Designing the 2007 Commodity Flow Survey

Scot Dahl, William Davie, Jr,
US Census Bureau[1]

Abstract

The Commodity Flow Survey produces data on the movement of goods in the US. This paper describes the activities conducted to improve the quality of the frame and the 2007 sample design. Auxiliaries are establishments of large companies that provide support to the rest of the company. In the past many sampled auxiliaries turned out to be non-shippers and our measures of shipping activity were poor. In 2006, we mailed a one-page survey form to all auxiliaries on the Census Bureau's Business Register. Responses to this survey were used to classify establishments as shippers or non-shippers and to determine a measure of size. For the sample design, we identified industries with large amounts of HAZMAT shipments to better stratify the sample to capture HAZMAT activity and, for auxiliaries, we used the data collected earlier. The design was optimized subject to geographic constraints.

Keywords: Commodity Flow, CFS, Sample design, Frame construction

1. Introduction

The 2007 Commodity Flow Survey (CFS), a component of the Census Bureau's Economic Census programs, is undertaken every five years through a partnership between the Bureau of the Census, U.S. Department of Commerce, and the Bureau of Transportation Statistics (BTS), U.S. Department of Transportation. The survey provides key information about the movement of goods in the United States and provides a crucial source of transportation statistics not available elsewhere, including value of shipments, weight of shipments, commodities shipped, mode(s) of transportation used, origin and destination of shipments, ton-miles, and average miles per shipment. The data from the CFS are used by public policy analysts and for transportation planning and decision-making to assess the demand for transportation facilities and services, energy use, and safety risk and environmental concerns. Previous Commodity Flow Surveys were conducted in 2002, 1997, and 1993.

The CFS covers business establishments in mining, manufacturing, wholesale, and selected retail and service industries. The survey also covers auxiliary establishments (e.g., warehouses) of multi-establishment companies. The survey coverage excludes establishments classified as farms, forestry, fisheries, governments, construction, transportation, foreign establishments, and most establishments in services and retail.

The CFS captures data on shipments originating from sampled business establishments located in the 50 states and the District of Columbia. Each sample establishment will be contacted four times throughout 2007 - once every quarter. We request that the respondent provide the following information about a sample of their establishment's shipments: domestic destination or port of exit, commodity, value, weight, mode(s) of transportation, the date on which the shipment was made, and an indication of whether the shipment was an export, or a hazardous material. For exports, we also ask the respondent to provide the mode of export and the foreign destination city and country. For hazardous materials shipments, we ask for the corresponding United Nations/North American Hazardous Materials (UN/NA) code.

2. Changes for 2007

Changes were made in several areas to improve the 2007 CFS. Specifically, changes were made in:
- Scope and coverage
- Frame development
- Questionnaire development
- Sample design
- Disclosure prevention

The remainder of this paper describes these changes.

3. Scope and Coverage

In 2005, a combined Census Bureau and BTS team began considering changes to the industry and geographic coverage for the 2007 CFS. As a result of these discussions, several decisions were made

---

[1] This report is released to inform interested parties of research and to encourage discussion. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

regarding the industry coverage of the 2007 CFS. These decisions were:

- Consideration was given to improving the coverage for Logs & Wood in the Rough (*Standard Classification of Transported Goods* (SCTG) category 25). However, most of the establishments involved in logging are classified as Agricultural industries and the Census Bureau no longer maintains data on establishments in those industries. Therefore it was decided to make no special effort to improve coverage of this commodity.
- In order to improve the coverage of petroleum products (SCTG 18 - Fuel Oils), retail fuel oil dealers were included in the 2007 CFS sample frame. This added about 11,000 establishments to the frame.
- In order to improve the coverage of newsprint, paper, and other printed products (SCTGs 27, 28, and 29), all publishers (North American Industry Classification System (NAICS) codes 5111 and 51223) were included in the 2007 frame. This is consistent with how publishers were handled in the 1997 CFS when they were classified as manufacturers. Publishers were not included in the 2002 CFS.
- An Advance survey was conducted to improve the coverage of auxiliary establishments with shipment activity and the quality of their data (described in detail below).

4. Frame Development

4.1 Auxiliaries
The major frame improvement for 2007 was improving the quality of the frame data associated with auxiliaries. Auxiliaries are establishments of multi-unit enterprises that provide support for the other establishments in the company. Examples are: headquarters offices, regional offices, data processing facilities, training facilities, warehouses, and distributions centers. Traditionally, the CFS has included auxiliaries classified in the following industries:

- Warehousing & Storage -NAICS 4931
- Corporate, Subsidiary, and Regional Managing Offices - NAICS 551114 (these have been included since they may perform a warehousing/shipping function as well as their primary management function.)

In past surveys, many of the auxiliaries selected into the CFS sample, especially those in NAICS 551114, turned out to be non-shippers. Further, for those that did have shipping activity, our initial measure of the establishment's size (MOS - a proxy for value of shipments - used to place these establishments in the

appropriate size strata and assign a sample weight) was frequently substantially different from the value of shipments derived from their CFS responses.

This was the case because these establishments do not report sales or receipts in the Economic Census since they provide support for the rest of the company and do not directly produce revenue for a company. Further, if data are reported for these establishments, it may represent more than just the shipping activity at the establishment, such as receipts for the entire company or it may include the payroll of company executives at a headquarters.

4.2 Advance Survey
In order to identify those auxiliaries with shipping activity, improve the quality of the frame data for auxiliaries that ship, and also to obtain improved contact information for the largest establishments in the in-scope industries, in 2006 the Census Bureau mailed a one page questionnaire to approximately 85,000 establishments (see Table 1). This included all 40,000 in-scope auxiliaries in NAICS 4931 and 551114 and 45,000 of the largest establishments in the other in-scope industries

Including telephone as well as mailed back responses, data were received for 87% of the auxiliaries and 75% of the large establishments.

Shipper status (whether or not an auxiliary establishment performed any shipping operations) was determined from the response to a four part question (see Figure 1). The respondent had to check "No" to all four parts of this question to be classified as a non-shipper.

Though these questions were not asked of non-auxiliary (large) establishments, comments on the returned questionnaire (such as those indicating that the establishment was out of business) may have been used to classify a large establishment as a non-shipper. Table 2 shows the final shipper status of the responding establishments. Those establishments that were classified as non-shippers were excluded from the sample frame. Non-respondents were considered to be shippers by default.

In addition, for those auxiliaries that were shippers, we asked for an estimate of their annual value of shipments using a checkbox type question with eight possible responses (see Figure 2).

Table 3 shows the number of the 10,716 shipping auxiliaries that checked each range box. Those auxiliaries that checked one of the boxes were

randomly assigned (from a uniform distribution) a MOS within the range of the box checked (using an upper bound of $10 billion for the last range). Non-responding auxiliaries (and those that failed to report an annual value of shipments) were assigned a MOS based on their 2002 Economic Census responses or other administrative data contained on the Census Bureau's business register as was done in the 2002 CFS.

5. Questionnaire Development

The Census Bureau and BTS conducted a year long process to improve the CFS questionnaire. This involved three rounds of cognitive interviews with employees of selected, cooperating establishments throughout the country. The major changes to the questionnaire for 2007 were:
- An improved layout and a more intuitive question flow.
- Greater standardization of font, indentation, answer blocks, question phrasing, etc.
- Improved sampling instructions. (For the CFS, the respondent is required to select a sample of shipments within an assigned sample week.)
- More consistent references between the questionnaire, the instructions, and the commodity code manual.
- A checkbox to identify shipments that are "intermodal" (transportation of freight in a container or vehicle, using multiple modes of transportation such as rail, ocean vessel, and truck, without any handling of the freight itself when changing modes)
- A series of questions regarding the use of Third Party Logistics (3PL) services that will be included on the fourth quarter 2007 questionnaire mailed to each establishment.

In addition respondents will be able to respond electronically by entering their data into an Excel spreadsheet (using a template provided by the Census Bureau that matches the shipment matrix in the questionnaire). This spreadsheet can then be sent to the Census Bureau.

6. Sample Design

As was done in past surveys, the 2007 survey was selected using a stratified three-stage design in which the first-stage sampling units were establishments, the second-stage sampling units were four reporting periods (1-week) spaced thirteen weeks apart, and the third-stage sampling units were individual shipments sampled by the respondent.

The primary objective of the 2007 CFS is to estimate shipping volumes (e.g., value, tons, and ton-miles) by geographic area, mode, and commodity. A secondary objective is to estimate the volume of shipments moving from one geographic area to another (e.g., state-to-state, metropolitan area-to-metropolitan area, etc.) by mode and commodity.

For the sample design process, we used an estimate of annual total value of shipments as the measure of size for each establishment. The geographic stratification was based on a combination of the 50 States and the District of Columbia and selected metropolitan areas (generally selected based on population in the 2000 Decennial Census or their transportation importance). As a proxy for commodity when designing the sample, we used combinations of 3- and 4-digit NAICS codes assigned to each establishment.

6.1 General Design Features
The design incorporated the following features:
- An average of 40-50 establishments per primary (first-stage) stratum was desired so that the Central Limit Theorem would apply to the distribution of our estimates and inferences from these estimates would be more robust.
- The sample size would be increased to 100,000 establishments from the 50,000 used in the 2002 survey.
- Primary strata were defined by geography and industry. In addition to the 50 States and the District of Columbia, some metropolitan areas were used to define the geographic strata. For 2007, greater geographic detail would be provided by adding ten additional metropolitan areas (see Table 4). Most of these are major sea or land transportation gateways.

6.2 Sample Frame
For 2007, the CFS sample frame consisted of approximately 750,000 establishments selected based on the following criteria:
- In-scope North American Industry Classification System industry,
- Active on the Census Bureau's Business Register with non-zero 2005 payroll, or a responding shipper from the Advance survey.

Industry classification was based on the 2002 NAICS. In addition to most mining, manufacturing, wholesale, and electronic shopping and mail order establishments, the 2007 frame included the additional industries discussed earlier:
- retail fuel dealers (NAICS 45431);

- newspaper, periodical, book, and directory publishers (NAICS 5111);
- music publishers (NAICS 51223),
- shipping auxiliaries (e.g. warehouses and management offices – NAICS 4931 and 551114) based on the Advance survey.

For each establishment we extracted census sales, payroll, and number of employees, name and address, a 6-digit NAICS code, and a primary identifier. Using census sales, payroll, and some additional data collected in the 2002 Economic Census, as well as data collected from auxiliary establishments during the Advance survey operation, a measure of size was computed for each establishment. The measure of size was designed to approximate an establishment's annual total value of shipments.

6.3 Geographic Strata
We defined 123 geographic strata based on the 50 states, the District of Columbia, and selected Core Based Statistical Areas (CBSA) and Combined Statistical Areas (CSA). These consisted of 73 metropolitan areas and 50 other strata containing the balance of each state. Metropolitan areas with significant representation in more than one state (such as the Washington-Baltimore CSA) were split into individual state parts.

6.4 Industry Strata
We created up to 45 industry strata within each of the geographic strata:
- 4-digit NAICS for Mining (3 strata)
- 4-digit NAICS for Wholesale (18 strata)
- 4-digit NAICS for Nonstore Retail (1 stratum)
- 4-digit NAICS for Retail Fuel Dealers (1 stratum)
- 3-digit NAICS for Manufacturing (21 strata)
- 5111 – Publishers, including music publishers (1 stratum)

If a three or four digit NAICS industry contributed at least 4% of the total value (based on sampling measure of size) or tonnage (based on 2002 CFS data[2]) for the geographic stratum or the nation, it was designated a "do not collapse" industry stratum within the geographic stratum. Industries not meeting this level of activity within a geographic stratum were grouped with other similar industries, as discussed below. Auxiliary

---

[2] Prior to computing the percentages based on tonnage for each industry grouping and geographic stratum, we adjusted the proportions estimated from the 2002 CFS sample so that the proportions of the total value estimated from the 2002 sample were equal to the proportions computed using the measure of size from the 2007 CFS sampling frame.

establishments were excluded from these computations, because we decided to create separate strata for auxiliary establishments in each of the 123 geographic strata (see section 6.6 below for details).

This procedure produced 1,306 "do not collapse" geography/industry strata. The remaining industry strata were collapsed to at most 10 "collapsed" industry strata within each geographic stratum as shown in Table 5.

The method used to collapse the remaining strata used 2002 CFS data as input to a Classification and Regression Tree (CART) procedure that related industries with commodities. The terminal nodes from the CART procedure were then grouped using a hierarchical clustering algorithm. Using the results from the hierarchical clustering algorithm, we manually regrouped some of the clusters to arrive at the final industry clusters shown in Table 5. This process resulted in 1,154 collapsed geography/industry strata.

6.5 Special HAZMAT Strata
To produce better estimates of the shipment of hazardous materials for 2007, we created 160 strata targeting HAZMAT shippers. Using 2002 CFS data, we identified those 6-digit NAICS industries that accounted for a large proportion of the estimated total value and/or total tonnage for six groups of hazardous materials:
(1) ammonium nitrate,
(2) ethanol,
(3) explosives,
(4) hydrogen,
(5) toxic by inhalation,
(6) all other hazardous materials.

The HAZMAT strata chosen for the 2007 CFS are shown in Table 6. A large proportion of estimated shipment value or tonnage for the listed hazardous materials group came from the NAICS industry indicated in column 2. The column labeled "Sample?" indicates whether or not we selected all establishments in the industry with certainty or if we selected a sample of them. The column labeled "Geo Strata?" indicates the level of stratification (national or state) used. When deciding whether or not a particular industry would be a "take-all" stratum and the level of geographic stratification, we considered the proportion of estimated HAZMAT value and/or tons accounted for by the particular NAICS industry and the number of establishments within the industry at the state and national levels.

6.6 Auxiliary Strata

The treatment of auxiliary establishments was modified for 2007 to take advantage of the data collected in the Advance survey. For auxiliaries that responded to the Advance survey and were considered to be shippers, we created 123 strata, one in each geographic stratum, combining both NAICS 4931 and 551114.

We also created two national strata for auxiliary establishments that did not respond to the Advance survey – one stratum for non-responding warehouses (those classified in NAICS 4931) and one stratum for non-responding management offices (NAICS 551114). Based on the results of the 2002 CFS, auxiliaries classified in NAICS 4931 are likely to actually be shippers, whereas about half of sampled auxiliaries in NAICS 551114 reported no shipping activity in 2002. Although we had concerns that the establishments that did not respond to the Advance survey were likely to be either non-shippers or unreachable, we wanted to ensure the non-responding establishments were represented in the sample.

6.7 Summary of Stratification

We started with a total of about 5,500 primary strata defined by geography and industry. After completing the steps outlined above, each establishment on the sampling frame was assigned to one of 2,745 primary strata. A summary of the final stratification is shown in Table 7. Of the 2,745 strata, 232 were designated as "take-all" strata because of the small number of establishments in the stratum and/or their importance.

6.8 Design Constraints

Sample sizes were computed to meet coefficient of variation (CV) constraints on estimated value of shipments totals for each primary stratum. We used a desired CV of 1.5% on the estimated total value of shipments for each primary stratum. We chose this CV because it produced total sample sizes of approximately 100,000 establishments.

The primary constraints are budget related, which are translated into an approximate fixed sample size for the survey. What we try to accomplish with our design is to allocate this fixed total sample size in a statistically efficient manner. The CV constraints are primarily used as a tool to allocate more of the sample to more important strata. We assume the cost of data collection does not vary by stratum.

We also imposed maximum sampling weight and minimum sample size constraints. For the CFS designs, we set the maximum first stage sample weight to 100

and the minimum sample size to 2 establishments per stratum.

6.9 Procedure for Determining Sampling Parameters and Sample Sizes

The procedure for determining sampling parameters is an iterative computerized process. The sample design programs used in the process are part of a group of generalized programs that have been modified to accommodate the needs of the survey, but use common methods such as the Dalenius & Hodges cumulative sqrt(f) procedure, Neyman allocation, and similar rules for determining acceptable designs.

For each (non-take all) primary sampling stratum, the survey designer specified as input to a Generalized Univariate Stratification (GUS) program:
- desired number of bins (for a frequency distribution used in the Dalenius & Hodges' cumulative sqrt(f) procedure)
- desired number of size strata
- desired number of certainty companies
- desired coefficient of variation for total value of shipments
- maximum sampling weight
- minimum sample size

The GUS program first identifies the desired number of certainty establishments within each primary stratum. This set of establishments comprises the take-all stratum. The noncertainty sampling frame consists of all establishments that are not in the take-all stratum.

The GUS program uses the Dalenius & Hodges procedure to determine noncertainty stratum boundaries. Sample sizes are computed using Neyman allocation (i.e., proportional to the number and variance of units in each size stratum). The designer tries many different possible combinations of the bulleted items shown above. Each combination will usually give a different sample size and allocation.

The "best" design for each primary stratum is determined using the following guidelines in order of importance:
1. the total sample size is within 20% of the minimum sample size of all designs considered for the primary stratum
2. certainty companies comprise 20-40% of the total sample size
3. none of the noncertainty strata are "Neyman" take-all strata

4. the sampling weights are monotonic decreasing as the average sales/receipts size of units in each stratum increases
5. one-third or fewer of the noncertainty strata have a Neyman sample size less than the minimum stratum size
6. the number of units added to meet the maximum sampling weight constraint is less than 10% of the total sample size for the primary stratum without the additional units

When choosing between 2 or more designs, we ranked the designs according to the 6 guidelines shown above. From the top-ranked designs, we chose the one with the smallest sample size and, if the sample sizes were equal, we chose the one with the smallest percent of certainty companies. In some cases, we did not find a design that met all of these guidelines. In those instances, we selected the design that met the majority of these guidelines.

Once designs were determined for each of the primary strata, the information from these designs was used as input to a program that attempts to more efficiently allocate the sample to meet the desired CV on each primary stratum ("P/S CV") and also determine the sample sizes needed to meet a national level constraint ("Natl CV"). Tables 8A and 8B summarize the designs considered. Designs with a national level constraint tend to allocate more sample to the larger states so there is a trade off between better national estimates and the quality of the more detailed geographic estimates.

In view of the BTS preference for improving the national level estimates for the 2007 CFS, the third option (primary strata CV of 1.7% and a national CV of 0.036%) was chosen.

6.10 Sample Selection
Within each size stratum, a simple random sample is selected without replacement. Each selected establishment is assigned a sampling weight equal to the reciprocal of its probability of selection into the sample.

7. Disclosure Prevention

One of the most detailed tables produced from the CFS is a four-way table (commodity by mode by origin by destination) that provides estimates of total value, tons, and ton-miles of shipments and the average miles traveled by shipments. A table with this level of detail that was published from the 2002 CFS contained nearly 400,000 cells (Massell and Russell, 2007). Cell suppression has been used in prior CFS surveys to provide disclosure protection. However, data users demand for more detailed statistics coupled with strict laws governing the protection of respondent data continues to drive research in the area of disclosure protection for tabular data. Additionally, cell suppression becomes much more difficult to implement for tables of 3 or more dimensions. To meet these demands, the Census Bureau conducted research on the application of EZS Noise (Evans, et al., 1998) to the 2002 CFS data (Massell and Funk, 2007). The EZS Noise method adds "noise" to the microdata used to construct the estimates rather than to the estimates themselves. Results of this research are promising and we intend to apply the EZS noise method for disclosure protection of 2007 CFS data. For a more detailed description of the disclosure protection research and its initial results, see Massell and Russell (2007).

8. Summary

While no major changes were made to the design or operation of the Commodity Flow Survey, several incremental improvements were made. These improvements were:
• Expanded the industry coverage of the sample frame,
• Use of an Advance survey to obtain better frame data for auxiliary establishments,
• More detailed geographic and industry (especially HAZMAT) stratification,
• A larger sample size,
• A national level constraint on the sample design,
• A redesigned questionnaire.
• Use of noise to improve the detail of published tables while at the same time preventing disclosure of individual shipment and company-level data.

References

Barnett, La Toya, Rebecca Morrison, and Grace O'Neill (2005), "Findings and Recommendations from the 2005 CFS Cognitive Testing, Round 1 of 3," unpublished report, Washington, DC: US Census Bureau

Barnett, La Toya, Rebecca Morrison, and Grace O'Neill (2006), "Findings and Recommendations from the 2005 CFS Cognitive Testing, Round 2 of 3," unpublished report, Washington, DC: US Census Bureau

Barnett, La Toya, Rebecca Morrison, and Grace O'Neill (2006), "Findings and Recommendations from the 2005 CFS Cognitive Testing, Round 3 of 3," unpublished report, Washington, DC: US Census

Bureau

Dahl, Scot (2006), "Recommendations of the Scope and Coverage Joint Investigative Team (JIT) for the 2007 CFS," unpublished report, Washington, DC: US Census Bureau

Davie, William (2006) "Overview of the Sample Design for the 2007 CFS," unpublished report, Washington, DC: US Census Bureau

Evans, Timothy, Laura Zayatz, John Slanta (1998), "Using Noise for Disclosure Limitation of Establishment Tabular Data", Journal of Official Statistics

Fowler, John (2006), "Executive Summary of the Scope and Coverage JIT Findings," unpublished report, Washington, DC: US Census Bureau

Massell, Paul and Funk, Jeremy (2007), "Advances in Disclosure Protection: Releasing More Business and Farm Data to the Public," paper presented at the Third International Conference on Establishment Surveys, Montreal, Quebec, Canada.

Massell, Paul and Russell, J. Neil (2007), "Protecting the Confidentiality of Commodity Flow Survey Tables by Adding Noise to the Underlying Microdata," paper presented at the 2007 Joint Statistical Meetings, Salt Lake City, Utah.

Office of Management & Budget, *North American Industrial Classification System: United States 2002* (NAICS Manual), Washington, DC

Sweet, Elizabeth and Richard Sigman (1995), "User Guide for the Generalized SAS® Univariate Stratification Program," ESM Report Series ESM-9504, Washington, DC: US Census Bureau

US Census Bureau (2006), "Commodity Flow Survey: Program Overview," Feb 10, 2007, www.census.gov/sssd/cfsdat/cfsoverview.htm

US Census Bureau (2006), "2007 Commodity Flow Survey Standard Classification of Transported Goods (SCTG)," Washington, DC: US Census Bureau, www.census.gov/svsd/www/cfsdat/cfs071200.pdf

Figure 1: Excerpt from Advance Survey Questionnaire (Shipping Question)

Figure 2: Excerpt from Advance Survey Questionnaire (Value of Shipments)

Table 1: Advance Survey		
Trade Area		Sample
Mining		2,258
Manufacturing		20,704
Wholesale		18,189
Retail		2,814
Service		1,010
Auxiliaries:	4931	4,845
	551114	35,152
Total		84,972

Table 2: Shipper Status of Advance Survey Responses			
Shipper Status	Aux	Large	Total
Shipper	10,716	32,275	42,991
Non-Shipper	24,220	1,561	25,781
Total	34,936	33,836	68,772

Table 3: Shipping Auxiliaries

Annual value of Shipments ($mil)	Number of auxiliaries
Not reported	1,130
Less than 5	4,451
5 - 20	1,652
20 - 50	1,020
50 - 200	1,225
200 - 500	566
500 - 1,000	322
1,000 - 5,000	304
More than 5,000	46
Total	10,716

Table 4: Metropolitan Area Additions for 2007 CFS

CSA	CBSA	CSA / CBSA Description
132		Baton Rouge-Pierre Part, LA CSA
	13140	Beaumont-Port Arthur, TX MSA
	16700	Charleston-North Charleston, SC MSA
204		Corpus Christi-Kingsville, TX CSA
	21340	El Paso, TX MSA
278		Hartford-West Hartford-Willimantic, CT CSA
324		Lake Charles-Jennings, LA CSA
	29700	Laredo, TX MSA
380		Mobile-Daphne-Fairhope, AL CSA
496		Savannah-Hinesville-Fort Stewart, GA CSA

Table 5: Collapsed Industry Strata

Industry Cluster	Industries in Cluster
Textiles and Furniture	313, 314, 315, 316, 337, 4232, 4243
Food and Beverage	311, 312, 4244, 4245, 4248
Machinery and Motor Vehicles	333, 336, 4231, 4238
Chemicals and Petroleum	324, 325, 326, 4242, 4246, 4247, 4543
Paper	322, 323, 4241, 5111
Computers and Electronics	334, 335, 4234, 4236
Metals	2122, 331, 332, 4235, 4237
Nonmetals	2121, 2123, 327
Wood and Lumber	321, 4233
Miscellaneous	339, 4239, 4249, 4541

Table 6: HAXMAT Strata

HAZMAT Group	NAICS	Sample?	Geo Strata?
Ammonium Nitrate	325311	Take all	n/a
	424690	Sample	State
	424910	Sample	State
Ethanol	311221	Take all	n/a
	325193	Take all	n/a
	424210	Sample	State
Explosives	325920	Take all	n/a
	332992	Take all	n/a
	336414	Take all	n/a
	336415	Take all	n/a
	325998	Sample	National
Hydrogen	325120	Sample	
	325188	Sample	National
	325199	Sample	
Toxic by Inhalation	325181	Take all	n/a
Other HAZMAT	324110	Take all	n/a
	424710	Sample	State

Table 7: Summary of Primary Stratification

Overall Primary Strata	Number
Do not collapse	1,306
Collapsed	1,154
Auxiliaries (Advance Survey responders)	123
Auxiliaries (Advance Survey non-responders)	2
Hazmat	160
Total Primary Strata	2,745
Take All Strata	Number
Do not collapse	155
Collapsed	51
Auxiliaries (Advance Survey responders)	10
Auxiliaries (Advance Survey non-responders)	0
Hazmat	16
Total Take-All Strata	232

Table 8A: Total Sample Size of Alternative CFS Designs						
		Total Sample Size				
		P/S CV=1.5%	P/S CV=1.6%		P/S CV=1.7%	
	Universe	No Natl CV	Natl CV=0.04%		Natl CV=0.036%	
Trade Area	Count	Count	Count	% chg	Count	% chg
Auxiliary	14,878	2,981	4,376	46.8	4,713	58.1
Manufacturing	327,826	44,835	45,367	1.2	45,285	1.0
Mining	6,789	3,003	2,926	(2.6)	2,846	(5.2)
Other	22,539	2,035	2,051	0.8	2,033	(0.1)
Retail	25,190	2,784	2,688	(3.4)	2,562	(8.0)
Wholesale	356,477	44,629	44,740	0.2	44,930	0.7
TOTAL	753,699	100,267	102,148	1.9	102,369	2.1

Table 8B: Total MOS of Alternative CFS Designs						
		MOS Estimate from Sample				
		P/S CV=1.5%	P/S CV=1.6%		P/S CV=1.7%	
	Universe	No Natl CV	Natl CV=0.04%		Natl CV=0.036%	
Trade Area	MOS ($mil)	MOS ($mil)	%diff	MOS ($mil)	%diff	MOS ($mil)	%diff
Auxiliary	2,177,288	2,180,204	0.1	2,177,501	0.0	2,173,949	(0.2)
Manufacturing	4,317,543	4,319,946	0.1	4,315,700	(0.0)	4,312,062	(0.1)
Mining	53,893	54,348	0.8	54,025	0.2	53,872	(0.0)
Other	181,541	179,729	(1.0)	179,584	(1.1)	179,580	(1.1)
Retail	157,904	157,958	0.0	157,817	(0.1)	158,804	0.6
Wholesale	3,799,490	3,794,935	(0.1)	3,804,258	0.1	3,807,708	0.2
TOTAL	10,687,659	10,687,120	(0.0)	10,688,884	0.0	10,685,975	(0.0)