

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 15<sup>th</sup>. 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.

	<b>NVRA</b>	<b>UOCAVA</b>	<b>EDS</b>
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.