

1225-0059 – QCEW Survey

Attachment: Interview Protocol

A. Introduction

Thank you very much for assisting us in this effort.

BLS is considering changing its disclosure limitation method for QCEW data. We currently use cell suppression but we are considering other options. We are looking to understand how disclosure limitation efforts might impact [data users/data providers], such as yourself. Specifically, we're looking to learn about your attitudes toward QCEW's current disclosure limitation method and how that impacts you as a [data provider/data user] as well as your thoughts about what characteristics of data quality and confidentiality are important to you as we consider changing our disclosure limitation methodology.

Read to all participants

Your voluntary participation is requested to help BLS understand data provider and data user attitudes toward data confidentiality and data quality, in designing a long-term approach to publishing more useful data. The purpose of this call is for internal purposes only. We estimate it will take you 60 minutes to respond to this collection of information and this collection is authorized under OMB Number 1225-0059.

Information related to this study is confidential and will not be released to the public in any way that would allow identification of individuals or establishments. The BLS, its employees, agents, and partner statistical agencies, will use the information you provide for statistical purposes only and will hold the information in confidence to the full extent permitted by law. In accordance with the Confidential Information Protection and Statistical Efficiency Act of 2002 (Title 5 of Public Law 107-347) and other applicable Federal laws, your responses will not be disclosed in identifiable form without your informed consent.

Results of this project will be used to evaluate alternative disclosure limitation methodologies. Information discussed in our meeting will be strictly confidential. We will not be sharing this information with anyone outside the BLS. Our report will only summarize the findings without identifying the people who participated. The report will be available to you if you wish.

B. Permission to Record

May I have your permission to record this conversation? This will allow me to devote my full attention to the discussion. No one outside the project team will have access to the recording, and it will be deleted after the project is finished.

C. Interviewee Characteristics

Before we start, I'd like to learn a little bit about your background. It should help me to keep our conversation productive and relevant to you.

- What is your job title?
- What are your main responsibilities?

Thank you for that information.

Data User protocols (administer only to Data Users)

D. Background

Now I'd like to learn about how you currently use QCEW data.

- What is the primary reason you use the data?
 - What analyses do you conduct? For example, what kinds of tables do you use?
 - How critical is this analysis to your overall work?
 - How frequently do you use this analysis?
 - Would you say that others in your line of work also do similar analyses?
 - Are there any other reasons you use the QCEW data?
(repeat these follow-up questions as necessary for additional uses if the data)

E. Cell Suppression

Let's talk about the current way that we limit disclosure for QCEW - BLS is required by law to protect our respondents. We currently suppress the numbers in some table cells. By suppressing some cells, we are able to publish the reported values for many other cells.

- As a method of disclosure limitation, what do you think about cell suppression?
- How do the suppressed cells affect your analysis?
- Do you have any strategies for working around the suppressions?
- What kinds of analyses do you wish you could do with the QCEW dataset that you cannot currently do because of cell suppression?
 - Why do you want to do that analysis?
 - What is the obstacle?

F. Data Perturbation alternatives

Cell suppression is just one method for avoiding disclosure. BLS needs to continue implementing disclosure avoidance methods, but we do not necessarily have to continue using cell suppression. If there is compelling evidence that an alternative method is superior, BLS would consider alternatives.

- Have you seen any other disclosure limitation methods that you think might work for QCEW?
 - What do you see as the advantages of that method?
 - What would you say are the disadvantages?
- We have to consider all the ways that our published data may expose our respondents. The more we know about ways that our respondents are vulnerable, the better able we will be to protect them in a smart way, rather than just suppressing all the data. Can you think of any ways that people might try to misuse the QCEW data?

One method used by some agencies to avoid disclosure is adding noise to the data, such that the data to be published all undergo a process by which some amount of noise or fuzz is added. The data would be fuzzed but you would then have a full dataset of tables without holes. And, at the higher levels of aggregation, like state totals, the reported data would be published without noise. *If express skepticism: The amount of fuzzing can vary but, for now, let's assume that the level of fuzzing is acceptable to the average data user.*

- What do you think about noise methods?
- Do you have any concerns about this method? What are they?
- Would this method prevent you from doing any analyses you currently do?
 - How important to you are those analyses?
 - Are there other ways you could run these analyses?
- Would this method allow you to do any analyses that you cannot currently do?
 - How valuable would those analyses be to you?

Another disclosure limitation method involves replacing all of the values with average values, such that we could provide a microdataset that includes every single respondent but all of them protected – a synthetic dataset. The published values would not be the reported values but you would then have a full microdataset. The cell values might not be precisely the reported values but you could then make custom tables.

If express skepticism: The parameters for calculating the averages can vary but, for now, let's assume that the values are acceptable to the average data user.

- What do you think about synthetic datasets?
- Do you have any concerns about this method? What are they?
- Would this method prevent you from doing any analyses you currently do?
 - How important to you are those analyses?
 - Are there other ways you could run these analyses?
- Would this method allow you to do any analyses that you cannot currently do?
 - How valuable would those analyses be to you?

Data Provider protocols (administer only to Data Providers)

D. Background

Now I'd like to learn about your workplace and how you currently are involved in the QCEW.

- Are you the person at your workplace who reports for the QCEW?
 - *Ask if not:* What is your involvement at your workplace in reporting for the QCEW?
 - Is there anyone else at your workplace who is involved in reporting for the QCEW?
- Do you report for any other locations or just the one that you are at?
- And you told me earlier that you work at [company name]. What industry is [company name] in?

And could you tell me about your workplace's employment and wages? Just ballpark estimates would be fine. This will help me to place your views in context.

If hesitant, remind participant that data will be held confidentially.

- What is the number of employees?
If hesitant, offer: Fewer than 10? Between 10-100? Between 100-1000? More than 1000?
- What is the average wages at your workplace? Any figure will do – just one that you would say is representative of your workplaces’ wages.

E. Cell Suppression

Let’s talk about the current way that we limit disclosure for QCEW - BLS is required by law to protect our respondents. We currently suppress the numbers in some table cells. By suppressing some cells, we are able to protect the identities of respondents who might be identifiable based on their data.

- As a method of disclosure limitation, what do you think about cell suppression?
- Does cell suppression affect your company or individuals at your company in any way?
- Are you concerned that intruders could identify your data from BLS published tables?
- Do you feel that cell suppression effectively protects your identity?
 - What makes cell suppression [effective/ineffective]?

F. Data Perturbation alternatives

Cell suppression is just one method for avoiding disclosure. BLS needs to continue implementing disclosure avoidance methods, but we do not necessarily have to continue using cell suppression. If there is compelling evidence that an alternative method is superior, BLS would consider alternatives.

- Have you seen any other disclosure limitation methods that you think might work for QCEW?
 - What do you see as the advantages of that method?
 - What would you say are the disadvantages?
- We have to consider all the ways that our published data may expose our respondents. The more we know about ways that our respondents are vulnerable, the better able we will be to protect them in a smart way. Can you think of any ways that people might try to misuse the data that you and others provide for the QCEW?

One method used by some agencies to avoid disclosure is adding noise to the data, such that the reported values all undergo a process by which some amount of noise or fuzz is added.

Everyone’s data would be published, but the reported values would be fuzzed so your reported value would be protected. At the higher levels of aggregation, like state totals, the reported data would be published without noise.

If express skepticism: The amount of fuzzing can vary but, for now, let’s assume that the level of fuzzing is acceptable to the average data provider.

- What do you think about noise methods?
- Do you have any concerns? What are they?
- Would this method affect your company or individuals at your company in any way?
- Are you concerned that intruders could identify your data from BLS these tables?
- Do you feel that noise methods would effectively protect your identity?
 - What makes noise [effective/ineffective]?

Another disclosure limitation method involves replacing all of the values with average values, such that we could provide a microdataset that includes every single respondent but all of them protected – a synthetic dataset. The published values would not be the precisely reported values but the data users would have a full microdataset with which they could make custom tables.

Data users could produce a value that they think is very close to your reported value.

If express skepticism: The parameters for calculating the averages can vary but, for now, let's assume that the values are acceptable to the average data provider.

- What do you think about synthetic datasets?
- Do you have any concerns about this method? What are they?
- Would this method affect your company or individuals at your company in any way?
- Are you concerned that intruders could identify your data from this BLS dataset?
- Do you feel that this method would effectively protect your identity?
 - What makes synthetic datasets [effective/ineffective]?

Closing (administer to all participants)

G. Discussion

I've asked you questions about disclosure limitation methods and how they might affect you. Is there anything else on the topic that you would like to share?

Do you have any thoughts generally about balancing data quality and data confidentiality?

H. Thank you

That is all the questions that I have. Thank you very much for your time. We appreciate your input.