

## Manuscript Title

Data Collection Mode Effect on Abortion Questions: A Comparison of Face-To-Face and Web Surveys

## Author(s)

Mingnan Liu

Michigan Program in Survey Methodology, Institute for Social Research, University of Michigan, USA

## Summary

*The manuscript was received on October 10, 2017 and was peer reviewed by three reviewers and an editor.*

*The initial recommendation of Major Revision was made on November 16, 2017.*

*The first revision was submitted on January 06, 2018 and was re-evaluated by the editor.*

*The manuscript was accepted for publication on January 30, 2018.*

## Peer Review Comments

### Peer Reviewer 1

This paper deals with an important issue in terms of how different methods of data collection could potentially affect the opinions and attitudes elicited from respondents. The paper makes reference to the heated debates concerning abortion rights and attitudes to abortion in the US and globally. This paper brought to mind the question of how and why certain kinds of “knowledge” come to be valued or ignored in debates concerning issues as politically charged as abortion attitudes, and how and why different kinds of information feed into policy making processes. Whilst the authors hint at this, the paper feels very limited in scope and by broadening their focus they can situate their findings within the wider discussions concerning how different forms of evidence inform public narratives and decision making in the field of abortion and abortion rights.

### Abstract

The abstract presents an accurate synopsis of the paper informing the reader that the study aims to assess the data collection mode effects (face-to-face vs. web surveys) on abortion-related attitudinal questions, using the 2012 American National Election Studies.

### Introduction and Aims

The authors make a general statement on page 1, line 3 that attitudes to abortion in the US are polarized. The paper would benefit from further discussion of the idea that the debate in the US about abortion is polarised. Given the heated debates over term limits and the conditions under

which different groups feel that abortion should be allowed, the use of the term polarised may not capture the nuances within the abortion debate. Further exploration of this topic would be useful.

On page 1, line 4 the authors note that studies on attitudes to abortion almost exclusively depend on survey data. Qualitative research and research based analyses of media narratives and policy documentation have provided useful insights into attitudes to abortion, though it may be that large sample surveys feature more prominently in public debate due to the public assumption that they are more objective than qualitative research. This is a topic that the researchers could explore in more depth.

*“A solid understanding of mode effect on abortion attitudes can also inform clinicians on choosing the appropriate approach to communicate with patients about abortion. It can also have a profound policy making impact when policies are informed by findings of public opinion survey research”.*

The above needs referenced and could be discussed in greater detail.

### **Methods and Results**

The authors used a t-test and chi square to compare proportions and distribution of responses. The paper makes good use of secondary data but I would question if there is enough evidence presented to conclude that ‘mode’ explains the differences in response. Perhaps different statistical tests could be used. It may be useful to include a logistic regression analysis with item non-response as the dependent variable. Is mode still statistically significant when it interacts with other demographic and attitudinal variables?

If the authors used analysis such as logistic regression they could explore in more depth the somewhat contradictory findings.

### **Discussion and Conclusion**

Within the discussion the authors point to the limitations of their study and most importantly they note the limited range of options in the item scales and how this could affect the conclusions we can draw from their analysis. The discussion could be better grounded within the existing literature on this topic and some of the statements would need further clarification, referencing and discussion. It would be useful to include a discussion of the long running debates concerning feminist research and quantitative methods, positivism, interpretivism and recent efforts to integrate feminist theory and quantitative methods. A section discussing the relationships between researcher and researched in terms of power dynamics and how the interplay between different research modes and researcher-researched relationships would be useful.

Recommendation: Major Revision. However, this is an important field of research as it touches upon the relationship between researcher and researched and the idea of objective knowledge concerning attitudes to abortion. With further statistical analysis and more engagement with the literature the paper has the potential to make an important contribution to discussions on the methodologies in abortion related research.

### **Peer Reviewer 2**

#### **General comments**

This is an interesting and understudied aspect of abortion research and I applaud the authors for their endeavor. I think additional efforts to convince the reader that observed mode differences are not a by-product of unmeasured differences across the samples would strengthen the paper, as would additional details in many of the article sections (see specific comments below). Addressing these issues would improve the article, but I am not certain the article can overcome some of the stated concerns.

### **Major comments**

The article could really benefit from expanding the background section, specifically with regard to prior literature on mode effects (e.g. Tourangeau and Yan 2007). Previous research has been done specifically in the area of abortion as well, primarily measuring abortion experience, not opinion, but still salient (Lensvelt-Mulders et al 2005; Jones and Forres 1992; Langaug et al 2011; Lara et al 2004). Additionally, no prior estimates related to public opinion on abortion are presented. Need to contextualize the estimates presented in results section, and need to at least briefly make a case for why we care about this (e.g. public health reasons?, public policy reasons?, because interesting?, etc.).

The response rates (38% and 2%) are incredibly low, particularly for the web survey. It is hard to place credibility in the comparability and representativeness of the results. Is this response rate deemed acceptable/normal for web-based surveys? Briefly address this concern in limitations but this seems to be a substantial issue; would be good to discuss potential biases.

Would like actual results with specific numbers referenced in the text of the results section. There is no need to write out everything in the tables, but it is helpful to highlight key findings in the context of the numeric results; results text currently contains no numbers. Please re-write.

The results section contains interpretation of the findings, which should be in the discussion section. Please remove.

The data and associated results at the two time points are never presented separately. Is there a reason both time points were collected? And if not analyzing the data longitudinally in any way, need to explain why not and why combining the data make sense. I believe the only reference to the trends is when the authors write in the last sentence of the results, "The trends of the level of favoring/opposing across all the items for both genders are also similar to the combined whole samples" but the trends overall are never presented, let alone by gender.

With regard to the analyses, I think it would make a stronger case that the observed differences are a result of the mode if multivariate analyses were done; I am not confident weights (and possibly not even multivariate analyses) can account for the potential sample differences as a result of unmeasured characteristics. This is especially true given likely differences in people who might take part in the two surveys, which could confound the relationship between the mode and the abortion attitudes of interest. In particular, people participating in a web *panel* may be even more different from people in a single face-to-face survey than a one-off web survey. Alternatively, the authors could use propensity scores to try to isolate the effect of the survey mode apart from differences in sample characteristics.

### **Minor comments**

In abstract, there is a typo in the second sentence; "face-to-face" is "fact-to-face".

In the first sentence of the introduction, “overtime” should be “over time”.

Not clear what “prenatal testing preference” refers to in relation to the Singer & Couper findings. Please explain further.

Please provide some additional details on the GfK KnowledgePanel sampling; reader needs more information in methods description to determine quality of that design and associated data. Also, what do you mean by “respondents for the 2012 ANES were selected from this probability panel”? How were they selected from the main GfK KnowledgePanel? What were the inclusion/exclusion criteria?

For the description of the face-to-face survey, I would also like a little more detail on the sampling in the text of the article. Were the strata urban/rural? And what were the stages in the multi-stage cluster sample? State, county, census tract? And the description seems incorrect; you would use an address based stratified multi-stage cluster sampling design “to select” the 125 census tracts, not “in” the 125 census tracts. And was the selection of the census tracts done using probability proportional to size or randomly?

Can you provide the reasoning behind why you randomized the first 7 questions and always had the women’s choice question last? And why was the response option “don’t know” provided in the face-to-face interview but not in the web survey?

The authors need to describe more detail (probably in the introduction) on the 8 attitudinal questions. It is not until reviewing Table 2 that I even knew what the questions were about, and even then I didn’t know what the actual question wording was. It would be helpful to also reference the appendix early on in the article, or even include this as the first table in the article.

Would prefer if authors present actual p-value instead of  $X^2$  value in the tables.

### **Peer Reviewer 3**

This paper attempts to compare the effect of mode of survey administration on responses to abortion-related attitudinal questions using the same instrument, administered in parallel through both modes (web-based and face-to-face). The topic of mode of survey administration on abortion related responses is worthy of investigation, and might be relevant to this journal depending on the presentation of findings, however, the paper fails to ground the analysis in the existing literature, present a coherent hypothesis, and to conduct an appropriate analyses of the existing data that would contribute anything new to the field.

Unfortunately, while the introduction describes the potential challenge of social desirability bias, and introduces the reader to the web-based survey literature, the authors fail to draw upon the extensive literature exploring ACASI or CASI survey administration. The introduction presents some information that should be presented in the methods section, but fails to present a coherent hypothesis and rationale for the analysis.

The methods for survey administration are fairly described, however, the statistical approach is simplistic and does not take advantage of the existing data to examine anything beyond gender as a potential effect modifier of the association between mode of administration and response to abortion-related items. The dataset is rich and could be used to explore anything from predictors of non-response to variation in response itself. The authors do not describe any potential differences

between the two populations (face-to-face or web-based) and do not address any potential issues of confounding. Given the simplistic analysis, the results presented do little to add nuance to our understanding of mode of survey administration on item response in abortion-related attitudinal questions, and, I would argue, in fact, the authors conclusion that differences do, in fact, exist, do not acknowledge the very preliminary nature of the analysis.