A Dual-Frame, Multi-Mode Sample Survey Design for Exit Polls in States

With Election Day in-Person, Early, and By-Mail Voting*

Geoffrey Cannon, Howard Christensen, David Magleby,

Joseph Olsen and Dan Williams

Brigham Young University

A Dual-Frame, Multi-Mode Sample Survey Design for Exit Polls in States

With Election Day in-Person, Early, and By-Mail Voting

Geoffrey Cannon, Howard Christensen, David Magleby,

Joseph Olsen and Dan Williams

Brigham Young University

Abstract

Adopting no-excuse absentee voting, vote-by-mail, and early voting centers will change how exit polling is done in the future. Conventional Election Day exit poll sample designs are incapable of reaching the early voter. Until 2010 the Utah Colleges Exit Poll relied on a sample based on a geographical frame consisting of strata/counties, and polling places within counties that ultimately connected the survey to the voters. With two large counties that have been part of the sampling frame for the exit poll adopting some form of vote-by-mail in 2014 we expanded our sample design adopting a dual frame, multi-mode approach.

This paper describes changes in sample design for the Utah Colleges Exit Poll and our 2014 experiences in using a second sample design. We had four different modes of data collection: the traditional, Election Day paper ballot, an iPad-based ballot, a telephone ballot, and Internet ballot. The latter two modes utilized data provided by the state on who had voted, with the data updated daily. We found some significant differences in who votes early in-person, by mail, and on Election Day. We also comment on lessons learned for our 2016 exit poll. Our findings from this paper are valuable for exit pollsters both at the state and national level as we look at the likelihood of more and more states offering or requiring vote-by-mail or other early voting options.

A Dual-Frame, Multi-Mode Sample Survey Design for Exit Polls in States

With Election Day in-Person, Early, and By-Mail Voting

Geoffrey Cannon, Howard Christensen, David Magleby,

Joseph Olsen and Dan Williams

Brigham Young University

Exit polls have been around for more than a half century. In writing a "brief history" of exit polling, Fritz Scheuren and Wendy Alvey reported the first exit poll may have occurred in the 1940s but had a more evident presence in the American political scene in the late 60s and early 70s (2008, 5). Warren Mitofsky, regarded as the father of exit polling, conducted his first exit poll in the 1967 Kentucky Gubernatorial election (Best and Krueger 2012, 3). Since then, exit polls have become common and have adopted modifications as voting has changed.

Exit polls have largely been conducted by the news media as a way of being able to project election winners more quickly and to assist in analyzing why people voted the way they did. The KBYU-Utah Colleges Exit Poll (UCEP) is an academic undertaking intended to help students in political science, statistics, and communications learn about elections, survey research, and survey sampling.

The UCEP is the longest continuous exit poll operating under a common leadership model in the United States. The UCEP began in 1982 and has been conducted statewide each biennial general election since then, along with a few primary and local elections. Until the last three election cycles, there were one or more media polls also conducted in these elections in Utah. Presently the UCEP is the only exit poll still in operation.

During this same time period several national organizations have also conducted nationwide exit polls. Initially, several major media organizations fielded their own exit polls,

but they later pooled resources with access to a common data set from the Voter Research and Surveys (VRS) from 1990 to 1992, the Voter News Service (VNS) from 1994 to 2003, and Edison Mitofsky since 2004 (Best and Kruger 2012, 6-9).

The UCEP's major purpose is to provide an educational experience for students at Brigham Young University (BYU), along with students from six to seven other Utah colleges and universities. First undertaken by David B. Magleby of the BYU Political Science Department, who enlisted the help of Howard B. Christensen, a member of the BYU Department of Statistics, the exit poll has garnered a reputation for accuracy. This collaboration has persisted since then for every general presidential and mid-term election up to the present. Though the faculty personnel have changed over time, the overall objective and organization has persisted. The project has a media partner, the PBS-affiliated station at Brigham Young University, KBYU, which has made the exit poll a central part of an Election Night program since the poll's inception. The poll has been cited in state policy debates, especially following votes on taxes and school vouchers. In some instances the poll has also addressed local elections, but the primary focus has been on statewide general elections. Data from the poll is used by students and faculty for academic projects and made available on the UCEP's website.[1]

The Context of the 2014 Utah Colleges Exit Poll

The 2014 general election was the first midterm election held after the 2010 U.S. census and the 2012 reapportionment and redistricting, which created some new voting district boundaries and new boundaries for all four of Utah's congressional districts. Boundary changes made predicting turnout difficult because important baseline data was unavailable for new precincts and new polling places. Because 2012 was a presidential election, and because

---

[1]The UCEP data from 1982-2014 is available at http://exitpolldata.byu.edu/

Republican candidate Mitt Romney had connections to Utah, turnout in 2012 was not seen as an adequate predictor of turnout in 2014.[2] However, individual-level vote history data was available from the Lieutenant Governor's office in the Utah Voter File. Therefore, the projected turnout estimates for 2014 were calculated using the 2014 voting district boundaries, with individual-level turnout from the 2010 midterm election as the baseline.

Utah, like many states, has seen increased use of absentee voting in the past decade. While it is too early to compare across all states for 2014, Michael McDonald of the University of Florida has tracked the rate of early voting in 2014 and compared it to total ballots cast in 2010. States range from a low of 5 percent of early votes in 2014 in Kentucky and Virginia. Colorado's 2014 early vote was 88 percent.[3] Utah had 218,446 early votes and in 2010 had a total vote of 653,272 for 33 percent (McDonald 2015). Utah first adopted no-excuse absentee voting in 2003 for the 2004 election cycle.[4] In 2006 state law was again changed to allow early voting centers, and starting in 2012, some counties substituted vote by-mail for in-person Election Day voting (Lee 2015). Until 2014, the counties using by-mail voting were sparsely populated rural counties often very far from population centers. But that changed in 2014 when Cache County adopted the practice, and an even larger county, Davis, also encouraged vote by-mail. Davis County encouraged by-mail voting by sending all voters a ballot but allowed voters the option of dropping off their ballot or voting in person on Election Day at a voting center

---

[2] Romney graduated from Brigham Young University, was chief executive of the 2002 Utah Winter Olympics, and previously had a residence in the state. As the first member of the Church of Jesus Christ of Latter-day Saints to be nominated for the presidency, many Utah members of the Church had this additional motivation to vote in 2012.

[3] McDonald compares early votes in 2014 against the total vote in 2010 because at the time he did the analysis the total vote for 2014 had not been released.

[4] Section 20A-3-304, *Utah Election Laws*, 2004 Edition. Prior to this voters needed to claim a legal disability, jury duty, or needing to be absent from the voting precinct on Election Day. We thank Salt Lake County Clerk, Sherrie Swensen for her helpful comments on this section.

instead. Of the state's ten most populated counties, the rate of early and by-mail voting in 2012 ranged from 30 percent in Utah County to 49 percent in Davis County, even when including provisional ballots in the mix (Cann et al. 2014, 11). Pre-election surveys of Utahns found 87 percent favoring keeping "the current system with a choice of absentee voting by mail, early in-person voting, or voting on Election Day" (Cann et al. 2014, 13). We asked a similar question in the 2014 UCEP but included the Internet as one possible voting option. The Internet would be preferred by 18 percent of Utah voters. Removing that option to make the data more comparable to the earlier study, we find that 66 percent of Utah voters in 2014 would prefer to vote in person on Election Day, 20 percent would prefer to vote through the mail, and 12 percent would prefer to vote early at a voting center.

As a result of the changes in how Utah voters can cast their ballots we have also changed our sample design and data collection strategy. This paper summarizes our sample design and data collection process, assessing how well it worked in 2014 and what changes we plan to implement in 2016.

The Election Day Sample Design

Over the history of the Utah Colleges Exit Poll, the sample design has varied in specifics but can best be described as a stratified multi-stage design where counties or parts of counties have served as a basic stratum and polling places in counties have served as primary sampling units (PSUs). The objective of the design from the outset has been to provide estimates of the major statewide candidate races, ballot initiatives, and congressional races in the state of Utah with sufficient precision to accurately call races on Election Night. Our emphasis on precision in the congressional races enhances our accuracy in estimating the vote in statewide races, such as the vote for president, governor, U.S. senator, and controversial ballot questions.

Since 1982, the electorate has grown and changed. With population growth, the number of U.S. representatives rose from 2 to 3 in 1982 following the 1980 census. Following the 2010 census, the number increased to 4 in 2012. Boundaries of voting precincts and congressional districts have changed frequently, especially with reapportionment and redistricting after each decennial census. The sample design has been modified to account for these changes. However, adding a new district in 2012 meant more substantial boundary changes than in 1992 or 2002, when no new districts were added. These changes presented a problem in trying to estimate the number of registered voters who would turn out, especially because several congressional districts cut across a county instead of aligning along county boundaries. Since the new districts overlapped in new and different ways, and since estimating voter turnout at the county and polling place level is a crucial part of our sample design, which uses probability proportional to size (PPS) sampling, we found that our models for estimating turnout had to be revamped.

Many of Utah's 29 counties are rural counties with sparse populations having little potential to influence the outcome of a state or U.S. Congressional election. Pollster staffing constraints, distance and its related costs of transportation, and the safety of student interviewers who would need to travel meant we often excluded many of these rural counties as too difficult to sample and of little likely impact in estimating voting results. Ninety-five percent of the registered voters in the state live in only 10 of the 29 counties. Using final election counts obtained from the state and reported by county we could measure the actual coverage bias due to dropping a county or several counties out of the definition of the population of interest. As a result, we found that we could sample roughly half as many counties and still have acceptable margins of error after taking the coverage bias into account. This has meant that in the past decade we have typically sampled in only the 10 largest counties. Having the 10 largest counties

serve as our universe resulted in only a 1 or 2 percent coverage bias for estimates in the various races. This also did not affect analysis and estimates seriously except when differences of percentages were so small that the races would be declared to be too close to call (not statistically different in any event).

Because we want to predict the vote for the U.S. House race in each congressional district, our sample design has multiple independent strata for each district. Within each congressional district we also have three different questionnaires, identified by color, with questions developed by the participating students and faculty.[5] Some questions are on all of the questionnaires, but others are only found on one questionnaire form. This allows coverage of a broader range of issues but with a reduced sample size for those unique questions associated with a specific questionnaire. The number of different questionnaires in the UCEP is consistent with the four different questionnaires in national exit polls in presidential years (Best and Krueger 2012, 14). Thus, for statistical purposes we could, for example, analyze questions found only on the blue questionnaires across all of the congressional districts, which would give us a statewide sample and estimates for those questions.

Compared to national exit polls that ask between 25 and 40 questions (Best and Krueger 2006, 14) the UCEP asks more questions, ranging from a low of 37 to as many as 75 (Mock et al. 2006, 32-35). In 2014 we asked between 39 and 45 questions on our three different questionnaires with vote questions tailored to each of the four congressional districts.[6] The

---

[5] In some election cycles we have used four different questionnaires, again color coded and affording the same aggregated statewide sample on common questions.

[6] One unusual feature of the UCEP questionnaires is that it measures party identification with all seven categories of party identification plus the "other" and "don't know" responses commonly used. The question is worded as follows: Strong Democrat, Not so strong Democrat, Independent leaning Democrat, Independent, Independent leaning Republican, Not so strong Republican, Strong Republican. This listing of all seven party identification categories yields a very similar

congressional candidates running varied by district but many of the other questions were

constant for a given color of questionnaire. This feature provided a rich data set, more detailed

than other exit polls, but it also added to the complexity of survey administration.

Studies of national exit polls find that responses drop on the back sheet of the one page

questionnaire by between 3 and 5 percent (Best and Krueger 2012, 15). A rough estimate for the

UCEP in 2014 would be that the drop would be no more than about 6 percent.

Distinguishing Elements of the UCEP are:

- It has a very large sample size for a statewide exit poll, varying from 1,834 respondents

  in 1982 to 13,022 in 2014; the latter number is about two-thirds of the size of the 2008

  and 2012 national exit polls (Best and Krueger 2012, 15).

- It involves multiple colleges and universities throughout the state of Utah collaborating

  with questionnaire topics, field interviewing, and data entry.

- It utilizes far more interviewers per voting place (2-5) than the single interviewer per

  voting place used in national exit polls (Best and Krueger 2012, 16).

- In 1994, it developed a model to accurately measure non-response rates as a function of

  the skip interval of the systematic sample of voters (Mock et al. 2006, 13) by including

  nonresponse information completed by interviewers on the questionnaire form.

- It involves a series of questionnaires with common statewide questions along with a wide

  variety of other issues of interest. These questionnaires are distributed to the voters so as

  to provide statewide estimates from the questionnaires without unduly burdening the

---

pattern of behavior as that found in branching questions used by the American National Election
study and the Cooperative Congressional Election study. National exit polls thus miss the key
differences between independent leaners and pure independents. For wording of national exit
polls' questions on party identification see Best and Kruger 2012, 103.

respondent with a long, multi-page questionnaire in order to cover the same

comprehensive set of topics. We began doing this in the 1986 off-year election, and ours

was probably the first exit poll to do this (Mock et al. 2006, 12).

- It pulls everything together between the summer and the first two months of the fall

  semester in which the election is held with a new cohort of students, most of whom have

  not had any previous experience with the exit poll.

Improvements in sampling software for analysis and error estimation have also

contributed to changes in certain design elements. For instance, in the early years we aimed at

developing a self-weighting sample using probability proportional to size (PPS) principles. This

became constraining in smaller counties with fewer registered voters as we needed at least 2

PSUs per stratum in order to allow for error estimation. Software like Sudaan, SAS, Stata, and

SPSS now allows us to incorporate weights in a more natural way and also provides improved

techniques for estimating standard errors. These were important innovations that have led to

changes in our design in recent years.

The sample frame used for Election Day sampling is geographically based and maintains

features of the sample frame used over the more than 30 years we have been doing this project.

To create this sample frame, we first formed groups of counties, individual counties, or portions

of counties to serve as strata in our sample design.[7] Polling places within counties served as the

primary sampling units (PSUs) and consisted of 1 or more voting precincts depending upon

decisions of county officials and available voting locations in the county. The secondary

sampling units (SSUs) are the voters exiting the polling places. These individuals were chosen

---

[7] We used voting precincts as our PSUs in 1982 and 1984. But the difficulty of identifying the
correct voters to interview in locations with multiple precincts prompted us to begin using
polling places in 1986 (Mock et al. 2006, 11-14).

using a systematic random sample with a random start based on turnout projections as well as the expected response rate for the polling place.

We use strata as the basic, first level of creating the sample design. Strata by definition are independent of one another. Using the information gathered in each stratum, we estimate the total vote and the total vote for a candidate in the stratum. We do this for all strata and since the strata when combined comprise the total universe, we add the total votes for the *candidate* together (for a state race we add across all strata, for a congressional race we add across the strata comprising the congressional district). Next we add the total votes for all candidates across the same set of strata. We divide the total vote for the candidate by the total vote and we get the voting proportion for the candidate. The estimates for each stratum are subject to sampling variation, so we compute the variance of the totals, stratum by stratum, using familiar statistical methodology. The total variance is the sum of all of these stratum variances. From this we compute the margin of error for each variable of interest. Stratification typically reduces error while clustering by PSU contributes to greater error. Our experience over time is that our extensive stratification more than compensates for the increased variation due to clustering found in the polling places.

The Election Day Sampling Frame: 2012--2014

In 2012 and 2014**,** we defined a total of 14 strata created from the 10 most populous counties in the state. This number occurred because of the way the four congressional districts (CDs) were created as a result of the 2010 census. Salt Lake County was subdivided among three of the congressional districts, districts 2, 3, and 4; Davis County, another large county, was split between districts 1 and 2; and Utah County was split between districts 3 and 4. These 14 strata and the CDs with which they are affiliated are presented visually in the figure below.

Figure 1. Counties as Sampling Units in Utah's Four Congressional Districts
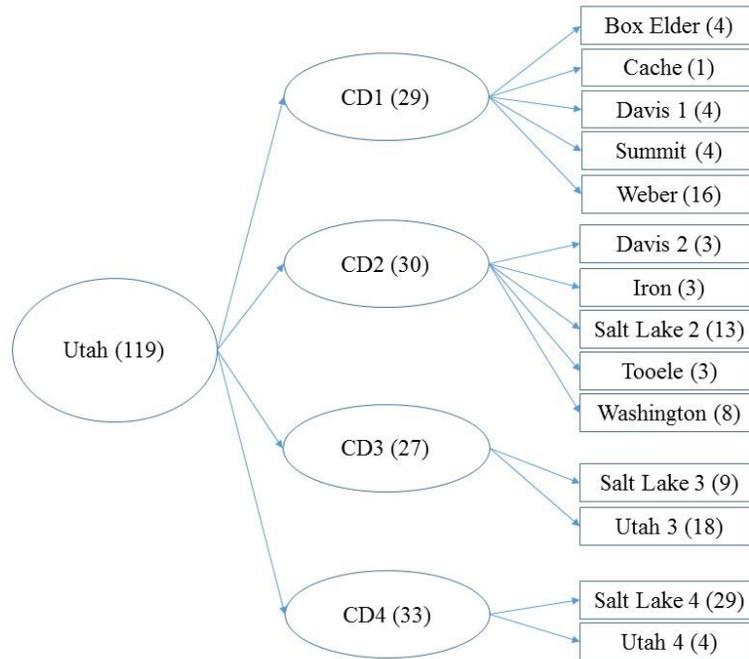
Note: The number in parentheses is the number of polling places selected in the county.

A Dual Frame Multi-Mode Strategy for Early and Election Day Voters

Election Day Sampling Frame

Historically, we have selected approximately 30 polling places in each congressional district from which voters are to be sampled on Election Day. In 2014, with four congressional districts, this policy would have led to a total of 120 polling places for our Election Day sample. However, we actually selected 29, 30, 27, and 33 polling places, respectively, from congressional districts 1 through 4. We increased the number of voting places in our sample in the 4th Congressional District from 30 to 33 because we anticipated the contest would be closer there, and we reduced the number to 27 in District 3 and 29 in District 1. We justified this based on a study of variances within strata, using a model developed during the summer of 2014 using the data from the Utah Voter File. This decision turned out to be helpful as the 4th district race was indeed the closest one in the state.[8]

Until 2006, interviewers were at their assigned polling locations the entire day. Since 2006, our sample has been divided into two shifts because research on our polls found no time-of-day effect in the way people vote (Veach 1989). This assumption was re-examined in an internal study done in the summer of 2006 during a primary election, which confirmed the results of Veach's study (Chipman 2006). In a given stratum, half of the polling places, chosen at random, were assigned interviewing teams for the first half of the voting day, where interviewers arrive at the assigned polling place at 6:45 a.m. (the polls open at 7:00 a.m.) and stayed until 1:00 p.m. The afternoon shift runs from 2:00 p.m. until the last voters leave between 8:30 and 9:00 p.m. The second shift would be assigned to different polling locations. This provides coverage of

---

[8] At 8 p.m. after the polls closed we projected the vote to be 50 percent for Republican Mia Love, 47 percent for Democrat Doug Owens, and 3 percent for other candidates. It turns out the projected and actual vote were the same.

all the sampled polling places in a congressional district with half as many interviewing teams and it allows us to require interviewers to only work a half-day shift. We have opted for this latter approach because it makes the recruiting of team members easier and takes less of their time.[9] These two shifts are to approach enough voters to produce about the same number of completed interviews per shift after accounting for nonresponse.

When Davis County, one of the larger counties in terms of population and registered voters, decided to strongly encourage voting by-mail in 2014, we were concerned that missteps on how we dealt with this change could have serious consequences when making estimates. Davis County established seven voting centers that would receive ballots of voters who preferred to drop them off rather than posting them by regular mail. These voters would show up on the Utah Voter File as having voted early. However, these same seven voting centers would also be open on Election Day to receive drop-off ballots or for those wanting to vote using a voting machine on Election Day. Therefore, Davis voters could consist of both early voters who dropped their by-mail ballots off on Election Day as well in-person Election Day voters.

Because Davis County is divided into two congressional districts and because Election Day voters from either congressional district could vote in any of the seven voting centers, it meant our paper questionnaire mode of data collection was not optimally suited for this setting because of the challenge of connecting the proper questionnaire to the voter. The paper questionnaires are specific to each congressional district, having only the congressional candidates listed who are running in that district. Rather than have an intrusive set of filtering questions about possible sets of candidates so that we could give a voter the correct form, we opted to use iPads for data collection in Davis County for in-person, Election Day voters. We

___

[9] It should be noted that interviewing teams come from universities and colleges that are as close to the polling place sites as we can possibly make it.

had found that a person's zip code would uniquely connect them to the correct congressional district with the exception of one zip code. In that instance we asked respondents which congressional district race they had voted in. Thus, the iPad could easily bring up the right questionnaire as soon as the zip code was entered on the screen.

This system had the potential to deal not only with connecting the voter to the correct questionnaire, but it also reduced another potential source of error when transferring the data to the database. For all other Election Day polling places, data was recorded by the voter on a paper questionnaire and then called into our phone center for data entry. An error in correctly coding the data could occur between the remote interviewer and the coder receiving the data over the phone and then another error could occur when the coder keys in the data. Though we have taken steps to reduce the potential for such errors, using the iPad eliminates both of these error sources and also reduces the time to convey the data from the field to the database.

Cache County, another county in our sample, had also decided to vote by mail and had a single voting center for those waiting until Election Day to drop off their ballot or to vote with a voting machine. We stationed student pollsters at the voting center, and they used the regular exit poll strategy for selecting the voters exiting the voting center. Early voters in Cache County were sampled as a part of the early voting procedure described in the following section.

The Early Voter and By-Mail Sampling Frame in 2014

With counties in our sample adding additional modes of voting, we adopted a dual-frame methodology to have access to early voters. Our frame, used to sample early voters, vote-by-mail, and absentee voters, was the Utah Voter File amended by the "Has Voted File," which we obtained from the Utah Lieutenant Governor's office on a daily basis starting October 11, 2014. This file is a list of all registered voters in the state of Utah and includes the voter's addresses,

voting precinct, congressional district, and, in some cases, a phone number.[10] This updated file

ultimately became our early voter, by-mail sampling frame.

We drew three samples that we called waves from this file. The first wave consisted of a

sample from all voters from our sample counties who were recorded as having voted between the

first day of early voting, which was approximately October 6, up to the early morning of October

15. The second wave consisted of all voters whose vote was recorded between October 16 and

the early morning of October 24. The third wave consisted of those voters whose vote was

recorded between October 24 and the early morning of October 30. This produced a coverage

gap between October 31 and Election Day, November 4, since there was not enough time or

resources to contact voters by mail to ask them to complete a questionnaire.

Next, each wave was further stratified using the same stratum definitions that our

Election Day sample followed. This produced 14 strata per wave for a total of 42 strata. A

systematic random sample of voters, with a random start, was selected from each of the 42 strata,

which we treated as an SRS for purposes of estimating voting percentages and their standard

errors. This extensive stratification would increase the precision of our estimates as well as

provide a seamless method for combining errors for various variables and domains of analysis.[11]

Our aim was to obtain 100 completes at each polling place on Election Day and 300 completes

for each early vote stratum. We met these targets for all of our early vote strata except for

Summit County and overall had more than 100 completed interviews in 34 percent of our polling

---

[10] In March of 2014 legislation took effect removing from the current voter file the voter's age.
In versions of the voter file we accessed later in 2014 to draw our sample of early and by-mail
voters we linked these later versions to the March 2014 version to obtain voter ages. We also
supplemented the downloaded 2014 voter files by submitting them to vendors to secure
telephone numbers.

[11] In their book on exit polls, Best and Krueger (2012, 20) err in their description of the preferred
sample design for exit polls. They incorrectly claim that a simple random sample is preferred to a
stratified random sample. The opposite is the case.

places. Table 2 shows our target numbers and the actual numbers across our strata for our early

voting sampling frame as well as the numbers for the Election Day sampling frame.

Table 1. Number of Expected vs. Actual Respondents, Early and Election Day

	Early		Election Day	
	Desired	Actual	Desired	Actual
CD 1				
Box Elder	41	44	400	177
Cache	300	357	100	200
Davis	300	341	400	754
Summit	46	43	400	339
Weber	213	229	1600	1175
CD 1 TOTAL	900	1014	2900	2645
CD 2				
Davis	300	439	300	579
Iron	27	39	300	195
Salt Lake	140	212	1300	954
Tooele	33	42	300	199
Washington	100	114	800	556
CD 2 TOTAL	600	846	3000	2483
CD 3				
Salt Lake	186	191	900	639
Utah	114	175	1800	1762
CD 3 TOTAL	300	366	2700	2401
CD 4				
Salt Lake	277	311	2900	2348
Utah	23	56	400	552
CD 4 TOTAL	300	367	3300	2900
TOTAL	2100	2593	11900	10429

Multiple Modes of Collecting Data

We basically used four modes to collect data from voters in 2014. Those sampled in the

early voting waves were first contacted by postcard to alert them that they had been selected as a

part of our sample. They were given information directing them to a website where they could

fill out our questionnaire online, or they could call a toll-free number to be interviewed over the

phone. For those who had not responded to either method, we established a phone center with

student interviewers contacting members of the sample and interviewing them over the phone. This represents two modes of data collection: phone interviewing and a self-reporting system over the Internet.[12]

There were two additional modes used for the Election Day voter. In the traditional polling place, interviewers approached a selected individual—one who had been identified by our systematic sampling interval as a member of the sample—and asked them to respond to our questionnaire. Those who cooperate are given a paper questionnaire on a clipboard and self-report their responses. These results are later called into a call-in center on the campus of Brigham Young University and stored in our database. Persons not wanting to participate were recorded as non-respondents and a questionnaire was marked with a few observable items of information, and the 1 in k sample count began again to identify the next person to be interviewed.

For Davis County Election Day voters, a team member asked a respondent to agree to respond to our questionnaire. If they agreed, they would be given a tablet, enter their zip code, and then respond to the appropriate survey. These results were then uploaded to our database. In Davis County our training was inadequate for handling the number of voters coming to the voting centers. Therefore, legitimate non-respondents were recorded as usual, but others were coded as "missed." [13]

In Cache County, which also largely moved to mail-ballots, Election Day voters could vote in person at the County Clerk's office. These voters were given a regular paper exit poll

---

[12] We contracted with a private vendor to augment our phone numbers from our samples drawn from the Voter File. This helped us have more working phone numbers. For example, in wave 2 it added 1,582 extra phone numbers.

[13] In 2014 our rate of missed voters as compared to refused voters outside of Davis and Cache counties was 17 percent. In Davis County the rate of missed voters as compared to refused voters was 71 percent.

form and the polling place functioned as any other traditional polling place with regard to respondents and non-respondents. Thus, the different modes of data collection were:

- Traditional Paper Questionnaire—Self Reported

- Online Questionnaire from persons responding on a personal computer, tablet, or smart phone—Self Reported

- Online Questionnaire using a tablet on Election Day—Self Reported

- Telephone Questionnaire for pre-election voters—Interviewer Reported

(Note: We continued to staff a telephone center on Election Day to try to increase our response rate for the last wave of the early voting sample.)

Those respondents using the paper questionnaire or the two online methods saw the same set of questions, so the only sources of bias beyond questionnaire bias would be bias due to an online environment in contrast to the person-to-person interaction and the self-reporting in the field on Election Day. We assumed these biases to be minor. With regard to the outbound phone survey, because respondents were not able to view the questions themselves, a telephone dialogue had to be prepared, but the same questions were asked over the phone as found in the other modes. Consequently, there may be an interviewer bias particularly with regard to sensitive questions. We assumed these are potentially real biases but likely to be small in most cases.

Computing the Sample Weights

We have an established computational process for computing sample weights within the traditional polling place Election Day procedure. We have models that project turnout at the stratum level and at the polling place level. Polling places are sampled using PPS sampling. We also have available from prior exit polls non-response figures collected over time. From this information, we computed a sampling weight including a non-response adjustment. For our 42

early voting strata (three waves with 14 strata per wave), we treated the systematic random sample within each stratum as if it were a simple random sample (SRS) of voters with respect to creating weights and then adjusted them for non-response. This led to 42 different weights that are attached to the respondents, determined by the wave and stratum to which they belong.

The weights in the traditional Election Day design are unique to the stratum (14 of them) and the polling place within each stratum (119 in total across the strata). Our system uses the inverse of the probability of selection as the weight before implementing the nonresponse adjustment. However, the by-mail voting system adopted in Davis and Cache Counties presented some unexpected problems.

Since we had no historical data for these two counties that would allow us to estimate how many voters would turn out on Election Day in comparison to the number of early voters, we imposed a complete enumeration (interview all voters) when interviewing in these two counties on Election Day. We assumed that most of the voters would vote by mail, and we could interview everyone who showed up on Election Day. Davis county officials projected 95 percent of the vote in the 2014 general election would be cast by mail. This was the percentage of vote by mail in their Republican Primary held in early summer, and they felt it would apply for the general election. Using documents requested from the Lieutenant Governor's office, we found the actual percent of the vote cast by mail in Davis County was 82 percent. (From those same documents, 99 percent of the Cache County vote was cast by mail.) But our attempt to interview all Election Day in-person voters in Davis County did not work because we did not provide enough iPads at the seven Voting Centers. This resulted in our being unable to handle the unexpected load, and our paper questionnaire backup system was not implemented. The full

impact of this failure is still being investigated, but one effect would be to change our weights for

Davis County as well as change our procedures in 2016.[14]

Performance of the Dual Frame, Multi-mode Design

An exit poll has some significant advantages over the typical opinion polling operation

when evaluating design and sampling features as well as modes of collecting data. In an exit poll,

we make our projections of winners, and then can look at the actual vote and compare our

estimates to the actual population result. In addition, with the Utah Voter File available to us

during and after the election, we are able to make some comparisons on a few variables that are

common to both datasets.

This allows us to answer some basic questions definitively and others with enhanced

objectivity. For instance, are those who vote by mail or early different from those who vote on

Election Day? This question is important because if there are not important differences it may be

possible to rely on fewer modes of data collection, removing the need for an early-voter strategy.

Are there any observable differences between modes of data collection? Again, if the answer is

no then we may feel comfortable in relying on a single mode of data collection.

In answering these questions, the first approach we have taken is to identify the variables

that are common to the sampling frame as well as the exit poll set of variables. We found that

---

[14] The theoretical basis for sampling everyone was sound and would produce weights of 1, the probability of selection, before adjusting for nonresponse. This sample design worked well in Cache County. We had 188 contacts out of the 233 who voted on Election Day, with only 1 percent of the voters voting on Election Day in the county. The sample design fell short of expectations in Davis County in part because we had more Election Day voters than anticipated and an inadequate number of interviewers and iPads to manage the 11,924 voters who showed up. We had 1,333 completed responses from Davis County or 11 percent of those who voted on Election Day. In future exit polls we will commit more resources to bolster the response rate on Election Day in Davis County.

both the Utah Voter File and the exit poll data set contain data for age and party affiliation. We also have access online to the actual vote counts by county, congressional district, and for a set of state races from the Utah Office of Elections.[15] We can examine features of our design using these known population quantities to see what differences they might reveal about the efficacy of the sampling design in producing estimates that match the respective population.

In the following we will look at comparisons between the two sampling frames, which distinguish between early voters and Election Day voters. We will also examine differences in the modes of data collection. It should be noted explicitly that two modes—phone and online, web-based responses on a personal computer, tablet, or smart phone—are unique to the early voter sampling frame using the Utah Voter File. The other two modes, paper and iPad, are both connected to the Election Day voter frame and involve a person-to-person contact between an interviewing team member and the individual voter. Thus, there are some confounding factors that occur between modes and frames. These items will be covered in the next sections of the paper along with a comparison of nonresponse rates for the various frames and modes for which we have data.

Response Rates for Early and Election Day Voters

One clear measure of difference between the two frames is with regard to the response rates for the Election Day voter frame and the early voter frame. The following table shows the response rates in the various CDs on Election Day compared to early voting. The difference from Election Day (paper) results compared to early voting is dramatic and indicates one of our primary concerns about the confidence we have for the early voting strategy using phones and web-based self-reporting. The response rate accounts for those who were missed and those who

---

[15] These vote counts can be found at http://elections.utah.gov/election-resources

refused at voting centers and polling places. Over the years, we have experienced response rates in the high 60s, but it has been showing a steady decline in recent years. This year's numbers for Election Day are about what we experienced in 2012 and are in the upper 40s and lower 50s. The iPad rates are half that, but we hope that by cleaning up our strategy for iPad use, we would see it approach the paper rates because the approach mechanism is the same. Furthermore, we will also make sure those voters not wanting to use the iPad are given a paper survey. So the real concern is what can and should be done to improve the response rates that, though typical for current web and phone surveys, are a serious concern for us when compared to the paper mode and the excellent accuracy it has provided over the years.

Table 2. Response Rates as Percentages Within CD's, Early vs. Election Day

	Election Day (Paper)	Election Day (iPad)	Early (Web and Phone)
CD1	47	25	7
CD2	52	21	7
CD3	53	—	7
CD4	50	—	7

Comparing the UCEP and Utah Voter File Frames: Early vs. Election Day Voters [16]

Party Affiliation

The Utah Voter file was examined for data that was also a part of the Utah Exit Poll set of questions. There were two possibilities to consider: (1) voter age and (2) party identification. The party identification was self-reported when a person registered to vote and so respondents could mark Republican or Democrat, or they could select a different party or leave it blank. In

---

[16]Because our purpose was to evaluate the properties of our sample design in contrast to estimating population parameters, we have used unweighted data throughout the remaining tables that use exit poll data. Additionally, all cell entries are percentages, and the rows sum to 100 percent. Finally, all statistical tests presented are likelihood-ratio chi-square tests.

contrast, the questions on the exit poll gave the following response options: Strong Democrat, Not So Strong Democrat, Independent-leaning Democrat, Independent, Independent-leaning Republican, Not So Strong Republican, Strong Republican, Other, and Don't Know. As with party identification measures nationally, this measure has shown remarkable stability over the more than three decades it has been used in the UCEP.

The perspectives of the person responding to the party affiliation question when registering to vote compared to the voter responding to the question on the exit poll survey are likely to be very different. This difference could prevent any meaningful comparisons of these questions to one another. For instance, if a person who registered as unaffiliated became a part of the exit poll sample, she may then report herself in any of a variety of ways. Ultimately, we have no way of knowing how such a person might have responded to the exit poll question.

However, the Utah Voter File population does show differences between Election Day voters and early voters, though the differences are generally quite small. (Note that because we are dealing with population data, a difference *is* a difference. The only issue is whether the difference is important enough to warrant discussion.) Table 3 gives results for the 10-county population of registered voters from the Utah Voter File. Note that differences are only 1 to about 5 percent when comparing early to Election Day voters.

Table 3. Party Registration Among Early and Election Day Voters, 2014

	Democrat	Unaffiliated	Other	Republican	Total	N
Early	10	31	1	58	100	226,911
Election Day	9	36	1	54	100	270,747
Total	10	34	1	56	100	497,658

Source: Utah Voter File, ten counties in 2014 Utah Colleges Exit Poll (UCEP)
Note: The total number of ballots cast in Utah in 2014 was 577,973. The total number of ballots reported in this table reflects the drop from a 29 county total to a 10 county total.

In Table 3, we see that over half of Utah voters are registered as Republicans, compared to only around 10 percent who are registered as Democrats. Additionally, there are over three times as many Unaffiliated voters as there are registered Democrats. This imbalance between the two major parties and the high proportion of unaffiliated voters is evident in only a few states such as Massachusetts and Rhode Island, where there about five times as many unaffiliated voters as there are registered Republicans (McDonald 2010).

The political affiliation data for early and Election Day voters from the Utah Colleges Exit Poll is found in Table 4. It shows a result that has been consistent over time with regard to voters at the polls—that there is a strong Republican voter base in the state of Utah—but it also shows that a slightly higher percentage of early voters describe themselves to be Democrats when compared to the percentage of Election Day voters. This difference is statistically significant and potentially important as we consider using an early voter frame in the future.

Table 4. Reported Party Affiliation of Early and Election Day Voters

	Democrat	Independent	Republican	Total	N
Early	31	10	59	100	2,126
Election Day	28	10	62	100	9,182
Total	29	10	62	100	11,308

$\chi^2(2) = 8.82, p = 0.012$
Source: 2014 UCEP data

Notice in the "Total" row that the majority of voters (62 percent) report themselves to be Republican in contrast to only 29 percent who declare themselves to be Democrats. This Republican majority in Utah has been present over time as shown by previous exit polls. However, in Utah, early and by-mail voters are about 3 percent more Democratic than are Election Day voters (31 percent of early voters were Democrats in 2014 compared to 28 percent of Election Day voters). Republicans were 59 percent of early voters and 62 percent of Election Day voters, while Pure Independents were 10 percent of both early and Election Day voters. For

this analysis we have combined Independent leaners with the self-reported members of the

parties towards which they lean (Keith et al. 1992).

Another example of differences in early and Election Day voting can be seen in the 2014

Utah 4th district congressional race. Early voting favored the Democratic candidate Doug Owens,

while Election Day voting favored the Republican candidate and eventual winner Mia Love in

one of the most closely watched races in the state (see Table 5).

Table 5. U.S. House Vote of Early and Election Day Voters, Congressional District 4

	Love (R)	Owens (D)	Someone Else	Total	N
Early	46	52	2	100	300
Election Day	57	40	3	100	2,775
Total	56	41	3	100	3,075

$\chi^2(2) = 17.21, \ p \ < 0.001$
Source: 2014 UCEP data

There are important differences between early and Election Day voters with respect to

age and party affiliation as far as the population is concerned. However, these differences tend to

be more pronounced in the exit poll data. We also see some of these differences showing up in

the various CD races as noted in Table 5.

Age

The comparison of the age distribution for early voters vs. Election Day voters is slightly

more dramatic both in the population at large (Utah Voter File) and in its reflection in the exit

poll data. Using population data from the Utah Voter File, we see that early voters, when

compared to Election Day voters, tend to shift to the higher age brackets. For example, only 10

percent of early voters are in the 18-34 year-old age group with the percentage steadily

increasing as you move to successively older age brackets, with 40 percent being 65+ years of

age (see Table 6). The Election Day age distribution consistently shifts towards the younger

voter.

Table 6. Age of Early and Election Day Voters

	18-34	35-50	51-64	65+	Total	N
Early	10	19	31	40	100	222,112
Election Day	15	31	30	23	100	262,342
Total	13	26	31	31	100	484,454

Source: Utah Voter File, ten counties in 2014 UCEP
Note: The drop of the total to 484,454 reflects the fact that the current voter file does not contain an age variable. In merging the current file to the March 2014 file we lost cases of voters who did not match on the age variable.

The exit poll data in Table 7 shows the same general tendency as the population, but the difference is larger than is seen in Table 6. Approximately 49 percent of the early voters are in the 65+ age group (compared to 40 percent for the Utah Voter File), and only about 7 percent of the early voters are in the 18-34 age range as compared to 21 percent of the Election Day voters who are in the same age range (see Table 7). We see this as an important characteristic to be considered in the future as we look at strategies to account for this and to adjust accordingly for it as well. If there are voting behaviors that correlate with age, it would be important to take that into account in future sampling and data collection strategies.

Table 7. Age of Early and Election Day Voters

	18-34	35-50	51-64	65+	Total	N
Early	7	15	28	49	100	2,210
Election Day	21	33	27	19	100	9,613
Total	19	30	27	24	100	11,823

$\chi^2(3) = 1000.00$, $p < 0.001$
Source: 2014 UCEP data

Comparing Modes: Web vs. Phone for Early Voters

It should be kept in mind that the comparison of web respondents with phone respondents applies strictly to the early voter sampling frame. Neither method was used on Election Day, though the iPad data collection process is similar in many respects to the experience of the web respondents. We have kept the iPad process separate, however, because there was a different

frame of voters—those who vote in person on Election Day—and there was an interviewer involved, in contrast with early voters taking the survey online, who had no contact with a person while filling out their form.

It should also be noted that the sample sizes are much smaller for the phone and web respondents compared to the sample size for the Election Day respondents. It is also important to note that the response rate for both of these groups is less than 10 percent, which leads to a concern about non-response bias.

With these caveats, we note that interesting differences can be found between web respondents and phone respondents in terms of party identification, age, and education. Phone respondents tended more to be Republican than web respondents (see Table 8).

Table 8. Party Affiliation of Web and Telephone Respondents

	Democrat	Independent	Republican	Total	N
Web respondents	36	11	53	100	1,114
Telephone Respondents	26	9	65	100	1,012
Total	31	10	59	100	2,126

$\chi^2(2) = 31.52, p < 0.001$
Source: 2014 UCEP data

With respect to the age distribution, there was a high percentage of web respondents in the age groups under age 65 and a much higher proportion of phone respondents age 65 and over.

Table 9. Age of Web and Telephone Respondents Among Early Voters

	18-34	35-50	51-64	65+	Total	N
Web respondents	10	18	34	38	100	1,147
Phone Respondents	5	11	24	61	100	1,063
Total	7	15	29	49	100	2,210

$\chi^2(3) = 128.01, p < 0.001$
Source: 2014 UCEP data

With respect to education, there was a significant difference between the web respondents and the phone respondents, with a general tendency for the web group to have a higher percentage of college and post college respondents compared to the phone respondents (see Table 10).

Table 10. Education of Web and Phone Respondents Among Early Voters

	Some High School	High School	Some College	College Graduate	Graduate Degree	Total	N
Web respondents	<1	6	23	36	34	100	1,161
Phone Respondents	2	14	32	32	21	100	1,076
Total	1	10	28	34	28	100	2,237

$\chi^2(4) = 96.36, p < .001$
Source: 2014 UCEP data

Comparing Modes for Election Day Voters: Paper vs. iPad

When comparing the paper vs. iPad respondents we must keep in mind that the iPad responses were restricted to the voters in Davis County, so when differences are observed in the exit poll data, we cannot attribute them to the iPad any more than to the differences between voters in Davis county compared to voters in other counties. As a result, making conclusions about iPad users cannot be differentiated from conclusions about Davis county voters. The iPad effects and Davis County voter effects are completely confounded.

As we look at Tables 11 and 12, we observe that paper respondents in District 1 and District 2 consistently show a higher Democratic edge compared to the iPad respondent.

However, this is more likely a reflection of the differences that we see in the actual vote, i.e.

Davis County tends to be more Republican in the vote for Congress than all but one of the other

counties in our sample in District 1.[17] Thus, this is not so much an iPad vs. paper issue as it is

Davis County versus everyone else in the 1st District. In fact, the iPad/Davis 1 vote is right in line

with the actual percentages for Davis 1 in the District 1 actual counts. (The exit poll Davis

County District 1 percentages are 67 percent Republican, 25 percent Democrat. The actual

percentages for Davis County are 69 percent Republican and 23 percent Democrat.)

Table 11. The Distribution of the Vote in the 1st District Race, Paper vs. iPad

	Bishop (R)	McAleer (D)	Someone Else	Total	N
Paper Respondents	52	44	5	100	1,779
iPad Respondents	67	25	7	100	726
Total	56	38	5	100	2,505

$\chi^2(2) = 82.65,\ p < 0.001$
Source: 2014 UCEP data

The same conclusions made in the previous paragraph can be made when examining

Table 12. That is, the Davis 2 percentages in the iPad row match (within the margin of error) the

actual percentages obtained from the final election counts reported by the state. (Exit poll: 70

percent Republican, 26 percent Democrat. Actual counts are 73 percent Republican, 22 percent

Democrat.) We can therefore see from this that the iPad does not appear to bias the vote in any

way since in both District 1 and District 2, the Davis County voting percentages matched up with

the percentages obtained in the final counts.

---

[17]The percent Republican vote in Davis County in Congressional District 1 was 66 percent compared to 52 percent in Weber County, another large county, and 58 percent in Cache County. Only Box Elder County had a higher percent Republican vote.

Table 12. The Distribution of the Vote in the 2nd District Race, Paper vs. iPad

	Stewart (R)	Robles (D)	Someone Else	Total	N
Paper Respondents	49	45	5	100	1,804
iPad Respondents	70	26	4	100	560
Total	54	41	5	100	2,364

$\chi^2(2) = 76.62, \; p < 0.001$
Source: 2014 UCEP data

Tables 13 and 14 compare paper with iPad respondents with regard to party affiliation and age distribution. Interestingly, the paper respondents tend to be more Democrat and less Republican by affiliation than do the iPad respondents. However, this is more likely a reflection of the Davis County voter vs. the non-Davis County voters rather than paper vs. iPad. Davis County tends to be more heavily Republican, and the non-Davis County voters across the state would balance that out by leaning more to the Democratic Party comparatively.

Table 13. The Distribution of Party Affiliation, Paper vs. iPad

	Democrat	Independent	Republican	Total	N
Paper Respondents	29	10	61	100	7,973
iPad Respondents	20	10	70	100	1,209
Total	28	10	62	100	9,182

$\chi^2(2) = 49.04, \; p < 0.001$
Source: 2014 UCEP data

When comparing the age distribution of the two groups, though the differences are significant, they probably are not that important, practically. And remember this is a state comparison, not a congressional district comparison. The 65+ age group is greater by 5 percent than the iPad but the 35-50 age group shows a 4 percent shift in the opposite direction. This may reflect a tendency for older voters to be non-respondents in Davis County when presented an iPad to answer the questions. But that is only a conjecture.

Table 14. Age Distribution, Paper vs. iPad

	18-34	35-50	51-64	65+	Total	N
Paper Respondents	21	33	27	19	100	8,325
iPad Respondents	21	37	27	14	100	1,288
Total	21	33	27	19	100	9,613

$\chi^2(3) = 21.14$, $p < 0.001$
Source: 2014 UCEP data

In summary, the paper questionnaire, the self-reporting mode of data collection we use—along with other exit polling organizations—is well proven and has produced good estimates for our exit polls since 1982. Our estimates show this to be true again this year. The comparison of paper to iPad in our study is possibly more a reflection of the differences in Election Day respondents in non-Davis Counties compared to Davis County voters than it is a difference due to modes.

However, there are aspects of iPad use that encourage us to consider extending it to future exit polls. First, we experienced no technology failures in the process. Secondly, it would appear that it does not bias the outcomes based on 2014 results. And for an increasingly tech savvy population, it has advantages we feel we can capitalize on. It could reduce the size of the staff needed to do telephone data entry and reduce the total time from when the voter completes the questionnaire to its input to the database, along with a reduction in data entry errors.

One decision we should have made differently was not to use iPads exclusively in Davis County simply because it solved a problem for us there. Instead, we should have distributed them to other locations with a plan in place for evaluating their advantages and disadvantages. In addition, the lower response rate for the iPads compared to the paper questionnaire does not necessarily reflect an iPad concern. Rather it reflects the need to have a better-trained team to handle both iPads and paper questionnaires in conjunction with one another.

Lessons Learned for 2016

We expect to conduct another exit poll in 2016. We do not know the extent to which more Utah counties in our sample of 10 urban counties will do as Davis and Cache counties did in 2014 and provide all voters with vote-by-mail ballots. We expect there will be some movement in that direction by other counties but do not know how much. We assume that we will continue to have a mix of counties using in-person Election Day voting as their presumed primary mode of voting. Therefore, we expect to continue as we have previously with a sample design for Election Day voting as described in this paper.

For early in-person and by-mail voters our 2014 experience reinforces our commitment to a multi-mode sample and data collection effort. Our highest priority is to make a stronger effort to reduce non-response rates, especially among early and by-mail voters. In 2014 we had a 50 percent response rate for Election Day in-person voters but a 7 percent response rate for our phone and mail based samples. In 2016 we may try to send two postcards in each wave, one only a day or two after the first to try and encourage participation. We may also begin calling early and by-mail voters more than ten days before the election to try and boost our telephone response rate. If counties mailing out paper ballots to all voters retain in-person voting centers for Election Day, we will boost the number of interviewers at those voting places to be able to interact with more voters. (We anticipate that we can do a better job of estimating voting center turnout in the future, based on the 2014 experience.) This will likely mean having more iPads at each voting center, using paper and pencil data collection when all iPads are in use or randomly interchanging them in order to compare the two approaches directly against one another at voting places. We should also place iPads in other strata under controlled experimental conditions so we can properly evaluate their effectiveness.

References

Best, Samuel. J. and Brian S. Krueger. 2012. *Exit Polls: Surveying the American Electorate, 1972-2010.* Los Angeles: CQ Press.

Cann, Damon, Thad Hall, and Quin Monson. 2014. *Evaluating the feasibility of vote by mail in Utah: A report to the Utah Lieutenant Governor's office*. Salt Lake City: Utah Lieutenant Governor's Office.

Chipman, Jonathan. J, Howard B. Christensen, and Paul J. Fields. 2006. Utah Colleges Exit Poll: 2006 Primary election within Congressional District Three. Undergraduate Research Project. Brigham Young University.

Grimshaw, Scott D., Howard B. Christensen, David B. Magleby, and Kelly D. Patterson. 2004. Twenty Years of the Utah Colleges Exit Poll: Learning by Doing. *Chance* 17, no. 2 (September): 32-38.

Keith, Bruce E., Candice J. Nelson, Elizabeth Orr, Mark C. Westlye, and Raymond E. 1992. *The myth of the independent voter*. Berkeley: University of California Press.

Lee, Justin. 2015. Interview by author David Magleby. Salt Lake City, UT. April 13.

Mock, Derek R., Howard B. Christensen, Clint L. Stevenson, Joseph A. Olsen, and Kelly D. Patterson. 2006. Utah Colleges Exit Poll: Design and methodology. Paper of the Departments of Political Science and Statistics. Brigham Young University.

McDonald, Michael P. November general election early voting. United States Election Project. United States Election Project. http://www.electproject.org/2014_early_vote (accessed April 16, 2015).

McDonald, Michael P. 2010. Partisan voter registration totals. *Huffington Post.* October 13.

Scheuren, Fritz J. and Wendy Alvey. 2008. *Elections and exit polling*. Hoboken, NJ: Wiley.

Veach, Jeffrey L. and Howard B. Christensen. 1989. The effect of time of voting on election

      results --based on exit poll data. http://www.amstat.org/sections/SRMS/Proceedings/

      papers/1989_150.pdf (accessed April 15, 2015).

Williams, Dan. Researchers use opinionology to implement new exit poll design. Opinionology.

      http://www.opinionology.com/docs/whatsnew/research/Researchers_Use_Opinionology_

      to_Implement_New_Exit_Poll_Design.pdf (accessed April 15, 2015).