

Elections, Voting, Technology

The series Elections, Voting, Technology examines the relationships between people, electoral processes and technologies, and democracy. Elections are a fundamental aspect of a free and democratic society and, at their core, they involve a citizenry making selections for who will represent them. This series examines the ways in which citizens select their candidates—the voting technologies used, the rules of the game that govern the process—and considers how changes in processes and technologies affect the voter and the democratic process.

Thad Hall is an associate professor of political science at the University of Utah and a research affiliate with the Caltech/MIT Voting Technology Project. He is the coauthor of several books on elections and voting, including *Point, Click, and Vote: The Future of Internet Voting* and *Electronic Elections: The Perils and Promise of Digital Democracy*, and coeditor of the book *Election Fraud*.

Confirming Elections: Creating Confidence and Integrity through Election Auditing
Edited by R. Michael Alvarez, Lonna Rae Atkeson, and Thad E. Hall

CONFIRMING ELECTIONS

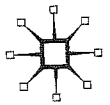
CREATING CONFIDENCE AND INTEGRITY THROUGH

ELECTION AUDITING

Edited by

R. Michael Alvarez, Lonna Rae Atkeson, and Thad E. Hall

palgrave
macmillan



CONFIRMING ELECTIONS
Copyright © R. Michael Alvarez, Lonna Rae Atkeson, and
Thad E. Hall, 2012.

All rights reserved.

First published in 2012 by
PALGRAVE MACMILLAN®
in the United States—a division of St. Martin's Press LLC,
175 Fifth Avenue, New York, NY 10010.

Where this book is distributed in the UK, Europe and the rest of the World,
this is by Palgrave Macmillan, a division of Macmillan Publishers Limited,
registered in England, company number 785998, of Houndmills,
Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan is the global academic imprint of the above
companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United
States, the United Kingdom, Europe and other countries.

ISBN: 978-0-230-34121-0

Library of Congress Cataloging-in-Publication Data

Confirming elections : creating confidence and integrity through election
auditing / Edited by R. Michael Alvarez, Lonna Rae Atkeson, and
Thad E. Hall.

pages cm.— (Elections, voting, technology)

ISBN 978-0-230-34121-0 (hardback)

- I. Elections—United States—Auditing. I. Alvarez, R. Michael, 1964–
- II. Atkeson, Lonna Rae, 1965– III. Hall, Thad E. (Thad Edward), 1968–
JK1976.C585 2012 20111050334
- 324.65—dc23

A catalogue record of the book is available from the British Library.

Design by Integra Software Services

First edition: June 2012

10 9 8 7 6 5 4 3 2 1

Printed in the United States of America.

CONTENTS

List of Figures	vii
List of Tables	ix
About the Editors	xi
List of Contributors	xiii
Introduction	1
<i>R. Michael Alvarez, Lonna Rae Atkeson, and Thad E. Hall</i>	
1 Confirming Elections: Creating Confidence and Integrity	5
<i>Thad E. Hall and Pamela Smith</i>	
2 Comprehensive Audits	23
<i>R. Michael Alvarez, Lonna Rae Atkeson, Thad E. Hall, and Jessica Tauerna</i>	
3 Confidence in the Electoral System: Why We Do Auditing	41
<i>Michael W. Traugott and Frederick G. Conrad</i>	
4 Summary of Election Audits in California	57
<i>Dean C. Logan</i>	
5 Machines versus Humans: The Counting and Recounting of Prescored Punchcard Ballots	73
<i>R. Michael Alvarez, Jonathan N. Katz, Sarah A. Hill, and Erin K. Hartman</i>	
6 The New Mexico Pilot Project	89
<i>Lonna Rae Atkeson, R. Michael Alvarez, and Thad E. Hall</i>	
7 Counting the Ballots: A Comparison of Machine and Hand Counts in New Mexico	101
<i>Yann Kerevel and Lonna Rae Atkeson</i>	

- Alvarez R. Michael and Thad E. Hall. 2006. "Controlling Democracy: The Principal-agent Problems in Election Administration." *Policy Studies Journal*, 34, 4, 491-510.
- Alvarez R. Michael, Thad E. Hall, and Betsy Sinclair. 2008. "Whose Absentee Votes Are Counted: The Variety and Use of Absentee Ballots in California." *Electoral Studies* 27, 4, 673-83.
- Paul Gronke, Eva Galanes-Rosenbaum, Peter A. Miller, and Daniel Toffey. 2008. "Convenience Voting." *Annual Review of Political Science*, 11, 437-55.

CHAPTER 12

USING RECOUNTS TO MEASURE THE ACCURACY OF VOTE TABULATIONS: EVIDENCE FROM NEW HAMPSHIRE ELECTIONS, 1946-2002

STEPHEN ANSOLABEHRE
AND ANDREW REEVES

AS NOTED BY SEVERAL OF THE AUTHORS IN THIS VOLUME, the now infamous 2000 presidential election exposed a surprisingly high level of inaccuracy in the tabulation of ballots. Differences between total ballots cast and votes counted were as high as 19 percent in some counties in Florida, and these discrepancies were widely attributed to the ballot formats, the handling of ballots, and machine operations.¹ Over the last 40 years the United States has introduced new technologies, especially punch card and optically scanned ballots, to improve vote tabulations. Thus, for those involved in the administration of elections the recount was particularly troubling. Certainly, the problems revealed in Florida suggested that these newer technologies may not in fact represent an improvement over traditional hand-counted paper ballots.

A number of important studies of the performance and accuracy of voting technologies have sought to measure the error rate of vote

tabulations. The main metric that emerges from these evaluations uses "residual votes"—the discrepancy between total ballots cast and votes cast for a particular office, such as president or governor. The incidence of residual votes should be unrelated to the type of voting technology used, and the difference in residual votes across technologies measures the extent of errors in the casting or tabulation of votes are attributable to the specific voting technology. Similar jurisdictions using different technologies ought to have the same residual vote rate, on average. By this metric, hand-counted paper ballots and optically scanned paper ballots have shown the better overall performance than punch cards, lever machines, and electronic voting machines.²

In this chapter, we examine a second measure of accuracy—the agreement between initial counts and recounts of ballots in contested elections. We term this the *tabulation validation rate*. Accordingly, the discrepancy between the initial count and the recount is the *invalidation rate*. When there is a legal challenge to an election or a mandated recount, state or local election offices conduct an audit of votes cast. For paper ballot systems, election officials reexamine the ballots to determine voter intention. Tabulations may change from the initial count to the recount for a variety of reasons: Ballots may be mishandled; machines may have difficulty reading markings; people and machines may make tabulation errors. Because recounts are used to certify the vote, greater effort is taken to arrive at the most accurate accounting of ballots cast. The initial count of ballots, then, is treated as a preliminary count, and the recount as the official count. The recount, then, validates the initial tabulation.

Recounts allow us to see more precisely whether the introduction of voting machines has improved the tabulation of votes. Do machines have higher validation rates (and lower invalidation rates) than hand counts?

To begin to answer this question, we examine data on recounts for New Hampshire towns. There are several important reasons for studying New Hampshire. First, New Hampshire uses hand-counted paper ballots extensively as well as optically scanned ballots. This allows us to contrast new and old methods of counting directly. Second, although recounts are generally rare, New Hampshire is an exception where about one in four ballots are recounted at each election—This provides a large number of cases to study.³ Third, New Hampshire has a uniform reporting system for recounts dating back at least to the 1940s. The data on recounts are, then, comparable and the historical record allows us to establish a solid baseline against which to contrast new tabulation methods.

Two caveats accompany the use of recounts to measure the validity of tabulations. First, this measure only concerns tabulation. It ignores other factors, such as voter confusion about how to mark a machine-readable

ballot, which might lead people to vote accidentally twice for an office or not at all. Such phenomena are captured by residual votes.

The metric differs for paper ballot systems and elections in which votes are cast on lever machines or electronic voting machines (also known as Direct Recording Electronics, or DREs). Mechanical and electronic voting machines do not retain a separate record of the voters' intention, so it is impossible, at least with current technology, to compare voters' intentions with the machines' recordings. Recounts with mechanical and electronic voting machines merely capture whether the election office made a recording error or whether the machine is functioning. With mechanical and electronic machines, it is impossible to gauge the degree of malfunction or the disparity between voter intentions and machine recordings.

HISTORICAL RECOUNTS

Historical recounts provide an important baseline for our study. Indeed by studying recounts across many offices, now that we have some historical distance, we could examine the percent difference between initial counts and recounts.

In our survey of the historical recount data in the state of New Hampshire from 1946 to 1962, we found that the first use of punch card machinery in the United States occurred in 1964. In the general elections in this span, 108 races had recounts. The offices involved range from registrar of voters to representative of Congress as well as town questions. The majority of the recounts (69) occurred in races for state representative. All recounts were aggregated to the town level, except those for Congressional races and state senate races that were aggregated to district levels.

These data sets provide complete and consistent information about the accuracy of hand-counted votes for nine elections over 16 years. Since all ballots during this time were hand counted once in the initial count, and again in the recount, we were able to evaluate accuracy rates of the counts independent of residual votes.

Our analysis consists of 415 cases where a case is a reported ballot tabulation of a town or (district) for each candidate running in a race. We computed the invalidation rate for each candidate for each office in each jurisdiction. That is, we computed the percent difference between the initial count and the recount for the votes recorded in each jurisdiction (i.e., town or ward) for each candidate seeking an office.

We found that two different average invalidation rates are of interest. First, the simple average invalidation reflects what happens in a typical

jurisdiction. In a state with many small towns and a handful of larger towns, the typical jurisdiction will be a small one. Second, the weighted average, weighting by total ballots cast, equals the percent of ballots cast that differed between the counts. This measure guards against the occasional aberrant tabulation in a small community. The weighted average measures the frequency with which one's vote is counted in the initial count with each type of equipment. We will focus on the latter, but will report both.

Historically, hand-counted paper tabulations have an invalidation rate just under 1 percent. The weighted average of the discrepancy between the initial count and the recount is approximately 0.83 percentage point with a standard error of 0.11.

HAND COUNTS VERSUS MACHINE COUNTS

Fast forward to 2000. While most counties and towns continue to use hand-counted paper, many of the towns in three counties in New Hampshire had shifted to optically scanned paper ballots. Two sorts of scanners were used: Sequoia-Pacific's OpTech and Global's (now Diebold's) Accuvote.⁴

Sixty races were recounted in the general and primary statewide elections of 2000, 2002, and 2004.⁵ The majority of the recounts (47) were for state representative races. Recounts were also conducted for state senate (7) and county commissioner (5) races as well as a single recount for a county attorney contest. In our study, data on the recount are reported for the town and sometimes ward level. Because each race typically involves many towns, the 60 recounted races yield 1,331 cases—where a case is a recounted vote total for each candidate in each town in the district. Again, we distinguished between the average percent difference and the percent of all ballots cast (regardless of the size of jurisdiction).

Our analysis here considers the changes between the initial counts and recounts for each method of vote tabulation. Again, we computed a weighted mean of the average absolute percent difference between the initial count and the recount. As a result, we present the data with and without two exceptional cases. In the 2000 general election, a faulty machine missed nearly 2,000 votes in the third ward of the town of Dover in a Strafford County commissioner's race—This error was discovered and corrected in the recount and it did not affect the outcome of the race. Likewise, a recount in the town of Bradford yielded tabulations differing from 15 to 20 percent.⁶

Table 12.1 presents the invalidation rates of hand-counted and scanned ballots in New Hampshire for the 2000, 2002, and 2004 general and

Table 12.1 Invalidation rates of hand-counted and scanned ballots: New Hampshire 2000, 2002, 2004 general and primary elections

	N	Weighted average	Median	Standard error	Maximum
Hand counted	464	1.98	0.83	0.14	42.11
Handed counted (without Bradford)	445	1.72	0.78	0.11	42.11
Machine counted	892	0.95	0.50	0.10	32.80
Machine counted (without Dover, Ward 3)	886	0.68	0.50	0.04	14.03
Accuvote	776	0.98	0.50	0.04	32.80
Accuvote (without Dover Ward 3)	770	0.68	0.50	0.04	14.03
OpTech	116	0.69	0.53	0.09	8.04

primary elections. Here, the invalidation rate for hand-counted paper ballots was 1.72 percent—higher than the historical average. The percentage difference for optically scanned paper was 0.68 percent, approximately 1 percent and significantly lower than the hand-counted paper.

The differences found in the table may owe to the type of equipment used for the election or the size of the community. Smaller communities are more likely to have a large discrepancy in their tabulations and they are more likely to use paper ballots. We untangled these effects by controlling for the office at stake and the size of the vote cast. Moreover, we performed two regression analyses to estimate the effect of Scan Ballot, which equals one if the town used optically scanned ballots and zero if the town used hand-counted paper ballots. We controlled for total votes cast in a town and the initial count for a candidate. We also controlled for the office at stake, as some elections spanned several towns. This allowed us to hold constant the candidates on the ballot. Table 12.2 presents two analyses, one of which removes the effect of office (see column labeled Fixed Effects) and one does not (see column labeled OLS). We present both as a robustness check.

We controlled for total votes cast in a town and the initial count for a candidate. We also controlled for the office at stake, the year of the election, and the type of election. This allowed us to hold constant the candidates on the ballot and control for any idiosyncratic attributes. Table 12.2 presents two analyses, one of which removes the effect of office (see column B) and one does not (see column A). We present both as a robustness check.

The key coefficient of interest is the effect of Scan Ballot. The coefficient in both specifications is approximately -0.8 , which means that

Table 12.2 Estimate effect of machine v. hand count on percentage change in votes counted: 2000–2004

	Least squares	Fixed effects for office
	b (S.E.)	b (S.E.)
Constant	1.828(0.108)	1.477(0.138)
Scan ballot	-0.820(0.137)	-0.777(0.138)
Town vote (1000s)	0.002(0.006)	-0.003(0.006)
Initial candidate vote (1000s)	-0.013(0.004)	0.027(0.017)

the percentage difference between the initial count and the recount was approximately eight-tenths of 1 percentage point lower in towns that used scanners than in towns that used hand-counted paper, holding constant the initial count, the total vote, and the office sought. These coefficients are slightly smaller than the observed difference between the mean for hand-counted and machine-counted ballots (which is $1.72 - 0.68 = 1.01$).

To see the practical effect of these estimates, consider an election with 10,000 votes where the candidate in question received exactly half of the votes.⁷ The predicted discrepancy between counts is 1.58 percent if the town tabulates by hand (i.e., $1.477 - 0.003 \times 10 + 0.027 \times 5 = 1.582$) and 0.81 percent if the town uses an optical scanner (i.e., $1.477 - 0.777 - 0.003 \times 10 + 0.027 \times 5 = 0.805$).

CONCLUSIONS

Recounts provide information on the reliability of voting tabulation methods. In recounts of paper ballots, election officials and judges make the best attempt to resolve voter intentions. The discrepancy between the initial (and preliminary) count and the recount captures the degree to which a tabulation method incorrectly tabulates the votes. Estimated validation rates from recounts supplements other metrics, especially residual votes. Residual votes capture a wide variety of ways that voting systems fail to capture the voter's intentions. These failures are prompted by voter confusion and errors in marking ballots, mishaps in handling ballots, machine failures, and tabulator errors.

As an example of how this information can be used to refine our understanding of the sources and causes of errors in voting, we must consider what percent of the residual vote might be tabulation error. We can calculate this using data on the residual vote in presidential races in New Hampshire from 1988 to 2000. In the counties and towns using only

paper ballots, the residual vote rate was 2.3 percent; in the towns using optical scanning, the residual vote rate was 1.7 percent. Using the data on the validation rate from the historical record and from the 2002 elections, the tabulation invalidation rate was 1.72 percent for paper and 0.68 percent for optical scanning. Between roughly two-fifths and as much as three-fourths of the residual vote, then, was pure tabulation error. The remainder consisted of either unrecoverable ballots (i.e., people who accidentally voted twice) or blank ballots.

While stepping back from the details of the data, New Hampshire's recounts speak directly to two important questions in election administration.

Have we made progress? Do machine counts improve on hand counts? At least in the comparison of optical scanning and paper, the answer is yes. Historically, there is about a 1 percent difference between initial counts and recounts when ballots are tabulated by hand. The discrepancy between initial counts and recounts falls to about 0.5 percent with the optically scanned ballots. Looking at recounts from 2002, controlling for total vote, initial vote, and office-sought, optically scanned paper produced a lower discrepancy between the initial count and the recount compared to hand-counted paper.

Considering these tabulation errors, how confident should we be in vote counts, and when should we have a recount? The tabulation invalidation rate in New Hampshire from 2000–2004 was low, especially for optical scanning. However, it was not trivial. In a US House election with 250,000 votes, the invalidation rate of 0.005 for scanners amounts to 1,250 votes. If the race is close, say split 50–50, then the tabulation might affect the margin between the candidates by as much as 1 percentage point of the vote (i.e., if 0.5 percent of the votes for candidate A went to candidate B and candidate B lost none of her votes).

Most states have no set standard for an automatic recount; the courts or state election officer decides whether a recount is appropriate. Some states, however, have provisions for automatic recounts—typically, if the election is closer than one-half of 1 percent of ballots cast (National Commission on Federal Election Reform, pp. 343–346). Yet, this standard appears inadequate. An automatic recount standard, such as 0.5 percent, should be treated as a minimum threshold, rather than an absolute standard for recounts. Random tabulation errors on average change the votes received by a candidate 0.5–1 percent. In a large number of races, then, recounts may be justified. The political problem in most states, though, is that discretion over when to have a recount lies in the hands of election officials, who may have a political stake in the outcome.

NOTES

Correspondence should be directed to Stephen Ansolabehere-email: sdamic.edu. Mailing address: Department of Political Science, MIT, Building E53-457, Cambridge, MA 02139. We wish to thank the Carnegie Foundation and the Knight Foundation for their support of this research and Anthony Stevens for his helpful comments.

1. The NORC coding of the ballots found that the most common problem with optical-scanned ballots involved people voting for a candidate, say Bush, and then writing in that candidate's name in the columns labeled "Write In." The machine tabulator failed to count these ballots, though voter intention was easily resolved.
2. See Caltech/MIT Voting Technology Project Report, no. 1 (2001) and Ansolabehere and Stewart (2005) for further definition and discussion of residual votes. Studies using similar measures include whose results are consistent with the Caltech/MIT study and who find similar problems with punch cards but suggest somewhat better performance for electronic equipment.
3. We wish to thank New Hampshire assistant secretary of state Anthony Stevens for providing us with this information.
4. Voting machinery of New Hampshire towns obtained from <http://www.nh.gov/sos/voting%20machines.htm>, accessed May 22, 2002, and March 1, 2006.
5. We exclude the recounted race for District 1 Executive Council in 2004 in the Democratic primary election because one of the candidates received only write-in votes.
6. We have confirmed this irregularity with the New Hampshire secretary of states office—They are looking into the matter.
7. We caution against making out-of-sample predictions. The largest community in 2002 was Manchester, which recorded approximately 25,000 votes. The effect of town vote likely tends to an asymptote and the linear specification used cannot capture the predicted discrepancy for larger communities.

REFERENCE

Stephen Ansolabehere and Charles S. Stewart (2005), "Residual Votes Attributable to Technology," *Journal of Politics*, 67: 365–389.

CONCLUSION: THE FUTURE OF ELECTION AUDITING

R. MICHAEL ALVAREZ, LONNA RAE,
ATKESON, AND THAD E. HALL

OUR COLLECTION OF ESSAYS PROVIDES AN OVERVIEW of the issues associated with auditing elections in the United States. It builds on discussions and papers that were presented at a workshop that we organized in Salt Lake City, Utah, from February 9th through the 10th in 2009. At the conclusion of the conference, we held a brainstorming session to discuss the key issues related to auditing elections. In this concluding chapter, then, we present the important points raised in this brainstorming session. We believe these points appropriately encapsulate a number of important future issues regarding election auditing.

UNDERSTANDING AN AUDIT

There is a difference between election audits and other specific activities that are part of the general election process: These other activities include reconciliation, canvassing, and recounts. Ballot reconciliation—comparing the total number of votes cast against the total number of ballots cast and the total number of voters who signed in to vote—is a critical part of an audit and a critical component of the overall ballot-counting process. Likewise, canvassing—which ensures that only the correct ballots were included in the vote totals and no proper ballots were excluded—is a critical part of the election process. Nevertheless, it does not constitute election auditing.

Finally, a recount is a postelection activity that can be a part of an audit (it is in some countries) or not. However, it is also a critical postelection activity that candidates, parties, and others engage in to determine if the