ANNUAL REVIEWS

Annual Review of Statistics and Its Application

Election Polls—A Survey, A Critique, and Proposals

Ron S. Kenett,[1,2,3] Danny Pfeffermann,[4,5,6] and David M. Steinberg[7]

[1]KPA Ltd., Raanana 43100, Israel; email: ron@kpa-group.com

[2]Department of Mathematics, University of Turin, 10124 Turin, Italy

[3]The Samuel Neaman Institute, Technion 32000, Israel

[4]Israel Central Bureau of Statistics, Jerusalem 95464, Israel

[5]Department of Statistics, Hebrew University, Jerusalem 91905, Israel

[6]Department of Social Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom

[7]Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69978, Israel

Annu. Rev. Stat. Appl. 2018. 5:1–24

First published as a Review in Advance on November 16, 2017

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-031017-100204

Keywords

election polls, election surveys, data collection methods, survey analysis, predictive analytics, information quality

Abstract

Election polls, also called election surveys, have been under severe criticism because of apparent gaps between their outcomes and election results. In this article, we survey election poll performance in the United States, United Kingdom, Canada, and Israel and discuss the current state of the art. We list the main data collection methods used in election surveys, describe a wide range of analysis techniques that can be applied to such data, and expand on the relatively new application of predictive models used in this context. A special section considers sources of error in election surveys followed by an introduction and a general discussion of an information quality framework for studying them. We conclude with a section on outlooks and proposals that require more research.

1. INTRODUCTION

Wherever open elections are held, public opinion polls, also called election surveys, are conducted before the elections. In this review we use the terms election polls and election surveys interchangeably. Extensive statistically based polling procedures began with the work of the Gallup Organization. Gallup successfully predicted that Franklin Roosevelt would be reelected by a wide margin in the 1936 US presidential election. Their success stood out in contrast to the much larger, but biased, Literary Digest poll, which pointed to his challenger Alfred Landon as the winner. Since that election 80 years ago, most election surveys have invoked statistical methods for selecting samples and analyzing the resulting data.

Historically, in spite of many success stories of election surveys, the record is mixed. Polls preceding the 1948 US presidential election gave Thomas Dewey the advantage over Harry Truman. Subsequent concern about the future of election surveys played an important role in establishing the American Association for Public Opinion Research (AAPOR). In Britain, preelection pollsters wrongly predicted a Labor victory in the 1992 parliamentary election. The Conservative Party won by eight percentage points, prompting an investigation by the British Market Research Association. For more background and examples see Frankovic et al. (2009).

Several recent failures have drawn much attention to the reliability and hence the necessity of election surveys. In the US presidential election in November 2016, pollsters were almost unanimous in showing Hilary Clinton with a clear lead; yet, Donald Trump won the election. Voters in Great Britain approached the vote for Parliament in May 2015 expecting, on the basis of the polls, a dead heat between the Conservative and Labor parties. However, when the returns were in, the Conservatives had a 7% edge and a solid majority in Parliament. Pollsters in Israel reported, in March 2015, that the Zionist Union party, led by Yitzhak Herzog, would win more seats than Benjamin Netanyahu's Likud party. However, final results showed Likud with 30 seats to 24 for the Zionist Union and a clear mandate to form the new government. Great Britain went back to the voting booth in June 2016 to cast ballots on the Brexit referendum. Contrary to the advance polls, the referendum passed in favor of the Brexit by a margin of 52% to 48%. In the current French presidential election, *Le Parisien Aujourd'hui en France* announced on January 3, 2017, their decision to suspend publishing the results of election surveys, favoring instead a return to journalism focused on field work and reporting rather than so-called horse races; the announcement referred specifically, and critically, to the election surveys noted in this review (**https://www.dailymotion.com/video/x577ump_le-parisien-renonce-aux-sondages-politiques_news**).

Despite these examples, recent election polls have not been uniformly off the mark. In the Canadian general election of October 2015, the polls accurately predicted Justin Trudeau's victory. We discuss all of these elections in more detail in Section 2.

Not surprisingly, election surveys attract significant public attention. And quite naturally, errors, like those noted above, generate much more discussion than the successes. In particular, they naturally lead to questions about the role of the polls and the validity of the methodology that supports them. In this article, we review the state of the art of election surveys, discuss the methods used, and raise our own suggestions for how to take the field forward.

We limit our coverage to preelection surveys and do not cover election-day exit polls. Although the two share common ground, there are also important differences between them. We focus only on polls or surveys made during election campaigns.

Traditionally, the goal of election surveys is to reflect the public view about standings of candidates, parties, and specific issues at the time the poll is conducted. As Traugott & Lavrakas (2008; chapter 1, p. 9) state, "polls provide. . .information about what the voters are thinking and how they are inclined to vote." This information is used to design strategies and plan tactical interventions by the parties and candidates contesting the election to influence current trends. A familiar debate is about whether the polls reflect public opinion or actually determine it to a large extent (see Simon 1954 and many sequels). For example, McAllister & Studlar (1991) studied British elections from 1979 to 1987 and found some evidence of a bandwagon effect, in which voters shifted support to the candidate who was ahead in the polls. They also found that survey results could encourage tactical voting in races with more than two candidates. In the context of strategic campaign decisions, temporal relevance and chronology of data and goal are critical attributes of the surveys. They need to be timely and provide the right information to the right people.

There is a subtle, but important, difference between reflecting current public sentiment and predicting the results of an election. Surveys have focused largely on the former—in other words, on providing a current snapshot of voting preferences, even when asking about voting preference as if elections were carried out on the day of the survey. In that regard, high information quality (InfoQ) surveys are accurately describing current opinions of the electorate. We define below the InfoQ framework and use the term accurate to mean both low bias and low variance, referring to the closeness of the surveys to the actual voter sentiment. However, the public perception is often focused on projecting the survey results forward in time to election day, which is eventually used to evaluate the performance of election surveys. Moreover, the public often focuses solely on whether the polls got the winner right and not on whether the predicted vote shares were close to the true results. Sometimes we refer more loosely to "accuracy" as whether the winner is correctly specified. Providing an accurate current picture and predicting the ultimate winner are not contradictory goals. As the election approaches, survey results are expected to increasingly point toward the eventual election outcome, and it is natural that the success or failure of the survey methodology and execution is judged by comparing the final polls and trends with the actual election results. Mosteller et al. (1949), in their study of the US election polls of 1948, proposed several possible measures of success, all of which were based on comparing the final poll predictions with the true outcome. See also Smith (1990).

An interesting development in recent years has been the advent of survey-based prediction models. The pioneer in this area has been Nate Silver and fivethirtyeight.com, an organization that does not conduct polls but is focused entirely on analysis and projection based on (but not only on) results from many different polling organizations. The quality of the prediction models depends on the ability of the polls to provide an accurate ongoing account of public opinion and voters' preference and how it changes during the course of an election. However, in addition to current data, other data sources and theories are also used in these models. In the 2016 US elections, fivethirtyeight.com actually computed two daily predictions, one based only on the data from the polls and one that also incorporated additional data, such as trends from past elections.

Section 3 presents the methods that are used in current election surveys. Section 4 describes sources of error in election surveys. In Section 5, we present a generic InfoQ framework and its relevance to election polls. We specifically refer to the eight dimensions of InfoQ proposed by Kenett & Shmueli (2014, 2016) and use them to analyze the information that is generated by election surveys for both the voting public and the candidates and their campaign managers. We close with a look to the future and a conjecture about how election surveys will develop in the years ahead, providing some suggestions and recommendations.

2. RECENT ELECTIONS

2.1. The Israeli General Election of March 2015

The elections of March 17, 2015, in Israel were held three months after the dissolution of the incumbent government in December 2014. Voting in Israel is a nationwide process, not local, and parties put up a list of candidates. Voters only choose which party gets their vote. The Israeli parliament, called Knesset, consists of 120 members with each party represented proportionally to its total vote. A new government is elected by the Knesset with the leader of the party with most votes being traditionally assigned the task of forming a government by the Israeli president. Election polls for the 2015 election were affected by several changes in the composition of parties. In the previous elections, the Likud party proposed a joint slate with the Israel Beytenu party, but in 2015 they ran separately. The Zionist Union was a partnership of two parties, Labor and Ha'Tnuah. Three different parties representing Arab citizens in Israel joined forces in a common integrated party. A new party, Kulanu, attracted many voters. A new law had set a higher threshold for receiving seats in the Knesset based on the popular vote, and support for two parties, Yachad and Meretz, appeared to be close to the threshold.

In the closing weeks before the election, the polls consistently showed either a slight lead for the Zionist Union party or a near tie with the Likud. Final polls were released on the 13th of March (Israeli law prohibits publicizing polls in the final 4 days before an election), with a consistent 3- to 4-seat advantage for the Zionist Union over the Likud (roughly 25 seats to 21). Nonetheless, the election results were 30 seats for Likud and only 24 for the Zionist Union. The polls were accurate in predicting how the vote would split between the parties on the right and those on the left; the additional seats for the Likud came largely from other parties at the same end of the political spectrum. However, the party with the most seats in the Knesset generally has considerable leverage to be chosen to form the next government, so the large disparity on the Likud vote between the polls and the actual election was of major political importance. For more details, see Fuchs (2017).

2.2. The UK General Election of May 2015

The elections in the United Kingdom were end of term and regularly scheduled. Four primary parties competed for the vote: the Conservatives (with David Cameron as the incumbent prime minister), Labor, the Scottish National Party (SNP), and the Liberal Democrats. The SNP was a relatively new party, with only six seats in the outgoing Parliament, but was expected to pull in considerable support from Scotland.

The elections in the United Kingdom are regional, with the winning candidate in each region representing the region in Parliament. Final election polls placed the Conservative and Labor parties neck and neck, at approximately 34% of the vote, and predicted that the Conservatives would have an approximate 10-seat advantage in Parliament. In the overall vote, the Conservatives had a 7% majority over Labor and achieved 330 seats in Parliament, approximately 50 more than had been predicted in the polls, enough to form the new government.

The British Polling Council, a professional body representing the polling organizations, and the Market Research Society responded to the big gap between the polls and the actual election results by forming an inquiry committee, headed by Professor Patrick Sturgis of the University of Southampton. We discuss later in this article many of the findings in the committee's report (Sturgis et al. 2016).

2.3. The Canadian General Election of October 2015

The Canadian elections were held on October 19, 2015. The primary contestants were the incumbent prime minister from the Conservative party, Stephen Harper, and his young challenger from the Liberal party, Justin Trudeau. The New Democratic Party (NDP), which held 95 seats in the outgoing parliament, was also a major contender.

Polls during the summer gave the lead in popular support to the NDP. By early September, the major polls showed nearly equal support for all three major parties. In mid-September, there was a steady decline in support for the NDP, with a corresponding increase in support for the other two parties. Initially, these results showed a close race between the Liberal and Conservative parties, but in the last two weeks before the election, the polls pointed to a slight drop in support for the Conservatives and a strong increase for the Liberals. The margins in the final polls were very close to the 7.5% margin in the final vote. The sharp drop in support for the NDP in the polls was also right on target with the final vote.

2.4. The Brexit Vote

On June 23, 2016, the United Kingdom voted on a proposal to leave the European Union (EU), commonly known as the Brexit referendum. Prime Minister David Cameron supported remaining in the EU; however, nationalist sentiment led to strong support for the proposal.

Opinion polls on the referendum began as early as September 2015. The polls consistently showed a majority opposing the referendum and preferring to remain in the EU. As election day approached, the polls showed the gap between "nays" and "yeas" narrowing but still found a majority in opposition. The actual vote contradicted those polls.

2.5. The US Presidential Election of 2016

The November 2016 presidential election campaign between Donald Trump and Hillary Clinton was followed intensively by polls for months preceding the election. The polls had a consistent lead for Clinton. Polls showed a near deadlock in July, right after Trump was nominated at the Republican convention, but Clinton's lead returned quickly. The gap in the polls narrowed as the election approached, but final polls still had Clinton approximately 3–4% ahead. State-by-state polls are also critical in the US presidential election, as the winner is determined by the electoral college, a group of delegates from each state with proportional representation, with the votes from each state cast for the candidate who carries the state. Those polls showed Clinton with a stronger lead.

Trump's victory came as a huge surprise to the pollsters, although they found some consolation in the fact that Clinton did win the popular vote (albeit by 2%, not the 4% of the final polls). The error in predicting the popular vote was well within the range of errors typical of past presidential elections. But the clear miss in predicting the electoral college result was seen as a major blow for the polls.

3. REVIEW OF ELECTION SURVEYS METHODS

In this section, we review data collection and analysis methods in the context of election surveys. For more details on these and other methods, see Frankovic et al. (2009). The section concludes with presentation of two new areas that have risen to prominence: prediction models and predictive analytics. Both areas involve innovative approaches to the analysis of survey data and, in turn, have implications for the collection of survey data.

3.1. Data Collection Methods in Election Surveys

The following subsections discuss various data collection methods. They range from the traditional and almost extinct mail surveys, up to the challenging social media–derived methods that rely on Twitter and other relatively new platforms.

3.1.1. Mail surveys. Mail surveys were used for many years as the preferred mode of data collection. These surveys have the advantages of low cost and lack of interviewer bias. A major drawback of mail surveys is their long turnaround time. Immediacy of information has great value for election surveys. Consequently, recent polls have invariably preferred alternative options with much shorter response delays. Also, response rates in mail surveys are typically low because of lack of interaction with interviewers.

3.1.2. Telephone surveys. Telephone surveys have been the primary mode for election surveys since the 1970s, when telephone penetration in the United States already exceeded 90%. Telephone surveys have significantly shorter execution time than mail surveys and are much less expensive than face-to-face surveys. They also allow centralized control over interviewers. These surveys typically use random digit dialing (RDD) to generate random samples. This is faster, easier, and cheaper than the enumeration and sampling procedures used for personal interviewing, but the use of RDD could become a very expensive operation if one attempts to control the representativeness of the sample with respect to important characteristics. Computer-assisted telephone interviewing (CATI) is known to reduce interviewer bias. Concerns about whether interviewers have physical access to respondents' addresses in urban areas also encouraged the use of telephones (Nathan 2001).

Telephone surveys have several disadvantages. In recent years, they have suffered from growing rates of nonresponse, partly due to innovations like caller ID, answering machines, and privacy managers. Traditionally, these surveys have relied on landlines, but with the growing proliferation of cell phones, many individuals no longer have landlines. Moreover, this phenomenon is especially true of the younger population, leading to potential bias in telephone surveys that use only landline telephone numbers. Interviews are usually limited to no more than 15 or 20 minutes.

3.1.3. Web-based surveys and Internet panels. The Internet has become a major platform for conducting surveys, including those for elections. One can distinguish between open-ended web surveys and Internet panels. In the former, a banner or another open call invites respondents to participate. Open-ended web surveys pose methodological challenges in assembling sampling frames for probability sampling and dealing with coverage issues and selection bias. Internet panels recruit thousands of volunteers through invitations on Internet sites or email messages. Specific surveys are then directed at a probability sample using the panel members as the sampling frame. The panel itself is subject to selection bias and the method assumes that this can be corrected through sophisticated weighting methods, such as propensity scores (Rosenbaum & Rubin 1983) or calibration (Deville & Särndal 1992). See Pfeffermann (2015) for a short review of bias correction methods with references.

As telephone response rates dropped, web-based surveys gained popularity. An advantage of web surveys is that they are self-administered and so avoid interviewer effects and reduce costs. Web-based surveys pose questions related to authentication procedures and differences in format and presentation across computing systems and browsers (Dillman 1978, 2000).

In the United Kingdom, a company called YouGov, founded in 2000, has been a major proponent of Internet surveys. YouGov maintains a large Internet panel for research and was successful in predicting the outcome of the 2001 parliamentary election, performing better than conventional polls. However, it was not as successful in the parliamentary elections of 2005 or 2016 nor in the US presidential elections of 2004 and 2016. As noted by Blumenthal (2005), preelection polls from self-selected Internet panels in the United States have done no better to date than telephone surveys.

Probability sampling for computerized surveys was originated by Willem Saris (1998), who developed this method in the Netherlands prior to the development of the Internet. In addition, the method utilizes computerized self-administered questionnaires (Schonlau & Couper 2017) and was implemented by the Telepanel of the Netherlands Institute for Public Opinion (NIPO or Dutch Gallup). Respondents were provided with computers and modems and were trained to download and fill out the questionnaire on the computer. Upon completing the questionnaire, each respondent uploaded it for data collection and processing. For more information on this and previous topics in this section, see Frankovic et al. (2009). A new approach for analyzing web panels that potentially avoids the problems associated with the use of propensity scores is proposed in Pfeffermann (2015). Schonlau & Couper (2017) review the state of the art on the design and analysis of web surveys and provide advice on tailoring the analysis to the survey goals. For more on web surveys see Kenett & Salini (2011).

3.1.4. Face-to-face interviews. Personal interviews are expensive and are less common in election polls. However, they have remained an effective way to collect detailed data that require longer questionnaires than are feasible in telephone or Internet surveys. Face-to-face surveys achieve higher response rates than other modes of data collection, and respondents are usually more engaged and forthcoming during the interview. Interviewers are also able to be more personal and interactive, use visual aids and clarify questions, and monitor nonverbal behavior. Computer-assisted self-interviewing may be used to reduce the potential for interviewer bias.

In the United States, the General Social Survey of the University of Chicago (GSS; **http://www.gss.norc.org**) and the American National Election Study (ANES; **http://www.electionstudies.org**) use face-to-face interviews. Both surveys use probability sampling to select respondents at the household level. For example, the GSS employs a multistage area sample to select blocks or segments and then adds a level of quota sampling at the block level. The response rate for the 2006 GSS was 71%; the preelection response rate for the 2008 ANES was 64%. The complex administration and cost of interviewing limits the frequency of face-to-face surveys. The ANES has typically included one or two preelection waves followed by one or two postelection waves. In 2012, they began to supplement the face-to-face sample with an Internet panel survey.

In modern surveys, there are also options to combine different response modes. For example, Internet surveys can be combined with telephone surveys, and telephone interviews can be complemented by home visits and personal interviews. Pfeffermann (2015) discusses the effects of the use of different response modes in a given survey.

3.1.5. Social media-derived data. Data from Tweets and other social media has been used to predict election results. An example is provided by the Italian start-up company Voices from the Blogs (**http://www.voices-int.com/**), which provided predictions for elections held in Italy and the 2016 US elections (**https://www.voices-int.com/home/usa2016**). In this context, a technique of opinion analysis is the iSA (integrated Sentiment Analysis) algorithm (Ceron et al. 2014, 2015), which extracts the sentiment from texts posted on social networks. This approach has also been used to capture instantaneous happiness from social media data (Curini et al. 2015).

3.2. Analysis of Election Surveys

Methods for analyzing surveys focus on adjustments aimed at reducing the bias that results from the failure of the sample to accurately represent the population of voters. A common approach is the use of reweighting, in which weights are computed on the basis of under- or overrepresentation, so that observations from sectors that are underrepresented in the sample get a weight larger than

1, and observations from overrepresented sectors get a weight smaller than 1. A comprehensive analysis of sampling error in the context of probability sampling is presented by Chambers (1999).

Examples of sources that lead to nonrepresentative samples include:

- Inadequate sampling frames that fail to include all voters
- Inclusion of interviewees who, ultimately, do not vote
- Nonresponse or lack of cooperation on the part of those selected for the sample
- In open-ended surveys, self-selection of the participants

These issues are classified and briefly discussed below. The classification is not mutually exclusive and includes some overlaps.

1. *Post-stratification.* One of the simplest methods of weighting is post-stratification, in which the observed sample is treated as if it were a stratified sample. The strata are determined using variables viewed relevant to the outcome(s) being assessed in the survey, for which data are available at the population level. Then, standard methods for estimating a mean from a stratified sample are used. For example, if our sample contains 60% men and 40% women, whereas the population has an equal number of men and women, then each man in the sample gets a weight of $0.5/0.6 = 0.833$ and each woman gets a weight of $0.5/0.4 = 1.25$. The total weight is unity, as $0.6 * \frac{0.5}{0.6} + 0.4 * \frac{0.5}{0.4} = 1$. Population averages are subsequently estimated by the weighted average of the responses for men and women in the sample. The choice of variables for the stratification is important. Party affiliation or past voting history are often used in election surveys, because of their obvious relevance to the current political attitudes. However, other polling firms have been reluctant to weight by party affiliation, arguing that it is not sufficiently stable, with significant short-term fluctuations.
2. *Calibration*. A related, but more general, approach is calibration. Here again, weights are sought for which weighted sample averages match known population averages for a number of variables for which population data are available. If z_i is the value for the *i*th respondent on the calibration variable z, then we seek weights w_i that satisfy $\sum w_i z_i = \bar{z}_{\text{pop}}$, the population mean of z. This constraint can be applied with respect to several variables simultaneously. Then, quantities of interest (such as the support for a given candidate) are estimated by similar weighted averages of the sample responses. Post-stratification is a special case of calibration in which the matching is to the fraction of the population in each of the subgroups formed by crossing the post-stratification variables. (A useful review of these methods can be found in AAPOR 2010).
3. *Inverse probability weighting.* Leading methods in this category include the propensity score (Rosenbaum & Rubin 1983) and Heckman's econometric approach (Heckman 1979). Both methods attempt to assign the self-selected survey respondents a propensity (or probability) reflecting how likely they are to join the survey. The scores are then used as inverse weights following the Horvitz–Thompson methodology for estimation from samples selected with unequal selection probabilities; the example of weighting by sex in point 1 above illustrates this idea. Here, the application is to surveys in which the inclusion probabilities are estimated, whereas the estimator was originally proposed for samples in which the unequal probabilities were part of the survey design. See Pfeffermann & Landsman (2011) for discussion of these and other related methods. Lee & Valliant (2009) found that propensity score adjustment could be usefully combined with calibration for reducing bias. Another approach based on decision trees was proposed by Yahav et al (2016).

The success of these methods is related to Rubin's (1987) well-known classification of missing data. The missing data here are those who declined to participate in the survey. When the probability of participation depends only on the calibration or propensity variables z_i in the sense that

$\Pr(i \in s \mid y_i, z_i) = \Pr(i \in s \mid z_i)$, where y is the study variable with value y_i for unit i and z_i represents the calibration (propensity score) variables, the nonresponse is missing at random (MAR). When the nonresponse cannot be explained by other observed data, the missingness is not at random (MNAR) and the nonresponse is informative. It is important to emphasize that calibration and propensity score weighting only eliminate the bias in the MAR setting. See Pfeffermann & Landsman (2011) for further discussion. Several nonresponse patterns are discussed below:

1. *Undecided voters.* Some voters decline to give a preference when asked. The fraction of undecided voters in election surveys can vary greatly across surveys and along stages of the political campaign. According to Visser et al. (2000), between 1988 and 1996, the proportion of undecided voters reported in polls ranged from 3% to 73% of the sample. Some polling organizations treat undecided voters as a separate category and report that percentage alongside support for the candidates, other organizations remove them and recalculate the percentage for each candidate or party, and others attempt to allocate them to the different categories of decided voters. Allocation schemes vary considerably. One approach is to assume that undecided voters who end up voting do so randomly, so truly undecided respondents should be allocated equally between candidates or parties. Application of this procedure could yield more accurate forecasts than eliminating undecided respondents altogether (Erikson & Sigelman 1995, Visser et al. 2000). Another approach allocates undecided voters disproportionately to new parties or candidates, reflecting an assumption that pre-election surveys systematically underestimate support for a challenging party (Panagakis 1989, Noelle-Neumann 1993).
2. *Voting intention*. Preelection surveys invariably include some eligible voters who do not vote in the election. Preferences of nonvoters may be different from those of actual voters. Thus, the data from (likely) nonvoters can bias the results of a survey. In the United States, pollsters have almost always first asked respondents whether or not they were registered to vote, and those who were not registered were automatically treated as nonvoters. The registered voters are then asked additional questions, designed to separate those who will vote from those who will not. One might ask whether the respondent voted in the past and would vote in the current election, or inquire as to their degree of political interest. For example, some screening measures in the past included a question on whether respondents knew the location of their polling places. A common solution has been to divide registered respondents into two groups. Respondents who scored beyond a specified cutoff have been designated as likely voters, and only their choices are counted in the tally (Daves 2000, Asher 2007). However, estimates of likely voters in the weeks and months prior to election day can reflect transient political interest with little relationship to behavior on the day of the election. An analysis of Gallup polls in the 2000 US presidential election indicated that the sorting of likely and unlikely voters is volatile and that much of the reported change is an artifact of classification (Erikson et al. 2004). An alternative that deserves further study is weighting the responses by the expressed likelihood of voting.
3. *Monitoring*. Repeated polls that monitor campaign dynamics on a daily basis are used to evaluate campaign events and the impact of political advertising. Typically, these so-called tracking polls take small samples of respondents (100–350) each day and estimate levels of support from rolling averages of two or three consecutive days to achieve samples large enough for reliable estimates. Thus, estimates may be based on 500–600 interviews aggregated across several days (Traugott & Lavrakas 2008). Tracking polls can be useful to assess campaign dynamics. The Washington Post tracking poll in the 2004 presidential election adjusted each day's sample of adults to match the voting-age population distribution by age, sex, race, and education, as reported by the Census Bureau, based on the Current Population

Survey. The Washington Post also adjusted the sample percentages of self-identified Democrats and Republicans by partial weighting to bring the percentages of those groups to within three percentage points of their proportion in the electorate, as measured by national exit polls of voters in the last three presidential elections (Washington Post 2004).

4. *Nonresponse adjustment*. Nonresponse is a critical issue in survey analysis. Selection bias due to nonresponse is an a posteriori effect that can make the set of completed surveys unrepresentative of the population of interest, in the sense that some groups are over- or underrepresented in the sample. The adjustment methods described earlier are useful tools to adjust for nonresponse bias, provided the nonresponse can be explained by weighting or calibration variables, as we noted at the end of point number 3, above. Bias due to self-selection poses a similar challenge in election surveys, in particular for open-ended Internet surveys. Kenett (1991) proposed a method for identifying significant nonresponse patterns in surveys. The method compares observed responses with expected responses in the case of no bias in various respondent groups defined by variables such as geographical location or socioeconomic status. If a significant nonresponse bias is determined, a weighting of the responses using weights determined by the target group is used (see Kenett & Salini 2011).

3.3. Prediction Models in Election Surveys

One of the most interesting developments in recent elections has been the use of prediction models. The goal of election surveys has traditionally been to reflect the current opinions of the electorate. Yet, much of the public interest in polls centers on viewing the results, not as a current snapshot, but as a prediction of the final results. The dominance of this view is evident in the wide consensus to assess the quality of polls by how closely they match the actual vote.

Prediction models are not polls but rather attempts to extrapolate to election day. Recent surveys are typically one of the major inputs to these models, but other inputs may also be relevant, including exogenous predictors like the state of the economy or voting trends from past elections and relevant variables like the lead time prior to the election. These models can also take explicit account of the extent of undecided voters and rates of nonresponse. Nate Silver and his fivethirtyeight.com blog was the first major player in the prediction field. Following Silver's success in predicting the outcome of the 2008 and 2012 US presidential elections (with correct predictions of the winner in 49 and 50 out of the 50 states, respectively), his ideas have set the tone in this direction. Silver's original model was based on a weighted average of leading US surveys. The weights were determined from statistical analysis of past election campaigns, with higher weights given to polling organizations that produced estimates that were closer to final results (Silver 2012). Over time, additional features were added. For the 2016 US election, fivethirtyeight.com presented two prediction models: one based only on polls and one that merged data from polls, demographics, historical voting patterns, and other relevant variables. For example, Silver's second model also reflects the extent to which voter sentiment can change in the time remaining until the election. Here, again, analysis of data from past elections is an important step in making this connection. See **https://fivethirtyeight.com/features/a-users-guide-to-fivethirtyeights-2016-general-election-forecast/** for details.

Prediction models, like that at fivethirtyeight.com, have a number of features that distinguish them from surveys.

- As noted from the outset, these models are focused on what will result when the ballots are counted, whereas surveys attempt to describe the current state of opinion. Consequently, issues related to extrapolation are highly relevant.

- Various inputs, including the latest survey results, are used as explanatory variables in a model that simulates the results of the vote.
- The predictions are probabilistic in nature. The model is used to generate replications of the election. The reported probability of winning is the fraction of simulation runs in which that candidate or party comes out on top. The replications also generate a predictive distribution of relevant quantities like the fraction of votes for a candidate. The reports to the media and the public emphasize the uncertainty.
- The simulations can also take into account possible systematic bias of the surveys. For example, in the days before the 2016 US election, Silver's daily columns at fivethirtyeight.com emphasized the possibility that if such a bias was present, it would likely have a similar effect on several swing states. This was one of the reasons why Silver's model gave higher probabilities to a Trump victory than did other models. Possible bias can be represented by a random effect, which adds variation to the simulation and hence uncertainty to the prediction. Or it can be represented by a fixed effect to generate a sensitivity analysis that examines whether the predictions are robust to the assumption that the surveys are not biased.

In the US elections of 2016, fivethirtyeight.com erred in several key swing states (Pennsylvania, Michigan, Florida, Wisconsin) that were instrumental to Trump's election. The survey errors in those states were a major reason for the prediction errors. In the 2015 UK election, the final fivethirtyeight.com prediction was substantially off target. As in the US election, the model largely reflected the final surveys. It showed a near draw between the Conservative and Labor parties and failed to predict the demise of the Liberal Democrats. See also Lauderdale (2015) and Bialik & Enten (2016) for sources of error in the prediction. See also Grajales (2015).

3.4. Predictive Analytics in Election Campaigns

Election campaigns are moving away from mass advertising via conventional media to targeted advertising that exploits social media. Data analysis plays a major role in that effort. Predictive analytics refers to the use of data to discover patterns and meaning in the data, with the goal of predicting outcomes or behavior. In the context of election campaigns, the analysis matches campaign messages and modes of presentation to subsets of voters in a way that is designed to maximize effectiveness in convincing voters to support a candidate. For general background on predictive analytics, see Shmueli et al. (2017).

Predictive analytics were successfully used by the 2008 Obama campaign in the United States. Following that initial success, the 2012 Obama campaign greatly expanded the use of predictive analytics, increasing the size of the analytics team by a factor of five and hiring a chief data scientist. Insights from the data analysis were an important ingredient in formulating the campaign strategy.

In the 2016 US election, both major presidential candidates were assisted by analytics teams. Much interesting information on how the analytics contributed to the strategy can be found on the website of Cambridge Analytica (**https://ca-political.com/index.php/casestudies/casestudydonaldjtrumpforpresident2016**), which advised the Trump campaign on how to effectively target and personalize its political messages. The company's data analyses apparently combined psychometric profiling, demographics, opinions on particular issues, voting intentions, and historical data to produce detailed predictions of how potential voters would respond to particular campaign messages. This is claimed to have led to targeted political campaigns tailored to specific population groups with some contradicting testimonials appearing in the press (Confessore & Hakim 2017).

The growing shift from campaigning via mass media to using personalized messages (by mail, text, social media, etc.) has also opened the door to running campaign experiments that have survey elements. For example, if a campaign has several alternative fund-raising messages to send by email, they can now compare them experimentally, sending each to a sample of likely supporters. Analysis of the subsequent response data can suggest not only which messages are most effective but even how best to match the message to the demographics of a recipient. Similar experiments can be performed with users of websites, following the A/B testing format (Kohavi et al. 2009). We read reports that the 2016 Clinton campaign made regular use of such experiments but have not seen any details or specific examples. In the past, with primary exposure via mass media, campaigners had to hope for some differential exposure to different messages to reveal such patterns. As this could not be designed into the experiment, methods of quasi experiments were needed to assess and compare effectiveness. Survey data are one of the key inputs to predictive analytics in an election campaign. However, the information needed requires more depth and detail than simply asking who an interviewee supports. The survey may involve psychological profiling, as apparently reported by Cambridge Analytica, and reactions to specific messages or issues. Invariably demographic information will be important to promote analyses that link campaign themes to voter groups.

4. ERRORS IN ELECTION SURVEYS

Many different factors contribute to errors in election surveys. These can be divided into sampling errors and systematic errors, or bias. Sampling error is related to issues of design (e.g., effective use of stratification) and sample size; statistical theory provides a complete guide to quantifying sampling error. Bias appears to be the more critical issue and is compounded by the fact that, although we can describe its sources, it is difficult to quantify and estimate them so as to account for them when summarizing the survey results. All the different sources affect the accuracy of the final estimates and so constitute what is known as total survey error (Groves & Lyberg 2010). In this section, we describe some of the major sources of bias and how they may have affected the survey results in recent elections.

4.1. Unrepresentative Samples

Election surveys attempt to reflect the current opinions and attitudes of the voting public. The design and analysis methods described in Section 3 are chosen to assure that the sample is representative, and the analysis methods provide ways to correct the estimates for any known failures to represent the voters. Nonetheless, failure of the surveys in this regard appears to be the major source of bias.

The UK election in 2015 provides an excellent case study. The inquiry report by Sturgis et al. (2016) evaluates and compares postelection surveys conducted by the major UK polling companies with surveys conducted by the British Election Survey (BES) and the British Social Attitudes (BSA) survey. The latter surveys are conducted for research purposes and used face-to-face interviews, with well-defined sampling frames and multistage, stratified random sampling. These surveys found leads of 7% and 6%, respectively, for the Conservative party over Labor, quite close to the actual election outcome. By contrast, the postelection surveys by the polling firms showed only a small margin in favor of the Conservatives, closer to the preelection survey results than the final vote. Thus, the evidence points to problems of accurate representation by the polling firms as a major source of bias.

Sturgis et al. (2016) also compared preelection results from the polling companies with those in the BES and BSA surveys. For voters aged 44 or under, all the polls had similar results for the Conservative lead. Among older voters, however, the BES and BSA postelection surveys showed a much stronger lead for the Conservatives than the preelection surveys had found. Assuming (as in Sturgis et al. 2016) that the BES and BSA surveys give an accurate picture of the voting patterns, the fact that gaps are specific to age is important. First, the presence of gaps at any given age implies that analysis methods that adjust only for age, ignoring other confounding variables, do not correct the bias. Further, it implies that the factors that are related to the failure to achieve a representative sample are themselves differential, leading to bias that is much more pronounced among older voters. Analysis by region found that the Conservative–Labor gap in the surveys was similar to the actual vote in most regions, but there was a tendency to underestimate the Conservative lead in the east, the southeast, and the southwest, areas where the Conservatives ran especially strong. Again, with region-specific bias, analyses that include marginal correction for regional imbalance in the sample do not remove the bias from the survey estimates. See also Farmer (2015).

Mercer et al. (2016) of the Pew Research Center raise similar criticism of the surveys preceding the 2016 US presidential election. They note that surveys have traditionally had difficulties interviewing a sufficient number of less-educated voters and that this was a key demographic group contributing to the Trump victory. At the time of this writing, postelection surveys like those in the United Kingdom are not yet available for comparison.

4.2. Under-Coverage of Target Population

Most current election surveys are conducted by either telephone or the Internet. In the former, the use of RDD has become increasingly problematic with (*a*) the shift from landlines to cell phones, (*b*) use of call-screening technologies, and (*c*) low rates of cooperation from those who do answer their phone. The net result is that there are sizeable segments of the voting population who are only partly accessible to phone surveys. Internet surveys offer improved cooperation but have their own problems of under-coverage. With a panel-based survey, the panel replaces the population of voters as the sampling frame. Panels are recruited in advance and consist of volunteers, so individuals who are not willing to volunteer—not to mention those who do not use the Internet—are clearly misrepresented. An interesting question, not answered in the Sturgis et al. (2016) report, is whether the differences between the preelection surveys and the BES and BSA surveys among older voters could be due, at least in part, to self-selection/cooperation bias that is specific to older voters.

4.3. Informative Nonresponse

Nonresponse refers to several categories: sampled subjects who cannot be contacted, subjects who decline to participate in the survey, and those declaring that they have no established preference among the candidates. Typically, rates of nonresponse differ across supporters of candidates/parties, so that the observed responses constitute a biased sample. It is important to assess whether the nonresponse is MAR or MNAR (Rubin 1987). As mentioned in Section 3.2, adjusting for informative nonresponse (MNAR) is a very difficult task and requires in general the use of statistical models, which are hard to test. Some survey organizations have attempted to correct for this problem with adjustments at the survey (not the individual) level, using models from past elections that compare information from interviewees willing to disclose their preferences with the final election results.

4.4. Misinformation

One of the problems cited by survey firms is that sample participants do not always truthfully reveal their preferences. In the 2015 Israeli elections, Fuchs (2017) reported intentional efforts to misinform so as to discredit the media and perhaps affect the election results. Kellner (2015) of YouGov cited “shy Tories” as one explanation why the YouGov surveys failed to predict the UK outcome in 2015. As noted by Mercer et al. (2016), this phenomenon may also have affected the 2016 US election prediction, with interviewees reticent to express preference for Trump. Other articles have noted a tendency toward political correctness—for example, reticence to express support for parties considered extreme or to oppose minority candidates. This is generally known as the Bradley effect, following the 1982 elections for governor of California when Tom Bradley, a black candidate, had a clear lead in the surveys but was defeated by George Deukmejian in the election. See also Anderson et al. (1988).

4.5. Poorly Designed Questionnaires

In a most entertaining TV episode of the BBC program “Yes Prime Minister,” Sir Humphrey Appleby demonstrates the use of leading questions to skew an opinion survey to support or oppose national service (military conscription) (**https://www.youtube.com/watch?v=G0ZZJXw4MTA&list=RDG0ZZJXw4MTA&t=21**). This stunning episode shows the importance of questionnaire design and the potential bias that can result from poorly designed questionnaires. In the present article we do not expand on this topic and refer interested readers to Kenett & Salini (2011) section 7.3.3, the Food and Agriculture Organization of the United Nations online document on questionnaire design (**http://www.fao.org/docrep/w3241e/w3241e05.htm**), and the Harvard University Questionnaire Design Tip Sheet (**http://psr.iq.harvard.edu/book/questionnaire-design-tip-sheet**). See also Smith (1978).

4.6. Late Swings

The 2015 Canadian election is an excellent example of a late swing. The shift toward Trudeau began several weeks before the election, giving the polls enough time to identify it, which stood out as one of the few successes of recent election surveys. Swings might reflect a true change in opinion, with voters switching their support from one candidate to another, or they might reflect differential support among those who were undecided until close to election day.

Sometimes swings occur in the last days before an election. As surveys cannot always be completed in a tight time frame, they may miss these swings. The problem is intensified in countries that prohibit publication of survey results in the final days preceding an election. For example, Israel prohibits publication of survey results in the last four days before an election. Fuchs (2017) reported that surveys taken during those days in 2015 already captured the late swing toward the Likud party. Conversely, Sturgis et al. (2016) concluded that in the 2015 UK election, late swings could not explain the discrepancies between the surveys and the actual vote. They based their conclusion mostly on the comparison (reported above) between post- and preelection polls.

5. AN INFORMATION QUALITY FRAMEWORK FOR STUDYING ELECTION SURVEYS

Election surveys and any other types of applied research are conducted to provide information. A fundamental question underlying such studies is therefore the InfoQ provided by them. In this section, we describe the InfoQ framework proposed in Kenett & Shmueli (2014) to answer such questions and explain its relevance to election surveys. We begin with discussion of the four InfoQ components—goals, utility, data, and analytic methods—and proceed with the eight InfoQ dimensions. Throughout this section, we discuss the InfoQ dimensions with a focus on election surveys.

5.1. The Information Quality Components

Kenett & Shmueli (2014, 2016) define the concept of InfoQ as the potential of a data set to achieve a specific goal (scientific or practical) when using a given empirical analysis method. InfoQ is different from data quality and analytic quality, but it is dependent on these components and on the relationship between them. The technical definition of InfoQ is the derived utility (U) from an application of a statistical or data analytic model (f) to a data set (X), given the research goal (g). This can be written in symbols as $\text{InfoQ}(f, X, g) = U[f(X|g)]$.

A key requirement for determining InfoQ is the nature of the goal of the study. In particular, we distinguish between explanatory, predictive, and descriptive goals. An explanatory goal is based on causal hypotheses or seeks causal answers. For example, "do higher income and education characterize voters of the democratic party in the United States?" A predictive goal is aimed at predicting future or new individual observations. For example, "predict the impact of a specific election campaign on voters of a given income." A descriptive goal aims to quantify an observed effect, using a statistical definition or another device. For example, "how do income and education levels correlate with the level of support for candidate X?"

5.1.1. Goals (g) in election survey analysis. Election surveys or election polls are of interest to many different stakeholders, so unsurprisingly, a variety of goals are relevant. These goals include:

- Giving the media news to publish
- Providing information on current preferences
- Providing a basis for accurate election predictions
- Setting up political campaign goals
- Deciding where, when, and how to launch specific campaign initiatives
- Identifying the drivers of electorate preferences
- Detecting positive or negative trends in voters' attitudes
- Highlighting good practices by comparing the impact of several campaign methods
- Generating reliable data for research and analysis of voter behavior

5.1.2. Utility (U) in election surveys. The common utility in election surveys seems to be prediction accuracy, which touches on all of the goals listed above. However, as with survey goals, utility varies across different stakeholders. For the public, utility is tied closely to the accuracy of the survey's results. The results can affect decisions about whom to support, whether to vote in line with one's own political alignment or to cast a tactical vote (e.g., in proportional representation systems, to vote for a small party in danger of not passing the lower threshold for gaining seats), or even whether or not to vote altogether. Accurate survey results are essential to making informed decisions. Candidates and political parties are interested in identifying issues that might increase their support and subgroups of voters that should be targeted in the campaign. The media wants news, which might lead to high utility for surveys that deviate from the consensus and provide provocative results that make headlines. At the same time, the media have an interest in reputation, so accuracy is also a major component of utility. Thus, all stakeholders have an interest in controlling total survey error over all its constituent parts. The need to obtain accurate estimates and the need to minimize sources of error are of course common to all statistical activity; they are not limited just to surveys.

5.1.3. Data (X) in election surveys. The different methods for collecting data in election surveys are described in Section 3.1. The sampling design underlying the survey determines in many ways the data collected. We described in Section 3.2 the severe problems of bias that can arise when samples are not representative and were collected in a way that does not permit appropriate adjustments by the methods described in Section 3.2. Their representativeness begins with the sampling design. Individuals who are contacted in an election survey are requested to express their political preference and attitudes about a range of issues. This is achieved by questionnaires ranging from approximately 5 to 100 questions. Different measurements of voter intentions are used to define the construct known as voting preference. The response (dependent) variables in voting intention models are typically expressed on an anchored scale that shows alternative voting options and strength of attitude. Bad questionnaire design and possible interviewer effects can also lead to biased results. As an example, consider the following question in a poll conducted by the Trump administration in the United States, not long after taking office: "Do you believe that the media unfairly reported on President Trump's executive order, temporarily restricting people entering our country from nations compromised by radical Islamic terrorism?" Describing the countries as suffering from "radical Islamic terrorism" is an obvious choice to invoke a positive response to this question about media coverage. (The full survey is available at **https://action.donaldjtrump.com/mainstream-media-accountability-survey/?utm_medium=email&utm_campaign=GOP_surveys_Mainstream-Media-Accountability-Survey&utm_content=021717-media-survey-fwd-djt-jfc-p-p-hf-e-1&utm_source=e_p-p**.)

5.1.4. Analysis (f): models for election survey data analysis. The analysis of election surveys can be based on a range of models, such as regression models, compositional models, or structural models (Kenett & Salini 2011). Section 3.2 discusses a wide range of issues related to election survey data analysis. An additional technical issue in the analysis of election surveys is the compositional structure of the collected data. Respondents provide preferences to a set of alternatives that represent all possible choices. The objective is to identify the relative proportion of choices in voter's preferences (Vives-Mestres et al. 2016). For an example of compositional analysis of election survey data, see Pawlowsky-Glahn & Buccianti (2011; section 2.2) on the November 2010 elections to the Parliament of Catalonia.

At a statistical strategic level, election surveys are designed to provide information to stakeholders such as candidates, political consultants, the media, and the public. Information generated by election surveys can be assessed with the InfoQ dimensions presented in the next subsection.

5.2. The Information Quality Dimensions

To assess the level of InfoQ in a particular study, Kenett & Shmueli (2014, 2016) propose eight dimensions of InfoQ. These dimensions are introduced below in the context of election surveys.

1. *Data resolution*. This dimension measures the adequacy of the measurement scale and level of aggregation of the data, relative to the study objectives. General survey methodology distinguishes between personal, household, and establishment surveys. Election surveys typically focus on capturing the personal opinion or attitudes of responders. The respondents' demographics such as voting location, socioeconomic status, and educational level determine the data resolution level. As noted in Section 3.6, predictive analytics requires higher-resolution data than those obtained in most election surveys conducted for media presentation.

2. *Data structure.* The analyzed data can consist of quantitative data and/or verbal textual data. Both types of data can be obtained by asking respondents to answer closed list questions and to provide open comments. In addition, survey data can be combined with social media data and past election results.
3. *Data integration.* Data is often spread out across multiple data sources. Hence, properly identifying the different relevant sources, collecting the relevant data, and effectively integrating them improves the InfoQ. For example, combining a range of surveys carried out by different polling organizations has become popular in the United States, although this approach is not always applicable (see Section 3.3). Predictive analytics may combine profiling data from a small to moderate survey of voters with social media data from a large group of voters, with the goal of using the social media data to infer the likely profile of the many voters who were not interviewed in the survey.
4. *Temporal relevance.* Data sets contain information that relates to a certain time window. The degree of relevance of that time window to the current goal must be assessed. For example, we could consider if patterns evaluated in a past election are still relevant in a current context or, in other words, if they are temporally relevant. Time effects can be incorporated in models analyzing election surveys, such that temporal relevance, as reflected by the data, can be fully represented. Temporal relevance is a characteristic of the data used in the analysis.
5. *Chronology of data and goal.* The chronology of data availability and analysis is a further component related to InfoQ. An election survey that requires one week to carry out and analyze may well be obsolete before it is even released. The need for immediacy is common to all the different stakeholders in surveys. Political campaigns are often driven by real-time decisions even at the expense of accuracy. An example of this is provided by recommender systems that provide a list of suggested alternatives. The immediacy of the display supersedes the accuracy of the recommendation. The message has to reach the right person at the right time, even if a more precise message can be generated with additional time. If the "temporal relevance" dimension characterizes the data at hand from a temporal perspective, then "chronology of data and goal" considers the operationalization derived from the data analysis with respect to the decision maker's needs.
6. *Generalizability.* Two types of generalizability can be considered: statistical generalizability and scientific generalizability. Statistical generalizability refers to the process of inferring from a sample to a target population. In election surveys, there is always a need to extrapolate from the sample to the voting population; thus, statistical generalizability is a constant concern. The sampling and analysis methods we reviewed in Section 3 are all dedicated to making an accurate leap from the sample data to the population. An important goal of statistical generalizability is to accurately quantify the level of uncertainty in the survey estimates. One of the clear problems with the US, UK, and Israeli election polls was their failure to provide (and to highlight) good assessments of uncertainty. In fact, the survey reports suggested overconfidence in the results. It is especially difficult to quantify the uncertainty related to sources of bias and more research on this problem is needed. Scientific generalizability refers to the application of a model developed for a particular target population to other populations. This can mean either generalizing an estimated population pattern or model to other populations, or applying the model developed for one population to predict individual observations in other populations. Issues of scientific generalization are relevant to a number of aspects of election surveys. First, they are fundamental to developing prediction models, which may exploit broad social and political trends that emerge from survey results. Norpoth (2016) developed such a model that predicted a Trump victory early in the

recent US election campaign. Second, they assist in drawing conclusions about the meaning of survey results, which is of major interest to campaign strategists, for whom it can suggest strategies, and to academic researchers. Recent publications have called for nonquantitative, intuitive-based methods to be applied as an alternative to election surveys (MSNBC 2016). With this approach, generalization of findings relies on context-specific knowledge like the so-called journalistic approach of the French newspaper *Le Parisien Aujourd'hui en France*.

7. *Operationalization*. We distinguish between construct operationalization and action operationalization. The choice of observable data represents a construct operationalization of underlying attitudes and positions. Voting preferences can be assessed via a questionnaire, by tracking attendance in political rallies, or by ratings for partisan TV chains. A major challenge is to identify the position of undecided voters. In social media contexts, covariate data can provide some estimates of voters' positions, whether or not they declare their preferences. Action operationalization, however, assesses the concrete actions that can be derived from the information provided from a study. In election surveys, linking covariates to voter responses can guide focused interventions in political campaigns.
8. *Communication*. Clear and timely communication of information is essential for achieving high InfoQ. Data visualization is important for good communication and is therefore directly related to the quality of the information. Poor visualization of findings can lead to degradation of the quality of information contained in the analysis performed on the data. Analysis of election survey data provides estimates that need to be interpreted. The survey analyst needs to create explicit statements on the basis of these estimates and ensure that they are understood correctly. In many cases, the analysis is supplemented by graphs and other visualization methods. Another important issue in communicating election survey results is the report of uncertainty. As we already stated, this is a notoriously difficult task. Some organizations, such as the New York Times, are careful to include a statement about the uncertainty associated with their survey results. Often, though, results are published as though there is no uncertainty at all. This practice damages the image of statistics as a scientific discipline. The recent use of prediction models has contributed significantly to public awareness of the uncertainty associated with election survey results. The probabilistic nature of these predictions has led the media to publish histograms of possible results (generated via repeated simulations of a generative model), not just verbal descriptions of uncertainty. The picture of uncertainty conveyed by such histograms can help readers to understand that, for example, a 70% chance of winning includes many possible outcomes in which the favored candidate comes out on the short end of the vote count.

In summary, one can compare and benchmark election surveys by computing an overall score that relates to all eight dimensions. See Kenett & Shmueli (2014, 2016) for details. The book by Kenett & Shmueli (2016) includes examples of InfoQ score assessments in education, customer surveys, health care, risk management, and official statistics. We emphasize next the important dimension of data integration.

5.3. Integration of Different Data Sources: The Third Information Quality Dimension

Election surveys are performed in different ways, employing different data sources. The main methods for conducting surveys include mail, telephone interviews, face-to-face interviews, online responses, and social media data mining. In principle, one could combine different surveys, which possibly use different modes, to collect responses into an integrated survey before analyzing the data. An example could be the combination of online and telephone surveys (Fricker et al. 2005). In fact, survey sampling applications often offer respondents different modes of response for their choice, or a single mode is initially offered to, say, respond via the Internet and then telephone calls are made to nonresponding sampled units (Pfeffermann 2015). Offering respondents a choice of response modes is supposed to increase rates of response. The prediction models described in Section 3.3 present an example of data integration in election surveys. A particularly promising area of research is the development of methodologies combining questionnaire-based surveys with data such as sentiment analysis indicators derived from social media.

Another area for further research is the combination of administrative data with survey data. Dalla Valle & Kenett (2015) propose a methodology for increasing InfoQ via integration of past survey with more recent surveys or administrative information, thus enhancing temporal relevance. The idea is in the same spirit of external benchmarking used in small area estimation (Pfeffermann 2013). In small area estimation, benchmarking makes the inference robust by forcing the model-based predictors in small areas to agree either with known population totals or proportions known from administrative data or with estimators obtained from a large independent sample. The calibration methodology by Dalla Valle & Kenett (2015) is based on qualitative data calibration performed by conditioning graphical models, in which past survey estimates are updated to agree with more recent surveys or concurrent administrative data. The calibration methodology is structured in three phases. Phase one consists of a multivariate data analysis of past survey results and new surveys or administrative data sets. Graphical models such as vines and Bayesian networks are used to describe the dependence structure among the variables and then to calibrate past data with recent data. In the second phase, common correlated variables are identified between the past and current data sets. In the last phase, the past data are updated via a calibration procedure, in which the Bayesian network of both data sets is conditioned on specific target variables. (For application examples and more details see chapter 10 in Kenett & Shmueli 2016).

A recent report of the US National Academy (Natl. Acad. Sci. Eng. Med. 2017) describes the use of multiple data sources in official statistics and methods to protect privacy when data are merged. Lohr & Raghunathan (2017) provide a thorough review and discussion of methods for combining survey data with data from administrative and other sources. In general, there is growing emphasis on the importance of fusing data sources in official statistics and thereby enhancing InfoQ.

5.4. Communication of Survey Polls: The Eighth Information Quality Dimension

Communication of results from election surveys carries an essential element of InfoQ. We noted earlier the importance of accurately quantifying the level of uncertainty underlying the results of a survey or prediction model. Equally important is to communicate that uncertainty to those who read or use the results. For example, in the US election of 2016, the consensus in the surveys was that Clinton would win the popular vote by approximately 3–4%, and in the final count, she won by 2%. That is within the level of uncertainty of the polls, but perhaps that uncertainty was not conveyed clearly to the public. Moreover, the public often judges the accuracy of the surveys by the single question: Did they get the winner right? This ignores the issue of uncertainty altogether and suggests that the public has not been trained by those reporting survey results to understand that there is always an element of uncertainty. We believe that presenting the graphical displays of potential outcomes that accompany prediction models (but not the surveys themselves) is an important step toward bridging this communications gap.

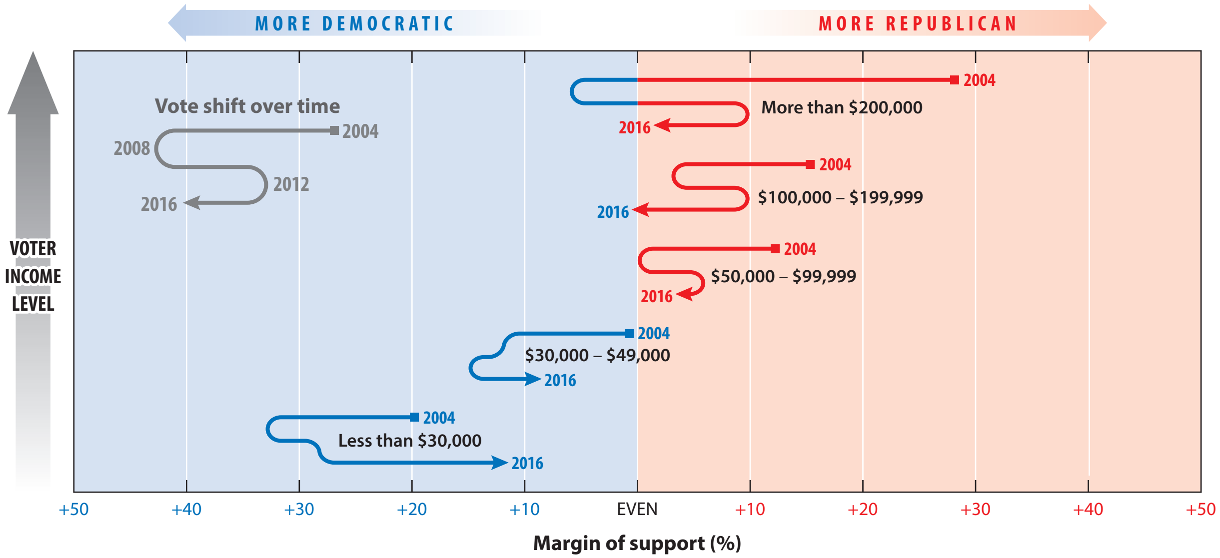

Figure 1

Vote shift over time. Effective graphs and proper communication of survey polls are obviously essential elements in determining the information quality of survey polls.

Many types of graphical displays can be used in communicating survey results. For examples of election maps see the website of Edward Tufte (**https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001AJ**). Political data is increasingly plotted and shared in the media. Online tools ranging from NationMaster.com to the NameVoyager (**http://www.babynamewizard.com/voyager**) are becoming increasingly accessible, with data dumps such as Hans Rosling's TED talk (**http://www.ted.com/index.php/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html**) becoming cult favorites.

The last 30 years have seen the development of a set of principles for sound graphical displays based on solid scientific research and experimentation (Tufte 1983, Cleveland 1994, Gelman et al. 2002). An important argument made by these authors is that an effective graph construction is designed to answer this question: How does the choice of graph affect the information perceived by the recipient of the graph? For many graphs, rearranging the values in decreasing or increasing order provides greatly enhanced pattern recognition. Moreover, two very commonly used displays, pie charts and divided bar charts, typically do a poor job of revealing patterns.

As a constructive example, we refer to snake-looking plots similar to those presented in Lai et al. (2016). These charts provide a graph of vote shifts over time. For additional data sources, see ACE Elect. Knowl. Netw. (2006). The graph in **Figure 1** shows the level of support for the Democratic and Republican Parties in the United States during the elections of 2004, 2008, 2012, and 2016 by level of income. The far ends of the graph indicate high levels of support, and the center vertical line represents an even level of support. In the lowest income group, one observes higher support for Democrats. In 2008 and 2012, the level of support for Democrats among those who earned less than $30,000 a year was approximately 30%. In 2016, it dropped to approximately 10%. In 2004, high earners with annual incomes of more than $200,000 a year favored Republicans by a margin of approximately 30%. This margin dropped to approximately zero in 2016. The plot shows voting dimensions and income level effects. Presenting election trends as done in **Figure 1** is highly informative.

6. FUTURE OUTLOOK

In this future-looking section we expand on two topics: The first concerns predictive methods, and the second is about the interplay between statistical and analytic methods with political science theories and tools. We conclude with a list of areas where election surveys seem to be heading.

The 2008 US presidential election marked a turning point in the use of election surveys as a tool to predict election outcomes. The striking success of fivethirtyeight.com in predicting the results of that election, followed by successful predictions in 2012, established prediction models as a natural companion to individual surveys. These models combine data from many different polls rather than summarize the results of a single survey. Further work will undoubtedly be carried out to elaborate and improve these models for future elections. The surveys are essential inputs, so the prediction models will fail if all the surveys are biased in the same direction, as happened in the UK election of 2015 and the US election of 2016.

We described in Section 3.4 the significant role that predictive analytics has assumed in the last decade in election campaigns. These methods use data from a variety of sources, including social media, to facilitate tailored messages aimed at targeted groups. Detailed survey data are important inputs to predictive analytics. We envisage a growing application of such methods with strong involvement of data scientists. In turn, the data needed for these models will be powerful drivers for future election surveys.

Potential errors in election surveys need to be better addressed at the design stage. Section 4 lists such errors. Properly addressing them requires the knowledge and skills of social and political scientists working in close collaboration with statisticians. We focus in this article on sources of bias in survey elections. More work is needed on how to reduce the bias and how to quantify the uncertainty resulting from it. As discussed in Section 5.4, it is equally important to effectively communicate to the public and decision makers the extent of the uncertainty.

Finally, we emphasize the need for a comprehensive approach to InfoQ beyond the technical domain of classical statistical survey methods. Considerations of data integration, chronology of data and goals, construct operationalization, generalization, and communication are essential dimensions in election polls. We present these dimensions in Section 5 with specific considerations applicable to election surveys. These considerations are pertinent to both organizations commissioning surveys and organizations conducting them. The InfoQ dimensions can be used as a checklist to assess the InfoQ of specific surveys.

This article was designed as a critical review pointing at future trends in election survey methodology. Our intention was to provide a solid foundation for practitioners and researchers interested in this important domain. In a general sense, the article is relevant also to the public at large, and as such, we hope that it contributes to a better understanding of the potential for and limitations of election surveys as a source of information.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This article is dedicated to the late Steve Fienberg who encouraged us to write it and who has been a major influence to statistical methodology in general and particularly to the authors. He was both a friend and an infinite source of knowledge and experience and is dearly missed.

LITERATURE CITED

AAPOR (Am. Assoc. Public Opin. Res.). 2010. Research synthesis: AAPOR report on online panels. *Pub. Opin. Q.* 74(4):711–81

ACE Elect. Knowl. Netw. 2006. *Comparative Data Summary*. **http://aceproject.org/epic-en**

Anderson B, Silver B, Abramson P. 1988. The effects of the race of the interviewer on race-related attitudes of black respondents in SRC/CPS national election studies. *Pub. Opin. Q.* 52(3)289–324

Asher H. 2007. *Polling and the Public: What Every Citizen Should Know*. Washington, DC: CQ Press. 7th ed.

Bialik C, Enten H. 2016. The polls missed Trump. We asked pollsters why. *FiveThirtyEight*, Nov. 9. **http://fivethirtyeight.com/features/the-polls-missed-trump-we-asked-pollsters-why/**

Blumenthal MM. 2005. Toward an open-source methodology: what we can learn from the blogosphere. *Pub. Opin. Q.* 69(5):655–69

Ceron A, Curini L, Iacus SM. 2014. *Social Media e Sentiment Analysis: L'evoluzione dei fenomeni sociali attraverso la Rete*. Sxl—Springer for Innovation Ser. 9. New York: Springer

Ceron A, Curini L, Iacus SM. 2015. iSA: a fast, scalable and accurate algorithm for sentiment analysis of social media content. *Inf. Sci.* 368(1):105–24

Chambers R. 1999. Probability sampling. In *Model Quality Report in Business Statistics*, ed. P Davies, P Smith, pp. 7–39. Newport, UK: Off. Natl. Stat. **https://users.soe.ucsc.edu/~draper/bergdahl-etal-1999-v1.pdf**

Cleveland WS. 1994. *The Elements of Graphing Data*. Summit, NJ: Hobart. 2nd ed.

Confessore N, Hakim D. 2017. Data firm says 'secret sauce' aided Trump; many scoff. *New York Times*, Mar. 6, p. A1. **https://www.nytimes.com/2017/03/06/us/politics/cambridge-analytica.html?referer=https://t.co/7VjOulz9Q5&_r=1**

Curini L, Iacus SM, Canova L. 2015. Measuring idiosyncratic happiness through the analysis of twitter: an application to the Italian case. *Soc. Indic. Res.* 121(2):525–42

Dalla Valle L, Kenett RS. 2015. Official statistics data integration to enhance information quality. *Qual. Reliab. Eng. Int.* 31(7):1281–300

Daves RP. 2000. Who will vote? Ascertaining likelihood to vote and modeling a probable electorate in pre-election polls. In *Election Polls, the News Media and Democracy*, ed. P Lavrakas, M Traugott, pp. 205–23. New York: Chatham House

Deville JC, Särndal CE. 1992. Calibration estimators in survey sampling. *J. Am. Stat. Assoc.* 87:376–82

Dillman DA. 1978. *Mail and Telephone Surveys*. New York: Wiley

Dillman DA. 2000. *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley. 2nd ed.

Erikson RS, Panagopoulos C, Wlezien C. 2004. Likely (and unlikely) voters and the assessment of campaign dynamics. *Pub. Opin. Q.* 68(4):588–601

Erikson RS, Sigelman L. 1995. Poll-based forecasts of midterm congressional elections: Do the pollsters get it right? *Pub. Opin. Q.* 59:589–605

Farmer B. 2015. Independent inquiry announced into what went wrong with election polls. *The Telegraph*. May 8. **http://www.telegraph.co.uk/news/general-election-2015/11592840/Independent-inquiry-announced-into-what-went-wrong-with-election-polls.html**

Frankovic K, Panagopoulos C, Shapiro RY. 2009. Opinion and election polls. In *Handbook of Statistics: Sample Surveys: Inference and Analysis*, ed. D Pfeffermann, CR Rao, pp. 567–95. Amsterdam: North-Holland

Fricker S, Galesic M, Tourangeau R, Yan T. 2005. An experimental comparison of web and telephone surveys. *Pub. Opin. Q.* 69(3):370–92

Fuchs C. 2017. Surveys and election forecasts in a world of social media and party dealignment. In *The Elections in Israel 2015*, ed. M Shamir, G Rahat, pp. 213–36. New Brunswick, NJ: Transaction

Gelman A, Pasarica C, Dodhia R. 2002. Let's practice what we preach: turning tables into graphs. *Am. Stat.* 56:121–30

Grajales C. 2015. So what can we learn from Nate Silver's mistakes? *StatisticsViews*. May 14. **http://www.statisticsviews.com/details/news/7945151/So-what-can-we-learn-from-Nate-Silvers-mistakes.html**

Groves RM, Lyberg L. 2010. Total survey error: past, present and future. *Pub. Opin. Q.* 74:849–79

Heckman JJ. 1979. Sample selection bias as a specification error. *Econometrica* 47:153–61

Kellner P. 2015. We got it wrong. Why? *YouGov UK*, May 11. **https://yougov.co.uk/news/2015/05/11/we-got-it-wrong-why/**

Kenett RS. 1991. Two methods for comparing Pareto charts. *J. Qual. Technol.* 23:27–31

Kenett RS, Salini S. 2011. *Modern Analysis of Customer Satisfaction Surveys: With Applications Using R*. Chichester, UK: John Wiley

Kenett RS, Shmueli G. 2014. On information quality. *J. R. Stat. Soc. A* 177:3–38

Kenett RS, Shmueli G. 2016. *Information Quality: The Potential of Data and Analytics to Generate Knowledge*. Chichester, UK: John Wiley

Kohavi R, Longbotham R, Sommerfield D, Henne RM. 2009. Controlled experiments on the web: survey and practical guide. *Data Min. Knowl. Discov.* 18:140–81

Lai KKR, Parlapiani A, White J, Yourish K. 2016. How Trump won the election according to exit polls. *New York Times*, Nov. 8. **http://www.nytimes.com/interactive/2016/11/08/us/elections/exit-poll-analysis.html**

Lauderdale B. 2015. What we got wrong in our 2015 U.K. general election model. *FiveThirtyEight*, May 8. **https://fivethirtyeight.com/datalab/what-we-got-wrong-in-our-2015-uk-general-election-model/**

Lee S, Valliant R. 2009. Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociol. Methods Res.* 37:319–43

Lohr SL, Raghunathan TE. 2017. Combining survey data with other data sources. *Stat. Sci.* 32(2):293–312

McAllister I, Studlar DT. 1991. Bandwagon, underdog, or projection? Opinion polls and electoral choice in Britain, 1979–1987. *J. Polit.* 53:720–41

Mercer A, Deane C, McGeeney K. 2016. Why 2016 election polls missed their mark. *Fact Tank*, Nov. 9. **http://www.pewresearch.org/fact-tank/2016/11/09/why-2016-election-polls-missed-their-mark**

Mosteller F, Hyman H, McCarthy PJ, Marks ES, Truman DB. 1949. *The Pre-Election Polls of 1948*. New York: Soc. Sci. Res. Counc.

MSNBC. 2016. Michael Moore joins wide-ranging election talk. *Morning Joe*, Nov. 11. **http://www.msnbc.com/morning-joe/watch/michael-moore-joins-wide-ranging-election-talk-806604867876**

Nathan G. 2001. Telesurvey methodologies for household surveys—a review and some thoughts on the future. *Surv. Methodol.* 27:7–37

Natl. Acad. Sci. Eng. Med. 2017. *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*. Washington, DC: Natl. Acad. Press

Noelle-Neumann E. 1993. *The Spiral of Silence: Public Opinion—Our Social Skin*. Chicago: Univ. Chicago Press. 2nd ed.

Norpoth H. 2016. Primary model predicts Trump victory. *The Primary Model*, Oct. 2016. **https://www.primarymodel.com/2016-forecast-full/**

Panagakis N. 1989. Incumbent races: closer than they appear. *The Polling Report*, Febr. 27. **http://www.pollingreport.com/incumbent.htm**

Pawlowsky-Glahn V, Buccianti A. 2011. *Compositional Data Analysis: Theory and Applications*. Chichester, UK: John Wiley

Pfeffermann D. 2013. New important developments in small area estimation. *Stat. Sci.* 28:40–68

Pfeffermann D. 2015. Methodological issues and challenges in the production of official statistics: 24th Annual Morris Hansen Lecture. *J. Surv. Stat. Methodol.* 3(4):425–83

Pfeffermann D, Landsman V. 2011. Are private schools really better than public schools? Assessment by methods for observational studies. *Ann. Appl. Stat.* 5:1726–51

Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55

Rubin DB. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley

Saris W. 1998. Ten years of interviewing without interviewers: the telepanel. In *Computer Assisted Survey Information Collection*, ed. NP Couper, RP Baker, J Bethlehem, CZF Clark, J Martin, et al., pp. 409–31. New York: Wiley

Schonlau M, Couper MP. 2017. Options for conducting web surveys. *Stat. Sci.* 32(2):279–92

Shmueli G, Bruce PC, Stephens ML, Patel NR. 2017. *Data Mining for Business Analytics*. Hoboken, NJ: John Wiley

Silver NA. 2012. *The Signal and the Noise: Why So Many Predictions Fail but Some Don't*. New York: Penguin

Simon HA. 1954. Bandwagon and underdog effects and the possibility of election predictions. *Pub. Opin. Q.* 18:245–53

Smith TW. 1978. In search of house effects: a comparison of responses to various questions by different survey organization. *Pub. Opin. Q.* 42:443–63

Smith TW. 1990. The first straw: a study of the origins of election polls. *Pub. Opin. Q.* 54:21–36

Sturgis P, Baker N, Callegaro M, Fisher S, Green J, et al. 2016. *Report of the inquiry into the 2015 British general election opinion polls*. London: Br. Poll. Counc. and Mark. Res. Soc. **http://eprints.ncrm.ac.uk/3789/1/Report_final_revised.pdf**

Traugott MW, Lavrakas PJ. 2008. *The Voter's Guide to Election Polls*. Lanham, MD: Rowman and Littlefield. 4th ed.

Tufte ER. 1983. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press

Visser P, Krosnick J, Marquette J, Curtin M. 2000. Improving election forecasting: allocation of undecided respondents, identification of likely voters and response order effects. In *Election Polls, the News Media and Democracy*, ed. P Lavrakas, M Traugott, pp. 224–60. New York: Chatham House

Vives-Mestres M, Martín-Fernández J-A, Kenett RS. 2016. Compositional data methods in customer survey analysis. *Qual. Reliab. Eng. Int.* 32(6):2115–25

Washington Post. 2004. Daily tracking poll methodology. *Washington Post*, Nov. 1. **http://www.washingtonpost.com/wp-dyn/articles/A9363-2004Oct5.html**

Yahav I, Shmueli G, Mani D. 2016. A tree-based approach for addressing self-selection in impact studies with big data. *Man. Inf. Syst. Q.* 40(4):819–48