I'm Not Voting for Her: Polling Discrepancies and Female Candidates

Christopher T. Stout · Reuben Kline

Published online: 16 September 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Although there is a large literature on the predictive accuracy of pre-election polls, there is virtually no systematic research examining the role that a candidate's gender plays in polling accuracy. This is a surprising omission given the precipitous growth of female candidates in recent years. Looking at Senate and Gubernatorial candidates from 1989 to 2008 (more than 200 elections in over 40 states), we analyze the accuracy of pre-election polls for almost the complete universe of female candidates and a matched sample of white male cases. We demonstrate that pre-election polls consistently underestimate support for female candidates when compared to white male candidates. Furthermore, our results indicate that this phenomenon—which we dub the Richards Effect, after Ann Richards of Texas—is more common in states which exhibit traits associated with culturally conservative views of gender issues.

Keywords Gender and politics · Polls · Female candidates · Public opinion

On the eve of the 2008 New Hampshire Democratic presidential primary, pre-elections polls predicted a sizable win for Barack Obama in the first presidential

Electronic supplementary material The online version of this article (doi:[10.1007/s11109-010-9137-6](https://doi.org/10.1007/s11109-010-9137-6)) contains supplementary material, which is available to authorized users.

C. T. Stout · R. Kline
Department of Political Science, University of California, Irvine, 3151 Social Science Plaza,
Irvine, CA 92697, USA
e-mail: kliner@uci.edu

C. T. Stout (✉)
Department of Political Science, Wellesley College, 106 Central Street, Wellesley, MA 02481, USA
e-mail: cstout2@wellesley.edu

primary of the year. A victory in New Hampshire would be the second consecutive defeat for the one-time front runner Hillary Clinton and essentially end her bid for President. In a surprising turn of events, Hillary Clinton defied pre-election poll predictions and won the New Hampshire primary. While some polling critics argued that this was evidence of the Bradley Effect, others claimed the poll discrepancies could be explained by an unexpected boost in turnout or possibly a large shift in undecided voters towards Clinton. What was omitted from this discussion, however, was the role that Clinton's gender played in these polling discrepancies.

While the claim that polls often *over*-state support for African-American candidates—a phenomenon known as the Bradley Effect—is a familiar one, women too may be susceptible to systematic polling biases. We can readily identify instances where polls have overestimated the support of female candidates such as Janet Napolitano in the 2002 Arizona Gubernatorial Election. However, there appear to be many more cases where the polls have *underestimated* the support of female candidates. For example, Hillary Clinton in the 2000 New York US Senate Election, Ann Richards in the 1990 Texas Gubernatorial Election, and Christine Todd Whitman in the 1993 New Jersey Gubernatorial Election were predicted to lose by sizable margins or be in very competitive contests, and yet all three won their respective elections with ease. This anecdotal evidence begs the following questions: To better understand polling inaccuracy, must we take into account the gender of the candidate, the same way we should consider race (Finkel et al. 1991; Traugott and Price 1992; Reeves 1997)? Or is the gender of the candidate inconsequential once we account for factors thought to influence polling accuracy, such as the party of the candidate, turnout, margin of victory, and the proportion of undecided voters in any poll?

Despite a very large literature on polling accuracy, there is almost no research examining the role that gender has on polling discrepancies. The one refreshing exception to this rule is Hopkins' (2009) study in which he examines a rich data set of female candidates who campaigned for either the US Senate or Governor from 1989 to 2006. In his analysis, which is the first to systematically examine polling bias for female candidates, Hopkins concludes that there is no "Whitman Effect" (the term he uses for polling inaccuracies with respect to female candidates) and that polls do not *overestimate* support for women. Despite the fact that Hopkins finds that polls do not overestimate support for female candidates, his analysis stops short of testing alternative hypotheses or offering any theoretical explanation why we would observe differences for female candidates. Therefore, a necessary addition to this literature is a test of whether polls *underestimate* support for female candidates and a theoretical explanation of why we might observe gender related polling discrepancies.

Moreover, we build on Hopkins' work by adding additional features to the empirical analysis. These features include the addition of a set of carefully selected white-male matched comparison cases and the introduction of other potentially important factors such as the election and social context, the incumbency status of the candidate, and gender specific turnout rates. If we find that the accuracy of polls for female candidates is significantly reduced when compared to polls for similarly situated traditional white male candidates, then this is strong evidence that such a decrease in polling accuracy is somehow associated with the candidate's gender.

In this paper, we analyze polling accuracy for female candidates. Based on previous research pertaining to gender and socially desirable response bias, we construct two competing hypotheses. The first hypothesis claims that polls will overestimate support for women, likely because respondents want to avoid being perceived as sexist. The second hypothesis argues that polls will tend to underestimate support for women because voters may fail to publicly endorse a female candidate out of reluctance to signal support for non-traditional views of gender roles. Using data on female and white male candidates collected from the US Census, university and newspaper polls, and numerous state election resources, we run several standard OLS regressions to test our hypotheses. The results indicate that pre-election polls for female candidates significantly *underestimate* their ultimate electoral support, a phenomenon we call the “Richards Effect” after former Texas Governor Ann Richards.

To further investigate the motives behind the systematic underestimation of female candidates’ electoral support, we analyze contextual factors for the universe of female candidates who were campaigning for governor or US Senator from 1989 to 2008. We find that female candidates are most likely to experience larger polling discrepancies in states where fewer women are in the labor force and in states where the Congressional delegation has a poor record on supporting progressive gender issues. The results from both sets of analyses suggest that voters tend to prove more supportive of female candidates in states in the voting booth than in responding to a pre-election poll, and this tendency intensifies in gender-conservative states.

Causes of Polling Inaccuracies

Election polls have evolved over the course of a century and have become increasingly precise instruments for measuring election outcomes. This is why most pre-election polls are very successful in predicting the eventual winner (Bolstein 1991; NCPP 1997; Mitofsky 1998; Traugott 2005; Keeter and Samaranayake 2007). Several authors have shown that the best predictor of whether or not a person will vote and who they will vote for can be found in their pre-election poll response (Mosteller 1949; Bolstein 1991; Mitofsky 1998). Despite the general accuracy of modern pre-election polls, there have been numerous instances in which the actual results do not match their predictions (the 1948 and 1996 Presidential elections are often cited as examples of particularly poor poll performance). Furthermore, under certain conditions related to the nature of the candidates and the race, the predictive power of these polls may be further diminished.

Studies of polling have identified several factors that could lead polls to be less accurate. Some of the usual culprits of imprecise polling are large numbers of undecided voters and larger than expected turnout rates (Perry 1979; Fenwick et al. 1982; Mitofsky 1998; Berinsky 1999; Durand et al. 2001; Traugott 2005). Others have also shown that non-competitive races and elections with incumbents can also lead less accurate polling, both pre- and post-election (Crespi 1988; Wright 1990; Gow and Eubank 1984). These factors have historically helped predict polling accuracy. We propose that certain salient characteristics of the candidate’s identity as a necessary addition to this literature.

One of the major shortcomings of the literature on polling accuracy is that it has not taken into account the gender of the candidate, with the only major exception to this rule being Hopkins' 2009 piece. Perhaps the chief reason for this is that until the 1990s few female candidates were ever nominated for races that got statewide or national attention. As a result, only a small number of firms conducted polling for these candidates. With the growth of female candidates in recent elections there has been a concomitant increase in the number of polls gauging their support. In what follows, we argue that polls predicting support for female candidates may be less accurate than for traditional candidates, even once context and other relevant factors have been taken into account.

Gender and Polling Inaccuracies

A commonly cited problem with polling and survey accuracy is that respondents may give misleading responses in order to appease the interviewer. This phenomenon is known as socially desirable response bias and may be expected to be most common among respondents who have the most to gain from this deception (Hatchett and Schuman 1975; Jackman and Muha 1984; Krysan 1998). In the race featuring the candidate who lends her name to the "effect" discussed in this paper—Ann Richards—gender figured into the contest in important and salient ways.¹ Because many issues regarding gender are sensitive, survey respondents may—even more so than is usually the case—perceive that they are being personally judged based on their responses (Ballou and Del Boca 1980; Lueptow et al. 1990; Northrup 1992). Such a psychological process thus could induce the Richards Effect that we document below.

While there is a dearth of literature directly theorizing regarding the direction, if any, of polling biases for female candidates, we can gain some insight into how voters will respond to these candidates based on related research. Several studies find that opinion polls measuring attitudes on topics pertaining to gender are often subject to inaccuracies (Ballou and Del Boca 1980; Lueptow et al. 1990; Northrup 1992; Huddy et al. 1997; Johnson and DeLamater 1976; DiCamillo 1991; Streb et al. 2008). This research, however, provides two competing hypotheses which predict biases in opposite directions. The principal objective of our analysis is to adjudicate, using the universe of female candidates for Governor or US Senator from 1989 to 2006, between these two contradictory hypotheses.

One strain of research indicates that men and women will often over-report their support for causes and policies related to gender equality (Huddy et al. 1997; Johnson and DeLamater 1976; DiCamillo 1991). These studies find that some men and women want to appear more supportive of issues such as pay equity, the women's movement, and non-traditional gender roles than they actually are in order to preclude being labeled a sexist (Huddy et al. 1997; Johnson and DeLamater 1976; DiCamillo 1991). Furthermore, Streb et al. (2008) find that there is a significant socially desirable response bias in discussions regarding hypothetical female presidential candidates. Using a list experiment, Streb et al. (2008) show that many respondents are willing to

¹ Richards' opponent, Clayton Williams Jr. likened poor weather to rape in a March 1990 interview and was suspected of visiting a brothel when he was a college student.

tell a pollster that they will vote for a female president when in actuality, they have a small probability of supporting such a candidate.

What incentives does a respondent have to say that they will vote for a female candidate or agree with a progressive gender ideology when they in fact do not? Similar to explanations of the Bradley Effect (see Finkel et al. 1991; Traugott and Price 1992), some respondents might tell pollsters that they support a female candidate, even when they do not, in order to preclude the appearance of sexism. There is evidence indicating that female candidates may be perceived as too liberal, weak on national security and crime, and biased toward women's issues (Sapiro 1983; Huddy and Terkildsen 1993; Sanbonmatsu 2002). As a result, we might find that respondents mislead pollsters in stating their support for the female candidate, but based on fears that she will not adequately represent their interests, vote against her. If this hypothesis is correct, we should see that female candidates perform better in the polls than in the actual election. Moreover, this effect should be exacerbated in states where residents generally hold more progressive views about gender roles. It is in this context where respondents may falsify their preference in support of female candidates in order to appear more progressive on gender issues so their friends and colleagues do not think less of them.

A second strain of research suggests that both males and females are more progressive on gendered issues than they admit (Breinlinger and Kelly 1994; Buschman and Lenart 1996; Jacobson 1981). For example, previous studies have shown that respondents will agree with the idea of progressive gender roles including pay equity and increasing the number of women in positions of power, but these respondents will not identify as being a “feminist” or anything related to that term (Burn et al. 2000). Moreover, Theriault and Holmberg (1998) provide evidence that suggests that respondents want to appear more conservative on gender issues than they actually are. They find that the respondents in their study who score high on the social desirability scale “seem to be portraying themselves as relatively conservative and relatively traditional [on gender issues]” (Theriault and Holmberg 1998, p. 108). These authors argue that, possibly as a cultural backlash to the feminist movement of the 1960s and 1970s, some respondents may publicly distance themselves from the values that feminist fought for, while supporting these issues privately.

Why would respondents want to appear more conservative on gender issues than they actually are? There are a variety of reasons why this may be the case, including fears that a gender progressive agenda would be biased towards women and a negative cultural perception of feminism. While most men and women may publicly agree with pay equity and having more women in positions of power in the abstract, they may be uncomfortable with the methods used to obtain this equity. For example, some voters may see affirmative action as a vehicle for ending gender disparities in the work place (Bielby 2000; Northrup 1992). However, policies like affirmative action are perceived by some as examples of feminists being too aggressive and creating policies which unfairly benefit women (Northrup 1992). Those who support such policies may be negatively perceived as being too “radical” or unfair. As a result, some voters may support these issues privately, but publicly may be ashamed to support policies which are perceived as privileging women.

A persistent belief about female candidates is that they are seen as biased toward gender issues, regardless of their partisanship, and thus they are often perceived as being supportive of a progressive gender agenda (Leeper 1991; Huddy and Terkildsen 1993; McDermott 1997; Sanbonmatsu and Dolan 2009). As a result, some voters may believe that female candidates, like feminists, may advocate for policies which unfairly benefit women. While some may privately agree with and support policies that address gender inequity, they may publicly want to distance themselves from female candidates in order to preclude the appearance of being too “extreme”. Polling accuracy for women may suffer as a result as these voters may mislead pollsters by saying that they do not plan to vote for a female candidate, but then vote for her in the privacy of the voting booth.

Additionally, the negative stereotypes of feminism may lead some voters to try to distance themselves from any progressive gender issues. In addition to being perceived as too aggressive, some people attribute the demise of the traditional American family to feminism and the women’s liberation movement (Burn et al. 2000; Williams and Wittig 1997; Twenge and Zucker 1999; Roy et al. 2007). As a result, voters may approve and agree with a progressive gender agenda privately, but may be too ashamed to support such issues in fear that they will be associated with feminism and these negative attributes.

If this is true, this backlash to feminism may also influence polling accuracy for female candidates. Female candidates are sometimes portrayed in the media as having the negative stereotypical qualities associated with feminism (Templin 1999). For example, Glenn Beck notes “[Hillary Clinton] is like the stereotypical...she’s the stereotypical bitch, you know what I mean? She’s that stereotypical nagging...” (National Organization for Women 2008). Democratic Strategist Paul Begala and Washington Post staff writer Tony Kornheiser have drawn comparisons between Florida US Senate candidate Katherine Harris and Disney villain Cruella De Vil (CNN 2004; Kornheiser 2000). These characterizations of female candidates as being too aggressive, cold, and calculating are often the same negative stereotypes associated with the feminist movement. Moreover, female candidates who run for elected office are often portrayed as being irresponsible home makers. Kantor and Swarns in a 2008 New York Times article identify this perception among voters “Many women expressed incredulity—some of it polite, some angry—that Ms. Palin would pursue the vice presidency given her younger son’s age...” (Kantor and Swarns 2008, p. 1). This negative dialogue about such candidates can make voters feel uncomfortable about voicing their support for these candidates publicly, even if, for ideological or other reasons, they support these candidates privately.

We expect that if such a behavioral response is exhibited, it will be more prevalent in areas in which traditional gender roles are the norm. In this context, signaling a preference for a female candidate would go against the prevailing view of traditional roles for women, and thus could be expected to lead to the loss of esteem among one’s peers. Therefore, if it is true that respondents falsify their true preference for the female candidate as a public signal of their commitment to traditional roles for women, then we should expect such behavior to be more prevalent in gender-conservative states, especially in states where the prevailing view of gender roles is the traditional one.

To summarize, previous research has shown that survey respondents may be particularly likely to give misleading information when discussing gender related topics. This makes it difficult to gauge respondents' true feelings about these issues. Studies related to gender have found mixed results in this regard. On one hand, respondents may want to appear more progressive on gender issues than they actually are. On the other hand, respondents may want to distance themselves from being perceived as supporting feminist issues. Taken together, these empirical findings lead us to the following two competing hypotheses.²

Hypothesis A Pre-Election Polls will consistently *overestimate* support for female candidates because respondents want to appear more progressive on gender issues. And, if true, this tendency should be greatest in more gender-progressive states.

Hypothesis B Pre-Election Polls will consistently *underestimate* support for female candidates because respondents want to appear more conservative on gender issues. And, if true, this tendency should be more prevalent in gender-conservative states.

In order to assess the validity of these hypotheses, we collect and analyze data on more than 100 major party female gubernatorial and US Senate candidates, and to serve as a baseline for comparison, nearly 100 cases of two white male opponents in similar contests.

Data

Our sample includes 215 distinct cases from over 40 states. It contains 123 female candidates and, to serve as controls, 92 races in which two white males were the major party candidates. We focus on prominent statewide races—those for US Senate and Governor—for several reasons. At this level of government, candidates tend to be better known than those at the local or congressional level. Also, there are many more female candidates at this level than in Presidential elections. Finally, there are more polling results available for US Senate and Gubernatorial candidates than there are for congressional and local level candidates.

The election results in this analysis were obtained from www.uselectionatlas.org. In addition to collecting information on the final election results, we also collected information on pre-election polls. This data was obtained from newspapers from the state in which the election took place. We used four main criteria in selecting these polls. First, we used the poll that was closest to the election date. Therefore, we had the most current—thus presumably the most accurate, at least on average—estimate of the candidate's support.³ Second, we only included surveys that were conducted by telephone. By relying on a single method of polling we can ensure that differences in our dependent variable are not driven by differences in methods of poll

² Hopkins (2009) tested Hypothesis A, and found no support for it, but did not consider Hypothesis B, though his paper does actually provide evidence consistent with this second hypothesis (see footnote 13 on p. 774).

³ If two or more polls were conducted within 3 days, the average score was calculated.

collection.⁴ Third, we only used surveys that were conducted by polling firms, newspapers, or universities, thus eliminating the potentially biasing factor of partisan polling. Finally, we only considered surveys that contacted more than 300 respondents.

Because our principal variable of interest is the gender of the candidate, we collected polling information for the universe of female candidates from 1989 to 2008 in which polling data was available.⁵ The data was collected primarily from Lexis-Nexis, Access News Archives, and Daniel Hopkins' publicly available data on polls for female candidates.⁶ For comparison cases, we also collected information on white male candidates from the similar sources.⁷ All white male candidates were selected according to matching criteria, so as to have one matched "male" case for every case of an election with a female candidate in our sample. This was done in order to minimize the potential confounds introduced by ever-changing polling and weighting formulas as well as to control for certain temporal and spatial effects on polling accuracy.

Based on previous research, we also expect polling inaccuracies to be tied to the circumstances of the election and the context in which the election took place. With respect to the election context, we considered whether the candidate of interest was a Democrat or Republican and whether the candidate or his opponent was an incumbent.⁸ Several social scientists argue that polls are more likely to overestimate Democratic candidates' support and encounter difficulties in estimating incumbents' vote shares (Crespi 1988). We also include a dummy for whether the election of interest was a gubernatorial election or a US Senate election. Research shows that some voters are more hesitant to support a female candidate for an executive position than they are for legislative office (Huddy and Terkildsen 1993; Dolan 1997). Therefore, the office being sought by the female candidate may influence the level of polling accuracy.

With respect to other features of the election, we also considered whether the election had an uncharacteristically high (or low) turnout which could lead to less accurate polls (Crespi 1988; Bolstein 1991; Mitofsky 1998). This variable was

⁴ Unfortunately, we are not able to know whether polls were conducted using Interactive Voice Response (IVR) polls. Our data was collected from newspapers which rarely specify how the poll was collected beyond whether it was collected by phone or internet.

⁵ African American female candidates and candidates who campaigned against African American males were excluded from the analysis. Several studies have shown that black candidates, especially prior to 1996, are susceptible to polling errors. Therefore, to ensure that race does not complicate our findings we excluded these female candidates from our analysis. See Hopkins (2009) and Stromberg (2008) for more information on this topic.

⁶ <http://dvn.iq.harvard.edu/dvn/dv/DJHopkins>.

⁷ It is difficult to find unique comparison elections in states where female candidates are very common, such as Washington, Texas, and California, therefore 30 white male comparison elections were matched to more than one female case. Linda Lingle of Hawaii (2006) did not have a matched white male candidate because there was no Senate or Gubernatorial race within 6 years of her election.

⁸ While we initially included variables for whether our candidate of interest is an incumbent and whether there is an incumbent in the race, due to collinearity issues our final model only includes the former variable. In both models, neither variable was significant and both had no effect on polling accuracy for the race or gender of the candidate.

constructed using voting age population (VAP) turnout rates for each state of interest from George Mason's United States election project (http://elections.gmu.edu/voter_turnout.htm). For each state, we calculate a standard score for both Presidential and non-Presidential elections from 1988 to 2008. Using standard scores allows us to see if the election of interest had much higher or lower turnout than expected and allows us to make comparisons across states. Turnout of certain subgroups may be particularly important in elections with female candidates, so we also calculated a standardized score for female turnout. If we do find that polling for female candidates is less accurate than polling for male candidates a potential culprit may be an unexpected increase in female turnout when there is a female candidate on the ballot. A large swell in female turnout may be particularly problematic for pollsters as all else being equal, female voters prefer female candidates (Brians 2005; Dolan 2004; Fox 1997).

We also include a variable for the number of undecided voters, because it is well established that large numbers of undecided voters can lead to polling inaccuracies (Perry 1979; Fenwick et al. 1982; Mitofsky 1998; Berinsky 1999). We also controlled for the percent of third party support and the percent of "Don't Know" voters in the poll by subtracting the candidate of interest's vote share, their opponent's vote share, and the percent undecided from 100. Next, we used the margin of victory in the pre-election polls as an indicator of competitiveness.⁹ Some research suggests that polling in uncompetitive races will be less accurate because there is often a regression toward the mean (i.e., toward a 50–50 split of the vote share between the two candidates), which continues past the final poll into the election (Hopkins 2009). As a result, candidates who are leading by a large margin will tend to perform significantly worse in the election than their poll-based expectations and candidates losing by large margins according to polls will likely outperform expectations. However, polls for candidates in close elections do not have much room to regress to the mean as the (estimated) vote shares are already near an equal split. Finally, we include a variable that measures the number of days away from the day of the election that the poll was conducted to ensure that our results are not driven by the time frame of the poll's collection of data.

We also sought to account for the social context in which each election occurred, since this may affect the decision making calculus of the survey respondents. We use four variables to measure the social context of the election: percent women in the state legislature, percent of women in the labor force, the political culture surrounding gender issues in the state, and the Democratic Presidential candidate's vote share compared to the his national vote share. The percent of women in the state legislature variable was collected from Rutgers' Center for American Women in Politics. The percent of women in the labor force was collected on an annual basis through the Bureau of Labor Statistics summaries of the Current Population Survey. The women in the labor force variable measures the percent of women who are employed in that state who are over 16 years old and not institutionalized.¹⁰

⁹ We use the predicted margin of victory as a measure of competitiveness rather than the election margin of victory, because voters make calculations on their vote choice with only the knowledge of polled support.

¹⁰ West Virginia, Louisiana, Alabama, Arkansas, and New York are the states with the fewest women in the labor force and Minnesota, North Dakota, Colorado, New Hampshire, and Vermont have the highest percent of women in the labor force.

To approximate the gender political culture in the state we use the average American Association for University Women (AAUW)¹¹ voting records for Congressional representative and US Senators from each state. The AAUW assigns scores for each US Senator and Congressional representative based on their support for bills which advance educational, economic, and social equity for women. High AAUW scores indicate that the particular representative supports a gender progressive agenda. Our AAUW measure is calculated by averaging the AAUW score for entire Congressional state delegation. Finally, our proxy for partisanship of the state was created by subtracting the Democratic presidential candidate's vote share in the state from his vote share in the general election. The partisanship score for each state is based on the previous Presidential Election. For example, the state's partisanship score for 2002 is based on that state's support for Gore in 2000 relative to Gore's national support.

If Hypothesis A is correct, we should expect that respondents in states with higher percentages of women in the labor force, women in the state legislature, and states with higher AAUW scores to be the most likely to *over-report* their intentions to support female candidates in an attempt to express views which better fit these presumably more gender-progressive areas. Conversely, we should expect a tendency to *under-report* in states with lower values of these variables if Hypothesis B is correct.

Methods

The measure we use as our dependent variable is often referred to as Mosteller #5 (Mosteller 1949). To construct this measure, we first calculate the margin of victory as measured by pre-election polling, meaning that we subtract the level of polling support for the opponent (P_b below) from the level of polled support for our candidate of interest (P_a below). We then repeat this calculation for actual election margin of victory (V_a and V_b below). Finally, we subtract the election margin of victory from the predicted margin of victory. If this difference is negative, then there is putative evidence that polls underestimate support for female candidates. That is, when compared to his/her opponent, the candidate's expected performance (measured by the pre-election poll) was worse than his/her actual performance. The reverse is true if this difference is positive. Equation 1 below provides a more formal definition of our dependent variable.¹²

Construction of polling inaccuracy measure

$$\text{Mosteller\#5} = (P_a - P_b) - (V_a - V_b) \quad (1)$$

¹¹ Alaska, New Jersey, Texas, Virginia, and Oklahoma are the states with the lowest average AAUW scores. Conversely, Vermont, New Hampshire, Rhode Island, Hawaii and Massachusetts consistently have the highest AAUW scores.

¹² The models were also run using a polling accuracy method, measure A by Martin et al. (2005). The results remain the same as Mosteller 5 (see Online supplementary appendix). We choose to use Mosteller 5 for this paper however, because this method is, almost without exception, used in discussion of polling error in everyday life, journalism, and the academic literature (Mitofsky 1998).

where P_i is the percentage of poll respondents indicating their intent to vote for candidate i , and V_i is the percentage of votes for candidate i , where $i = A, B$. In our application, candidate A is our candidate of interest, i.e. the candidate with respect to whom we are assessing the polling discrepancy.

This measure of polling inaccuracy suits our purpose for several reasons. It can be calculated with respect to a specific candidate, it is unaffected by the pre-election margin of victory, (i.e., competitiveness) because it is a measure based on absolute rather than relative error. Moreover, it is not a function of the level of support for the candidate of interest, making it less susceptible to variation due to changes in third-party support or undecided voters. Finally, it is the measure of (in)-accuracy most often reported by the media and preferred by social scientists (Mitofsky 1998).

To understand the relationship between gender and the accuracy of pre-election polls we use several statistical methods. To obtain a basic understanding of this relationship, we first review the descriptive statistics and perform two difference of means tests. The first difference of means test measures whether male and female candidates' support in pre-election polls is significantly different from their actual election support. In order to isolate the effect of the candidate's gender on polling accuracy, we utilize a second difference of means test. This time we use a paired two sample t -test measuring differences in polling accuracy for female candidates and their white male comparisons. As described above, for each female candidate we chose a comparison statewide election (i.e., one which involves two white males). The polls for the comparison cases are often collected by the same polling firm as that which conducted the polls for the election of interest.¹³

Using this matching method, we are able to control for geographic, temporal, and polling firm-specific effects to the maximum extent possible. Various studies have identified many different types of general survey mode effects (Tourangeau and Smith 1996; Krysan et al. 1994; Fricker et al. 2005; Kreuter et al. 2008), and moreover, the way in which a candidate's gender interacts with these previously identified modes effects is entirely unknown. In light of this, the use of matched cases becomes even more crucial. Without them, we could not be sure that any polling inaccuracy that exists for female candidates is not attributable to the mode in which the survey was conducted. Utilizing the matched cases we can isolate the effect of gender and obviate any confounds due to mode effects. Finally, this technique allows us to compare the average polling bias for one set of candidates of interest (white females) with the average for another set of carefully chosen candidates of interest (within carefully chosen elections of interest) with the result being that the two groups are as similar as possible, except the polar differences in the gender identity of the candidates. Without identifying a candidate of interest in a particular election the average polling-performance gap (to borrow Hopkins' (2009) phrase) is always zero, as an overestimate for one candidate perfectly cancels an underestimate for the other. However, because our interest is in the effect of gender on polling accuracy, the female candidate is our default candidate of interest, allowing us to calculate a female-specific bias.

¹³ For information on the comparison cases, see "Appendix".

Because there are multiple factors, aside from the candidate's gender, which we expect will affect polling accuracy we estimate four different OLS regressions which predict the accuracy of pre-election polls. In the first regression, we analyze the differences between white male and white female candidates without any controls. This model will allow us to determine if polling accuracy for female candidates are significantly different than polls for males. The second and third models will test the same relationship, but in these models we control for the context of the election (Party ID, non-competitiveness, standardized voting age population turnout, female turnout, percent undecided, incumbent in the race, and a gubernatorial election indicator variable). What distinguishes the third from the second model is in the third model we account for female turnout instead of voting age population turnout.¹⁴ Our final model will test the effects of both election and state (social) context variables on the accuracy of polls.

Results

Table 1 presents descriptive statistics for the size distribution of the polling inaccuracy for our groups of interest. For each group, we report the mean dependent variable score (the average difference between the predicted and actual margins of victory), its standard deviation, the percent of cases in which the polls fall outside the margin of error, and, of those which fall outside the margin of error, the percentage of cases which over-predict and under-predict the electoral performance of the candidate. Figure 1 is a box plot of levels of polling accuracy for male and female candidates which presents the descriptive statistics in Table 1 and the distribution of the data visually.

As indicated by the first column, in the aggregate there is essentially no difference between the pre-election polls and the final results for white male candidates. The average difference between the pre-election polls and the final result for these candidates is well under a half of one percentage point. Therefore, the discrepancies for the individual cases are not biased in either a positive or negative direction, so they, for the most part, cancel each other out. The same table shows that pre-election polls tend to underestimate the final results for female candidates by 2.12% points. Furthermore, the difference between the pre-election poll and the final result is significantly different, which indicates fairly unequivocally that polls significantly underestimate support for female candidates.¹⁵

¹⁴ We use separate models because Voting Age Turnout and Group Specific Turnout are highly collinear.

¹⁵ The results from our paper differ from the results in Hopkins (2009) for three main reasons. First, instead of using multiple polls over the course of the election, we only use the most recent poll or an average of the most recent polls for estimates of polling accuracy. We believe this is advantageous because under this method each candidate in our data equal number of scores (one) and this score is the most recent, thus presumably the most accurate estimate of the candidate's support. Second, we have slightly different samples. Our data set includes cases from 2008 and excludes cases with black candidates to ensure that the so called Bradley Effect does not influence our result. Finally, the difference in the results may also be attributed to our different measures of polling inaccuracy. Despite these differences, Hopkins (2009) does find that female US Senate candidates suffer from a Richards Effect as noted in a footnote 13 on p. 774 of his manuscript.

Table 1 Summary of polling discrepancies for male and female senate and gubernatorial candidate W/T-test of means

Candidate type	N	Avg. poll accuracy	Standard deviation	Outside margin of error (%)	% Over predict	% Under predict
Female	123	−2.12***	6.98	54	34.32	65.67
Male	92	0.43	8.16	57.60	56.60	43.39

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Comparison Variables: Predicted Margin of Victory and Election Margin of Victory. Positive results indicate the candidate performed better in pre election polls than in the election. Negative results indicate a Richards Effect. The margin of error in the fifth column of this table varies from $\pm 3\%$ to about $\pm 5\%$ depending on the poll. Of those which fall outside the margin of error, the sixth and seventh columns display the percent of cases which over-predict and under-predict the electoral performance of the candidate. The table shows that polls for female candidates more often than not underestimate their election performance. Moreover, almost two-thirds of the polls outside of the margin of error for female candidates underestimate their electoral support

Fig. 1 Box plot of polling accuracy for male and female US Senate and gubernatorial candidates from 1989 to 2008. The box-plot displays the results in Table 1 graphically. The results of the box-plot indicate that the median female candidate generally outperform her pre-election poll prediction. Moreover, the graph shows that most female candidates in the upper quartile of the distribution perform better in the election than polls predict. The box plot also shows that the there is not a large discrepancy between the median white male candidates' polled support and their electoral support

The difference in polling accuracy between white male and female candidates is also indicated in Fig. 1. According to the figure, polls tend to *underestimate* female candidates' electoral support. This appears to occur for a majority of the women in our

Table 2 Matched paired T tests comparing women candidate to white male candidates from the same state and time period

Candidate type	N	Avg. poll inaccuracy	Standard error	Standard deviation
Female poll discrepancy	122	−2.11	0.63	7.01
Male poll discrepancy	122	1.36	0.73	8.11
Difference	–	−3.47***	0.902	9.97

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Negative results indicate that candidates perform better in the final election than pre-election polls would predict. All comparison polls for women candidate are white males from the same state and many from the same poll. See “Appendix” for more information on the matched candidates

sample, a fact which becomes very apparent when examining the upper middle quartile of the plot. Even female candidates *above the median* in terms of polling accuracy mostly have values which are *less than* zero indicating that polls underestimate their support. Figure 1 thus dramatically demonstrates that polls for female candidates are much more likely to underestimate their support compared to polls for white male candidates.

While polls may be quite accurate in predicting the winning candidate, about half of the polls included in this analysis did not predict the margin of victory within the margin of error. In spite of the similarities in terms of the proportion of elections predicted within the margin of error, there appears to be a clear directionality to the nature of the error for female candidates. In nearly 70% of the female cases, the pre-election polls *underestimate* performance in the election for these candidates. This is in contrast to the set of white males, for which polls are much closer to being equally split (43–57%) between underestimating and overestimating support.

As we discussed earlier, female candidates may not be the only ones susceptible to polling inaccuracies. Therefore, we use the paired sample t-test to compare the total number of the female candidates to their male comparisons (elections with two white male candidates). The results of this *t*-test are reported in Table 2.¹⁶ When compared to white male candidates in their state who ran for US Senate or governor around the same time (within 6 years), and mostly under similar circumstances and from the same party, female candidates perform significantly better in the final results than would be predicted by pre-election polls to the tune of about three and half percentage points. This difference in polling accuracy for male and female candidates is well below the 0.01 level of significance. This finding provides further evidence that a candidate’s gender is associated—in a systematic way—with the accuracy of pre-election polls. However, we would like to be confident that possible confounds such as competitiveness, partisanship, and turnout do not affect our conclusion that the Richards Effect is real. In order to do so, we turn to a standard Ordinary Least Squares Regression model.

¹⁶ We also compared the Mosteller 5 scores for female candidates to the score of 0 for each paired test. The results show that female candidates’ scores are significantly less than zero with a T score of −3.36 and a *p*-value which is less than 0.01.

A Comparison of Female and Male Candidates: Polling Versus Electoral Performance

The results of our regression models are reported in Table 3.¹⁷ The baseline model again shows that female candidates perform significantly worse—by almost 2.7% points¹⁸—in pre-election polls than the final results when compared to polling for white male candidates. The magnitude of this variable weakens slightly but remains significant when we take into account the election context. According to Model 2, the election context model, female candidates perform 2.15% points worse in their respective polls than in the election when compared to white male candidates. When one controls for the same factors as Model 2 but substitutes voting age population (VAP) turnout for female turnout, female candidates perform better, in fact by 2.12% points, in the final election than polls predict. Finally, when we account for social context in addition to the election context, women continue to out-perform their pre-election polls by 2.41% points compared to polls for similarly situated white males. This social context model was also the strongest model used in this analysis as it explained 24% of the variance in our dependent variable.¹⁹

The only other variables that were significant in our analysis were the competitiveness measure and the size of the female labor force in the state. For every point increase in the predicted margin of victory, polls underestimated the candidate's support by about 0.1% points in the election context models and 0.092% in social context model, holding all else equal. This suggests that candidates who are losing by a large margin of victory are more likely to perform (slightly) better in the final election than polls predict. Polls also performed worse in states where women make up a large proportion of the labor force. For every one percentage point increase in the percent of women in the labor force, polls overestimate support of the candidate of interest by 0.4% points. Contrary to our expectations, the proportion of undecideds, the proximity of the poll to the election date, and turnout had no significant effect on polling inaccuracy.²⁰

¹⁷ Outliers beyond three standard deviations were excluded to ensure that they do not bias our results. This amounted to only four cases. The regression results with the outliers included are substantively similar and can be found in the Online supplementary appendix.

¹⁸ While our data show that, on average, these discrepancies are in some cases “within” the common margin of error for most polls, the operationalization of our variables and our research design, render this concern moot. Because we find significant and substantial differences in the accuracy of polling for female candidates, *when compared to* white male candidates and *on average*, our findings are nonetheless substantive. Though these findings suggest that a small proportion of survey respondents may be misrepresenting their voting intentions, even a small proportion of such respondents may alter the accuracy of predictions in close elections.

¹⁹ The results hold even with controls for yearly effects. The results also do not appear to be contingent on the matching technique. Compared to a random sample of white candidates, rather than a matched set used in the analysis in Table 3, female candidates significantly ($p < 0.1$) outperform their pre-election poll predictions. See the Online supplemental appendix for more information.

²⁰ We tested many other variables which are not included in our final models shown below. These variables include a polling firm effect, the number of respondents in each poll, a regional effect, and the Margin of Error in the poll. None of these variables were statistically significant nor did any of these variables significantly alter our results.

Table 3 Predicting the Richards Effect: OLS regression predicting the difference between the latest pre-election poll and the final results for male and female candidates from 1988 to 2008

	Baseline	Election context model	Female turnout model	Social context model
Female	–2.650 (0.949)**	–2.155 (0.921)*	–2.121 (0.920)*	–2.415 (0.900)**
Competitiveness		0.100 (0.020)**	0.100 (0.020)**	0.092 (0.020)**
% Undecided		–0.027 (0.102)	–0.035 (0.102)	0.030 (0.103)
Days to election		0.007 (0.047)	0.007 (0.047)	–0.001 (0.048)
Incumbent in race		–1.239 (0.938)	–1.203 (0.936)	–1.003 (0.914)
Third party/DK		0.015 (0.062)	0.334 (0.947)	0.010 (0.061)
Democrat		0.416 (0.947)	1.401 (0.907)	0.599 (0.932)
Governor		1.454 (0.908)	0.008 (0.062)	1.394 (0.890)
VAP turnout		–0.252 (0.523)		–0.281 (0.508)
Female turnout			–0.526 (0.506)	
Women in state leg				–0.002 (0.023)
AAUW score				0.020 (0.016)
Partisanship of state				–0.095 (0.084)
Female labor force				0.422 (0.112)**
Constant	0.526 (0.718)	0.456 (1.715)	0.619 (1.718)	–26.69 (7.157)**
Observations	211	205	205	205
R ²	0.04	0.18	0.18	0.24

** $p < 0.01$, * $p < 0.05$, + $p < 0.1$. DV: (for all models) Latest pre-election poll margin of victory – Final result margin of victory. Positive results indicate that candidate performed better in pre election polls than in the election. Negative results indicate a Richards Effect. Standard errors are presented within parenthesis. No Cook's D score was greater than 0.06

Context and Polling Discrepancy

The regression results from the previous section provide consistent evidence that voters may be more apprehensive about admitting their support for female candidates than they are for male candidates. In the theory section, we argued that this result may be explained by men and women trying to appear more conservative on gender related issues to distance themselves from feminism. Our findings thus far are consistent with hypothesis B. Nonetheless, the data presented in Table 3 alone does not provide sufficient evidence that voters are falsifying their preferences for female candidates in order to appear more gender conservative. Therefore, we explore the state context in search of additional support for this hypothesis.

As we mentioned in the theory section, if voters falsify their preference to support male candidates when they intend to vote for female candidates, we should see larger polling discrepancies in context where residents adhere to traditional gender roles and in states where voters are not supportive of advancing a gender progressive agenda, as potential voters in such areas may be more reluctant to disclose their (true) support for a female candidate.

While there are not direct measures of these conditions, there are several proxies that we utilize to approximate these contexts. First, to measure whether states adhere

to traditional gender roles we can examine the percent of women in the work force and the number of women in state government. Several studies have shown that voters are much more supportive of women in politics in states where a large percentage of women are in the workforce (Iversen and Rosenbluth 2008; Matland and Studlar 1998; Andersen 1975; Andersen and Cook 1985). When more women enter the work force, support of traditional gender roles for women generally declines and voters are more approving of women in politics (Andersen 1975; Andersen and Cook 1985). Based on this research, we should expect that states with fewer women in the work place would be more conservative about traditional gender roles in politics and be the most apprehensive about professing their support for female candidates. Therefore, there should be additional evidence for hypothesis B if the results demonstrate that polls tend to more drastically under-predict female candidates' performance in states where fewer women are in the workforce.

We also expect that states with fewer women in the state legislature will be less accepting of women in politics. The amount of female representation in the state thus should be a valid proxy of the progressiveness of a state regarding women and politics. If polls underestimate support of female candidates at a greater rate in states with fewer women in the state legislature, then we consider this as additional support for Hypothesis B.

Finally, to measure the culture toward women's issues in a state we will look at the representatives voting records towards women's issues in that state. While there may be a slight disconnect between voters and their representatives, there is a vast literature which shows that US House and Senate representatives follow the public opinion of their constituency when voting (Page et al. 1984; Erikson 1978; Bartels 1991). Therefore, representatives in states where voters tend to be more conservative with regards to gender issues should have poor voting records on bills which are considered to be supportive of women's issues.

To further test Hypothesis B, we run several OLS regressions predicting polling accuracy for only women US Senate and gubernatorial candidates from 1989 to 2008. This method will provide information on *which* female candidates are most susceptible to polling inaccuracies. The first regression model only includes variables that measure the percent of women in the labor force, the average American Association of University Women voting record scores of US Senators and Congressional Representatives in each state and year, and the percent of women in the state legislature. The second model uses these same variables but also includes controls for several election and social context variables to isolate the effect of the state's culture toward gendered issues on polling accuracy for female candidates. In addition to running an OLS regression for only female candidates, we run one model with all the controls for male candidates to determine whether the partisanship of the candidate and the social context effects on polling accuracy are unique to female candidates.

The results from all three models provide additional support for Hypothesis B. In the first model, polls are significantly more likely to underestimate female candidates' support when there are fewer women in the labor force. For every one percentage point increase in the number of women in the labor force, polls overestimated support for female candidates by a half of a percentage point. Both the average AAUW scores

for the state and the percent of women in that state legislature were inconsequential without the appropriate controls.

In the second set of models, polls in states where women make up a small portion of the labor force continue to significantly under predict support for female candidates. In the full model, for every one percentage point decrease in the percent of women in the labor force, polls underestimate support of female candidates by about a fourth of a percentage point. Also, when controlling for the election context, polls for women candidates in states where Congressional representatives and US Senators have low AAUW scores are significantly more likely to underestimate female candidates' electoral support.²¹ The final model on Table 4, which replicates the election context model for only white male candidates, demonstrates that the male candidate's polling accuracy is not significantly influenced by the percent of women in the labor force nor the average AAUW scores for representatives in their state. This indicates that these variables have a unique effect on female candidates.²²

Contrary to expectations, the percent of women in that state legislature had no effect on polling accuracy for female candidates. While this result goes counter to our expectations based on Hypothesis B, there are several reasons why percent women in the state legislature may not have the influence that we expected. First, because most Americans are unaware of how many women are in the state legislature (Sanbonmatsu 2003), female representation at the state legislature may not have a significant effect on altering perceptions about women in politics. Second, the partisanship of the women in the state legislature may influence this result. If a large proportion of the women in the state legislature are Republicans who advocate for traditional gender roles, then this measure may not accurately gauge how progressive the state is on gender issues.

The only other variable that was significant in any of the specifications is the predicted margin of victory. For every one percentage point increase in the female candidate's support, she performed 0.12 points worse in the actual election. For male candidates, polls understated their support by 0.06% points for every percentage point lead they had on their opponent.

Discussion

In the 1990 Texas Gubernatorial election, many polls predicted Clayton Williams to replace the retired Bill Clements as the Governor of Texas. To everyone's surprise, Williams lead (which was as much as 8 points in some polls) dissipated on election

²¹ A fourth model was not included in this analysis but is available in the appendix which substitutes for Female Turnout for VAP Turnout in the state.

²² We tested several interactions to determine whether female candidates who run for Governor or who run in states where the Democratic Presidential candidates perform well, and states with higher percentages of women in the labor force or with higher AAUW scores are more susceptible to a Richards Effect. None of the interaction variables was significant. Due to the few cases in our data set, we have inadequate statistical power to properly test these interactions. In other words, the difference between male candidates, female candidates and the interacted variables would have to be very large in magnitude to produce significant results with such a small sample size.

Table 4 Context, candidates characteristics, and polling inaccuracy: OLS regression predicting sources of polling inaccuracy for only women candidates from 2000 to 2008

	Baseline model (female only)	Full model (female model)	Full model (male only)
Female labor force	0.468 (0.157)**	0.408 (0.148)**	0.306 (0.225)
AAUW score	0.021 (0.016)	0.036 ⁺ (0.019)	0.051 (0.044)
Women in state leg	0.081 (0.073)	0.011 (0.072)	0.004 (0.121)
Competitiveness		0.119 (0.028)**	0.065 (0.032)*
% Undecided		0.135 (0.126)	−0.075 (0.184)
Days to election		−0.019 (0.058)	0.047 (0.093)
Incumbent in race		−1.048 (1.094)	−1.573 (1.781)
Third party/DK		−0.004 (0.068)	−0.009 (0.121)
Governor		1.061 (1.081)	1.345 (1.697)
VAP turnout		0.115 (0.623)	−0.936 (0.938)
Partisanship		−0.135 (0.159)	−0.198 (0.158)
Constant	−33.707 (9.266)**	−29.604 (8.940)**	−19.997 (12.996)
Observations	121	117	87
R ²	0.12	0.29	0.16

** $p < 0.01$, * $p < 0.05$, ⁺ $p < 0.1$. DV: (for all models) Latest pre-election poll margin of victory – Final result margin of victory (positive direction signals that polls overestimated margin of victory). A positive number indicate that the candidate performed better in pre election polls than in the election. Negative results indicate a Richards Effect. Standard errors are presented within parenthesis

day and Democratic candidate Ann Richards was victorious. When a similar polling discrepancy occurred a year earlier in Virginia, where black gubernatorial candidate Douglas Wilder's large estimated pre-election lead proved wildly incorrect as he won by less than one percentage point, many people attributed this polling discrepancy to his race. In the 1990 Texas Gubernatorial election, gender was omitted from the list of potential causes of polling error. Our study provides new insight into this puzzle. Perhaps it was not only the traditional polling problems that led polls to be less accurate, Ann Richards' gender may have also played a vital role in these polling discrepancies. Our results indicate that female candidates, and in particular female candidates from gender-conservative states, like Ann Richards in Texas, tend to do worse in pre-election polls than in actual elections.

The findings in this article suggest that there may be a stigma attached to supporting a female candidate. While polling biases for gender related issues are nothing new (e.g. Ballou and Del Boca 1980; Lueptow et al. 1990; Northrup 1992) the finding that female candidates are also susceptible to a polling bias is a novel one. In spite of the growth of high profile female candidates and elected officials in recent years, some Americans still appear to be hesitant to offer their true opinions regarding women in positions of power. While some have argued that respondents will want to appear more supportive of women politicians than they actually are (Streb et al. 2008), our results suggest the opposite. The findings in this paper indicate that voters—and especially those living in states with more traditional views—may want to appear *less* supportive of female candidates. The results in this paper suggest that pollsters and

researchers should be concerned about the gender of the candidate as much or possibly even more than they are about race.

In addition to showing that female candidates tend to see their vote shares systematically underestimated by pre-election polls, we were also able to show that it is more likely to occur in areas with a more conservative view of gender roles. While both analyses lend support to our hypotheses that voters may want to appear less supportive of female candidates as an attempt to distance themselves from feminists—i.e., they tend to falsify their true preference when responding to polls—we can only make inferences based on aggregate data. Due to the proprietary nature of polling data and the secret ballot process, testing claims regarding *why* we observe these systematic polling inaccuracies for these groups is extremely difficult. While we acknowledge that inferences about individual behavior based on aggregate data are not entirely satisfactory, this analysis offers the first explanation of these empirically robust results.

Therefore, more research into this topic is necessary, not least because 2010 has been dubbed the “year of the woman”, with female candidates in many major races across the country. Future research should focus on the role that undecided voters and partisanship play in the underestimation of electoral support for female candidates. Because the data limitations inherent in proprietary polling and the secret ballot process often prevent a more disaggregated analysis of polling response and vote choice, perhaps the best methodology for this type of an investigation would be a survey experiment. Such an experiment would allow us to isolate the factors that may affect the willingness of respondents to announce their support of a female candidate.

Moreover, future studies should conduct a content analysis of elections with female candidates to determine whether the electoral context, like social context, also influences the magnitude of the Richards Effect. It may be that races where gender is a prominent issue such as the 2010 South Carolina Republican primary, 2000 New York US Senate Election, or the 1990 Texas Gubernatorial Elections where the Richards Effect is most prominent. Further investigation of this issue is especially important given the increasing number of female candidates, and we hope this manuscript lays a foundation for such an understanding.

On the surface, the Richards Effect appears to be innocuous and in some ways may appear beneficial for the increasing number of female candidates, but under-reporting female candidates’ support can have negative consequences. There is a well documented “bandwagon” effect in American and international politics where likely voters change their preferences from an underdog candidate to support a leading candidate in order to conform into what they perceive to be the political norm (Marsh 1985; McAllister and Studlar 1991). If support for female candidates is consistently and systematically underestimated because of polling discrepancies, some voters who would have otherwise supported a woman for political office may adjust their preference to support her competitor. As a result, the Richards Effect may cause female candidates to be less competitive than they would be if polls were correctly gauging their support.

The Richards Effect could also be problematic for female candidates who are trying to attract financial backing. Candidates who perform better in pre-election polls often have an easier time raising resources for their campaigns (Gross et al. 2002).

Potential donors may be less willing to give money to female candidates if their polling support appears to be low. This may be especially problematic for female candidates, who, because they are relative newcomers to the political scene, struggle to demonstrate their viability (Carroll 1994). By underestimating support for female candidates, polls may thus weaken female candidates' appeals for financial support. Taken together, these arguments suggest that polling discrepancies for female candidates adds to an already extensive literature about disadvantages women face in the political system.

Acknowledgments We are grateful for helpful comments on early drafts of this paper from Bernard Grofman, Kelsy Kretschmer, Louis DeSipio, Katherine Tate, Carole Uhlaner, David Lublin, Tasha Philpot, Peter Wielhouwer and participants of the 2008 PREIC summer colloquium in Irvine, CA, the 2008 APSA annual meeting, and especially to the anonymous referees and Daniel Hopkins for providing very useful comments on the original submission.

Open Access This article is distributed under the terms of the Creative Commons Attribution Non-commercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

The matched cases were selected in the following way. We first selected the most recent US Senate or gubernatorial election (between two white males). Ideally, the comparison case would be from the same year as the election of interest (e.g. the election of interest is a US Senate election in 2004 then the comparison election is the Gubernatorial election in the same state and year). If this was not available, we selected an election in the preceding 2 years with two white male candidates of first the same election type (e.g. Senator at time t vs. Senator at time $t - 2$) and then we selected the other prominent statewide election with two white male candidates (Senator at time t vs. Governor at time $t - 2$). If neither of these were available we selected cases following the same order but in the following election year ($t + 2$). In the rare cases two white male candidates had not run in the same election during this time frame (only 23.6% of the comparison candidates were selected outside of this time frame), we selected cases first from the previous election cycle ($t - 4$ years) and then the following election cycle ($t + 4$ Years). If there were still no comparisons available, we repeated this approach for 6 years from the election of interest (only 1 case had a matched comparison more than 4 years from their election date).

Once we select the election of interest we then match the cases, with priority going to incumbency status. We decided to match on incumbency status first, because the candidate's vote share was a strong predictor of whether polls underestimate or overestimate their support. We believe and our data shows [Pearson's R (incumbent, Predicted Margin of Victory²³) = 0.43] that incumbency status is a strong predictor of candidate support. Female incumbents were matched with male incumbents and

²³ The predicted margin of victory is the difference between the candidate of interest's polled support and their opponent's polled support.

Table 5 Information on matched cases

Gender	Democrat (%)	Challengers (%)	Incumbent (%)	Absolute predicted margin of victory (%)
Female	66	33	25	19.08
Male	66	36	33	17.10

Comparisons are in all cases a white male in an election in which two white males compete

Table 6 Female turnout model for female candidates

	Female turnout model (female only)
Female labor force	0.414 (0.147)**
AAUW score	0.036 ⁺ (0.019)
Women in state leg	0.018 (0.072)
Competitiveness	0.117 (0.028)**
% Undecided	0.165 (0.125)
Days to election	−0.019 (0.057)
Incumbent in race	−1.275 (1.098)
Third party/DK	−0.002 (0.067)
Governor	1.274 (1.077)
Female turnout	0.726 (0.638)
Partisanship	−0.131 (0.105)
Constant	−30.482 (8.914)**
Observations	117
R ²	0.3

** $p < 0.01$,

* $p < 0.05$, ⁺ $p < 0.1$. DV:

Latest pre-election poll margin of victory – Final result margin of victory (positive direction signals that polls overestimated margin of victory. A positive number indicate that the candidate performed better in pre election polls than in the election. Negative results indicate a Richards Effect. Standard errors are presented within parentheses

female challengers with male challengers. In the case in which one type of contest had an incumbent and the comparison did not, the candidates were then matched on partisanship. Table 5 provides information on the matched cases. For results based on matching first by partisanship, rather than by incumbency status see the Online supplemental appendix. These results indicate that the findings in this paper are not overly dependent on our matching criteria (Table 6).

References

- Andersen, K. (1975). Working women and political participation, 1952–1972. *American Journal of Political Science*, 10, 439–455.
- Andersen, K., & Cook, E. (1985). Women, work, and political attitudes. *American Journal of Political Science*, 29, 606–625.
- Ballou, J., & Del Boca, F. K. (1980). *Gender interaction effects on survey measures in telephone interviews*. Paper presented at the 1980 meetings of the American Association for Public Opinion Research, Mason, OH.
- Bartels, L. (1991). Constituency opinion and congressional policy making: The Reagan defense build up. *American Political Science Review*, 85, 457–474.

- Berinsky, A. J. (1999). The two faces of public opinion. *American Journal of Political Science*, 43, 1209–1230.
- Bielby, W. T. (2000). Minimizing workplace gender and racial bias. *Contemporary Sociology*, 29, 120–129.
- Bolstein, R. (1991). Predicting the likelihood to vote in pre-election polls. *The Statistician*, 40, 277–283.
- Breinlinger, S., & Kelly, C. (1994). A test of social identity theory. *Psychology of Women Quarterly*, 18, 1–16.
- Brians, C. L. (2005). Women for women?: Gender and party bias in voting for female candidates. *American Politics Research*, 33, 357.
- Burn, S. M., Aboud, R., & Moyles, C. (2000). The relationship between gender social identity and support for feminism. *Sex Roles*, 42, 1081–1089.
- Buschman, J. K., & Lenart, S. (1996). “I am not a feminist, but”. College women, feminism, and negative experiences. *Political Psychology*, 17, 59–75.
- Carroll, S. J. (1994). *Women as candidates in American politics*. Bloomington: Indiana University Press.
- CNN. (2004). Religion and politics. Retrieved on July 1, 2010 <http://edition.cnn.com/TRANSCRIPTS/0401/08/cf.00.html>.
- Crespi, I. (1988). *Polling sources of accuracy and error*. New York: Russell Sage.
- DiCamillo, M. (1991). *Exploring gender effects in the June 1990 California gubernatorial primary election*. Paper presented at the 1991 meetings of the American Association for Public Opinion Research, Phoenix.
- Dolan, K. (1997). Gender differences in support for women candidates: Is there a glass ceiling in American politics? *Women and Politics*, 17, 27–41.
- Dolan, K. A. (2004). *Voting for women: How the public evaluates women candidates*. Boulder, CO: Westview.
- Durand, C., Blais, A., & Vachon, S. (2001). Review: A late campaign swing or a failure of the polls? The case of the 1998 Quebec election. *Public Opinion Quarterly*, 65, 108–123.
- Erikson, R. (1978). Constituency opinion and congressional behavior: A reexamination of the Miller-Stokes representation data. *American Journal of Political Science*, 22, 511–535.
- Fenwick, I., Wiseman, F., Becker, J. F., & Heiman, J. R. (1982). Classifying undecided voters in pre-election polls. *The Public Opinion Quarterly*, 46, 383–391.
- Finkel, S. E., Guterbock, T. M., & Borg, M. J. (1991). Race-of-interviewer effects in a preelection poll: Virginia 1989. *The Public Opinion Quarterly*, 55, 313–330.
- Fox, R. L. (1997). *Gender dynamics in congressional elections*. Thousand Oaks, CA: Sage.
- Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly*, 69(3), 370–392.
- Gow, J., & Eubank, R. B. (1984). The pro-incumbent bias in the 1982 national election study. *American Journal of Political Science*, 28, 224–230.
- Gross, D., Goidel, R., & Shields, T. (2002). State campaign finance regulations and electoral competition. *American Politics Research*, 30, 143–155.
- Hatchett, S., & Schuman, H. (1975). White respondents and race-of-interviewer effects. *Public Opinion Quarterly*, 39, 523–528.
- Hopkins, D. (2009). No more wilder effect, never a Whitman effect: When and why polls mislead about Black and female candidates. *Journal of Politics*, 71, 769–781.
- Huddy, L., Billig, J., Bracciodieta, J., Hoefler, L., Moynihan, P., & Pugliani, P. (1997). The effect of interviewer gender on the survey response. *Political Behavior*, 19, 197–220.
- Huddy, L., & Terkildsen, N. (1993). The consequences of gender stereotypes for female candidates at different levels and types of office. *Political Research Quarterly*, 46, 503–525.
- Iversen, T., & Rosenbluth, F. (2008). Work and power: The connection between female labor force participation and female political representation. *Annual Review of Political Science*, 11, 479–495.
- Jackman, M. R., & Muha, M. J. (1984). Education and intergroup attitudes: Moral enlightenment, superficial democratic commitment, or ideological refinement? *American Sociological Review*, 49, 751–769.
- Jacobson, M. B. (1981). You say potato and i say potahto: Attitudes toward feminism as a function of its subject-selected label. *Sex Roles*, 7, 349–354.
- Johnson, W. T., & DeLamater, J. D. (1976). Response effects in sex surveys. *Public Opinion Quarterly*, 40, 165–181.
- Kantor, J., & Swarns, R. (2008). A new twist in the debate on mothers. Retrieved on July 1, 2010 <http://www.nytimes.com/2008/09/02/us/politics/02mother.html#>.

- Keeter, S., & Samaranayake, N. (2007). Can you trust what polls say about Obama's electoral prospects. Retrieved March 12, 2008, <http://pewresearch.org/pubs/408/can-you-trust-what-polls-say-about-obamas-electoral-prospects>.
- Kornheiser, T. (2000). The good, the bad, and the ugly. *Washington Post*, November 19, p. F01.
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72(5), 847–865.
- Krysan, M. (1998). Privacy and the expression of white racial attitudes: A comparison across three contexts. *Public Opinion Quarterly*, 62, 506–544.
- Krysan, M., Schuman, H., Scott, L. J., & Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly*, 58, 381–399.
- Leeper, M. S. (1991). The impact of prejudice on female candidates: An experimental look at voter inference. *American Politics Research*, 19, 248–261.
- Lueptow, L. B., Moser, S. L., & Pendleton, B. F. (1990). Gender and response effects in telephone interviews about gender characteristics. *Sex Roles*, 22, 29–42.
- Marsh, C. (1985). Back on the bandwagon: The effect of opinion polls on public opinion. *British Journal of Political Science*, 15, 51–74.
- Martin, E. A., Traugott, M. W., & Kennedy, C. (2005). A review and proposal for a new measure of poll accuracy. *Public Opinion Quarterly*, 69, 342–369.
- Matland, R., & Studlar, D. T. (1998). Gender and the electoral opportunity structure in the Canadian provinces. *Political Research Quarterly*, 51, 117–140.
- McAllister, I., & Studlar, D. T. (1991). Bandwagon, underdog, or projection? Opinion polls and electoral choice in Britain, 1979–1987. *The Journal of Politics*, 53, 720–740.
- McDermott, M. L. (1997). Voting cues in low-information elections: Candidate gender as a social information variable in contemporary United States elections. *American Journal of Political Science*, 41, 270–283.
- Mitofsky, W. J. (1998). Review: Was 1996 a worse year for polls than 1948? *The Public Opinion Quarterly*, 62, 230–249.
- Mosteller, F. (1949). Measuring the error. In F. Mosteller, et al. (Eds.), *The pre-election polls of 1948*. New York: Social Science Research Council.
- National Council on Public Polls (NCPP). (1997). Polling council analysis concludes criticisms of 1996 presidential poll accuracy are unfounded. Press release, February 13. National Council on Public Polls, Fairfield, CT.
- National Organization for Women. (2008). Glenn Beck on ABC's Good Morning America. Retrieved on July 1, 2010 http://www.now.org/issues/media/hall_of_shame/.
- Northrup, D. A. (1992). *Gender, gender, gender: The effect of interviewer's gender on respondent's answers to affirmative action items*. Working paper. North York, ON: Institute for Social Research.
- Page, B., Shapiro, R., Gronke, P., & Rosenberg, R. (1984). Constituency, party, and representation in congress. *Public Opinion Quarterly*, 48, 741–756.
- Perry, P. (1979). Certain problems in election survey methodology. *The Public Opinion Quarterly*, 43, 312–325.
- Reeves, K. (1997). *Voting hopes or fears?: White voters, black candidates & racial politics in America*. New York: Oxford University Press.
- Roy, R., Weibust, K. S., & Miller, C. T. (2007). Effects of stereotypes about feminists on feminist self identification. *Psychology of Women Quarterly*, 31, 146–156.
- Sanbonmatsu, K. (2002). Gender stereotypes and vote choice. *American Journal of Political Science*, 46, 20–34.
- Sanbonmatsu, K. (2003). Gender-related political knowledge and the descriptive representation of women. *Political Behavior*, 25, 367–388.
- Sanbonmatsu, K., & Dolan, K. (2009). Do gender stereotypes transcend party? *Political Research Quarterly*, 62, 485–494.
- Sapiro, V. (1983). *The political integration of women*. Urbana Champaign: University of Illinois Press.
- Streb, M. J., Burrell, B., Frederick, B., & Genovese, M. A. (2008). Social desirability effects and support for a female American president. *Public Opinion Quarterly*, 72, 76–89.
- Stromberg, D. (2008). How Large is the Bradley Effect and Does it Matter for Obama? November 1, 2008. Draft. Institute for International Economic Studies, Sweden.
- Templin, C. (1999). Hillary Clinton as threat to gender norms: Cartoon images of the first lady. *Journal of Communication Inquiry*, 21, 20–36.

- Theriault, S. W., & Holmberg, D. (1998). The new old-fashioned girl: Effects of gender and social desirability on reported gender-role ideology. *Sex Roles*, 39, 97–112.
- Tourangeau, R., & Smith, T. W. (1996). The impact of data collection mode, question format and question context. *Public Opinion Quarterly*, 60, 275–304.
- Traugott, M. W. (2005). The accuracy of the national preelection polls in the 2004 presidential election. *Public Opinion Quarterly*, 69, 642–654.
- Traugott, M. W., & Price, V. (1992). A review: Exit polls in the 1989 Virginia gubernatorial race: Where did they go wrong? *The Public Opinion Quarterly*, 56, 245–253.
- Twenge, J. M., & Zucker, A. N. (1999). What is a feminist? *Psychology of Women Quarterly*, 23, 591–605.
- Williams, R., & Wittig, M. A. (1997). “I’m not a feminist, but...”: Factors contributing to the discrepancy between pro-feminist orientation and feminist social identity. *Sex Roles*, 37, 885–904.
- Wright, G. C. (1990). Misreports of vote choice in the 1998 NES senate election study. *Legislative Studies Quarterly*, 15, 543–563.