BLS Response to "Working Children: Federal Injury Data and Compliance Strategies Could Be Strengthened" (GAO-19-26)

David Biagas, Sarah Garland, Robin Kaplan, Kristen Monaco
U.S. Bureau of Labor Statistics
2 Massachusetts Avenue NE
Washington, DC 20212

(correspondence regarding this report should be directed to Kristen Monaco at 202-691-7527, Monaco.kristen@bls.gov)

1. Introduction/Executive Summary

GAO's Report "*Working Children: Federal Injury Data and Compliance Strategies Could Be Strengthened,"* (GAO-19-26), Recommendation 1 (p. 50) states, "The Acting Commissioner of the Bureau of Labor Statistics should, upon completion of the pilot Household Survey Occupational Injuries and Illnesses, evaluate the feasibility of measuring injuries and illnesses to certain worker populations—specifically, children aged 17 and under, child household workers, and those employed on farms with 10 or fewer workers—in any final survey scope, or determine a way to gather information on these populations. (Recommendation 1)."

In response, the Bureau of Labor Statistics (BLS) indicated that it would deliver a report with, "an assessment of the suitability of the existing Household Survey of Occupational Injuries and Illnesses (HSOII) instrument to gather data on working children" as well as a review of the literature on questionnaire design focused on sensitive topics and proxy reporting.

This report covers those topics and addresses the broader issue of survey design for the domains identified by Government Accountability Office (GAO), namely children workers, including householder workers and children working on small farms. The structure of the report is as follows: section 2 presents background on occupational safety surveillance and existing data collection efforts at BLS, section 3 presents information on survey design for the HSOII as well as the domains recommended by GAO, section 4 presents operational challenges involved in collecting data that cover sensitive issues related to children, and section 5 summarizes the conclusions.

The report concludes that GAO-identified domains can be classified as mini or rare domains given their low prevalence in the U.S. working population. For this reason, direct estimates from the HSOII in a design-based approach may not be effective. These smaller domains may be more suited to a model-based approach which combines data from multiple sources using statistical modeling techniques. The operational challenges from surveying parents about the safety of their children further reinforces the conclusion that model-based estimates to generate reliable estimates for the GAO-identified domains, rather than modifying a design-based approach such as the HSOII, is the preferred approach.

2. Background

2.1 The Occupational Safety and Health Act

The Occupational Safety and Health (OSH) Act of 1970 recognized the importance of a broad and systematic occupational safety surveillance system and charged the Departments of Labor and Health and Human Services with collecting comprehensive statistics on workplace safety, researching occupational safety surveillance, and establishing and enforcing workplace safety standards.

Stemming from this, the Secretary of Labor charged BLS with collecting comprehensive statistics. The Occupational Safety and Health Administration (OSHA) was charged with regulation and enforcement, and the Secretary of Health and Human Services created the National Institute of Occupational Safety and Health (NIOSH) to focus on research.

The BLS had collected data on occupational safety and health before the OSH Act. Special studies on workplace safety were conducted by BLS in the 1900s and expanded in the 1940s to include broader data collection efforts. These efforts were still narrow in scope – considering severe injuries that resulted in death, permanent impairment or temporary disability. After the OSH Act, BLS developed the Survey of Occupational Injuries and Illnesses (SOII).

OSHA coverage is a bit complex and merits some explanation. The OSH Act covers most private sector workers and some, but not all, public sector workers.[1] OSHA coverage stems from two sources: Federal OSHA or OSHA-approved state plans. For states that have their own occupational safety and health programs (currently 22), OSHA monitors the plans in the state, but does not directly provide enforcement. In cases where state plans do not include all workers, Federal OSHA provides coverage.

Out of scope for the OSH Act are the self-employed, immediate families of farm workers, and workplaces regulated by a different federal agency, MSHA and Department of Energy, to name two. All employers are mandated to report the following case types to OSHA: worker death from a work-related incident and work-related inpatient hospitalizations, amputations, and eye loss. Additionally, OSHA requires recordkeeping of work-related injuries and illnesses on OSHA Forms 300, 300A, and 301 for employers at establishments with more than 10 employees in many industries. Those employers are also required to post the Summary of Work-Related Injuries and Illnesses in the workplace every year.

2.2 SOII Background and Overview

The SOII was established in 1972 as a Federal-State cooperative program, focused on publishing estimates of injury and illness counts and rates by industry at the state level and for the nation annually. The initial survey design did not capture detailed information about the characteristics of the injuries and illnesses. Fatal injuries were also collected through survey.

In 1987, a National Academies of Science panel recommended that SOII be expanded to include detailed case characteristics and that the collection of information on fatal cases be moved away from a survey approach to a full accounting of fatal injuries through a census that relied on multiple source documents, including administrative records.[2]

[1] For more information about OSHA coverage, see https://www.osha.gov/Publications/all_about_OSHA.pdf and https://www.osha.gov/law-regs.html.

[2] "Counting Injuries and Illnesses in the Workplace: Proposals for a Better System" https://www.nap.edu/catalog/18911/counting-injuries-and-illnesses-in-the-workplace-proposals-for-a

These recommendations were implemented in 1992. The underlying scope of these collections, however, has remained relatively constant since the OSH Act. The definitions used by the SOII for data collection mirror OSHA's definitions. The rationale for this is two-fold: to limit respondent burden and to leverage employer knowledge of OSHA recordkeeping definitions. The criteria for a case to be counted are workplace injuries that result in any of the following:

1. Loss of consciousness;
2. Days away from work;
3. Restricted work activity or job transfer; or
4. Medical treatment beyond first aid.

Employers are also required to record illnesses or injuries diagnosed by a physician or heath professional, cases involving needlesticks or cuts that are contaminated by a another person's blood, tuberculosis infection, hearing tests that fall below certain thresholds, or other OSHA criteria.

2.3 HSOII Design, Domains of Interest and Fielding Summary

Since 2009, the BLS has conducted research into the completeness of the SOII, stemming from concerns that not all in-scope cases were included in data submitted by establishments to BLS.[3] There is no consensus on the reasons for establishment under-reporting, but they include lack of consistent recordkeeping over the year, gaps in understanding OSHA's recording criteria, and potential disincentives for both establishments and workers to report injuries and illnesses. One proposal for overcoming potential filters between the worker, establishment recordkeeping, and establishment reporting to BLS was to approach workers directly to collect information on workplace injuries and illnesses.

In fiscal year (FY) 2015, BLS contracted with Westat to review related surveys and recommend potential designs for a survey with the goal of capturing work-related injuries and illnesses directly from workers that result in nationally representative estimates that could be compared to the SOII.[4] The Westat report provided analysis of the pros and cons of different modalities (face-to-face, mail, telephone) and structures (stand-alone survey, adding a module to an existing survey, developing a follow on survey). Westat recommended a pilot survey noting that the pilot structure would likely differ from the final survey structure. They recommended a pilot would have utility if it provided information on response rates, differences in recall periods, differences in responses by modes of collection and measures of proxy reporting quality.

In FY 2016-2017, BLS contracted with NORC to further evaluate the suitability of existing surveys and frames to capture data on work-related injuries and illnesses.[5] They also provided reports on survey design options and an evaluation of existing questions/instruments on workplace injuries and illnesses. This contract also involved developing and cognitively testing a survey instrument. The goal of the survey instrument was to capture the same elements as the SOII and provide additional details that

[3] For additional information on this research, see https://www.bls.gov/iif/data-quality.htm.
[4] "Pilot Study Design for the SOII Employee Survey" https://www.bls.gov/iif/westat-worker-survey-pilot-design-final-report.pdf.
[5] "Designing a Household Survey of Injuries and Illnesses" https://www.bls.gov/iif/norc-final-report.pdf.

would inform data quality research. These additional measures included constructs to capture whether the worker reported the injury or illness, received medical care, or filed for workers compensation. There were three rounds of cognitive testing used to both adjust survey questions for comprehension and test for the quality of proxy reporting.

2.4 HSOII Pilot Results

In FY 2018, BLS contracted with ICF to do a large-scale pilot test of the survey instrument developed by NORC. The pilot test targeted a sample of roughly 4,000 individuals and was conducted via telephone between August 2017 and July 2018, using a mix of landline (10%) and cell phone frames (90%). The pilot included a pretest of 50 completed interviews and was used to identify questions that were confusing or difficult to answer and identify any problems with skip patterns in the survey instrument. The interviewers for the pilot received classroom training specifically related to the survey and completed practice interviews.

The scope of the survey was individuals 18 years of age or older who had worked in the 12 months prior to the interview. The screener question was, "In the last 12 months did you do ANY work for pay?"[6] Among the 7,943 individuals reached by ICF interviewers, 666 had unknown eligibility, 3,100 were ineligible based on their response to the screener, and 4,177 were eligible for the survey. Of the 4,177 eligible, there were 3,798 completed surveys, 314 break-offs, and 65 incomplete due to refusal or language barrier.[7] Based on the total number of contacts attempted, the 3,798 completed interviews yielded a response rate of 14.3%. This is based on American Association of Public Opinion Research, Response Rate 3, which include assumptions of eligibility for cases of unknown eligibility.[8]

To capture occupational injuries or illnesses, two questions were asked immediately following the screener question. The first was, "In the last 12 months have you experienced any injuries or illnesses related to any job you held?" There were 4,103 respondents to the first question, 340 of whom answered "yes," 3,737 "no," and 26 "don't know" or refused to answer the question. Those who did not respond "yes" were asked, "Just to make sure, I'm going to read some examples of work-related injuries and illnesses … In the last 12 months, have you experiences any of these, or other types of injuries or illnesses, related to any jobs you had?" There were 3,706 responses to respondents to the second question, 417 of whom answered "yes", 3,256"no" and 33 "don't know" or refused to answer the question. This is notable for two reasons. First, the follow up question captured more "yes" answers than the initial question. Second, there were 47 break-offs between the two questions.

While the total number of workers answering yes to either question suggests an unweighted injury and illness rate of 18% (757/4,177), these questions were not designed to correspond to the definition of OSHA-recordable injury or illness. After the pilot was complete, two experienced BLS staff members ("coders") who specialize in coding SOII cases were tasked with reading the narrative description from each respondent and identifying whether the case was OSHA recordable. Many narratives did not contain adequate information to determine OSHA-recordable status. For this reason, the staff focused on cases determined to be OSHA-recordable and combined the not-recordable and unknown cases into

[6] The probe for this is, "this would include salary, compensation, or other profit."
[7] The instrument was programmed in CATI in both English and Spanish.
[8] For further information on AAPOR response rates, see The American Association for Public Opinion Research. 2016. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition*. AAPOR. https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

a second classification. The coders agreed that 145 cases of the 757 were OSHA-recordable (19%). This suggests an overall, unweighted prevalence of 3.5% (145/4177). It is important to note that the coders had disagreement in 24% of cases, with one coder determining the case was OSHA-recordable and the other determining the case was either not recordable or unknown. These coding disagreements were fairly symmetrical. Calculating the prevalence rate from the coder who most frequently assigned the OSHA-recordable code suggests an unweighted prevalence weight of 5.8%.

Break-offs were a substantial problem at early points in the survey, an issue not identified in the pilot test. Of the 757 "yes" responses to the injury and illness questions, 131 were break-offs and many of these (64) occurred prior to the question that asked about the nature of the injury or illness, resulting in no information about these cases, including a determination of whether they were OSHA-recordable. The break-off rate for cases where a respondent indicated they had experienced an injury or illness was 17%, versus 6% for those who responded "no" to both questions related to an injury or illness.

The higher rate of break-offs among those reporting workplace injuries or illnesses, along with the high rate of unknown OSHA-recordable codes assigned by the BLS coders, suggests that item non-response and response quality in the workplace injury or illness narratives must be further examined and addressed in further HSOII tests. For this reason, we will not produce weighted estimates of these prevalence numbers.

3. GAO Recommended Domains and Implications for Survey Design

In this section, we consider the size of the GAO-recommended domains, the prevalence of injuries and illnesses within these domains, and the estimation methodologies appropriate for producing estimates of injuries and illnesses for these domains. We consider both design-based and model-based methods. Design-based estimation methodology relies on a probability-based sampling method to estimate the target value based only on data collected from sampled units. Model-based estimation methodology leverages data collected for a domain through a probability-based sample design as well as auxiliary data to estimate the target value.

3.1 Domain Identification and Size Classification

To identify the size of the relevant domains, this study relies heavily on the American Community Survey (ACS). The ACS is a continuous national household survey of the U.S. Census Bureau used to produce population, demographic, and housing unit estimates.[9] According to the ACS's published results, in 2017, there were an estimated 155 million workers in the United States, aged 16 years and older. The ACS published estimates do not represent children workers less than 16 years of age. This is close to the target population of U.S. workers for the HSOII, though the HSOII pilot only collected data for those 18 years of age and older.

This study focuses primarily on industry domains, as industry domains are the focus of the HSOII and can be used as a starting point to understand the relative size of the domains. Adding additional dimension to the domains, such as age categories, will decrease the size of domains and increase the complexity of sampling and estimation.

[9] https://www.census.gov/programs-surveys/acs/about.html

Table 1, column 2 presents the 2017 ACS estimates of number of workers in the United States aged 16 and older by industry sector.[10,11] Column 4 presents an estimate of the share of workers in each industry. Column 5 identifies the domain size for each industry for all U.S. workers. Kish defines domain sizes as major for 10% or more of the population, minor for 1% to 10% of the population, mini for 0.1% to 1% percent of the population, and rare for less than 0.01% percent of the population (Kish, 1980). In general, sufficient sample sizes to estimate the overall population, major domains, and some minor domains are part of the survey design of national surveys.

Table 1: 2017 ACS Estimates of U.S. Employment by Industry, Aged 16 and Above

Industry*	Estimate*	Margin of Error*	Percent of U.S. Workers†	Domain Size Classification†
Civilian employed population 16 years and over	155,058,331	+/-139,282	100%	
Agriculture, forestry, fishing and hunting, and mining:	2,637,326	+/-29,825	1.7%	Minor
Agriculture, forestry, fishing and hunting	1,906,427	+/-25,184	1.2%	Minor
Mining, quarrying, and oil and gas extraction	730,899	+/-16,275	0.5%	Mini
Construction	10,292,425	+/-64,278	6.6%	Minor
Manufacturing	15,631,115	+/-62,066	10.1%	Major
Wholesale trade	3,984,192	+/-33,563	2.6%	Minor
Retail trade	17,342,338	+/-69,643	11.2%	Major
Transportation and warehousing, and utilities:	8,343,526	+/-48,159	5.4%	Minor
Transportation and warehousing	7,070,552	+/-44,942	4.6%	Minor
Utilities	1,272,974	+/-18,427	0.8%	Mini
Information	3,135,019	+/-33,512	2.0%	Minor
Finance and insurance, and real estate and rental and leasing:	10,227,159	+/-55,413	6.6%	Minor
Finance and insurance	7,177,561	+/-46,301	4.6%	Minor
Real estate and rental and leasing	3,049,598	+/-31,137	2.0%	Minor
Professional, scientific, and management, and administrative and waste management services:	17,865,131	+/-71,214	11.5%	Major
Professional, scientific, and technical services	11,092,877	+/-60,019	7.2%	Minor
Management of companies and enterprises	201,692	+/-7,672	0.1%	Mini
Administrative and support and waste management services	6,570,562	+/-45,898	4.2%	Minor

[10] The ACS uses the North American Industry Classification System (NAICS) for industry classifications, as does the SOII. https://www.bls.gov/bls/naics.htm

[11] The entire U.S. population is the full scope of the ACS. However, for the purposes of this feasibility study, we will consider the full scope to be U.S. workers only, not the working and non-working components of the population. It is possible to complete this section's analysis on the full U.S. population, which would reduce the sizes of the domains, but would present additional challenges in methodology discussions. Therefore, only the population of U.S. workers will be used here.

Educational services, and health care and social assistance:	35,805,182	+/-115,862	23.1%	Major
Educational services	14,172,748	+/-85,813	9.1%	Minor
Health care and social assistance	21,632,434	+/-74,715	14.0%	Major
Arts, entertainment, and recreation, and accommodation and food services:	15,071,444	+/-79,742	9.7%	Minor
Arts, entertainment, and recreation	3,397,600	+/-31,254	2.2%	Minor
Accommodation and food services	11,673,844	+/-78,538	7.5%	Minor
Other services, except public administration	7,596,464	+/-53,443	4.9%	Minor
Public administration	7,127,010	+/-53,167	4.6%	Minor

*Source: U.S. Census Bureau, 2017 American Community Survey 1-Year Estimates.
†Percent and Domain Size columns added based on ACS point estimates.

While HSOII focused on industry domains for comparability to SOII, GAO has expressed interested in domains for child workers. GAO-19-26 provided some estimates on the size of these domains. The Current Population Survey (CPS) contains data on children aged 15 through 17 who work, including household children working 15 or more hours per week with or without pay. In 2017, there were an estimated 2.5 million children working during summer months and an estimated 2.9% of working children (about 72,000) working in agriculture in the summer months (GAO-19-26, pp.18, 23). Based on these estimates, working children would be classified as a minor domain and children working in agriculture as a rare domain.

The NIOSH Childhood Agricultural Injury Survey (CAIS) collects data on working children under 18 years of age on U.S. farms. In 2014, the estimated number of children working in agriculture was about 524,000. Roughly 375,000 were working household children, and, of these, an estimated 246,000 were aged 14 years or less (GAO-19-26, pp. 23-4). These estimates correspond with a mini domain.

While the estimates vary between the CPS and the CAIS due to differences in scope and variable definitions, we can conclude that the size classifications for the GAO-identified domains are minor to rare. Mini and rare domains generally require model-based estimation methods, rather than design-based (Valliant, Dever, & Kreuter, 2013; Kalton, 2009; Mohadjer & Curtin, 2008; Marker, 2001).

3.2 Injury and Illness Prevalence

This study next considers the prevalence of the character trait in the population (Valliant, Dever, & Kreuter, 2013). The number of workplace injuries and illnesses is the character trait of interest for the HSOII. To calculate sample sizes required to produce publishable estimates, we use the expected prevalence of this character trait (*i.e.*, how often the trait occurs in the population or subpopulation). For this, we use the injury and illness rates from the SOII. Though the SOII has different scope than the HSOII, and a less expansive definition of occupational injury or illness, it is the only national source of data on prevalence by industry.

Table 2: Incidence rates of nonfatal occupational injuries and illnesses by industry, 2017

Industry	Total Recordable

	Cases per 100 FTE
All Private Industry	2.8
Agriculture, forestry, fishing and hunting, and mining:	
Agriculture, forestry, fishing and hunting	5.0
Mining, quarrying, and oil and gas extraction	1.5
Construction	3.1
Manufacturing	3.5
Wholesale trade	2.8
Retail trade	3.3
Transportation and warehousing, and utilities:	
Transportation and warehousing	4.6
Utilities	2.0
Information	1.3
Finance and insurance, and real estate and rental and leasing:	
Finance and insurance	0.5
Real estate and rental and leasing	2.4
Professional, scientific, and management, and administrative and waste management services:	
Professional, scientific, and technical services	0.8
Management of companies and enterprises	0.9
Administrative and support and waste management services	2.2
Educational services, and health care and social assistance:	
Educational services	1.9
Health care and social assistance	4.1
Arts, entertainment, and recreation, and accommodation and food services:	
Arts, entertainment, and recreation	4.2
Accommodation and food services	3.2
Other services, except public administration	2.1
State and local government	4.6

*Source: Bureau of Labor Statistics, U.S. Department of Labor, Survey of Occupational Injuries and Illnesses, https://www.bls.gov/iif/oshwc/osh/os/summ1_00_2017.htm

Assuming a simple random sample with replacement of U.S. workers, the number of injury and illness cases, the character trait of interest, expected for a given sample size of size n, is $p * n$ (Valliant, Dever, & Kreuter, 2013). The private sector worker expected prevalence rate is 2.8. If we took a simple random sample of 1,000 U.S. workers with replacement, the expected number of workplace injuries or illnesses is 28, lacking power to produce reliable estimates. A simple random sample size in the tens of thousands is required to capture enough workplace injuries and illnesses to generate estimates at the broad industry group, the domains of interest for HSOII.

Turning to the GAO-recommended domains, there is limited information on prevalence in GAO-19-26. The CAIS average annual estimated number of injuries to children working in agriculture is 2,372 (4,744 cases in 2012 and 2014, p.35). In 2014, there were an estimated 524,000 children working in agriculture, according to the CAIS (GAO-19-26, pp. 23-24). This gives an expected prevalence rate for the character trait of 0.5%. In 2016, the NEISS-Work estimate of 26,300 workplace, emergency department-treated injuries and illness to children less than 18 years of age leads to a prevalence rate of 1.0% if using the 2.5 million CPS estimate for number of working children (GAO 19-26, p. 93).

Both examples suggest a sample within the desired domain in the tens of thousands. Unlike the HSOII which requires a sample of comparable size, among a population of all workers, the samples for the GAO-recommended domains suggest a large sample within a population of working children, which is a challenging population from which to identify a large sample. This raises logistical and cost concerns given than these are small domains relative to the overall population. Collecting a sufficient number of observations within these minor and rare domains requires either a more sophisticated sampling approach or model-based approaches. Each is considered in turn.

3.3 Approaches for Generating Estimates of Minor and Rare Domains

3.3.1 Design-Based Approaches

To successfully sample the GAO-identified domains for reliable direct estimates of workplace injuries and illnesses, the HSOII's sampling methods based on industry domains is insufficient. However, probability-based sampling methods for rare domains might enable the use of direct estimation methodologies (Kalton, 2009) (Marker, 2001).

Stratification

To produce estimates for multiple domains, statisticians use stratified sampling to allocate the sample such that estimates with a pre-specified reliability thresholds are likely to be produced within each domain of interest (Valliant, Dever, & Kreuter, 2013). There are well-established sampling and analysis methodologies associated with stratification, however these require that the sampling frame contains information that can be used to identify the strata relevant to each domain. Absent a large-scale frame that includes both industry and age, which are required for the GAO-identified minor and rare domains, it is not feasible to generate reliable estimates for these domains using stratified sampling.

Oversampling

Oversampling permits sampling a domain at a rate higher than the group occurs in the general population via frame information or adding an additional stage in the survey design. Large-scale household surveys often employ oversampling to study domains of interest (Valliant, Dever, & Kreuter, 2013). For example, the 2006 sample of the National Health and Nutrition Examination Survey (NHANES) oversampled several subpopulations such as pregnant women, the elderly, adolescents, and low-income individuals, among others (Mohadjer & Curtin, 2008). However, when using oversampling to gain precision for one domain (or set of domains), given a fixed total sample size for the study, there will be a loss of precision across all other domains. Oversampling proportions must balance the need to estimate the minor or rare domain with the need to produce reliable estimates for the other domains.

Using the HSOII sampling approach and incorporating the GAO-identified domains requires sampling from the population of households identified as having at least one worker. There are multiple methodologies that might be used to oversample the GAO-identified domains from this population.

- Screening Methods, Two-Phase Design: If a sampling frame does not contain identifiers for the domains of the HSOII, nor for the GAO-identified domains, a set of screener questions is necessary at the point-of-contact to determine to which domains the respondents belongs. To leverage the largest sample size possible in this situation, one can also sample multiple people in a household who pass the screener, such as more than one working child. The size of the required sample depends on multiple factors, such as how small the domain is and how rare the characteristic we are measuring. Screeners can become expensive relative to the rarity of the domain and trait characteristics. In other words, the rarer the group is, the larger the net required to capture the group (Kalton, 2009). An additional concern in using screeners for small domains data collection is noncoverage (*i.e.*, the frame coverage of the mini and rare domains is simply inadequate). For example, in some commercially-available address-based frames, there is documented undercoverage in rural areas (Valliant, Hubbard, Lee, & Chang, 2014).

- Multi-Year Design: A less expensive method to generate reliable direct estimates for smaller domains is to combine multiple years of a continuously running survey. The estimates would no longer be representative of a point in time. For mini or rare domains, multi-year estimates may need to span many years, however, the characteristics of these groups can change substantially in short periods of time. The potential changes across times and the effect on interpretations and uses of the estimates requires consideration (Kalton, 2009). From GAO-19-26 we know that the number of working children fluctuates within each year, with more children working during the non-school months, while the proportion of children working in all seasons has fluctuated across time.

- Area-Based Methods: A small deviation from an area-based sampling method (Valliant, Dever, & Kreuter, 2013) is supplementing an address-based sampling frame with auxiliary frame data, such as Census Bureau information or farm information available from the National Agricultural Statistical Service (NASS), by classifying each address on the frame as in an urbanized area, an urban cluster, or rural.[12] Then for the GAO-identified domain for small farms, oversampling could occur in the rural classification addresses, as the probability of locating small farms wth working children would increase. However, for children working outside of the agriculture industry this method has limited benefit.

- Multiple Frame Design: To incorporate an adequate number of small farms into the HSOII, one approach is to use more than one sampling frame. If the HSOII's main sampling frame is an address-based frame, then we can use one (or more) additional frames to oversample the mini and rare GAO-identified domains. For example, the NASS combines multiple sources of information to create a list frame all U.S. farms and supplements this with land-segment sampling from an area frame (National Academies of Sciences, Engineering, and Medicine, 2017). Using this frame in addition to an address-based sampling frame for the HSOII, would allow for targeted oversampling of farms to identify small farms, any working household children on those farms, and the injuries and illness associated with working children. However, while this frame would have a higher probability of finding small farms, finding the character

[12] https://www.census.gov/geo/reference/urban-rural.html

trait of a workplace injury and illness would still present its own sampling challenges to identify enough respondents, and larger sample sizes remain necessary (Marker, 2001) (Kalton, 2009). For working household children outside of the agriculture industry, we would need to identify other sampling frames or employ other oversampling methods.

- Multipopulation Design: We could consider the GAO-identified domains that do not align with the target population of the HSOII as separate populations and use a multipopulation survey. In a multipopulation study, each population has a separate, flexible, probability-based sampling design, utilizing the sampling frames, methods, and budget allocations that accommodate each population. In a multipopulation design, the desired statistical comparisons are a driving force in sample size calculations. If we define the GAO-identified and HSOII populations as separate populations, these populations could simply be studied in separately designed and analyzed studies, without affecting the goals of either study (Kish, 1994). Note that this also implies that the same statistical agency need not field both surveys (i.e. BLS could administer the HSOII and NASS or NIOSH could administer the child worker study).

3.3.2 Model-Based Approaches

Domains in the mini and rare size classes present challenges that we cannot always address through sampling methods. Model-based estimation methods are a useful set of tools for estimating target values in these small domains when there is insufficient collected data for direct estimates. The model-based methods leverage auxiliary data to increase the inference capabilities for the domains of interest. Rao, *et. al.* provides a detailed foundation for the field of small area estimation, also known as small domain estimation (Rao & Molina, 2015).

There are two broad types of small area models: area-level and unit-level. Area-level models generate estimates for the small domains as a whole, while unit-level models work at the microdata level. Small area models leverage similarities in auxiliary data to generate small domain estimates, but have the advantage of understanding additional sources of variation. Small area models can become fairly complex, but also often result in reliable estimates for small domains where direct estimates were impossible (Rao & Molina, 2015).

Datasets that might provide auxiliary data for small area estimation models of the GAO-identified domains include:

- NEISS-Work: This is a national probability-based survey of U.S. emergency room departments. and collects data on occupational injuries for civilian, non-institutionalized workers treated in emergency-departments in the United States. Worker demographics are available in the data as well as injury-related variables; however, the survey does not collect industry and other job characteristics.[13]

- Childhood Agricultural Injury Survey (CAIS): This is the major survey for the Childhood Agricultural Injury Prevention Initiative at NIOSH. The survey, joint between NIOSH and NASS, studies youth farm workers, gathering information about the type of farm, type of children

[13] https://wwwn.cdc.gov/wisards/workrisqs/Default.aspx

workers, and injuries and illnesses occurring during work. The survey's results are available for years 1998, 2001, 2004, 2006, 2009, 2012, and 2014.[14]

- ACS: While the ACS does not collect information on workplace injuries and illnesses, there is substantial data available at smaller domains for employment and demographics.[15]

- Farm Labor survey: NASS collects employment and wage data on all U.S. farms that have more than $1,000 in annual sales.[16]

While models would need to be evaluated and modified for workplace injuries and illnesses estimates within the GAO-identified domains, such models lay a sufficient foundation to start the work required to generate viable estimates. Much more can be learned about these subpopulations from small area models than from a study that is relying heavily on oversampling for direct estimates. The models typically require some limited collected survey data. The HSOII could be used to collect a small amount of data in these domains to help inform these models.

4.Nonresponse and Data Quality Considerations

While section 3 focused on challenges in identifying a sample design to generate reliable estimates for the GAO-identified domains, it did not address operational complexities associated with collecting data from respondents in these domains.

Williams and Brick (2018) find response rates have decreased over time for household surveys. The trend is not as severe for in-person surveys as telephone surveys. The response rates for in-person interview surveys ranges from less than 70 percent to almost 80 percent and appear to still be declining. (Williams & Brick, 2018). While there is some information on response rates by industry from the HSOII pilot test, the expected nonresponse to the GAO-identified domains requires additional study. However, it is reasonable to assume higher levels of nonresponse to questions involving injuries among children. This nonresponse not only affects sample allocation but can be a source of bias in the resulting estimates.

Nonresponse issues are intertwined with other operational challenges associated with the collection of child worker injuries, and these challenges are generally more prevalent within the domain of child workers in agriculture. These issues pertain to (a) proxy reporting, (b) question sensitivity, and (c) logistical concerns in the context of an agricultural survey.

Proxy reporting

Proxy reporting occurs when a single member of a household or business reports on behalf of other members. This is commonly done in surveys to save on time, costs, and nonresponse in the context of collecting data on youth, when the population of interest is unable to report for themselves. The accuracy of proxy reports is difficult to assess due to a lack of systematic studies, however most research indicates proxy reports between spouses and close family members tend to show at least moderate agreement (Krosnick et al., 2015). However, proxy reporting is associated with measurement error due

[14] https://www.cdc.gov/niosh/topics/childag/CAIS/default.html
[15] https://www.census.gov/programs-surveys/acs/about.html
[16] https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Farm_Labor/index.php

to knowledge and recall issues, as well as privacy and confidentiality concerns, all of which can negatively affect data quality (Lee et al., 2004; Tourangeau, 1984; Bickart et al., 1990).

Proxy reporting in the context of collecting youth injury and illness data may lead to additional challenges. Because many farms are family-owned businesses, the head-of-household is also likely to be the farm owner or co-owner. This may lead to conflicts where the respondent reporting on injuries and illnesses for all employees at the farm is also the parent of the youth(s) working there. The validity and reliability of parental proxy reports on their children's health conditions and injuries is mixed, showing moderate concordance and sometimes significant discrepancies (Garbarski, 2014; Ecob et al., 1993). For example, parents often underreport on negative aspects of their children's health status and overreport positive aspects (Erhart et al., 2009; Reynolds & Wenger, 2012). Proxy reporting on children is seen as a sensitive topic and can elicit privacy and confidentiality concerns that lead to threats to data quality (Mingay et al., 1994). Often, parents report feeling uncomfortable providing information about their young children for fear of disclosure to third parties (King et al., 2012). This can lead to non-response, underreporting, and poor data quality (Todorov and Kirchner, 2000).

Another challenge in proxy reporting relates to knowledge and recall issues. Unless the respondent has kept very good records on all employees' injuries and illnesses, questions about these injuries are likely to rely on recall. Depending on the severity of the injury and how long ago it occurred, injuries could be underreported (Warner et al., 2005). For instance, recent and severe injuries are more likely to be noticed and recalled than minor and more distant injuries via proxy reports (Sudman et al., 1996).

Question sensitivity

Topics that are sensitive, perceived as threatening, or elicit fear of disclosure to third parties also tend to be underreported (Tourangeau & Yan, 2007). Because proxy reporting on children, as well as child work-related injuries and illnesses are also considered sensitive topics, this leaves the question open as to the data quality that could be obtained from collecting data on childhood agricultural injuries and illnesses. For instance, parents may not be willing to accurately report this information about their children or other children who may be employed at their farm. This was seen in the CAIS administered by CDC, NIOSH, and NASS. They found that the respondents tended to be female head of households and that this may have led to an underreporting of hired youth injuries. This was likely worsened for large farms or those employing a large number of workers or their own children. Motivated underreporting likely occurred in the CAIS, but there was no way to validate the accuracy of the results. Thus, the sensitive nature of reporting on youth injury and illness in the workplace is likely to create data quality issues, including underreporting and item non-response.

Logistical concerns in collecting agricultural data

The SOII collects injury and illness data via self-administered modes from employers. Employers report only on farms with 10 or more employees. Thus, small farms are currently out of scope for collection in the SOII. Collecting youth agriculture injuries would require a significant shift in the scope and sample that SOII uses. Other agencies, such as NASS and NIOSH are better structured to collect such data.

NASS reports suggest that surveying farm populations has become increasingly difficult across time. Response rates to several NASS surveys, including the Agricultural Resource Management, Crops Stocks, and the June Area Survey, have been on the decline in recent years (e.g. Mitchell, Ott, Ridolfo, and McCarthy, 2015; Ridofo, Boone, and Dickey, 2013). NASS has examined several strategies to combat this

trend (e.g. incentives, caller id experiments, and targeted data collection efforts), which have been met with mixed results.

In addition to dealing with increasing unit non-response, the Agricultural Resource Management Survey, which asks for sensitive financial information, has high item non-response on over 30 questions of interest (Miller and O'Connor, 2012). Although difficult to fully assess, researchers have identified some reasons for non-response to a non-invasive agricultural survey. The top five reasons for not responding among the agricultural population of interest were that 1) they do not like or do surveys, 2) they are too busy and do not have the time, 3) the information requested is too personal and none of the government's business, 4) I believe surveys hurt farmers, and 5) I will have nothing to do with the government (Gerling, Tran, and O'Connor, 2010). The reasons provided speak to the respondent burden experienced by agricultural respondents, their perceptions that the questions currently asked of them are already overly sensitive, and their general reluctance to complete surveys administered by the federal government.

Given the aforementioned challenges associated with proxy reporting and sensitive topics, which are compounded by the general reluctance of farmers to participate in federal surveys, BLS is not poised to collect information regarding the injuries experienced by child workers in agriculture.

5. Conclusions

GAO's Report "*Working Children: Federal Injury Data and Compliance Strategies Could Be Strengthened,"* (GAO-19-26), Recommendation 1 (p. 50) states, "The Acting Commissioner of the Bureau of Labor Statistics should, upon completion of the pilot Household Survey Occupational Injuries and Illnesses, evaluate the feasibility of measuring injuries and illnesses to certain worker populations—specifically, children aged 17 and under, child household workers, and those employed on farms with 10 or fewer workers—in any final survey scope, or determine a way to gather information on these populations. (Recommendation 1)."

This report provides a feasibility assessment of including the GAO-identified domains in the HSOII to directly collect data on working children, including those working on farms. This report first assesses survey design approaches required to produce GAO-identified domains and concludes that even complex survey design approaches are highly unlikely to generate estimates through direct data collection. This conclusion is supported through an analysis of the operational challenges of collecting data on these sensitive topics, including a discussion of how non-response and proxy reporting can lead to bias.

While the HSOII is not a suitable platform for producing estimates of the GAO-identified domains, section 3.3.2 notes that model-based approaches might be used to achieve the measures of interest for GAO and suggests potential data sources from NIOSH and NASS that could be used as auxiliary data.

References

Bickart, B. A. (1990). *Cognitive Aspects of Proxy Reporting of Behavior*. Retrieved from ACR North American Advances: http://acrwebsite.org/volumes/7021/volumes/v17/NA-17

Ecob, Russel, Sally Macintyre, and Patrick West. (1993). Reporting by parents of longstanding illness in their adolescent children. *Social Science & Medicine 36, no. 8*, 1017-1022.

Erhard, Michael, Ute Ellert, Barbel-Maria Kurth, and Ulrike Ravens-Sieberer. (2009). Measuring adolescents' HRQol via self reports and parent proxy reports: an eveluation of the psychometric properties of both versions of the KINDL-R instrument. *Health and Quality of Life Outcomes 7, no. 1*, 77.

Garbarski, Dana. (2014). Comparing self and maternal reports of adolescents' general heakth status: Do self and proxy reports differ in their relationships with covariates? *Quality of Life Research 22, no. 7*, 1953-1965.

Gerling, M.W., Tran, H.N, and O'Connor, T.P. (2010). *The road to understanding nonresponse in the June Area Survey.* Washington, DC: National Agricultural Statistics Service.

Kalton, G. (2009, December). Methods for oversampling rare subpopulations in social surveys. *Survey Methodology, 35*(2), 125-141.

King, T., Sarah Cook, and Jennifer Hunter Childs. (2012). *Interviewing proxy versus self-reporting respondents to obtain information regarding living situations*. Retrieved May 23, 2016, from In Proceedings for the Joint Statistical Meetings, Survey Research Methods Section: https://www. amstat. org/sections/srms/proceedings/y2012/files/400243_500698. pdf

Kish, L. (1994, August). Multipopulation Survey Designs: Five Types with Seven Shared Aspects. *International Statistical Review, 62*(2), 167-186.

Krosnick, Jon A., Stanley Presser, Kaye Husbands Fealing, Steven Ruggles, and David L. Vannette. (2015). *The future of survey research: Challenges and opportunities*. Retrieved from The National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences Subcommittee on advancing SBE Survey Research: http://www. nsf. gov/sbe/AC_Materials/The_Future_of_Survey_Research. pdf

Lee, Sunghee, Nancy A. Mathiowetz, and Roger Tourangeau. (2004). Perceptions of disability: the effect of self and proxy response. *Journal of Official Statistics 20, no. 4*, 671.

Li, W., Kelsey, J. L., Zhang, Z., Lemon, S. C., Mezgebu, S., Boddie-Willis, C., & Reed, G. W. (2009, MArch). Small-Area Estimation and Prioritizing Communities for Obesity Control in Massachusetts. *Am J Public Health, 99*(3), 511-519. doi:10.2105/AJPH.2008.137364

Marker, D. A. (2001, December). Producing Small Area Estimates From National Surveys: Methods for Minimizing use of Indirect Estimators. *Survey Methodology, 27*(2), 183-188.

Miller, D. and O'Connor, T.P. (2012). *Item response rates for the Agricultural Resource Managment Survey III in 2006 and 2007.* Washington, DC: National Agricultural Statistics Service.

Mingay, David J., Steven K. Shevell, Norman M. Gradburn, and Carl Ramirez. (1994). Self and proxy reports of everyday events. In *In Autobiographical memory and the validity of retrospective reports, pp* (pp. 235-250). New York, NY: Springer.

Mitchell, M. Ott, K., Ridolfo, H., and McCarthy, J. (2015). *Targeted data collection efforts for the 2013 ARMS III.* Washington, DC: National Agricultural Statistics Service.

Mohadjer, L., & Curtin, L. R. (2008, June). Balancing sample design goals for the National Health and Nutrition Examination Survey. *Survey Methodology, 34*(1), 119-126.

National Academies of Sciences, Engineering, and Medicine. (2017). *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps.* Washington, D.C.: The National Academies Press. doi:10.17226/24893

Pierannunzi, C., Xu, F., Wallace, R. C., Garvin, W., Greenlund, K. J., Bartoli, W., . . . Town, M. (2016, July 14). A Methodological Approach to Small Area Estimation for the Behavioral Risk Factor Surveillance System. *Preventing Chronic Disease, 13*, 150480. doi:10.58888/pcd13.150480

Rao, J., & Molina, I. (2015). *Small Area Estimation* (2nd ed.). Hoboken: John Wiley & Sons, Inc.

Reynolds, Jeremy, and Jeffery B. Wenger. (2012). He said, she said: the gender wage gap according to self and proxy reports in the current population survey. *Social science research 41, no. 2*, 392-411.

Ridofo, H., Boone, J., and Dickey, N. (2013). *Will they answer the phone if they know it's us? Using caller id to improve response rates.* Washington, DC: National Agricultural Statistics Service.

Sudman, S., Bradburn, N.M., & Schwartz, N. (1996). *Thinking about answers: The application of cognitive process to survey methodology.* San Francisco: Jossey-Bass.

Todorov, Alexander, and Corinne Kircher. (2000). Bias in proxies' reports of disability: data from the National Health Interview Survey on disability. *American Journal of Public Health 90, no. 8*, 1248.

Tourangeau, R. (1984). Cognitive sciences and survey methods. *Cognitive aspects of survey methodology: Building a bridge between disciplines*, 73-100.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin*, 133(5), 859.

Tzavidis, N., Zhang, L.-C., & Luna, A. (2018). From start to finish: a framework for the production of small area official statistics. *J. R. Statist. Soc. A, 181*(Part 4), 927-979.

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

Valliant, R., Hubbard, F., Lee, S., & Chang, C. (2014, June 1). Efficient Use of Commercial Lists in U.S. Household Sampling. *Journal of Survey Statistics and Methodology, 2*(2), 182-209. doi:10.1093/jssam/smu006

Warner, M., Schenker, N., Heinen, M. A., & Fingerhut, L.A. (2005). The effects of recall on reporting injury and poisoning episodes in the National Health Interview Survey. *Injury Precention*, 11, 282-287.

Williams, D., & Brick, J. M. (2018, June 1). Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort. *Journal of Survey Statistics and Methodology, 6*(2), 186-211. doi:10.1093/jssam/smx019