

Collaborative Research: The American Mass Public in The 1930s and 1940s
Adam Berinsky, Massachusetts Institute of Technology
Eric Schickler, Harvard University

Intellectual Merit

The study of individual-level public opinion and behavior has flourished in recent decades. But our understanding of the dynamics of mass opinion prior to the 1950s has been undermined by the absence of high-quality, individual-level data. Over 450 national opinion polls were conducted from 1935-1945. However, these surveys have not been exploited by the social science community because they are not easily usable. The data often contains numerous miscodings and other errors. In addition, the surveys employed now-discredited quota sampling procedures. The principal investigators will recode the datasets and implement and disseminate methods appropriate for analyzing this historical public opinion data. By applying methods that make quota sample data usable, this project will open up a new realm for public opinion research in the U.S. and other countries, such as England, that have used quota-sampling methods.

Recovering this early opinion data promises to illuminate critical questions concerning the role of the mass public in the political system. The public opinion polls from this time contain a plethora of valuable items measuring attitudes concerning economic policy, World War II, and racial relations. The principal investigators will use the polling data on attitudes toward international intervention to determine how the mass public guided and reacted to foreign policy decisions during the Second World War, thereby expanding the systematic study of public opinion and war to include the most significant international conflict in American history. The principal investigators will also use the opinion data on domestic policy as a resource for examining the relationship between the mass public and both the programmatic accomplishments and limitations of the New Deal.

Broader Impacts Resulting from the Proposed Activity

This project will make available to the social science research community a trove of public opinion data that has largely been ignored in the past. The principal investigators will compile and produce readily usable computer files for roughly 450 opinion polls undertaken in the United States from 1935 to 1945. These files are available through the Roper Center in Storrs, CT. However these data are almost impossible to use in their existing condition. The principal investigators will recode the individual files, prepare documentation, and make available a series of weights to mitigate the biases resulting from quota-control sampling. The investigators will also compile the individual polls into a series of cumulative datasets. These efforts will make the data more easily accessible to the larger social science research community.

The project also promises to expand the field of political behavior by promoting the study of historical public opinion. It will provide to the community of scholars a wealth of individual-level opinion data in the pre-1950 era, allowing researchers to gain new insights into an array of substantively important topics, such as changes in Americans' racial attitudes and public support for the early American welfare state.

The American Mass Public in The 1930s and 1940s project has four primary aims:

- To compile and produce readily usable computer files for roughly 450 public opinion polls undertaken from 1935 to 1945 by the four major survey organizations active during that period: George Gallup's American Institute of Public Opinion (AIPO); Roper; Hadley Cantril's Office of Public Opinion Research (OPOR); and the National Opinion Research Council (NORC). These files are available through the Roper Center. However these data are almost impossible to use in their existing condition. As a result, only a handful of scholars have used the data over the last 50 years. We will recode the files, compile the data into a series of cumulative datasets, and prepare documentation for the recoded files, thereby making this immense resource usable for the larger social science research community.
- To implement and disseminate methods appropriate for analyzing public opinion data collected using quota sampling methods. Though it has largely been discredited in the academic community, almost all of the public opinion data collected in the U.S. before 1950 used quota sampling procedures. By implementing methods that make quota sample data usable, this project will open up a new realm for public opinion research.
- To expand the field of political behavior by promoting the study of historical public opinion research. While scholars have dissected the wealth of opinion data collected in the U.S. since the 1950s, studies of mass political attitudes and behavior in earlier periods have been limited by the lack of high-quality, individual-level data. Our project will provide a wealth of individual-level opinion data in the pre-1950 era.
- To determine how the mass public guided and reacted to foreign policy decisions during World War II, thereby expanding the study of public opinion and war to include the most significant international conflict in American history. By examining a broader set of topics relating to public opinion and World War II, we will better understand the relationship between the mass public and government during wartime.

Introduction

The study of political behavior has flourished in recent decades with the development of long-term, time series data collected by both academic survey researchers and commercial pollsters. Our understanding of the dynamics of mass opinion and behavior prior to the 1950s has, however, been undermined by the absence of high-quality, individual-level data. As a result, scholars interested in such phenomena as the New Deal realignment, ethnic voting patterns, and the public's policy mood have been forced to rely primarily upon state and county-level election results. But such aggregate data cannot tell us why certain types of voters favored Democrats while others favored Republicans. The historical study of mass political behavior promises to enhance our understanding of the interplay between public opinion and elite policy-making. Fulfilling this promise requires capitalizing upon available individual-level data. This project will promote the historical study of political behavior by rendering usable a plethora of polling data from a critical era that has not yet been the subject of rigorous individual-level research.

Quota-Controlled Opinion Poll Data: Promise and Problems

The decade from 1935 to 1945 was like none other in American history. The Great Depression and World War II transformed American politics. The New Deal revolutionized the relationship between the federal government and its citizens, even as an emerging conservative coalition limited liberal policy innovations after 1937. In the foreign arena, U.S. involvement in World War II ended a long period of American isolationism and set the stage for the global

policies of the Cold War. The relationship of public opinion to government policy during these years is of considerable importance. Fortunately, there is a great deal of data concerning the public's views during this time. Starting in the mid-1930s, polling companies surveyed the public about important issues on a monthly basis.

These polls contain valuable questions concerning government policy. For example, in the two years preceding Pearl Harbor, Gallup repeatedly asked respondents if it was more important to help England or to stay out of the war. Opinion polls also illuminate the role played by the mass public in ending the rush of New Deal programs in the late 1930s. Starting in 1935, Gallup frequently asked respondents their opinions about government relief and regulation. Finally, pollsters asked numerous questions relating to racial politics, including school integration, employment practices, and biological racism. These data enable us to determine how racial attitudes have changed over the last sixty-five years.

Beyond the important cross-sectional analyses that these data make possible, several of the most important survey questions were repeated multiple times, allowing studies of change over time. For example, most Gallup polls asked about presidential vote intention and/or recent presidential vote choice, with little variation in question wording. In addition, Congressional vote was included in about one quarter of the polls. Presidential approval was included in at least 88 polls from 1935-45, again with little variation in wording. Gallup also asked a party identification question at least 39 times from 1937 to 1945. In our preliminary and (thus far) incomplete assessment of repeated questions, we identified 20 policy items that were repeated at least six times during this period, including questions concerning the war and labor policy. We summarize the repeated questions that we have identified thus far at <http://web.mit.edu/berinsky/www/surveyquestion.pdf>.

It is surprising, then, that these data have largely been overlooked by modern researchers. Some researchers – most notably Page and Shapiro (1992) – have used the aggregate poll data to study patterns of stability and change in public opinion. But this work is the exception. For example, Erikson, Mackuen, and Stimson's (2002) pathbreaking study of macropolitical trends begins in the early 1950s. Furthermore, contemporary studies of individual-level behavior using poll data collected before 1952 are rare (though see Baum and Kernell 2001; Caldeira 1987; Schlozman and Verba 1979; Verba and Schlozman 1977; Weatherford and Sereyev 2000).

One reason for this relative neglect is that early polls are extremely difficult to work with. Most of these surveys have not been touched for almost sixty years, and as a result, the datasets are not easily usable. The data contain numerous miscodings and other errors. For example, in one dataset from 1940 we ferreted out twenty different keypunch errors. In addition, some codebooks do not include the codes necessary to decipher important questions, such as the respondent's occupation and education. The short list of articles that have made use of the pre-1950s surveys attest to the problematic nature of the data in its current form. However, with a great deal of preparatory work, we have been able to make the data suitable for analysis. Moreover, by comparing and compiling variable codes across datasets we have been able to reconstruct the proper coding for almost all variables. A sample "before" and "after" codebook can be found at <http://web.mit.edu/berinsky/www/codebooks.html>.

Another reason political scientists do not use the data arises from the manner in which they were collected. Modern opinion polls are conducted using probability sampling to ensure that every citizen has an equal chance of being interviewed. However, polls in the U.S. before the 1950s were conducted using quota-controlled sampling methods, where pollsters sought to interview certain predetermined proportions of people from particular segments of the population

(see Berinsky 2005b for a description of the quota sampling practices). While some pollsters used quotas in seeking a descriptively representative group of citizens (Roper 1940), others designed quotas to produce sample proportions that differed systematically from the population. George Gallup was most interested in predicting elections, so he drew samples to represent each population segment in proportion to the votes it usually cast in elections. Because Southerners, African Americans, and women turned out at low rates in this period, these groups were deliberately underrepresented in opinion polls. For example, the 1940 Census found that 50 percent of the U.S. population was women, 10 percent was African American, and 31 percent lived in the South. By contrast, a December, 1940 Gallup poll included only 34 percent women, 3 percent African Americans, and 13 percent Southerners.¹ Thus, the Gallup data that scholars use to represent the voice of the mass public, in fact, comes from a skewed sample of that public.

The practice of quota sampling also introduced unintended distortions. Apart from having to fulfill certain demographic quotas, interviewers were given much discretion to select particular citizens to interview. Since interviewers preferred to work in safer areas and tended to survey approachable respondents, the “public” they interviewed often differed markedly from the public writ large. For example, the 1940 census indicated that about 10 percent of the population had at least some college education, while almost 30 percent of a typical 1940 Gallup sample had attended college. Similarly, polls conducted by Gallup and Roper tended to include more “professionals” than identified by the Census. The skew in these variables is not surprising, given that education and occupation were not quota categories. It is likely that the highly-educated and professionals were more willing to be interviewed, and, as a result, comprise a disproportionately large share in these samples.

Many political scientists have rejected out of hand polls conducted before 1950 because of their sampling procedure. For example, Converse (1965) concludes that the Gallup and Roper data “were collected by methods long since viewed as shoddy and unrepresentative.” Rivers argues that quota sampling is “a methodology that failed” (quoted in Schafer 1999). Nonetheless, while early opinion polls have substantial flaws, the critical information those polls contain should not be abandoned. Instead, we should recognize the inherent value of this data while taking into consideration the flaws of the collection procedures. In this project, we will “clean” the data to make it usable for researchers, and devise a system of weights, which we will use in our own analysis. We will also make these weights, and the auxiliary information used to create them, publicly available. These weights will allow us to make more accurate aggregate-level inferences about public opinion in this critical era.

A Method for Analyzing Surveys using Quota-Control Sampling

A major goal of our project is to implement and disseminate methods that directly address the biases introduced by quota-controlled sampling. The central problem is that many of the survey samples do not represent certain groups in proportion to their population share. But though the quota-controlled sample data were collected in ways that appear from a modern vantage point to be haphazard, the data collection process introduced *predictable* deviations between the characteristics of the sample and that of the population. We can therefore employ

¹ These figures are typical of the polls we have examined through the early 1940s. By the mid-1940s, however, Gallup adjusted his gender quotas to interview equal numbers of men and women. This change in the composition of the sample makes it difficult to track real changes in opinion over time.

methods designed to account for these measurable differences to make reasonable inferences.

A Weighting Solution

The quota-controlled sampling procedures introduced a unit non-response problem – certain classes of individuals were either willingly or inadvertently underrepresented in the samples. When we have detailed information about the characteristics of non-respondents, we can employ selection bias techniques to account for the differences between the sample and the population (Achen 1986; Breen 1996; Heckman 1979). But when we only have information about the population relative to the sample – auxiliary information taken from the census – weighting adjustments are typically applied to reduce the bias in survey estimates that non-response can cause (Holt and Eliot 1991; Lohr 1999; Kalton and Flores-Cervantes 2003).²

We employ a model-based poststratification weighting scheme.³ Under the model-based approach to sampling – which provides the foundation for our weighting strategy – the values of the variables of interest are considered to be random variables. Model-based approaches therefore involve employing a model of the joint probability distribution of the population variables (Thompson 2002). It is necessary to take a model-based approach to draw inferences from quota samples because, as Lohr (1999) notes, we do not know the probability with which individuals were sampled.

Though the use of weights to adjust for non-response is common, there is controversy about the best way to implement weighting (Lohr 1999). We therefore implement four solutions recommended by the survey weighting literature (see Bethlehem 2002; Deville and Sarndal 1992; Deville, Sarndal, and Sautory 1993; Kalton and Flores-Cervantes 2003; Little 1993; Gelman 2005; Gelman and Carlin 2002). These methods are cell-weighting, raking, regression estimation, and a regression modeling approach advanced by Gelman (2005). Each of these techniques has its own strengths and weaknesses, and we use all four methods here in order to gauge their robustness. Although we prefer cell weighting because it is simple to employ and requires minimal assumptions, other researchers may have other preferences. Thus, we will provide the auxiliary information necessary to employ each of the weighting methods.

Cell Weighting

Cell weighting is a simple way to bring the sample proportion in line with auxiliary information – namely census estimates of the population proportion. We stratify the sample into a number of cells (J), based on the characteristics of the population deemed important (the matrix of X variables). If the distribution of demographic variables in the sample differs from the distribution in the population, poststratification weights are used to combine the separate cell estimates into a population estimate by giving extra weight to groups underrepresented in the sample and less weight to overrepresented groups. Using cell weighting to adjust for non-response requires us to assume that the respondents within a given cell represent the non-respondents within that cell. That is, we must assume that the data is missing at random (MAR)

² The use of weighting adjustments can lower the precision of survey estimates because we trade off a reduction in bias for an increase in variance (Kalton and Flores-Cervantes 2003). Given the compositional imbalance in the quota samples, we believe that this is a worthwhile tradeoff.

³ The poststratification weights we employ are different from probability weights, which are known at the time the survey is designed and are used to adjust for non-constant probability of sample inclusion. It is not possible to employ probability weights for the quota samples we examine here.

(Little and Rubin 2002; we discuss possible violations of this assumption below). Under post stratification, the estimate of the population mean for our quantity of interest (for instance, support for FDR) is:

$$\hat{\theta} = \sum_{j=1}^J \frac{N_j}{N} \hat{\theta}_j$$

where N_j refers to the population size in each poststratification cell j , N is the total population size, $\hat{\theta}$ is the weighted estimate of the mean, and $\hat{\theta}_j$ is the sample mean for poststratification cell j (Lohr 1999).

A well-known practical limitation of cell weighting is that as the number of stratification variables increases, the number of weighting cells becomes larger. With fewer cases in each cell, the aggregate estimates derived from the weighting estimator are less stable. Still, cell weighting has advantages. First, as Lohr notes, cell weighting requires minimal assumptions beyond that the data is MAR (an assumption common to all forms of weighting, including those discussed below). For instance, as Kalton and Flores-Cervantes (2003) note, unlike other methods, cell weighting requires no assumptions regarding the structure of response probabilities across cells. In addition, cell weighting allows the researcher to take advantage of information concerning the joint distribution of weighting variables – information we have from census data.

Raking

While cell-by-cell weighting allows researchers to use only a limited amount of auxiliary data owing to sample size issues, raking – also known as iterative proportional fitting (Deming and Stephan 1940; Little and Wu 1991) or rim weighting (Sharot 1986) – allows researchers to incorporate auxiliary information on several dimensions. Raking matches cell counts to the *marginal* distributions of the variables used in the weighting scheme. For example, say we wish to weight by gender and a three-category age variable. Picture a three by two table, with age representing the rows and gender representing the columns. We have the marginal totals for each age group, as well as the proportion of men and women in the sample. The raking algorithm works by first matching the rows to the marginal distribution (in this case age), and then the columns (gender). This process is repeated until both the rows and columns match their marginal distributions (see Little and Wu 1991 for a description of the algorithm).

Raking allows for more weighting variables to be included than under cell weighting. Little (1993) demonstrates that if the data have an approximately additive structure, raking is an appropriate response to the small cell problem. However, raking ignores information available in the joint distribution of the weighting variables. Other methods are more appropriate if we expect there to be meaningful interactions among the weighting variables, a problem addressed by both cell weighting and the regression weighting methods we consider next.

Regression Estimation

An alternative method of including auxiliary information is regression estimation, which uses variables that are correlated with the variable of interest to improve the precision of the estimate of the mean (Lohr 1999). In this method, an ordinary least squares regression model is estimated using the sample data.⁴ The dependent variable is the quantity of interest – be it vote

⁴ Strictly speaking, since our dependent variables of interest are often binary (e.g. vote choice), using logit or probit would be more appropriate in those situations than OLS. But OLS allows us to interpret the poststratified estimate of the mean generated from regression estimation method

for the President or support for a given policy. The independent variables are the variables in the dataset that can be matched to measured quantities in the population (e.g., census data on age and gender). The population information is then used to estimate the population means of the dependent variable in the regression. Specifically, the mean population values of the independent variables are combined with the coefficients calculated from the sample model to derive fitted value estimates of the population mean of the quantity of interest.

$$\hat{\theta} = X_{pm} \hat{\beta}$$

Where X_{pm} is the vector of population mean values, and $\hat{\beta}$ is the least squares estimator $\hat{\beta} = (X'X)^{-1}X'y$. A disadvantage of regression weights is that a new set of weights must be calculated for each dependent variable analyzed.

Gelman Regression Model

The problem with classical linear regression is that the resulting weights often suffer from a large variability as more variables and interactions are added. We therefore turn to Gelman Hierarchical Regression Weighting. By theorizing a hierarchical model with a Bayesian approach, the Gelman method is less vulnerable to this traditional problem created by linear regression weighting schemes (see Gelman 2005 for a full description of this method).

To arrive at the Gelman regression weights, we transform the interaction indicator terms into multi-level regression terms. The model remains normal with $y \sim N(X\beta, \Sigma_y)$, where Σ_y is the variance of y , and uses a Normal prior distribution on β with mean of 0 and variance of Σ_β . The inverse of the variance of the prior is a diagonal matrix with the inverse variance of the individual X term for hierarchical terms and zero for all other terms (including the constant). The poststratified estimate is similar to the above for regression estimation:

$$\hat{\theta} = \frac{1}{N} (N^{pop})^t X^{pop} \hat{\beta}$$

Where $\hat{\beta}_j = (X^t \Sigma_y^{-1} X + \Sigma_\beta^{-1})^{-1} X^t \Sigma_y^{-1} y$, and N^{pop} is a vector of cell size in the population.

What to weight on? The art of selecting weighting variables

Poststratification weighting rests on a firm statistical foundation, but in practice requires a series of data analytic choices. All four weighting methods rely on auxiliary information to arrive at valid inferences. Kalton and Flores-Cervantes (2003) note that it is important to choose auxiliary variables that predict the response probabilities of the different cells. To make the case that our methods capture and correct for differences between the survey samples and the population, we discuss how auxiliary information from the census can be used to correct the problems introduced by quota sampling.

The distortions introduced by quota sampling can be divided into two types: “non-representative strata size” and “non-random selection within strata.” Non-representative strata

and the Gelman method (discussed below) as a weighted average of the data (see Gelman 2005). We believe this justifies using the linear model in order to ease the interpretation of our results.

size distortions arose through the instructions the central survey offices of the polling firms gave to the field staff -- for example, by setting the quota of female respondents below their true population proportion, the pollsters deliberately skewed their samples. By contrast, “non-random selection within strata distortions” are a result of interviewer discretion in respondent selection. This discretion ensured that the citizens who were interviewed differed in systematic ways from citizens who were not. Though these distortions arise through different processes, both can be addressed with a common solution. By employing auxiliary information about the population, we can correct for the known differences between the sample and the population.

The use of auxiliary information is an especially powerful way to correct for non-representative strata size distortions. There is no reason to suspect that the members of deliberately under-represented groups – such as women and Southerners – who were interviewed were systematically different from the members of those groups who were not interviewed (conditioning on the interviewer-induced differences that we address below). After all, the sample imbalance exists because the pollsters deliberately drew non-representative samples based on these characteristics. Thus, using the cases of the underrepresented groups who were interviewed to represent those respondents who were not interviewed is appropriate. We can simply use the observable characteristics of the respondents to reweight the data.

Auxiliary information can also be used to correct for distortions arising from non-random selection within strata. The latitude given to interviewers in selecting their subjects ensured that the probability of being interviewed depended on respondent characteristics that interviewers found attractive. Thus, within quota categories, those citizens who were interviewed were not necessarily representative of the population in that category, potentially violating the MAR assumption. Correcting for non-random selection within strata is difficult because we do not have information on the people who were not interviewed. However, we do sometimes have important information that can be employed in analysis – namely the education level of the respondent. While interviewers did not explicitly select respondents on the basis of their schooling, education is the best proxy the surveys have for the “observables” that made an interviewer more likely to pick one individual from a given demographic group than another individual. The key is that education: (1) is a powerful predictor of who is a desirable interview subject, (2) affects politically relevant variables, (3) was *not* used as a quota control, but (4) was often measured by survey organizations. Therefore, by utilizing auxiliary information on education levels, we can account for at least some of the problems introduced by non-random selection within strata. Although education measures were not included in every survey in this period, a respondent’s occupation can serve a similar purpose. For the same reason that interviewers gravitated to highly educated respondents, they tended to interview professionals at the expense of laborers and other members of the work force. In addition, occupation, like education, was not used as a firm quota variable. Thus, even when education is not available, we can correct for non-random selection within strata with auxiliary information on occupation.

The use of auxiliary data – particularly on education and occupation – to account for differences between the sample and the population is imperfect. It is possible that the low-education respondents selected by interviewers were not fully representative of the population of low education citizens. Thus, controlling for education does not completely solve the non-random selection within strata problem because the use of weights may multiply the influence of respondents who are “unusual.” However, education captures many of the important interviewer-induced differences between respondents and non-respondents. While quota-controlled sampling procedures reduced the probability that certain individuals – women, southerners, non-

professionals, and those with low education – would be interviewed, no individuals were excluded from the sampling scheme. Every individual therefore had some probability – no matter how low – of being included in the survey samples of this era. By using auxiliary information on education and occupation, we can take advantage of the residue of interviewer-induced distortions to correct at least some of the problems caused by non-random selection within strata. Controlling for some of the problem through the use of proxy variables, such as education, is preferable to completely ignoring the problem. In essence, by conditioning on the variables that affect the probability that a given individual would be interviewed, we can better fulfill the conditions required by the MAR assumption. Without detailed information on non-respondents, this strategy is the best solution available to modern researchers. The “check data” on vote for FDR and telephone use presented below suggests that our weighting system moves us closer to an accurate depiction of the full population.

In sum, the use of auxiliary information can mitigate the deficiencies of quota controlled sampling procedures. Thus, in aggregate analysis, we argue that researchers should weight the data on education levels and occupation and those quota category variables – such as gender, region, and age – that can be matched to census data. The necessary population counts for the 1940 census are available from the Integrated Public Use Microdata Series (Ruggles et al. 2004). Even when weighting makes only a modest difference in conclusions, it nonetheless provides more confidence that our estimates are not attributable to problematic sample design.

Results

The utility of the weighting strategy is demonstrated by calibrating the quota-control survey data to known quantities, notably election returns and telephone penetration. Because sampling error is a concern, it would be unrealistic to think that we could predict external quantities exactly with either the weighted or the unweighted estimates. But by examining the effect of the different weighting schemes on the estimates of known quantities, we can gauge how well the weighting process might bring us closer to the truth on opinion poll items.⁵

We first employed the four weighting schemes on Gallup’s post-election poll of FDR’s share of the two-party vote in the 1940 and 1944 elections.⁶ As Table 1 demonstrates, in both 1940 and 1944, Gallup underestimated FDR’s vote (we have two separate post-election polls for 1940). Using as our measure of error the difference in percentage points between the Democratic candidate’s share of the two-party vote in the poll and the election-day vote (Mosteller et al 1949, Mitofsky 1999), we find that Gallup had an average error of 2.7 percent. Though applying the different weighting schemes to this data provides somewhat different estimates, depending on the method used, the weighted estimates are roughly in the same range.⁷ The error rate ranges

⁵ It would, in theory, also be useful to calibrate the quota-sampling polls to probability sample polls in the late 1940s. Unfortunately, with one exception, we have not been able to find probability sampling and quota-sampling polls that targeted the same population during the same field period. Furthermore, in that one case – the 1948 NES and the Gallup election polls – there are no survey questions that overlap aside from the vote choice question. Because we can compare the Gallup numbers to election results, these data are not useful for calibration.

⁶ We do not employ data from 1936 because Gallup apparently supplemented his sampled polls with mail-back questionnaires (Erikson 1976). We are currently attempting to recover this data.

⁷ The cell weighting method uses census data on gender, region, and occupation (or education, when available). The other methods use these data as well as data on age.

from 0.4 percent using cell weights – our preferred method – to 1.7 percent using the Gelman regression weights. From this demonstration it appears that the cell weighted estimates get us closest to the final election results, but all four methods have a lower error rate than the unweighted estimates.⁸

Table 1: Aggregate Data Analysis

Poll	Two-Party Vote	Gallup Raw Data	Cell Weight	Raked	Regression Weight	Gelman Regression
Early Nov. 1940	55.0%	51.9%	54.1%	54.7%	54.4%	54.3%
Mid Nov. 1940	55.0%	52.0%	54.7%	56.4%	55.4%	56.2%
Nov. 1944	53.8%	51.8%	54.3%	55.3%	56.6%	56.9%

We can also use information about phone penetration to gauge the utility of weighting. According to Census Bureau estimates, 33 percent of American households had telephone service in 1936, a figure that rose to 46 percent by 1945. Polls conducted by Gallup routinely asked respondents if they had a telephone in their household. Given the class and education biases in the sampling procedure, it should not be a surprise that the phone penetration in the unweighted sample of respondents would be greater than the population at large. In fact, these unweighted estimates overestimate phone penetration rates by about 10 to 15 points. Across a variety of polls, applying the weights brings the estimate phone penetration closer to the true number (for a full description, see Berinsky 2005b). The analysis of the voting and telephone penetration data is especially important because it indicates that our statistical methods allow us to better estimate the true characteristics of the mass public.

Our analysis also indicates that weighting can change our view of collective public opinion on a number of issues important in the 1930s and 1940s (see Table 2). The January 1941 OPOR survey contained several questions concerning interest in and knowledge of world events. Given that the sample of respondents overrepresented the highly educated, men, and non-southerners – all groups that tended to be more involved with politics – it is not surprising that the unweighted marginals overstate the true political interest and knowledge of the American public relative to the weighted estimates. For instance, respondents were asked, “can you name a country where Greece and Italy are fighting?” In unweighted analysis, a majority gave the correct answer, Albania. But when the weights are introduced to correct for the sample selection and interviewer-induced biases, a minority gave the correct answer, no matter which method is used. Similar differences are found in the proportion of respondents who say they have been following Lend-Lease, and the proportion that correctly identify Hitler’s tenure in power at 8 years (see Table 2). Finally, sample biases can affect the shape of political attitudes. For example, an October 1944 OPOR survey asked respondents, “Do you think Russia can be trusted to cooperate with us when the war is over?” In the unweighted sample, 59% say they trust Russia, but applying the weights drops trust of Russia about three to five percentage points.⁹

⁸ Analysis of pre-election polls from this period yield similar results.

⁹ Weighting will have the greatest impact on aggregate public opinion estimates when: (1) the sample imbalances are related to the dependent variable of interest and (2) the biases induced by these imbalances work in the same direction (that is, there are no countervailing biases). In the information items discussed above, these conditions hold. But on policy items concerning war, weighting often has little effect on aggregate estimates because southerners (who tended to be hawkish) and women (who tended to be dovish) are each underrepresented in the Gallup quota

In sum, we recommend using weights to adjust the aggregate estimates of opinion from the 1930s and 1940s. Though we prefer to use the cell weights, as this section demonstrates, using different weighting schemes gives us similar pictures of public opinion.¹⁰ Researchers employing different weighting methods will come to similar pictures regarding the shape of public opinion from this era – conclusions that are sometimes different than those reached using the raw survey marginals. Thus, through the use of weights, we can have greater confidence that the inferences we draw about public opinion more accurately reflect underlying public sentiment.

Table 2: OPOR data

Poll	OPOR Raw Data	Cell Weight	Raked	Regression Weighting	Gelman Regression
Information Question					
Greece /Italy	54.7%	46.3%	44.1%	45.9%	45.5%
Follow LL?	72.1%	66.0%	66.0%	67.2%	66.9%
Hitler in Power?	47.3%	41.2%	41.7%	42.7%	43.3%
Political Question					
Trust Russia?	58.9%	55.9%	54.4%	56.4%	54.7%

Application to Individual Level Analysis

Thus far we have outlined a strategy to estimate aggregate statistics, such as population means. Nevertheless, we are also interested in estimating more complex relationships among the variables available in the data through individual level regression analysis. Whether to include weights in regression analysis is a source of ongoing controversy in the literature.¹¹ However, several authors caution against the use of such techniques (see Lohr 1999, pp. 362-365 for a review). This admonition is especially pertinent here because our weights are poststratification weights, not sampling weights. Winship and Radbill (1994) note that when weights are solely a function of observed independent variables that can be included in a regression model – as is the case with our data – unweighted OLS will yield unbiased parameter estimates. Thus, the most straightforward method of dealing with the potential bias created by quota sampling is simply to include the weighting variables as independent variables in the regression model (Gelman and Carlin 2002).¹² In this case, the problem is similar to omitted variable bias: the oversampling of certain types of respondents – namely highly educated white males – may mask the true relationship among other predictors if these variables are not controlled for. In this way, the

samples. Even in this case, however, the weighted opinion estimate is clearly preferable since it does not depend on such errors canceling one another out.

¹⁰ In typical applications, raking and regression estimation is preferred to cell weighting because a great deal of auxiliary information can be incorporated into the weights. In our case we only have a limited amount of auxiliary information, thus the gains from the alternative methods on this dimension are small.

¹¹ The concern here with weighted regression is distinct from the regression estimation methods discussed above.

¹² As Gelman counsels, “In a regression context, the analysis should include as ‘X variables,’ everything that affects sample selection or nonresponse. Or, to be realistic, all variables should be included that have an important effect on sampling or nonresponse, if they also are potentially predictive of the outcome of interest” (2005, p. 3).

individual and aggregate analysis are closely related, in that in order to get aggregate estimates, we average over the proportions of different types of respondents present in the population. Just as the cell weighting and the regression estimation methods incorporate information about the joint distribution of the sample, introducing the quota variables – and relevant interaction terms – as independent variables allows us to control for the sample imbalances introduced by the quota sampling methods of the time.

Democracies, Public Opinion, and War

To demonstrate the promise of this early opinion poll data, we describe an empirical project– presented in greater detail in Berinsky (2005a) – exploring how the mass public guided and reacted to foreign policy decisions during World War II. While a number of political scientists have conducted systematic studies of public opinion concerning international interventions (Foyle 1999; Holsti 1996, Mueller 1973, 1994; Sobel 2001), these analyses have focused on the Cold War and post-Cold War periods and almost completely ignored World War II. As shown below, including World War II in the scope of analysis could change our view of the relationship between the mass public and government.

Most scholars contend that the course of events in a conflict determines public support for war. The most prominent line of argument in this vein is what Burk (1999) calls the “casualties hypothesis.” The conventional wisdom that has emerged over the last 30 years – growing out of work by Mueller (1973) – holds that the American people will shirk from international involvement in the face of mounting war deaths. This theory has been modified to include local casualty rates (Gartner and Segrua 1997, 2000), cost-benefit calculations incorporating the perceived stakes and objectives of the conflict (Larson 1996, 2000), and the ongoing success of the mission (Kull and Ramsay 2001; Feaver and Gelpi 2004). These theories share the belief that “events” directly determine public support for war, arguing that the collective mass public is rational and will support war only if the perceived benefits of a successful outcome outweigh the costs of military action.

Though event-based accounts of support for war have made important contributions, these theories have several potentially serious conceptual problems. Existing theories rest on tenuous notions of collective rationality. Larson, for instance, argues that the aggregate mass public will support war “if the aims are clear,” but he does not describe the conditions under which individuals, much less the aggregate public, make such complex calculations. Without individual-level analysis, it is impossible to specify the mechanism by which the mass public processes information concerning war. Second, almost all the work described above completely ignores the political process. Treating the mass public as an undifferentiated whole, innocent of political attachments, leaves no room for the effect of domestic politics. But, popular perceptions notwithstanding, politics has never stopped at the water’s edge. Foreign policy is often as contentious and partisan as domestic politics. Theories of war and politics must account for the effects of the domestic political process (Baum and Groelling 2004; Schultz 2001)

Elite Conflict and Mediated Reality

Event-driven theories are, however, not the only explanation for the dynamics of public support for war. Another possibility is to examine the influence of competition among political elites. The leading proponent of this theory in the foreign policy context is Zaller (1992), who contends that elite discourse is the key to explaining war support (see also Brody 1991). Zaller argues that the balance of persuasive messages carried in the political media determines the contours of opinion on a given policy controversy. Politically knowledgeable individuals are

most likely to receive and accept political messages that accord with their personal political predispositions. The greater the volume of elite discourse favoring a particular policy position from elites of a particular political stripe, the more likely it is that the members of the mass public of the same political stripe will adopt that position. Zaller makes his case in the context of the Vietnam War, arguing that a change in the balance of elite discourse drove the decline in support for the war. In the early phase of the war, when political elites were almost uniform in their support for the U.S. policy, Zaller found a monotonic relationship between political awareness and support for the war; those most attentive to elite discourse were most supportive of the current policy, regardless of their individual predispositions. Zaller terms this the “mainstream pattern.” In the later phases of Vietnam, when elites began to disagree, a “polarization pattern” emerged. Here, the effect of political awareness on support for the war was conditional on an individual’s political values. Citizens attentive to politics followed the path of those leaders who shared their political views. For the Vietnam War, greater awareness led to higher levels of support among hawks and higher levels of opposition among doves.

The casualties hypothesis (and its extensions) and elite-driven theories provide contrasting explanations for the dynamics of public support for war. These theories also carry very different normative implications; whether partisan political actors lead or follow opinion concerning war is a question with profound consequences for democratic practice. However, it has been difficult to assess the relative validity of the two approaches because scholars have focused on the Cold War and post-Cold War American experiences, namely war failures and short-term military excursions (Larson 1996; Sobel 2001). Consider, for instance, the Korean and Vietnam wars. Both the elite conflict theory and the event-driven theory predict that public support would decline as the conflicts unfolded. In the first view, as divisions among elites widened over time during both Korea and Vietnam, public opinion became polarized, thereby decreasing overall support for war. At the same time, since most scholars have used cumulative casualties as a measure of the war’s cost (Larson 1996; though see Gartner et al 1997; Gartner and Segura 1998), and cumulative casualties – as Gartner, Segura, and Wilkerning (1997) note – are collinear with time, the casualties hypothesis predicts a secular decline in support for war over time. Thus, for both theories of public support, time is correlated with the explanatory variables of interest: real-world events and how those events are discussed by elites.

By expanding the study of war to include World War II, it is possible to determine which of these mechanisms is most powerful. Over the almost four years of U.S. involvement in the War, support never waned, even as casualties mounted and as the U.S. suffered both important setbacks and gains (Campbell and Cain 1965). This suggests that the casualties hypothesis is flawed. Authors who have examined support for World War II, such as Larson (1996), have attempted to rescue Mueller’s theory by arguing that World War II was inherently “different,” referring to the unique moral issues at stake. However, opinion polls from that period demonstrate that a sizeable minority of the public could not identify a clear aim for the war (see Berinsky 2005a). More importantly, Zaller’s elite discourse theory can explain the patterns of public support in World War II. As time marched on, cumulative U.S. casualties increased, but political elites remained unified behind the war (Legro 2000; see Berinsky 2005a for further discussion). Examining the U.S. experience during World War II will allow us to disentangle the independent variables at the heart of the two theories.

Analysis

To determine whether the mainstream pattern or the polarization pattern best

characterizes public opinion, we need individual-level measures of predispositions, political information (which proxies attentiveness to elite discourse), and support for the war. In this period, support for FDR seems the most appropriate proxy for predispositions, given FDR's role in pushing the United States to aid England and the prevalence of isolationist tendencies among his Republican opponents (for a discussion of the measurement strategy, see Berinsky 2005a). To measure political information, we rely on questions concerning knowledge of political leaders, geography, and current events that are similar in form to measures of political information used today (Delli Carpini and Keeter 1996). Finally, turning to the dependent variable – support for war – different strategies need to be adopted for different periods of the conflict. Before the U.S. entered World War II, pollsters often asked if the U.S. should become involved in the war and attempted to gauge the conditions under which the public would be willing to risk entry into war. However, measuring support for the war after Pearl Harbor is more difficult than in later conflicts because pollsters never asked respondents if becoming involved was a “mistake.” As a result, indirect measures are necessary. There are a number of items appropriate to such a strategy. Pollsters measured support for the U.S. diplomatic and military aims, both contemporaneously and in the future. These questions can be used to measure underlying support for the military and governmental objectives of the war effort. For instance, several organizations asked respondents if the U.S. should adopt an internationalist posture and take an active role in world affairs after the war, thereby embracing the dominant orthodoxy in foreign policy that emerged after Pearl Harbor (Legro 2000). Admittedly, these are not perfect measures of support for war. We also used alternative measures that arguably are more direct; several polls during this time asked if the U.S. should make peace with Germany.

As a first cut, we have examined how predispositions and political engagement determined support for the war from 1939 to 1944. We expect to see the polarization pattern before U.S. entry into the war. As information levels increase, supporters of FDR should be more willing to adopt his position that the U.S. should be willing to risk war to aid the Allied countries. Opponents of FDR, on the other hand, should be less likely to support aiding the allies as their levels of attention to elite discourse increases (or at least should be no more supportive of such a course of action). On the other hand, once the U.S. entered World War II in December 1941, elite discourse unified behind the President. We therefore expect the mainstream pattern to emerge. Regardless of an individual's political predispositions, those citizens with higher levels of political information should express greater support for administration policies.

We have examined data from nine polls. We present the results from four polls here (see Berinsky 2005a for the full results). The first two polls were conducted by Gallup in November 1939 and January 1941. The second set of polls was conducted by Roper in March 1943 and OPOR in January 1944. Our expectation was that public opinion, as measured in 1939 and 1941, would exhibit the polarization pattern, while opinion measured in the later polls would exhibit the mainstream pattern. For each poll, we followed the analytic strategy outlined by Zaller (1992; 1994) and regressed the dependent variable – various indicators of support for administration policy – on (1) pro/anti FDR predisposition, (2) information levels, (3) interactions between partisanship and the information terms, and (4) a series of demographic variables to control for biases arising from sampling concerns (as described above). Instead of presenting the regression results, we present graphs of the predicted effects of information and partisanship on respondents' support for war.

Figure 1 demonstrates that, as predicted, the polarization pattern characterizes opinion through early 1941. Consistent with the elite competition model, the pattern changed greatly

after the U.S. entered the war: individuals more attuned to elite discourse were more supportive of an active role, regardless of their views of FDR. Figure 2 presents results from the 1943 and 1944 polls. As discourse moved from a two-sided flow to a one-sided flow in 1941, the public followed suit. These results suggest that elite discourse plays a critical role in shaping popular responses to war and challenge the view that events on the battlefield are sufficient to explain the dynamics of public opinion.

Figure 1: Evidence of Polarization Effect

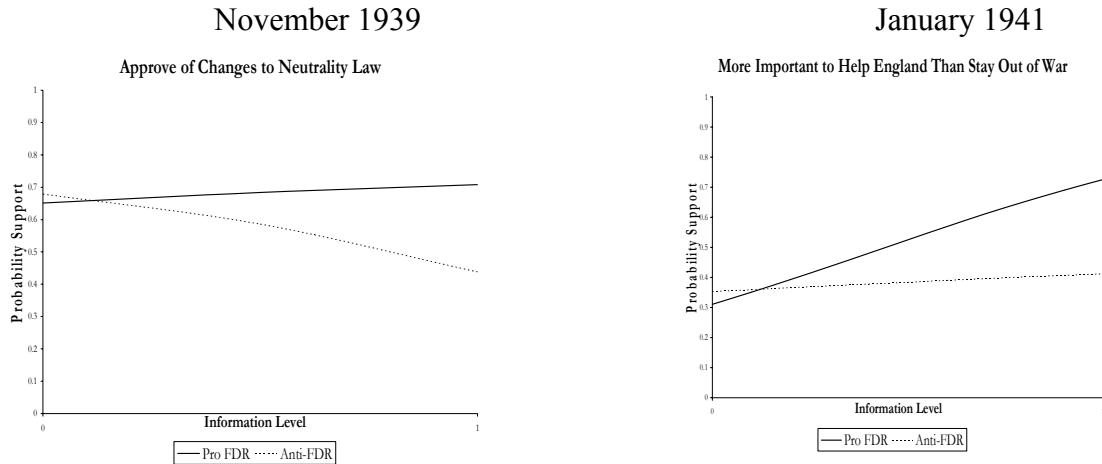
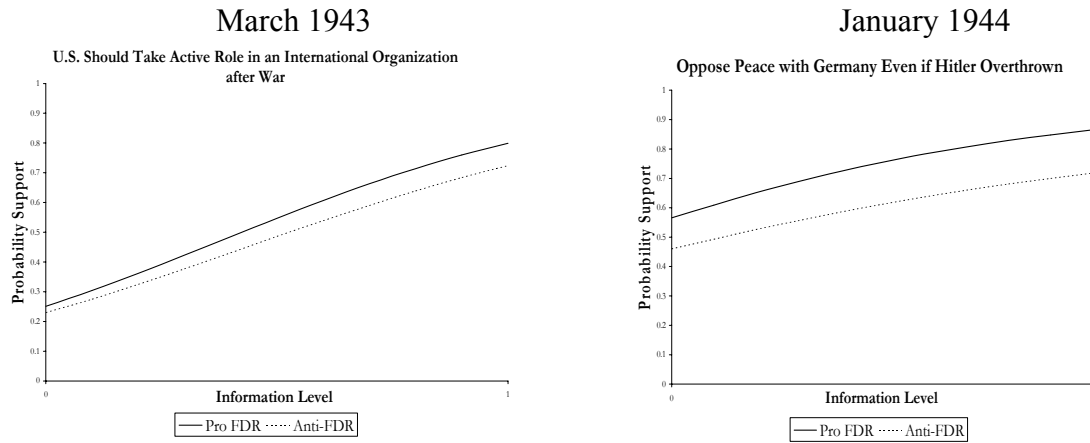


Figure 2: Evidence of Mainstream Effect



Contribution

The American Mass Public in the 1930s and 1940s Project will have a broad impact both in providing data resources for the social science community and in the education of students. First, the project will make a valuable set of data accessible to political scientists, economists, and sociologists. Second, the project will collect auxiliary census information and develop statistical methods that can be used to appropriately analyze these data. Third, the project will promote the use of these data by graduate students interested in the historical study of political behavior. Finally, our project will expose undergraduates to social science research methods through coursework and employment as research assistants.

The most important contribution of this project is the cleaning and recoding of the surveys to make them usable. In the process of completing the project, we will process

approximately 450 datasets currently archived at the Roper Center. We have already recoded 200 surveys. However, much work remains to be done. The most labor-intensive facets of the project are recoding the data, compiling the data into cumulative datasets, and creating codebooks that correspond to the recoded datasets. But this work will also carry the greatest benefit to the profession because it will make it much easier for scholars to use the polls. We intend to make our recoded data and edited codebooks available to the broader research community. The Roper Data is available only through the Roper Center itself, just as data from ICPSR is only released to member institutions (though both Roper and ICPSR allow individuals at non-member institutions to obtain data after paying an access fee). Beginning in Fall 2005, the Roper Center is adopting a system like that of ICPSR, in which membership includes unlimited rights to download and use data from the collection. We intend to follow the lead of Brady and his colleagues (Brady et al. 2000), who used NSF funding to compile 207 datasets from Roper into a single file, available through the Roper Center. We have reached an agreement with the Roper Center to make our data available in a similar manner (see memorandum of understanding posted at <http://web.mit.edu/berinsky/www/memo.pdf>).

Second, we will refine and disseminate the statistical methods appropriate for analyzing quota survey data from the pre-1950 period, described above. Methods for analyzing data collected through random samples are taught in a number of settings, most notably through the Survey Methodology program at the University of Michigan and the University of Maryland. However, there are no widely accepted techniques to analyze the data that have already been collected using quota-sampling methods. Aside from a few texts from the mid-1950s – for instance, Stephan and McCarthy (1958) – there are no guides for interested researchers. If we simply discard these data because they were collected in a non-standard manner, we abandon a rich source of data regarding American history. By disseminating methods to properly interpret quota-sampled surveys, our project will benefit other social scientists interested in those data. Berinsky and Schickler will teach survey research classes, and, as appropriate, will conduct workshops on the use of these data for larger audiences at ICPSR and APSA.

This leads to the third broad impact of our research project. The 1930s and 1940s were critical years in the American political system. Our particular interests lie in the intersection of domestic and international politics during World War II and in the dynamics of domestic policy-making in the late 1930s and 1940s. A noteworthy conservative revival began in Congress starting in 1937. Prevailing accounts of this transformation have focused on the role of the south (Katznelson, Geiger, and Kryder 1993), elite-level ideological battles (Brinkley 1996; Hawley 1966), and the pressure for FDR to moderate in the wake of the onset of World War II (Brinkley 1996). We believe that changes in mass attitudes—in particular, concerns about the growing influence of labor unions—played an important role in shaping the limits of New Deal liberalism. We plan to use the polling data to assess the dynamics of mass opinion—both at the aggregate and individual levels—toward New Deal liberalism. However, there are a number of important questions that could be addressed by other researchers given the early opinion data. By bringing modern tools and theories of political behavior to these data, we can explore many questions of great historical significance. Why, for example, did democracy prevail in the United States during the Depression years of the 1930s, given its failure in Germany and Italy? How did attitudes towards racial policy evolve in the 1930s and 1940s? The work we are doing to prepare the survey data and develop methods to analyze those data will provide valuable resources for scholars of mass political behavior during this pivotal era in American history.

- Achen, Christopher H. 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.
- Baum, Matthew. A. 2002. "The Constituent Foundations of the Rally-Round-the-Flag Phenomenon." *International Studies Quarterly* 46: 263-298.
- Baum, Matthew. A. and Tim Groeling. 2004. "What Gets Covered?: How Media Coverage of Elite Debate Drives the Rally-'Round-the-Flag Phenomenon, 1979-1998." Paper presented at the Annual Meeting of the American Political Science Association.
- Baum, Matthew A.. and Samuel Kernell. 2001. "Economic Class and Popular Support for Franklin Roosevelt in War and Peace." *Public Opinion Quarterly* 65: 198-229.
- Berinsky, Adam J. 2005a. *Assuming The Costs of War: Events, Elites, and American Public Support for Military Conflict*. Working Paper. Available at: <http://web.mit.edu/berinsky/www/war.pdf>
- Berinsky, Adam J. 2005b. *American Public Opinion in the 1930s and 1940s: The Analysis of Quota-Controlled Sample Survey Data*. Working Paper. Available at: <http://web.mit.edu/berinsky/www/QCS.pdf>
- Belknap, George and Angus Campbel. 1951. "Political Party Identification and Attitudes Toward Foreign Policy" *The Public Opinion Quarterly*, 15(4): 601-623.
- Bethlehem, J.G.. 2002. "Weighting Nonresponse Adjustments Based on Auxiliary Information." In Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (Eds): *Survey Nonresponse*, Wiley, New York.
- Brady, Henry E., Robert D. Putnam, Andrea L. Campbell, Laurel Elms, Steven Yonish, and Dorie Apollonio. 2000. *Roper Social and Political Trends Data, 1973-1994*. Storrs, CT: Roper Center for Public Opinion Research
- Breen, Richard. 1996. *Regression Models: Censored, Sample Selected, or Truncated Data*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-111. Thousand Oaks, CA: Sage.
- Brinkley, Alan. 1995. *The End of Reform: New Deal Liberalism in Recession and War*. New York: Knopf.
- Brody, Richard A. 1991. *Assessing the President: The Media, Elite Opinion and Public Support*. Stanford, CA: Stanford University Press.
- Burk, James. 1999. "Public Support for Peacekeeping in Lebanon and Somalia: Assessing the Casualties Hypothesis." *Political Science Quarterly* 114 : 53-78.
- Campbell, Joel T. and Leila S. Cain. 1965. "Public Opinion and the Outbreak of War." *The Journal of Conflict Resolution* 9(3): 318-333
- Converse, Philip E. 1965. "Comment on: Public Opinion and the Outbreak of War." *The Journal of Conflict Resolution*, 9:330-33.
- Delli Carpini, Michael X. and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. New Haven: Yale University Press.
- Deming, W.E. and Stephan F.F. 1940. "On a least squares adjustment of a sample frequency table when the expected marginal total are known". *Annals of Mathematical Statistics*, 11, 427-444.
- Deville, J. and C. Sarndal. 1992. "Calibration Estimators in Survey Sampling". *Journal of the American Statistical Association*, 87, 376-382.
- Deville, J., C. Sarndal, and O. Sautory. 1993. "Generalizing raking procedures in survey sampling." *Journal of the American Statistical Association*, 88, 1013-1020.

- Edwards, George C., and Tami Swenson. 1997. "Who Rallies? The Anatomy of a Rally Event." *Journal of Politics* 59:200–212.
- Erikson, Robert S., Michael B. MacKuen, and James A. Stimson. 2002. *The Macro Polity*. New York, NY: Cambridge University Press.
- Erikson, Robert S. 1976. "The Relationship between Public Opinion and State Policy: A New Look Based on Some Forgotten Data." *American Journal of Political Science* 20: 25-36.
- Feaver, Peter D. and Christopher Gelpi. 2004. *Choosing Your Battles: American Civil-Military Relations and the Use of Force*. Princeton, NJ: Princeton University Press).
- Foyle, Douglas C. 1999. *Counting the Public In: Presidents, Public Opinion, and Foreign Policy*. New York: Columbia University Press.
- Garnter, Scott Sigmund and Gary M. Segura. 2000. "Race, Casualties, and Opinion in the Vietnam War." *The Journal of Politics* 62: 115-146.
- Gartner, Scott Sigmund, Gary M. Segura, and Michael Wilkening. 1997. "All Politics are Local: Local Losses and Individual Attitudes toward the Vietnam War." *The Journal of Conflict Resolution* 41(5): 669-694.
- Gelman, Andrew. 2005. "Struggles with Survey-Weighting and Regression Modeling". Department of Statistics and Department of Political Science, Columbia University, New York.
- Gelman, Andrew and John B. Carlin. 2002. "Poststratification and Weighting Adjustments." in *Survey Nonresponse*. Ed. Robert M. Groves et al. New York: John Wiley and Sons.
- Gelpi, Christopher, Peter D. Feaver, and Jason Reifler. 2004. *Paying the Human Costs of War*. Working Paper.
- Hawley, Ellis W. *The New Deal and the Problem of Monopoly*, Princeton University Press, 1966,
- Heckman, James.J., "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1):153-161.
- Holsti, Ole R. 2004. *Public Opinion and American Foreign Policy*. Ann Arbor: University of Michigan Press.
- Holt, D. and D. Elliot. 1991. "Methods of Weighting for Unit Non-Response". *The Statistician* 40(3):333-342.
- Jentleson, Bruce W. 1992. "The Pretty Prudent Public: Post Post-Vietnam American Opinion on the Use of Military Force." *International Studies Quarterly* 36: 49-73.
- Jentleson, Bruce W. and Rebecca L. Britton. 1998. "Still Pretty Prudent: Post-Cold War American Public Opinion on the Use of Military Force" *Journal of Conflict Resolution*. 42: 395-417
- Kalton, Graham. 1983. *Introduction to Survey Sampling*. Beverly Hills: Sage Publications.
- Kalton, Graham and Ismael Flores-Cervantes. 2003. "Weighting Methods" *Journal of Official Statistics*. 19(2): 81-97.
- Katznelson, Ira, Kim Geiger, and Daniel Kryder 1993 "Limiting Liberalism: The Southern Veto in Congress, 1933-1950," *Political Science Quarterly*, 108 (1993): 283-306.
- Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley and Sons.
- Klarevas, L. 2002. "The 'Essential Domino' of Military Operations: American Public Opinion and the Use of Force." *International Studies Perspectives* 3(4): 417-37.
- Kull, Steven and Ramsey Clay. 2001. "The Myth of the Reactive Public: American Public Attitudes on Military Fatalities in the Post-Cold War Period." In Phillip Everts and

- Pierangelo Isneria, (Eds.). *Public Opinion and the International Use of Force*. London: Routledge.
- Larson, Eric V. 1996. *Casualties and Consensus: The Historical Role of Casualties in Domestic Support for US Wars and Military Operations*. Santa Monica: The RAND Corporation, MR-726-RC.
- Larson, Eric V. 2000. "Putting Theory to Work: Diagnosing Public Opinion on the U.S. Intervention in Bosnia." In Miroslan Nincic and Joseph Leppgold, eds. *Being Useful: Policy Relevance and International Relations Theory*. Ann Arbor, University of Michigan Press.
- Legro, Jeffrey W. 2000. "Whence American Internationalism?" *International Organization* 54(2): 253-289
- Little, R.J.A. 1993. "Post-stratification: A Modeler's Perspective." *Journal of the American Statistical Association* 88, 1001-1012.
- Little, R.J.A. and D.B. Rubin. 2002. *Statistical Analysis With Missing Data* (2nd ed.). New York, Wiley.
- Little, Roderick J.A. and Mei-Miau Wu. 1991. "Models for Contingency Tables with Known Margins when Target and Sampled Populations Differ." *Journal of the American Statistical Association* 86: 87-95.
- Lohr, Susan. 1999. *Sampling: Design and Analysis*. Pacific Grove, Calif.: Duxbury.
- Mitofsky Warren J. 1999. "Was 1996 a Worse Year for Polls than 1948?" *Public Opinion Quarterly* 62:230-249
- Mosteller, Frederick, Herbert Hyman, Philip J. McCarthy, Eli S. Marks, and David B Truman, eds. 1949. *The Pre-Election Polls of 1948*. New York, NY: Social Science Research Council.
- Mueller, John E. 1973. *War, Presidents, and Public Opinion*. New York: Wiley.
- Mueller, John. E. 1994. *Policy and Opinion in the Gulf War*. Chicago: University of Chicago Press
- Orren, Karen and Stephen Skowronek. 2002. "The Study of American Political Development." In Ira Katznelson and Helen V. Milner (eds.), *Political Science: The State of the Discipline*. New York: W. W. Norton & Company.
- Page, Benjamin I. and Robert Y. Shapiro. 1992. *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. Chicago: University of Chicago Press.
- Pierson, Paul and Theda Skocpol. 2002. "Historical Institutionalism in Contemporary Political Science." In Ira Katznelson and Helen V. Milner (eds.), *Political Science: The State of the Discipline*. New York: W. W. Norton & Company.
- Roper, Elmo. 1940. "Sampling Public Opinion." *Journal of the American Statistical Association* 35(210, Part 1):325-34.
- Ruggels, Steven and Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, Chad Ronnander. 2004. "Integrated Public Use Microdata Series: Version 3.0". Minneapolis, MN: Minnesota Population Center.
- Schafer, Sarah. 1999. "To Politically Connect, And Profitably Collect; Ron Howard's Web Site Wants to Get People Involved -- and to Sell the Results Of Its Internet Polls. The Question: Will It Work?" *Washington Post*. p. F10. December 13, 1999
- Schlozman, Kay Lehman, and Sidney Verba. 1979. *Injury to Insult: Unemployment, Class, and Political Response*. Cambridge: Harvard University Press.

- Schultz, Kenneth. 2001. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press
- Sharot, T. 1986. "Weighting Survey Results." *Journal of the Market Research Society*, 28, 269-284.
- Stephan, Frederick and Philip McCarthy. 1958. "Sampling Opinions, An Analysis of Survey Procedure". John Wiley & Sons, Inc. New York.
- Sobel, Richard. 2001. *The Impact of Public Opinion on U.S. Foreign Policy Since Vietnam: Constraining the Colossus*. New York: Oxford University Press.
- Thompson, Steven K. 2002. *Sampling*. John Wiley & Sons, Inc. New York.
- Verba, Sidney, and Kay Lehman Schlozman. 1977. "Unemployment, Class Consciousness, and Radical Politics." *Journal of Politics* 39: 291-323.
- Weatherford, M. Stephen and Boris Sergeyev. 2000. "Thinking About Economic Interests: Class and Recession in the New Deal." *Political Behavior* 22(4):311 - 339
- Winship, Christopher and Larry Radbill. 1994. "Sampling Weights and Regression Analysis". *Sociological Methods & Research* 23 (2) 230-257.
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.
- Zaller, John R. 1994 "Elite Leadership of Mass Opinion: New Evidence from the Gulf War." In Lance Bennett and David Paletz (Eds.). *Taken By Storm: The Media, Public Opinion, and U.S. Foreign Policy In the Gulf War*. Chicago, IL: University of Chicago Press.