

March 20, 2003

Mr. Rick Kowalewski
Acting Director
Bureau of Transportation Statistics
400 7th Street, S.W.
Washington, D.C. 20590

Dear Mr. Kowalewski:

We are pleased to transmit this third letter report of the Committee to Review the Bureau of Transportation Statistics' (BTS) Survey Programs. This committee was convened by the Transportation Research Board and the Committee on National Statistics in response to a request from BTS. The membership of the committee is shown in Enclosure A. The committee has been charged with reviewing the current BTS survey programs in light of transportation data needs for policy planning and research and in light of the characteristics and functions of an effective statistical agency. This letter presents the committee's consensus findings and recommendations concerning the Commodity Flow Survey (CFS).

The committee held its third meeting on October 31–November 1, 2002, at the National Academies facilities in Washington, D.C. The purpose of this meeting was review of the CFS by the committee. To this end, the committee heard presentations from representatives of the CFS partnership, namely, BTS and the Census Bureau; from a representative of the Oak Ridge National Laboratory (ORNL), which plays a key role in analyzing the survey results; and from a range of public and private sector users of CFS data. A list of the presentations at the meeting is provided in Enclosure B. Following the data-gathering sessions, the committee met in closed session to deliberate on its findings and recommendations and begin preparation of this report, which was completed through correspondence among members. In developing these findings and recommendations, the committee drew on the information gathered at its third meeting, articles in the technical literature,¹ examples of CFS products, and the experience and expertise of individual members. The committee would like to thank all those who contributed to this review through their participation in the third committee meeting and their responses to follow-up questions.

In summary, the committee found that the CFS plays a unique role in providing data on domestic freight movements to inform a wide range of economic and policy analyses and related investment decisions. However, gaps in shipment and industry coverage and a lack of geographic and commodity detail limit the usefulness of the CFS data for a growing number of applications. Moreover, limitations due to the lack of detail are being compounded by the shrinking sample size. A national freight data architecture could eventually result in a more comprehensive national picture of freight flows. In the meantime, the committee recommends that the CFS be continued at least until an improved alternative has been established. In an effort to make future editions of the CFS more useful and more cost-effective in providing data for a range of users, BTS and the Census Bureau should (a) investigate opportunities to update the survey

¹ A list of all nonproprietary materials considered by the committee is available from the Public Records Office of the National Academies (e-mail: publicac@nas.edu).

methodology, with emphasis on the use of new technologies to support increases in sample size through more cost-effective data collection; (b) establish improved mechanisms for soliciting suggestions and feedback from users to inform design decisions and prioritize survey modifications; and (c) re-evaluate their roles and responsibilities within the CFS partnership to ensure adequate and timely funding for future surveys.

The remainder of this report commences with a brief overview of the CFS. The committee's findings on (a) data use and data users and (b) the design of the 2002 CFS are then presented. The report concludes with the committee's recommendations to BTS and the Census Bureau for approaches to providing a variety of users with improved freight data. These recommendations address three main areas: the future of the CFS, meeting user needs, and the CFS partnership.

OVERVIEW OF THE CFS

The CFS, which is undertaken through a partnership between BTS of the U.S. Department of Transportation and the Census Bureau of the U.S. Department of Commerce, aims to provide information on the flow of goods by mode of transport within the United States.² All methods of freight transportation (air, motor carrier, rail, water, and pipeline) and intermodal combinations are covered. The survey was conducted in 1993 and 1997, and data collection for the 2002 CFS was ongoing at the time of the committee's third meeting.³ The budget for the 5-year cycle of the 2002 CFS is \$13.03 million, of which 80 percent is provided by BTS and 20 percent by the Census Bureau.

The CFS captures data on shipments originating from manufacturing, mining, wholesale, and selected retail establishments located in the 50 states of the United States and the District of Columbia. The sampling frame is drawn from the Census Bureau's Business Register of 6 million employer establishments, of which approximately 750,000 are in industries covered by the CFS. The 2002 CFS is collecting data from 50,000 establishments. The sample sizes for the 1993 and 1997 editions of the CFS were 200,000 and 100,000 establishments, respectively. As in 1993 and 1997, the 2002 survey is being conducted entirely by mail.⁴

Because the CFS is administered by the Census Bureau as part of the 5-yearly Economic Census, survey response is mandatory under Title 13 of the U.S. Code. The response rate for the 1997 CFS was 75 percent, and as of October 2002, that for the 2002 CFS was estimated at approximately 70 percent.⁵ Respondents are required to report their total numbers of outbound shipments as well as information on value, weight, commodity, domestic destination or port of exit, and mode(s) of transport for a sample of these shipments. For the 2002 CFS, each establishment was assigned a 1-week reporting period every quarter, for a total of 4 weeks in the calendar year. By assigning different reporting periods to different establishments, the

² For a more detailed discussion of the CFS, the reader is referred to the websites of BTS (www.bts.gov) and the Census Bureau (www.census.gov) and to the final report on the 1997 CFS (Census Bureau 1999).

³ The CFS restores a data program on commodity flows that the Census Bureau conducted from 1963 through 1977 as part of its 5-year economic census program. The Census Bureau conducted a smaller commodity transportation survey in 1983 but did not release the results because of problems with data reliability.

⁴ Guidance on completing the questionnaire is available on the Census Bureau website or through a toll-free telephone number.

⁵ As reported to the committee by Census Bureau staff, October 31, 2002.

sample covers all 52 weeks of the year. It is anticipated that the 2002 CFS will gather information on a total of 2.7 million shipments.⁶

The Census Bureau makes a range of CFS data products available to the public in printed reports, on CD-ROM, and on the Internet. In accordance with federal law governing Census Bureau reports, no data are published that would disclose information about the operations of an individual firm or establishment. Thus, data at the level of individual establishments, known as microdata, are kept confidential, although researchers may on occasion be permitted very limited access to these data at the Census Bureau's Research Data Centers.

The published CFS data at the national level tabulate information on shipment characteristics by mode of transport (including intermodal combinations) and by commodity. Data are provided on tons, miles, ton-miles,⁷ value, shipment distance, commodity, and weight. Additional reports provide geographical breakdowns for flows between census divisions and regions, individual states, and major metropolitan areas. Reports on movements of hazardous materials and on exports are also published.⁸

The final report on the 1997 CFS was published in December 1999, 2 years after the completion of data collection (Census Bureau 1999). A similar schedule is anticipated for the 2002 CFS. The final report is expected at the end of 2004, with preliminary national data available at the end of 2003.⁹

FINDINGS

Data Use and Data Users

Finding 1: Analysts and researchers in both the public and private sectors use data from the Commodity Flow Survey (CFS)—often in conjunction with data from other sources—for a variety of purposes, including:

- *Analyzing trends in goods movement over time,*
- *Conducting national, regional, and sectoral economic analyses,*
- *Developing models and other analytical products to inform policy analyses and management and investment decisions,*
- *Forecasting future demand for goods movement and associated infrastructure and equipment needs,*
- *Cross-checking data from other sources and establishing benchmarks for estimating national accounts, and*
- *Analyzing and mapping spatial patterns of commodity and vehicle flows.*

⁶ For the 1993 CFS, each establishment was assigned a 2-week reporting period every quarter and information was gathered on a total of 10.3 million shipments. For the 1997 CFS, each establishment was assigned a 1-week reporting period every quarter and information was gathered on a total of 5.3 million shipments.

⁷ The Center for Transportation Analysis at ORNL computes shipment mileages from the CFS data by means of an intermodal transportation network modeling system. These mileages are used in preparing the values of ton-miles provided in the CFS reports.

⁸ Further information on CFS products is available on the Census Bureau's CFS website (www.census.gov/econ/www/cfsmain.html).

⁹ As reported to the committee by Census Bureau staff, October 31, 2002.

CFS data are widely used by federal government agencies, including those outside the U.S. Department of Transportation; by academic researchers; and by consulting companies, whose clients include a range of businesses, state departments of transportation, federal government agencies, and associations. In their presentations to the committee (see Enclosure B), several users made the distinction between “power” users, who employ CFS data in their own analyses and models, and “regular” users, who include CFS-based facts in briefing papers and reports but do not undertake extensive calculations with CFS data.

Most power users make use of all the information provided by the CFS at all levels of geographic detail, with emphasis on states and metropolitan areas. This emphasis reflects the growing interest of states and metropolitan planning organizations (MPOs) in freight issues. As a result of this interest, freight transportation data are needed at a finer level of geographic detail than in the past to inform policy and investment decisions relating to economic development and environmental goals. CFS data are aggregated at the level of states and Bureau of Economic Analysis regions to maintain statistical validity and protect the confidentiality of data providers. However, these aggregate data are of limited use for most state and metropolitan planners and engineers, who need to assign commodity and vehicle flows to corridors—and if possible to major highways and rail lines.

Examples of investigations using CFS data that were reported to the committee are (a) research on the geographic organization of production and trade in the United States; (b) a study of the economic impacts of highway construction in California; (c) benchmarking of the input–output accounts developed by the Bureau of Economic Analysis that show how industries provide input to, and use output from, each other to generate Gross Domestic Product; (d) forecasting of motor carrier equipment requirements on the basis of information about length of haul and commodity carried; and (e) development of the Federal Highway Administration’s Freight Analysis Framework, which made extensive use of CFS data to build a comprehensive picture of national freight flows for policy analysis purposes.

Another example of the use of CFS data is of particular interest in the present context. Reebie Associates uses the CFS data, together with data from public and proprietary sources, to develop its Transearch database.¹⁰ Like the CFS, the Transearch database aims to provide a reasonably comprehensive picture of the flow of goods by mode of transport in the United States. Data are available for purchase at many different levels of modal, geographic, and commodity detail. Many power users reported that they use the Transearch database extensively, often because it provides greater geographic detail than the CFS and because it is updated annually. The Transearch database is generated with proprietary methods, and information about data reliability is not reported. In contrast, the data reliability and sources of error for the 1997 CFS are discussed in the final report on the survey (Census Bureau 1999). Data that fail to meet certain reliability criteria are excluded from publication by the Census Bureau, resulting in gaps in the picture of national freight flows. With a larger sample size, these gaps would be fewer in number.

Many users reported to the committee that they frequently use CFS data in conjunction with data from other sources. For example, a recent study of the potential for economic integration between Canadian and U.S. regions (Brown and Anderson 2002) combined data from the CFS and from Statistics Canada’s Trade Information and Retrieval System to obtain interregional

¹⁰ Before 1993, when the CFS was first conducted, data from the 1977 Commodity Transportation Survey were used to develop the Transearch database.

trade flows. Such combining of data is often problematic, because differing data collection strategies and data definitions raise concerns about data quality and comparability.

On the basis of the committee's discussions with CFS users, it would appear that no single source of freight data is ever likely to meet all the needs of all data users. Participants in the 2001 Saratoga Springs meeting, *Data Needs in the Changing World of Logistics and Freight Transportation*, concluded that a national freight data architecture is needed to "streamline future data collection efforts and facilitate compatibility of various data sources at different levels of aggregation" (Meyburg and Mbwana 2002, p. 23). These issues are under consideration by the Committee on Freight Transportation Data—A Framework for Development, which expects to issue its report by mid 2003. In the absence of a national freight data architecture, the CFS is widely used—despite its deficiencies—because it goes some way toward meeting user requirements for data that provide a comprehensive picture of national freight flows.

Finding 2: Data from the CFS—a periodic 5-yearly survey of domestic shipper establishments—are of limited use for a number of applications because of

- *Gaps in shipment and industry coverage,*
- *Insufficient geographic and commodity detail at the state and local levels,*
and
- *The inability to capture rapid changes in economic cycles.*

Although the CFS attempts to provide reasonably complete data on the movement of goods in the United States, there are some notable gaps in both shipment and industry coverage. Some of these gaps have become increasingly significant in recent years because of (a) changes in the national economy, including greater emphasis on international trade and freight logistics, and (b) the need for an improved understanding of freight movements, particularly at state and local levels, to inform many policy, planning, and investment decisions.

Comprehensive information on international shipments is increasingly needed because of the growing importance of international trade to the U.S. economy. Because the CFS samples domestic shipper establishments, it cannot capture information on shipments from foreign establishments. Imported products are included in the CFS at the point that they leave the importer's domestic location (which is not necessarily the port of entry) for shipment to another location in the United States. Thus the first leg of import shipments is excluded. Export shipments are included in the CFS, with the domestic destination defined as the port of exit from the United States.

Shipper surveys have traditionally focused on firms in the mining, manufacturing, and wholesale sectors of the economy, on the assumption that such surveys capture information on the majority of goods transported by freight carriers. With the advent of freight logistics and a focus on finding the most efficient way to source, manufacture, and distribute products, trans-shipments between warehouses, distribution centers, and transportation terminals have grown in importance. The CFS covers selected auxiliary establishments, such as warehouses, but excludes transportation and service establishments and most retail establishments. These and other gaps in the CFS industry coverage—for example, in agricultural shipments from the farm to the first point of assembly—have become increasingly important as analysts and transportation planners try to develop a better understanding of freight movements to inform a range of policy and investment decisions.

There is widespread agreement, particularly among the power users, that increased geographic and commodity detail at the state and local levels would greatly enhance the usefulness of the CFS. The availability of such data depends on two major factors: (1) the survey sample size and (2) the statutory obligation to maintain the confidentiality of individual establishments. As the sample size decreases, the statistical variability of the data increases. If the sample size is too small, the data may not be sufficiently reliable to be useful for analysis at the required level of geographic detail. Although the 1997 CFS, with a sample size of 100,000 establishments, collected potentially useful local-level data, these microdata cannot be made available to the public because their release could compromise the confidentiality of data providers. The publicly available CFS data are aggregated to avoid any possibility of disclosing information about individual establishments. Even if the sample size were increased to provide more reliable data at finer levels of geographic detail, large well-known companies could still be relatively easy to identify. Importantly, the reduction in sample size to 50,000 establishments in 2002 further restricts the availability of disaggregate data.

Some users expressed a need for data on transportation costs and service characteristics, which would be especially useful for tracking service quality and modeling mode choice. However, many shippers surveyed in the CFS are unlikely to be able to supply reliable reports of transportation service characteristics. Thus, meeting this need may require a survey of carriers, which is currently beyond the scope of the CFS.¹¹

The CFS is unable to capture rapid changes in economic cycles because of the 5-year interval between data collection cycles. The lack of coverage of the intervening 4 years means that time trends in freight activity, such as the effects of emerging from a period of recession or of a severe drought, cannot be studied satisfactorily using the CFS data alone. Several CFS users noted in their presentations to the committee that the 2-year time lag between the completion of data collection and release of the final data, combined with the 5-year interval between surveys, results in CFS data whose timeliness is “less than ideal.” Nevertheless, users generally indicated that they would be willing to sacrifice improved timeliness for greater richness in the data.

Design of the 2002 CFS

Finding 3: The design of the 2002 CFS appears to have been compromised in important ways by the lack of a clear understanding between BTS and the Census Bureau about ownership of the CFS; the responsibility for ensuring sufficient funding to produce a useful, quality product; and the respective roles of the two agencies in developing survey methods, which combined led to the following effects:

- *Because of uncertainty about the availability and level of funding, key design decisions were delayed until late in the survey planning process, which hampered advance preparation and problem solving.*
- *The reduction in survey sample size to 50,000 establishments has adversely affected the anticipated usefulness of the 2002 CFS data for many applications.*
- *The technical rationale for the survey design was not documented in a sufficiently clear and timely fashion for data users to (a) understand the trade-*

¹¹ While carrier surveys are useful in capturing data about shipments, shipper-based surveys are needed to obtain some important items of data, such as mode-independent flows.

offs involved and the resulting implications for data quality and (b) provide input to the design process to help ensure the usefulness of the resulting data.

On the basis of experience gained from the 1993 CFS, some changes were made to the design and questionnaire for the 1997 CFS (Black et al. 2000).¹² In addition, an automated editing system was introduced that enabled data collection staff and survey analysts to identify and correct problematic reports quickly. Census Bureau staff reported to the committee that they had hoped to implement further improvements in 2002 on the basis of experience with the 1997 CFS. It was not clear to the committee to what extent BTS staff were involved in identifying and prioritizing such improvements. However, because of uncertainties about the availability and level of funding for the 2002 CFS until very late in the design process, opportunities for improvement and innovation were severely limited.

The lead time for developing a new CFS questionnaire is on the order of 3 years, and guidance on data collection priorities is needed 2 years before the survey is fielded for changes to be implemented.¹³ Pilot studies of new data collection methodologies also require time for their planning and cannot be implemented with only a few months' notice. The 2002 CFS design had to be finalized at short notice because of funding uncertainties. Therefore the design makes only very limited use of statistical information from earlier editions of the survey to improve sampling strategies and other features, and does not incorporate any pilot studies of innovative techniques, such as web-based data collection, that offer the potential to improve data quality and response rates.

Because of its effect on overall cost and data reliability, the choice of sample size is one of the most important design decisions for any survey. The budgets and sample sizes in numbers of establishments for the 1993, 1997, and 2002 editions of the CFS are

CFS Survey Year	Budget (\$ million)	Sample Size
1993	15.0	200,000
1997	19.0	100,000 ¹⁴
2002	13.0	50,000

For the 2002 CFS, the survey budget appears to have been the dominant factor determining the sample size. The estimated cost for a sample size of 100,000 was \$17.7 million, whereas the estimate for a sample size of 50,000 was \$13.0 million.^{15,16} Thus, a 36 percent increase in cost

¹² The design changes focused on reducing (a) respondent burden, (b) the influence of large and infrequent shipments, and (c) the time between completing data collection and releasing the survey results to the public. Changes to the questionnaire aimed at facilitating the shipment sampling task for respondents.

¹³ As reported to the committee by Census Bureau staff, November 1, 2002.

¹⁴ Black et al. (2000) report that the reduction in the CFS sample size from 200,000 in 1993 to 100,000 in 1997 was to allow for intensive follow-up of problem reports early in the survey and thereby improve data quality and accuracy. With the larger sample, the delay in processing the data decreased the effectiveness of such follow-up efforts. The automated editing system introduced in 1997 also aimed at improving data quality through more rapid identification of problem reports.

¹⁵ As reported to the committee by Census Bureau staff, November 1, 2002.

¹⁶ Census Bureau staff reported to the committee that they were also asked by BTS to produce cost estimates for sample sizes of 10,000 and 30,000. The Census Bureau indicated to BTS that it would not participate in the survey if the sample sizes were reduced to these levels, because the reliability of the

would have resulted in a 100 percent increase in sample size. Furthermore, the cost per establishment would have dropped from \$260 to \$177—less than the 1997 cost per establishment of \$190.

In 1997, each of 100,000 establishments was sampled four times over the course of the year for a total of 400,000 reports. For the 2002 CFS, one of the major design decisions was whether to obtain 200,000 reports over the course of the year by (a) sampling each of 50,000 establishments four times or (b) sampling each of 100,000 establishments twice. BTS and the Census Bureau jointly decided that the first option was preferable because it gave lower estimated coefficients of variation for freight flows. Nonetheless, reducing the sample size from 100,000 to 50,000 establishments degraded the quality of the publishable data.

A number of users of CFS data shared with the committee their concern that reductions in sample size are adversely affecting the data's usefulness. For example, such reductions limit the ability of the Bureau of Economic Analysis to estimate an interstate trade index and develop regional multipliers to measure the effect of changes in demand on industries and local economies. Similarly, the 1993 CFS data were used to estimate the ton-miles of trucking activity over the nation's highways (TranStats 1997). There are concerns that the fourfold reduction in sample size to 50,000 for the 2002 CFS will result in much greater uncertainty in such estimates of infrastructure use, particularly for through-state shipments.¹⁷

The fairly extensive investigations of possible sampling schemes for the 2002 CFS conducted by the Census Bureau do not appear to have been shared with users of CFS data or made available to the public in any form. The decision to reduce the sample size from 100,000 to 50,000 establishments was taken by the Census Bureau and BTS with apparently little consideration of the advantages and disadvantages of the different sampling options from a user perspective. Thus, many users are not aware that a relatively small increase in funding for the survey (\$4.7 million over 5 years) could have offered very real benefits for data users by maintaining the sample size at the 1997 level of 100,000 establishments. In failing to share this information, the CFS partnership deprived itself of opportunities to enlist the support of users in seeking additional funding for the 2002 survey.

RECOMMENDATIONS

Scope

The committee recognizes that the CFS is only one source of data on freight movements and that not all the current deficiencies of freight data can be remedied by changes to the CFS. The following recommendations are intended to assist BTS and the Census Bureau in making the CFS a more useful data source for a range of users. These recommendations also provide a basis for developing successor survey(s) to the CFS, since they focus on meeting the need for cost-effective surveys that generate quality data and are responsive to user requirements.

associated data would be unacceptably low and the survey would not provide general purpose statistics on commodity flows.

¹⁷ As reported to the committee by Frank Southworth, ORNL, November 1, 2002.

The Future of the CFS

Recommendation 1. In view of the widespread use of CFS data for a diversity of applications, BTS should continue to provide data on the flow of goods by mode of transport within the United States. These data should be updated at intervals of no more than 5 years. To ensure that ongoing user needs are met, the CFS should be continued—with some modifications—at least until such time as a viable alternative source of national freight data has been established.

Although the CFS has been criticized, primarily because of gaps in data and a lack of geographic and commodity detail, a large—and growing—user market in the United States requires information on freight movements to inform economic and policy analyses and related investment decisions. The CFS currently plays a unique role in providing such data. In the committee's view, therefore, the CFS should be continued at least until an improved alternative is implemented to ensure the continuing availability of data on domestic freight flows.

Future versions of the CFS would benefit greatly from modifications that update the methodology and make the survey more responsive to the needs of data users. A re-evaluation of the roles and responsibilities of the CFS partners—BTS and the Census Bureau—would also benefit the conduct of future surveys. The following recommendations address these items in more detail.

Recommendation 2. BTS and the Census Bureau should proceed with planning for the 2007 CFS. This effort should explore opportunities for conducting pilot studies of new methods in parallel with established designs. These new methods should be selected on the basis of their potential to reduce survey costs through more cost-effective data collection techniques and sampling strategies; reduce respondent burden; improve data quality; and provide more useful data for a range of users. Every effort should be made to investigate opportunities for achieving economies to permit much-needed increases in sample size.

In view of the long lead time necessary to implement changes in a major survey such as the CFS, the committee urges BTS and the Census Bureau to initiate work on the design of the 2007 CFS without delay. These early design initiatives should include investigations of the potential of new technologies for improving data quality and reducing both the respondent burden and the costs of data collection. The 2007 CFS provides an excellent opportunity to conduct pilot studies of web-based surveys and the like and to compare the results with those obtained from more conventional approaches.

In view of the widespread user concerns about the implications of the reduced sample size for the 2002 CFS, the committee urges BTS and the Census Bureau to make every effort to increase the sample size for the 2007 CFS. The proposed uses of the data drive both the sample size and the sampling scheme. Therefore, it is essential for the CFS partnership to work closely with users in developing a survey design that will meet user needs.

Changes that allow the CFS to be conducted more cost effectively offer the potential to support increases in sample size. The committee was pleased to learn that the Census Bureau has tentative plans to provide the option of a web-based questionnaire for the 2007 CFS. This approach would build on the bureau's experience of electronic reporting for the 2002 Economic

Census and other surveys. In addition to lowering the reporting costs incurred by many large establishments, such electronic reporting would reduce costs of data entry for the Census Bureau. The committee urges the CFS partnership to pursue the resulting opportunities to increase sample size as a matter of high priority.

Recommendation 3. The CFS partnership should initiate a research program to investigate survey methods for the CFS. To help stimulate creativity and innovation, organizations outside the federal government, including universities and small businesses, should be encouraged to participate in this research program. Topics to be investigated should include data collection, sample design, survey nonresponse, statistical estimation, and data processing.

Aside from some limited changes in the design and questionnaire for the 1997 CFS (Black et al. 2000), the CFS methodology has remained largely unchanged since the survey was initiated in 1993. In the committee's view, neither BTS nor the Census Bureau has taken a sufficiently active role in investigating opportunities to improve the overall quality of the survey and use available funds more effectively. The committee urges the CFS partnership to invest in research into possible improvements in CFS methods to stimulate creative thinking about new approaches to the survey, particularly in the areas of data collection, sample design, survey nonresponse, statistical estimation, and data processing.

Data Collection

A number of users cited as a deficiency of the survey its inability to capture rapid changes in freight activity trends. The CFS also needs to provide effective coverage of evolving shipment patterns, such as those associated with the growth of e-commerce. Research aimed at developing a better understanding of the rates of change of freight flows and trip characteristics (e.g., mode of shipment by shipment size and distance for selected commodities) over time could be helpful in informing decisions about how frequently to collect various types of data.

The CFS partnership should also investigate the possibility of eventually moving to a system of continuous data collection in which data are collected every month (or year), drawing new sample establishments monthly (or annually). Such continuous data collection affords more timely data than a periodic 5-yearly survey and could also provide greater geographic detail by accumulating data over longer time periods. An additional advantage, particularly from the perspective of the Census Bureau, is that the heavy workload associated with CFS data collection and processing would be broken into smaller tasks over a longer time period and would no longer be concurrent with a similarly burdensome period for the 5-yearly Economic Census.

The CFS partnership should investigate options for using mixed-mode data collection methods to reach different establishments in different ways. Thus, establishments equipped to provide data electronically—through electronic data interchange systems, for example—could provide CFS data by e-mail, diskette, web data entry, or other electronic media. Such electronic filing of survey data may be far more convenient for firms that have the necessary equipment and expertise, but would not preclude the use of mail-in questionnaires for other survey participants. Other firms could choose to enter data using a telephone key pad data entry system analogous to the Touchtone Data Entry used by the Bureau of Labor Statistics for its Current Employment Statistics Survey (see, for example, Rosen et al. 1999).

Sample Design

Investigations of ways to extend the industry and shipment coverage should be conducted in consultation with data users (see Recommendation 5). Although a survey of shipper establishments cannot fill all the current gaps identified by users, there may be opportunities to provide greater coverage in some areas. Possibilities for facilitating the linkage of CFS data with other data sources are also worthy of investigation.

Sampling existing establishments more effectively could improve data quality and reduce costs. Possible areas for investigation include the following:

- Alternative within-firm sample designs. These may offer opportunities to reduce reporting errors. For example, an approach that involves randomly selecting a starting point in terms of shipments and taking the next n records may be less susceptible to error for some firms than the current systematic sampling method (Black 1997).¹⁸
- Stratification of shipments by size. This may reduce the variability of estimates and thereby provide more accurate estimates of flows. For example, all large shipments could be included over a longer period (1 month, 1 year) than the current 1-week reference period.

Nonresponse

Data provided by the CFS partnership do not enable the committee to obtain insights into CFS nonresponse such as the reasons for nonresponse, unit response rates by type of firm or other grouping (e.g., stratum), or item nonresponse rates. It was not apparent to the committee that the CFS partnership has conducted any detailed analyses of the 25 percent of establishments that failed to respond to the 1997 CFS and of the likely impact of this level of nonresponse on the final data set. Such analyses would offer valuable insights into bias in the survey results and could also help focus research efforts on specific problems.

In response to the large number of complaints when the survey was first conducted, efforts have been made to reduce the burden on CFS respondents. Nonetheless, research into the level of burden that firms are willing to tolerate may offer useful guidance for reducing survey nonresponse. Such respondent burden research could be investigated through focus groups or survey research among firms.

Statistical Estimation

The Census Bureau is already doing some research into the effect of large shipments or large firms on the precision of survey estimates (Black et al. 2000). These efforts should be pursued, because large firms that make either large numbers of shipments or high-value shipments can contribute disproportionately to estimates and increase the associated variances. Research on the distribution of shipment sizes and values, and on their impact on precision, may be useful in developing new, more efficient sample designs.

Data Processing

The committee encourages the Census Bureau to continue its work on developing data editing systems for the CFS. If combined with a web-based questionnaire, such systems could be

¹⁸ The current sampling method involves using a lookup table to translate the total number of shipments into a “take-every” number that, properly applied, results in a sample of shipments.

useful in identifying potential problems and providing feedback to assist respondents while they are in the process of completing the questionnaire. Such automated edits could be used to check that critical items (e.g., total number of shipments) are reported or that the respondent reports the expected number of individual shipments as determined from the total number of shipments. Experience in developing electronic data collection forms and automated editing systems for the Economic Census (Murphy et al. 2001) should yield important benefits for the CFS.

Meeting User Needs

Recommendation 4. BTS should establish a process to facilitate dialogue between private- and public-sector CFS users and technical professionals at the Census Bureau and BTS. This dialogue would assist both agencies in developing an in-depth understanding of the diversity of uses of CFS data and associated limitations. Such an understanding would

- ***Assist BTS in identifying the role of the CFS, or its successor(s), in the broader context of efforts to develop a national freight data architecture; and***
- ***Assist BTS and the Census Bureau in making future surveys more responsive to user needs by targeting particular content and problem areas and by prioritizing improvements.***

BTS solicits comments on the CFS from data users through a monthly customer feedback survey and has also sought comments from the modal administrations in DOT on their use of CFS data. In November 2000, the agency convened a meeting with both public- and private-sector CFS users and Census Bureau representatives to initiate a dialogue on freight data needs for planning and policy purposes.¹⁹ However, the committee is not aware of any efforts to continue this dialogue on a regular basis, other than through informal discussions between users and individual staff members at BTS and ORNL.

The committee observed that many users of CFS data are extremely interested in the future of the survey and have valuable suggestions to make about a range of possible improvements. The establishment of an appropriate process for dialogue would provide a forum for regular user discussions with BTS and the Census Bureau on freight data needs in general and the CFS in particular. Through these discussions, the CFS partnership would benefit from the knowledge and expertise of informed users and would be better positioned to respond to evolving data needs. The CFS users could also provide valuable support to the partnership in its efforts to obtain the stable funding needed for effective survey planning and development.

Recommendation 5. In developing future versions of the CFS, or its successor survey(s), BTS and the Census Bureau should

- ***Solicit user input to the design process through dialogue with CFS users and other outreach mechanisms; and***
- ***Ensure that the rationale for major design decisions—notably those affecting sample size—is documented in such a way as to provide openness in decision making.***

¹⁹ As reported to the committee by Felix Ammah-Tagoe, BTS, November 1, 2002.

A key decision in designing future editions of the CFS is determining the level of geographic detail the data will provide. Given that the purpose of the survey is to provide useful data for a variety of applications, the committee urges the CFS partnership to obtain substantive input from a range of public- and private-sector users about the levels of geographic detail and associated data reliability that they require to support their proposed uses of the CFS data. This input from users should be used to establish target levels of geographic detail to guide subsequent decisions about design and, notably, sample size. While the availability of funding will inevitably influence survey design, the committee believes that a sound technical rationale is needed to inform design decisions, ensure that the best use is made of available resources, and provide a basis for seeking sufficient resources for future surveys.

On the basis of its discussions with users, the committee anticipates that future editions of the CFS will need to provide usable data at least at the state level. The extent to which the CFS can realistically provide useful data at a finer level of geographic detail requires further investigation. Clearly a single national survey cannot meet all the needs of all users. For example, the CFS may not be able to provide measurements of vehicle flows in metropolitan areas in support of investigations of options for relieving bottlenecks. In the longer term, a national freight data architecture that facilitates the integration of data from various sources probably offers the most promise for many local needs. For example, data from a national database, such as the CFS, could be combined with metropolitan area data to inform policy, planning, and investment decisions at the MPO level.

BTS receives requests from users for additional data items not currently available from the CFS. A number of users also commented to the committee that they would like to see the CFS provide additional information on shipments (e.g., transportation costs) as well as greater industry coverage (e.g., agricultural movements from farm sites to processing centers or terminal elevators and freight movements by service industries, such as lawn-care companies). It is not clear that the CFS, or any survey of shipper establishments, can provide all the additional data users would like. Nevertheless, the committee urges the CFS partnership to solicit user suggestions for additional data items and to use these suggestions in targeting improvements to future editions of the CFS. In some instances, relatively minor modifications to the CFS may facilitate the linkage of CFS data with that from other sources (e.g., quality of service data) to meet user needs. Such linkages could greatly increase the value of the CFS for many users.

The committee was unable to obtain any formal documentation on the 2002 CFS apart from the survey questionnaire and other materials provided to respondents. This situation is in marked contrast to both the National Household Travel Survey and the Omnibus Survey, for which publicly available reports describe the survey methodology and supporting rationale. Although delays in finalizing the design of the 2002 CFS likely contributed to difficulties in documenting the process, the lack of openness in decision making for a major national survey is of serious concern to the committee. Improved documentation of critical design issues for future editions of the CFS would provide greater opportunities for users to participate in and influence survey development and thereby enhance the usefulness of the final data set.

Recommendation 6. The CFS partnership should investigate options for improved delivery of CFS data to users. In particular, BTS should work with the Census Bureau to investigate technical and administrative options for increasing access to the CFS microdata, while continuing to maintain the confidentiality of data providers.

The Census Bureau, in common with other federal statistical agencies, uses both technical and administrative procedures to protect the confidentiality of data providers. Technical methods for statistical disclosure limitation have been reported in the literature for more than 20 years, and research in this area is continuing in an effort to find better ways of accommodating the needs of statistical agencies, data providers, and data users.²⁰ Administrative approaches involve restricted access procedures, as implemented through the Census Bureau's Research Data Centers, for example.

Restrictions on the release of CFS microdata are a major source of frustration for many power users seeking to perform detailed analyses of freight activity. A National Research Council report on confidentiality and accessibility of government statistics recommended that federal statistical agencies should "strive for a greater return on public investment in statistical programs through [...] expanded availability of federal data sets to external users" (National Research Council 1993, p. 224). The same report also advocated a policy of "responsible innovation" in expanding access for external data users and recommended "experiment[ing] with some of the newer restricted access techniques, with appropriate confidentiality safeguards and periodic reviews of the costs and benefits of each procedure" (p. 224). The committee urges the CFS partnership to follow this advice and examine the extent to which disclosure limitation methods can mask the identity of individual establishments in the CFS microdata.

The committee also encourages BTS to work with the Census Bureau to facilitate user access to the CFS microdata through the Census Bureau's Research Data Centers. The aforementioned National Research Council report recommended that statistical agencies should "make access conditions more affordable and acceptable to users" in instances for which restricted access procedures are needed (National Research Council 1993, p. 225). BTS needs to take an active approach in encouraging the Census Bureau to implement this recommendation.

Although the final report on the 1997 CFS discusses data reliability and sources of error (Census Bureau 1999), the committee believes that power users could benefit from additional documentation about the quality of CFS data in the form of an error profile. More detailed reporting about the frequency of data imputation would also be helpful for users seeking an in-depth understanding of data reliability. Census Bureau staff reported to the committee that they use various imputation procedures to compensate for partial nonresponse—for example, if an establishment fails to provide data for one of the four reporting periods—but the level of imputation does not appear to be reported. Flagging all microdata values that are imputed rather than reported may be appropriate in the event that at least some of these data can one day be released.

Several users commented to the committee that the form in which the 1997 CFS data were made available to the public lacked versatility, particularly for analysts seeking to use the data as input to their own models and calculations. For example, many of the data are provided in summary tables rather than in a database format that provides access to basic origin–destination flow patterns. The Census Bureau has indicated that it will try to improve the CD-ROM for the 2002 CFS for power users. The committee urges BTS to play a role in investigating alternative formats for the public data file in an effort to provide power users with the versatility they require, while continuing to provide regular users with the information they

²⁰ For further information, the reader is referred to *Information about Statistical Disclosure Methods*, American Statistical Association, Committee on Privacy and Confidentiality, <http://users.erols.com/dewolf/protect/sdlinfo.htm>.

need in a user-friendly format. The CFS partnership should also consider investigating alternative delivery mechanisms, such as a web server, for making CFS data available to a broader set of users.

The change in commodity coding system between the 1993 and 1997 editions of the CFS and the lack of alignment with coding systems used for trade and production data were highlighted by some users as a source of difficulty in using CFS data. However, the issue of standardization among commodity coding systems has implications well beyond the CFS, and the committee decided against addressing such a complex issue in this report.

The CFS Partnership

Recommendation 7. BTS and the Census Bureau should re-evaluate their roles and responsibilities within the CFS partnership to build on the expertise and experience of both parties, including in the following ways:

- ***As an element of the CFS partnership, BTS and the Census Bureau should work together to obtain the necessary funding for future versions of the CFS.***
- ***In view of the linkage between the CFS and the Economic Census, the Census Bureau should assume an appropriate share of the responsibility for survey innovation.***
- ***BTS should focus on developing priorities to guide the evolution of the CFS within the broad context of a national freight data architecture.***

The CFS partners need to work together as a team to avoid repetition of the 2002 CFS scenario in which delays in committing funds eliminated most opportunities for survey improvement and innovation and almost resulted in the cancellation of the survey itself. Both partners have a role to play in obtaining commitments for funding the 2007 CFS at a level appropriate to providing useful data at the chosen level of geographical detail and in a time sufficient to permit the preliminary investigations needed to inform decisions about survey design and methodology.

Although BTS provides much of the funding for the CFS (80 percent), the Census Bureau has a statutory requirement to conduct the Economic Census,²¹ to which the CFS is linked, and also maintains the register of employer establishments from which the CFS sample is drawn. Therefore, if the CFS is to be continued and improved, the Census Bureau needs to consider itself as a partner in the CFS program rather than a contractor. Given the Census Bureau's role as the data collection agency for the CFS, the committee considers it appropriate that the Census Bureau, drawing on relevant experience that it has gained in other Census Bureau surveys, take a major role in proposing and investigating new survey designs and data collection methods.

As an agency in the U.S. Department of Transportation, BTS is well positioned to develop an understanding of freight data needs in general and the extent to which future editions of the CFS, or its successor survey(s), can contribute to the development of a comprehensive national picture of freight flows. As a federal statistical agency, BTS clearly has a role to play in researching new methodologies for the CFS and in investigating issues of data analysis and

²¹ Title 13 of the U.S. Code directs the Census Bureau to take the economic census every 5 years, covering years ending in 2 and 7.

data delivery. However, the committee suggests that BTS' major role within the CFS partnership should be to establish priorities to guide future development of the survey. Developing these priorities will require BTS to do the following:

- Engage in an active dialogue with data users (see Recommendation 4), and
- Establish alliances with other data providers in the federal and state governments and the private sector to coordinate data collection efforts in the context of a national freight data architecture.

CLOSING REMARKS

The committee appreciates this opportunity to review and comment on the CFS and looks forward to preparing its final report on cross-cutting issues relating to the three BTS surveys it has reviewed—the National Household Travel Survey, the Omnibus Survey, and the CFS.

Sincerely yours,

Joseph L. Schofer
Chair
Committee to Review the Bureau of Transportation Statistics' Survey Programs

cc: Tom Zabelsky, Census Bureau

Enclosure A: Committee membership
Enclosure B: Presentations at the third committee meeting

References

- Black, J. 1997. Can Respondents Construct a Frame and Draw a Sample? Experiences from the 1993 Commodity Flow Survey. American Statistical Association, Proceedings of the Survey Methods Section, 216–221. Alexandria, VA: American Statistical Association.
- Black, J. R., W. C. Davie, Jr., and J. R. Jonas. 2000. An Evaluation of Sample Design Changes for the 1997 Commodity Flow Survey. American Statistical Association, Proceedings of the Survey Research Methods Section, 244–249. Alexandria, VA: American Statistical Association.
- Brown, W. M., and W. P. Anderson. 2002. Spatial markets and the potential for economic integration between Canadian and U.S. regions. *Papers in Regional Science*, 81, 99–120.
- Census Bureau. 1999. 1997 Economic Census, Transportation: 1997 Commodity Flow Survey. Report EC97TCF-US. U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau, Washington DC.
- Meyburg, A. H., and J. R. Mbwana. 2002. Conference Synthesis: Data Needs in the Changing World of Logistics and Freight Transportation, Saratoga Springs, New York, November 14–15, 2001. Albany, NY: New York State Department of Transportation.

- Murphy, E. D., E. M. Nichols, A. E. Anderson, M. D. Harley, and K. D. Presley. 2001. Building Usability into Electronic Data-Collection Forms for Economic Censuses and Surveys. U.S. Census Bureau, Washington DC. (www.fcsm.gov/01papers/Murphy.htm)
- National Research Council. 1993. Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics. National Academy Press, Washington DC.
- Rosen, R. J., C. D. Manning, L. J. Harrell, Jr., and D. A. Skuta. 1999. Data Collection Issues Related to Implementing the Redesigned Current Employment Statistics Survey. American Statistical Association, Proceedings of the Survey Research Methods Section, 169–174. Alexandria, VA: American Statistical Association.
- TranStats. 1997. Truck Movements in America: Shipments From, To, Within, and Through States, BTS/97-TS/1. U.S. Department of Transportation, Bureau of Transportation Statistics, Washington DC, May.

ENCLOSURE A

COMMITTEE TO REVIEW THE BUREAU OF TRANSPORTATION STATISTICS' SURVEY PROGRAMS*

JOSEPH L. SCHOFER, *Chair*, Northwestern University, Evanston, Illinois
THOMAS B. DEEN (National Academy of Engineering), Consultant, Stevensville, Maryland
WILLIAM F. EDDY, Carnegie Mellon University, Pittsburgh, Pennsylvania
T. KEITH LAWTON, Metro, Portland, Oregon
JAMES M. LEPKOWSKI, University of Michigan, Ann Arbor
ARNIM H. MEYBURG, Cornell University, Ithaca, New York
DEBBIE A. NIEMEIER, University of California, Davis
ALAN E. PISARSKI, Consultant, Falls Church, Virginia
STANLEY PRESSER, University of Maryland, College Park
G. SCOTT RUTHERFORD, University of Washington, Seattle
EDWARD J. SPAR, Council of Professional Associations on Federal Statistics, Alexandria,
Virginia
RONALD W. TWEEDIE, Consultant, Delmar, New York

*The committee was composed and reviewed according to National Academies procedures and was judged to be free of potential conflicts of interest.

ENCLOSURE B

PRESENTATIONS AT THE THIRD MEETING OF THE COMMITTEE TO REVIEW THE BUREAU OF TRANSPORTATION STATISTICS' SURVEY PROGRAMS, OCTOBER 31 – NOVEMBER 1, 2002, WASHINGTON, D.C.

Objectives of the Commodity Flow Survey

Michael Cohen, *Bureau of Transportation Statistics*

The Commodity Flow Survey—Key Issues

John Fowler, *Census Bureau*

Frank Southworth, *Oak Ridge National Laboratory/Bureau of Transportation Statistics*

Felix Ammah-Tagoe, *Bureau of Transportation Statistics*

User Perspectives on the Commodity Flow Survey

Martin Labbe, *Martin Labbe Associates, Ormond Beach, Florida*

Paul Ciannavei, *Reebie Associates, Stamford, Connecticut*

Bill Anderson, *Boston University, Boston, Massachusetts*

Paul Bingham, *Global Insight, Inc., Washington, D.C.*

Bruce Lambert, *Federal Highway Administration, Washington, D.C.*

Rick Donnelly, *PB Consult, Albuquerque, New Mexico*

Sue Okubo, *Bureau of Economic Analysis, Washington, D.C.*

Russ Hillberry, *U.S. International Trade Commission, Washington, D.C.*

Joel Palley, *Federal Railroad Administration, Washington, D.C.*

Agnes Muszynska, *Cambridge Systematics, Washington, D.C.*

Ed Weiner, *Office of the Secretary of Transportation, Washington, D.C.*