Measuring the Improvement (or Lack of Improvement) in Voting since 2000 in the U.S.

January 14, 2006

Charles Stewart III
The Massachusetts Institute of Technology
cstewart@mit.edu

Abstract

This paper summarizes what systematic evidence exists about the performance of the American voting process in 2004 and proposes a comprehensive system of performance measures that would allow citizens and officials to assess the quality of the voting system in the U.S.

Despite the great deal of attention paid to voting reform from 2000 to 2004, and billions of dollars spent, there is surprisingly little systematic evidence of improvement in how elections are conducted in the United States. The best evidence of improvement comes in assessing the overall quality of voting machines that were used, and here the news is good. Nonetheless the measures used to assess voting machines could be greatly improved. There is little systematic, nationwide evidence of whether registration problems declined, polling places were administered better, or whether voter tabulations were more accurate.

In thinking about how to improve data gathering about the election system, we first need to specify four principles guiding data gathering (uniformity, transparency, expedition, and multiple sources) and three major obstacles (federalism, state and local officials, and disputes over the purpose of elections). With these principles and obstacles in mind, I sketch out a basic data gathering agenda intended to allow the public to assess the quality of voting in the United States.

Measuring the Improvement (or Lack of Improvement) in Voting since 2000 in the U.S.

Charles Stewart III
The Massachusetts Institute of Technology
cstewart@mit.edu

Following the 2000 presidential election, states throughout the country reformed their voting procedures, primarily in response to the debacle in Florida. These reforms were spurred by two related developments. The first was a series of reform commissions that convened through the authority of state officials — governors, legislatures, and secretaries of state (Coleman and Fischer 2001). These commissions recommended a host of reforms tailored to the needs (or especially loud and organized interests) in the particular states, ranging from the institution of Election Day registration to the decertification of punch card voting devices. The second development was the passage of the Help America Vote Act (HAVA) in October 2002, which mandated a range of reforms for federal elections and made available nearly $4 billion in federal funds to help retire punch card and mechanical lever voting machines, and generally to help improve the administration of elections (Coleman and Fischer 2004).[1]

This concerted effort at reforming the mechanical aspects of voting, the likes of which the nation had never before seen,[2] cries out for evidence of its effectiveness, or lack thereof.

---

[1]HAVA is P.L. 107-252. A comprehensive summary, along with links to the actual legislation, is available on the web site of the National Conference of State Legislatures at the following URL: http://www.ncsl.org/programs/legman/elect/nass-ncslsummaryw-orecs.htm. A briefer summary can be found at election.org (2003).

[2]There have been reform waves in the past, but they have not been as comprehensive, either with respect to geography or the process of voting. The closest in geographic scope to HAVA was the Voting Rights Act (VRA), even though its provisions were focused on the South and on gaining access to previously disenfranchised voters to the polls. The VRA did not address the question of voting machines, for instance, and it was agnostic to most voting procedures, so long as they did not hinder minority access to the polls. The National Voter Registration Act (NVRA), or "Motor Voter," focused only on voter registration and no other aspects of the voting chain than runs from registration to the certification of elections. The

There is mostly bad news here, with a smattering of good. The bad news is that the current wave of reform has not succeeded in establishing a comprehensive set of performance measures to help the public and policymakers judge whether election reform has met its goals of improving the access of voters to the polls, improving the experience of voters once at the polls, and improving the administration of elections. Because of the polarization of election reform that arose after 2000, efforts to assess voting systems performance have regressed on some fronts, both through cutting off information that was previously available and by flooding the system with claims that are based on unscientific methods.

The slight ray of good news is that on one widely-reported measure of system performance, the residual vote rate, the 2004 presidential election appears to have been administered better than 2000 (Stewart 2006). This is an imperfect, partial measure of system performance, and thus the good news is imperfect and partial. Until election administrators and reformers become more serious about documenting the performance of the election system, our understanding of reform efforts will be murky at best.

The purpose of this paper is two-fold. The first is to summarize what systematic evidence exists about the performance of the American voting process in 2004. The second is to propose a comprehensive system of performance measures that would allow citizens and officials to assess the quality of how the franchise is exercised in the United States.

The following two sections of the paper parallel these purposes. The next section systematically examines the voting system in 2004, looking for evidence that 2004 was

---

introduction of mechanical lever machines required the change of election laws in individual states, which was a painstaking process that consumed roughly three-quarters of a century; many states never approved the introduction of mechanical lever machines.

administered better than 2000. The section after that takes a broader view, by postulating a set of criteria for establishing a systematic monitoring system for the United States and then proposing an agenda for the future. A conclusion summarizes the entire argument.

I. A Quantitative Assessment of Voting in 2004

Were elections run better in 2004 than in 2000? By one measure, newspaper accounts about presidential elections, 2004 looked significantly better than 2000. In a *Lexis/Nexis* search of five major newspapers[3] across the United States on the terms "election", "problem*", and "president*", we retrieve 963 hits between November 1 and December 31, 2000 and only 470 hits for a comparable period in 2004. (To calibrate things, a similar search for 1996 generated 442 hits). If we add the word "Florida" to the search, we get 34 hits for 1996, 579 for 2000, and only 58 for 2004.[4] So, while in the minds of some the election of 2004 was just as fraudulent as 2000,[5] by the newspaper evidence, the level of concern with election problems returned to a pre-2000 baseline.

And yet by these same reporting measures, things did not look so rosy. Although Florida, by the newspaper accounts, improved significantly between 2000 and 2004, Ohio backslided. The Buckeye State, which generated 31 hits for electoral problems in 1996 and 39 in 39 in 2000, generated 59 in 2004 — more than Florida.

---

[3]These papers, chosen to be geographically dispersed and not located in the states that were the focus of so much national press attention in either 2000 or 2004, were the *Atlanta Journal, Chicago Sun-Times, Denver Post, Los Angeles Times,* and the *New York Times.*

[4]A similar search for Iowa, which has generated little national attention, generated 17 hits in 1996, 29 in 2000, and 26 in 2004).

[5]One collection of such sentiments can be found at the following URL: http://www.crisispapers.org/topics/election-fraud.htm.

If we change our search strategy to focus on particular kinks in the voting chain, a different pattern emerges, as well. If we search these same papers for stories about voting machine problems, we get a total of 19 stories in 1996, 128 in 2000, and precisely 200 in 2004.[6] The number of stories about voter registration problems rose from 96 in 1996, to 112 in 2000, and 221 in 2004.[7] The number of stories about long lines at the polls in the election season increased from 7 in 1996 to 41 in 2000 to 50 in 2004.[8] The only bright point here is that the number of stories about vote fraud in the presidential election fell back to 14 in 2004, after rising to 28 in 2000.[9] (The number was 5 in 1996.)

Results in the press such as these are but one piece of evidence about why it is important to establish a series of systematic and objective benchmarks against which to assess improvements and deteriorations in the voting process. The normative issues here are significant. At a middle level of normative concern, with billions of dollars at stake, it is important to know whether dollars have been allocated effectively in the past and how they should be allocated in the future. At a higher level of concern, the legitimacy of the electoral process is at stake. It makes a difference whether American elections are regarded on a par with Canada or Zimbabwe. With partisans of all stripes eager to use disconnected piece of evidence to uphold or challenge the legitimacy of any election outcomes, attentive citizens must have

[6]The time frame has now shifted to the entire calendar years of 1996, 2000, and 2004. The search terms here are "voting machine" and "problem*".

[7]The search terms here are "voter registration" and "problem*".

[8]The search terms here are "election" and "long lines" and "president". The time frame is November 1 to December 31 of each year.

[9]The search terms here are "vote fraud" and "presidential election".

access to facts about the electoral process that can withstand partisan and ideological election interpretations.

The answer for why assessments of the election of 2004, compared to 2000, were so mixed becomes clearer when we explore the voting process as it unfolded in 2004 and ask what independent evidence we have about the performance of the system at every step along the way. To aid in such an exploration, it is useful to be explicit about the chain of procedures that must be successful in order for a voter's votes to be accurately cast and counted.[10] These procedures start with the registration of voters and end with the tabulation of results. The important links in the chain of procedures are shown in Table 1. Also included in Table 1 are a summary of how the Help America Vote Act (HAVA) affected that step of voting and a list of methods that are being used, or could be used, to assess the quality of each of these steps, from the perspective of the voter.

[Table 1 about here]

Step 1: Establishing the Voting Rolls

The voting chain begins with establishing the voting rolls, normally through the registration of voters. This chain fails when the registration process itself is incomplete, such as when a registration post card mailed in by an eligible voter never makes it to the election office, or when the election office makes a clerical error, such as mistaking parents and children with identical names who live at the same address. The main problem here for the voter is that without a

---

[10]This is similar to the voting system as specified in the original Caltech/MIT Voting Technology Project report *Voting: What Is/What Could Be* (2001b, pp. 12–16), with the exception that I do not deal explicitly with ballot security, which appears to be intrinsically linked with each of the steps that I do explicitly examine.

procedure such as provisional ballots, the voter may have no recourse on Election Day and be

denied the opportunity to vote. Even with provisional ballots, the problem may be irremediable,

such as when the registration post card gets lost in the mail.

HAVA's main provisions pertaining to voter registration supplemented the 1993 NVRA,

by requiring that all states have an integrated, computerized voter registration system. The stated

purpose of this provision was that centralized, computerized state systems would help to deal

better with the high level of transience among American voters and subject each state's voters to

the same degree of administrative capacity in dealing with voter registration statewide. It also

was intended to nudge states to automate more effectively the blizzard of voter registration cards

that grew in response to the more liberal registration provisions of the NVRA.[11] Unstated

reasons behind this provision are equally important for considering how to monitor the

functioning of the registration system. The main one was a compliance problem that many states

had with local election officials in managing the voter rolls properly, particularly in fast-growing

exurban areas where newcomers are both an administrative burden and political threat.

How would a civic-minded voter, social scientist, or Justice Department investigator

know when registration problems have risen or fallen in a state? The most rigorous method, and

least likely to be implemented, would be regular, systematic audits of the "paper trail" involved

---

[11]With the passage of the NVRA, county courthouses are no longer the place where citizens typically register to vote. According to the most current figures compiled by the U.S. Election Assistance Commission, only 25% of new voter registrations between 2003 and 2004 were "in-person" registrations. This contrasts with 32% that came through the mail, 33% that came through motor vehicle offices, and roughly 10% that came from other state agencies. In raw numbers, this amounts to over 16 million mail-in registrations during the two-year period and over 20 million registrations coming from agencies whose major function is not administering elections (U.S. EAC 2005b).

in the voting registration system. For instance, investigators could follow a series of "dummy" registration cards to see what fraction of them eventually led to a properly-entered registration for fictitious individuals. For those uneasy with the prospect of using fictitious individuals to test the integrity of the registration system, it would be possible to deploy investigators to randomly-chosen places where registration cards were typically filled out (like Department of Motor Vehicles offices), to have them tag a certain fraction of those cards, and then to follow them through the process. As far as I know, no state implements such a program on a regular basis.

The Election Assistance Administration (EAC) has taken an initial step toward documenting the administrative implementation of the registration requirements under HAVA, and registration procedures more generally, through its Voter Registration Survey, which forms the core of the data for its biennial NVRA report (U.S. EAC 2005b). However, the questions in the survey tap administrative procedures, not performance measures like accuracy. Consequently, the EAC Voter Registration Survey, at best, can provide measures of independent variables that might help explain variations in the performance of registration systems (as experienced by voters), but not document the performance of the systems themselves.

Intensive systematic auditing of the registration system would be the best way to identify problems with registration and to document improvements that might be associated with changes in the law. However, the expense of such a procedure, accompanied with the lag time between most investigations of this sort and reporting the results, suggest the value of relying on other data that are generated for other purposes. One source of such data are the election returns. Another source is national surveys.

A natural starting place for measuring the effectiveness of a jurisdiction's registration procedures is the number of provisional ballots requested on Election Day. Assuming a properly implemented provisional balloting law, having more provisional ballots cast in a jurisdiction may be a sign of more registration problems.

Or it may not. The assumption of a "properly implemented provisional balloting law" may be heroic. Even states that have reputations for taking their provisional balloting laws seriously have compliance problems. For instance, in reviewing the election returns from North Carolina in 2000, I noticed that three counties (Ashe, Hyde, and Polk) reported precisely zero provisional ballots. When I called one county's election office to see why this was, the official stated that "we don't like 'em, so we don't use 'em." A North Carolina state official later confirmed that this attitude of non-compliance (what political scientists would call a classical "principal-agent problem") was significant in the implementation of their "failsafe" voting law.[12]

Compliance issues were no doubt significant in states in 2004 that were newly implementing provisional ballot laws required by HAVA. In Georgia, a state demographically similar to and geographically proximate to North Carolina, 30 of 159 counties reported precisely zero provisional ballots cast in 2004, compared to none of North Carolina's 100 counties.[13]

---

[12]See Alvarez and Hall (2003) for a general discussion of the principal-agent problem in election administration.

[13]These numbers are based on results reported in the EAC's 2004 Election Day Survey, data tables available at http://www.eac.gov/election_survey_2004/toc.htm. Previous drafts of this paper reported numbers contained in the official state reports. While the two sets of numbers are similar, they are different. Because the EAC figures are more accessible, I rely on them here. The fact that states report one set of numbers to their citizens and another to the EAC is only one of many piece of evidence that the data associated with election administration is not tightly gathered and reported. I discuss the Election Day Survey below.

Measured another way, 12,895 Georgians cast provisional ballots in 2004, compared to 77,469

North Carolinians, even though both states had roughly the same total turnout (3.3 million in

Georgia and 3.6 million in North Carolina). While it is possible, it is unlikely that the

administration of voter registration rolls in Georgia is better than North Carolina by a factor of

six.

Furthermore, the implementation of provisional balloting laws itself may be a political

variable. The use of provisional ballots may fluctuate because election officials may be

instructed (or otherwise feel compelled) to make it easier or harder to use provisional ballots.

For the moment, the use of provisional ballots is so poorly understood, that it is not clear

whether their use helps or hurts certain types of candidates. For instance, in the 2004 election,

some civil rights advocates attacked the use of provisional ballots, arguing that their use

substituted for "real" ballots. Other civil rights advocates encouraged the use of provisional

ballots, arguing that their use substituted for voters being turned away at the polls. In the future,

as provisional ballots are better understood, civil rights advocates will come to a less varied

interpretation of their use, and thus we should expect the number of provisional ballots to vary

simply with the number of registration problems on Election Day.[14]

As well, provisional ballots may also be an indication of problems in other parts of the

voting process chain, such as polling place administration. For instance, a harried precinct

---

[14]In North Carolina, which has the most comprehensive and transparent record keeping of a large state about how voters vote, 48% of ballots cast provisionally in 2004 were cast by Democratic registrants, compared to 47% of ballots cast in-person that were cast by Democratic registrants. Republicans accounted for 37% of all in-person voters but only 32% of provisional voters. Unaffiliated voters accounted for 20% of provisional ballots but only 17% of in-personal ballots. This suggests that while provisional ballot users are less Republican than average voters, the counterbalance is offered by the unaffiliated, not by Democrats.

warden may be more likely to offer a provisional ballot to a voter than to try and resolve the registration issue with a call to the county office if the line to vote is out than door than if business has been slow that day. And, it is likely that counties with greater-than-average registration problems will have greater-than-average problems managing their precincts. To the degree that the number of registration problems is correlated with the number of polling place problems at the geographic level being analyzed, the number of provisional ballots used will be an imprecise, and possibly biased, measure of registration problems.

Finally, provisional ballots may be an indication that election officials are doing their jobs in the face of the challenges that give rise to registration ambiguities in the first place. For instance, state registration deadlines often overlap with deadlines for preparing pre-election and Election Day materials at the local level. It is common for counties, in the midst of performing the exacting procedures to get ready for Election Day, to be inundated with new voter registration cards, spurred on by a last-minute flurry of interest in the upcoming contest. Faced with the choice of not entering into the computer the names of the last remaining registrants or not being ready to open scores of precincts on time, county officials understandably focus on opening the polling stations, at the expense of entering last-minute registrations. Consequently, it is possible for provisional ballots to surge in a county because of last-minute interest in the

race by voters (or activist groups who often generate large surges of new registrants),[15] not

because registration procedures have suddenly broken down.

Figure 1 reports the number of provisional ballots counted in North Carolina counties in

2000 and 2004, as a percentage of all ballots cast. (The circles are in proportion to the turnout in

the counties.) The fraction of provisional ballots counted in North Carolina went up between

2000 and 2004, from 1.0% to 1.3%. Does this overall increase in provisional ballots reflect more

problems with registration in North Carolina or the greater prominence of the voting provision?

Without comparisons with other states, that is difficult to say. Across counties in North

Carolina, did the ones that used more provisional ballots in 2004 have more registration

problems than before? Without further probing of the why provisional ballots are actually used,

it is difficult to say.[16] The fraction of ballots cast provisionally across the two elections is

correlated at a moderately high level (r= .40 if we do not weight by population and r= .65 if we

do). Thus, it is likely that measuring the use of provisional ballots will tell us something about

---

[15]In formal and informal conversations with local election officials over the past four years, the issue of "bundled" registration forms has been one that has come up frequently. Most officials seem to have stories of groups that held a registration drive, bundled the post cards together, and then forgot about them as they languished in car trunks. Often these cards are mailed in as the election is imminent. It is impossible to judge whether these stories are genuine or part of urban legend — probably a bit of both. But the existence of the stories illustrates that in the minds of local election officials, registration problems usually arise due to the behavior of people over whom they have no control, but for whose behavior they are nonetheless held responsible when things go wrong.

[16]I conducted a preliminary statistical analysis, in an attempt to explain both the cross-sectional use of provisional ballots in North Carolina and the change in their use from 2000 to 2004. Neither variable was strongly correlated with factors like a county's race, turnout (level or change), change in number of registrations, or change in turnout.

the administration of elections in particular counties, but it is unclear at the moment precisely what that would be.

[Figure 1 about here]

Another source of information that could be used to judge the effectiveness of the election system is national surveys. The most direct evidence for how smoothly registration proceeded would be to contact a randomly-chosen group of voters and ask them if they had experienced a range of common registration problems on Election Day.[17] Even though registration lists are public records, and most states make these lists available in easily-used electronic form, it seems that no such investigation has ever been performed.

The closest thing to such a national survey is the Voting and Registration Supplemental File of the Census Bureau's Current Population Survey (CPS). The CPS, which typically involves over 50,000 households, distributed across each state, is best known as the instrument that helps to estimate the monthly unemployment rate. The Voting and Registration Supplement (VRS) is added to the survey in even-numbered Novembers. The VRS asks respondents whether they voted in the November election. If the answer is "no," it asks why not.

Beginning in 2000, one of the choices offered respondents for not voting was "registration problems." In 2000, 6.8% of non-voters listed registration problems as their reason for not voting, compared to 4.1% in 2002 and 6.9% in 2004. Expanding the denominator to all

[17]Of course, such surveys would omit people who had been turned away from voting, possibly because of registration problems, so there would be limits to what one could learn from this technique. However, if linked to a companion survey of all eligible voters, we could learn a lot about the quality of the voter registration process.

registered voters, we find 0.9% of all registered voters reporting they did not vote in 2000 due to registration problems, 1.1% in 2002, and 0.7% in 2004.

Figure 2 shows the scatterplots comparing the prevalence of registration problems in keeping voters from the polls in 2000, 2002, and 2004. On a statewide level, there was a moderate degree of year-to-year correlation in these figures,[18] which suggests there are likely slow-changing factors within each state that throw up registration barriers to a state's voters. If we trust that this correlation is due to real underlying problems with a state's registration process, then a factor analysis of this data could at least identify states with overall "good" and "bad" registration. Applying such a procedure to this data reveals the District of Columbia, Oregon, Washington, South Carolina, and Oklahoma as the five states with the greatest registration problems and Wisconsin, Maine, Minnesota, New Hampshire, and North Dakota as the five states with the fewest problems across the last three federal elections. Of these latter states, four had election day registration (EDR) and North Dakota had no voter registration at all. This pattern lends a certain degree of validity to this measure as tapping into levels of registration problems in the cross-section of states, although without further research it is unclear

---

[18]Here are the intercorrelation matrices associated with these graphs, weighting each state by the (geometric average) number of observations in each year's VRS:

Election year	Election year		
	2000	2002	2004
2000	—		
2002	.47	—	
2004	.43	.65	—

whether we should trust *changes* in this measure from election-to-election as anything more than

random noise.[19]

[Figure 2 about here]

Where does this leave us with respect to measuring the quality of voter registration in

general, and the change in that quality over the past quadrennium? At the moment, we have very

little to go on if we want to answer either question. Provisional ballot data are so fugitive at this

point, and our understanding of their use is so primitive, that even thinking about using these

data as a measurement strategy is still in its early stages. The CPS-VRS seems to have promise

for developing a reliable measure of cross-sectional performance, even though the question

wording of the instrument is blunt, at best.[20] In any event, none of these measures has been

---

[19]It is also interesting that of the states that reported the highest levels of registration
problems, two (Oregon and Washington) were among the states with the highest level of mail-in
voting. In general, the correlation between the fraction of registered voters reporting registration
problems and the fraction of voters who used mail-in procedures is a moderate .30. In a multiple
regression setting, both the presence of election day registration/no registration and the
percentage of ballots cast by mail are significant predictors of how many non-voters blamed
registration problems in 2004 (standard errors in parentheses):

Election day registration or no registration (dummy var.)	-0.054 (0.014)
Pct. of ballots cast by mail	0.064 (0.028)
Constant	0.067 (0.005)
N	51
R^2	.32

[20]See the next section for a discussion of the shortcomings with the VRS supplement
question wording.

developed sufficiently to give us confidence in using them to assess whether we have made

progress in improving voter registration since 2000.

Step 2: Checking-in voters at polling places

The voting chain continues when voters arrive at the polling place and are checked in. This link

in the chain fails when a qualified voter appears at a polling place and is unable to complete the

check-in. A major reason for failure at this step is related to failures in the previous step: if a

voter's registration has been erroneously processed, she or he will show up at the correct

precinct and not have her or his name on the voting list. A problem that is probably equally

prevalent is showing up at the wrong precinct. Most communities that have more than one

voting location do not have a comprehensive directory of all voters at each polling place, which

would direct errant voters to the correct location. When a voter arrives at the wrong voting

place, many things can be done, which are more or less effective. The standard procedure in

most places is for a precinct worker to call the central voting office to inquire where the voter

belongs. Because of the peak load problems associated with handling so many phone calls on

Election Day, voters often do not get redirected to the correct precinct.[21] Large numbers of

registration problems at check-in cause lines to form at the polls. If the lines get long enough,

voters walk away without voting.

---

[21]The Los Angeles County Registrar-Recorder/County Clerk's Office handled over 64,000 calls on Election Day 2004, which is roughly 2% of turnout in the county. Common Cause's account of activity on their 866-MYVOTE1 telephone line reported that over 55% of the voters that contacted them on Election Day who had tried to reach their own local election departments had been unable to do so (Common Cause 2004, p. 2).

As before, there are straightforward ways to study how prevalent polling place problems are, and therefore of measuring improvement in polling place practices. The discussion in the previous subsection about using the number of provisional ballots as an indicator for registration problems could easily be adapted for this subsection, too. It is possible that a spike in provisional ballot usage in a jurisdiction could be an indicator of added troubles with polling stations.

The most direct measurement would be systematic observation of polling places by trained researchers, who would note things like the number of people who approached the check-in desk, the number of people who were successfully checked in, the problems that emerged, and how problems were resolved. While there have been pilot projects done to test the feasibility of doing such large-scaled research, a nationwide program has yet to be attempted.

On the surface, it appears that numerous activist groups and law schools conducted projects in 2004 that utilized methodologies similar to this approach. Probably the best-known was the Election Incident Reporting System (EIRS), which was associated with groups such as the Verified Voting Foundation, Computer Professionals for Social Responsibility, the National Committee for Voting Integrity, the Electronic Frontier Foundation, and the Online Policy Group (see https://voteproject.org). The centerpiece of this effort was an web-based data aggregating tool that allowed election observers who were affiliated with numerous of "election protection" organizations to report voting problems they encountered, for later action and analysis. Another effort was the collaboration between Common Cause and Votewatch (now the Election Science Institute). This effort involved surveys of voters leaving the polls in New Mexico and Ohio and a nationwide survey of voters about their voting experience. There were also numerous efforts

centered in law schools to monitor the conduct of election, a good example of which was the one located at the University of North Carolina School of Law (UNC Law School 2005).

The EIRS and the Common Cause/Votewatch projects, which relied on self-motivated voters to communicate their experiences, were conceded by these organizations to produce results that were suggestive at best, since the samples were of unknown quality.[22] Therefore, the data from the EIRS that are easily accessible through the voteproject.org web site, are probably not useful for assessing the quality of polling place operations nationwide in 2004.

The national survey that was part of the Votewatch project has greater potential as a scientific instrument, but results from that study have yet to be released and its sample size (900 voters) is undoubtedly too small to detect differences at the state or local levels.[23]

The EAC commissioned an Election Day Survey, which was administered to state election officials, seeking information from local election administrators about a host of Election

---

[22]This type of sampling is referred to a "convenience sampling," and includes a variety of techniques in which the statistical properties of the sample are unknown. ("Man on the street" interviewing is the best known of convenience sampling techniques. All of these projects that encourage voters or election observers to record election incidents are the electronic equivalent of man on the street sampling.) Convenience samples are often valuable in the preliminary stages of research, but they are useless for making inferences back into the population they are meant to represent.

[23]A common mistake made by many people in trying to assess the performance of the election system is in over-estimating the number of "incidents," whether they be simple errors or foul play, and therefore under-estimating the size of a sample needed to detect problems and changes in the frequency of problems across time. What little systematic data we have about voting projects nationwide — from residual vote studies and from studying the CPS-VRS — suggests that the percentage of voters who have any particular type of problem at the polls is probably in the single digits. Therefore, it is possible that even a sample of 900 voters nationwide will yield only a handful of voters with problems. Hence, a national sample to detect serious polling place problems would have to have a sample size of many thousands. I expand upon this point in the next section.

Day practices and outcomes.[24] Among these were data about pollworkers and voting machine

malfunctions. The voting machine malfunction results were the most discouraging of the whole

survey, with information returned concerning voting machine malfunctions from less than 10%

of voting jurisdictions. Twenty-one states failed to respond to the question altogether and two

states reported the implausible claim that precisely none of their voting machines experienced

problems on Election Day (U.S. EAC 2005b, p. 11-2).

The Election Day Survey also asked about pollworker staffing, including the number of

workers required under state law and the number of workers who were actually employed.

Overall, roughly 5% of polling places had inadequate staffing on Election Day, with reports of

inadequate staffing more often coming from localities that were poorer, more minority, more

urban or rural than the rest of the nation. Interestingly, jurisdictions in "battleground states" and

those with mechanisms to facilitate voting before Election Day had fewer reported cases of

having an inadequate number of polling place workers (U.S. EAC 2005b, pp. 12-4 – 12-6).

The Voting and Registration Supplement of the CPS is a possible source of hope for

gathering systematic data about polling place practices. Since 1996, one of the reasons non-

voters have been allowed to give for not voting is "inconvenient polling place or hours or lines

too long."   In 1996, 1.3% of non-voters chose this as their principal reason for not voting,

compared to 2.7% in 2000, 1.5% in 2002, and 2.9% in 2004. When we use all registered voters

as the denominator, these percentages were 0.3% in 1996, 0.4% in 2000, 0.4% in 2002, and 0.3%

---

[24]The data requested was encyclopedic, including voter registration, ballots counted, sources of turnout, absentee ballots, provisional ballots, drop-off, over- and under-votes, voting equipment usage, poll workers, polling places, and disability access. The report is valuable not only for the data that are reported, but also for the catalogue of obstacles faced by the researchers in obtaining the data that were requested.

in 2004. Thus, it appears that the fluctuations in the fraction of non-voters who use this explanation is mostly due to the changing composition of non-voters, rather than changing barriers that face registered voters who might possibly go to the polls on Election Day.

Figure 3 illustrates the inter-correlations among the states over time on this measure.[25] If we focus on the presidential election years, the inter-correlations are similar to the "registration problem" item we previously considered, which again suggests there is something persistent in most states that cause some to regularly have more troubles at polling places than others. Like before, we can subject these data to a factor analysis to combine the four year's worth of data into a single scale that measures the level of problems with polling places. When we do that, we find that North Carolina, Arizona, Georgia, Indiana, and South Carolina were the states with the worst polling place experiences and Oregon, Alaska, Iowa, New Mexico, and Virginia were the best.

[Figure 3 about here]

---

[25]Figure 3 excludes Nevada, with 25% of non-voters citing this as the reason. Here is the intercorrelation matrix illustrated by Figure 3:

	Election year			
Election year	1996	2000	2002	2004
1996	—			
2000	.34	—		
2002	.06	.05	—	
2004	.38	.42	.10	—

Where does this leave us in assessing whether polling place practices actually improved in the United States between 2000 and 2004? On the one hand, there were certainly more news accounts of polling place problems — long lines, insufficient machines, etc. — in 2004 than in 2000. It is likely that this increase in reports was endogenous to the electoral controversy itself. The most obvious example of this endogeneity was the Votewatch project, which was used by NBC news to generate stories about Election Day voting problems.[26] Votewatch (and similar efforts) did not exist in 2000, and therefore it must be the case that the rise in reported incidents in the press and on blogs was due to this greater scrutiny, especially in states where the heat of the election was higher than average. The CPS-VRS survey suggests that polling place problems may have been steady in 2004. On the whole, then, even the best evidence we can adduce gives us little basis on which to judge whether the administration of polling places improved between 2000 and 2004. The best that can be said is that 2006 and 2008 may see better monitoring of polling places, based on preliminary studies that were conducted in 2004.

Step 3: Using Voting Machines

The next step in the process is actually using voting machines. Failures at this point were the focus of much of the Florida controversy in 2000, both the "butterfly" and "caterpillar" ballots (which represented a failure of human factors engineering) and "hanging chad" (which represented a failure of engineering, period). As was so well-documented in Florida, failures at this point can lead to one of two things, either the failure of a correctly cast vote to register

---

[26]The press release from the National Constitution Center, which participated in the project is located at the following URL: http://www.constitutioncenter.org/PressRoom/PressReleases/2004_10_26_12749.shtml.

outright or for a voter to be confused and have a correctly registered vote counted for an unintended candidate.

Catching failures in voting machines at this point is probably the most difficult task of election auditing, because of the secret ballot. The most direct way of testing for failures and documenting improvements across time would be to observe voters in the voting booth, and then ask them their intentions afterwards. This, of course, is unlikely to happen.

As a consequence, researchers have had to be indirect about measuring the performance of voting machines. The principal measure of voting machine performance that has emerged has been the "residual vote rate," which is the percentage of ballots that are over-voted or under-voted for a candidate (Caltech/MIT VTP 2001a,b; Ansolabehere and Stewart 2005).

Despite its widespread use, the residual vote rate has its limitations. There are first conceptual issues that arise in using residual vote as a measure of machine failure. As a one-time measure, it conflates over-/under-votes that arise because of intentional abstention as well as machine malfunction. It also does not measure votes that were counted that were nonetheless cast in error — the 2,000 votes cast by mistake for Pat Buchanan by Democratic voters in Palm Beach County in 2000 (Wand, et al. 2001) were counted as "successes" for these voters. Finally, the residual vote rate is based on any discrepancy that arises between the number of total ballots cast and the number of ballots counted, which are calculated at geographical and temporal removes that vary across jurisdictions. In other words, the component parts of the residual vote rate calculation are not generated the same way across all states. For instance, in some jurisdictions, the number of total voters is calculated by the number of times an electronic voting machine is activated, whereas in other jurisdictions the turnout figure is calculated from the

number of names crossed off the voter registration list; in some places the turnout number is

reported at the same time the election results are reported, whereas in others turnout is reported

(and calculated) months after the election returns.

The second set of issues with using residual vote rate have to do with state laws that vary

how, or even if, turnout is calculated and how, or even if, write-in ballots are tabulated.[27] In

2004, fourteen states did not report the total number of voters who appeared on Election Day.

Thus, it is not possible to calculate the residual vote rate at all in those states. Perhaps even

worse, some states report figures that appear to be turnout when in fact they are not.[28] Finally,

---

[27]Added to this is variability in reporting the incidents of over- and under-voting separately. Because there is much less ambiguity about whether an over-voted ballot is an error than an under-voted ballot, measuring the over-vote rate would perhaps be a better indicator of voting machine problems. However, Florida is the only state that mandates such reporting. See Florida Division of Elections (2003, 2005).

[28]A good example of this is South Carolina, whose turnout figures are reported at http://www.state.sc.us/scsec/election.html. The page claims to allow one to lookup "the number of voters actually taking part in the election." In fact, the "turnout" figures available on this site represent the number of registered voters in a county who were still resident in that county several months after the November general election. This results in a systematic under-count of turnout. Numerous counties end up with negative residual vote rates, as a consequence. Georgia, which now reports turnout based on the number of "ballots counted" by their electronic machines, also has a separate procedure that is similar to South Carolina's. After each general election, Georgia generates the "Credit for Voting Report," (CFV) which also systematically under-reports actual Election Day turnout. When an election reform activist discovered the discrepancy between turnout reported in the CFV Report and turnout reported using actual ballots cast, the Georgia Secretary of State's office pulled the CFV report from the web.
    More typical are states like Kansas and Pennsylvania. The Kansas Secretary of State's office informally polls its county officers on Election Day to get a turnout count, but this is not an official figure, and it rarely includes ballots counted after Election Day, like absentees and provisional ballots. As a consequence, 8 of Kansas's 106 counties had negative residual vote rates in 2004; in general, the Kansas residual vote rate would be biased downward quite a bit by using the Secretary of State's turnout figures. In Pennsylvania, the state does not collect turnout figures, but almost all counties do, using to their own methods. As a consequence, the turnout figures in Pennsylvania are based on inconsistent methodologies across counties. See Alvarez, Ansolabehere, and Stewart (2005).

some states do not regularly count write-in votes, or count them inconsistently, which artificially inflates the residual vote rate for those states.[29]

Because states vary so much in the procedures they use to count votes and calculate turnout, and because candidates will induce varying levels of intentional abstentions across different geographic units, the residual vote rate has its least utility as a cross-sectional indicator of voting machine performance. Its greatest utility comes in applying it across a period of time, either by simply taking first differences or by using a multivariate statistical technique such as fixed effects regression.

Nationwide, among the 38 states and the District of Columbia for which it was possible to calculate residual vote rates in the presidential contest in both 2000 and 2004, the aggregate residual vote rate fell from 1.89% in 2000 to 1.09% in 2004. Figure 4 shows the scatterplot that compares the residual vote rate among these states. The diagonal line traces out a 45-degree angle, so that states above it had higher residual vote rates in 2004 than in 2000, and states below had lower residual vote rates. With the exception of the four states in the lower right-hand part of the graph, there is a moderate correlation in residual vote rates between the two presidential election years. Three of the four states that had exceptional drops between 2000 and 2004 — Florida, Georgia, and Illinois — saw a significant amount of activity in upgrading voting machine in the intervening years, and it is likely that this activity helped to significantly lower the residual vote rates in these previously poor-performing states.[30]

---

[29]See the exchange between Miller (2005) and Alvarez, Ansolabehere, and Stewart (2005) concerning Wyoming's informal method of reporting write-in ballots in 2000.

[30]In Florida, 45% of the counties, representing 65% of the voters, used different voting machines in 2004 compared to 2000; in Georgia, all counties used new machines in 2004; in

[Figure 4 about here]

The most expensive policy intervention in election reform over the past quadrennium has been buying new machines, and therefore it is important to tie these residual vote changes to specific machines and, most importantly, to *changes* in machines. This is where the decentralized nature of voting administration in the United States causes further headaches to policy analysis.[31] There is simply no comprehensive, freely-available listing of the precise voting machines used by localities in the United States. The most comprehensive list is available through Election Data Services (EDS), for a fee. Although the fee is reasonable (a few hundred dollars), its proprietary nature hinders widespread analysis of the performance of specific machines. The EAC's Election Day Survey, which was conducted by EDS, reports broad machine types for most states, but not specific models or when they were adopted.[32] Verified Voting maintains the most comprehensive freely-available dataset, but it does not cover every county, and some of the data are imputed.[33] This latter comment is not meant to disparage

---

Illinois, 60% of the counties representing 46% of the voters used new machines. Nationwide, 15% of counties, representing 35% of voters, used new machines. These election return figures, and others used in this paper to report residual vote rates, were gathered directly from state election officials and are available at the following URL: http://web.mit.edu/cstewart/www/election2004.html. Data about the use of voting machines was purchased from Election Data Services.

[31]At the same time, the decentralization of voting administration is a boon for introducing variation into the methods of voting. If all jurisdictions used the same voting machines nationwide, there would be no variation in machines on which to leverage cross-machine performance analysis.

[32]As well, twenty states did not report the number of different machines, as requested by the EAC, and many of the numbers that were reported were implausible. See U.S. EAC 2005b, pp. 12-1 - 12-2.

[33]This dataset may be accessed at http://www.verifiedvoting.org/verifier/.

Verified Voting, since getting this information often requires contacting directly thousands of

local election officials, many of whom claim not to know what kind of voting machines they use,

or refuse to report the information.

The bottom line here is that the best data we have to track how the use of voting

machines is evolving in the United States is still imprecise and incomplete. This has many

unfortunate consequences that will be discussed in the next section. For now, what is important

is that we *can* track the evolution of voting machine use at a local level if we are satisfied with

crude categorizations, like "mechanical lever machine" and "optical scanning." We cannot gain

much purchase in answering questions like whether Diebold's Accuvote-OS performs better than

ES&S's Model 100 Precinct Ballot Counter, both optical scanners.

Returning to the analysis, the previous conjecture that improvements in the residual vote

rate between 2000 and 2004 can be associated with voting machine upgrades is solidified with

Figure 5, which show residual vote rates at the county level, comparing counties that changed

machine types with those that did not. (The top graph is counties that kept their old machines;

the bottom graph is counties that got new machines.) To facilitate legibility, the graphs are

shown on a log-log scale. In each graph, the solid line is the 45-degree line; the dashed line is

the least squares regression line.[34]   The difference in the regression slopes illustrates that

counties that kept their old voting machines generally had similar residual vote rates in 2004 as

in 2000; counties that changed machines generally improved, especially those counties with

exceptionally high residual vote rates in 2000.

[Figure 5 about here]

In research that has explored changes in residual vote rates between 2000 and 2004 in

greater detail, this general pattern has been illustrated, nationwide and in the states of Georgia

and Massachusetts (Caltech/MIT Voting Technology Project 2003; Stewart 2004a, 2005, 2006).

For the nation as a whole, Stewart (2005, 2006) shows that a significant portion of the drop in

the residual vote rate between 2000 and 2004 can be tied directly to changes in voting machines.

The biggest improvements in residual vote rates came in communities that abandoned punch

cards and mechanical lever machines in favor of the controversial direct register electronic

(DRE) machines, though communities that newly adopted optical scanners also saw

improvements.

---

[34]The following tables report the regression parameters used in Figure 5 (dependent variable = log of residual vote rate in 2004; standard errors in parentheses):

	Did not change voting method	Changed voting method
log(residual vote rate, 2000)	0.55 (0.01)	0.15 (0.04)
Intercept	-2.30 (0.06)	-4.34 (0.16)
N	3,457	638
R^2	.31	.03

The residual vote rate has the advantage of being easy to calculate with voting statistics that are commonly reported and (relatively) easy to access. It has the disadvantage (which is true of virtually all of the measures being discussed here) of being generated by many processes that occur before and after the voting machines have been used. The type of statistical analysis described above (Ansolabehere and Stewart 2005, Stewart 2004a, Stewart 2005, Stewart 2006) is leveraged off an assumption that when we see a high correlation between voting machine changes and residual vote rate drops, the residual vote rate drop is due to the machine changes, per se, and not to other changes that might accompany a change in voting machines, such as in voting administration or education.[35]

Georgia provides an interesting case that illustrates how residual vote rate improvements that are associated with adopting new machines may not be just about the machines. After having a residual vote rate in 2000 that was worse than Florida's, Georgia's Secretary of State Cathy Cox pushed through a measure in the Peach State that eventually required the statewide use of Diebold's Accuvote-TS in the 2002 election. Following that, the residual vote rate in the gubernatorial election went from 2.8% in 1998 to 1.1% in 2002; the presidential residual vote rate went from 3.5% in 2000 to 0.4% in 2004. Improvements were even greater in "down ballot" races; they were also greatest in counties with low average incomes and high minority

---

[35]Contrary to the intuition of many people, when counties adopt new voting machines, voters are not so confused by the new contraptions that residual vote rates go up immediately after the implementation of the new machines. In general, adopting new machines, regardless of the type, seems to reduce the residual vote rate of a county, which is undoubtedly due to the extraordinary efforts that county officials and vendors make in educating voters about how to use the equipment. The last thing an election director or vendor needs is for the rollout of new voting equipment to be met with objective evidence of voter confusion.

populations (Stewart 2004a). These results provide *prima facie* evidence that new voting machines improve elections.

Still, the implementation of the new machines was accompanied by an unprecedented amount of vendor support and centralized attention to training by the Secretary of State's Office. The state attempted to train at least two workers per precinct. Videotapes were circulated to reinforce the training and to educate voters more generally. Diebold itself allocated "more than 360 professionals, including 190 field technicians, 160 county support technicians, and a dozen regional support managers" throughout the state.[36] Only time will tell whether the improvements were due to the machines or to the 360 Diebold professionals working in the state.

In general, although the failures of some voting machines throughout the country on Election Day 2004 caught the eye of the attentive public, the big story was the overall improvement in machine performance from 2000 and 2004. The gains were not as great as once hoped, but they were real. If "lost votes" as understood in 2000 were all we cared about, the problem of assessing voting machine performance would appear to be on a firm footing, though still in need of improvement.

However, the controversy that has arisen since then, about the lack of paper back-ups to electronic machines, introduces a new set of performance issues that were not widely anticipated four years ago. In principle, we can use established techniques to see whether the addition of "voter verifiable paper trails" will complicate voting machines so much that they cause an

---

[36]URL: http://www.diebold.com/news/newsdisp.asp?id=2909.

increase in spoiled ballots,[37] but problems of voting machine "hacking" may be, in principle,

unmeasurable. That, at least, is a topic to be considered in the next section.

Step 4: Counting ballots

After the voting machines have been used on Election Day, the counting begins. This part of the

process fails whenever the intention of a voter to choose Candidate X does not lead to an

additional vote being counted for Candidate X. The biggest causes of this failure are due to

ambiguities on the ballot (incomplete check-marks, for instance), or failures in the voting

medium (ripped optically scanned ballots, for instance.) It is these sorts of failures that have led

election administrators, at various times, to gravitate toward mechanization. If they function

properly, mechanical lever machines and DREs leave no room for ambiguity, assuming the voter

makes no mistakes; if they function improperly, it is often undetected, so is not an issue. Even

the migration from hand-counted paper ballots to optical scanners was intended to relieve the

tedium of counting, so that fewer counting mistakes would be made.

The gap between casting ballots and counting them has always been the Twilight Zone of

election administration, both because the ballot is anonymous and because ballots are often

counted in a confined space with (in practice) limited opportunity for scrutiny from outsiders.

This fact has led some researchers in recent years to experiment with cryptographic schemes that

---

[37]A hint that the rush to provide "voter verifiability" to electronic equipment may have induced new performance headaches is provided in the sketchy and incomplete evidence revealed in the EAC's Election Day survey (U.S. EAC 2005b, chap. 11). In that survey, the 191 DRE-using jurisdictions that reported machine malfunctions on Election Day reported 439 instances of "printer failures," which significantly exceeded the number of failures associated with the defining features of DREs, which are screen failures (320 cases) and power failures (169 cases).

might allow for voting "receipts" that could allow individual voters to verify that their vote was properly accounted for, without anonymity of the ballot being compromised (Chaum 2004).

In the absence of a voter-verifiable, anonymous "receipt," we are left with highly indirect ways of auditing the counts of elections. The most direct method of assessing election counts would be to do them a second time, presumably using a different method from the first. One state, California, requires a "1% audit" of the ballots in its precincts, which is in the spirit of this method.[38] Controversy over purely electronic voting machines has led many states to consider similar auditing practices. Additionally, most traditional recounts are also a type of audit that uses a different method to count the votes.[39] The shortcoming with recounts as audit devices is that they are episodic and are rarely applied to large-turnout races like the presidential election. Finally, some have suggested using exit polls as a method of auditing the count.

The *California Election Code* makes no provision for the publication of the 1% mandatory recounts, and therefore it has been hitherto impossible to use that procedure to gauge whether vote tabulation in that state has gotten better or worse over the past four years. A paper by Alvarez, Katz, and Hill (2005) used data from 1% recounts in Los Angeles County, California, for the general elections of 2000 and 2002, plus the recall election of 2003. (Los Angeles County used pre-scored punch card ballots in these elections.) They examined

---

[38]*California Elections Code,* Section 15360, accessible through http://www.leginfo.ca.gov/calaw.html.

[39]Even when hand-counted paper ballots are recounted, the method is different, since non-recounted offices are generally ignored and "many eyes" painstakingly scrutinize each ballot.

differences in the total number of ballots counted under the two tabulation procedures (mechanical initial count followed by a manual recount), but not differences in vote margins.

Their most important finding is that the manual recount generally found more ballots than the initial mechanical count, but the number of ballots was small. In the 2000 recount, for instance, the manual recount of the presidential race found 38 more ballots, among a sample of over 13,000; the 2002 manual recount of the gubernatorial race found 21 more ballots among a sample of over 10,000 total. Ballots were rarely lost in the recount, but they were occasionally — such as in the recount of the 2003 gubernatorial recall question, where 3 votes "disappeared" in a sample of over 13,000 votes. In percentage terms, the difference between the number of ballots originally counted and the number counted in the 1% audit are typically around ½%, with a standard deviation across precincts of approximately 1/10 %.

The Alvarez, Katz, and Hill paper asks whether finding new ballots in these audits could have led to changed election outcomes, had the "overlooked votes" been counted in the first place. They conclude that the differences were too infinitesimal to have any substantive consequences in the races that they examined. However, if we are interested in improvements to election administration, the important question is whether the number of "overlooked votes" has changed over time. Mechanical machines degrade over time, and widely-published results from 1% audits could show which jurisdictions were taking care of their election machines and which were not. Los Angeles County has a reputation for high-quality maintenance of its voting machines, and the infinitesimal differences shown in the Alvarez, Katz, and Hill paper — which was about machines that had been used for decades — is testimony to the county's efforts. At the same time, computer scientist Doug Jones has shown, through a series of simple mechanical

tests, that the types of deficiencies found in the Florida punch cards in 2000 — such as "pregnant chad" — were possible only if machines had been poorly maintained over a long period of time.[40] Had Florida counties been required to conduct 1% audits and publish them widely, much of the recount fiasco of 2000 might have been avoided.

Papers by Ansolabehere and Reeves (2004) and Herron and Wand (2005) rely on the public data from New Hampshire recounts to assess whether hand counting is better than machine counting (Ansolabehere and Reeves), and whether different types of optical scanners performed differently in 2004 (Herron and Wand). Like the California study, both papers find that ballots recounted by hand tend to yield more counted ballots the second time, regardless of the voting technology used the first time. Ansolabehere and Reeves also show that when computers were used to count ballots the first time, the differences with the second count were smaller than if the first count was also by hand.

These three papers show the possibilities of using 1% random sample and traditional recount studies to examine the accuracy of election tabulations. They also illustrate the shortcomings. No state recounts all of its ballots; the 1% recounts are confined to California and the results that have issued from them are fugitive. The Ansolabehere and Reeves paper is rare, in that it covers a long period of time (1946 to 2002), but its idiosyncratic collection of state and local races make it impossible to use to gauge tabulation accuracy across time, at any level of analysis.

---

[40]See "Douglas W. Jones's Chad Page,"
http://www.cs.uiowa.edu/~jones/cards/chad.html.

The current lack of widely available 1% audit reports has led some to propose using exit polls to audit the tabulation of elections. Exit polls have been used in the past in countries with shaky democracies to ensure that the election was conducted fairly, and have played critical roles in overturning obviously corrupt elections, such as in the 2004 Ukrainian presidential election.[41]

Exit polls had not been a central part of election auditing in the United States, until 2004, when the discrepancies between the National Election Pool (NEP) exit polls and official election returns were shown to be so pronounced. This led to charges of fraud that have circulated around the Internet,[42] repudiations of these charges by others on the Internet,[43] a murky explanation of what went wrong from the two organizations that actually administered the poll (Edison/Mitosfky 2005), a blue ribbon academic panel empaneled by the Social Science

---

[41]The official returns from the first round of voting in Ukraine had incumbent prime minister Yanukovych leading the challenger Yushchenko by a margin of 49.4% - 46.3% (Finn 2004a). An exit poll conducted by the Democratic Initiatives Foundation gave Yushchenko a 54%-43% lead over Yanukovych; a second conducted by Socis gave Yushchenko a 49.4%-45.5% lead.

It is also important to recall that the Democratic Initiatives Foundation exit polls were funded by western donors, and that Yushchenko's supporters linked the early release of the exit poll results with demonstrations aimed at casting doubt on the official count. This observation is not meant to deny the fraud in Ukraine, or even the accuracy of the poll, just to note the Ukrainian case is probably not a model of exit polls that should guide independently monitored election counts in the United States.

[42]The most relentless attack on the exit polls has been by Steven F. Freeman at the University of Pennsylvania. See Freeman (2004).

[43]The clearest critique of the early Freeman paper is contained on the blog operated by Mark Blumenthal, whose *nom de plume* is "The Mystery Pollster." See http://www.mysterypollster.com/main/2004/11/faq_questions_a.html.

Research Council[44] that spawned, among other things, a comprehensive review of the exit poll controversy (Traugott, Highton, and Brady 2005).

As someone who contributed to the Samizdat literature that sought to quell the conspiracy theories that arose from this controversy (Stewart 2004b), my opinions about this specific case are well established: all of the evidence points to a disastrously implemented exit poll effort in 2004, rather than massive fraud of historic proportions. However, the controversy that has grown up around the 2004 exit polls can be illuminating, by showing how current exit polls are poorly constructed to audit election tabulations.[45]

Even if they were properly run, the current construction of the network exit polls make them poorly designed to detect anything but the most egregious types of fraud. This is important to remember when we recall that much of the push for using exit polls to audit the vote count has come from critics of electronic voting machines. The fraud schemes that most concern computer security experts are those are subtle and difficult-to-detect. For instance, a common supposed scheme would be to program a DRE to hijack every 100th vote for the Democrat and allocate it to the Republican instead. The idea, in general, is that electronic election machine hackers, being more cleaver than prior generations' political hacks, would corrupt the count just enough to affect the outcome, and no more.

---

[44]The following web site documents the work of the SSRC's National Research Commission on Elections and Voting: http://elections.ssrc.org/commission/research/.

[45]The Election Science Institute conducted pilot projects in New Mexico and Ohio in 2004, with the intention of discovering whether exit polls can be used in this manner. These projects have led to some thoughtful reports about that experience that should be read by anyone interested in using exit polls to monitor vote tabulation. See ESI (2004b; 2005a,b).

Let us suppose, then, that such a hack was in fact instituted. Let us suppose it was

unleashed on a fictional state named "Ohio," whose voters had, in fact, given John Kerry a

victory by a margin of 50.1% to 49.9% over Bush. Suppose, further, that hackers had intervened

to divert 1% of Kerry's vote, leading to an official count of 49.6% for Kerry and 50.4% for

Bush. Finally, suppose that if we ask people in an exit poll how they voted, they reply truthfully.

How many voters would we have to interview to expose the subterfuge?

If we use the simplest approach, which is to randomly sample the exit poll participants

and then conduct a simple test of the probability that an exit poll sample of 50.1% for Kerry

could come from an actual population that had voted 49.6% for Kerry, we would need over

30,000 respondents to reject the null hypothesis at the traditional probability value of 5%.[46]

However, exit polls do not randomly sample voters, they randomly sample precincts and then

randomly sample voters. Because voters in a given precinct tend to share political

characteristics, responses within a precinct are not independent, which is a critical requirement

of most classical statistical tests. Statisticians have long dealt with this sort of "clustering

problem," and a well-known finding in this field is that one would take *even more* respondents to

overcome the problem of the non-independence of respondents within precincts. The rule of

thumb is to double the number of respondents needed.[47] That gets us up to 60,000 voters we

---

[46]I leave aside the standard that courts and election officials should use in rejecting the null hypothesis that the exit poll results deviation from the official ones. The 5% criterion has been standard in the social sciences for years, but it is unclear whether this is too generous or conservative for the application to uncovering election fraud.

[47]The precise nature of the clustering problem is very fact-specific, and the documentation of the NEP public-access data set is not especially helpful in understanding how to correct for its sample design. A discussion of the sampling issue in the context of the 1996 exit poll suggests a 30% increase in sample error because of the cluster design (Merkle and

would need to interview, just in this one fictional state of "Ohio." This is approximately the

number of respondents *nationwide* in the 2004 NEP exit poll.[48]

As this admittedly contrived example suggests, even under the best of circumstances, exit

polls are unlikely to be economically viable options for detecting the types of mischief that many

anti-DRE activists are worried about, nor would they be very efficient for the more prosaic task

of documenting whether jurisdictions are getting better or worse in tabulating outcomes.

The most interesting idea that arises from the longing to use exit polls as an independent

check on election tabulation is that of providing an independent stream of information about

voters' intentions against which the official tabulation can be checked. That is the idea behind

the 1% audit laws, which could be beefed up in states that adopt voter-verifiable paper trails for

their DREs. If *all* paper ballots were counted and compared against the electronic results, that

would provide not only a good check against fraud, in addition to information to officials and the

public about how well the county or town conducted its elections.

Where do we stand for 2004? How good was the voting tabulation and was it better or

worse than 2000? The sad truth is that there is no solid evidence, at this point, to provide an

answer to these questions. The lack of widespread use of post-election audit procedures by

states, and the failure of states that have them to publicize the results widely, is the greatest

---

Edelman 2000, p. 72).

[48]The NEP public-release data can be accessed through the Inter-University Consortium for Political and Social Research and the Roper Center. According to study description available from the ICPSR, the entire national sample includes 77,006 cases. See ftp://ftp.icpsr.umich.edu/pub/FastTrack/General_Election_Exit_Polls2004/readme.txt.

source of our ignorance. As a consequence, citizens have been tempted to rely on a fatally

flawed exit poll instrument, which has only muddied the waters further.

Summary

The problems associated with counting the vote in Florida in 2004 educated us all about the

myriad ways in which elections could be run poorly, and pointed to the many ways in which they

should, and could, be improved. The most intense scrutiny in Florida was on the specific

problem of voting machines, but all of the major postmortems on 2004 educated the public about

problems that exist in all aspects of running elections in the United States, and all advocated the

improvement of the complete voting system, from registration to certification.

For all of the attention focused on the problem since November 2000 and all the money

thrown at improving voting in the United States, it is impossible to demonstrate anything but the

most basic improvements in voting, nationwide, using systematic data. Claims have been made

by a host of interested parties about why things have gotten better or worse, but it would be an

abuse of the hard evidence we do have to make strong claims about 2004.

Rather than wallow in despair about the state of the evidence, we should be clear about

what we need to know in order to assess the quality of election administration in the United

States, going forward. With the experiences of 2000 and 2004 fresh in our memories, the

following section suggests the type of agenda the nation should follow in improving our

knowledge of how well elections are run in the United States.

II. What Is to Be Done?

In the previous section, I walked through the most critical steps of the election process that need

to be monitored, so that we can better know whether elections are being administered better. In

that excursion, I made reference to numerous data sources that are currently available for use in

this enterprise, and to sources and techniques that might be developed in the future. In this

section, I revisit the data source issue in a more focused way, in an attempt to lay out an agenda

for data gathering that would assist in documenting the integrity of electoral counts in the United

States.

Preliminary to this discussion, we should identify some general principles guiding this

data gathering and some important obstacles in the way of achieving our goals.

The four principles I would highlight are uniformity, transparency, expedition, and

multiple sources. By *uniformity,* I mean the adoption of a common language about election data.

For instance, "turnout" should mean the same thing in Massachusetts as it does in Georgia as it

does in South Carolina. I do not mean that all jurisdictions should run elections in the same

way — quite to the contrary. Most of the heterogeneity of election administration in the United

States is probably appropriate, given different practical circumstances facing a continental

nation. This heterogeneity provides a "laboratory" for further innovation. Still, when it comes

to certain critical terms, like "turnout," or critical practices, like when to count write-in votes and

how to report them, all jurisdictions in the United States should talk the same language and

follow the same rules.[49]

---

[49]Aside from providing the first comprehensive set of data about election administration
for the use by the general public, the EAC's Election Day Survey report is essentially a 266-
page-long plea for greater uniformity. See particularly its Part 3 ("Recommendations on Future

By *transparency,* I mean that *all* data pertaining to elections should be readily available in electronic form. For elections data, this includes not only the basics, like turnout and vote counts, but also precise information about voting machines used and precise documentation of post-election audits.[50]

By *expedition*, I mean that election data should be released soon enough to ensure against fraud in a particular election, and to allow election officials time to analyze the data so that improvements can be made in the future. Of the three principles mentioned thus far, this is the one that needs to be constrained the most by other values. Even under the best of circumstances, quick election counts are Janus-faced. On the one hand, a quick release of election returns is among the best checks against fraud. On the other hand, a quick release of election returns invites more errors.[51] That is why even the best data-release process would be staged over time. That is also why transparency and uniformity are so important, too. The most careful election counts always change preliminary election counts. Therefore, it is important that non-officials trust that these changes are proper.

Data Collection"), though the introduction to virtually every data chapter contains reports of significant problems due to the lack of even a modicum of a common language among election administrator nationwide.

[50]ESI's experience with trying to obtain such information in Ohio, a month after the 2004 general election, illustrates the current problems with meeting the transparency standard. See ESI (2004a).

[51]Indeed, I am convinced, by my informal discussion with officials in state election divisions following the 2004 election, that one of the reason why residual vote rates were so high in the past, and why they improved so much in 2004, is that election officials were just less rushed and more careful in 2004. In one instance, I called a state election official to ask why the state was taking so long to release its county- and precinct-specific turnout figures. His answer was "the Verified Voting people will hammer us if anything's wrong, so we have to make sure the numbers are absolutely right." Taking longer to produce final election return reports would probably be a good thing, up to a point.

The final principle is that of relying on *multiple sources* of information in assessing the quality of elections. The core of election data are the returns, but we need to establish independent means of checking those returns. This is the kernel of valuable insight from the desire to use exit polls to audit election tabulations. It is also the value of relying on survey instruments, like the Current Population Survey, when appropriate.

The three obstacles to focus on are federalism, state and local election officials, and the nature of elections themself. *Federalism* is an obvious obstacle, since it has given us so many of the problems we currently face in understanding elections in the United States and documenting their improvement. Because of federalism, we have different electoral vocabularies and procedures. These differences make understanding election returns at the required level of specificity difficult. Many of the panicked alarms sent up during the 2004 electoral count in Ohio, for instance, were due to differences in the temporal order of counting the vote of different communities — these differences could easily make it appear that some communities had seen thousands of "phantom voters," while others had had voter suppression of historic proportion. At the same time, federalism makes it so that many of the changes that are necessary for improving our understanding of elections must be enacted serially, one state at a time.

State and local election officials are another obstacle in improving the independent monitoring of elections. This charges can easily be seen as unfair, since election officials (especially those in the counties and towns) usually labor under inadequate budgets and political support to pull off a heroic task of running elections to the standards that voters expect. However, the fact is that since the presidential election of 2000, the job of election officials has

changed, traditional approaches to running elections are inadequate, and therefore the jobs of many election officials are at risk.

This, in turn, produces two problems. The first is a series of principal-agent problems, alluded to before, in which local officials are reluctant to implement changes mandated by the state or federal governments that would increase transparency and uniformity.

The second problem is that the job of election administrator, as understood by the administrators themselves, has diverged mightily from the job as understood by the average citizen. For most election officials, the job comes down to calling the winner correctly without calling attention to how the election was administered. For most voters, the job comes down to recording the voter's vote correctly. The two tasks, while highly correlated, are not identical. Obviously, if voters' individual votes are tallied correctly, the likelihood that the proper winner will be identified is certain. However, virtually everyone agrees that the proper winner is already being identified virtually all the time right now, if we consider the thousands of elections run in the United States every four years. Therefore, even if energy is spent making counts more precise, election outcomes will rarely change. This leads many election officials to be obstructionist about efforts to increase the accurate of vote counts. As a consequence, election officials are now often on the defensive and seen as being in the "anti-reform" camp in many states.

The greatest symbol of the reluctance of state and local officials to increase uniformity and transparency in reporting election information was the vote by the National Association of Secretaries of State that advocated the abolition of the U.S. Electoral Assistance Commission,

which has a mission to encourage greater uniformity and transparency in election administration (NASS 2005).

The final obstacle to increasing the quality of data for monitoring American elections has already been alluded to — *disagreements over what elections are about.* Is a well-run election one that chooses the right winner, or one that records the choices of individual voters the most accurately? It is trivially true that elections that record the votes of each voter perfectly accurately and then tabulate all the votes perfectly always call the right winner. Yet election administration is governed by the economic laws of diminishing marginal returns and accelerating marginal costs to effort. The economic costs of *literally* ensuring that "every vote is counted" are so great in large electoral jurisdictions that every vote is in fact not counted, for reasons that have nothing to do with election fraud. Election officials conduct an implicit cost-benefit analysis when they decide how much effort to put into counting, double-checking, and reporting election returns.

Said more directly, while a perfectly-counted election will always return the right winner, a less-than-perfectly-counted election will *almost always* return the right winner. The practical question for election administrators is how much less-than-perfect is good enough.

Voters are typically horrified when presented with this line of thinking. It seems cavalier and even sacrilegious. This horror is evidence that voters care about something beyond winners and losers in elections. While the constitution of a democratic state regards elections instrumentally, voters regard them expressively.

This disagreement over what elections are intended to achieve has raised the costs of reform as the past four years have unfolded. Election officials have largely dug in their heels on

greater transparency and uniformity, which has moved political activity up the political food chain and even into the initiative arena. This, in turn, introduces collective action problems (Olson 1965) that favor election officials over organized voter groups.

Thus, achieving progress in the assessment of elections will be difficult, but not impossible. Technology has actually been a boon to transparency and uniformity, since most electronic tabulation systems automatically generate the number of ballots counted and produce electronic reports.[52] Now that the standard-setting process for elections is situated in the National Institute of Standards and Technology (NIST), there is great hope that in the future, the evolution of election systems will facilitate even greater uniformity and transparency in the reporting of election systems. In addition, many well-established national foundations, like Ford and Carnegie, have taken an interest in sponsoring efforts that are intended to provide other independent audits of elections, like the Votewatch/Election Science Institute project.

Therefore, assuming that it is actually possible to improve the quantitative monitoring of elections in the United States, both to detect fraud and to document successes and failures in electoral innovations, what should be done?

In answering this question, we should be explicit about identifying the general sources of information about election performance. All have been introduced already. They are (1)

---

[52]For this reasons, attacks on tabulation systems by anti-DRE groups (McCullagh 2004) seem to be wrong-headed, *especially* in those cases where the systems are counting scanned ballots, and in the prospective cases where the systems will be aggregating votes from DREs that have a "paper backup." The easy availability of detailed, precinct-specific election returns is a godsend to election auditing and monitoring.

election returns, (2) systematic surveys of voters, (3) systematic observation of electoral

processes, and (4) systematic surveys of election officials.[53]

Election returns.

To provide direct evidence of potential registration, polling place, voting machine, and

tabulation problems, every jurisdiction that administers elections should make the following

information available *for each precinct* in that jurisdiction, after the election.

- Total number of voters allowed to vote (i.e., turnout)

- Total number of people who presented themselves to vote but were turned away

- Complete vote count for all candidates, including write-ins[54]

- Precise manufacturer and model description of all voting machines used, and the

  number used in that precinct

- Number of absentee votes attributable to that precinct[55]

- Number of provisional ballots requested and the number eventually allowed.

In addition, there will be reports that are more appropriately given at the jurisdiction

level, both because of administrative practicality and concerns that reporting some things at the

---

[53]The following list of necessary data is consistent with Caltech/Voting Technology Project (2004b); the attempt here is to motivate such a list more thoroughly. Two thoughtful reports that discuss what research is needed to provide for a thorough understanding of the voting system in the united states are Frankel (2004) and National Research Commission on Elections and Voting (2005).

[54]Alternatively, a summary count of all write-ins, or a "scattering" category could be included. The important thing here is to account for all votes.

[55]Breaking out election returns by absentee ballots at the precinct level is unlikely to be a good thing, because here we approach problems of anonymity of the ballot. Still, knowing the number of absentee voters that belong to particular precincts would be helpful in monitoring election performance.

precinct level could compromise the anonymity of voters. Therefore, jurisdictions responsible

for elections should report the following *for the jurisdiction as a whole* following each election:

- Number of absentee ballots mailed out and returned (on time and after the
deadline)

- Number of absentee ballots disallowed, and the reason for the disallowance

- Complete vote count for all candidates, including write-ins, for ballots cast
provisionally and absentee (as separate categories)

- Complete counts of provisional and absentee ballots that were *disallowed* [56]

Systematic surveys of voters.

As discussed previously, exit polls and other surveys are probably not well-suited for ferreting

out most fraud schemes, but they are likely well-suited for generally monitoring other aspects of

election administration, such as experiences at the polling place, problems encountered with

registration, malfunctioning election machines, etc.

Unfortunately, the two major academic political surveys, the National Election Study and

the General Social Survey, have not added questions about voters' experience with the voting

process.[57] They should. These academic surveys are independent enough of election officials

---

[56]An important question that arises in the use of both of these procedures is whether they are used by central election administrators to help rig elections. Therefore, knowing whether disallowed ballots contained different votes from allowed ballots is important to know. This information could be used in tandem with the precinct-level data about the number of provisional ballots that were requested and eventually denied.

[57]It is unclear whether the National Annenberg Election Survey 2004 contained any questions about the experience that voters had with the administration of the election system in 2004.

and well-respected that their findings would quickly become the gold standard in national monitoring. They are also able to ask questions that a government-sponsored survey like the CPS cannot ask, such as whether respondents trust that their vote was properly counted.

We currently have two general efforts that could be better adapted for election monitoring than they are now, the VRS of the CPS and exit polls. The current shortcoming of the VRS arise because it is optimized for understanding the administration of the NVRA, whereas it should now move its attention to the administration of HAVA. The only current question within the VRS that taps a voter's experience on Election Day is addressed only to non-voters, when it asks why a non-voting respondent did not vote. People who voted need to be asked whether they encountered substantial problems with registration, polling place administration, and with voting machines. This, of course, would add some complexity to the VRS, but if the CPS is going to be the only government-sanctioned survey of experience with the election system, more complexity is called for.

Similar things could be said about the exit polls. In 2004, the NEP exit poll contained the following question: "How confidence are you that votes in your state will be counted accurately?" The response categories range from very confident (50% of valid responses) to "not at all confident" (0.4%).[58] However, the public-access data available through the ICPSR are not consistent with the codebook, so it is difficult to assess the quality of this data for the task of understanding how voters perceived election administration in their states.[59] It appears that only

---

[58]These marginal frequencies are taken from the codebook for the NEP exit poll, obtained through the ICPSR (study number 4181).

[59]The primary problems are two: (1) According to the codebook (p. 53), this question only appears on Version 3 of the national questionnaire; according to the data as downloaded, all

23% of respondents answered this question in the nationwide survey, for an average of only 63 respondents in each state. Respondents in eight states[60] were asked the question in the larger-sample state surveys, but the states chosen were hardly representative of the nation as a whole. This question about counting accuracy should become a staple of nationwide exit polls; three more that probe the respondent's experience with each step in the voting chain (registration, polling place experience, and voting technology), should become regular features on future exit polls, asked in all the state samples.

Systematic observation of polling places.

Survey respondents often mis-perceive or mis-remember their experiences, and therefore national surveys that ask voters about their experiences with the voting system will be of limited utility, especially in the cross-section. It is therefore important that the electoral process be observed directly, by trained observers, who can document the presence, or absence, of problems at polling places.

A national effort is needed that deploys observers to a sample of polling places on Election Day. The purpose of this effort would be to document changes in polling place operations at the *state level,* therefore the number of sampled precincts in each state would need to be in the hundreds, rather than the thousands, which would be necessary if our goal was to

---

data are generated by Version 3 questionnaires. (2) According to the codebook, there are 10,527 missing values; however, there is no information in the codebook about why the data should be missing from 77% of the observations.

[60]California, Florida, Georgia, Louisiana, Maryland, Ohio, Oregon, and Washington.

monitor individual jurisdictions within states, as well. The poll observers would report, for the sampled precincts, the following information:

- The number of people who had presented themselves, asking to vote

- The number of people who had been allowed to vote

- The number of people who had been turned away, and the reasons for turning them away

- The number of people who completed a provisional ballot

- Length of the line to (a) check-in at the precinct and (b) to use a voting machine, at specified times during the day

- Number of people staffing the polls

The question arises about who would sponsor such a national survey. Two organizations come immediately to mind, the EAC and/or the General Accountability Office, both of which have conducted national surveys of election practices, but none in real-time. Another model would have having the National Election Pool sponsor such a study, as a part of their exit poll efforts.

Systematic surveys of election officials.

The three previous efforts are aimed at measuring outcome metrics about the voting process. Recall that the broad goal in this enterprise is to document the performance of the election system so that it can be improved. Therefore, it is important that we also establish a system that measures policy inputs, including information about administrative practices and about changes in election administration.

A previous effort was made in this direction, following the 2001 election, when the

General Accounting Office conducted a mail survey of a randomly-chosen sample of 513

jurisdictions, out of a sample size of 607 (GAO 2001, pp. 348–372). This survey asked the

jurisdictions a comprehensive set of questions concerning topics such as the voting technology

used in the jurisdiction, reasons for changing technologies, voter education, vote tabulation

procedures, and the like. Like most GAO surveys, however, the raw results were not made

public, nor was the survey repeated, so it was impossible to tie these results to outcome

measures, or to document the effects of policy change over the past quadrennium.

The U.S. Election Assistance Administration has embarked upon an ambitious research

agenda, attempting to go beyond the simple biennial NVRA voter registration reports it inherited

from the Federal Elections Commission. Currently, however, both of its major survey projects

are addressed to state administrators, who are requested to gather massive amounts of

information from their counties and towns that administer elections.[61] The report of the Election

Day Survey (U.S. EAC 2005b) is instructive, both for the substance of the data that are reported

and for the account of the great difficulties in acquiring the data.

Although it would seem to be a daunting task to gather information about the

administration of each polling place in the United States, several states already publish summary

information about precincts, such as the number of precincts in each county, and states are

increasingly publishing detailed election return data at the precinct level. It would be trivial, on

[61]The two survey instruments may be found at the following URLs:
http://www.eac.gov/docs/Election%20Day%20Survey.pdf ("Election Day Survey" of Election
Day practices and incidents) and http://www.eac.gov/docs/nvra%20survey%202005.pdf ("Voter
Registration Survey" of registration statistics and processes). The results of these surveys have
now been released. See U.S. EAC (2005a,b)

the margin, to add to these reports other basic data, such as the number of workers who actually worked during the day (as opposed to the number assigned) and the number of people in line waiting to vote at a few designated times during the day.

*   *   *

Virtually all of these proposed efforts would produce data that are rather crude. They would probably be ill-suited for uncovering subtle election fraud, both because of the crudeness of the measures and the time between gathering all the data together and reaching conclusions about the data. Therefore, efforts specifically aimed at ferreting out fraud are a related, but separate, enterprise. Still, as the widespread use of the crude residual vote measure has shown, even a highly aggregated measure can tell us a lot about the performance of the electoral system if it is used carefully and in parallel with other measures.

III. Conclusion

The major purpose of this paper has been to motivate efforts toward greater care in documenting how elections are conducted in the United States. Starting with the 2004 election itself, we see what even though billions of dollars have been spent on improving elections over the past four years, we only dimly know whether those dollars were well spent. We still have barely any systematic evidence about the quality of registration systems, polling place practices, or tabulation accuracy. We have a better sense about the quality of voting machines, but that is based on one measure (residual vote) that needs to be supplemented, and needs more information about the voting machines that are being used, in order to be useful.

Therefore, we should be confident in concluding that Americans, on average, used better voting machines in 2004 than in 2000, and that continued attention on retiring older machines will improve tings even more. The controversy over whether adding voter verifiable paper trails will make things more confusing for voters will be partially answered in the future using the same techniques that judged 2004 to be better than 2000.

Because the decline in the residual vote rate between 2000 and 2004 is also due, in part, to greater care given in tabulating results, it is also likely that the "back office" part of counting ballots was handled better in 2004, but here the evidence becomes very thin. Finally, we simply do not know whether registration got better, polls were administered better, or whether requiring provisional ballots in all states improved the quality of elections in 2004.

As 2004 demonstrated, rumors can fly when election data are bad, and the integrity of an election can quickly be cast into doubt when the evidence is murky. That is why all citizens of all political stripes, and not just political scientists, should insist that governments and private organizations work to improve how elections are documented in the United States.

Cited Works

Alvarez, R. Michael, Stephen Ansolabehere, and Charles Stewart, III. 2005. "Studying Elections: Data Quality and Pitfalls in Measuring Effects of Voting Technologies." *Policy Studies Journal* 33(1).

Alvarez, R. Michael, and Thad E. Hall. 2003. Why Everything That Can Go Wrong Often Does: An Analysis of Election Administration Problems. *VTP Working Papers,* no. 10, November 26, 2003 (rev.). http://vote.caltech.edu/media/documents/wps/vtp_wp10.pdf

Alvarez, R. Michael, Jonathan N. Katz, and Sarah A. Hill. 2005. Machines Versus Humans: The Counting and Recounting of Pre-Scored Punchcard Ballots. *VTP Working Papers,* no. 32, July 2005. http://vote.caltech.edu/media/documents/wps/vtp_wp32.pdf

Ansolabehere, Stephen, and Andrew Reeves. 2004. Using Recounts to Measure the Accuracy of Vote Tabulations: Evidence from New Hampshire Elections, 1946-2002. *VTP Working Papers,* no. 11, January 2004. http://www.vote.caltech.edu/media/documents/wps/vtp_wp11.pdf

Ansolabehere, Stephen, and Charles Stewart, III. 2005. "Residual Votes Attributable to Technology." *Journal of Politics* 67(2): 365-89.

Caltech/MIT Voting Technology Project. 2001a. Residual Votes Attributable to Technology. *VTP Working Papers,* no. 2, March 2001. http://vote.caltech.edu/media/documents/wps/vtp_wp2.pdf

———. 2001b. Voting: What Is/What Could Be. http://vote.caltech.edu/media/documents/july01/July01_VTP_Voting_Report_Entire.pdf

———. 2003. Voting in Massachusetts. http://www.vote.caltech.edu/media/documents/VotinginMass.pdf

———. 2004. Insuring the Integrity of the Electoral Process: Recommendations for Consistent and Complete Reporting of Election Data. October 2004. http://vote.caltech.edu/media/documents/auditing_elections_final.pdf

Chaum, David. 2004. "Secret-Ballot Receipts: True Voter-Verifiable Elections." *IEEE Security and Privacy* 2(1): 38-47.

Coleman, Kevin J., and Eric A. Fischer. 2001. Elections Reform: Overview and Issues. *CRS Report for Congress,* April 30, 2001. http://www.electionreform.org/ERMain/resources/docs/crs/rs20898.pdf

———. 2004. Elections Reform: Overview and Issues. *CRS Report for Congress,* January 21, 2004. http://fpc.state.gov/documents/organization/28920.pdf

Common Cause. 2004. Report from the Voters: A First Look at 2004 Election Data/Common Cause Agenda for Reform. December 2004. http://verifiedvoting.org/downloads/CommonCauseReporttotheVoters.pdf

Edison/Mitofsky. 2005. Evaluation of Edison/Mitofsky Election System 2004 Prepared by Edison Media Research and Mitofsky International for the National Election Pool (NEP). January 19, 2005. http://www.exit-poll.net/election-night/EvaluationJan192005.pdf

Election Science Institute (ESI). 2004a. Votewatch Exit Poll, Albuquerque General Election 2004. November 30, 2004. http://www.votewatch.us/Members/StevenHertzberg/report.2004-12-13.1986469362/report_contents_file/

———. 2004b. Votewatch Finds Ohio Not Currently Auditable, Due to Issues of Timeliness, Completeness, and Quality of Documents from County Election Officials. *Votewatch research report.* December 7, 2004. http://www.votewatch.us/Members/StevenHertzberg/report.2004-12-15.2641756478/report_contents_file/

———. 2005a. Can Exit Polls Be Used to Monitor Elections? *ESI Brief,* May/June 2005. http://www.votewatch.us/Members/stevenhertzberg/report.2005-05-10.0988159730/report_contents_file/

———. 2005b. Ohio 2004 Exit Polls: Explaining the Discrepancy. *ESI Brief,* June 6, 2005. http://www.votewatch.us/Members/stevenhertzberg/report.2005-07-19.7420722886/report_contents_file/

electionline.org. 2003. Ready for Reform? *Election reform briefing,* March 2003. http://www.electionline.org/Portals/1/Ready%20for%20Reform.pdf

Finn, Peter. 2004b. "Widespread Vote Fraud Is Alleged in Ukraine." *Washington Post*, November 22, 2004, p. A01.

———. 2004a. "Partial Vote Results Show a Tight Race in Ukraine Runoff." *Washington Post*, November 22, 2004, p. A15

Florida Division of Elections. 2003. Analysis and Report of Overvotes and Undervotes for the 2002 General Election Pursuant to Section 101.595, Florida Statutes. January 31, 2003. http://election.dos.state.fl.us/reports/pdf/03OverUnderVotes.pdf

———. 2005. Analysis and Report of Overvotes and Undervotes for the 2004 General Election. January 31, 2005. http://enight.dos.state.fl.us/reports/pdf/Over_Under_Report_04.pdf

Frankel, Mark S. 2004. Making Each Vote Count: A Research Agenda for Electronic Voting. October 2004. http://www.aaas.org/spp/sfrl/evoting/report2.pdf

Freeman, Steven F. 2004. The Unexplained Exit Poll Discrepancy. *Research Report.* December 29, 2004. http://center.grad.upenn.edu/center/get.cgi?item=exitpollp

Herron, Michael C., and Jonathan N. Wand. 2005. "Assessing Partisan Bias in Voting Technology: The Case of the 2004 New Hampshire Recount." http://www.dartmouth.edu/~herron/nh.pdf

McCullagh, Declan. 2004. "E-Voting Critics Report New Flaws." *C-Net news.com*, September 22, 2004. http://news.com.com/E-voting+critics+report+new+flaws/2100-1028_3-5378199.html

Merkle, Daniel M., and Murray Edelman. 2000. "A Review of the 1996 Voter New Service Exit Polls from a Total Survey Error Perspective." In *Election Polls, the News Media, and Democracy*, edited by P.J. Lavrakas and Michael Traugott, pp. 68-92. New York: Chatam House.

Miller, Geralyn. 2005. "Methodology, Statistics, and Voting Error: An Exploration of 2000 Presidential Election Data in Two States." *Policy Studies Journal* 33(1): 1.

National Association of Secretaries of State (NASS). "NASS Position on Funding and Authorization of the U.S. Election Assistance Commission." http://www.nass.org/EAC%20Position%20Cover%20Ltr.pdf

National Research Commission on Elections and Voting. 2005. Challenges Facing the American Electoral System: Research Priorities for the Social Sciences. March 1, 2005. http://elections.ssrc.org/research/FinalReport030105.pdf

Olson, Mancur. *The Logic of Collective Action*. Cambridge, Mass.: Harvard University Press, 1965.

Stewart, Charles, III. 2004a. The Reliability of Electronic Voting Machines in Georgia. *VTP Working Papers,* no. 20, October 2004. http://www.vote.caltech.edu/media/documents/wps/vtp_wp20.pdf

———. 2004b. Addendum to Voting Machines and the Underestimate of the Bush Vote. December 5, 2004. http://www.vote.caltech.edu/media/documents/Addendum_Voting_Machines_Bush_Vote.pdf

———. 2005. Residual Vote in the 2004 Election. *VTP Working Papers,* no. 25, February 2005. http://vote.caltech.edu/media/documents/wps/vtp_wp25.pdf

_____. 2006. Residual vote in the 2004 Election. *Election Law Journal,* forthcoming.

Traugott, Michael, Benjamin Highton, and Henry E. Brady. 2005. A Review of Recent Controversies Concerning the 2004 Presidential Election Exit Polls. March 10, 2005. http://elections.ssrc.org/research/ExitPollReport031005.pdf

U.S. Election Assistance Commission. 2005a. *The Impact of the National Voter Registration Act of 1993 on the Administration of Elections for Federal Office.* Washington, D.C., June 30, 2005. http://www.eac.gov/docs/NVRA-2004-Survey.pdf

_____. 2005b. *Final Report of the 2004 Election Day Survey.* Washington, D.C., September 27, 2005. http://www.eac.gov/election_survey_2004/pdf/EDS-Full_Report_wTables.pdf

U.S. General Accounting Office. 2001. Elections: Perspectives on Activities and Challenges across the Nation. *GAO Report to Congressional Requesters,* no. GAO-02-3 October 2001. http://www.gao.gov/new.items/d023.pdf

UNC School of Law. 2005. Final Report: 2004 Election Protection, North Carolina. no. March 31, 2005. http://www.law.unc.edu/PDFs/EPreport.pdf

United States Election Assistance Commission. 2005. The Impact of the National Voter Registration Act. June 30, 2005. http://www.eac.gov/election_resources/NVRA-2004-Survey.pdf

Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr., Michael C. Herron, and Henry E. Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95(4): 793–810.

Figure 1. Provisional ballots counted in North Carolina counties, 2000 and 2004 general elections. (Circle proportional to 2004 turnout)
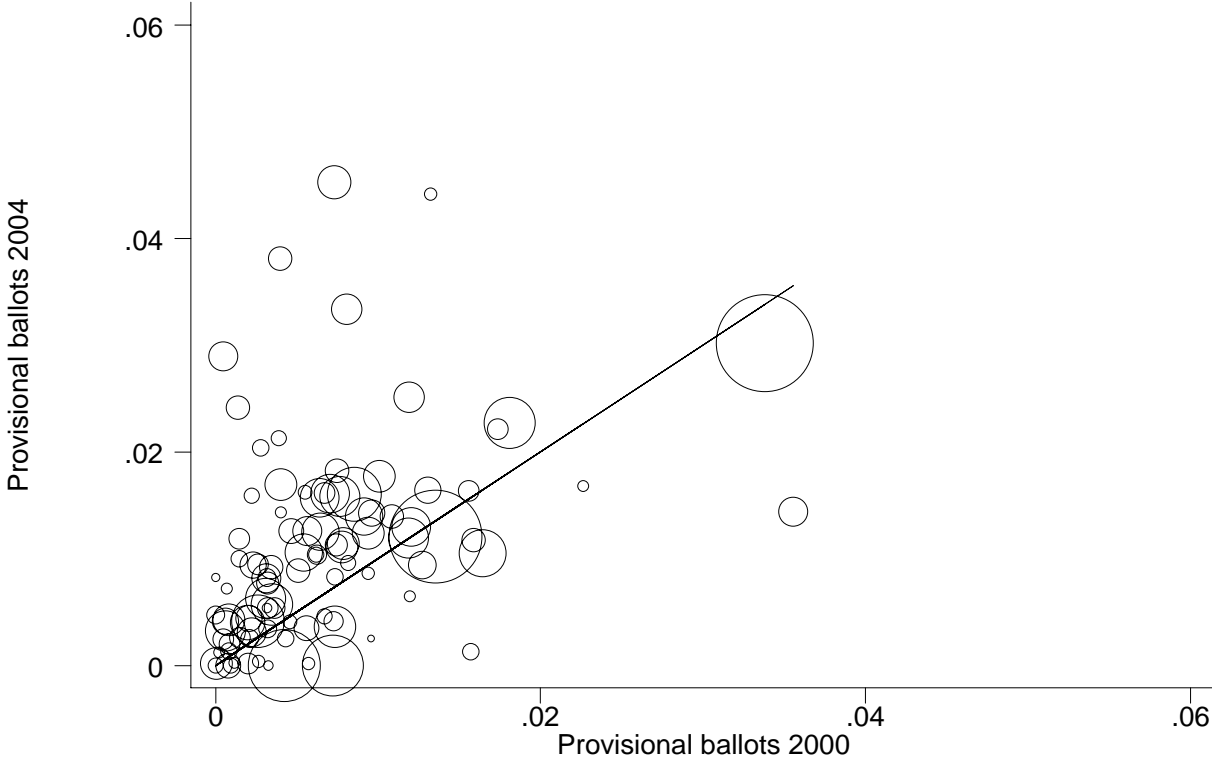
Figure 2. Percent of non-voters naming "registration problems" as the principal reason for not voting, 2000–2004, by state.
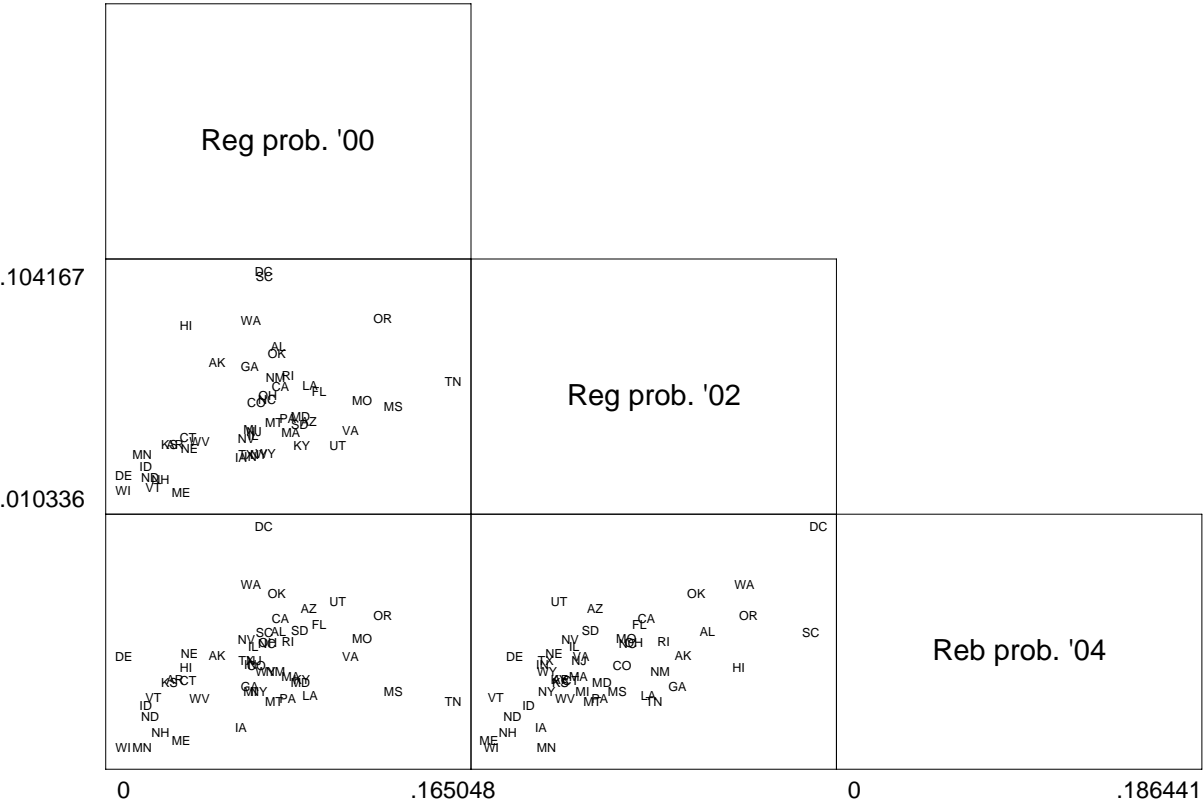
Figure 3. Percent of non-voters naming "inconvenient polling place or hours or lines too long" as the principal reason for not voting, 1996–2004, by state.
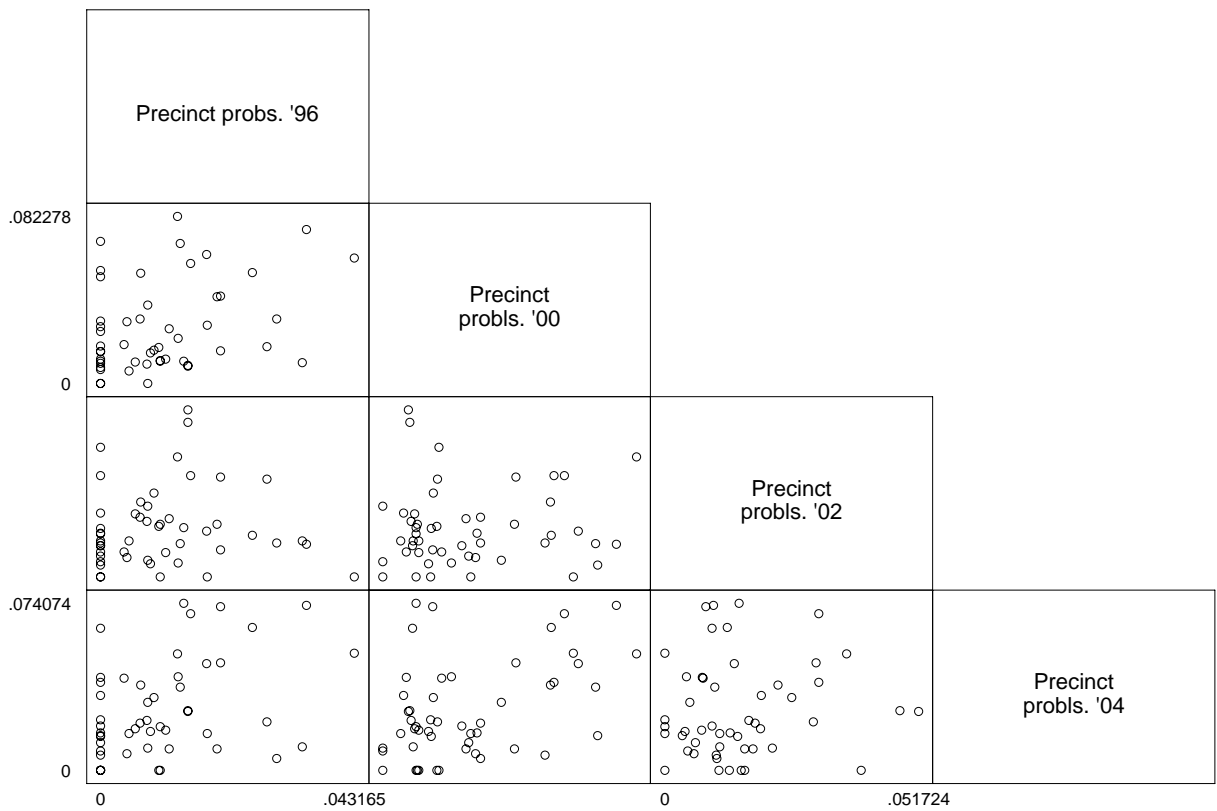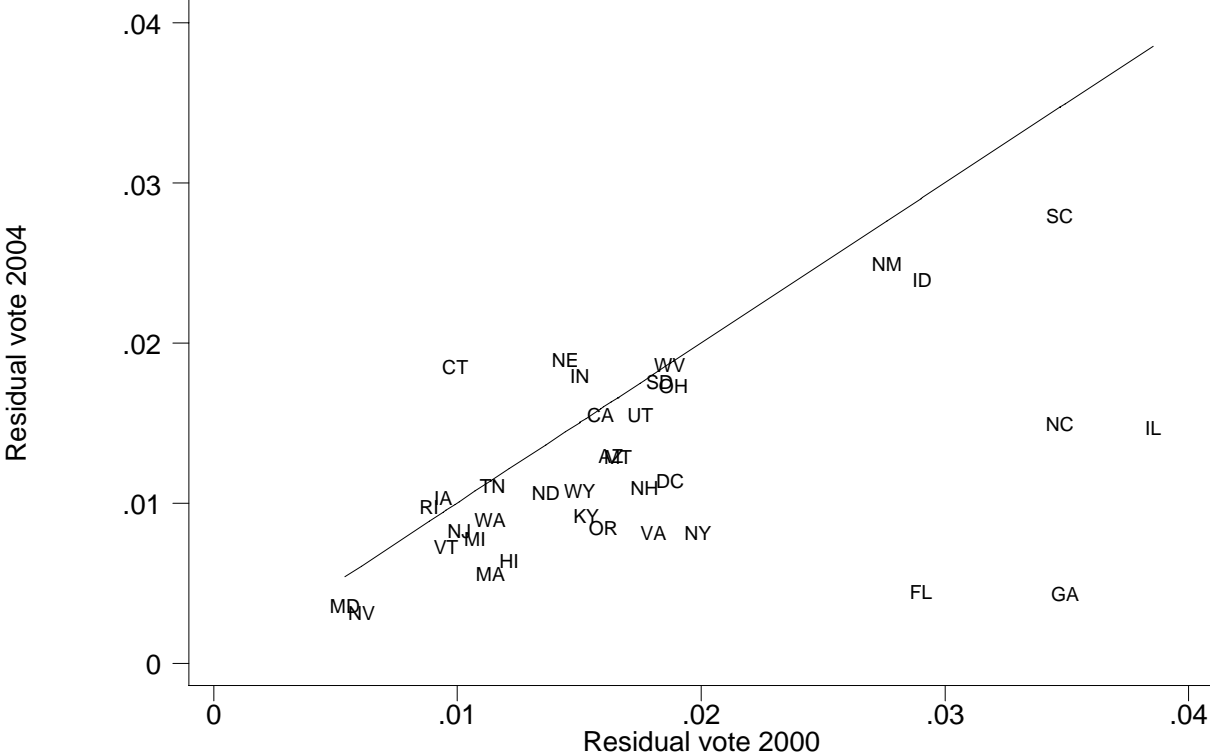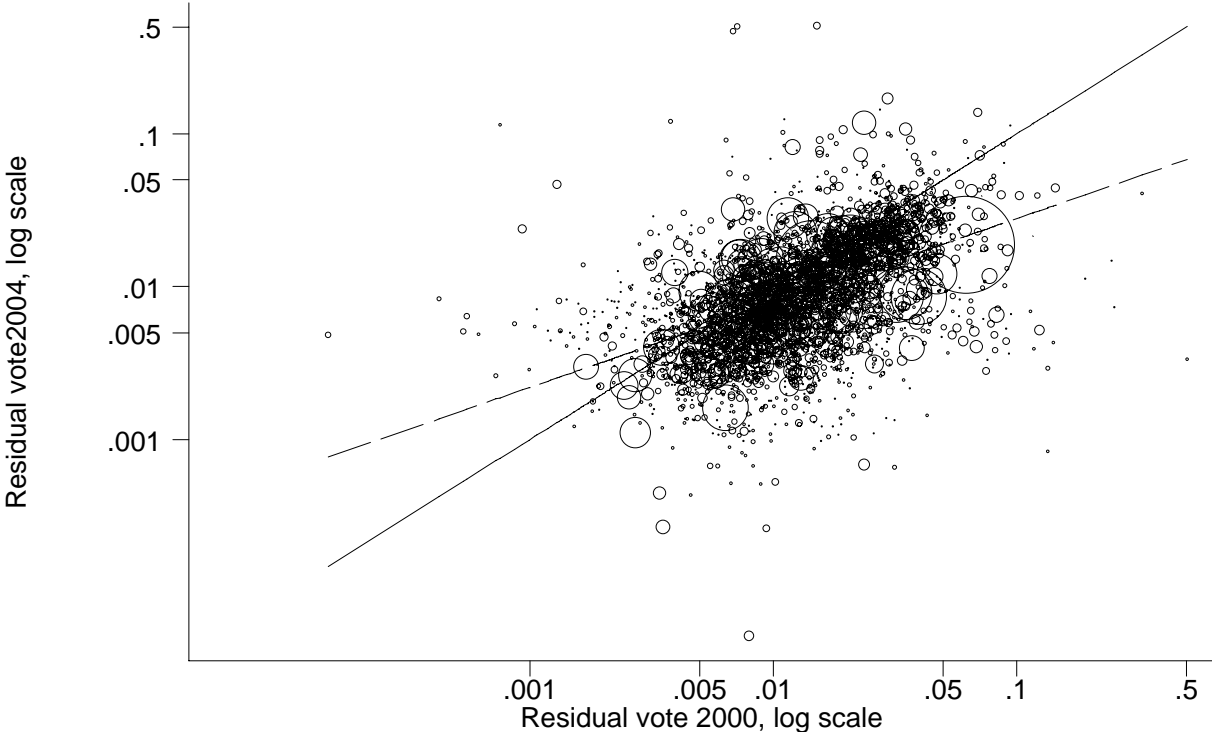
Figure 4. Residual votes rates by state, 2000 and 2004.

Figure 5. Residual vote rates by counties, divided by those that changed voting equipment between 2000 and 2004 and those that did not.

a. No voting machine change from 2000 to 2004
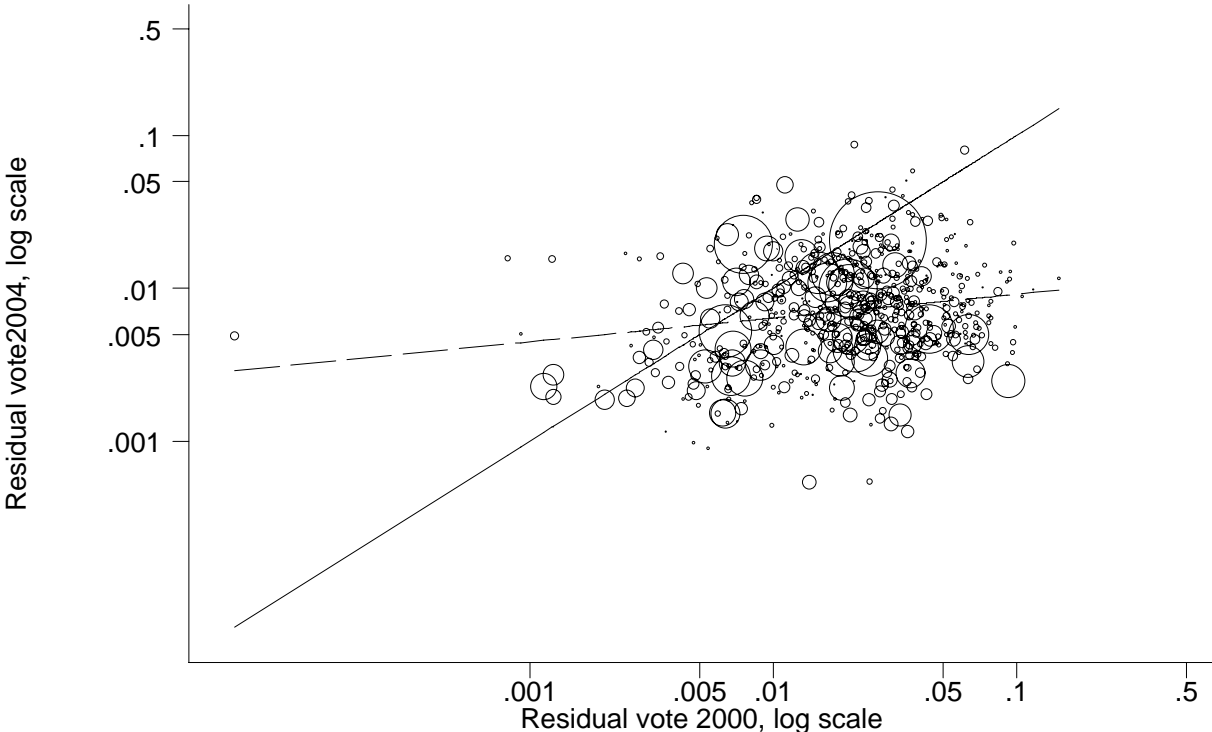

b. Voting machine change from 2000 to 2004

Table 1. Links in the election process chain.

Procedure	How affected by HAVA	Current and potential methods of assessing
Establishing the voting rolls (registration)	•Mandated integrated state-wide registration databases	•Focused audits of "paper trail" •Number of provisional ballots issued •Reports of registration problems at the polls •Exit polls
Checking-in voters at polling place	•Mandated "provisional ballots" to address Election Day disputes over registration	•Observer audits of polling places •Number of provisional ballots issued •Reports of registration problems at the polls •Exit polls
Use of voting equipment by voter	•Mandated handicapped accessibility •Outlawed punch cards and mechanical lever machines •Paid for most equipment upgrades (with local matches) •Provided role for EAC and NIST in assessing machines	•Residual vote rates •Observer audits of polling places •"Ease of use" surveys of voters
Counting ballots	Mandated state-wide criteria for counting ambiguous ballots	•Random-sample auditing •Study of recounts •Exit polls