

**Cognitive Interview Evaluation of the
Federal Statistical System Public Opinion Monitoring Survey
Round 2, Focus on Administrative Record Linkage Questions:
Results of interviews conducted in March, 2012**

Stephanie Willson¹
Paul Scanlon^{1,2}

Introduction

This report documents results from a second round of cognitive interviewing of the Federal Statistical System Public Opinion Monitoring Survey (FSS POMS). This survey represents an effort to understand public opinions and knowledge of the FSS, specifically, trust in the FSS, the credibility of federal statistics, and attitudes toward and knowledge of the statistical uses of survey linkage to administrative records. To this end, a working group was formed to design and implement a cross-agency survey of public attitudes about federal statistics and statistical agencies. The working group consisted of members from Office of Management and Budget (OMB), the Census Bureau, National Agricultural Statistics Services (NASS), the Internal Revenue Service (IRS), and the National Center for Health Statistics (NCHS). The first round of testing focused on general attitudes about and trust in the FSS and federal statistics. The second round of testing, the topic of this report, focused on questions about administrative record linkage. The goal of the questions is to measure respondents' attitudes toward the federal government linking to various administrative records such as medical records, credit card records, and records from the Social Security Administration. Not all questions were analyzed for this report. Section one includes findings for questions 15, 16, 25, and R1 – R6, which were asked separately (and slightly differently) for Census, BLS and NCHS. The second section of the report covers question 8 – 11. These questions presented respondents with different messages regarding the general concepts of administrative data use and data sharing and were not intended to be fielded in the actual survey. Rather, they were included as a qualitative exploration to better understand respondents' thoughts on privacy and confidentiality pertaining to federal statistics and data sharing.

The next section briefly describes the qualitative methodology of cognitive interviewing, including the procedure for sampling respondents, the data collection method, and analysis plan.

Methodology

Sampling and Respondent Demographics

Staff from the Census Bureau, NCHS, and IRS participated in cognitive interviewing. A total of 43 interviews were completed. Respondents were selected using a purposive sample. The goal of a purposive sample is not to obtain a statistically representative sample. Instead, emphasis is on coverage of the survey questions and topics, not the survey population. The goal of the sample is to capture as many different patterns of interpretation as possible. Respondents were

¹ Questionnaire Design Research Lab, National Center for Health Statistics

² American Community Survey Office, United States Census Bureau

selected according to whether they reported having an interest in or following statistics on health, unemployment, and/or population counts. Demographic diversity among respondents was also important, especially with regard to education, as this may influence people’s level of knowledge of the FSS and, by extension, their opinions on the topic. The goal was to include respondents with high and low levels of educational attainment (college graduates and those with a high school diploma or less). Some interviews (10) took place in the lab at NCHS. However, the Census Bureau and the Internal Revenue Service conducted the remainder of the interviews off-site in an effort to create an environment conducive to respondents expressing distrust of government and government institutions. Interviews were designed to last 60 minutes and a \$40 token of appreciation was given to respondents at the conclusion of the interview. A summary of respondents appears in Table 1. The sample was about evenly split between male and female, and the age distribution was such that 26 respondents were between 30 and 59 years of age, while seven were under 30 and eight were 60 or over. Half of the respondents had a high school diploma or less and half had some type of college degree (2-year, 4-year or graduate degree).

Table 1: Summary of respondents and number of interviews by agency (n = 43)

<u>Interviewing Agency</u>	<u>Total</u>	<u>Percent</u>
Census	18	42%
IRS	15	35%
NCHS	10	23%
<u>Gender</u>		
Female	23	53%
Male	20	47%
<u>Age</u>		
Under 30 years	7	16%
30 – 59 years	26	60%
60 years and Over	8	19%
Missing	2	5%
<u>Education</u>		
High school diploma or less	13	30%
2-year or 4-year college degree	7	16%
Graduate degree	6	14%
Missing	17	40%

Data Collection

Cognitive interviewing, as a qualitative methodology, offers the ability to understand the interpretive process behind answers to survey questions. Different types of cognitive interviewing techniques exist. Respondent narrative and intensive follow-up verbal probing were the primary methods used for data collection in this project. With these techniques, interviewers administer the survey question, obtain an answer, and then probe the respondent for information relevant to his or her responses. Follow-up probes are initiated when contradictory information is given by the respondent, as this may indicate points of confusion and misinterpretation.

Probes are also useful for exploring pre-identified areas of concern in the instrument. In this study, the working group was interested in knowing not just about the question-response process, but also what respondents' feelings and impressions were regarding administrative record linkage. As a result, interviewers used a standard probe asking respondents why they answered the way they did – that is, why they agreed or disagreed with the question.

All data were entered into and are currently stored in Q-Notes, a software application maintained by NCHS. This software was designed specifically for cognitive interview methods and helps to facilitate a common approach among different agencies – agencies that may normally have very different techniques – who are participating in a single project.

Method of Analysis

Data analysis proceeded according to the grounded theory approach which does *not* aim to test existing hypotheses, but instead generates explanations of response error and various interpretive patterns that are closely tied to the empirical data. The process of analysis is a constant comparison of data in three distinct steps. The first step occurs within the interview as the interviewer attempts to understand how one respondent has come to understand, process, and then answer a survey question. Response errors can be identified by comparing respondents' answers to the survey questions to the narrative they provide during the interview. When logical contradictions are evident between the narrative and the survey answer, the interviewer explores why these contradictions occurred.

The second step in analysis occurs once the interview is over, and is a systematic comparison *across* all interviews. This level of comparative analysis reveals *patterns* in the way people answer survey questions. At this level, it's possible to identify the construct that's being captured by the survey question and illustrate the substantive meaning behind the survey statistic. The final phase of analysis is a comparison of patterns across sub-groups, identifying whether particular *groups* of respondents interpret or process a question differently from other groups. At this level of analysis, that is, identifying patterned differences among subgroups, we begin to understand where potential for bias would occur in survey estimates.

Section I: Results for Questions 15, 16, 25 and R1 – R6 for Each Agency

Questions 15 and 16 were attitude questions asking respondents whether they strongly agreed, somewhat agreed, somewhat disagreed, strongly disagreed or neither agreed nor disagreed with a statement about federal statistics or statistical agencies. Question 15 was, “If a mistake had been made when producing federal statistics, the public would not be told about it.” Question 16 was, “Federal statistical agencies collect data to benefit the public.”

Question 25 was also an attitude question, but phrased as a yes/no question. It was, “Next a question about the U.S. Government as a whole. Do you think federal government agencies share a single central database of the name, address and date of birth of U.S. residents?”

Questions R1 – R6 were all related to personal attitudes about administrative record linkage. Each question focused on a different aspect of the potential benefit of record linkage, such as assuring privacy (R2), saving money (R3), saving time (R4), improving statistical accuracy (R5), or serving the public good (R6). The first question (R1) was simply a general question about favoring or not favoring administrative record linkage. Additionally, R1 through R6 were asked (in slightly different ways) for three different agencies, the Census Bureau, the Bureau of Labor Statistics, and the National Center for Health Statistics.

Overview of Findings

Respondents’ Lack of Knowledge

One of the most salient findings of this research project is that most respondents did not have any conception of federal statistical agencies or their data systems. They did not realize that there exists a formal Federal Statistical System and, indeed, give little thought to how statistics are produced in general. As a result, there were no preexisting attitudes for the survey questions to capture. Respondents answered the questions by formulating an opinion on-the-spot, based not on existing knowledge but on what was presented to them (as they interpreted it) in the question. This phenomenon was explored more thoroughly in the first round of testing but was evident in both rounds.³

Inconsistent Interpretations

In the second round of testing, questions on record linkage asked respondents the extent to which they supported or opposed the federal government obtaining information about people via record linkage rather than collecting the information through the use of surveys. The questions offered different scenarios. In addition to a general question that asked respondents if they supported record linkage, other questions offered caveats that provided more context for record linkage

³ The final report from the first round of testing focused on attitudes about the Federal Statistical System in general (excluding the topic of administrative record linkage) and documents the finding that many (even most) respondents did not have any conception of federal statistical agencies or their data systems. For more about that, see [Cognitive Interview Evaluation of the Federal Statistical System Trust Monitoring Survey, Round 1: Results of interviews conducted in October, 2011](http://wwwn.cdc.gov/qbank/home.aspx), <http://wwwn.cdc.gov/qbank/home.aspx>.

activity. These caveats included: 1) privacy concerns, 2) costs, 3) time, 4) data accuracy, and 5) uses and purpose of data.

Respondents thought of each of these factors when answering the questions, but not necessarily in equal proportion and not always in relation to what the question was asking. For example, if the question asked about respondents' opinion about record linkage in the context of saving money (e.g., see question R3a, R3b, or R3c), respondents also thought about other considerations such as privacy, uses of the data, and respondent burden. As a result, the questions do not consistently measure attitudes in relation to the factor specifically asked about in the question.

The Importance of the Privacy Consideration

Respondents were often guided to certain considerations based on the topic of the question (e.g., some questions specifically ask the respondent to think about accuracy, burden or costs), but respondents took privacy considerations into account to some extent in all questions. In other words, respondents' *degree* of concern over privacy varied by question and by the agency referenced in the question (i.e., Census, BLS or NCHS). However, concern over privacy was present in respondents' evaluations and answers to every question on administrative records. The centrality of privacy was especially prominent in the first question for each agency (R1a, R1b, and R1c).⁴ Each of these questions received the fewest number of those answering 'in favor' because of respondents' concerns about privacy. Neither the ordering of the questions (context effects) nor the emphasis of the questions (on cost, burden, or accuracy) erased the importance of privacy in respondents' minds. They may have incorporated additional considerations into their reasoning, but privacy was rarely absent altogether.

Not only were privacy concerns highly prominent, they were also multidimensional. Respondents thought of and defined privacy in three distinct but sometimes overlapping ways. Privacy was defined as 1) information about a person that is not public, 2) information that is defined as sensitive, and 3) information as a personal possession. These are described next.

Dimensions of Privacy

- a) Information that is inherently not public: Some respondents talked about privacy in terms of information that only the respondent knows and which should not be public information. As one respondent said, "Some things are private." Some respondents felt that certain information should be kept confidential; to make it public would be violating a person's privacy. One specific manifestation of this was to think about personal identifiers. Respondents gave name, address, and social security number as examples of information that identify them individually. One person said, "Someone's name and social security number is confidential and no agency should be able to give that information out. They should get that information directly from the person."
- b) Sensitive information: This included information that respondents considered personal and sensitive. The essence of information being personal or sensitive means that it has the potential to harm the person to which it belongs. As one person said, "I think you

⁴ See the question-by-question analysis for more detailed findings.

should ask first. Sensitive things you might not want them to know.” A common example was information (such as criminal or health records) that, if fallen into the “wrong hands”, might prevent a person from getting a job. Hence, information that others can use to harm a person is considered sensitive. Information that cannot be used to harm a person is not considered sensitive (and, therefore, acceptable to link to). For example, one respondent gave this reasoning, “This is private information between me and my doctor. Just because I get a job with someone doesn’t mean they need to know my health. I don’t want them to find out and then I don’t get hired.”

- c) Information as a personal possession: This was the most common privacy dimension and involves the idea that individuals have the exclusive right to control their own information and that others need consent before having, sharing, or using it. As one respondent said, “I would prefer that they go straight through me and not behind my back.” Another said, “I think what you choose to spend your money on is private, and should only be able to be disclosed at your discretion.” Another stated, “This is very personal information and should only be obtained from a person if they are willing to give it.” And another personal argued, “You wouldn’t know what they were taking or why. You have no control. On a survey [vs. administrative record linkage] you can refuse an answer if you don’t want to disclose that information.”

Question-By-Question Analysis

15. If a mistake had been made when producing federal statistics, the public would not be told about it. *Strongly agree, somewhat agree, neither, somewhat disagree, strongly disagree*

Findings: There was a fairly even split in the way respondents answered this question. Fourteen disagreed while 19 agreed. Seven answered neither agree nor disagree. Respondents generally understood the premise of the question; however, not all respondents were thinking about statistical agencies. Although many were thinking of statistics when answering this question, just about as many respondents were thinking of the federal government in general. This was especially true for respondents not familiar with federal statistical agencies – they would tend to think of the federal government as a whole. For those who agreed with the statement, they tended to think that the government *in general* does not admit to mistakes. Some quotes illustrating this idea include, “I feel like the government doesn’t like to admit its mistakes.” “They want everything that is going on to look good.” “I don’t think they would want a mistake publicized.” “They’re afraid to tell people the truth in that they made a mistake – sort of like sweeping it under the rug.”

An important finding for this question is that many respondents had difficulty answering because it depended on the type of mistake. As one person said, “Depends on the agency and what the situation was.” Similarly, another person said, “I’m thinking it may be dependent on what it is.” Respondents were not knowledgeable enough to give examples of what a statistical mistake might be, so some focused on the idea of the severity of the mistake. As one respondent put it, if the mistake affected the entire population, then the government would report it, but if the mistake affected only a small number of people they probably would not. Someone else said, “I believe it depends on what the statistic is and the degree of misinformation or mistake. If it was huge,

they would probably put that information out. Otherwise, if it was minor, they wouldn't bother." Some respondents who had an "it depends" interpretation answered "neither agree nor disagree" to the question.

Another finding is that the question may assume that not telling the public is a bad thing, when some respondents didn't necessarily judge it that way. For example one respondent said she agreed but didn't see it "as a conspiracy theory or treason". She just thinks mistakes are made and fixed without always telling the public.

Finally, at least one person was confused by the double negative (i.e., having to agree/disagree with a negatively worded statement). She heard the question as saying that the public would be told about mistakes and answered "I can't agree with it." In reality, she should have answered 'agree' to the question.

16. Federal statistical agencies collect data to benefit the public. *Strongly agree, somewhat agree, neither, somewhat disagree, strongly disagree*

Findings: The answers to this question were fairly skewed compared to the previous question – only seven disagreed with the statement and 32 agreed. One person answered neither agree nor disagree and one said 'don't know'. One pattern of interpretation had to do with the difference between the stated intent of statistical data vs. what actually happens with the data. Some respondents answered on the basis of what is supposed to happen with statistical data. As one person said, "I believe the reason they collect statistics is to show where needs exist and the public benefits from the collection of the information." Another person talked of benefitting the public as the "core mission" of statistical agencies. Others answered on the basis of what they think actually happens with uses of statistical data. One respondent expressed the opinion, "I think the public is last on the priority list." Still other respondents considered both scenarios (i.e., what is supposed to happen with statistical data vs. what actually does happen) which makes the question more difficult to answer. One person expressed the idea best:

"I think that the purpose of the data they collect is to inform government programs or sometimes even private organizations and employers as well as funding agencies. So I think the purpose behind them is to help the public. I think there are instances in which the statistics have been used in the past or even now...that they have unintended consequences or they're not used in a way that is the best way that they could be used. But I think the goal behind their collection is to benefit the public."

A second interpretation pattern included some respondents thinking that the question was asking whether statistical agencies collect data to benefit the public vs. someone else (usually the government itself). One person said, "They don't collect things to benefit the public, they collect things to benefit themselves." However, when asked who "they" were she said "the government." This respondent was thinking of government as a whole, but others were thinking of statistical agencies. One person said, "I think they do a lot to benefit themselves to get more money." Another person had the same opinion, "A lot of it is for their own benefit, but they wouldn't do it if it did not benefit the public somewhat."

A third pattern of interpretation was respondents understanding the question as asking whether statistical information is collected to benefit vs. harm the public. One person said, “I would think that they would be trying to do things that would help the American people.” Someone else echoed, “They don’t do it to screw people over!” With this interpretation, the question is nonsensical.

25. Next, a question about the U.S. government as a whole. Do you think federal government agencies share a single, central database of the name, address and date of birth of U.S. residents? (yes/no)

Findings: There was an even split of about half the respondents answering yes and half answering no to this question. Most respondents understood the “central database” premise to be referring to the idea that government agencies have access to each other's data, if not through, literally, a common database, then at least through cooperation and communication.

After answering yes or no, respondents were asked why they answered the way they did. Respondents who answered ‘no’ tended to think that the logistics of creating, maintaining, and accessing such a database were not realistic. For example, one person said, “It seems like a lot of information to manage. And I think it’s not unreasonable, but I just don’t think the cooperation between the agencies is there to share all that data.” Another person said, “I think they have different databases, there is no one central location.”

Additionally, not all respondents judged a central database as a bad thing to have. One person said, “I think that would make sense but I can’t imagine that they actually do.” Someone else said, “I don’t think different parts of the government have interoperable systems. I also think this is a big problem.” Similarly, another person said, “I don’t think they’re that organized. I think it would be a good idea.” She believed that competition for funding prevents cooperation among government agencies. This lack of cooperation was expressed by another respondent, “It seems like agencies don’t really talk to each other. They each collect their own information.”

Others did think a central database is undesirable. A few respondents expressed deep concerns over the government keeping tabs on citizens. “They have to have that to keep tabs on people.” One person suggested that this is justified through the lens of national security. “I think it’s in the best interest of national security in their minds – I think they believe that it’s necessary.” Another person said, “Yes, definitely – there is one button that they all push.”

Question designers should consider the fact that some respondents judge the “central database” premise as a good thing to have and others see it in a negative light. In other words, if the question was written with the understanding that a central database is inherently negative, the evidence here shows that to be a false assumption.

Questions on Administrative Record Linkage: R1 – R6

Similar questions on administrative record linkage were asked in relation to three federal agencies, Census, BLS, and NCHS. Question numbering and wording reflect this grouping such that:

Group a = Census
 Group b = BLS
 Group c = NCHS

Questions were asked in different rotation. In some cases questions on Census were asked first (followed by BLS and NCHS question), in others BLS questions were asked first, and others started with the NCHS questions. There was, however, no discernible effect on how respondents answered the questions or on the considerations they took into account when arriving at their answers. That’s not to say that the agency asked about made no difference – it did. But the ordering of agencies did not seem to fundamentally change the way respondents answered.

The U.S. Census Bureau was by far the most recognized agency with a well-understood purpose and mission being linked to actual census taking (other surveys conducted by the Census Bureau were not well known). This knowledge informed and shaped respondents’ understandings of and answers to the questions. This partially explains why the Census questions had the fewest “against” responses – people generally agreed that taking a national census is important. Almost no one knew of NCHS, but respondents could at least relate to and understand the concept of health statistics. Thinking about the importance of health data grounded many of the interpretations to the NCHS questions.

The BLS questions were the most confusing set of questions for respondents. Neither BLS, nor the CPI, nor the CES were things that respondents had heard of. This made understanding the questions more difficult and contributed to the BLS questions producing the highest rate of “against” and “neither” answers. In other words, it was difficult for respondents to be in favor of something they knew nothing about. See Table 3 for a breakdown of how respondents answered the questions.

Table 3: Summary of Respondent Answers to the Survey Questions on Administrative Record Linkage⁵

	Census Version (a)			BLS Version (b)			NCHS Version (c)		
	Favor	Against	Neither	Favor	Against	Neither	Favor	Against	Neither
R1 General	20	13	7	11	27	2	19	19	3
R2 Privacy	30	7	2	11	18	7	23	16	1
R3 Costs	28	10	1	16	18	4	28	11	1
R4 Time	28	9	1	16	20	2	22	14	4
R5 Accuracy	29	8	1	15	18	4	26	10	2
R6 Use of Data	27	6	3	16	15	5	31	6	3

⁵ See question-by-question section for full text version of each question.

R1a. For the next census in 2020, if the Census Bureau could obtain your name and age from the Social Security Administration, instead of asking you for this information on a questionnaire, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*?

Findings: Respondents were somewhat more in favor than against this question, 20 (in favor) to 13 (against). In fact, Census questions produced the least amount of ‘neither’ responses overall. The evidence suggests this is because people are familiar with the purpose of Census (and, therefore, have an actual opinion of the agency) and because the question gives specifics such as the type of information to be collected.

Respondents were thinking of a variety of considerations in this question, not just name and age. For example, some mentioned social security number and address. Respondents also thought about privacy, purpose of the data, accuracy, cost and burden even though the question makes no specific reference to these issues. However, the majority of respondents were thinking about privacy concerns, either in favor or against it. Different dimensions of privacy were raised, often depending on how respondents felt about name and age – or other identifiers that seemed similar, such as social security number. Some felt that these should always remain private. Others felt name and age are not sensitive and pose no potential for harm, but if the Census Bureau also obtained social security number (since the information was being obtained from SSA), this would be problematic. For example, one respondent said, “Someone’s name and social security number is confidential and no agency should be able to give that information out. They should get that information directly from the person.” Others had a different evaluation. “I don’t think that’s very difficult information to obtain. I don’t have such a feeling that it’s ‘private’ if that’s the right word.”

Respondents also considered the purposes or uses of the data being collected. Some were in favor because it was “important information” for the Census to have. Others were against it precisely because they didn’t know why the Census needs this information. “Maybe if I know the reason...”, “I am more skeptical about how they are going to use the information.”

Some people saw the question as nonsensical because the examples seemed like something the Census already knows. Several respondents expressed this confusion, but one person said it best, “You already know! How do you know to contact me for the Census? It just doesn’t make sense. They used to show up at your door. Now they call me on the phone and ask for me by name.” She couldn’t answer the question. However, others with the same level of confusion did answer the question. One person was in favor, reasoning that “They already have the information anyway.” Others were in favor for the same reason.

Accuracy was another consideration in this question. Some were in favor thinking that SSA would have more accurate information than what Census could obtain from individuals. These respondents were thinking of people who lie to Census or who simply get missed in the count. However, others were against it for the opposite reason – they believed that more accurate data would come directly from individuals. Their main premise is that SSA records can be out of date or otherwise wrong. “I would rather they ask me because I know they would get correct information. There would be no mistake.”

Reducing costs also came up a few times. A couple respondents were in favor thinking that it could help reduce costs (time and money) for Census. But one person was against it for the same reason. He suggested it was a form of “cutting corners”. He felt as though the Census needs to continue going door to door. This overlaps with concerns for accuracy.

Finally, a few respondents considered burden when answering the question. One person was in favor “so they wouldn’t have to bother me.” In a similar vein, one respondent said, “It takes time to sit and complete the form.” Another person felt as though they were “intruding in my life”. On the other hand, some were against the question because they felt as though it was not a burden for them to provide the information, so they would rather do it.

R1b. To produce the Consumer Price Index, the Bureau of Labor Statistics collects information on people’s living expenses and purchases they make through the Consumer Expenditure Survey. If the Bureau of Labor Statistics could obtain some of your purchases from a credit card company instead of asking you on a questionnaire would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*?

Findings: This question is the BLS version of question R1a (for Census). While respondents tended to be in favor of the Census version (because they were familiar with the Census Bureau), they were against this question 27 (against) to 11 (in favor). Only two answered ‘neither’. This was the linkage question that produced the most ‘against’ responses.

Additionally, respondents tended to be more confused by this question than by the others. The question is long and, unlike the Census, many people do not know the CES or what the CPI is. The people who were confused often answered ‘neither’ (or ‘against’). This trend is observed for all questions relating to BLS.

Privacy was by far the main concern, heavily out stripping other considerations. Moreover, all dimensions of privacy were mentioned. Some examples include, “I feel like how you choose to spend your money is your own choice and the government should not be able to get this information whenever they want it.” “That’s like prying into your personal affairs.” “I would rather that information be voluntarily submitted.” “As long as they ask permission.” “It depends on if they knew it was MY purchases or just A purchase.”

The use of the data – specifically uses for public good – was another consideration for some respondents. Some said they were against it because they couldn’t see the value in it. One person who was in favor of this for the NCHS question but against it here explained that “I believe more in health.” Others expressed the same view, “Health is an important issue as is the government knowing the population count.” But it’s not so important to know what he buys. Another indicated, “I could also see it being an issue depending on how the data would be used.” On the other hand, some thought it was a good thing to know, “It would be very interesting information to have.”

A couple respondents answered ‘in favor’ because they believed the government is doing this (record linkage) already.

R1c. The National Center for Health Statistics collects information on people's health and medical care through a variety of surveys. If the National Center for Health Statistics could obtain some of your medical information from your health care provider instead of asking you on a questionnaire would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it?*

Findings: This is the NCHS version of question R1a (for Census) and R1b (for BLS). There was an even split on responses: 19 were in favor and 19 were against. Two main considerations factored into respondents' interpretations: privacy and uses of the data for public good. Additionally, respondents thought of different types of information that might be revealed in their health records and how this information may or may not be private.

Privacy was the main consideration, followed by the use and purpose of the data. All were considered in the context of the type of health information in a record. Much like BLS, respondents did not know NCHS (like they know Census) but, unlike BLS and the CPI, they could identify with health information which was seen as both sensitive *and* potentially beneficial to serving the public good. As one person put it, "There's a limit." If they collect age and medical conditions like migraines, that's fine. "But not personal information like name, address, social security number...driver's license stuff." Someone else said, "That's a tough question...it's wise to know what the agency is dealing with AND how to deal with the individual. The person may not like it, but it's important for the agency to know."

Despite the potentially sensitive nature of health information, many people saw a public benefit to government knowing this information. One respondent answered 'somewhat against' it (and not strongly against it) because "health information is necessary for the public good." Another person said, "I'm less cynical about health information. I think that they are going to use this information in a good way." "I think it would be of good benefit to the country."

However, the sensitive nature of some medical information also weighed into people's considerations, but not all medical information was seen as equally sensitive. For example, one person said, "Like people with HIV. They say it's secret, but I don't really think it should be. People should know." On the other side, someone said, "I don't feel comfortable people knowing that I have hepatitis C or AIDS. I don't want them to know." "Like telling the government that I had gonorrhea this year is not something I really want to do!" "Some things are private." Some respondents thought that if this sensitive information became public, it could harm them in some way (like getting hired for certain jobs) while a couple respondents did not think this to be the case. "Health information is important and there's not a lot of bad you can do with health information." A few respondents said they would be in favor if they knew they could not be personally identified.

Additionally, because of the potentially sensitive nature of health information, many respondents wanted to be able to give consent first. "I think you would want to ask my permission first. Wouldn't that be a breach of HIPPA?" "I think you should ask first." One person noted for all the questions that "None of them say 'with your consent.' You wouldn't know what they were taking or why. You have no control. On a survey you can refuse an answer if you don't want to disclose that information."

Finally, the issue of accuracy came up several times. Respondents said they can't remember things about their health and that their medical record would be far more correct than their memory.

R2a. If you knew that this information was only being obtained from the Social Security Administration to produce statistics, and that your personal information would remain unavailable to the public, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*??

Findings: This Census question garnered more favorable support than the previous Census question (R1a). Thirty respondents said in favor, 7 were against, and 2 said neither. (The previous question's answers were 20, 13, and 7 respectively). This is likely because it directly addresses the main consideration respondents thought of in answering questions on linkage – privacy. Specifically, by mentioning confidentiality the number of those in favor increases. Conversely, the number of those against decreased.

As with the previous question, the main consideration was privacy, regardless of how the respondent answered the question. In other words, even when respondents were in favor of linkage, many were weighing their answers in relation to privacy. For example, one person answered 'strongly in favor' and explained, "Even if someone got this information, there is no big risk. So what if they know my name and address?" Other respondents thinking of the same topic answered 'against' because they understand their name as a personal identifier that they don't want made public. Still others were not limiting their thinking to name and address. To some extent this is because this question (unlike the previous) generically refers only to "personal information".

A variety of privacy dimensions were mentioned, including the type of information (whether it was sensitive information or not), who "owns" the information, and the fact that consent must be given. Privacy, however, was not the only consideration. Some respondents also talked about accuracy, burden, and cost in explaining how they arrived at their survey answers.

R2b. If you knew that this information was only being obtained from credit card companies to produce statistics, and that your personal information would remain unavailable to the public, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*?

Findings: Respondents were less in favor of this question on BLS than with similar questions on other agencies, R2a (for Census) and R2c (for NCHS). Only 11 were in favor, 18 were against and 7 said neither. However, fewer respondents were against this question in relation to the previous one for BLS. In the previous question (R1b) 11 said they were in favor, 27 were against, and 2 said neither. The number of those saying either against or neither decreased in this question largely because the question addresses people's main concerns about privacy. Various dimensions of privacy were mentioned by respondents, and privacy was the most-often mentioned consideration. This is not surprising, given that the phrase "personal information" suggests that the information may be sensitive.

In addition to privacy considerations, accuracy, uses of data and burden were also mentioned. In addition, the fact that linkage would be to a credit card company was also something respondents weighed in their mind. When thinking of this aspect, most respondents answered against linkage, citing reasons such as, “Credit card companies will say one thing and do another.” “I don’t want my credit card company dealing with my information.”

R2c. If you knew that this information was only being obtained from health care providers to produce statistics, and that your personal information would remain unavailable to the public, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*?

Findings: Similar to the pattern found in the previously discussed questions, more respondents were in favor of linkage in this NCHS question than the first (R1c) because it addressed the issue of privacy. In the first NCHS linkage question (R1c), respondents were evenly split between being in favor or against (19 to 19). However, in this question 23 said they were in favor, 16 said they were against, and only one answered neither.

In arriving at an answer, most respondents thought of some aspect of privacy. For the health topic in particular, the nature of the data was an important factor. For example, some judged health information to be sensitive and others did not. It depended on the type of health information a respondent was thinking about (e.g., having the flu is deemed different from having HIV). The uses of data for public good also factored into respondents’ answers. In fact these were the two dominant considerations mentioned by all respondents. Accuracy was the only other consideration mentioned, and that was by only one respondent. In sum, this question topic produced fewer considerations than similar questions for Census and BLS.

R3a. The 2010 Census cost over \$10 billion dollars. To reduce this cost for the 2020 Census, the Census Bureau could get your name and age from the Social Security Administration. If this method could save government money, *are you strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*?

Findings: Respondents were in favor (28) more than against (10) this question. The most common consideration that shaped respondent answers was cost, which is not surprising given the topic of the question. However, the different figures -- \$10 billion here, \$40 million for BLS, and \$100 million for NCHS – had no discernable effect on what respondents were thinking or how they answered. In other words, respondents were not more likely to agree with the question proposing to save more money (\$10 billion for Census) than they were to agree with the question proposing to save less money (\$100 million for NCHS). In fact, the numbers in agreement were exactly the same (28 in favor for both Census (R3a) and NCHS (R3c)).

That’s not to say that respondents did not think of cost – many did. Most were just generically in favor of saving money, “It saves money, so I’m for it.” Those in favor thought it might help reduce the deficit or perhaps save them in taxes.

On the other hand, some respondents disagreed with the premise that linkage would save money to begin with. “I don’t see where the savings is.” Another person asked, “How would you do that?” She reasoned that the Census would have to go the extra step to get your permission first, which would mitigate any cost savings. Note that some answers of “against” do NOT mean that they are against reducing costs. It means that they are skeptical of the basic premise of the question, that record linkage can save money. This is a different construct, and suggests the question is potentially double-barreled.

Two other considerations, privacy and accuracy, were also mentioned, both in equal measure to each other. Regardless of cost savings, some respondents were still weighing the private nature of name and social security number. Others expressed doubts that the social security administration would necessarily have correct information on everyone, while others thought social security would have better information because of non-response on Census forms.

R3b. The Consumer Expenditure Survey costs over \$40 million dollars. To reduce this cost, the Bureau of Labor Statistics could get some of your purchases from a credit card company. If this method could save government money, are you strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it?

Findings: Unlike the Census and NCHS versions of this question (R3a and R3c respectively), respondents were more likely against (18) than in favor (16), though the split is nearly even.

Unlike the Census and NCHS versions of the question, cost was NOT the main consideration weighing into respondent’s answers, despite the topic of the question. Instead, respondents considered many different factors, with privacy concerns being the most frequently mentioned. In addition to privacy, respondents thought of uses of data and respondent burden when formulating their answers.

Finally, in at least four cases, respondents were outright confused by the question. The confusion was prompted mostly by the fact that people were not familiar with BLS, the Consumer Expenditure Survey, or both.

That respondents factored more considerations into their answers and were more confused by this question than the same questions for Census and NCHS could partially explain why they were less likely to favor it vis-à-vis the same question for other agencies.

R3c. The National Center for Health Statistics spends over \$100 million on surveys every year. To reduce this cost, it could get some of your medical information from your health care provider. If this method could save government money, are you strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it?

Findings: Similar to the Census version of this question (R3a), respondents were in favor (28) more than against (11) record linkage.

Four considerations were mentioned, with cost being the most prominent, followed by privacy and uses of data. Accuracy was also mentioned, but only by one respondent. Like other NCHS linkage questions, respondents mentioned both the need for privacy of their personal information (“some health things you just don’t want out there”) and the fact that personal health information can serve a public good (“boy, that’s a tough one – if it’s for legitimate reason like cancer, then yes”). Part of what respondents’ considered was that some types of medical information (like having the flu) are less sensitive than others (like having HIV).

R4a. Various government agencies ask you to provide the same information on a number of forms. If the Census Bureau could save you the time and hassle by getting your name and age from the Social Security Administration, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*?

Findings: Respondents were in favor (28) more than against (9) this question. Respondents took four different factors into consideration when answering this question, with privacy being the most prominent, followed by burden, accuracy and cost. Moreover, not everyone limited their thinking to name and age. Some were thinking of social security number and/or address. Given that the previous question shifted respondents’ focus to cost because it asked about cost, it was surprising that burden was not a stronger consideration in this question.

R4b. Completing the Consumer Expenditure Survey takes a lot of time and effort. It requires you to keep a diary of your purchases over a period of time. If the Bureau of Labor Statistics could save you the time and hassle by getting some of your purchases from a credit card company, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*?

Findings: Unlike the Census and NCHS versions of this question (R4a and R4c respectively), respondents were more likely be against (20) than in favor (16) of this one. This pattern is similar to that observed in the previous BLS question.

Respondents considered many different factors when thinking of this question. The considerations they took into account included privacy, burden, accuracy, uses of the data, and the type of organization engaging in the linkage. However, privacy was the most commonly mentioned consideration that shaped respondents’ answers.

As in other questions related to BLS, this one created some confusion. At least four respondents demonstrated confusion in understanding the question. Some were silent misunderstandings, but one person expressed outright confusion, “I’m kind of confused on this. Please read it again.” Another respondent was so confused by the BLS questions that the interviewer gave up and skipped to the next set of questions so as not to exasperate the respondent. That respondents had difficulty comprehending what this question was asking is further evidence that this series of questions should be simplified. For example, it is probably unnecessary to reference a specific survey like the CES.

R4c. If you are selected for one of these health surveys, they can take a lot of time: they often ask people to recall detailed medical histories, doctors' visits, and treatments. If the National Center for Health Statistics could save you the time and hassle by getting some of your medical information from your health care provider, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*?

Findings: Similar to the Census version of this question (R4a), respondents were in favor (22) more than against (14) record linkage. Four answered neither.

Respondents mentioned four different considerations, but privacy was by far the most common. Burden, accuracy, and uses of data were also mentioned. Even though the question specifically referenced burden, most respondents were persistent in thinking more about privacy concerns, and in fact, only six people mentioned thinking about burden at all. Like the other questions on health information, when considering linkage to their health records, they thought about the extent to which health information is sensitive and private – something to be known only by their health care providers. When agreeing to linkage, they often concluded that the information is not overly sensitive, that personal identifiers would not be released, or that they would be asked for consent to link.

R5a. Many Americans don't return their census forms, and as a result the census numbers may be incomplete. If the Census Bureaus could improve the accuracy of their numbers by asking the Social Security Administration for the names and ages of the missing people, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*?

Findings: As with the previous Census question, respondents were in favor (29) more than against (8) this question.

Accuracy and privacy were the two most mentioned considerations. This is not surprising to the extent that privacy was a consistent and prominent consideration in all the questions on record linkage and that the question directs respondents to think about accuracy issues. Cost and burden also came up as considerations, but only by one respondent each.

R5b. Many of the purchases asked on the Consumer Expenditure Survey may be difficult for a person to recall accurately, but a credit card company may already have this information. If the Bureau of Labor Statistics could improve the accuracy of the Consumer Price Index by asking a credit card company for some of your living expenditures or purchases, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it*?

Findings: Unlike the Census and NCHS versions of this question (R5a and R5c respectively), respondents were more likely be against (18) than in favor (15) of this one. Four respondents answered neither. This is the same pattern as observed in the previous questions. As with other questions, privacy was mentioned most often, but accuracy was the second most mentioned consideration. One person mentioned burden and one other mentioned uses of the data.

R5c. A lot of the information asked on these health surveys may be difficult for a person to recall accurately, but a health care provider may already have this information. If the National Center for Health Statistics could improve the accuracy of their numbers by asking your health care provider for some of your medical information, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it?*

Findings: Similar to the Census version of this question (R5a), respondents were in favor (26) more than against (10) record linkage. Only two answered neither. Privacy and accuracy were considerations mentioned in equal measure. Burden was also mentioned several times.

R6a. Census numbers determine where local, state, and federal agencies build new schools, roads, and firehouses. If the Census Bureau could obtain names and ages from the Social Security Administration to get a better idea of where these types of services should be located, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it?*

Findings: As with the previous Census question, respondents were in favor (27) more than against (6) this question. Respondents were largely focusing on the examples given (i.e., schools, roads, and firehouses). As a result, the most common consideration for this question was the purpose or use of the data being collected. When the purposes and uses of collecting data via record linkage are explained, respondents are more likely to favor linkage. For example, one person said, “Here is the answer to my question about wanting to know what they intend to do with the information. As soon as I hear the purpose of the question and how the data will be used, then I can say strongly agree.” Another person had also changed his views from other questions for similar reasons, “I didn’t think about planning for schools and things like that.” This is an important finding to the extent that it relates to the fact that most people do not know much – if anything at all – about the FSS or administrative record linkage. Therefore, questions that provide some information about what statistics are used for allow respondents to give a more meaningful response.

Privacy and accuracy were the second most mentioned considerations. Burden was mentioned as well, but only by one respondent.

R6b. The Consumer Price Index is used to help calculate living costs, pensions, and wages across the country. If the Bureau of Labor Statistics could obtain some of your living expenditures and purchases from a credit card company to get a better sense of how people really live, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it?*

Findings: This question had an even split between those in favor (16) and those against (15). Five answered neither. In the series of record linkage questions, the questions on BLS by far elicited the most ‘against’ responses. Respondents consistently demonstrated a wide range of considerations when answering the BLS questions and were confused by them more often than they were by the questions about Census and NCHS. That pattern continued in this question.

Respondents considered privacy, accuracy, uses of data, and burden when answering this question. Five respondents were confused and had trouble understanding the meaning. However, of all the BLS record linkage questions, this one had the fewest ‘against’ answers. To a large extent, this is because the purposes and uses of the data are included in the question. Respondents often do not have awareness of or understand the purpose of federal statistics. This makes it difficult for them to answer questions on the topic. However, questions that provide some information – such as this one – give respondents a better grasp of what’s being asked and allow them to formulate a more meaningful answer. One respondent sums this up nicely, “I am struggling with this. Earlier questions had me asking myself about how much I trust the government. Are the statistics compromised? What are the politicians doing with the information? I tend not to be concerned about that. I am concerned about what the government is doing with the information. Now [I’m] understanding the question a little better and sensing earnestness of the need for information.” Others also mentioned that this question “lays out more clearly what the CPI is used for.”

R6c. Information from these health surveys help make funding decisions for cancer research, elderly medical care, and hospitals. If the National Center for Health Statistics could obtain some of your medical information from health care provider to get a better idea of the health care needs of the nation, would you be *strongly in favor of it, somewhat in favor of it, neither in favor nor against it, somewhat against it, or strongly against it?*

Findings: Of all the questions on record linkage, this was the question with those answering most in favor (31). Only six were against and three answered neither.

The factors taken into consideration were privacy, uses of the data, accuracy and burden. When answering the NCHS questions in particular, respondents were more likely to weigh privacy concerns about their health information against the potential benefit that health statistics offer the public. Because this question specifically addresses a public benefit to collecting individual information, respondents were more likely to favor record linkage. For some, concerns over privacy still prevailed, but for many the public good outweighed privacy concerns. Respondents said things like, “It’s for the good of the nation”, “It might save my kids; saving a life is a good deal”, “If they needed the information to determine more nursing homes or more medical clinics, then okay”.

Section II: Results for Questions 8 – 11

This section discusses Questions 8, 9, 10, and 11 of the FSS POMS Cognitive Interviewing Questionnaire. These questions presented respondents with messages related to administrative data use and data sharing and was a qualitative exploration of these ideas. These questions are not, and were not designed as, survey questions. Instead of functioning as a cognitive interview, this section was designed to be administered as a semi-structured interview exploring the various ways FSS members could communicate with the public about data sharing and administrative data use. The respondents were given four different messages and then asked the same six probes following each statement in order to start a conversation about each message. Following the question-by-question analysis below, the major emergent themes and concerns for each of the

four messages are detailed and summarized. Again, the overall purpose of this section of the interview was not to necessarily prepare questions for the FSS POMS, but rather to better understand the contours of privacy and confidentiality as applied to federal statistics and data sharing.

Question 8

Question 8 asked the respondents to consider the following message: “The United States government will only ask you for a piece of information one time.” After hearing this message, respondents were asked the following structured follow-up probes: *1) What do you think that statement would mean? 2) How does it make you feel? 3) Does it imply data sharing? 4) Does it imply efficiency? 5) How does all that make you feel? 6) Do you support this statement?*

Overall, respondents supported this statement, however, a variety of reactions and themes did emerge from the discussion.

Confusion: Some respondent expressed confusion, which manifested in two distinct ways. First, a few individuals did not understand the wording of the message at all. For instance, one respondent who did not seem to understand what this statement was saying explained that “I have no idea what this could mean.” However, this respondent went on to say that they “did not believe it to be true.”

The second type of confusion was less about language and more about not understanding the process behind this message. Some respondents questioned why would the government ask only one time, or to what types of information does this apply? One respondent said this statement meant the government will only ask her for info once but is not clear on what type of info this might be, saying, “Not sure why they are asking *one* time.”

As evident throughout the analysis of the FSS POMS qualitative research, it appears that respondents want to understand the processes behind data gathering, sharing, and usage.

Government Limitations: This message exposed seemingly opposite sentiments about the federal government that were evident in the interpretations of the questions in Section I of this report. On the one hand, some respondents thought that the government already had a centralized database of personal information to which various agencies have access, saying for instance that “It implies there is a big central data bank.”

On the other hand, some other respondents thought that the federal government is too disorganized and inefficient to pragmatically achieve what this message suggests. One respondent explained that:

This statement makes me feel good if they could do it...but considering how you interact with the government in different ways I am a bit dubious that it would actually happen.

This division notwithstanding, overall the respondents understood that the message implied that some sort of centralized government data system would be necessary.

Discomfort and *Fait Accompli*: Some respondents did appear to be somewhat uncomfortable with this message, though no single source of this discomfort emerged as a dominant theme. The reasons ranged from worrying about the lack of control an individual would have if this single government database existed, to distrust in the integrity of all government workers who would have access to their information.

Furthermore, some respondents displayed the “*fait accompli*” attitude seen throughout other phases of the FSS POMS qualitative research. This view is exemplified by a respondent who explained their positive answer by saying, “All right, since there is nothing I can do about it. I’m Okay with it.” Another respondent answered in a similar vein:

It makes me feel good that they’re trying to become more efficient especially now with the internet, like Google...they already know my address and employment and all that information so I don’t really have a problem with the government doing it if the goal is to make the data more accurate.

Therefore, while individuals expressing this attitude might also feel uncomfortable about the prospects of the government sharing data between its agencies, they also appear to understand that as society and technology change, the way the government collects and handles information must change as well.

Burden Reduction and Efficiency: Respondents generally understood this message to imply a reduced amount of burden. One way the respondents approached this burden reduction was generally selfish—they understood the message to mean that they would receive fewer calls from the government and have to spend less time giving the government information. Respondents described this personal burden as if dealing with the government and its agents was a hassle, saying things like “...they aren’t going to keep calling you and asking you for the same information year after year.” Another respondent explained that she supported the statement because, “[I feel better that I] won’t be bothered frequently.”

The other view was the reduction of burden on the government itself, or government efficiency. The respondents often described this idea of government efficiency in terms of either saving or wasting government time or money. For instance, one respondent said, “It would save a lot of time and money” when defending his answer in favor of the statement. Another explained that “...I don’t like a waste of time or money—if that’s what we’re talking about, then I’m in favor of it.”

Accuracy: Accuracy emerged as a weak theme throughout the respondents’ remarks about this message. Some respondents pointed out that by using non-survey sources of data, the final statistics would be more accurate. However, more individuals questioned whether using stored information would actually take away from the accuracy of the final data. For example, one respondent explained her hesitation by saying:

I also don't know how much of that information would actually stay the same. Your medical history is constantly changing, where you live, your employment, your groceries, things like that, those things are always changing so I don't know how much is actually out there that doesn't change.

Another respondent had a difficult time choosing either a positive or negative response to Q8's statement, saying:

It is both positive and negative since if information is inaccurate somewhere it will be repeated inaccurately everywhere or it is lost or compromised somewhere it will be lost or compromised everywhere.

Here we see an interesting consequence in the respondents' minds to the repeated (and more efficient) use of the same information—that it will “degrade” over time.

Question 9

Question 9 asked the respondents to think about the statement: “The United States Government will not release your personal information to non-Governmental entities.” After hearing this message, respondents were asked the following structured follow-up probes: *1) What do you think that statement would mean? 2) How does it make you feel? 3) Does it imply data sharing? 4) Does it imply efficiency? 5) How does all that make you feel? 6) Do you support this statement?*

The strongest themes to emerge from the discussion of this message focused on data security and trust in government, though some confusion was evident as well.

Confusion: Of particular note in regards to the confusion was the word “entities.” One respondent had to specifically ask what "entities" meant. Another respondent expressed his confusion by questioning

What do they mean by non-government entities? Does that mean not to the states or would that be implied? What other governmental entities are they referring to? FBI?

Another individual pointed out that the phrase “government entities” was vague, and like the quote above, wanted to know if there were intra-governmental limitations.

Discomfort: Following the vagueness noted above, this message did make some respondents uncomfortable or hesitant. Chief among the concerns was that some agencies that the respondents are not comfortable or familiar with would have access to individual-level data that they otherwise would not have. For example, one respondent (who answered the question in the negative) said:

That's good, but I assume that the information is being released to governmental agencies which I'm not comfortable with.

Another worry appears to be a process-based corollary, and relates to a concern that emerged with the message in Question 8 and in many interpretations of the questions in Section I—a loss of control. Because the statement did not provide details on what agencies would have access to the data, and how the agencies could get it, some people felt that this loss of control over their data was unacceptable. For instance, one respondent expressed stress because “you don't know whether they are going to give it [his data] out.” Another was worried about potential personal harm, explaining that:

I could see a problem of financial information being shared with the government for bad reasons...like if you were searching for foreclosure information on Google and then that was used to prevent you from getting a loan or that would be bad.

This worry, therefore, has its roots in the fact that some respondents are concerned that the agencies that now would have access to their data would harm them in some way. Alongside this worry, a distrust of government emerged as a secondary theme.

Trust and Confidence: Interestingly, trust and confidence in government processes and its agents emerged as a theme much more strongly than distrust of them did. A number of respondents believed that the message implied that the government would put more of an emphasis than it already does on preventing the unauthorized dissemination of personal information. One respondent remarked, “I like that. I'm a stickler for information.” Another respondent explained her favorable attitude towards the statement by saying, “It would reduce some of my anxiety if indeed none of it would be allowed to get out.” In general, the respondents who applied this pattern of interpretation expressed that a heightened effort by the government to protect data would make them feel more comfortable.

Data Security: Data security and safety strongly emerged as a theme throughout the respondents' reactions to the message in Question 9. They appeared to understand this statement as ensuring data safety in two different ways: the security of the data and the confidence in government processes and government workers.

The respondents understood data security to mean that the government protected their data and kept it from other members of the public. For example, one respondent noted that, “This statement means that everything is within the government, all personal information is safe.”

The respondents also related security to their confidence and trust in the government procedures and workers to whom they were signing over their information. One respondent explained this confidence saying, “Well, it means, if I signed a release, they are to be trusted.” Another respondent, who also answered the question positively, said that “It [the statement] makes me feel okay, it gives me that it will be protected and that it won't be given out improperly.”

Limitations on Data Sharing: Related to the two ways the respondents envisioned data security is the fact that they clearly understood this statement to mean that the government was limiting

data sharing within the government and not allowing their information to reach not only other members of the public, but also private businesses. When asked whether the statement “implied data sharing,” a number of respondents did not simply reply that it did, but also qualified their answer by explaining the limitations the message implied. For example, one respondent explained that there were boundaries between the government and the non-governmental realms, and said, “It does imply data sharing within the government but not outside.” Another respondent agreed, saying, “[It implies sharing] among government agencies it does, but not in private sector.”

While the message explicitly stated this limitation, it is important to know here that the respondents understood that it allows intra-governmental sharing. For instance, a respondent said she understood the statement to mean, “[It means] that your information won’t be shared with anybody other than federal agencies.” Others expressed this opinion using by again positing a barrier between government and non-governmental entities. For example, one respondent said, “the information would be held within the government itself and not outside of it.”

Efficiency and Accuracy: Efficiency and accuracy emerged in the discussion of this message, though in a less prominent way than did data security or the perceived limitations on sharing. Respondents tended to believe that the message implies that the government is using its resources wisely in order to provide a better service

Question 10

Question 10 asked the respondents to respond to the statement: “Federal Statistical Agencies will not release your personal information to non-Statistical agencies.” After hearing this message, respondents were asked the following structured follow-up probes: 1) *What do you think that statement would mean?* 2) *How does it make you feel?* 3) *Does it imply data sharing?* 4) *Does it imply efficiency?* 5) *How does all that make you feel?* 6) *Do you support this statement?*

Though some respondents understood this message to mean that limitations would be placed on data sharing, most respondents did not understand or agree with this statement.

Confusion: A total of 17 of the 29 responses to this question expressed some level of confusion or misunderstanding. First and foremost, there was clear confusion over what a statistical or a non-statistical agency was—a consistent finding throughout the questionnaire, which has already been discussed above in Section 1. Some respondents thought that the phrase referred to the government as a whole, while others tried to determine what made an agency a statistical or “non-statistical” agency.

The confusion about what constitutes a “statistical agency” is especially noteworthy in light of the fact that these message questions were asked at the *end* of the full FSS POMS, and statistical agency was defined previously and then used consistently throughout. In addition to the general public being largely unaware of the FSS, it also seems possible that the term “non-statistical” agency contributed to this comprehension issue. Two likely explanations for this misunderstanding are 1) that adding the new term “non-statistical agency” after the whole survey

had consistently referenced “statistical agencies” simply cognitively tripped up the respondents; and 2) that the respondents do not naturally break the government up into statistical and non-statistical spheres in their mind, and were therefore unable to quickly make this distinction. Both the Round 2 focus groups (see Smirnova and Scanlon, 2012) and the other rounds of the FSS POMS cognitive interviews back up this idea that respondents do not understand the government in statistical/non-statistical terms.

A further point of confusion came from the overall meaning of the statement itself. A number of respondents did understand this statement to provide the same limits to data sharing that they did in the previous statement in Question 9. One respondent, for example, thought that this statement was breaking down the barrier he had previously perceived between the government and non-governmental agencies. He explained his thinking by saying that:

That would imply to me that other non-governmental agencies doing statistical work would have access to the information, I’m not sure what other groups they would be talking about.

This interpretation is quite literal, as the message itself references federal statistical agencies, but not *federal* non-statistical agencies.

One last point of confusion that should be mentioned is confusion over the purpose of this message. One respondent expressed this confusion by saying that:

I think it is great that they are starting to share data but I don’t understand the point of differentiating the Bureau of Labor statistics from the Bureau of Labor for example I am not sure where the differentiation occurs statistically, there is a defined distinction, but I am not sure what that distinction is trying to accomplish.

This comment seems to be again related to the fact that the public does not think of the government in statistical/non-statistical terms, and that they do not understand the pragmatic reason for separating them (i.e. that statistical agencies can provide confidentiality whereas non-statistical ones cannot necessarily).

All in all, this message performed the worst out of the four tested in this round. Without significant wording revisions, it appears to be unworkable. However, by exposing the fact that the public does not necessarily see the government in statistical and non-statistical terms, it does show that the statistical agencies might need to dedicate more effort explaining how they are different from the rest of the federal government—or that they exist at all.

Question 11

Question 11 provided respondents with the statement, “The United States Government uses your information to evaluate programs and make them more efficient.” After hearing this message, respondents were asked the following structured follow-up probes: 1) *What do you think that*

statement would mean? 2) How does it make you feel? 3) Does it imply data sharing? 4) Does it imply efficiency? 5) How does all that make you feel? 6) Do you support this statement?

By far, this message elicited the greatest number of positive reactions from the respondents, particularly as applied to good governance and data driven decision-making.

Good Government: Many respondents believed that using data to make programs more efficient would lead to better decisions and better government. Some respondents focused on the “evaluate programs” part of the message, and interpreted it as saying that the government would use their data to see whether or not programs have been useful. As one respondent said, “This statement means that they want to know if their programs are accomplishing what they are supposed to accomplish.”

Besides simply evaluating programs, a number of respondents took this message to mean that the government would do a better job of evaluating the needs of the population and designing and funding programs to fit. For example, one respondent explained his thinking by saying:

It means that if 50% or a large portion of the population uses a program then the government would increase that program. It means to tailor programs and funding to how much it will be used.

Other respondents focused on the idea that their data could be used to determine what needs are. For instance, one respondent said, “Well, I think the information gathered is used to formulate policy and allocate funds, which is necessary.” This is a theme that emerged throughout the entire qualitative testing of the FSS POMS—that people want good, accurate data to be used to determine the needs of the public.

Data Driven Decisions: Respondents did not just *want* the government to use good quality, accurate data to determine the needs of the population, but they appeared to be ready to place a great deal of *trust* in the government if it did so. Using data to make business decisions is seen as normative, and it appears that respondents expect the same out of their government. The respondents who interpreted the statement this way believed that better data led to better services. Some respondents took this idea of data driven decisions a step further, combining it with the (unsaid in this message) concept of data sharing. For example, one respondent said:

I think government agencies can be very “siloes” and they can evaluate the one thing they’re supposed to do but what they are working on is all related, like health related to education related to workforce or income, all these things, so that sharing that information is key to being able to analyze these programs. You can’t just say that if 95% of the students in this program tested well, what does this mean? In a vacuum? I mean what is the point of teaching? The point of teaching is to become a good citizen of society but in order to measure those outcomes you need to know outcomes beyond education. So I think sharing is critical to evaluations.”

Other respondents emphasized the fact that they want the government to make decisions using all of the information they possibly can, with the belief that better decisions stem from analyzing

more data. This realization that data sharing will lead to data driven decision-making, and thus government efficiency is the root of what the Federal Statistical System is trying to get across with these messages. While it is undeniable that the public sometimes has an inflated opinion of what is possible to accomplish with databases and data mining, by understanding their “baseline” opinion, statistical agencies can adjust their messages to mesh as closely as possible.

Concern over Data Misuse: While much of the response to this statement was positive, some respondents believed that the message was implying adverse governmental actions. One respondent believed that the statement meant the government was going to make cuts. This in turn would make her feel insecure. Others brought politics into in, and worried about one group having access to all of this person data. For example, one respondent said:

I think this is where the politics can easily become part of these programs and the information could be used to justify what politicians would want, if it is really used non-politically it really could be very beneficial.

The main concern here seems to be the misuse of information to benefit a single person or a group (such as a political ideology). This fear fits in well with a number of comments to previous messages and questions throughout the qualitative testing of the FSS POMS that indicate that people trust the government *writ large*, but that they are wary of the actions of a single rogue employee or group of employees.

Confusion and Vagueness: In addition to this concern over the potential misuse of the data, there was some level of confusion evident. The great majority of this confusion stemmed from what the respondents thought was the “vagueness” of the message. A number of respondents wanted more details on which programs and which data specifically would be used in these evaluations.

Summary

While it is clear that the these messages, in their current form, are not designed to be used in the production FSS Public Opinion Monitoring Survey, this qualitative examination provides a number of themes, and concerns that ought to be considered in future work regarding administrative data use across the Federal Statistical System.

1. Efficiency: Respondents understood that these messages were suggesting new ways to achieve government efficiency, and overall this was viewed as a positive thing. Some messages, such as the first one, also suggested individual-level efficiency in the form of burden reduction, which was also viewed quite positively. Efficiency appears to be a key component behind the respondents’ desires for more data driven decision-making at the federal level. In particular, people wanted to know that the correct amount of time and resources are being allocated to those projects and programs that people use and that make an impact.
2. Discomfort: There was a noticeable amount of discomfort as related to data sharing, both within and outside of the government. This discomfort appears to come from a perceived risk of social and economic harm that could be a result of data breaches and the misuse of

data. Respondents particularly focused on their distrust not in the institutions of government, but rather on individuals—“rogue agents”—who could access and or disseminate their personal information for personal economic or political gain. In some sense, however, the discomfort with data sharing and administrative record use is mitigated by the “fait accompli” mindset wherein people realize that similar processes are already being used in both the public and private spheres.

3. Data Security: A key mitigating factor to the discomfort that respondents commented about regarding administrative record usage and data sharing is the security of that data. By and large, respondents wanted to know that their information was only being used for the purposes for which they consented. Because of their previously-mentioned worries about individuals breaking that trust, a large part of what the respondents wanted in terms of data security was to know who has access to their information and what they are doing with it.
4. Limitations on Data Sharing: The limitations to administrative record usage and data sharing implied in the middle two messages were interpreted by the respondents as adding to the security of their information, and were therefore seen in a positive light. Respondents appeared to draw distinct boundaries between the government or public sphere and corporations or the private sphere. They were much more comfortable with administrative record use and data sharing when these boundaries were seen as being respected, and were more uncomfortable when they perceived that this division was a risk of being violated.
5. Data Driven Decision Making: Respondents liked, and in some ways expected, the idea that the government makes its decisions based on accurate, unbiased data. The extensive use of data has become a fixture in the minds of the public. However inflated, using data to make decisions appeared to reduce the risk of the individual or political misuse of personal information in the eyes of the respondents.
6. Confusion: Two general types of confusion emerged. The first is simply related to fluency, as some respondents had a hard time with terms such as “entities” and “non-statistical.” The second, and more complex perhaps, type of confusion was over the processes behind data sharing. Across all four messages, some respondents indicated a desire to understand both how administrative records would be shared and why this was a positive or necessary action.
7. Understandings of the Government: An important cognitive issue that any future messaging that the Federal Statistical System creates in regards to data sharing is how the public view the government and mentally divide it. As seen in the confusion surrounding the third message (Question 10), the division between statistical and non-statistical agencies is not a natural one. Other qualitative work on the FSS POMS indicates that people tend to divide the government into potentially harmful and “safe” agencies (i.e. enforcement vs. programmatic agencies). While there is obviously a clear divide between the missions (and qualities, such as the ability to promise confidentiality) of

statistical agencies and the rest of the government, individuals do not appear to grasp it right away.