Electoral Studies 27 (2008) 732–739

Contents lists available at ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

Procedural transparency and the credibility of election surveys

Arthur Lupia*

Department of Political Science, University of Michigan, 4252 Institute for Social Research, 426 Thompson Street, Ann Arbor, MI 48104-2321, USA

Keywords:
Elections
Voting behavior
Survey research
Misspecification
Replication
Transparency

A B S T R A C T

The success and influence of survey-based electoral research is fueling the ambitions of survey analysts and producers. As a result, many new forms and uses of survey data are emerging. These new activities bring with them important questions about credibility. I address several of these questions by discussing common practices in the production and analysis of election surveys. I contend that the continuation of some of these practices threatens the credibility of individual studies and, in some cases, the election-oriented survey enterprise as a whole. In all of these cases, however, I argue that a commitment to increased transparency about analytic and/or production decisions can enhance credibility. In the process, I provide suggestions and examples of how transparency can be increased.

1. Introduction

Election surveys, and the insights provided by an international cadre of skilled analysts, have informed and transformed democratic governance around the world. In many countries, surveys affect what people believe about the legitimacy of electoral and legislative processes. Surveys also influence elected officials' perceptions of citizens' beliefs and desires and are used by many leaders to gauge the effectiveness of legislative programs and political strategies.

As the influence of election surveys grows, so do the ambitions of survey analysts and producers. Today, analysts use surveys to address an expanding range of questions. Traditionally, survey research was used to measure behavior and attitudes. Newer work uses surveys to examine deeper questions about cognition and emotion.

On the production side, complementary aspirations are manifest. Technological advances and changes in the scale economies of the survey industry are yielding new kinds of election surveys. Some researchers are using the Internet's video, audio, and interactive capabilities to redefine the survey interview experience. Others are using the Internet to accumulate large numbers of interviews at very low prices. While not yet common, it is only a matter of time before researchers administer election surveys through mobile communication devices such as cell phones.[1] The ability to survey people instantly, merge these data with demographic and related behavioral information, and quickly begin analyses is a logical next step in survey-based election research.

While the ambitions and modes of survey research evolve, one thing about this work remains the same: every survey-based claim about elections follows from arguments whose conclusions depend on the truth values of important assumptions. These assumptions apply to matters such as the representativeness of the sample, what respondents are thinking when they hear certain kinds of questions, and how the variables that are used to generate a particular result are (and are not) interrelated. Consider, for example, the claim that a certain number, say sixty-seven percent, represents "Americans approval of the war in Iraq" or the claim that a specific margin of

* Tel.: +1 734 647 7549.
E-mail address: lupia@umich.edu

[1] Already, we have seen papers at professional meetings using data on advertising exposure gathered through passive listening devices embedded in people's cell phones (Jackman et al., 2007).

0261-3794/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.electstud.2008.04.011

error, say "plus-or-minus four percent," applies to that numerical claim. The reliability of each claim depends on the truth value of a number of assumptions – some purely statistical and others based on the design of, and the methods employed when gathering, the data.

The ability of contemporary election survey analysts and producers to accomplish their ambitious goals depends on the relationship between the claims they generate and the truth value of the assumptions they make along the way. Are we doing what we should to fortify these relationships? To address this question, this essay focuses on common practices and key assumptions made in the production and analysis of election surveys. In the process, I argue that increased *procedural transparency* by those who produce, and those who analyze survey data can increase the credibility of survey-based election research.

By *procedural transparency in the analysis of survey data*, I mean *clarity and publicity about the steps that a scholar adopts in the process of assigning a particular meaning to, or drawing a particular inference from, a given data set*. In examining this kind of procedural transparency, I focus on the relationship between oft-used statistical models and the causal dynamics that such models are supposed to explain. Many analysts offer statistical models whose theoretical origins are unclear (and in some cases suspect). For example, many researchers use models that presume linear, additive, and independent relationships amongst key variables even when theory suggests (or evidence shows) such presumptions to be false. Others use the same kinds of models to characterize the mental states of citizens or voters, despite the fact that psychological research on related topics provides no clear basis for asserting such causal structures. I contend that the continuation of such practices can undermine the continuing credibility of survey-based election research. As such practices are fueled by scholarly norms that do not place a premium on transparency of the procedures that lead to the selection of a particular statistical model, I conclude that increasing procedural transparency in certain ways provides a means of increasing credibility.

By *procedural transparency in the production of election surveys*, I mean *clarity and publicity about the steps that researchers adopt in the process of converting capital (e.g., grant funding) and human labor into a given data set*. In examining this kind of transparency, I focus on procedural questions pertinent to matters of questionnaire content and debates about sampling procedures. The content focus pertains to whether and how leading surveys, particularly national election surveys, develop widely-used survey instruments in ways that are truly responsive to ideas from the scientific community. The sampling debate pertains to whether traditional telephone surveys or newer Internet-based surveys provide "better" platforms for administering election surveys. At issue is the extent to which familiar ways of recruiting respondents (that are experiencing declining response rates), or newer sampling procedures (that entail the use of volunteer samples), can be called "representative" of populations of interest.

The continuing credibility of survey-based election research depends on rigorously addressing questions raised in debates on topics such as questionnaire content and population sampling. To achieve such rigor, producers of survey data – in particular, producers of widely-used "national election studies" – must be increasingly transparent about their production procedures. Otherwise, needed methodological comparisons and evaluations cannot occur.

In my treatment of analysis and production issues, there is a common theme: *the benefits of increased procedural transparency*. As King et al. (1994:7–8) contend, scientific research has the following four properties:

1. The goal is inference.
2. The procedures are public.
3. The conclusions are uncertain.
4. The content is the method.

For this essay, point 2 takes center stage. I contend that limited introspection about key methodological assumptions in the production and analysis of election surveys puts scholars at risk of promulgating false claims. In each case, however, there is a feasible means of reducing the risk. It entails increasing transparency. I will argue that feasible means of increasing transparency – such as lab books – can help researchers design more effective data collection instruments, draw more reliable inferences from existing datasets, facilitate the all-important practice of replication and increase the speed and confidence with which knowledge about elections accumulates.

The essay continues as follows. First, I use a dose of formal logic to generate a standard for evaluating practices in the domain of election survey analysis and production. Using this standard, I review a set of common practices in contemporary scholarship. As I proceed through these practices, I propose means of enhancing credibility by increasing procedural transparency.

2. An evaluative standard

I want to argue that increasing procedural transparency in the production and analysis of election surveys can improve the credibility and reliability of survey-based claims. But to offer a means of improvement implies that there is a standard upon which improvement (or the lack thereof) can be measured. In the context of survey research, there are multiple standards of which one could conceive. Since my goal is to affect the credibility of current and future survey-based election studies, I will derive a standard from a common way in which people use election surveys – to make arguments.

Arguments are the primary currency of social science research. Whether researchers attempt to state trends, defend statements of fact, or support causal hypotheses, arguments are the means by which they convey their ideas to others. Arguments have two principal components: premises and conclusions. A conclusion is the point of an argument. It is the claim that the argument is intended to defend. Premises are reasons given to support the conclusion.

Many logicians classify arguments by relations between premises and conclusions. Deductively valid arguments are

seen by many scholars as having the most desirable relationship. The desirability comes from the fact that if all of the premises in an argument are true, then the conclusion *must be true*. For example,

> "If A, then B. If B, then C. If A, then C."

is a deductively valid argument. A deductively valid argument is one in which the logical connection between premises and conclusion is one of *necessity*.

Deductively valid arguments are compared to a more common relation between premises and conclusions, one of inductive validity. An inductively valid argument is one in which if all of the premises are true, then the conclusion *may be true*. For example,

> "If A, then B. If B, then C. If C, then A." "If A, then B. If C, then D. If A, then D."

are inductively valid arguments. An inductively valid argument is one in which the logical connection between premises and conclusion is one of *possibility*. Most arguments made in the course of normal conversation are of the inductive variety (e.g., one person attempts to persuade a colleague of the truth value ofa claim by offering reasons to support it – the reasons are typically related to the claim but the relation is not one of necessity).

How do matters of logical validity affect the credibility of survey-based election scholarship? When an election scholar uses a survey to make a claim such as "73 percent of likely voters prefer candidate Smith (with a four percent margin of error)" or "the coefficient of the relationship between education and turnout is .673 with a standard error of .257 and a specific underlying distribution of error variance," she is basing a conclusion on a particular set of assumptions about properties of the data. These properties are a result of the procedures by which the data was produced and a set of assumptions about the analytic techniques that she used. Whether her offering is meant to satisfy deductive or inductive standards, the truth value of her conclusion depends on the truth value of assumptions used to support the conclusion.

This fact about the use of election surveys is why scholarly debates about fact and causality in elections often focus on questions of data quality (including asking new or different questions) and analytic technique. Such debates can be quite constructive. At their best, they provide the means for analysts to produce conclusions with higher truth values in specific cases and/or to yield equivalent truth values in an expanded range of cases. Indeed, existing work in which previous assumptions about analysis were shown to be false have constituted watershed moments in the accumulation of knowledge (see, e.g., Achen, 1982 for an early review).

With the argument-quality-as-a-function-of-assumption-quality as our standard, I now ask you to reconsider the truth values of common assumptions made about the analysis and production of survey-based election studies. If the assumptions are false, then the reliability of the many conclusions that are based upon them is at risk. When I find problematic assumptions, however, I will argue that an increase in procedural transparency can act as a remedy to the credibility problems that the problems can cause.

3. The benefits of increased transparency in analysis

Analyses of election survey data are presumed to provide meaningful descriptions of what citizens are thinking and doing. Many survey claims, in turn, are generated by a scholar's use of a particular set of statistical techniques. Inherent in any statistical technique is a set of assumptions about a range of relationships that variables can have with one another. When a scientific discipline accepts a particular set of assumptions as fundamental to knowledge accumulation, the assumptions become paradigmatic. Henceforth, these assumptions tend to go unquestioned (see Kuhn, 1970: especially Chapter 3).

In the analysis of election survey data by political scientists, there exists a focal paradigm. The paradigm entails analyzing survey data through a commercial statistical package and assuming that the underlying causal model has a particular structure. The most common structure is $Y = \alpha + \beta' X + \epsilon$, where Y is the factor that the analyst is attempting to explain, X (a vector) is a set of factors that may affect the value of Y, α (a scalar) and β (a vector), are values derived from the analysis that are presented as explaining the relationship between X and Y, and ϵ is a measure of what is left unexplained by the model. In its most common form, each component of the vector X is presumed to have an independent, additive (and often linear) effect on Y. Common variants of this method include the introduction of a limited number of interaction terms, where elements of the vector X, $x_1,\ldots,x_N$, are multiplied with one another ($x_i{}^*x_j$, $i \neq j$), and with themselves (e.g., x_i^2) to make certain kinds of non-linear inferences.

Such mechanisms can provide a constructive way to generate inferences about complex causal relationships. But there are increasing questions about the reliability of the claims that they produce. Many of these questions are classified as instances of *misspecification*.

Consider, for example, the claim that a particular factor, X_1, has no effect on the dependent variable of interest, Y. (Typically, this claim is offered to support the argument that some other factor $X_2 \neq X_1$ is actually the key cause of Y.) The evidence often offered in support of the claim that X_1 does not matter, is that the coefficient of X_1 does not achieve a particular level of statistical significance. In such a case, however, what has been demonstrated is that the variable in question does not have a statistically significant effect *within the stated model*.[2] So, if all of the assumptions underlying the stated model are true, then we have reason to believe that the lack of a statistically significant relationship in the actual causal process is true.

But what if the structure of the true underlying data generating mechanism is not the same as the structure of the stated statistical model? This question is about more than omitted variable bias, a statistical ailment whose effects are well known. It is about the prospect that the effect of X_1 on Y may not be linear (or some other

[2] This critique puts aside the important question of whether the specific values of statistical significance often used as the basis of such claims are really meant to be used as binary criteria for whether or not a prospective causal variable matters entirely or not does not matter at all (see, e.g., King, 1986: 684).

pre-specified shape) or independent of other variables as is assumed in common interpretations of the results of a large number of published regression results. It may be that X_1 affects Y in one way if a variable X_2 has a value between 10 and 741 and affects Y in a completely different way if X_2 takes on any other value and either $X_3 = \cos(X_4)$ or the interview is conducted on a Tuesday. If the true data generating process and the posited empirical model are sufficiently different, then it is possible that X_1 in fact has a very important relationship to Y. However, because the model is *misspecified*, it fails to produce evidence of this relationship.[3]

Such problems of misspecification have been known, particularly to political methodologists, for decades (see, e.g., King, 1986). And yet, claims of the types just described are common because important aspects of the analytic procedures that produce them are treated as paradigmatic. Hence, key assumptions underlying these procedures are opaque, rather than transparent, to many readers. While it is true that some regression results of this kind are robust, in the sense that even when some key assumptions are not satisfied, unbiased estimates can emerge, there are other many cases when faulty (or, at a minimum, tenuous) conclusions are drawn because many analysts continue to believe that important underlying assumptions about regression analysis can go safely unquestioned.

What can be done to increase the credibility of such work? To be sure, the efforts of political methodologists throughout the world have provided many good options. Not only have they produced a host of new estimators that allow scholars to conduct analyses of a broader range of causal structures, but the penetration of their fundamental inferential critique have made certain kinds of misspecification common knowledge throughout political science.

That said, many people who want to work with election surveys either do not have, or lack the opportunity to acquire, the kind of advanced statistical training that many political methodologists possess. But *everyone has a powerful credibility-enhancing option* when it comes to making decisions about which model specification to use. That option is procedural transparency.

To see how greater transparency can aid credibility, I begin with the premise that all models (allowing for rounding error on the meaning of "all") are misspecified. Since, by definition, a model is typically offered as a simplification, in all likelihood a given statistical model of electoral phenomena is not an exact match of the true data generating function. So the credibility problems associated with much current research are not due to misspecification itself. Rather, it comes from doubts about why a particular specification should be regarded as a credible simplification or representation of an underlying data generating process.

A common manifestation of such doubts surrounds the practice called "stargazing." A stargazing analyst begins with a hunch that a particular variable, or perhaps a small set of variables, has an unappreciated association with, or causal impact upon, an important factor such as voting behavior or turnout. But the hunch does not offer details about the exact structure of that relationship. So a standard regression is run using an off-the-shelf statistical program. "Control" variables are added to the regression even though their relationship to voting behavior, and to the variable that is the subject of the hunch, is not really part of the hunch. Coefficients of key variables are then observed. The analyst looks for the "stars" that are symbolic of a statistically significant relationship between variables. If the stars are where the analyst wants them, the examination stops. Otherwise, additional regressions are run using alternate model specifications. New control variables may be added. Others may be dropped. No easily stated theory guides such decisions. Instead, the model's evolution is guided by the analyst's desire to see stars. The process stops when the analyst finds a regression in which the stars align (focal coefficients achieve conventional significance levels) with the initial hunch or some slightly altered version of it. The reason it stops here is because analysts who want to make headlines or publish in leading journals perceive a publication bias. He or she believes that statistical results in which stars do not align are less likely to be published. Stargazing clearly persists today (Gerber et al., 2001) and a growing number of scholars believe that this analytic practice is reducing the credibility of the electoral studies field as a whole.

One way to provide readers of our work with a reason to find a chosen model specification (and our conclusions) credible, is to engage in the practice of *keeping lab books*. In scientific disciplines such as laboratory-based subfields of chemistry, best practices entail the use of lab books that provide a methodical documentation of work, failures, and progress. Political science, and in particular the survey-based study of elections, has no such tradition. But we easily could – and we should.

For survey analysts, a typical entry in a lab book would entail stating the theory or theories one wishes to evaluate, explaining how and why specific hypotheses were derived from those theories, and describing (perhaps with its own concrete theory) the criteria by which data were selected or created for evaluating the focal hypotheses. Then one would state the empirical model to be used for the evaluation with an explicit defense of how and why a given set of control variables is included. Next, the analyst reports the results of the initial estimation. If the finding is not as anticipated, or suggests a revision to the theory, the need for different data, or an alternative empirical modeling specification, the analyst would explain why. They would detail how their observation changed their thinking and why this change necessitated a new approach. They would report subsequent observations and decisions as needed – documenting every estimation that they conducted, recording its attributes and, if an alternate estimation model is ultimately chosen, providing an argument as to why the change was necessary or sufficient to generate a more reliable inference.

Lab books could have multiple uses. In addition to keeping records for the researcher, they could aid in scholarly evaluations of new work (i.e., replication). For example, just as leading journals have begun to require replication

[3] The same critique applies to claims that X_1 affects Y in the specific way denoted by the relevant coefficients, standard errors, and goodness of fit statistics *because* it has a significant coefficient.

files (documents that describe how to derive a specific set of results from a given dataset) from analysts whose articles they consent to publish, they can also begin to require public access to lab books. So while lab books would not be published in journals, the expectation that they would be available could evolve. With a lab book and a replication file in hand, readers could replicate the logic that led to model selection and produced the stated conclusions. If lab books were honest, readers would not have to guess whether post-hoc rationalizations of observed analytic findings (e.g., stargazing) was at hand. Or, better yet, they could engage in open and constructive conversations about the extent to which particular conclusions are sensitive to key analytic assumptions. Such introspection happens in isolated cases now. Lab books would provide the materials needed to increase the range and value of such constructive conversations.

In other words, I contend that the regular availability of such lab books would lead to a fruitful expansion of current replication practices. In addition to using replication data to determine if a particular set of statistical claims can be recreated, scholars could use lab books to determine if the logic underlying the statistical model itself can be recreated. Expanding the domain of replication into the realm of underlying theoretical frameworks would give scholars who now "talk past each other," because they have limited means to replicate the logic or evidence of others, new opportunities to engage in more constructive conversations.

For scientific practice generally, such lab books are not a new idea. As Ramon y Cajal (1916), widely considered to be one of the founders of modern neuroscience, notes:

> "What a wonderful stimulant it would be for the beginner if his instructor, instead of amazing and dismaying him with the sublimity of great past achievements, would reveal instead the origin of each scientific discovery, the series of errors and missteps that preceded it – information that, from a human perspective, is essential to an accurate explanation of the discovery."

Such conversations are particularly important relative to the growing interest in using survey data to make claims about the cognitive or psychological state of citizens in electoral context. At present, many such claims are derived from traditional linear additive regression models. An entirely fair question that broad audiences should ask about such efforts is, "What are these models' theoretical origins?" In my experience, it is sometimes difficult, and often impossible, to trace the roots of many statistical models that are presented as explaining voter psychology to credible evidence of any such structural relations in any branch of psychology. And, as Graber (2001:5) argues (in the context of political learning – but the point applies more generally),

> "Unfortunately, failure to broaden the research approach beyond traditional political science domains has seriously harmed and impoverished the intellectual debate about political learning… Without a grounding in biological realities, judgments about the public's political acumen float in the never-ever land of unrealizable wishes, rather than in the far earthier reality of flesh-and-blood twenty-first century citizens struggling with the complexities of all aspects of life."

Greater procedural transparency about the selection of empirical models could lead to a more rigorous and constructive conversation about the most effective ways to draw from surveys reliable inferences about election-related psychology. Such documentation could clarify how causal structures offered by survey-based election researchers relate to causal structures identified in relevant fields such as psychology, the neurosciences, and education. It could help scholars determine whether the claims being made about citizen psychology *must be true* given a set of clearly stated assumptions (a deductively valid argument) or whether the claim is possibly true given those foundations (an inductive argument). Such inquiries could also help bring to light logically-sensible interactions between validated psychological structures and the many contextual forces identified by electoral scholars who focus on the incentive effects of institutions. In addition, greater procedural transparency of the forms discussed above could also lead to a more constructive conversation about the conditions under which surveys are better or worse than, say laboratory experiments, for evaluating particular hypotheses (see. e.g., Krosnick and McGraw, 2002).

More generally, the boost to credibility from practices (such as lab books) that increase procedural transparency need not just accrue to individual studies. If regularly expected and frequently available, they can also encourage or expand the domain of replication, which is an ideal way to boost the credibility of the field. Such practices can increase a reader's confidence that new model specifications are derived from defensible logic and evidence rather than a desire to be amongst stars.

4. The benefits of increased transparency in production

At the beginning of the survey era, most scholars who wanted to use surveys in their scholarship had to rely on surveys designed and fielded by others. Surveys were expensive and few researchers had the means to field a survey that could be devoted to their own research interests. As the survey-based study of elections matured and findings accumulated in a variety of substantive areas, this arrangement became increasingly unsatisfactory. There was only so much survey time available on leading surveys and with every passing election cycle a new set of interesting questions emerged. Individual scholars wanted to have greater input into survey design and content so that they could tailor the instruments and questions to contexts of greatest interest to them.

With the passage of time, emerging modes of survey production made new kinds of survey research more affordable. In the United States, changes in the regulatory environment in the 1970s and 1980s led to a massive drop in the price of long distance telephone calls. This was one of several factors that made phone-based surveys

affordable and led to an increase in the use of telephone-based surveys in election scholarship. In the 1990s and 2000s, the rise of the Internet enabled large-scale surveys to be conducted through that medium. Indeed, the Internet radically altered the economies of scale inherent in communicating with and coordinating large numbers of people who are distributed over broad geographic regions (Lupia and Sin, 2003). As a result, the evolution of Internet-based surveys have further driven down the cost of, and expanded the range of people who can run, large-scale surveys.

Technology and regulation have altered the scale economies of survey production. Which of the many new surveys that are emerging will allow scholars to draw *credible inferences* about election topics? Part of the answer to this question comes from the fact that in every survey there is a procedural path that converts capital and human resources into data points. The elements of this path are a series of decisions and actions by individuals (typically principal investigators and the data collection firms to whom they subcontract).

Many of these decisions affect data quality. Most are unseen by most users of the data. Should producers strive to be more transparent about their procedures?

Today, there is only a limited demand for increased transparency. Many analysts simply approach election survey data as though it provides valid measures of respondent beliefs and attitudes. But a growing literature on the psychology of surveys (see, e.g., Krosnick, 1999; Tourangeau et al., 2000) raises questions about the extent to which survey responses reflect respondent attitudes. These questions have generated more general inquiries about how many aspects of survey production affect the meaning of survey datapoints and the range of inferences that can be reliably drawn from survey data.

If survey-based election scholars want to respond to such questions effectively, they will need to know about decisions made during the process of producing the survey data. These decisions include matters of respondent recruitment, question wording, respondent conversation, interviewer incentives, interviewer characteristics (including the race and gender of an interviewer), and mode of interview (telephone, face-to-face, Internet, etc.). The more that analysts know about such decisions, the less that they will have to speculate about whether and how the data they have corresponds to the respondent's actual attitudes and beliefs. Generally speaking, the less that analysts rely on speculation about survey production procedures, lower is their credibility risk.

To allow researchers to gain knowledge about such procedures, survey producers must provide certain kinds of documentation. Producers of election surveys that will be used by many people, such as academic "national election studies," have a special responsibility in this regard. While the responsibilities for such producers are heavy and the number of production-oriented decisions is high, the stakes of proper interpretations of the data they produce are paramount. Therefore, greater transparency about procedures is in their projects' long-run interest.

At present, the rationale for many of the decisions made during the development of election surveys, "national" and otherwise, is not public. While users of these surveys can see codebooks that allow them to attach questions and answers to letters and numbers in a dataset, there is typically no documentation of why particular questions and response options were chosen. Moreover, while producers of election surveys typically release brief explanations of the sampling framework and related design elements, they seldom offer detailed information about how those decisions were made.

The lack of transparency about production procedures has been consequential. Though generally hailed for its transparency and attempts to solicit public input, many suspicions have been raised about how long-running surveys such the American National Election Studies (ANES)[4] and the British Election Studies (BES) chose which questions to include on past surveys. Some observers were upset about some issues being covered too little and others being covered too much. Some observers wanted the election study to focus on issues pertinent to a particular election year, while others wanted the election study to focus on issues that would fuel interesting comparisons across election years. Other scholars just wanted to know that points of view different than those of a small group of decision makers were being heard when the studies were being constructed. While "national election studies" such as the BES and the ANES cannot please all prospective constituents (at least not until scientific agencies agree to fund surveys of infinite length), they can increase the perceived legitimacy of their operations by including more information about how they choose content.

The ANES has recently taken a step in this direction through its development of the ANES Online Commons.[5] The Online Commons is a key part of a strategy to improve the breadth and quality of scientific input that goes into ANES survey development while increasing the public accountability of ANES decision makers.

In its initial incarnation (2006–2008), the ANES Online Commons was structured to encourage participation from a broad community of scholars. Any individual faculty member, student, survey researcher, or social science professional could participate. Individuals or teams of researchers were able to submit questions for placement on the survey, to comment on the proposals that had been submitted, and to express the level of support for the inclusion of specific proposals on the ANES survey instrument. All proposals that were sent to the Online Commons, and all comments about such proposals that were posted to the Online Commons, remain available for everyone who has an Internet connection to see.

This process was designed to produce (1) decisions about study content whose logic is transparent and public, and (2) substantial public debate and input about conceptualization, theory, and measurement *before* each study is fielded. This process constituted a stark contrast with normal practice in the development of many surveys, in which public conversations about survey development

[4] Disclosure: I (along with Jon Krosnick) am a Principal Investigator of the ANES.

[5] This description of the Online Commons follows that of Loftis et al. (in press).

have been more sporadic and occur mostly *after* the data are released, if they occur at all. In sum, the Online Commons was designed to increase participation in, and the transparency and accountability of, the ANES decision-making process. In the first two years of the program, over 600 scholars participated in the program, with over a third hailing from disciplines other than political science. Collectively, these researchers submitted over 3000 questions for consideration and subsequent users of the ANES data benefited from their many contributions.

That said, election surveys make numerous decisions in the process of converting dollars and human energy into data points. These include decisions about how to train interviewers, which sampling method to use, which weighting algorithm to use to adjust for prospective respondents who cannot be contacted or who refuse to participate, and how to code open-ended responses. At present, such decisions are typically described to users as *fait accompli*. But survey producers can increase their own credibility, and accelerate the accumulation of knowledge about important matters of survey production, by providing more information about these decisions and how they were made. Again, the analogy I draw is to lab books. We know that most producers of election surveys start with an ideal of what they want to accomplish, but along the way they have to (or choose to) adjust their implementation strategy. To be sure, some survey producers may be reticent about admitting to changes in design (e.g., after a change is made, the previous decision may seem embarrassing in retrospect) leading to a bias towards "publishing" descriptions of their own procedures that focus only on their final set of decisions – as opposed to how they arrived at those decisions. But if such documentation were provided regularly, user communities as a whole might recognize the complexity of assembling election surveys. Perhaps a significant number of scholars would read such reports in a constructive manner – realizing that each new election presents unanticipated challenges and understanding that they can benefit by learning more about the survey production process.

If survey production "lab books" were produced on a regular basis, more information would be available to inform debates about how current data should be interpreted and how subsequent studies should be run. For example, learning that survey producers were concerned about whether respondents would view a particular question in more than one way might help survey analysts make more effective uses of those questions (e.g., analysts might realize that common interpretations are less likely to be consistent with respondent attitudes). Similarly, learning that an important aspect of a survey's weighting algorithm (i.e., the algorithm supplied by producers of surveys to account for particular patterns of non-participation were debated during the production process, or learning that a series of algorithms was tried before a final selection was made, could inform users of weighted data about how sensitive particular inferences are to alternate weighting schemes. Such information could lead analysts to ask more constructive and focused questions about what existing data really means.

While the creation of lab books documenting decisions about procedures such as the selection of a weighting algorithm may seem arcane, they are essential to helping the field effectively manage one of its current hot debates. As of this writing, the use of volunteer samples in election-based scholarship is controversial. A growing number of private firms and large visible academic studies, such as the BES and the (US) Cooperative Congressional Election Study, are conducting surveys of large numbers of people who are not recruited through familiar methods. Familiar methods include random digit dialing, list-based sampling, or area probability sampling where survey firms purposefully contact respondents who themselves are randomly selected from a large representative (or exhaustive) list of citizens in an attempt to solicit their participation. In the new breed of Internet surveys, respondents often volunteer to be in the respondent pool. These respondents agree to participate either directly or indirectly (through their association with another organization that may issue survey invitations on a firm's behalf). The firms, in turn, maintain a set of these opt-in respondents from which they subsequently draw samples to complete specific surveys.

At the center of the controversy regarding such endeavors is whether and when claims drawn from such data can be considered representative of the population as a whole. Working against representativeness claims is the idea that people who volunteer to participate in surveys are systematically different from the people who do not volunteer. If the differences are sufficiently large or systematic, then such samples would be non-representative in many respects. But the argument is more complex than this. For at the same time that this volunteer Internet surveys have been rising in visibility, response rates on more traditional telephone and face-to-face based surveys have been falling. If the people who continue to participate for these surveys (after being contacted by one or more of the traditional methods described above) are systematically different than those who are now refusing to participate (or are impossible to contact), then these samples too would be non-representative in many respects.

Clearly, there are conditions under which samples that have identifiable patterns of non-participation can serve specific research hypotheses just as well as samples that are more representative. Also clearly, there are conditions under which the inferences generated from analyzing certain volunteer samples should not have any credibility. Years from now, election scholars will understand much more about these conditions than we do today. To minimize the number of years until important and higher levels of such knowledge are achieved, it is important to have clear, strong, and rigorous evaluations of many aspects of the production process. The effectiveness of these evaluations will depend on the amount and quality of available evidence. Greater transparency will give evaluators expanded opportunities to base their conclusions on more accurate assumptions about survey production processes.

Increasing transparency by survey producers will fuel increasing curiosity about the implications of production decisions by survey analysts, which, in turn, will induce scholars to demand more data on, and evaluations of, these decisions. So, for example, in the case of questions raised

about fading response rates and volunteer sampling endeavors, if survey producers are transparent about their methods, the properties of their recruitment strategies, sampling frameworks, and patterns of non-response, then important properties of survey data can be effectively evaluated. Such documents can also facilitate replication, which is becoming more feasible for many investors as key survey production costs drop. My hope is that increasing transparency about production procedures will induce the broad community of election scholars to discover for themselves the conditions under which survey samples have needed analytic properties rather than risk having unpleasant surprises about such matters being sprung upon them by subsequent generations.

5. Conclusion

> "Science at its best is a social enterprise. Every researcher or team of researchers labors under limitations of knowledge and insight, and mistakes are unavoidable, yet such errors will likely be pointed out by others. Understanding the social character of science can be liberating since it means that our work need not to be beyond criticism to make an important contribution…As long as our work explicitly addresses (or attempt to redirect) the concerns of the community of scholars and uses public methods to arrive at inferences that are consistent with rules of science and the information at our disposal, it is likely to make a contribution." (King et al., 1994:9)

New and coming advances in survey production and analysis have the potential to transform electoral politics and what we understand about it. Whether and how future efforts have such effects will depend on a set of decisions that people make. When analyzing data or producing survey data (or any election-related data, for that matter), documenting these decisions and making these documents public can have great benefits. A direct benefit of transparency in production will come to analysts. With such documentation in hand, analysts will have greater opportunities to conduct constructive replication studies and better determine the truth values of key conclusions and assumptions. If analysts can make these gains, then audiences of survey-based election claims will benefit as well. They will have new opportunities to substitute evidence for suspicion when attempting to determine the credibility of new work.

Given the many uses to which election surveys are put, there is great social value in seeking to bolster the credibility of survey-based claims about elections. Increased procedural transparency is a path by which the growing ambitions of survey researchers can increase the descriptive power, credibility and, hence, the social relevance of the work they do. For progress to continue, everyone who desires the survey-based study of elections to have growing and lasting credibility has an incentive – and people who aspire to be leaders of this field have an obligation – to provide honest, fair, and public accounts about the quality of their data and the assumptions on which their analyses are based.

Acknowledgements

I thank Lynn Vavreck, Elisabeth R. Gerber, and Yanna Krupnikov, and two anonymous reviewers for very helpful comments on earlier drafts of this essay. Lynn, in particular, made suggestions about the wording of key passages that I found enormously useful. I thank Harold Clarke, David Sanders, Marianne Stewart, and Paul Whiteley for extending an invitation for me to attend the Essex conference from which this Special Issue is drawn and for which I originally prepared these remarks. That said, I alone am responsible for all claims made in this essay.

References

Achen, C., 1982. Interpreting and Using Regression. Sage, Beverly Hills.
Gerber, A., Green, D.P., Nickerson, D., 2001. Testing for publication bias in political science. Political Analysis 9, 385–392.
Graber, D., 2001. Processing Politics: Learning from Television in the Internet Age. University of Chicago Press, Chicago.
Jackman, S.D., Lewis, J.B., Vavreck, L. 2007. Latent class modeling: network preference, advertising exposure, and channel-switching during the 2006 midterm election Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 13.
King, G., 1986. How not to lie with statistics: avoiding common mistakes in quantitative political inference. American Political Science Review 30, 666–687.
King, G., Keohane, R.O., Verba, S., 1994. Designing Social Inquiry: Scientific Inference in Qualitative Research. Princeton University Press, Princeton.
Krosnick, J.A., 1999. Maximizing measurement quality: principles of good questionnaire design. In: Robinson, J.P., Shaver, P.R., Wrightsman, L.S. (Eds.), Measures of Political Attitudes. Academic Press, New York.
Krosnick, J.A., McGraw, K., 2002. Psychological political science vs. political psychology true to its name: a plea for balance. In: Monroe, K.R. (Ed.), Political Psychology. Erlbaum, Mahwah, NJ.
Kuhn, Thomas, 1970. The Structure of Scientific Revolutions, second ed. The University of Chicago Press, Chicago.
Loftis, Vauthier, K.V., Lupia, A. Using the Internet to create new research opportunities: the new virtual communities of TESS and the American national election studies. PS: Political Science and Politics, in press.
Lupia, A., Sin, G., 2003. Which public goods are endangered? How evolving communication technologies affect the logic of collective action. Public Choice 117, 315–331.
Ramon y Cajal, S., 1916. Advice for a Young Investigator. 1999, (N. Swanson, L.W. Swanson, Trans). MIT Press, Cambridge MA.
Tourangeau, R., Rips, L.J., Rasinski, K., 2000. The Psychology of the Survey Response. Cambridge University Press, New York.