

To: Heather Gerken

From: Peter Miller

11/30/07

Re: First memo on data used in NVRA, UOCAVA and EDS reports

Attached: Memo 1 data.xls

Heather,

Here is the first memo regarding data submission rates for the 50 states, the District of Columbia and 4 territories (Puerto Rico, Guam, American Samoa and the U.S. Virgin Islands) related to three reports submitted to Congress this summer: the National Voter Registration Act (NVRA) and the Uniformed and Overseas Citizen Absentee Voting Act. A third report, the Election Day Survey (EDS), covers election administration data.

The attached spreadsheet is split into three segments: survey questions related to NVRA, UOCAVA and EDS reporting requirements. In each segment, the 55 jurisdictions are sorted on the vertical axis, and survey items are arranged along the horizontal axis. The first column identifies the state, the second column establishes the maximum number of reporting jurisdictions. Thus, a response rate can be found by dividing the number in any cell by the number in the second cell. For example, you can see that 100% of jurisdictions in Alabama provided data on the total number of active voters in the 2004 election, but 65 of the 67 jurisdictions in Alabama reported the number of inactive voters in 2004.

The purpose of this memo is to identify potential pitfalls in the data contained in each report. For this purpose, I will consider five groups of data in the NVRA section (active and inactive voters, election day registration, registration manner and location, registration rejections and registration removals), three groups in the UOCAVA section (ballots cast and counted, absentee ballots and ballots rejected) and five groups in the EDS section (ballots cast and ballots counted, reasons for rejecting provisional ballots, reasons for rejecting domestic absentee ballots, undervotes and overvotes and poll worker, polls and disabled voters). Based on the response rates in the accompanying spreadsheet it will be possible to rank states according to completeness of responses to the survey questions.

At the outset, I will make a few observations about the data and general features. First, the response counts consider any number as valid (i.e. if one jurisdiction responded with "0" for a survey item, that state would have a "1" for that survey item). Secondly, the meaning of a jurisdiction varies across states. In Alaska, all data was submitted at the state level, so there is one jurisdiction for the entire state. New Hampshire has 333 jurisdictions because that state collects data at the township; there are 13 counties in the state, however (so the response rate for election day registration is closer to 70% than 2%). Third, states can be roughly categorized three ways based on response rates to survey items: states that returned complete or nearly complete data on all survey

questions; states that returned complete or nearly complete data on general items, but only partial data for minor survey items (for example, a state where most jurisdictions would report data on the number of domestic civilian absentee ballots received, but not include data on why some of these ballots would be ultimately rejected and not counted); and, lastly, states that would provide data on major survey items, but no data on other survey items (Alabama is a good example). Lastly, across the three reports, the response rates for the territories were particularly low.

NVRA Report

The NVRA report includes data on new and removed voter registrations. The data on inactive voters is most problematic, due to a lack of a uniform definition of an inactive voter across all states. The NVRA report states that 10 states do not collect data on inactive voters, treating all voters as active.

Active and Inactive Voters

These survey questions ask about the number of voters on state rolls categorized as active (meaning all registered voters except those who have been sent but have not responded to a confirmation mailing sent in accordance with NVRA (42 U.S.C. 1973gg-6(d)) and have not yet since offered to vote) and inactive (meaning registered voters who have been sent but have not responded to a confirmation mailing sent in accordance with NVRA (42 U.S.C. 1973gg-6(d)) and have not yet since offered to vote).

Generally speaking, the response rate across states is good on this critical survey item. In 2004, two states had low response rates for active and inactive voters, but high response rates for total number of voters: Mississippi (60 out of 82 jurisdictions for active voters, 43 of 82 for inactive voters) and Utah (3 of 29 jurisdictions for active voters, and 0 for inactive voters). While 10 states do not collect data on inactive voters, most states provide a top-level count of registered voters in 2004. North Dakota does not have voter registration, so it is exempt from this reporting requirement. In 2006, the quality of data improved as more jurisdictions replied to the survey, but inactive voters continue to be a category which states do not uniformly collect. 2006 is the first year centralized, statewide voter registration rolls are available for the state of Wisconsin. Among the states that provided a partial count of inactive voters, response rates improved from 2004, but eight states and North Dakota did not supply this data.

Election Day Registration

The survey instrument presented every jurisdiction with a question asking for the total number of election day registrants, while only seven states permit registration at the polls: Idaho, Maine, Minnesota, Montana, New Hampshire, Wisconsin and Wyoming. Fortunately, jurisdictions that did respond to this survey item reported “0” cases of registration on election day.

Registration Manner and Location

The next group of survey items are related to the manner and place voter registration forms were submitted in each jurisdiction between 2004 and 2006. Registration forms can be submitted either in person or by mail and can be submitted to a variety of public agencies in accordance with the NVRA text. While most states report the total number of registration applications received, the data becomes less complete for other items. Five states and North Dakota reported nothing else but a top-level count of registrations. Other states, such as Massachusetts, reported complete or nearly complete data for some of the survey items, not did not respond to other survey items in this group.

Registration Rejections and Removals

The final section of the NVRA survey dealt with voter registrations which were rejected (and specify a series of reasons) or removed from the voter rolls (again, including a reason as required by statute), and the total number of new registrations added to the voter rolls.

UOCAVA

The UOCAVA report dealt primarily with military and overseas citizen voters. However, interpretation of the results are complicated by low response rates. The report states, "Response rates varied across the UOCAVA-related questions, but in general, they are at a level that makes it difficult to monitor compliance with the basic requirements

of UOCAVA." For one example, at least 900,000 ballots were requested by UOCAVA voters (56% to 65% of jurisdictions responded to this question, depending on the category of voter as military or civilian) but only about 333,000 ballot cast or counted. It is also hard to discriminate between the number of ballots cast and the number of ballots counted, because for many jurisdictions, only one of these two numbers was provided by jurisdictions. Thus, we're left with one of two conclusions: either turnout among UOCAVA voters is very low in general elections, or data on this question is not being collected and submitted to the Election Assistance Commission.

Ballots Cast and Counted

The UOCAVA report requires jurisdictions to report the number of ballots cast and counted at the polls, by the demographic groups targeted by this legislation (i.e. overseas and domestic military voters and overseas civilian voters). While every state allows absentee voting (some in special cases such as illness or travel and others allow absentee voting without a reason), it is hard to believe that in some states there were no cases of absentee ballots or early votes cast, or that 15 states and the territories did not

have any cases of overseas military votes cast. Survey data on ballots counted among those ballots discussed above are also partial. The UOCAVA report uses a “maximum ballot” variable to compare turnout in cases where states submitted data on ballots cast or counted, but not both.

Absentee Ballots

The UOCAVA survey includes questions about the number of absentee ballots requested by the voter, sent by the election officials, and received by the election officials. As stated above, the data suggests that about one third of those voters who requested a ballot under UOCAVA returned a ballot. The concluding statement of the UOCAVA report is perhaps most illustrative of the data collection and submission process: “Too many local election authorities continue to fail to track the precise number of ballots they mailed or transmitted to their domestic military, overseas military, and overseas citizens, as required by HAVA
”

Ballots Rejected

The UOCAVA report suggests that about 48,600 ballots were rejected. The most common reason for rejecting ballots was that ballots were often returned as undeliverable. The second-most common reason for rejecting UOCAVA ballots was “other,” suggesting that the already long list of reasons for rejecting a ballot was not long enough.

EDS

The Election Day Survey includes items on general election administration. The first EDS was conducted following the 2004 elections, and work is currently underway to prepare a 2008 survey. Unlike the NVRA and UOCAVA reports, the EDS report has not yet been formally submitted to Congress, so these data are not yet finalized. The data related to overvotes and undervotes is particularly hazardous to rely upon.

Ballots Cast and Counted

As explained above, the data on ballots cast and counted are often incomplete due to states not providing one of the counts.

Reasons for Rejecting Provisional Ballots

The data on reasons for rejecting provisional ballots is, like many other sections of these three reports, plagued by poor response rates. Vermont, for example, only returned data on the total number of provisional ballots which were rejected, and did not provide a reason for rejecting those ballots. Most of the remaining states follow the case of Vermont: providing complete or nearly complete data on a top-level count of rejected provisional ballots, but the response rates drop when considering the reasons for those rejections.

Reasons for Rejecting Domestic Absentee Ballots

Like the previous section on provisional ballots, domestic absentee ballots are also subject to poor data on reasons for rejecting ballots, even if most jurisdictions know how many absentee ballots were ultimately rejected. Across reasons, response rates range from 33% to 51% of jurisdictions.

Undervotes and Overvotes

Unlike every other column in the spreadsheet, the data on undervotes and overvotes reflect numbers of residual ballots for each state. The survey instrument question asked for counts of undervotes and overvotes in each Federal race in each jurisdiction. However, because some states sort Federal election data by Congressional district and because most races had many candidates, the resulting data was virtually unusable on a comparative basis. Instead, the report reflects the total number of undervotes and overvotes for House and Senate elections in each state. Note that 18 states and the 4 territories did not have a Senate race.

Polls, Poll Workers and Disabled Voters

Lastly, the EDS survey instrument asked about the number of precincts and polling places, data related to poll workers, and accessibility of voting places for disabled voters. Rather than follow the strategy adopted by many states in the 2004 survey and uniformly report the mandated number of poll workers, the 2006 survey includes many gaps in these data. Only about 40% of jurisdictions reported the number of precincts or polls that did not have the required number of poll workers. About 80% of jurisdictions responded to questions about disabled voter access. Between 85% and 94% of jurisdictions responded to questions about the number of precincts and polling places.

Conclusion

There are many holes in the data related to these three reports. Average response rates for each survey item in the three reports range from about 54% to 65%. For the entire dataset, 15 variables are complete or nearly complete, while 54 variables have less

than a 50% response rate. While inactive voters, ballots cast and counted and undervotes and overvotes are particularly problematic to interpret, the remaining data is by no means beyond reproach.

To: Heather Gerken

From: Peter Miller

1/16/08

Re: Supplemental Data following my second memo

Attachments: adjusted versions of the second data table, and final data table

Heather,

This memo is meant to answer questions remaining from the second memo and your email from January 15th. I will cover the topics of inactive voters, the quality of the data at day 1 (with a few anecdotes) and the five questions from your email. As a result of your questions, I adjusted the North Dakota scoring, the change is slight, but I will describe it below.

The Importance of Inactive Voters

I talked with Paul Gronke about this and the other subjects in this memo. Please feel free to contact him for any further information related to this memo, or prior information related to the EAC survey. In particular, he feels that most of the survey items are included because the three pieces of legislation I covered earlier include them *by implication*, and not in an explicit list.

He explained that, in his view, voting is a two-stage process where citizens first register to vote and then vote at a later date. Inactive voters are the group of people who, for whatever reason, pass the first requirement (they register to vote) but do not actually vote when an election comes around. Tracking the number of inactive voters can give some indication of the efficacy of efforts to register more citizens, even if these citizens do not ultimately vote. You may also want to contact Michael McDonald at George Washington University, as Gronke mentioned he is an expert in questions of turnout.

The State of the Data at Day 1

The data we worked with in the NVRA, UOCAVA and EDS surveys was based on the EAC survey instrument. This web-based instrument was sent to each county in the country, with instructions for county election officials to input requested data to this form. Most states were able to upload data using the web-based form (about 10%, or 5 states, were unable to use the form, and that data was input manually by staff involved in the contract. Myself and one other person entered all the data provided from Connecticut, for example).

The survey form included both county-level data forms, and statewide forms. Ideally, each county in a state would enter data, which would then be aggregated to the state-level. However, this aggregation was never done successfully. Some states would provide certain survey items at the county level, and other survey at the state-level only. I became the person responsible for conducting the collapse process (taking the county data, summing it for each state and then passing along the state-level data for each survey item) and every time I ran a collapse, I had to take account for the 5 to 7 states that included some statewide data that had to be manually included after the collapse was complete. Generally, these states were in New England, with Wisconsin included as well. The states that are organized by township (instead of county) tended to be the states that returned data at the state-level in addition to, or in place of, county-level data.

Many states initially did not respond to the survey, or responded with obvious errors. It was only when contracted staff contacted the individual jurisdictions, enumerating the specific survey items they had neglected, or erroneously entered, that the jurisdiction officials replied with data we were confident was correct.

At day 1, the dataset we worked with to produce three reports was in poor shape. It was replete with holes and obvious errors in the data. With time and effort, we were able to build a dataset that could be analyzed and used to produce those three reports.

Five Additional Questions

In my first memo, I stated that there are 15 variables across the three datasets with complete or nearly complete coverage (99% or better coverage). These 15 variables are over-arching subjects, like the number of registered voters in 2004 and 2006, the number of registration applications received and the number of registration applications rejected, or the number of ballots cast and the number of ballots counted. All of these 15 variables can be thought of as top-level data subjects. When greater detail is sought (for example, the number of provisional ballots cast, or the number of inactive voters) the response rates drop across the survey items. The table below charts the breakdown in response rate for the remaining survey items.

	NVRA	UOCAVA	EDS
90s	0	1	3
80s	2	1	2
70s	0	4	6
60s	9	1	3
50s or less	16	27	45

As you can see, a small number of the survey data either has very good coverage, or most of the data has less than 60% coverage. The full dataset can answer a few, broad questions concretely, such as the number of registered voters, but is probably unable to definitively say how many ballots were rejected because a ballot was returned with insufficient postage.

Your second question asks for more on why I dropped the overvotes and undervotes from the scoring technique. The original survey instrument asked for the number of overvotes and undervotes in House and Senate elections in every jurisdiction. For reasons that I do not understand, the answers to these questions were returned in two spreadsheets with 55 columns. I am not familiar with the method used to arrive at the statewide counts of undervotes and overvotes. Though the data is reported in Tables 31a and 31b of the final report, the 2006 Election Administration and Voting Survey. I dropped these survey items from the scoring procedure because the number of jurisdictions that responded to the survey item is not available. Only six states provided data on undervotes cast in Senate elections, while only two provided data on overvotes in Senate elections. Data for the House elections is better, but without a jurisdiction count, I could not response rates for these two items.

Your third question asks for a list of states that did not report having any military overseas votes. Looking over the data I sent with memo 1, I see that 14 states have a zero in the “total number of ballots cast by overseas military voters.” The states are: Alabama, Connecticut, Kentucky, Main, Massachusetts, New Hampshire, New York, Oklahoma, Oregon, Pennsylvania, Rhode Island, Tennessee, Vermont and Wisconsin. The District of Columbia, Guam and Puerto Rico also did not reply to that survey item.

Your fourth question asks for a list of states that reported no absentee/early votes cast. Two survey items that answer that question: the number of domestic civilian absentee ballots cast, and the number of ballots cast by early voting. There are 9 states that did not respond to the domestic civilian absentee survey item: Alabama, Connecticut, Kansas, Massachusetts, Minnesota, Oregon (though all ballots in Oregon are reported as cast at the polling place, while they are all sent to registered voters. Oregon should probably not be included in this list), Pennsylvania, Tennessee, and Vermont (as well as DC, Guam and Puerto Rico). 13 states did not answer the early voting survey item: Alabama, Connecticut, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Hampshire, New York, Pennsylvania, South Carolina, Tennessee and Wisconsin. While these states may claim that they do not have the legal provision for early or absentee voting, all states allow for absentee voting in certain cases, such as illness, travel or attending school outside of the state.

Your last question asked if North Dakota is unfairly punished in the scoring because that state does not have any voter registration. You are correct; I had failed to take account for that when calculating each state’s score. The adjusted score for North Dakota is now 82.3%, up from 79.8%. I have attached updated tables with this memo.

To: Heather Gerken

From: Peter Miller

1/2/08

Re: Second memo on data used in NVRA, UOCAVA and EDS reports

Attached: Memo 2 data.xls and Word documents of HAVA, NVRA, and UOCAVA

Heather,

Here is the second memo regarding data response rates. In this memo, I will do XXX things: first, I'll cover specific reporting requirements from the National Voter Registration Act (NVRA), the Uniformed and Overseas Citizen Absentee Voting Act (UOCAVA) and the Help America Vote Act (HAVA). I will describe the aggregation technique used to arrive at a GPA-like score for each state in terms of response rates. I am waiting for Paul Gronke to return from vacation before I can answer questions about the patterns of response rate for the survey questions related to poll workers, but expect that question to be answered soon.

Statutory Requirements

I have included the text of the three Acts mentioned above in Word documents with my comments, indicating reporting requirements of each Act where I could find them when I read through the Acts. As far as I can tell, there is not a itemized list of what data is required to be collected in accordance with these three Acts, and occasionally, pertinent which is conceivably collected is not included in the survey instrument (see my comments in the NVRA and UOCAVA text for two examples).

Briefly, the NVRA requires that citizens who use public assistance agencies, register residence or acquire a driver's license be allowed to register to vote at the same time. There is a requirement for a report to Congress, however.

Likewise, the UOCAVA text does not include a list of data to be collected, but does require a report to Congress

HAVA also has reporting requirements, such as a hotline to call to find out if a provisional ballot was counted and, if not, the reason for its exclusion. HAVA is also the only time I see a reference to residual votes, the over- and under-votes.

Justification for Survey Items

As you asked in our last phone conversation, I've been recording my thoughts on why each survey item is important for the survey's purposes. By survey, here are my reasons:

NVRA:

Active (or total) voters are often a denominator for other rates (such as turnout or the percentage of the population that are registered to vote) that make it an obvious count to record. Counts of inactive voters are less relevant, but I think that's because states are not sure what to do with that data.

Election Day Registration is also a high-profile count, given its experimental use in a small number of states. I'm sure other states are watching that count closely.

Questions asking for registration manner and location give an idea of where and how people are registering to vote. While these questions are broad enough to cover most cases of registration, there is not a question for people who registered by means other than via a public assistance agency (which is admittedly beyond the purview of NVRA).

Similarly, questions about vote rejection and voter removal give an idea of how the system of casting and counting votes is working properly. Data on the number of voters removed from the rolls because of death, for example, can be compared to the total number of voters removed from the rolls to assess if voters are being removed from the list inappropriately.

UOCAVA:

Questions about casting and counting ballots among uniformed and overseas citizens are a small piece of the election administration pie, but also a critical group that should not be denied the right to vote due to living outside the United States. These are more diagnostic tools to evaluate the efficacy of counting ballots cast by Americans abroad.

EDS:

The Election Day Survey is similar in subject matter to the UOCAVA survey, but vastly larger in scope.

In this survey, the battery of questions related to reasons for not counting provisional and absentee ballots is more expansive, including interesting items like the number of ballots that were rejected because the voter already voted some other way (a count which could be used as to show incidents where potential voter fraud was prevented) or the count of ballots rejected because of untimely receipt (which could be used as evidence to transition to a national policy of "postmark" requirements, where if a ballot is postmarked by Election Day, it is counted instead of mandating that ballots be received on Election Day in order to be counted).

Questions about poll workers, polls and access for disabled voters provide accountability measures to ensure polling places are properly administered, though there is some question if states are replying to these items accurately.

Overvotes and undervotes are technical items, but also important for policy debates involving questions of voting technology.

Aggregation Technique and Scoring

As you will recall, the first memo on this data included the number of counties and local jurisdictions that responded to the survey instrument circulated by the Election Assistance Commission. In that data, a “0” meant that no counties responded to a given survey item, while a response rate of 100% indicated that every county responded to the survey item, even if that response was 0 for any given item. Given the lack of a jurisdiction response count for the undervote and overvote items, I dropped these items from my calculations.

This scoring report is based on that earlier dataset. I have divided the number of counties that responded to each survey item by the number of counties in the state to arrive at some percentage, which is the response rate. Cases where 0 counties replied to a given survey item are replaced with blank cells. This response rate is then graded along a standard grading scale, from A to F (a specific grading policy is included in the “Total” tab in the data). The aggregate scores are included in the “Total” worksheet, while each survey is separated into its own worksheet. I have separated out the four territories and the District of Columbia from the main tabulation in the final score report, but calculated the results for these five areas too.

Let me walk you through the case of Alabama. You will see that Alabama has a “1” in the 2004 active voters column (NVRA tab), meaning that every county in Alabama responded to this question from the NVRA survey. 65 counties in Alabama responded to the 2004 inactive voter prompt, so the quantity in that cell is about 0.97, or indicating a 97% response rate. You will notice that the “Statewide response rate using total number of survey items (31/32)” column scores Alabama at about 0.29 and if you scroll to the right, you will see that Alabama’s good response rate on these first few survey items drops off considerably for the remainder of the NVRA survey. The same trend is replicated for the UOCAVA and Election Day Surveys. Alabama scored 0.24 and 0.14 on those surveys, respectively. Lastly, you will see in the “Total” tab that of the possible score of 133 (or a score of 1 for each of the 133 survey items Alabama could have responded to), Alabama scored just over 28, meaning that Alabama responded to about 21% of the survey items, and 21% is a failing grade. You’ll notice that the average score among the states is a D-. About half the states are reporting less than 60% of the survey items.

I considered two alternate aggregation techniques that I ultimately abandoned, or adjusted. First, I had mentioned to you that I would calculate a score based on the response rates within survey items that states responded to at any level, but decided that measure would be meaningless. Consider Alabama, which provided good data on a small number of survey items and Delaware, which provided good data across most of the survey items. I do not think these cases are comparable, since this measure would in effect artificially inflate the score of states like Alabama. So I dropped that technique entirely.

Secondly, I adjusted my technique to account for the states that do not provide for

Election Day registration. You will see that the total number of survey items is either 133 or 134 and that there are either 31 or 32 survey items in the NVRA section. The extra survey item is counts for EDR in the seven states that provide it. I altered the aggregation technique so that the 43 states that do not provide for EDR are not penalized in the final calculation.

I have one word of caution for the case of Hawaii. There are five legally defined counties in Hawaii, but one of them – Kalawao – is designated as a leper colony, and in a state of quarantine. As far as I can tell, there is a tiny county government (a sheriff and some officials from the Health Department), and this county accounts for most of the occasions in the data where 4 of the 5 Hawaiian counties responded to a given survey item. According to the 2000 census, there are 147 people living in this county, so it may be accurate to say that Hawaii's score is unfairly deflated because of one small case with unique circumstances did not reply to the survey instrument when it was circulated. If Hawaii's score is adjusted to remove this county, the adjusted score is .674 (an increase of .145), or a D+.

Rank	State	Score
1	Delaware	98.25%
2	Montana	97.37%
3	Georgia	96.41%
4	Florida	96.11%
5	Ohio	95.47%
6	Texas	91.80%
7	Michigan	90.23%
8	Arizona	85.46%
9	Alaska	84.21%
10	North Dakota	82.30%
11	Wyoming	81.34%
12	Washington	81.24%
13	Maryland	75.13%
14	Missouri	74.19%
15	Arkansas	71.04%
16	Idaho	70.42%
17	Iowa	70.21%
18	New Jersey	69.89%
19	Utah	69.77%
20	North Carolina	69.54%
21	Kentucky	69.22%
22	Colorado	67.70%
23	Nevada	66.83%
24	Louisiana	65.07%
25	Oregon	64.45%
26	South Dakota	62.58%
27	Mississippi	54.54%
28	New York	53.28%
29	Hawaii	52.93%
30	Kansas	52.62%
31	Oklahoma	51.87%
32	California	50.74%
33	Maine	49.63%
34	New Mexico	49.33%
35	Nebraska	47.64%
36	Rhode Island	46.77%
37	West Virginia	43.65%
38	Indiana	43.31%
39	Pennsylvania	42.00%
40	Minnesota	41.65%
41	Virginia	40.43%
42	Illinois	39.67%
43	Tennessee	38.16%
44	South Carolina	33.02%
45	Connecticut	28.01%
46	Vermont	27.12%
47	Massachusetts	24.81%
48	Wisconsin	23.13%

49 New Hampshire	21.68%
50 Alabama	21.34%