

Measuring The Nature and Effects of Campaign Advertising

Travis N. Ridout¹
Department of Political Science
University of Wisconsin-Madison
tnridout@polisci.wisc.edu

Michael Franz
Department of Political Science
University of Wisconsin-Madison
mmfranz@polisci.wisc.edu

Kenneth Goldstein
Department of Political Science
University of Wisconsin-Madison
goldstei@polisci.wisc.edu

Paul Freedman
Department of Politics
University of Virginia
freedman@virginia.edu

November 12, 2002

¹ Travis N. Ridout and Michael Franz are doctoral students in political science at the University of Wisconsin-Madison. Kenneth Goldstein is Associate Professor of Political Science at the University of Wisconsin-Madison. Paul Freedman is Assistant Professor of Politics at the University of Virginia. This research was funded by a grant from the Pew Charitable Trusts. Direct all correspondence to Travis N. Ridout, 110 North Hall, 1050 Bascom Mall, Department of Political Science, University of Wisconsin – Madison, WI 53706. Email: tnrident@polisci.wisc.edu. Phone: (608) 263-2414. FAX: (608) 265-2663.

Abstract

One reason scholarly debates over the effects of campaign advertising still continue is that good measures of the frequency and content of the advertising environment and valid indicators of advertising exposure have been unavailable. In this paper, we review six approaches to the measurement of campaign advertising, discussing the strengths and weaknesses of each in turn. We then highlight the use of advertising tracking data for measuring the campaign information environment, a technique that has many of the advantages of traditional approaches without many of the disadvantages. We offer assessments of the approach's validity and reliability before showing how it can be used in combination with survey data to test hypotheses and make claims about advertising's influence on political attitudes and voting behavior.

Measuring Exposure to Campaign Advertising

Television advertising is one of the primary means by which modern political campaigns communicate with potential voters. In a typical presidential, congressional, or gubernatorial election, spending on television advertising comprises the lion's share of a campaign's budget. In a perfect world, scholars seeking to describe the nature of a political ad war and its effects would have precise information on exactly what ads were aired and who saw them, an ideal that has escaped us since the introduction of this campaign tactic.

In fact, the lack of comprehensive data on the content, timing, volume, and targeting of political advertising has limited what scholars can report about the strategies employed by campaigns and the balance of advertising in particular contests. But perhaps more important, the lack of comprehensive data on political advertising has made it difficult for scholars to measure citizen exposure to advertising and to study the effectiveness of these communications. Put simply, without comprehensive and reliable data on the targeting, volume, and content of advertising by all the players involved, it has been difficult to study the effect of the main persuasive tool utilized by modern electoral campaigns.

The purpose of this paper is threefold. First, we review and critique six traditional approaches to the study of political advertising. Three of these focus on measuring the information environment of a particular campaign through the use of aggregate campaign spending data, archival collections of television commercials, and logs from the public files of television stations. The other three approaches—experimentation, self reports by survey respondents, and proxy measures of exposure—attempt to measure the *effects* of advertising. As we will demonstrate, each of these methods has weaknesses that make it difficult for scholars

both to characterize the information environment and to draw inferences about how campaign messages influence the attitudes and behavior of citizens.

Our second goal is to present an alternative approach to measuring the content, volume and targeting of television advertising through the use of advertising tracking data. These data describe the frequency of all ads aired in the largest media markets in the United States and are coded on a number of important variables. We analyze the reliability of these frequency data and the validity of the coding. We show that the data are an excellent source of information about the activities of major political actors and the content of campaign messages. Indeed, these data represent the most comprehensive and systematic collection of political advertisements ever collected. Assessing the reliability and validity of the data is paramount given their impending public release in early 2003.²

Although these data are unmatched in their ability to trace the content, location and timing of political advertising, they are not ideal for estimating the effects of advertising exposure unless they are combined with survey measures detailing what television programs people are watching. Thus, the objective of the paper's third section is to suggest how the tracking data can be used to create an individual-level measure of advertising exposure.

Approaches to Measuring Advertising

Campaign Spending

One common proxy for the campaign information environment is candidate spending. By measuring campaign expenditures, scholars have sought to make descriptive inferences about the impact and relative volume of candidate messages. For instance, some scholars have examined the relationship between expenditures and election outcomes (Green & Krasno, 1990;

² The complete datasets for 1998 and 2000 will be available for download at www.fcc.gov in early 2003.

Jacobson, 1992; Gerber, 1999); others have explored money's impact on voter knowledge and affect (Coleman & Manna, 2000; Coleman 2001). While candidate spending may be a reasonable "quick and dirty" proxy for the intensity of campaign communications, the measure is far removed from the actual messages that voters receive and to which they respond. Indeed, researchers have recognized this mismatch, referring to a "black box" through which campaign money is translated into electoral outcomes. As Coleman and Manna acknowledge, "campaign money must work through campaign strategy, advertising content, advertising frequency, and other intermediaries" (2000, p.759).

Except perhaps in a few very corrupt places, money does not directly buy votes. Rather, money affords candidates the means by which to spread their message or bring their supporters to the polls, two activities designed to increase the candidate's vote share. The important point, as Ansolabehere and Gerber note, is that "total campaign spending may not be a good measure of expenditures devoted to actual campaigning" (1994, p. 1107). Making use of Fritz and Morris' (1992) comprehensive analysis of FEC spending reports, Ansolabehere and Gerber separate campaign expenditures into three types: "direct communications with voters," such as radio or television commercials; other campaign activities, such as polling or the hiring of a consultant; and spending unconnected to a candidate's own campaign, such as a donation of money to another candidate. The authors find that House challengers devote, on average, only 58 percent of total expenditures to campaign communications. For House incumbents, the comparable figure is just 42 percent (p. 1110).

More generally, there are three drawbacks to the use of aggregate spending data in tapping the information environment and testing for campaign effects. First, the use of aggregate spending assumes that every citizen in a particular constituency is exposed to the same volume

of campaign messages. Such an assumption is just plain wrong. Political advertising is not evenly distributed across the U.S.—or even across one state. Scholars have demonstrated substantial variation across media markets in the volume of candidates advertising, largely due to differences in the competitiveness of a race in the area (Just, et al., 1996). These differences are particularly striking in presidential campaigns, in which some television markets receive no advertising at all, while others receive thousands of paid spots (Goldstein & Freedman, 2002; Hagen, Johnston & Jamieson, 2002).

Similarly, how much advertising a campaign dollar will buy varies geographically as well. Quite simply, \$100,000 will purchase several times more spots in Cheyenne, Wyoming, or Alpena, Michigan, than it will in Los Angeles or New York City. Spending measures, then, are not comparable across media markets. Granted, this problem can be addressed by weighting spending by measures of gross ratings points (which account for the differential cost of air time), but such ratings information is not readily available, and few scholars, if any, take the time to make such corrections.

A final drawback of using aggregate campaign expenditures as a measure of campaign advertising exposure is that such figures ignore the spending of non-candidate actors, including parties and interest groups. This is an increasingly important limitation because “soft money” expenditures have skyrocketed over the past few years. Thus researchers who make use of candidate spending measures as reported to the FEC may be fundamentally understating the extent of campaign spending in a race. Moreover, the error in the measure is likely systematic, not random, because both party and interest group spending is generally targeted at a small number of very competitive races (Herrnson, 2001).

Archival Data

A second common approach to measuring the content of campaign messages makes use of archived political advertisements. For example, Finkel and Geer (1998), who estimate the effect of campaign tone on voter turnout, utilized a detailed content analysis of presidential advertisements obtained from the political commercial archives at the University of Oklahoma. Kahn and Kenney (1999) took this same approach in their study of negative advertising in U.S. Senate elections, while Kaid and Johnston (1991) used the Oklahoma archive to assess the negativity of presidential campaign advertising over time. If a scholar wants to describe the characteristics of the advertisements that a campaign *produces*, this approach is a good one. But if one's purpose is to study the effects of advertising *exposure*—as Finkel and Geer and Kenney and Kahn seek to do—or even to describe the information environment, then this measure has many problems.

Specifically, even if archives have a complete collection of advertisements produced in a particular election (a dubious assumption challenged by Jamieson, Waldman, and Sherr 1998, p. 7) archival collections contain no information on how many times each ad was broadcast. By default, then, an advertisement that was aired one hundred times receives the same weight in analyses as an advertisement that aired one thousand times. Indeed, some spots held by archives may never have been broadcast at all. Prior (2001) addresses this problem in one market by showing that one's conclusions may depend on whether one examines advertisements aired or advertisements made. Archival data, then, allow researchers to comment on the content of advertisements produced, but leave them unable to speak about the actual distribution of advertisements on the air or the effects of viewing these advertisements in a campaign context.

A second problem with archival data is the lack of state- and market-level data on the geographical distribution of airings, forcing the assumption that all voters in a given year were

exposed to the same volume and mix of advertisements. As argued above with respect to campaign spending, this is surely not the case.

Ad Buys

Alternatively, some scholars have collected advertising data directly from television stations, examining station logs, advertising purchase contracts, or billing invoices. The advantage of this approach is that one can get a good sense of the volume and timing of ads aired in a given market during a campaign. Magleby (2001) and Shaw (1999) are among the relatively few scholars who have measured advertising exposure by obtaining documents from television stations.

This approach, however, has some drawbacks. Although station logs detail when an ad actually aired on a television station, they are not public records and thus may be unavailable to researchers; moreover, patterns of availability may not vary randomly. Purchase contracts, which all stations must keep in their public files, are agreements between buyers and stations to air ads, but all advertisements listed on a purchase contract are not necessarily broadcast, nor broadcast in the time period indicated on the contract. Television stations often pre-empt a commercial's broadcast, a practice that is especially common close to the day of the election when candidates may engage in a price war for scarce commercial time. A third set of documents, billing invoices, accurately report what was aired and the price paid for each ad, but most television stations do not put these documents in their public files. Moreover, some stations omit billing invoices and contracts issued to political parties and interest groups. And so while the ideal situation would be one in which a station keeps detailed invoices for all candidate, party and interest group activity, this is rarely the case.

Another major problem with collecting television station records is that the task is time-consuming in the extreme. If one is interested in a single campaign (and hence a single or a handful of media markets), this approach may be tractable, but any analysis of multiple campaigns would require intensive data collection, and such efforts—given the existence of over 200 media markets in the U.S.—would be nearly impossible for a presidential campaign or for the universe of congressional campaigns. This limits the generalizability of claims drawn from an analysis of a handful of television stations. In addition, television stations provide the researcher no information about the content of ads, making it impossible to say anything about the tone of the campaign or the issues mentioned.

Experimental Manipulation

The three methods discussed above have been used both to describe the volume or content of political advertising in a campaign, and sometimes, in turn, to estimate the effects of exposure to this advertising. The three methods that follow bypass direct measures of the campaign information environment in their attempts to gauge the individual-level effects of advertising exposure.

This research has made sporadic but significant use of experimental design (Ansolabehere et. al. 1994, Ansolabehere and Iyengar 1995, Garramone et. al. 1990, Kahn and Geer 1994, Noggle and Kaid 2000). The allure of experimentation is obvious: by enabling researchers to control which subjects are assigned to which treatments, the nature of the stimuli to which subjects are exposed, and the conditions under which such exposure takes place, experiments afford an unparalleled degree of internal validity. Moreover, by manipulating specific components of a stimulus, experimental researchers can achieve a high degree of

specificity in the causal inferences they make. As Kinder and Palfrey argue in calling for more experimentation in political science, experiments offer an “unrivaled capacity...to provide decisive tests of causal propositions,” and constitute “a tool of unexcelled power and precision” (1993, p.11). In a series of elegantly crafted experiments, Ansolabehere and Iyengar (1995) manipulate very specific elements of fictional campaign advertisements to show that, in their experiments, it is negative tone and not some other feature of the ads (or more broadly, of candidates or races) that leads to demobilizing effects on voter turnout.

As is well known, however, the clear advantages of experimentation are offset by potential pitfalls. There is a direct tradeoff between the internal validity made possible by the rigorous control of the laboratory, and *external* validity, the ability to move outside the lab and beyond a particular experimental context in making inferences. Put simply, we care little about how a particular group of subjects responds to a particular set of stimuli; experimental findings are useful only to the extent that they allow us to make inferences about processes that occur *in the real world*. For many questions—including some asked in the social sciences—the fit between what goes on in the laboratory and the analogous process in the real world may be close enough to draw meaningful inferences. When it comes to measuring the impact of campaign ad exposure, however, external validity may be so compromised that researchers must proceed with caution in drawing inferences on the basis of experimental findings. Specifically, there are concerns about the nature of the treatment, the setting in which the treatment is administered, and the measurement of the outcome variables of interest.

When it comes to the realism of the treatment, there are a number of concerns. First, researchers must take care that the stimuli—the manipulated ads themselves—successfully mimic the kinds of spots that candidates and their allies are actually producing and broadcasting.

This, of course, is easy enough to accomplish with modern digital editing techniques. Actual ads can be spliced and diced in convincing ways without doing too much damage to realism. This is done in the better studies. Second, in the real world ads are embedded in shows—most often news broadcasts—and therefore experiments should not test spots in isolation. Once again, better studies are careful to present ads in the context of actual programming. A third issue, however, concerns the number, the intensity, and the pacing of the spots to which subjects are exposed. In the real world, people are exposed to a given spot dozens, even hundreds of times over the course of increasingly long campaign period. In the lab, however, it is unusual to find subjects exposed to a spot more than once or twice during an experiment. Thus, experiments must make inferences about overall exposure effects on the basis of spots seen only a handful of times (at best).

An additional concern emerges from the setting in which ad exposure typically takes place. In the real world, people see campaign ads in all sorts of places: in bars, bowling alleys, and at the gym. But by far the most common circumstance for encountering a political advertisement is in one's own living room, usually some time after dinner but before prime time. People see ads while they are sitting on couches or in favorite chairs, talking on the phone, chasing children or dogs, finishing a meal, or reading a magazine. In short, they encounter campaign spots while going about their own lives in their own homes. This is, obviously, a far cry from the relatively sterile, decidedly artificial environments in which even the best laboratory experiments are conducted. To be sure, researchers may take pains to alleviate such artificiality; providing couches, reading material, doughnuts or other snacks in an effort to make the viewing experience as normal as possible. Moreover, with the advent of web-tv and other new technologies, researchers have the ability to deliver advertising and other stimuli to subjects in

their own homes. This is a potentially significant step forward, although scholars have only recently begun to take advantage of these opportunities. Of course, even such advances do nothing to address issues just raised about the nature of the treatment itself.

In addition to concerns about the delivery of the experimental treatment, i.e., the independent variable, experimental approaches also raise questions about the validity of the outcome, or dependent variable. Particularly when researchers are interested in the effects of an exposure on political behaviors such as voter turnout (or for that matter, actual vote choice), it is difficult to find studies that go beyond hypothetical and prospective reports such as intended turnout. Obviously, there is reason to be concerned about subjects' ability to accurately evaluate and to fairly report the probability of future (especially hypothetical) behavior, and these concerns are exacerbated when such behavior is clearly subject to social desirability biases.

Given these fundamental impediments to external validity, researchers engaged in the study of campaign advertising must find ways to move outside the laboratory. Doing so involves real challenges, and ultimately, many of the most common approaches suffer from their own fundamental limitations, as we discuss below.

Individual-Level Recall

Another method of measuring advertising exposure asks survey respondents if they recall having viewed a political advertisement, and if so, what the advertisement was about (Wattenberg & Brians, 1999; Brians & Wattenberg, 1996). The advantage of this approach is that it creatively measures the campaign environment by looking at a presumed effect of advertising exposure—the ability to recall the advertisement. Thus it is something of a “bottom-up” rather than a “top-down” measure. Although the approach has more external validity than

experimental studies, the internal validity of the recall measure is questionable, making it difficult to establish the causal chain connecting advertising exposure and behavior.

First, researchers agree that, in general, people's ability to recall information is poor (Niemi, Katz & Newman, 1980). This appears to hold true in the context of campaign advertising as well. Ansolabehere, Iyengar and Simon (1999), for example, demonstrated through an experiment that over half of subjects failed to recall a television ad they had seen just 30 minutes prior. And as adherents of the on-line model of information processing would argue, even though a viewer may not recall an ad, it still may have had an effect on his or her evaluation of the featured candidate (Lodge, McGraw and Stroh 1989).

Second, and more seriously, there is potential endogeneity between ad recall and political behavior (Ansolabehere, Iyengar and Simon, 1999). That is, while seeing campaign advertisements may influence one's probability of voting, for example, one's propensity to vote may influence how much attention one pays to campaign messages. At the same time, while some differences in people's abilities to recall information surely have to do with varying levels of exposure, some of the variation can be explained by differences across individuals. In other words, recall is a function of more than just exposure. Indeed, in their re-analysis of the debate surrounding negative campaigns and turnout, Goldstein and Freedman (1999) include measures of the "perceived negativity of ads"—one measure of recall—and found no effect on turnout when controlling for exposure as measured by the advertising tracking data described later in this paper.

In addressing the problems of endogeneity in recall measures, Ansolabehere, Iyengar and Simon (1999) propose a two-stage estimation procedure in an effort to control for "actual"

exposure. As Goldstein and Freedman (2002) argue, such a correction is at best extremely difficult to pull off; at worst it may simply be impossible.

Other Proxy Measures

Because self-reported exposure to the news media can be unreliable, Zaller (1992) and Price and Zaller (1993) argue that message reception (being exposed to a message and accepting it) can be best tapped by an individual's level of political awareness, as measured by a political information scale constructed from a set of factual questions about current events and leaders.

This measurement strategy may make sense when applied to messages from the news media or to a total campaign environment, but it is less useful when applied to the specific case of television advertising. To measure the effect of *television advertising*, it is necessary to utilize a variable that measures exposure to and reception of *television advertising*. Knowing who Boris Yeltsin or William Rehnquist is, for example, should theoretically have little relation to whether one watches television programs during which many campaign commercials typically air. Although viewers may learn such facts while watching news programs, during which many political ads do air, advertising appears during many types of shows, including those unlikely to convey political information (game shows and afternoon talk shows, for example). More to the point, such information has even less to do with the content of most political advertising. Political knowledge, then, is a tenuous proxy for campaign ad exposure.

[Table 1 about here]

Table 1 summarizes our arguments about the advantages and disadvantages of existing approaches to measuring the political advertising and its effects.

Advertising Tracking Data

We now turn to a unique and relatively new source of information about television advertising: advertising tracking data obtained from the Campaign Media Analysis Group (CMAG), a commercial firm that specializes in providing detailed tracking information to campaigns in real time. The CMAG campaign advertising data represent the most comprehensive and systematic collection of political advertisements ever assembled. These data include two types of information: frequency information about when ads aired and in which markets, and information about each ad's content. CMAG, using a satellite tracking system, collects the larger set of broadcast data. The company has "Ad Detectors" in each of the 100 largest media markets in the U.S.³ These detectors track advertisements on the major national networks, as well as national cable networks. The system's software recognizes the electronic seams between programming and advertising and identifies the "digital fingerprints" of specific advertisements. When the system does not recognize the fingerprints of a particular spot, the advertisement is captured and downloaded. Thereafter, the system automatically recognizes and logs that particular commercial wherever and whenever it airs.

CMAG also creates a storyboard for each unique ad. The storyboard contains transcripts of all audio and a still capture of every fourth second of video. For the 1998 data, students at Arizona State University coded all storyboards; students at the University of Wisconsin-Madison coded storyboards in 2000, 2001, and 2002. Coders were asked to document the advertisement's tone, purpose, sponsorship, and even commonly used adjectives, among a number of other attributes.

³ This figure is for 2002. In 2001, CMAG recorded advertising in 83 markets, and in 2000 and earlier years, the company recorded advertising in the nation's top-75 markets.

To create the dataset, these coded data were merged to the larger dataset containing information about the ad airings. In the final dataset, each case represents the airing of one ad. Each case also contains information about the date and time of the ad's airing, the television station and program on which it was broadcast, as well the coding of its content.

These data improve on the six approaches discussed above in several fundamental ways. Because the unit of analysis is an ad airing, with information on the timing (both the day and time of day) and media market in which it aired, scholars can tell precisely how many ads (of whatever tone, sponsor, or other classification) aired on particular days in particular markets, sponsored by particular political actors. These data can then be aggregated to the level of the unique ad, and can be aggregated on market, ad type or some other variable.

Addressing Potential Critiques

The use of ad tracking data is nonetheless open to critique from several angles. One potential criticism of the CMAG data is that they are not comprehensive. The company's system tracks advertising in only the 100 largest of the 210 media markets in the U.S. But according to Nielsen Media Research, those largest 100 markets cover 86 percent of the television households in the country (Nielsen Media Research 2002). Although there are gaps in CMAG's coverage of the country – advertising in the hotly-contested 2002 South Dakota Senate race is untracked, for example – scholars will find that almost all of the respondents in any nationally-representative poll are covered.⁴

⁴ One concern is whether one's results might be biased by excluding people who live outside those 100 markets, who would tend to reside in more rural areas.

The CMAG data can tell scholars what political advertisements aired on the vast majority of American television sets. But are the data valid? That is, would one get the same results by visiting television stations and digging through their logs or billing invoices?

To test this, we visited eight television stations in five different markets: WJRT in Flint, Michigan; WISN in Milwaukee; WTAE and KDKA in Pittsburgh; WRIC in Richmond, Virginia; and WJLA, WRC and WUSA in Washington, D.C. At each station, we obtained photocopies of a sample of billing invoices prepared for various political advertising buyers. Each invoice included the name of the advertiser (e.g., Gore/Lieberman, Inc., the Virginia Democratic Party) and a list of the dates on and times at which spots aired. Next, we matched the spots listed on the invoices with those listed in the CMAG dataset.

With one exception, we found the tracking data to be highly accurate (Table 2). At WJRT in Flint, for example, we collected a total of 41 invoices representing 618 aired spots. Six spots listed on the invoices were not found in the CMAG dataset; the remaining 612 were accurately listed to within 1 minute. At WISN in Milwaukee, the error rate was slightly higher. On the 39 invoices collected, 977 airings were documented. For five of those, the time stamp given by CMAG differed from the television station's records by more than one minute, and we were unable to locate 22 of the airings recorded on the invoice.⁵ Still, over 97 of the time, the CMAG data and the station invoices agreed within one minute on when a spot aired. At WTAE, WRIC and the three Washington, D.C., stations, accuracy was similarly high.

[Table 2 about here]

The one trouble spot was KDKA in Pittsburgh. For reasons unknown to us, we were unable to locate 62 of the 308 spots listed on the invoices obtained from the television station.

⁵ This discrepancy could just as easily be due to inadequate record keeping by the station as inaccurate coverage by the CMAG system.

The missing spots, however, were not related to any candidate or any period of time in particular, suggesting that even the data from this one market are not biased systematically.

In sum, there seems to be little need for researchers to invest the time and resources into traveling to television stations to collect advertising data; advertising tracking data provide an accurate and reliable substitute.

We next explored the accuracy of the human coding. Many of the objective codes were not likely to contain significant coding errors, but we did have concerns about the more subjective questions, most importantly, the tone of the advertisement. To examine this question, we randomly selected 150 unique advertisements from the over 4,000 advertisements airing in 2000.⁶ As part of a larger project, we asked five undergraduate students to code the ads using our coding sheet, thus allowing us to compare our original coding with five independent assessments. We then compared each coder's results with those of the CMAG database. In general, we found inter-coder reliability to be high.

Below, we report the results for four questions of interest.

1. Is the ad positive, negative, or contrasting in tone?
2. Is the primary focus of this ad the personal characteristics of either candidate, policy matters, or both?
3. Does the ad direct the viewer to take any action?
4. Does the favored candidate appear on screen narrating his or her ad?

⁶ It should be noted that of the 150 randomly selected ads, some were advertisements that were never coded in the original dataset. This is because these ads were not political, but were nonetheless included in the data provided by CMAG. We dropped these ads from the inter-coder reliability tests but did not replace them with additional random advertisements. Thus, the N for the tests below is not 150, but is sufficiently large to merit statistical analysis where appropriate.

Most assessments of inter-coder reliability report only the percentage of times coders were in agreement. We do this in addition to providing correlation coefficients, specifically, Kendall's tau-b statistic⁷. For our tests relating to the advertisement's tone, we report both correlation coefficients and the percentage agreement with the original coding. This is reported in Table 3. The numbers in parentheses for the first column are the number of ads each individual coded. The numbers are lower than 150 because the CMAG data do not have codes for some issue-related ads—ballot propositions, for example, and the numbers are slightly variable across individuals because the coders skipped some ads. Finally, the numbers in parentheses for the other three columns are the number of CMAG-coded ads in each category. Thus, percentage agreement expresses the percentage of those ads for which the coder and the original data were in agreement.

In all five cases, the two-tailed correlation test is significant, and there is overwhelming agreement among the five coders on positive ads. There is less agreement, however, for contrast and negative advertisements; this represents a curious finding, but not one completely unexpected. More specifically, our observation in dealing with the data is that many contrast ads are overwhelmingly negative in tone. Thus, there may be some confusion among coders in separating out a negative ad from an ad that is contrasting, but highly negative.

[Table 3 about here]

We next report results for ads that deal with policy issues or candidate characteristics (Table 4). The interpretation of the numbers in parentheses is the same as above. The number of

⁷ Weisberg (1974) reviews models of statistical relationships, noting the different assumptions underlying alternative measures of correlation. Kendall's tau-b (like Pearson's R) is appropriate when testing for independence (as in a Chi-square test) and "strong monotonicity." Essentially, it is a strict test of correlation, where a *perfect* relationship between two (identical) variables exists only when all observations match up on the diagonal of a cross-tabulation.

ads reported in the first column is low because this test only considers election-related ads that were originally coded as dealing with policy or personal issues.

While the results are less encouraging, they are not terribly troubling. That is, the correlation coefficients are not extremely high (when compared to the tests for tone), but they are nonetheless significantly different from zero. And when we examine the percentage agreement, we can assess the source of the problem. Coders were able to easily determine ads that dealt with both policy issues and personal characteristics, but had a much harder time dealing with ads that were strictly policy or strictly personal. Our experience with coding ads tells us that strictly personal ads often implicitly deal with policy issues and that policy ads can be interpreted as somewhat personal.⁸ Thus, there is expected confusion in assessing the difference.

[Table 4 about here]

In many ways, such confusion is inherent in these subjective questions. We should expect this to be less so for relatively objective measures. We report reliability results for two such questions in Table 5, and predictably, the correlations are high and statistically significant at conventional levels.

[Table 5 about here]

In the end, the results reported here are encouraging. When using a storyboard to code, visual frames are sometimes difficult to read. And with each successive photocopy of the storyboard, words in the visual frames become less and less legible. Thus, accurate coding is hampered with this format, and in particular, our efforts to ascertain reliability are likely biased

⁸ On the former point, consider an ad criticizing a candidate about his divorce. While this should be coded as dealing with personal issues, the implication may be that the candidate has no right discussing policy issues that push certain family-related concerns. On the latter point, consider an ad highlighting a candidate's policy commitment to social security. The candidate may say something such as, "I have cared about this issue for over twenty years, especially as I get older." While this should be coded as a policy ad, it hints at a personal issue.

downwards, as the original coders used fresh storyboards. And while we do not have unanimity across coders, there is every indication that the coding of the CMAG data produces results that are reliable and unbiased. And on such subjective measures as an advertisement's tone – a question of extreme importance – we see that reliability is very high.

Individual-Level Analysis

Advertising tracking data do an excellent job of describing the volume, content and timing of political advertising in a given media market. But many scholars have a larger ambition: examining the effects of advertising on individual behavior. One might ask, for example, how exposure to different types of ads might affect citizens' perceptions of candidates and citizens' levels of political activity. In this section, we demonstrate how scholars can use the CMAG data to create a measure of individual-level exposure.

The dataset records the media market in which each spot aired, allowing scholars to aggregate the data to that level. Given an interest in a certain type of advertisement (by tone or sponsor, for example), one can determine how many spots of that type aired in each media market. And because most surveys contain information on the respondents' zip code, county of residence or area code, one can effectively determine the number of ads aired in the viewing location of most survey respondents. The potential exists to create highly detailed and specific exposure measures – exposure to party negative ads vs. candidate negative ads. In cases where researchers have questions that warrant this comparison, the CMAG data allow for it.

The aggregate market-level counts can be thought of as the upper limit on the number of ads a respondent could have seen over the course of the campaign. However, assuming that all respondents within a certain media market saw all ads aired in that market is unrealistic. Individuals have varying viewing habits, and of course, few people watch television 24 hours a

day, let alone manage to watch simultaneously the several stations on which spots air. Thus, to operationalize an effective exposure measure, one needs additional information on the viewing habits of each respondent.

Some surveys, including the 2000 National Election Study and some 2000 ABC News polls, ask respondents how many times a week they watch certain television programs such as morning news shows, game shows, the national news or talk shows.⁹ Because the tracking data also record the television program during which the ad aired, one can easily determine for each media market the number of ads aired during Jeopardy, Wheel of Fortune, the Today Show, and so on. By multiplying the frequency that a survey respondent watches each type of program by the frequency with which political ads aired during these programs, and then summing across all programs, we can create a relative measure of advertising exposure for each individual.

What makes this procedure easier is that the great majority of campaign commercials are aired on just a small number of television programs. Table 6 reports on the number of ads airing during ten different television programs. This table is based on an analysis of over 800,000 television ads aired in federal election races between June 1, 2000, and Election Day.

[Table 6 about here]

Almost 44 percent of the spots aired during local and national news programs, making news viewership a good proxy for ad exposure. The networks' morning news programs, including Today, Good Morning America, and the Early Show, saw another 11 percent of the spots run. But that is not the whole story. Syndicated game shows and daytime talk shows also

⁹ For the 2000 NES, question wording was:

How many times in the last week have you watched "Wheel of Fortune (Jeopardy)?"

How many days in the past week did you watch the local TV news shows such as 'Eyewitness News' or 'Action News' in the late afternoon or early evening?

How many days in the past week did you watch the local news shows in the late evening?"

How many times in the last week have you watched morning news programs such as "Today," "Good Morning America," or "The Early Show"?"

attracted large numbers of political spots. All told, the top ten programs for political advertising attracted about 64 percent of the campaign spots run. We recommend that researchers fielding their own surveys ask respondents the frequency with which they viewed each program. If time constraints prevent the inclusion of the entire battery of questions, we recommend asking respondents how often they watched television in the past week during the morning, afternoon, early evening, and late evening, in combination with specific questions about local news viewing.

Three points are worth noting. First, many surveys do *not* ask questions about viewing habits for particular television programs. This makes estimating exposure more difficult and less powerful, although not impossible. One substitute to using the show-specific or time-specific measures is to rely on a question about how often a respondent watches the local news because a plurality of political advertisements air during local news broadcasts. This measure is available on a wide variety of surveys. It should be noted that viewership of local news *is* a good proxy for advertising exposure, but it is an inferior one when compared to an exposure measure that uses the local news question in *conjunction* with the market-level data. More specifically, by combining the CMAG data with a local news question, researchers can create an exposure measure that varies at both the individual and aggregate level, providing more variation than when relying solely on news viewership.¹⁰ Second, the measure is correctly interpreted as a *relative* exposure measure. It is most useful in comparing levels of exposure across respondents, as opposed to saying with certainty how many ads a respondent saw. Third, the structure of the tracking data allow for estimating exposure up to the date of the survey interview. Thus, if a

¹⁰ For example, a typical survey question might ask how much attention a respondent pays to local news broadcasts about the presidential election. Answers to such a question might range from “none” to “a great deal,” with a total of about five categories. Including the aggregate market-level CMAG data allows the measure to take on potentially 100 (media markets) times 5 (local news viewership) values, at the extreme.

respondent was interviewed thirty days before Election Day, the data allow one to construct a measure of exposure that excludes ads aired after the date of interview. This fact allows the researcher to use almost any pre-election survey, provided there exists some television viewing question.

Thus, these data present researchers with a number of opportunities to test hypothesized links between exposure and political activity, and in ways that avoid the inherent problems and biases that result from using advertising data in the ways outlined at the beginning of the article. There are some legitimate concerns, however, about the validity of this individual-level measure. For instance, one component of the measure, television viewing, depends on the accuracy of respondent recall, something about which we expressed doubts earlier. Several reasons, however, lead us to believe that the recall of television viewing is much more accurate than recall of political advertising. First, the act of watching a television program is often intentional. People decide to sit down to watch certain programs, while the viewing of advertising is incidental. Second, television programs last 30 or 60 minutes, in contrast to the 30 or 60 seconds of the typical candidate ad. Moreover, television watching is often a regular activity, while exposure to a political spot is not. Watching certain programs is part of people's daily routines, and thus they are acts likely to be recalled. Finally, it is worth noting that Price and Zaller (1993), two of the main critics of measures of self-reported media use, acknowledge that "it is still possible that they are valid and reliable indicators of simple exposure, even if not of actual reception" (p. 159). This statement is comforting given that our purpose in inquiring about media use is not to tap whether people received messages from the media, but to measure potential exposure to advertising when they turned on the television.

Another potential concern with our individual-level measure of advertising exposure is endogeneity between watching certain programs and the behavioral outcome of interest. To illustrate, an investigator may find, for example, that exposure to television advertising and knowledge of candidates are positively related. One explanation for this finding is that exposure increases knowledge, but a second explanation is that more knowledgeable people are more likely to watch certain programs. We believe the likelihood of the second explanation is small. First, our measure depends on the viewership of several programs, not just news. It seems implausible, for instance, that Wheel of Fortune watchers would be systematically more informed about candidates than people who do not watch the program. Second, if analysts believe there is endogeneity between watching a certain program and behavioral outcomes, then they should control for watching that program by including it as an independent variable in their models.¹¹ Finally, to the extent that the second explanation for the relationship is true, the effect of advertising exposure on knowledge should be *underestimated* in any statistical model. Because the exposure measure is a combination of aggregate market-level data (surely exogenous to individual behaviors and attitudes) and individual viewing habits, the endogeneity between exposure (as a concept) will be less severe than when directly asking respondents if they were exposed to political advertisements.

Admittedly, our measure of individual-level exposure to television advertising is not perfect. Some people may change the channel when a political advertisement comes on the air, and others may tune out. Our measure does not take into account such peculiarities, but neither

¹¹ We examined the 2000 NES and ran a number of bivariate logit models of turnout on viewership of Jeopardy, Wheel of Fortune, morning news shows, early and late evening news, and talk shows. As one might expect, news viewership is strongly related to turnout, but this is precisely why local news viewership should be included as a control in models with exposure as the primary independent variable. Of the remaining shows, there was no effect for “Wheel,” a positive effect for “Jeopardy,” and a negative one for “talk shows.” Thus, there appears to be no inherent and damaging endogeneity in the non-news components of the exposure measure.

does any other measure. Moreover, we are unaware of any other measure of advertising exposure that provides the researcher with so much information about the volume, content and timing of the political advertising that citizens see and that reduces the risk of inaccurate respondent recall.

Conclusion

Whether and how campaigns matter has been the focus of enduring debate. Yet scholars are still far from a resolution to these questions, in part, because they lack good measures of campaign activity. We have demonstrated the accuracy of using the CMAG data as a method for measuring exposure to political advertising. The approach has several advantages over other alternatives. In contrast to archival research, the approach allows one to gain an understanding of where, when and to what extent a particular ad aired. And in contrast to examining television station records, using tracking data involves much less collection effort and allows one to know the characteristics of the advertisements that did air. The advantage of tracking data over measures of individual recall is the ability to construct measures of exposure based on objective ad airings, not the subjective—and often inaccurate—recollections of survey respondents. Tracking data are also more closely related to the viewing experience than measures of political awareness. Finally, the tracking data outshine experimental methods in terms of external validity while retaining a fairly high level of internal validity.

In the end, the CMAG data allow us to ask important and unanswered questions about the impact of political advertisements, and as a consequence, open the black box of political campaigns.

References

- Ansolabehere, S., & Gerber, A. (1994). The mismeasure of campaign spending: Evidence from the 1990 U.S. House elections. *Journal of Politics*, 56, 1106-1118.
- Ansolabehere, S., Iyengar, S., Simon, A., & Valentino, N. (1994). Does attack advertising demobilize the electorate? *American Political Science Review*, 88, 829-38.
- Ansolabehere, S., & Iyengar, S. (1995). *Going negative: How political advertisements shrink and polarize the electorate*. New York: Free Press.
- Ansolabehere, S., Iyengar, S., & Simon, A. (1999). Replicating experiments using aggregate and survey data: The case of negative advertising and turnout. *American Political Science Review*, 93, 901-909.
- Brians, C. L., and Wattenberg, M. P. (1996). Campaign issue knowledge and salience: Comparing reception from TV commercials, TV news, and newspapers. *American Journal of Political Science*, 40, 172-93.
- Coleman, J. J., & Manna, P. F. (2000). Congressional campaign spending and the quality of democracy. *Journal of Politics*, 62, 757-89.
- Coleman, J. J. (2001). The distribution of campaign spending benefits across groups. *Journal of Politics*, 63, 916-34.
- Finkel, S.E., & Geer, J. G. (1998). A spot check: Casting doubt on the demobilizing effect of attack advertising. *American Journal of Political Science*, 42, 573-95.
- Fritz, S., & Morris, D. (1992). *Handbook of campaign spending*. Washington, D.C.: Congressional Quarterly Press.
- Freedman, P. & Goldstein, K. (1999). Measuring media exposure and the effects of negative campaign ads. *American Journal of Political Science*, 43, 1189-1208.
- Garramone, G.M., Atkin, C.T., Pinkleton, B.E. & Cole, R.T. (1990). Effects of negative political advertising on the political process. *Journal of Broadcasting and Electronic Media*, 34, 299-311.
- Goldstein, K., & Freedman, P. (2002). Lessons learned: Campaign advertising in the 2000 elections. *Political Communication*, 19, 5-28.
- Goldstein, K., & Freedman, P. (2000). New evidence for new arguments: Money and advertising in the 1996 Senate elections. *Journal of Politics*, 74, 5-32.
- Green, D. P., & Krasno, J. S. (1990). Rebuttal to Jacobson's 'New evidence for old arguments'. *American Journal of Political Science*, 34, 363-372.

- Hagen, M., Johnston, R., & Jamieson, K.H. (2002, May). *Effects of the 2000 presidential campaign*. Paper presented at the 57th Annual Conference of the American Association for Public Opinion Research, St. Pete Beach, Florida.
- Herrnson, P. (2001, April). *Political party and interest group issue advocacy advertising in the 2000 congressional elections*. Paper presented at A Conference on Televised Campaign Advertising, Chicago.
- Jacobson, G. C. (1992). *The politics of congressional elections*. 3rd Edition. New York: Harper Collins.
- Jamieson, K.H., Waldman, P., & Sherr, S. (1998, April). *Eliminating the negative? Defining and refining categories of analysis for political advertisements*. Presented at the Conference on Political Advertising in Election Campaigns, Washington, D.C.
- Just, M. R., Crigler, A. N., Alger, D. E., Cook, T. E., Kern, M., & West, D. M. *Crosstalk: Citizens, candidates, and the media in a presidential campaign*. Chicago: University of Chicago Press.
- Kahn, K. F., & Geer, J. (1994). Creating impressions: An experimental investigation of political advertising on television. *Political Behavior*, 16, 93-116.
- Kahn, K. F., & Kenney, P. J. (1999). Do negative campaigns mobilize or suppress turnout? Clarifying the relationship between negativity and participation. *American Political Science Review*, 93, 877-889.
- Kaid, L. L., & Johnston, A. (1991). Negative versus positive television advertising in U.S. presidential campaigns, 1960-1988. *Journal of Communication*, 41, 53-64.
- Kinder, D.R. & Palfrey, T.R. (1993). On behalf of an experimental political science, in Donald R. Kinder and Thomas R. Palfrey (eds.), *Experimental Foundations of Political Science*. Ann Arbor: University of Michigan Press, pp. 1-39.
- Lau, R. R., Sigelman, L., Heldman, C., & Babbitt, P. (1999). The effects of negative political advertisements: A meta-analytic assessment. *American Political Science Review*, 93, 851-876.
- Lodge, M., McGraw, K.M., & Stroh, P. (1989). An impression-drive model of candidate evaluation. *American Political Science Review*, 83, 399-419.
- Magleby, D. (2001). *Soft money and issue advocacy*. Salt Lake City, Utah: Brigham Young University Center for the Study of Elections and Democracy.
- Nielsen Media Research. (2002). www.nielsenmedia.com/DMA.html. Accessed on June 19, 2002.

- Niemi, R., Katz, R. S., & David Newman. (1980). Reconstructing past partisanship: The failure of the party identification recall. *American Journal of Political Science*, 24, 633-651.
- Noggle, G., & Kaid, L. L. (2000). The effects of visual images in political ads: experimental testings of distortions and visual literacy. *Social Science Quarterly*, 81, 913-27.
- Price, V., & Zaller, J. (1993). Who gets the news? Alternative measures of news reception and their implications for research. *Public Opinion Quarterly*, 57, 133-164.
- Prior, M. (2001). Weighted content analysis of political advertisements. *Political Communication*, 18, 335-345.
- Shaw, D. R. (1999). The effect of TV ads and candidate appearances on statewide presidential votes, 1988-96. *American Political Science Review*, 93, 345-61
- Wattenberg, M. P., & Brians, C. L. (1999). Negative campaign advertising: Demobilizer or mobilizer? *American Political Science Review*, 93, 891-899.
- Weisberg, H. F. (1974). Models of statistical relationship. *American Political Science Review*, 68, 1638-1655.
- Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. Cambridge, England: Cambridge University Press.

Tables

Table 1: Summary of Advantages and Disadvantages of Approaches to Measuring Advertising Exposure		
Approach	Advantages	Disadvantages
Campaigns Spending	-Easily available measure	-Taps more than advertising -Assumes equal exposure across individuals -Costs not comparable across markets -Ignores “soft” money
Archival Data	-Ad content available	-Assumes each ad aired equal number of times -No market-level data
Ad Buys	-One knows when and where spots aired	-Purchase contracts don’t necessarily reflect what aired -Invoices and logs are not always available -No information on ad content -Collection is time-consuming
Individual-Level Recall	-“Real-world” measure	-People forget much of what they see -Endogeneity between recall and behavior
“Awareness” proxy	-Demonstrated valid measure	-Works better for political news than advertising
Experiments	-High internal validity	-Do not examine real behavior - low external validity

Table 2: Assessing the accuracy of tracking data with billing invoices

Television Station	Percent correct within 1 sec.	Percent off by more than 3 sec.	Percent not located	Total spots
WJRT-Flint	.99	.00	.01	618
WISN-Milwaukee	.97	.01	.02	977
KDKA-Pittsburgh	.80	.00	.20	308
WTAE-Pittsburgh	.97	.00	.03	773
WRIC-Richmond	.96	.02	.02	161
WJLA-Washington	.98	.01	.01	698
WRC-Washington	.99	.00	.00	363
WUSA-Washington	.97	.00	.03	1433

Table 3: Accuracy of Advertising Tone

	Kendall's Tau-b Correlation	Percentage Agreement		
		Positive	Contrast	Negative
Coder 1	.822 (127) **	95.6% (68)	72.4% (29)	90.0% (30)
Coder 2	.740 (124) **	97.0% (67)	74.1% (27)	70.0% (30)
Coder 3	.677 (129) **	98.5% (68)	54.8% (31)	70.0% (30)
Coder 4	.683 (129) **	98.5% (68)	80.6% (31)	53.3% (30)
Coder 5	.820 (120) **	94.0% (67)	83.3% (24)	86.2% (29)

Numbers in parentheses: for the first column, this is the number of relevant ads. For other columns, number of ads coded as such in original dataset.

Coefficient is the correlation with code in CMAG data.

Percentage is agreement with original data codes.

** $p < .01$ (two-tailed test).

Table 4: Accuracy of Advertising Focus (Policy or Personal Issues)

	Kendall's Tau-b Correlation	Percentage Agreement		
		Policy	Personal	Both
Coder 1	.413 (115) **	60% (10)	40% (40)	84.6% (65)
Coder 2	.438 (115) **	40% (10)	45% (40)	87.7% (65)
Coder 3	.457 (115) **	40% (10)	45% (40)	78.5% (65)
Coder 4	.454 (115) **	70% (10)	45% (40)	73.8% (65)
Coder 5	.329 (115) **	50% (10)	52.5% (40)	70.8% (65)

Numbers in parentheses – for first column, this is the number of relevant ads

For other columns, number of ads coded as such in CMAG data

Coefficient is the Correlation with code in CMAG data

Percentage is agreement with CMAG data

** $p < .01$ (two-tailed test)

**Table 5: Kendall's Tau-b Correlation
for Two Objective Measures**

	Does the ad direct the viewer to take action?	Does the candidate narrate his or her ad?
Coder 1	.719 (136) **	.796 (127) **
Coder 2	.690 (139) **	.832 (124) **
Coder 3	.759 (139) **	.903 (129) **
Coder 4	.705 (139) **	.856 (129) **
Coder 5	.557 (137) **	.895 (120) **

Numbers in parentheses indicate the number of relevant ads

Coefficient is the Correlation with code in CMAG data

** $p < .01$ (two-tailed test)

**Table 6: Television Programs During Which Most
Ads Were Aired**

News	355,802
Today	35,863
Good Morning America	30,901
Early Show	19,350
Wheel of Fortune	15,689
Jeopardy!	14,106
Oprah Winfrey	13,050
Live With Regis	11,841
Judge Judy	11,623
Nightline	11,230
Other programs	294,567
Total	814,022

Includes all spots aired in federal races between June 1, 2000, and
November 6, 2000.
