Web Survey Data Scrutinizes Administrative Big Data: The Hidden Pitfalls of Statistical Inference from Korean Hypertension and Diabetes Health Insurance Data

Survey & Health Policy Research Center

Technical Report

August 2023

Web Survey Data Scrutinizes Administrative Big Data: The Hidden Pitfalls of Statistical Inference from Korean Hypertension and Diabetes Health Insurance Data

Jaehoon Kim[1] and Sunwoong Kim [2]

[1]Survey and Health Policy Research Center, Dongguk University, Seoul, South Korea
[2]Survey and Health Policy Research Center and Department of Statistics, Dongguk University, Seoul, South Korea

Corresponding Author:
Sunwoong Kim, Dongguk University, 30 Pildong-ro, 1-gil, Jung-gu, Seoul, South Korea 04620
Email: sunwk@dongguk.edu

Abstract
The recent rise in research interest in big data has inspired many researchers in South Korea to use administrative data to obtain more insightful national statistics. In response to this change, some public institutions and public enterprises have spontaneously increased the use of administrative data for statistical purposes and disclosed administrative data to the public for research purposes. Among these, the Korean National Health Insurance Service (KNHIS), which is responsible for the mandatory insurance system offering healthcare to all citizens of about 52,000,000, has released numerous national health statistics based on the insurance demands data. Also, it recently started the National Health Insurance Data Sharing Service (NHIDSS) for big data samples. The NHIDSS made nearly 20 years of health insurance database accessible to the public via the web. Solely relying on national health insurance data and treating it like de facto census data or representative sample data, however, may result in significantly biased estimates depending on what population statistic the researcher hopes to estimate. This paper addresses the undercoverage problem, which is rarely given due consideration for big data-related research. The undercoverage in this study is defined as the entire adult population which does not visit a medical institution despite having hypertension and diabetes for a given period (e.g., a year). Such a population is excluded from the national health insurance claims data, and this exclusion may result in non-coverage bias. We estimate the total number (or rate) of adults not covered among those diagnosed with hypertension and similarly among those diagnosed with diabetes. For this, we compared the survey statistics obtained from the data collected through the RDD smartphone web survey (self-administered survey) conducted by Kim and Kim (2023) as well as the Korea National Health and Nutrition Examination Survey and the Korean Community Health Survey (interviewer-administered CAPI surveys) with insurance statistics produced from national health insurance claims data. We found out that insurance statistics from such big data may suffer from serious undercoverage problems. In particular, it was estimated, based on survey results from the RDD

smartphone web survey, that about 4.8 million (41%) were not covered among about 11 million adults diagnosed with hypertension, and about 1.7 million (34%) were not covered among about 5 million diagnosed with diabetes in 2020. This implies that using a big data sample from the NHIDSS could inherit the same problem, even assuming that the big data sample is representative of the entire big data. Moreover, this would demonstrate the large potential of self-administered modes such as the RDD smartphone web surveys, which is a cost-effective mode, in evaluating the coverage bias of big data.

Keywords
administrative data, big data, health insurance claims data, diabetes, hypertension, undercoverage, coverage bias, smartphone web survey, CAPI, self-administered survey, interviewer-administered survey

Introduction

Decreasing response rates and increasing implementation costs in sample surveys have been a troublesome ongoing global trend for survey researchers (e.g., see Heer 1999; AAPOR, 2019; Groves & Harris-Kojetin, 2017, pp. 23-27). In response, researchers have turned to relatively cheaper data sources to supplement or replace survey data, one prominent source being administrative data, a distinctive form of big data. Traditionally, as administrative data is derived from the operation of administrative systems, such as those used in healthcare, taxation, vehicle licensing, etc., instead of a specific research design, it tends to be very messy, demanding costly cleaning effort before usage. In recent years, however, accessibility has been improved to the point that more researchers have attempted to utilize administrative data by establishing linkages between administrative data and survey data (Datta, Ugarte, & Resnick, 2020; Song, & Thomas, 2020).

Despite such accessibility, administrative data is often self-selective and incomplete. It may also contain measurement errors, reporting errors, problems with record matching, and many others. These multiple sources of problem or error in administrative data need to be systematically separated in a similar manner to a “total survey error (TSE)” framework, which identifies all the major sources of error in surveys as measurement error, processing error, coverage error, sampling error, nonresponse error, and adjustment error. For this, a “big data total error (BDTE)” framework, which distinguishes the errors similar to those in surveys plus additional errors unique to big data, can be applied to administrative data. It will aid in our understanding of the limitations of administrative data, leading to better-informed analyses and applications of the results (AAPOR, 2015; Connelly et al., 2016).

Among various errors inherent to administrative data with respect to the BDTE framework, the potential for coverage error from undercoverage attributable to the data definition and generation process may be a primary cause for concern from the perspective of a survey statistician. This is contrary to the idea that as Connelly et al. (2016, p8) note citing Groen (2012), administrative data are generally less likely to contain coverage error because they are not based on intermittently collected samples. Witnessing the sheer size of administrative data and lacking guidance from the data provider, the uninformed might unintentionally ignore the possibility of undercoverage, overestimating the representativeness of the data and failing to make adjustments.

In this study, we demonstrate such a possibility in the context of using the national health insurance data (national health insurance claims data), one of the biggest individual-level public administrative data in South Korea. To investigate the level of undercoverage and coverage bias from using the national health insurance data, we selected and studied hypertension and diabetes, the most common chronic diseases. We focus on estimating the total number (or rate) of adults not covered among those diagnosed with hypertension and similarly among those diagnosed with diabetes. For this, we compare the survey statistics obtained from the data collected through the RDD smartphone web survey (self-administered survey) conducted by Kim and Kim (2023) as well as the Korea National Health and Nutrition Examination Survey and the Korean Community Health Survey (interviewer-administered CAPI surveys) with insurance statistics produced from national health insurance claims data.

National Health Insurance Data

Many academic and nonacademic researchers in South Korea have been recently encouraged to use administrative data to obtain more insightful national health statistics. In response to this change, some public institutions and public enterprises have spontaneously increased the use of administrative data for statistical purposes and disclosed administrative data to the public for research purposes. Among these, the Korean National Health Insurance Service (KNHIS), which is responsible for the mandatory insurance system offering healthcare to all citizens of about 52,000,000, has released numerous national health statistics based on the insurance claims data. Also, it recently started the National Health Insurance Data Sharing Service (NHIDSS) for big data samples. The NHIDSS made nearly 20 years of health insurance database accessible to the public via the web, with potentially sensitive individual data being given restricted access with approval. Solely relying on national health insurance data and treating it like de facto census data or representative sample data, however, may result in significantly biased estimates depending on what population statistic the researcher hopes to estimate, as presented later. No matter the size of the insurance claims data sets available, one must always take into consideration the circumstances surrounding the data collection and remain vigilant against various sources of errors, especially coverage error, with respect to the BDTE framework.

Hypertension and Diabetes

Hypertension and diabetes are chronic diseases increasingly prevalent in developed countries including South Korea. They are common comorbidities and ongoing treatment can result in substantial costs over the course of a lifetime. To make matters worse, they are leading independent risk factors for cardiovascular disease (CVD), thus the likelihood of medical complications later in a person’s life increases when these diseases are left untreated. They are likely to lead to higher future healthcare costs down the road (Wang et al., 2017). This is especially a great concern for countries such as South Korea whose budgets are increasingly constrained by growing aging populations and the consequent reduction in tax revenues (World Health Organization, 2022). Naturally, managing these diseases garners great interest when developing public health policy. Therefore, it is important to provide accurate and well-presented factual reports or documents to raise public awareness, promote prevention and timely treatment, or be of practical assistance in developing policies around them.

Description of National Surveys

In this study, we used three surveys conducted at the national level in 2020 to find out the level of undercoverage and coverage bias in using the national health insurance data. The two surveys called the Korea Community Health Survey (KCHS) and the Korea National Health and Nutrition Examination Survey (KNHANES) each are conducted by the Korea Disease Control and Prevention Agency (KDCPA) and are interviewer-administered CAPI surveys. The third survey called the National Survey of Life and Health (NSLH) is a self-administered smartphone web survey conducted by Survey and Health Policy Research Center (SHPRC) at Dongguk University.

We briefly describe each survey data. We refer the reader to Kim & Kim (2023) for a much-detailed explanation of the technical details, sample design, and data collection methodology of KCHS and NSLH.

The KCHS is a community-based large annual survey covering the adult population in households for the purpose of gathering information that could be used to plan, implement, monitor, and evaluate community health promotion and disease prevention programs. This survey was jointly conducted by 255 community health centers located in cities and counties across the country in cooperation with universities within the communities. A sample of households was selected by stratified two-stage cluster sampling in each community. It ran for 11 weeks from August 16 to October 31.

The NSLH used a sample of cellphone numbers selected by using an unstratified and unclustered single-stage equal probability of selection method (EPSEM) from a cellphone RDD frame covering Korean adults. The sample numbers were contacted by an SMS text message with a link to a web survey. Data collection with enough follow-up reminders lasted for a total of 7 weeks from October 12 to November 28, which partially overlapped with the KCHS (Kim & Kim, 2023).

The KNHANES consists of a health interview survey, a health behavior survey, a nutrition survey, and a health examination. It intends to cover residents aged 1 year or older for the purpose of monitoring the health of the Korean population through the collection and analysis of data on a broad range of health and nutritional characteristics categorized by many demographic and socioeconomic characteristics. A sample of households was selected by stratified two-stage cluster sampling. It was conducted throughout the whole year.

Undercoverage and Coverage bias

The undercoverage in this study is defined as the entire adult population which does not visit a medical institution for a year despite having hypertension or diabetes. Such a population is excluded from the national health insurance claims data, and this exclusion may result in coverage bias, which is expressed as the product of the undercoverage rate and the statistical difference between the covered and non-covered persons. See Groves et al. (2009, pp. 54–56) for the details. The undercoverage problem always occurs when analyzing national health insurance claims data since there exists a certain number of individuals with hypertension or diabetes who do not visit a medical institution. The resulting coverage bias highly depends on the undercoverage rate. The higher the undercoverage rate, the higher the coverage bias.

Data Analysis for Responses to Sensitive Questions

For the NSLH and the KCHS, we analyzed the raw data using survey weights for responses to the two questions "Have you ever been diagnosed with high blood pressure by a doctor? (Yes, No)" and "Have you ever been diagnosed with diabetes by a doctor? (Yes, No)" to estimate the proportion of individuals diagnosed with hypertension or diabetes among a total of 43,526,824 adults. It should be noted that these questions are sensitive topics, and the results of the survey may differ depending on whether or not an interviewer is present. Sensitive questions address highly personal topics, including substance use, sexuality, delinquency, victimization, health, income, and voting habits. Surveys on issues conventionally perceived as sensitive tend to benefit from a switch to modern technologies; particularly when respondents are interviewed alone without the presence of interviewers such as in web-based surveys (Chang & Krosnick, 2009, 2010; Ye et al., 2011; Gnambs & Kaspar, 2015). Considering these aspects, it is expected that since hypertension or diabetes is a personal and socially undesirable attribute, the respondents asked by an interviewer would underreport. Accordingly, the estimated proportion of individuals diagnosed with hypertension or diabetes in the KCHS (interviewer-administered mode) would be lower than in the NSLH (self-administered web mode), whose accuracy and quality were demonstrated by Kim & Kim (2023).

Results from KNHANES

The survey results from the KNHANES are available in the report titled '2020 National Health Statistics' published by the Korea Disease Control and Prevention Agency (2022). We referenced the estimated prevalence rates (the proportion of individuals with hypertension or diabetes) obtained through a health examination, regardless of self-awareness of hypertension or diabetes. The prevalence rate of hypertension was 29.0% with a margin of error of 1.8 percent at a 95 percent level of confidence. Of these, 69.8% with a margin of error of 1.8 percent were aware that they had hypertension (ever diagnosed by a doctor), and 65.2% with a margin of error of 2.0 were receiving treatment. The prevalence rate of diabetes was 13.9% with a margin of error of 1.2 percent at the same level of confidence. Of these, 65.1% with a margin of error of 2.7 percent were aware that they had diabetes (ever diagnosed by a doctor), and 60.6% with a margin of error of 2.9 were receiving treatment.

Health Insurance Statistics

The Health Insurance Review and Assessment Service (HIRA) and the KNHIS jointly provide various reports to the public online. Of these reports, the following are the most widely known: The Healthy Lifestyle Information and the National Health Insurance Statistical Yearbook. The number of patients in these reports is defined as the number of patients who received healthcare service from providers in a given year adjusted for duplication. According to the National Health Insurance Statistical Yearbook in 2020, there were 6,733,000 hypertension patients and 3,344,000 diabetes patients who received healthcare services in 2020 (see Health Insurance Review and Assessment Service, 2021). These numbers include nonadults less than 1 percent of those numbers in each of hypertension and diabetes, but cannot be separated.

Results

Undercoverage in National Health Insurance Data

As we mentioned above, we expected that since hypertension or diabetes is a personal and socially undesirable attribute, the respondents asked by an interviewer would underreport, resulting in a lower estimated proportion of adults diagnosed with hypertension and diabetes in the KCHS (interviewer-administered mode) and a higher estimated proportion in the NSLH (self-administered web mode). The results were obtained as expected. The estimated proportion of those diagnosed with hypertension in the KCHS was 21.2% with a margin of error of 0.4 percent at a 95 percent level of confidence, whereas the one in the NSLH was 26.4% with a margin of error of 3.7 percent at the same level of confidence. The two estimates differed significantly ($p<0.01$) between the two surveys. While the estimated proportion of those diagnosed with diabetes in the KCHS was 9.1% with a margin of error of 0.2 percent, the one in the NSLH was 11.7% with a margin of error of 2.7 percent. The estimates showed significant differences ($p<0.05$) likewise.

We completed Table 1 by using the survey results from the KNHANES (prevalence of hypertension, 29.0%; prevalence of diabetes, 13.9%), estimated proportions of diagnosed adults in the NSLH and the KCHS, and 6,733,000 hypertension patients and 3,344,000 diabetes patients in health insurance statistics from administrative data, and a total number of adults (43,526,824) in the nation. Each number at the first line for each disease in the table was obtained by 43,526,824 multiplied by each estimated proportion in KNHANES, NSLH, and KCHS. The estimated undercoverage is the difference between the number in NSLH and the one in administrative data.

The table shows the number of adults with hypertension and diabetes in descending order from the KNHANES to administrative data. The number from the KNHANES is the highest because it included those unaware that they had hypertension (ever not diagnosed by a doctor). The undercoverage for those diagnosed with hypertension according to the use of the national health insurance data was estimated as 4,758,082 with a number and 41% with a rate. It means that when analyzing using national health insurance data, such a very large number or rate of individuals cannot be covered among those diagnosed with hypertension. The undercoverage for those with diabetes was estimated as 1,748,638 with a number and 34% with a rate.

Although these numbers are less than those for hypertension, they are very large numbers likewise. Clearly, so many hypertension and diabetes patients diagnosed by a doctor do not visit a medical institution, and cannot be covered by the administrative data. Thus, the results of the analysis of administrative data do not reflect a significant portion of the entire adult population diagnosed with hypertension and diabetes, inevitably resulting in serious bias. A similar bias would also appear when using big data samples provided by the NHIDSS.

Table 1. Estimated Undercoverage for Individuals Diagnosed with Hypertension and Diabetes in National Health Insurance Data

Hypertension	KNHANES	NSLH	KCHS	Administrative Data
Number of patients	12,622,779 (110%)	11,491,082 (100%)	9,227,687 (80%)	6,733,000 (59%)
Estimated undercoverage				4,758,082 (41%)
Diabetes	**KNHANES**	**NSLH**	**KCHS**	**Administrative Data**
Number of patients	6,050,229 (119%)	5,092,638 (100%)	3,960,914 (78%)	3,344,000 (66%)
Estimated undercoverage				1,748,638 (34%)

Note. Each number in KNHANES, NSLH, and KCHS was obtained by 43,526,824 multiplied by the estimated proportion. The estimated undercoverage is calculated by the number in NSLH minus the one in administrative data.

Undercoverage for Subpopulations

We compared the undercoverage between subpopulations (gender and age groups) according to the use of the national health insurance data. In Table 2 and Table 3, the numbers of patients in the administrative data came from the National Interest Disease Statistic in the HIRA Bigdata Open Portal (https://opendata.hira.or.kr/home.do), which is a different source for national health insurance data. The numbers of patients in the NSLH were estimated by subpopulation. As shown in Table 2, the undercoverage rate of the male group is 10% and 13% higher than the female group for both hypertension and diabetes, respectively (42% versus 32 and 36% versus 23%). As given in Table 3, the younger age groups had the higher undercoverage rate for both hypertension and diabetes (75%, 40%, 32%, and 50%, 30%, 28%). It is noted that the youngest age group (20-39) had a much higher undercoverage rate than other age groups for both hypertension and diabetes.

Table 2. Estimated Undercoverage by Gender Groups for Individuals Diagnosed with Hypertension and Diabetes in National Health Insurance Data

Hypertension	NSLH		Administrative Data	
	Male	Female	Male	Female
Number of patients	6,260,862 (100%)	5,216,224 (100%)	3,605,215 (58%)	3,549,341 (68%)
Estimated undercoverage			2,655,647 (42%)	1,666,883 (32%)
Diabetes	**NSLH**		**Administrative Data**	
	Male	Female	Male	Female
Number of patients	3,058,062 (100%)	2,014,179 (100%)	1,970,879 (64%)	1,559,025 (77%)
Estimated undercoverage			1,087,183 (36%)	455,154 (23%)

Note. The estimated undercoverage is calculated by the number in NSLH minus the one in administrative data. The total number of patients in administrative data is slightly different from the one in Table 1 because the source is different from each other.

Table 3. Estimated Undercoverage by Age Groups for Individuals Diagnosed with Hypertension and Diabetes in National Health Insurance Data

Hypertension	NSLH			Administrative Data		
	19-39	40-59	60 or above	20-39	40-59	60 or above
Number of patients	949,965 (100%)	4,299,830 (100%)	6,227,296 (100%)	235,417 (25%)	2,575,288 (60%)	4,343,851 (70%)
Estimated undercoverage				714,548 (75%)	1,724,542 (40%)	1,883,445 (32%)
Diabetes	**NSLH**			**Administrative Data**		
	19-39	40-59	60 or above	20-39	40-59	60 or above
Number of patients	315,895 (100%)	1,771,536 (100%)	2,984,815 (100%)	156,635 (50%)	1,235,187 (70%)	2,138,258 (72%)
Estimated undercoverage				159,260 (50%)	536,349 (30%)	846,557 (28%)

Note. The estimated undercoverage is calculated by the number in NSLH minus the one in administrative data. The first age group in administrative data does not include the age of 19. The total number of patients in administrative data is slightly different from the one in Table 1 because the source is different from each other.

Expected Coverage Bias

The bias related to undercoverage is called coverage bias. As described above, coverage bias can be expressed as the product of the undercoverage rate and the statistical difference between the covered and non-covered persons. Since undercoverage rates for individuals diagnosed with hypertension and diabetes are very high (41% for hypertension, and 34% for diabetes), the coverage bias would be high regardless of the statistical difference between the covered and non-covered persons. The exact coverage bias cannot be addressed because we do not have any information on the statistical difference between the covered and non-covered persons. But an appropriate example can be given. Assume that the average annual medical expense of covered persons is 10,000 dollars, and the one of non-covered people is 1,000 dollars. In this case, the statistical difference is 9,000 dollars. The coverage bias for the average annual medical expense of all persons is 3,726 dollars (=9,000·0.414) when the undercoverage rate for hypertension is applied. This large coverage error can occur analyzing administrative data.

Discussion

In this paper, using survey data from the NSLH, which is an RDD smartphone web survey, conducted by Kim & Kim (2023) as well as those from KCHS and KNHANES conducted by a government, we demonstrated that there exists a very high national level undercoverage of both diabetics and hypertension patients in the national health insurance data due to those patients who did not visit the hospital for treatment. We specifically showed that there can be a large coverage bias because of such an undercoverage. We also presented that the undercoverage rate of the male group is higher than the female group and the younger age group is a higher undercoverage rate for both hypertension and diabetes. Therefore, any statistical inference at the national level or subpopulation level drawn from the national health insurance data or a big data sample from the NHIDSS is highly likely to be biased. This implies one should be careful in analyzing the national health insurance data or reporting the analysis results. It would be best for data providers to mention the issue of undercoverage in a report or a statement on the use of big data. On the other hand, this study demonstrated the large potential of self-administered modes such as the RDD smartphone web surveys, which is a cost-effective mode, in evaluating the coverage bias of big data.

Author Biographies

Jaehoon Kim is a research associate at the Survey & Health Policy Research Center, Dongguk University. He is currently involved in several research projects of survey and data science methodology. E-mail: kimx4591@umn.edu

Sunwoong Kim is a professor at the Department of Statistics and the director of the Survey & Health Policy Research Center, Dongguk University. E-mail: sunwk@dongguk.edu

Data Availability

The anonymized data for the smartphone web survey are available from an author at a given email address. The reports and microdata for the CAPI survey can be downloaded by request from the KCHS website (https://chs.kdca.go.kr/chs/main.do).

Software Information

The analyses were conducted using SAS 9.4 and R.

Funding

The SHPRC's own funds were used for this study.

References

American Association for Public Opinion Research (2015). *AAPOR report on big data*. https://aapor.org/wp-content/uploads/2022/11/BigDataTaskForceReport_FINAL_2_12_15_b.pdf

American Association for Public Opinion Research (2019). *Report of the AAPOR task force on transitions from telephone surveys to self-administered and mixed-mode surveys*. https://aapor.org/wp-content/uploads/2022/11/Report-of-the-Task-Force-on-Transitions-from-Telephone-Surveys-FULL-REPORT-FINAL.pdf

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, 73, 641-78.

Chang, L., & Krosnick, J. A. (2010). Comparing oral interviewing with self-administered computerized questionnaires: An experiment. *Public Opinion Quarterly*, 74, 154-167.

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*, 1-12.

Datta, A.R., Ugarte, G., & Resnick, D. (2020). Linking survey data with commercial or administrative data for data quality assessment. In C.A. Hill, P.P. Biemer, T.D. Buskirk, L. Japec, A. Kirchner, S. Kolenikov, & L.E. Lyberg (Eds.), *Big Data Meets Survey Science*. (pp. 99-129). Wiley.

Gnambs, T., & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: A meta-analysis. *Behavioral Research Methods*, 47(4), 1237-1259.

Groen, J. A. (2012). Sources of error in survey and administrative data: The importance of reporting procedures. *Journal of Official Statistics*, *28*(2), 173-198.

Groves, R.M., & Harris-Kojetin, B.A. (2017). *Innovations in federal statistics: Combining data sources while protecting privacy*, The National Academies Press.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. John Wiley & Sons.

Health Insurance Review and Assessment Service (2021). *2020 National Health Insurance Statistical Yearbook*. https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA020045020000&brdScnBltNo=4&brdBltNo=2313&pageIndex=1

Heer, W. D., (1999). International response trends: Results of an international survey. *Journal of Official Statistics*, 15(2), 129-142.

Kim S.W., & Kim, J.H. (2023). A comparison of accuracy using administrative data between national smartphone web and CAPI surveys in the age of mobile technology and COVID-19. *Dongguk University Survey & Health Policy Research Center Technical Report*. https://shprc.dongguk.edu/article/reports/detail/178482?pageIndex=1&

Korea Disease Control and Prevention Agency (2022). *2020 National Health Statistics*. https://knhanes.kdca.go.kr/knhanes/sub04/sub04_04_01.do

Song, Xi, & Coleman, T. (2020). Using administrative big data to solve problems in social science and policy research. *University of Pennsylvania Population Center Working Paper* (PSC/PARC), 2020-58. https://repository.upenn.edu/psc_publications/58.

World Health Organization (2022). *Global spending on health: Rising to the pandemic's challenges*. https://www.who.int/publications/i/item/9789240064911

Wang, G., Zhou, X., Zhuo, X., & Zhang, P. (2017). Annual total medical expenditures associated with hypertension by diabetes status in U.S. adults. *American Journal of Preventive Medicine*, 53(6), S182-S189. https://doi.org/10.1016/j.amepre.2017.07.018

Ye, C., Fullton, J., & Tourangeau, R. (2011). More positive or more extreme?: A meta-analysis of mode differences in response choice. *Public Opinion Quarterly*, 75, 349- 365.