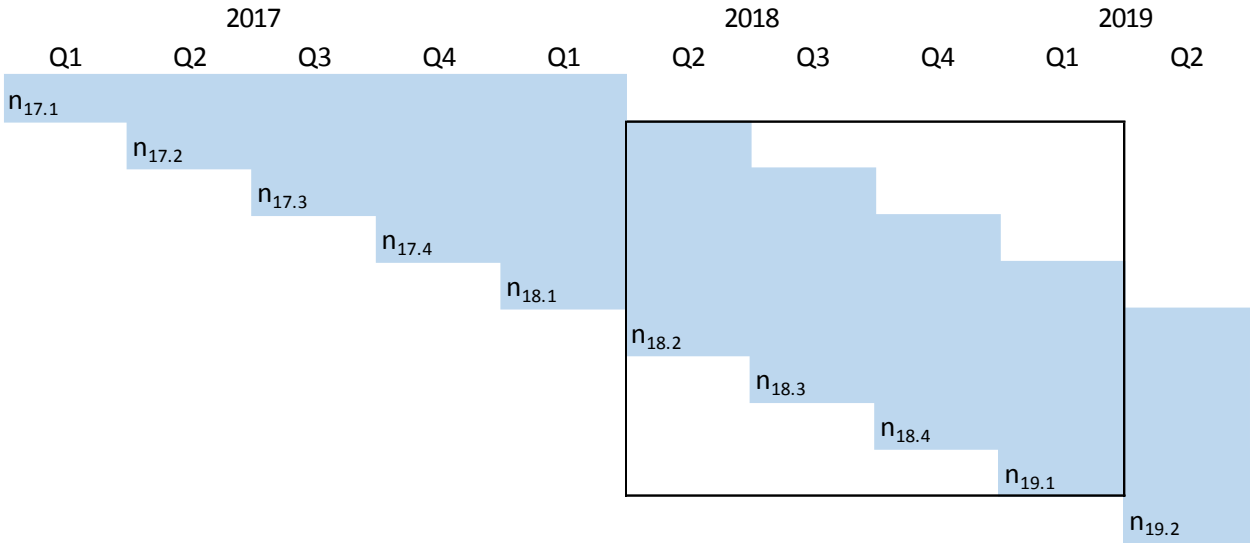
The advantage of the multi-year survey design is reduced data collection costs for a given sample size. In the illustration above, the data collection costs for each year would be roughly half those incurred to achieve an equivalent one year sample size of 2nYY for 20YY.

The consideration is that the estimates released each year would be for a two year period rather than for one calendar year. Given injury and illness incidence rates tend to be stable across time, this approach may be acceptable. One variation of this approach would be to generate national level, overall incidence rate estimates from the sample interviewed within the calendar year, and generate lower level estimates from the sample combined across years.

Option 7: Implement any of Options 2-6 as a panel survey

Under this option, selected sample units are interviewed multiple times across a designated period. This approach is analogous to CPS, wherein sample units are interviewed four consecutive months, and then interviewed the same four months the following year. Each month, new sample units are selected into CPS (a new rotation group, RG), and sample units having completed their eighth interview are dropped from the sample (a retired RG).

Illustrated below is a panel design with quarterly interviews. Each RG is interviewed 5 quarters before being retired, with a new RG selected and introduced into the survey each quarter. Thus in any given quarter, interviews would take place with sample from 5 RG's, with sample cases being asked about injuries and illnesses occurring in the prior quarter. Estimates for 2018 would be based upon data from 8 RG's, with two RG's contributing data for all 4 quarters in 2018, while the other RG's would contribute data for the number of quarters they reported for 2018. The net sample size would be $5n_{YY.RR}$.



The advantage of the panel survey design is reduced recall error. There may also be some reduced data collection costs as the locating and gaining cooperation would require less effort when interviewing sample cases a second, third, etc., time than when interviewing sample cases for the first time.

Appendix B. Links to Home Pages for Existing Surveys

Survey	Home Page Link
American Community Survey (ACS)	http://www.census.gov/programs-surveys/acs/
Current Population Survey (CPS)	http://www.bls.gov/cps/home.htm
National Health Interview Survey (NHIS)	http://www.cdc.gov/nchs/nhis.htm
Behavioral Risk Factor Surveillance System (BRFSS)	http://www.cdc.gov/brfss/
Medical Expenditure Panel Survey - Household Component (MEPS-HC)	http://meps.ahrq.gov/mepsweb/survey_comp/household.jsp
National Health and Nutrition Examination Survey (NHANES)	http://www.cdc.gov/nchs/nhanes.htm
National Longitudinal Survey of Youth (NLSY)	http://www.bls.gov/nls/
National Immunization Survey (NIS)	http://www.cdc.gov/nchs/nis.htm
Survey of Income and Program Participation (SIPP)	http://www.census.gov/sipp/
Survey of Consumer Finance (SCF)	http://www.federalreserve.gov/econresdata/scf/scfindex.htm
National Survey of Family Growth (NSFG)	http://www.cdc.gov/nchs/nsfg.htm
General Social Survey (GSS)	http://www3.norc.org/GSS+Website/
AmeriSpeak (AS)	http://www.norc.org/Research/Capabilities/Pages/amerispeak.aspx
KnowledgePanel (KP)	http://www.knowledgenetworks.com/knpanel/index.html

Appendix C. Definitions for Criteria Ratings

While the detailed information profiles the survey design environment for each existing survey, it is not conducive to easily comparing surveys nor to readily assessing the extent to which each existing survey fulfills needs for implementing the HSOII. To that end, the following definitions were created to provide for summarization of the survey design environment for each existing survey. These definitions are intended to represent relative, rather than exact, summary ratings. These ratings reflect the existing surveys as they are now configured, and do not imply modifications to existing surveys are not possible.

Criteria 1: Population Representation

Rating	Coverage		Sample Selection Frequency	Sample			Response Rate
	Frame Coverage Gaps	Estimated Coverage of Workers in Eligible Sample		Screened	Completed		
				HU's	HU's	Workers	
+++++	None	100%	Monthly or Quarterly	500.000+	350,000+	500.000+	90%+
++++	Non-telephone	90%-99%	Annually	100,000-499,999	70,000-349,999	100,000-499,999	75%-89%
+++	N/A	75%-89%	N/A	50,000-99,999	35,000-69,999	50,000-99,999	60%-74%
++	N/A	50%-74%	Biennial+	15,000-49,999	10,000-34,999	15,000-49,999	40%-59%
+	N/A	20%-49%	As Needed	5,000-19,999	4,000-14,999	5,000-14,999	25%-39%
0	N/A	<20%	N/A	<10,000	<7,000	<5,000	<25%

Criteria 2: Survey Mode and Use of Proxy Respondents

Rating	Data Collection	
	Sampled Persons	Respondent Rules
+++++	All persons 18+	Self-respondent
++++	N/A	N/A
+++	Sample of persons 18+	Self-report, if possible
++	N/A	HH respondent
+	N/A	N/A
0	N/A	N/A

Criteria 3: Availability of Key Information

Rating	Employment Information						Injury/Illness Information	
	Employment Status	Multiple Jobholders	Occupation	Industry	Employment Relationship	Sector	Injury	Illness
+++++	Prior 12 months	Yes	All jobs	All jobs	Yes	Yes	Accidents/ injuries at work	N/A
++++	Prior 3 months	N/A	Up to 2 jobs	Up to 2 jobs	N/A	N/A	Work days missed	Work days missed
+++	Current	N/A	One job	One job	N/A	N/A	Pain interfered w/ work activities	N/A
++	N/A	N/A	Optional module	Optional module	Optional module	Optional module	ER/provider visits; hospital stays; falls w/in 12 mo	ER/provider visits; hospital stays
+	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey	If asked as part of a given survey	Selected Chronic; If asked as part of a given survey
0	Not obtained	Not obtained	Not obtained	Not obtained	Not obtained	Not obtained	Not obtained	Not obtained

Criteria 4: Timeliness of Calendar Year Estimates

Rating	Collection Frequency for Sample Unit	Interview Length	Estimate Periodicity	Data Release Lag
+++++	Multiple times w/in 12 months	<15 min	Annual or more frequent	≤7 months
++++	Once w/in 12 months	16-25 min	N/A	8-9 months
+++	N/A	26-40 min	N/A	10 months
++	Once w/in 24+ months	41-59 min	Less frequently than annual	N/A
+	N/A	60+ min	N/A	11+ months
0	N/A	N/A	N/A	N/A