

WHO WATCHES PRESIDENTIAL DEBATES? MEASUREMENT PROBLEMS IN CAMPAIGN EFFECTS RESEARCH

MARKUS PRIOR*

Abstract To examine whether a campaign event affected candidate preferences or candidate knowledge, survey researchers need to know who was exposed to the event. Using the example of presidential debates, this study shows that survey respondents do not accurately report their exposure to even the most salient campaign events. Two independent methods are used to assess the validity of self-reported debate exposure. First, survey estimates are compared to Nielsen estimates, which track exposure automatically. Second, the temporal stability of self-reports across independent daily estimates in the National Annenberg Election Survey is analyzed. Both approaches indicate low validity of self-reports. Self-reported debate audiences are approximately twice as big as comparable Nielsen estimates. Independent random samples generate widely divergent audience estimates for the same debate depending on when the survey was conducted. The second finding demonstrates low validity of self-reports without assuming validity of Nielsen estimates (or any other benchmark). The low validity of self-reported debate exposure poses a major obstacle for campaign effects research. Without valid measures of who was exposed to a campaign event, research cannot establish the causal impact of the event.

Candidate debates are important campaign events. Research aims to characterize the difference in attitudes and preferences between debate viewers and non-viewers (e.g., Benoit and Hansen 2004; Fridkin et al. 2007; Holbert 2005) or explain who watches debates and why (e.g., Baum and Kernell 1999; Kenski and Stroud 2005). Yet, as this study demonstrates, many non-viewers inaccurately claim to have watched debates, raising doubts about research that treats self-reported debate exposure as valid.

MARKUS PRIOR is Associate Professor of Politics and Public Affairs at Princeton University, Princeton, NJ, USA. Shim Reza provided research assistance on this project. *Address correspondence to Markus Prior, Woodrow Wilson School and Department of Politics, Princeton University, Princeton, NJ 08544-1013, USA; e-mail: mprior@princeton.edu.

doi:10.1093/poq/nfs019

© The Author 2012. Published by Oxford University Press on behalf of the American Association for Public Opinion Research. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

This study assesses the validity of self-reported debate exposure using two independent validation strategies. The first compares self-reports to Nielsen audience estimates. The second does not rely on Nielsen data, but instead compares survey estimates of the same debate taken at different times. If self-reports are valid, the time of the report should not affect the audience estimate.

Behavioral Self-Reports and Misreporting

A large literature shows that behavioral self-reports are fraught with error (e.g., Bradburn, Rips, and Shevell 1987; Hadaway, Marler, and Chaves 1993, 1998; Schwarz and Oyserman 2001). Regarding campaign exposure, research shows that respondents often inaccurately report exposure to news (Price and Zaller 1993; Prior 2009a) and campaign advertising (Ansolabehere and Iyengar 1998; Vavreck 2007).¹ However, these findings may not extend to self-reported exposure to presidential debates. Reporting exposure to a discrete and relatively salient presidential debate involves different cognitive mechanisms than reporting exposure to regularly available news or short, discrete advertisements.

Self-reports can be based exclusively on enumeration of discrete occurrences of a behavior or on estimation of the rate at which one engages in the behavior. Survey methodologists have suggested different sources of error for reports of frequent and infrequent behaviors (Conrad, Brown, and Cashman 1998; Burton and Blair 1991; Menon 1993). When respondents believe that they have recalled some but not all episodes of the behavior, they will estimate its frequency. Estimation is important for regular behaviors, such as network news exposure, because respondents are unlikely to recall all specific instances of exposure (Prior 2009b).

Trying to recall exposure to a debate might be easier because it constitutes a discrete, infrequent event. Several studies suggest that estimation tends to generate higher self-reports than does enumeration (Brown and Sinclair 1999; Burton and Blair 1991). If easier recall makes estimation unnecessary, overreporting might be less severe. Even then, Schwarz (1999, p. 97) concludes, self-reports will be unreliable “unless the behavior is rare and of considerable importance.” Although presidential debates are relatively rare, they may not be important enough for many people to be reliably recalled.

Self-reported debate exposure may also suffer from social-desirability bias if respondents perceive debate viewing as a component of “good citizenship.” Social-desirability pressure is the most prominent explanation for

1. Ansolabehere and Iyengar (1998) and Vavreck (2007) examine the validity of self-reported exposure to ads that were shown in a setting created by the researchers, not naturally occurring ads.

turnout overreporting (Belli et al. 1999; Bernstein, Chadha, and Montjoy 2001; Holbrook and Krosnick 2010; Presser 1990). Even though one study did not find any support for social-desirability bias in self-reporting of news exposure (Prior 2009b), the high salience of presidential debates may amplify the pressure to report exposure.

Comparing Self-Reports to Independent Benchmarks

The first part of this analysis compares survey-based estimates to Nielsen estimates of audiences for the 2000, 2004, and 2008 presidential debates. It uses the National Annenberg Election Survey (NAES), which consists of phone interviews with RDD samples of U.S. residents (Romer et al. 2006). Survey data are weighted to adjust for the number of adults and phone lines in the household, and to match census distributions of race/ethnicity, age, sex, and education. Respondents were asked, “Did you happen to watch the [vice-presidential/presidential] debate on [date] between [names of candidates]?” Exact wordings are given in figures 2–4. (In 2008, the NAES modified its debate question on October 22; see the online appendix and table 2.) NAES estimates are compared to national estimates by Nielsen compiled through “people meters” that monitor television viewing in a random sample of at least 5,000 U.S. households. Nielsen estimates use post-stratification weights to match the Current Population Survey on age, gender, race/ethnicity, and education.

Between 47 and 63 percent of voting-age residents reported watching the presidential debates in the NAES. Self-reported audiences for vice-presidential debates have a wider range, from 32 percent in 2000 to twice that in 2008. Table 1 presents NAES audience estimates for all twelve debates.

Even with generous allowances for small audience segments that Nielsen might miss, the self-reported audience is about *double* the Nielsen estimate for most debates. For this comparison, Nielsen ratings—the average per-minute audience of a program or channel, expressed as a percentage of all households or persons with television—are inappropriate. Instead, it is necessary to use cumulative audience or “reach”—the number of people who watch some part of a program—which exceeds the average audience, and may exceed it by a lot for a ninety-minute program.

For the first presidential and vice-presidential debates in 2008, Nielsen released cumulative audience estimates (Nielsen Company 2008). Of individuals aged 18 years and older, 28.6 percent watched at least six minutes of the first presidential debate on September 26, 2008, on a commercial network. The average audience for this debate in the same population was 21.9 percent. The vice-presidential debate drew a cumulative audience of 35.6 percent and an average audience of 29.4 percent. The cumulative audiences were thus 31 and 21 percent larger than the average audiences, respectively.

Table 1. Debate Audiences 2000–2008, Nielsen and National Annenberg Election Surveys

	Nielsen			NAES	
	(Viewers P2+ plus PBS) × .943/VAP	Rating	Reach	Mean	<i>N</i>
2000				<i>Oct. 4–29</i>	
First	22.5			47.7 [46.5; 49.0]	6318
Second	17.9			47.3 [45.7; 48.9]	3912
Third	18.3			49.5 [47.9; 51.1]	3657
VP	13.0			31.7 [30.5; 32.9]	5777
2004				<i>Oct. 1–Nov. 1</i>	
First	28.0			63.0 [62.1; 63.9]	11329
Second	21.2			57.0 [56.0; 58.1]	8620
Third	23.3			58.8 [57.6; 59.9]	7053
VP	18.6			49.7 [48.7; 50.7]	9568
2008				<i>Sept. 27–Nov. 1</i>	
First	22.4	21.9	28.6	62.3 [61.3; 63.3]	8675
Second	27.0			62.9 [61.7; 64.1]	6306
Third	24.4			59.4 [58.0; 60.8]	4512
VP	30.0	29.4	35.6	63.8 [62.7; 64.9]	7232
<i>Reach by Age Segment (2008)</i>					
	First	age 18–24	10.4	50.2 [44.2; 56.2]	269
		age 25–34	20.6	51.8 [48.3; 55.2]	809
		age 35–44	25.3	60.9 [58.3; 63.4]	1427
		age 45–54	30.2	64.2 [62.1; 66.3]	2004
		age 55+	41.4	70.4 [69.0; 71.8]	4166
	VP	age 18–24	13.2	43.3 [36.9; 49.8]	230
		age 25–34	27.2	58.4 [54.6; 62.1]	669
		age 35–44	34.3	60.9 [58.1; 63.7]	1180
		age 45–54	39.0	64.2 [62.1; 66.3]	1697
		age 55+	47.9	71.5 [70.0; 73.0]	3456
	First or VP	age 18–24	19.5	62.0 [55.6; 68.3]	230
		age 25–34	36.6	71.2 [67.8; 74.7]	669
		age 35–44	44.2	74.5 [72.1; 77.0]	1180
		age 45–54	49.3	78.6 [76.7; 80.6]	1697
		age 55+	59.5	81.8 [80.5; 83.1]	3456

NOTE.—Nielsen measured audiences on ABC, CBS, NBC, CNN, MSNBC, and Fox News in 2000. Audiences on the FOX network are added in 2004 and 2008. In 2008, CNBC, BBC America, Telemundo, Telefutera (first debate), and Univision (second and third) were also added.

The first column adds PBS projections to Nielsen's average audience estimates for viewers aged 2 and older, multiplies by .943 (the share of viewers 18+ for the two available 2008 debates), and divides by the size of the voting age population.

Rating and Reach estimates are live and same day from Nielsen Company (2008). Rating is the percentage of viewers aged 18 and older as a percentage of Nielsen population estimate. Reach is the percentage of viewers aged 18 and older who watched at least six minutes. Rating and Reach estimates do not include PBS audiences. See text for more information.

Nielsen estimates include only commercial networks, thus omitting PBS and CSPAN viewers. PBS provided its own audience projections² of 2.6 million and 3.5 million PBS viewers, respectively, for the first presidential and the vice-presidential debate in 2008. Assuming that PBS viewers were aged 18 years or older, average audiences for these debates would rise by 1.2 and 1.6 percentage points to 23.1 and 31.0 percent of adult residents. It would be technically incorrect to add PBS projections to the reach because they are not cumulative audience estimates, and because projected PBS viewers may have watched portions of the debate on other channels and would thus already be included in Nielsen's cumulative audience estimate. These biases operate in opposite directions. As a rough approximation, I add PBS projections to Nielsen's cumulative estimates.

Combining Nielsen and PBS data yields a cumulative audience of 29.8 percent for the first presidential debate. In the NAES, in contrast, 62.3 percent of respondents reported watching at least some of this debate. For the vice-presidential debate, the Nielsen/PBS cumulative estimate is 37.2 percent. According to NAES, the audience was 63.8 percent. Self-reported audiences thus exceed automatically tracked³ audiences by factors of 2.1 and 1.8, respectively. Judging by self-reports, both debates had about the same audience. According to Nielsen, however, the audience for the vice-presidential debate exceeded the audience for the first presidential debate by a quarter, or about 16.5 million viewers.

This mismatch is unlikely to be explained by Nielsen undercounts or atypical NAES survey procedures. Some self-reported viewers may have watched the debate online, but the online audience was probably less than one or two percent of the adult population in 2008⁴—clearly not large enough to change the diagnosis of massive overreporting. The NAES is no more prone to overreporting than other surveys. Seven of eight other polls that measured debate viewing in 2008 generated *higher* estimates than the NAES in the same period (see the online appendix).

For other debates, Nielsen released only the average audience estimates for individuals older than one (P2+). I use the relationships observed for the two 2008 debates with more detailed Nielsen data to impute the average and cumulative P18+ audiences. Those two debates had 3.1 million and

2. According to Robert Flynn, Vice President of Communications and Marketing at MacNeil Lehrer Productions, PBS projections are based on Nielsen data for local markets (phone conversation, October 1, 2010).

3. Television viewing is automatically tracked in Nielsen households, but individuals still need to identify themselves by pushing a button. Prior (2009a) shows that substantial overreporting of news exposure occurs among members of one-person households, for whom failure to sign in does not generate measurement error.

4. According to Stelter (2008), 285,000 live streams of the 2008 vice-presidential debate were initiated on MSNBC.com, and "CNN reported 2.1 million streams of live video from 9 p.m. to 11 p.m." The CNN figure appears to include post-debate analysis, and neither estimate excludes viewers outside the United States.

3.9 million viewers between ages 2 and 17, according to Nielsen average audience estimates, which represent 5.9 percent and 5.5 percent of the P2+ audience. Hence, I impute average P18+ audiences for other debates by multiplying the sum of the Nielsen P2+estimate and PBS audience projections by .943 ($= 1 - (.059+.055)/2$) and then dividing by the size of the voting age population. Table 1 (first column) shows these estimates. To impute cumulative audiences, I multiply the average P18+ audience imputation by 1.26, the factor by which cumulative and average audiences differ for the two debates for which I have data.

Averaged across the ten debates for which Nielsen did not release P18+ or cumulative estimates, self-reported debate viewing in the NAES exceeds imputed average audiences by a factor of 2.5 (ranging from 2.1 to 2.8) and imputed cumulative audiences by a factor of 2.0 (ranging from 1.7 to 2.2).

For 2008, Nielsen also provided cumulative audience estimates for different age groups and estimates of audience overlap between the first presidential and the vice-presidential debates. Only 18.6 percent of adults watched at least six minutes of both debates. As shown in figure 1, a majority of NAES survey respondents, 50.6 percent, reported watching both debates, 2.7 times the Nielsen estimate. Self-report inflation is thus even higher for joint exposure to both debates.

The second half of table 1 shows comparisons of Nielsen estimates and self-reports by age (for individual debates and for watching at least one debate). Debate viewing increases with age. All age groups overreport, but overreporting is much greater among young people and decreases monotonically with age. Overreport factors are 4.8 (presidential) and 3.3 (vice-presidential)

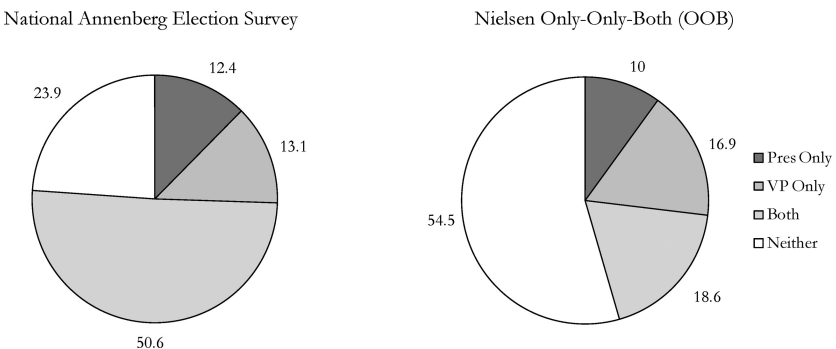


Figure 1. Overlap Between Exposure to First Presidential and Vice-Presidential Debate, 2008. National Annenberg Election Survey estimates are based on 7,232 respondents. Standard errors are approximately +/- 1 percent. Nielsen estimates are OOB estimates with six-minute thresholds from Nielsen Company (Sept./Oct. 2008).

among adults under 25. Omission of online debate viewing probably affects young people the most, yet greater overreporting among younger people has also been observed (for network news) at a time (in 2000) when newscasts were not yet available online (Prior 2009a).

The Temporal Consistency of Self-Reported Debate Audiences

Due to lack of information about the quality of Nielsen's sample and the reliability of its tracking method, skeptics might dismiss a mismatch between self-reports and Nielsen data as an indication of low validity of the Nielsen benchmark. A second approach therefore assesses validity of self-reported debate viewing without relying on external benchmarks.

NAES surveys were conducted as rolling cross-sections. By randomly releasing new phone numbers every day and retrying those that did not generate contacts for an equal number of days, this design yields independent random daily samples because "the date on which a respondent is interviewed is as much a product of random selection as the initial inclusion of the respondent in the sample" (Johnston and Brady 2002, p. 283). If respondents reported their debate exposure accurately, the daily means should be statistically indistinguishable because time of interview does not affect true exposure.

Figures 2–4 plot daily estimates of self-reported debate viewing. Locally weighted regressions summarize the time trend for each debate. In 2008, separate trends are fitted for the two question wordings. Some of the daily variation reflects sampling error. Daily sample sizes are between 257 and 323 in 2000, 228 and 519 in 2004, and 191 and 270 in 2008. The 95-percent confidence interval for the daily survey estimates is between ± 4.1 percent (for $N = 519$) and ± 6.9 percent (for $N = 191$). Table 2 compares confidence intervals of the highest and lowest daily estimates for each debate. The highest daily estimate for the first presidential debate in 2000 is 58.9 percent. With .95 probability, the self-reported audience is between 53.4 and 64.4 percent. The lowest estimate for this debate, 38.5 percent, has a confidence interval bounded by 32.9 and 44.1 percent. The difference between daily maximum and minimum for this and all other debates is highly statistically significant. Logit models using individual-level responses with fixed effects for each day reject the null hypothesis of no dependence on time for six of the twelve debates at $p < .1$ and for nine out of twelve debates at $p < .15$. Self-reports thus exhibit more daily variation than we would see if respondents reported their exposure accurately.

For most debates, plots of the self-reported audience suggest systematic time trends. Although linear declines seem most common, a few debates show short initial increases followed by later declines suggesting a quadratic trend. To verify that these patterns constitute more than random fluctuations, table 2 provides estimates of the lowest statistically significant time polynomial for each debate (up to cubic, fitted to the individual-level data, with a dummy for question wording in 2008).

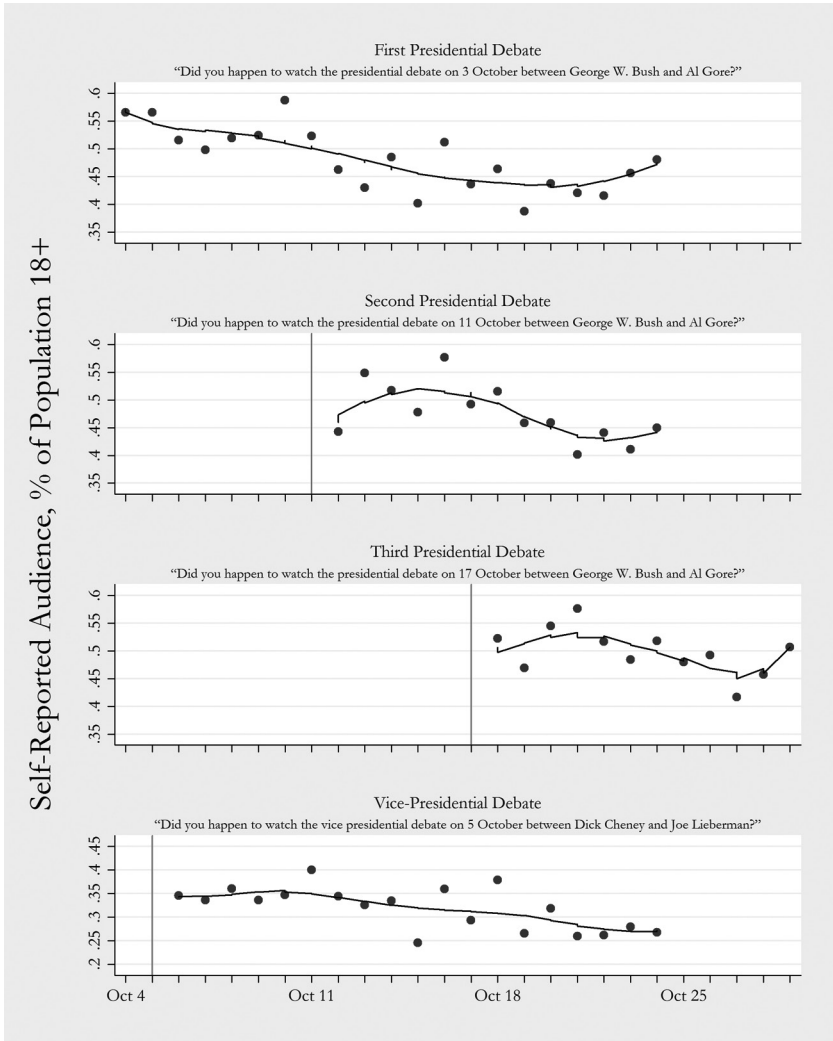


Figure 2. Debate Viewing over Time, 2000 (National Annenberg Election Survey). Trend lines are generated using locally weighted regression of daily averages on time with a bandwidth of .5. Vertical line indicates date of debate.

For six of the twelve debates, self-reported debate audiences exhibit a statistically significant negative linear trend. For another one, the negative linear trend is marginally significant. A quadratic trend fits the audiences of the first debate in 2008 and, marginally, the third debate in 2004. Finally, the self-reported audience for the second debate in 2004 is characterized by a significant cubic time trend. For only two of the

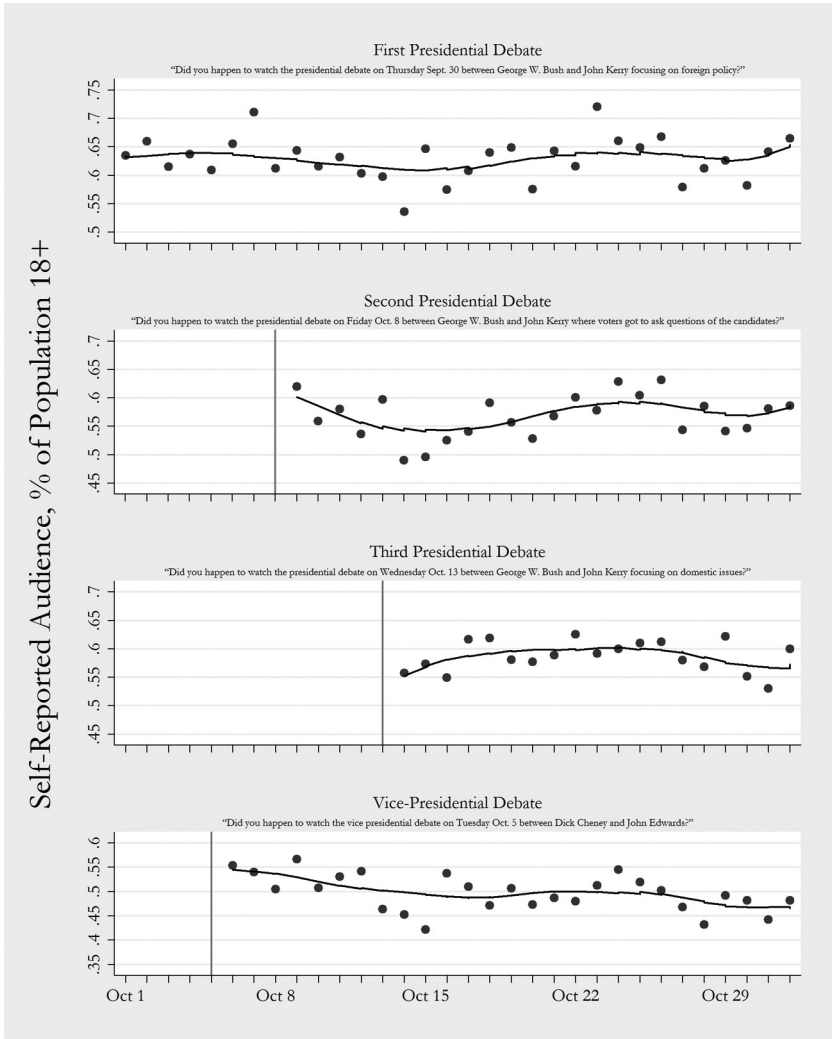


Figure 3. Debate Viewing over Time, 2004 (National Annenberg Election Survey). Trend lines are generated using locally weighted regression of daily averages on time with a bandwidth of .5. Vertical line indicates date of debate.

twelve debates does the self-reported audience not follow a significant time trend.

In surveys that are in the field for several weeks or that ask about debate exposure after the election, temporal variation in misreporting can add up to a large difference. In 2000, for example, audience estimates decrease by 20 to 30 million self-reported viewers (out of a maximum of 110 million) over the

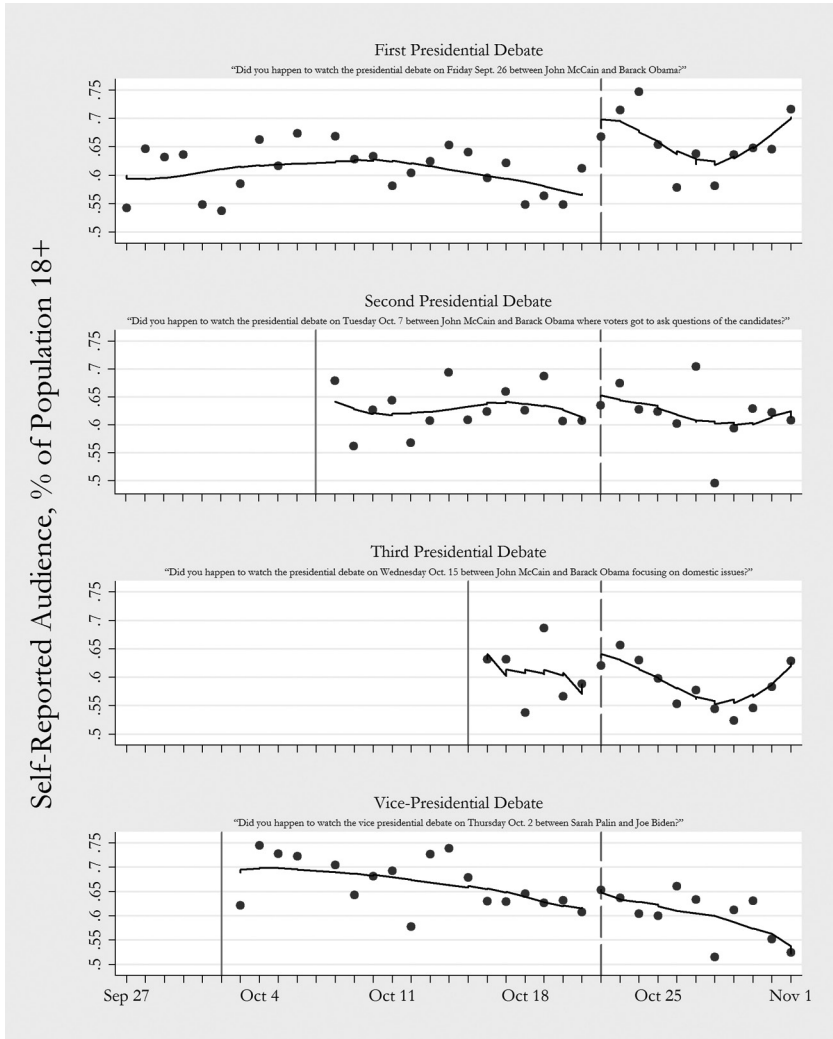


Figure 4. Debate Viewing over Time, 2008 (National Annenberg Election Survey). Trend lines are generated using locally weighted regression of daily averages on time with a bandwidth of .8. Solid vertical lines indicate date of debate. Dashed vertical red lines indicate change in question wording (see the online appendix). Starting October 22, NAES asked, “Which of the following election debates, if any, have you watched?” and then cycled through the four debates, accepting a separate answer for each.

Table 2. Temporal Dependence of Self-Reported Debate Viewing, National Annenberg Election Survey

	Lowest Daily Mean	Highest Daily Mean	Lowest-Sign. Time Polynomial
2000			
First	38.5 [32.9; 44.1]	58.9 [53.4; 64.4]	Linear: $\chi^2(1) = 30.0, p = .000$
Second	39.5 [33.9; 45.1]	56.9 [51.0; 62.8]	Linear: $\chi^2(1) = 10.2, p = .001$
Third	41.6 [36.2; 47.0]	57.5 [51.8; 63.1]	Linear: $\chi^2(1) = 3.9, p = .05$
VP	24.4 [19.6; 29.1]	39.8 [34.4; 45.2]	Linear: $\chi^2(1) = 18.0, p = .000$
2004			
First	53.6 [48.2; 59.1]	72.2 [67.6; 76.8]	n.s.
Second	49.0 [43.6; 54.4]	63.2 [58.0; 67.7]	Cubic: $\chi^2(3) = 9.0, p = .03$
Third	53.0 [48.7; 57.3]	62.6 [57.3; 67.9]	Quadratic: $\chi^2(2) = 3.9, p = .14$
VP	42.2 [35.7; 48.6]	56.7 [51.1; 62.1]	Linear: $\chi^2(1) = 8.1, p = .004$
2008			
First	53.8 [47.1; 60.4]	67.4 [61.2; 73.5]	Quadratic: $\chi^2(2) = 10.9, p = .004$
Second	56.2 [49.8; 62.6]	69.4 [63.2; 75.5]	n.s.
Third	53.8 [47.3; 60.2]	68.7 [63.0; 74.3]	Linear: $\chi^2(1) = 1.6, p = .21$
VP	57.8 [51.2; 64.2]	74.5 [68.9; 80.2]	Linear: $\chi^2(1) = 11.7, p = .001$

NOTE.—Cell entries are means with 95-percent confidence intervals in brackets. In the 2008 NAES, the format of the debate questions changed on October 22. For each of the four debates, 3,071 respondents answered the question in the second format. Estimates of functional form for 2008 include a dummy for question format.

course of three weeks. For the three vice-presidential debates, regressing people's self-reported debate viewing on a linear time trend generates statistically significant logit coefficients between $-.009$ and $-.032$. For every week that passes between the debate and the self-report, the probability of reporting debate exposure thus declines by about $.04$.

There is no indication that the variation in self-reports is caused by systematic differences in the composition of the sample. Polynomial time trends to predict education and age yield non-significant fits up to cubic. The share of registered voters or respondents who cast a vote in a primary did not vary significantly as a function of interview date, either. The most likely explanations for the significant time dependence of self-reported debate exposure thus involve changing recall, modified estimation rules, or varying social-desirability pressure.

Conclusion

For every survey respondent who correctly reports exposure to a presidential debate, there is one who claims to have watched but did not in fact do so. For young adults, overreporting is considerably higher yet. There are several reasons, unmeasured online viewing among them, to doubt any precise estimate of overreporting, yet the magnitude of the mismatch between self-reports and Nielsen data is so large that nibbling around the edges does little to question the bottom line: Surveys dramatically exaggerate debate viewing.

Serious doubts about validity of self-reported campaign exposure also emerge from the rolling cross-section design of the National Annenberg Election Survey. Independent random samples—drawn and interviewed according to the same procedures—provide widely divergent estimates of exposure *to the same campaign event*. The average difference between the highest and lowest daily estimates for a debate is 17 percentage points. Averaged across debates, the daily high exceeds the daily low by 40 percent. These differences are much too large to be explained by sampling error alone.

If self-reports of debate viewing were valid, they would not depend on the date of the interview. However, independent random samples yield estimates of the same debate audience that drop by up to 30 million viewers over a few weeks. At first glance, one might conclude that waiting a few weeks helps reduce the overreport problem, but it is entirely possible that people who actually watched the debate become less likely to say they did. The negative time trends are inconsistent with the claim that some respondents who did not see a debate use the question to report exposure to debate excerpts in later news reports, through social media, or on the Internet. The opposite is true for most debates examined here: Self-reports are highest right after a debate.

Low validity of self-reported debate exposure challenges conclusions about debate impact derived from survey research. If those who falsely report debate exposure learn about candidates from other campaign events, surveys might incorrectly credit debates with voter education. If the overreporters have little political information and change vote intention more frequently than actual viewers, studies using self-reports will underestimate learning effects and overestimate effects on candidate preference.

Other studies of debate impact forego self-reports of exposure and instead compare candidate support before and after a debate (e.g., Holbrook 1996, pp. 106–14; Shaw 1999). These aggregate estimates are not strictly estimates of debate impact and cannot distinguish between competing explanations for a before/after shift in vote intention. Did a debate change viewers' opinions? Did post-debate analysis affect vote intention of viewers and non-viewers alike? Or did other events that occurred at about the same time cause the before/after shift? Arbitrating between different causal explanations requires measuring debate exposure without bias. Understanding the bias better—through techniques such as list experiments and baseline primes (see Miller 1984; Prior 2009b), for example—is a first step. Ultimately, increasing validity of self-reports may require automatically and unobtrusively recording media exposure of survey respondents.

Supplementary Data

Supplementary data are freely available online at <http://poq.oxfordjournals.org/>.

References

- Ansolabehere, Stephen, and Shanto Iyengar. 1998. "Message Forgotten: Misreporting in Surveys and the Bias Toward Minimal Effects." Unpublished manuscript.
- Baum, Matthew A., and Samuel Kernell. 1999. "Has Cable Ended the Golden Age of Presidential Television?" *American Political Science Review* 93(1):99–114.
- Belli, Robert F., Michael W. Traugott, Margaret Young, and Katherine A. McGonagle. 1999. "Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring." *Public Opinion Quarterly* 63(1):90–108.
- Benoit, W. L., and G. J. Hansen. 2004. "Presidential Debate Watching, Issue Knowledge, Character Evaluation, and Vote Choice." *Human Communication Research* 30(1):121–44.
- Bernstein, Robert, Anita Chadha, and Robert Montjoy. 2001. "Overreporting Voting: Why It Happens and Why It Matters." *Public Opinion Quarterly* 65(1):22–44.
- Bradburn, Norman M., Lance J. Rips, and Steven K. Shevell. 1987. "Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys." *Science*, (April 10):157–61.
- Brown, Norman R., and Robert C. Sinclair. 1999. "Estimating Number of Lifetime Sexual Partners: Men and Women Do It Differently." *Journal of Sex Research* 36(3):292–97.
- Burton, Scot, and Edward Blair. 1991. "Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys." *Public Opinion Quarterly* 55(1):50–79.
- Conrad, Frederick G., Norman R. Brown, and Erin R. Cashman. 1998. "Strategies for Estimating Behavioral Frequency in Survey Interviews." *Memory* 6(4):339–66.

- Fridkin, K. L., P. J. Kenney, S. A. Gershon, K. Shafer, and G. S. Woodall. 2007. "Capturing the Power of a Campaign Event: The 2004 Presidential Debate in Tempe." *Journal of Politics* 69(3):770–85.
- Hadaway, C. Kirk, Penny Long Marler, and Mark Chaves. 1993. "What the Polls Don't Show: A Closer Look at U.S. Church Attendance." *American Sociological Review* 58(6):741–52.
- . 1998. "Overreporting Church Attendance in America: Evidence That Demands the Same Verdict." *American Sociological Review* 63(1):122–30.
- Holbert, R. L. 2005. "Debate Viewing as Mediator and Partisan Reinforcement in the Relationship Between News Use and Vote Choice." *Journal of Communication* 55(1):85–102.
- Holbrook, Allyson L., and Jon A. Krosnick. 2010. "Social-Desirability Bias in Voter Turnout Reports: Tests Using the Item Count and Randomized Response Technique." *Public Opinion Quarterly* 74:37–67.
- Holbrook, Thomas M. 1996. *Do Campaigns Matter?* Thousand Oaks, CA: Sage Publications.
- Johnston, Richard, and Henry E. Brady. 2002. "The Rolling Cross-Section Design." *Electoral Studies* 21(2):283–95.
- Kenski, K., and N. J. Stroud. 2005. "Who Watches Presidential Debates? A Comparative Analysis of Presidential Debate Viewing in 2000 and 2004." *American Behavioral Scientist* 49(2):213–28.
- Menon, Geeta. 1993. "The Effects of Accessibility of Information in Memory on Judgments of Behavioral Frequencies." *Journal of Consumer Research* 20(3):431–40.
- Miller, Judith D. 1984. "A New Survey Technique for Studying Deviant Behavior." Doctoral dissertation, George Washington University, Washington, DC.
- Nielsen Company. 2008. "Nielsen Examines the TV Viewership to the Presidential and VP Debates." <http://blog.nielsen.com/nielsenwire/wp-content/uploads/2008/10/2008-debate-tv-ratings-analysis-final.pdf>.
- Presser, Stanley. 1990. "Can Changes in Context Reduce Vote Overreporting in Surveys?" *Public Opinion Quarterly* 54:586–93.
- Price, Vincent, and John Zaller. 1993. "Who Gets the News? Alternative Measures of News Reception and Their Implications for Research." *Public Opinion Quarterly* 57(2):133–64.
- Prior, Markus. 2009a. "The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure." *Public Opinion Quarterly* 73(1):130–43.
- . 2009b. "Improving Media Effects Research Through Better Measurement of News Exposure." *Journal of Politics* 71(3):893–908.
- Romer, Daniel, Kate Kenski, Ken Winng, Christopher Adasiewicz, and Kathleen Hall Jamieson. 2006. *Capturing Campaign Dynamics 2000 & 2004: The National Annenberg Election Survey*. Philadelphia: University of Pennsylvania Press.
- Schwarz, Norbert. 1999. "Self-Reports: How the Questions Shape the Answers." *American Psychologist* 54(2):93–105.
- Schwarz, Norbert, and Daphna Oyserman. 2001. "Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction." *American Journal of Evaluation* 22(2):127–60.
- Shaw, D. R. 1999. "A Study of Presidential Campaign Event Effects from 1952 to 1992." *Journal of Politics* 61(2):387–422.
- Stelter, Brian. 2008. "Record Audience for Debate." *New York Times*, Oct. 4
- Vavreck, Lynn. 2007. "The Exaggerated Effects of Advertising on Turnout: The Dangers of Self-Reports." *Quarterly Journal of Political Science* 2(4):287–305.