From: SoS Rulemaking
To: SoS Rulemaking
Subject: FW: Objection to all rules pertaining to RLA as unnecessary and inferior to current methods
Date: Wednesday, July 5, 2017 12:00:14 PM
Attachments: Exhibit A.PDF
Exhibit B.PDF
Exhibit C.PDF
Exhibit D.PDF
Exhibit E.PDF
Exhibit F.PDF
Exec Sum - Final.pdf

From: Merlin Klotz [mailto:mklotz@douglas.co.us]
Sent: Wednesday, July 5, 2017 10:56 AM
To: Wayne Williams <Wayne.Williams@SOS.STATE.CO.US>; Judd Choate <Judd.Choate@SOS.STATE.CO.US>; Dwight Shellman <Dwight.Shellman@SOS.STATE.CO.US>
Cc: amyers@larimer.org; Pam Bacon <baconp@logancountyco.gov>; 'Beth Zilla' <beth.zilla@phillipscounty.co>; 'Beverly Wenger' <bwenger@co.yuma.co.us>; boots.campbell@rbc.us; Krystal Brown <BrownK@co.teller.co.us>; Christy Beckman <cbeckman@sedgwickcountygov.net>; cgomez@saguachecounty-co.gov; 'Chuck Broerman' <chuckbroerman@elpasoco.com>; ckoppes@weldgov.com; clengel@lincolncountyco.us; clerk@co.routt.co.us; clerk@hinsdalecountycolorado.us; Sherryl Steving <clerk@riograndecounty.org>; Ladonna Jaramillo <clerk@sanjuancountycolorado.us>; Colleen Stewart <gcclerk@co.gilpin.co.us>; custerclerk@custercountygov.com; dallas.schroeder@elbertcounty-co.gov; 'Debra Johnson' <debra.johnson@denvergov.org>; delisa.weeks@state.co.us; dgreen@parkco.us; La Rita Randolph <dolorescounty@hotmail.com>; 'Faye Griffin' <fgriffin@jeffco.us>; Garland Wahl <gwahl@co.washington.co.us>; 'Gilbert (Bo) Ortiz' <ortiz@co.pueblo.co.us>; 'Hillary Hall' <hhall@bouldercounty.org>; jalberico@garfield-county.com; janice.vos@pitkincounty.com; Hayle Johnson <jc_clerk@hotmail.com>; jcoen@prowerscounty.net; 'Jim Candelarie' <jcandelarie@broomfield.org>; June Madrid <jmadrid@archuletacounty.org>; Karen Garcia <karen.garcia@costillacounty.net>; kathleene@sanmiguelcountyco.gov; Kathy Neel <kathyn@co.summit.co.us>; Katie Barr <katie.barr@fremontco.com>; Kim Percell <kpercell@co.montezuma.co.us>; Kathy Simillion <ksimillion@gunnisoncounty.org>; Lawrence Gallegos <lawrence.gallegos@co.conejos.co.us>; Lila Herod <lherod@moffatcounty.net>; Lucile Nichols <lnichols@crowleycounty.net>; Lori Mitchell <lmitchell@chaffeecounty.org>; 'Matthew Crane' <mcrane@arapahoegov.com>; Merlin Klotz <mklotz@douglas.co.us>; mineralcountyclerk@hotmail.com; mnauer@ouraycountyco.gov; mwoodward@alamosacounty.org; Nancy Cruz <nancy@huerfano.us>; 'Pam Anderson' <cccaexecutivedirector@gmail.com>; 'Patricia Daughtery' <cheyclerk@gmail.com>; Patti Nickell <patti.nickell@bentcounty.net>; Patty Berger <pberger@co.lake.co.us>; peach.vigil@lasanimascounty.org; Pam Phipps <pphipps@co.clear-creek.co.us>; Regina OBrien <regina.obrien@eaglecounty.us>; sbailey@co.morgan.co.us; Sharon Sisnroy <ssisnroy@oterogov.org>; sheila.reiner@mesacounty.us; SMartin@adcogov.org; Sara Rosene <srosene@co.grand.co.us>; 'Susan Corliss' <susan.corliss@kitcarsoncounty.org>; tguynes@montrosecounty.net; Tiffany Parker <Tiffany.Parker@co.laplata.co.us>; tstephenson@deltacounty.com; Brett Mers <bmers@douglas.co.us>
Subject: Objection to all rules pertaining to RLA as unnecessary and inferior to current methods

To: Secretary Williams and all my fellow Clerks:

I attended most of the RLA discussion group meetings. I was struck by the blind faith in Professor Stark's theory and the objection be damned, concentrated effort to implement it, despite the fact that in a white paper Dr. Stark noted it was nearly impossible to apply in a central scan environment.

It was our calculation that if the Stark method was applied on all races, in Douglas County we could find it necessary to pull as many as 300% of the batches in order to randomly locate enough selected ballots in all 50 ballot issues. The committee seemed to recognize this problem and opted to only Stark test 1 statewide race, if there was one. How could ignoring all other races provide confidence in any one of them?

The target seemed to be a 95% confidence index with an interval of 3. Since it could be assumed that every vote that was cast for Trump would have otherwise been voted for Clinton, a voting tabulation error for one would likely have gone to the other or an interval of 3 would equate to a 6 point differential. Since a 6 point differential was the final Colorado difference between the two candidates in my mind using the professor's methodology it was only safe to say that we were 95% confident that one of them won the election in Colorado. I'd not expect such a low level of confidence to be accepted favorably with the voters of Colorado.

In Douglas, we began to discuss a question that was never raised in the working group of what was the confidence index on the current audit protocol. Intuitively I expected it to be similar so we turned to an operational statistics professor at the Air Force Academy and contracted with him to compare and contrast the confidence index of both the Current and Professor Stark audit protocols.

To our surprise the professor determined that the current post election audit protocol yielded a confidence index well in excess of 99%.

Before changing election processes on the whim of a few activists and a Berkley theoretician, it is only prudent to quantify the current protocol as a benchmark to establish that the proposed change adds value. This critical step appears to not have been taken until Douglas County stepped forward and contracted for such a comparison.

The attached detail substantiates that application of the Stark theory by rule should be abandoned as we already comply with the statutory RLA definition of using statistical methods in the audit protocol.

I don't know about your voters and the press, but I have no interest in trying to convince either that I used to be 99% confident that the correct person won the race but now with a lot more labor and cost I could only assure them that I am 95% confident in the posted outcome!

Please read attached documents and comment on the proposed RLA rule(s) and ask the Secretary of State to abandon this inferior concept.

Merlin Klotz
Douglas County Clerk and Recorder

Regarding Imposition of Berkley Theoretical Audit Process

Executive Summary

Introduction

I request that proposed rules requiring implementation of the Dr. Stark theory (Berkley Theory), Proposed Rule 1.1.10 and 25, be eliminated in their entirety as unnecessary, inferior to and costlier to execute than existing processes. Further and for the same reasons, I request that amendments to Rule 11.3 be rescinded as well.

CRS: 1-7-515 (5) (b) "Risk-limiting audit" means an audit protocol that makes use of statistical methods and is designed to limit to acceptable levels the risk of certifying a preliminary election outcome that constitutes an incorrect outcome.

I. The current Colorado SOS Audit Protocol (CSAP) meets the statutory definition and requirements of "Risk-limiting audit".

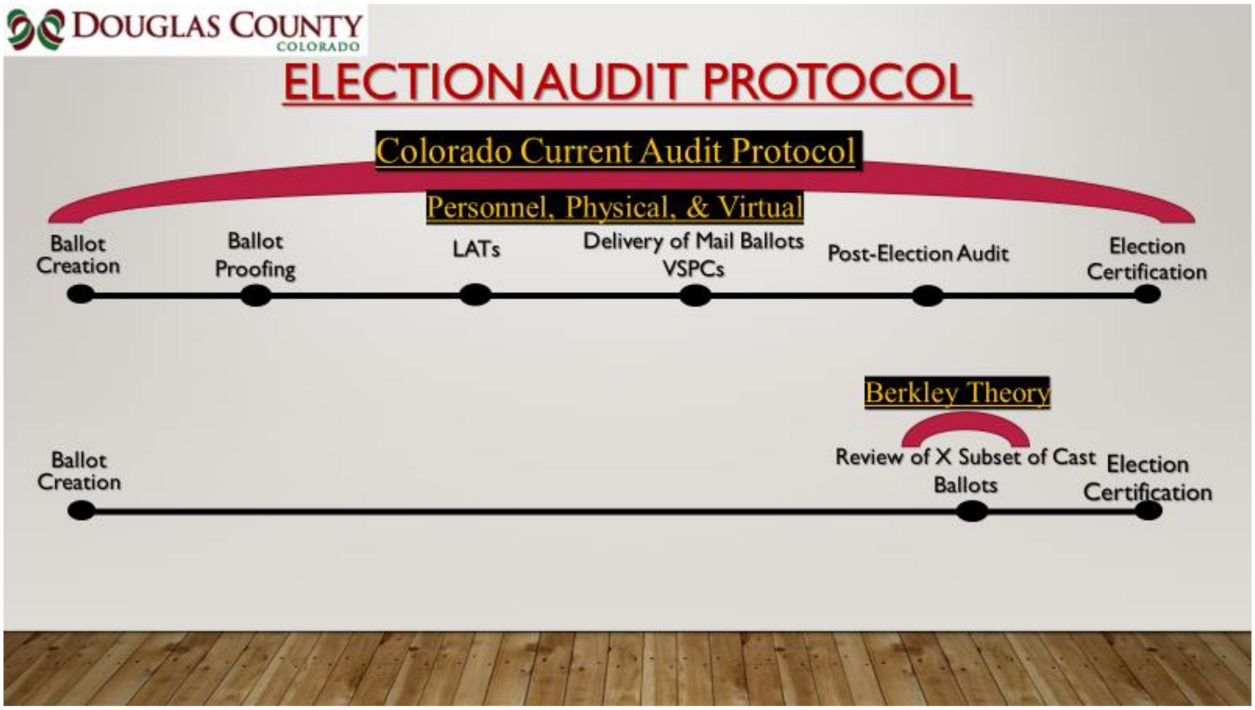
A. Implementation of the Berkley Theory (AKA: RLA by Colorado) is nothing more than chasing a shiny object with a romantically deceiving name. From multiple perspectives, the Berkley Theory by Dr. Stark is inferior to and provides no statistical benefit to the small element of the current audit protocol that it addresses. Further, the current audit protocol complies with the definition of Risk Limiting audit in CRS: 1-7-515 and no change in rule is required to comply with the 2017 RLA deadline. Statistical methods are currently used in selection of batches for testing in the Post-Election Audit.

B. The Berkley Theory, attempts to utilize a method not suited for multiple style, central scan election processes, as utilized in Colorado, to select and test electronically reported data to actual ballots. The inapplicability of applying this theory to Colorado is evidenced in the RLA working group's default decision to limit this experiment to a single statewide contest. Instead, it should be eliminated from rules entirely.

C. The attached exhibits consider the issue from Legal, Statistical and Operational viewpoints. In all cases, the Dr. Stark concept not only offers no advantage but produces inferior results with higher operational cost and time demands than the current audit protocol.

D. As a CPA, I have designed and implemented many audit protocols. An "audit protocol" includes consideration of the entirety of processes including software, physical security, technological security, quantitative tests and random statistical sampling of transactions. It is the entirety of these elements that provides confidence in the process rather than reliance on an isolated element, such as testing of selected transactions that limits to acceptable levels the risk of certifying an incorrect outcome. For example, successful testing of cash transactions from purchase order to payment is a failed complement to the audit process if a bank reconciliation would reveal serious issues.

CSAP covers the entire Election Process:



D. The Berkley Theory is analogous to calculation of the probability that your house will be broken into based on testing of the front door lock irrespective of whether the windows are open or other variables. In contrast, the totality of the Colorado election protocol is analogous to checks not just on whether all doors and windows are locked, but that the house is surrounded with 10-foot chain link fence topped with electrified razor wire that is surrounded by a moat filled with piranha. In Colorado, this theory is meaningless and adds nothing.

II. Berkely Theory Based on Faulty Assumption

A. "Computer software cannot be guaranteed to be perfect or secure, so voting systems should be software independent:" (Lindeman and Stark, A Gentle Introduction to Risk-limiting Audits, pg. 1), Exhibit E).

a) This is a straw-man argument that does not take into account any potential for human error in counting and calculating very large number sets.

b) Practical experience by every Clerk and Recorder who has performed a Post-Election Audit substantiates that multiple human counts are frequently required on a universe as small as 100 ballots to get the same answer. In contrast, I never recall receiving multiple answers from ballot scans of the same data.

B. Levy (MIT) and Murmane (Harvard) noted two conditions that must be met for a task to be better suited to a computer than a human. They are:

a) An Information Condition: All information necessary to carry out the task can be identified and acquired in a form that computers can process.

b) A Processing Condition: The information processing itself can be expressed in rules (Levy, Frank, Murnane, Richard J. 2013. *Dancing with Robots, Human Skills for Computerized Work*, pg. 7, Retrieved from http://dusp.mit.edu/uis/publication/dancing-robots-human-skills-computerized-work).

C. Colorado SOS Training notes that hand tallying votes is a less certain method for counting large sets. Regarding the procedures required for the Logic & Accuracy Testing required by Colorado statute, the computerized total is not assumed as the primary source of error. "If the totals don't match, re-check the hand tallies. Often, human error can account for the discrepancy" (Section 1-7-509(4), C.R.S.; Election Rule 11).

This is an accurate description of the sort of work involved in tabulating totals in large data sets. The key to accuracy in such protocol is to test and ensure to the greatest extent possible, that the tabulation systems are configured and operating correctly.

III. Berkley Theory Fails:

To date, we are unaware of any statistical or operational comparison, testing or analysis that has been done between the Berkley Theory as mandated by proposed Rule 25 and the CSAP in its current form and effect other than what has been done by Douglas County and is documented in this package.

A. Legally – The CSAP is accomplished within statutory timelines and obligated funds. The Berkley Theory may likely have timeline requirements that will push past statutory guidelines.

B. Statistically – The CSAP uses an applicable statistical process to ensure a 99% confidence rating that no ballot has been miscounted (Horton, Review of Douglas County Elections Process, pg. 1). The Berkley Theory uses a less applicable process to one portion of the elections process and recommends a 5% risk-limit.

C. Operationally: The CSAP utilizes relevant statistical methodology to meet all statutory requirements within statutory timelines. The Berkley Theory will extend work hours into the weekend prior to Thanksgiving and may very well extend into the Thanksgiving holiday making it difficult to obtain participants in Post-Election Audits and Canvass. Additionally, it will require labor expenditures beyond currently obligated funds.

IV. Legally

A. The current CSAP allows for the Post-Election Audit and Canvass to be conducted within the statutory timeline of 17 days (1-10-103; 1-1-106(5). The Berkley Theory will likely require greater timelines than allowed within Colorado statutory guidance (Douglas County Time Study Assumptions, Exhibit A).

V. Statistically

A. The Berkley Theory assumes some sort of homogeneity in the data of which the relevance is not readily apparent. The fact that X subset of the set of cast-votes matches the proportions of the entire set of cast-votes does not seem clearly illustrative of accuracy as X + 1 subset may or may not match said proportions (RLA Options, Slide 7, Exhibit D).

B. A particular process to generate randomness is required for the Berkley Theory. Randomness is defined as lacking a particular purpose or pattern (Merriam-Webster, RLA Options, Slide 10 Exhibit D). Therefore, any set is both random and ordered at the same time, i.e., random by one measure and ordered by another. In order to determine randomness, one must first define the purpose or pattern against which it is being judged. Given a standard by which to judge against, it is random or not. "A simple random sample of ballots would be difficult and time-consuming to collect, while adding no significant benefit to the statistical process" (Horton, op cit, pg. 3). Using the process described by the Berkley Theory will increase labor costs and timelines while not providing any definitive added value.

VI. Operationally

A. The Berkley Theory may add significant labor costs and time to the period allowed for Post-Election Audit and Canvass and may not allow for timely completion of tasks required prior to Certification (Douglas County Time Study, Exhibit A).

B. Based on Colorado statutory requirements and timelines, there will be approximately 3 days to conduct the Berkley Theory. Given the possible recommendation of number of ballots to review via the Berkley Theory, the time currently allotted may not be nearly enough (Douglas County Time-Study Assumptions, Exhibit F).

Conclusion

A. The Berkley Process is not a process that is readily applicable or statistically beneficial to Colorado's election activities. This is stated by the authors. "Some jurisdictions' heavy use of vote-by-mail ballots can complicate batch-level audits"
(Bretschneider, Jennie, Flaherty, Sean, Goodman, Susannah, Halvorson, Mark, Johnston, Roger, Lindeman, Mark, Rivest, Ronald L., Smith, Pam, Stark, Philip B. Oct 2012. *Risk-Limiting Post-Election Audits: Why and How*, pg. 26, Retrieved from https://www.stat.berkeley.edu/~stark/Preprints/RLAwhitepaper12.pdf).

B. The current CSAP is multilayered and multifaceted covering requirements and potential challenges from the personnel, physical, and virtual realms. This Protocol has been described by a local banker as being more secure than a bank vault (RLA Options, Slide 24, Exhibit D). By following the guidelines of Colorado statute, the current Protocol provides the voters of Colorado with 99% confidence that no ballot has been miscounted.

C. Being the first to use the Berkley Theory in an election cycle may well make Colorado the first to increase labor costs, fail statutory certification timelines and aggravate rather than mitigate any public concerns by reduction of the perceived confidence that we certified correct election results from the current 99% to the Berkley Theory target of 95%.

D. I request that proposed rules requiring implementation of the Dr. Stark theory (Berkley Theory), Proposed Rule 1.1.10 and 25, be eliminated in their entirety as unnecessary, inferior to and costlier to execute than existing processes. Further and for the same reasons, I request that amendments to Rule 11.3 be rescinded as well.

Merlin Klotz, CPA
In Consultation with:
Dr. Brett Mers, DM in Global Management
Dr. Kenneth Horton, PhD, Biostatistics
Mr. Kyle Rulli, Logistics & Technology Supervisor,
Douglas County Elections

DOUGLAS COUNTY COLORADO

Comparison Test of Amendment 72 - 2016

Colorado SOS Audit Protocol

% of Votes	95%	99%
51%	298	457
52%	149	228
53%	99	151
54%	74	113
55%	59	90
56%	49	75
57%	42	64
58%	36	56
59%	32	49
60%	29	44

CSAP – 99% Confidence Rating of No Incorrectly Counted Votes

Berkley Theory

Amendment 72 (CONSTITUTIONAL), Vote For 1

Yes/For	89,864	48.27%
No/Against	96,317	51.73%
Cast Votes:	186,181	96.72%

Initial sample size

Contest information

Ballots cast in all contests: 192617 Smallest margin (votes): 6,453. Diluted margin: 3.35%.

Contest 1. Contest name: Amendment 72

Winners: 1

Reported votes:

Candidate 1 Name: Yes/For Votes: 89864

Candidate 2 Name: No/Against Votes: 96317

Add candidate to contest 1 | Remove last candidate from contest 1

Add contest | Remove last contest

Audit parameters

Risk limit: 5% Expected sample size: 5,368.

DOUGLAS COUNTY
COLORADO

Comparison Test of Amendment 72 - 2016

Sample Race from 2016 Election
-- Assuming Audit of 5,368 as Noted by the Berkley Theory Calculator

To Meet Statutory Timeline (Need Accomplish w/in ~3 Days)
- -- Total Judges - 32
- -- Total Staff - 20
- -- Total Work Hours (Prep + RLA)
  --- 1,187
- -- Total Work Days - 2.85
- -- Total Cost - $17,137.48

Sample Race from 2016 Election

Amendment 72 (CONSTITUTIONAL), Vote For 1

Yes/For	89,864	48.27%
No/Against	96,317	51.73%
Cast Votes:	186,181	96.72%

Initial sample size

Contest information

Ballots cast in all contests: 192617 Smallest margin (votes): 6,453. Diluted margin: 3.35%.

Contest 1. Contest name: Amendment 72
Winners: 1

Reported votes:

Candidate 1 Name: Yes/For Votes: 89864
Candidate 2 Name: No/Against Votes: 96317

Add candidate to contest 1 | Remove last candidate from contest 1

Add contest | Remove last contest

Audit parameters

Risk limit: 5% Expected sample size: 5,368.

CSAP – 99% Confidence Rating of No Incorrectly Counted Votes in Noted Race

DOUGLAS COUNTY
COLORADO

Comparison Test of Amendment 72 - 2016

Sample Race from 2016 Election
-- Assuming Audit of 5,368 Ballots as Noted by Berkley Theory Calculator

To Limit Stress on Workforce and Logistic Requirement (Likely Not Meet Statutory Timeline)

-- Total Judges - 12
-- Total Staff - 20
-- Total Work Hours (Prep + RLA)
--- 1,583
-- Total Work Days – 7.61
-- Total Cost - $22,464.66

Sample Race from 2016 Election

Amendment 72 (CONSTITUTIONAL), Vote For 1

Yes/For	89,864	48.27%
No/Against	96,317	51.73%
Cast Votes:	186,181	96.72%

Initial sample size

Contest information

Ballots cast in all contests: 192617 Smallest margin (votes): 6,453. Diluted margin: 3.35%.

Contest 1. Contest name: Amendment 72
Winners: 1

Reported votes:

Candidate 1 Name: Yes/For Votes: 89864
Candidate 2 Name: No/Against Votes: 96317

Add candidate to contest 1 | Remove last candidate from contest 1

Add contest | Remove last contest

Audit parameters

Risk limit: 5% Expected sample size: 5,368.

CSAP – 99% Confidence Rating of No Incorrectly Counted Votes in Noted Race

MEMORANDUM FOR Douglas County (DC) Clerk & Recorder

Subject: Question & Comment Regarding the Risk Limiting Audit (RLA) Process as Described by Mark Lindeman and Professor Philip B. Stark

1. Assumption of Homogeneity in the Data

The RLA process as described is conducted on the cast ballots until an acceptable level of certainty is achieved. This is accomplished by reviewing random ballots until the sample approximates the election results (to within defined parameters) as noted by the Voting System (VS). This assumes a homogeneity of the data at X level (represented by the selected sample) which may or may not be valid. Additionally, the presence or absence of such homogeneity at X level neither supports nor detracts from the election results as tallied by the VS considering the totality of the ballots cast as opposed to a potentially representative sample.

2. Assumption of Value of Generated versus Actual Randomness to Process

Random versus ordered is an analytical criterion based on a selected standard. In other words, what may be random by one measure is ordered by another. The members of a high-school basketball team standing in a row may be random by height but ordered by grade-point-average. So, the first question that needs to be asked is "random by what standard and for what purpose"?

Ballots are returned to Elections for processing by a variety of means. Some are placed in one of 9 Ballot Collection Boxes placed across the county. Some are sent in the mail. Some are dropped off or marked in one of 6 Voter Service Polling Centers also spread across the county. These ballots are not ordered in any relevant manner or fashion other than that they are all from Douglas County legally registered voters. They come into the facility in random fashion based on multiple factors of everyday life. They come from various areas of the county. They are dropped off at all hours of the day, by people of all genders, races, religions, ages, income levels, and voting perspectives. The order and randomness inherent in the ballots extant is relevant to understanding the requirements and outcome of a fair and impartial election.

What is the evidence to support the contention that the randomness extant and inherent in the cast ballots is less relevant to the process than that generated by a seed and a pseudo-random number generator (PRNG)? The added labor costs of using the seed and PRNG may not be justified by the value-added.

3. Assumption of Fallacy in the Current Audit Process

The Internal Logic & Accuracy Test (I-LAT) is run by DC Elections personnel. This test randomly marks practice ballots for entry into the VS prior to the election. The results are then hand tabulated to ensure accuracy of throughput. This test is the re-done in the Public Logic & Accuracy Test (P-LAT). The election is then conducted and valid ballots from registered voters are entered, processed, and tabulated. Following this, a Post-Election Audit (PEA) is conducted under that same procedures as that of the I-LAT and P-LAT on 500 randomly selected ballots from the randomly ordered body of the cast ballots. This is to detect any processing error in the system that may have been missed.

In order for an entity to tamper with the VS tabulations and not have it noted in the PEA, one must first tamper with the Voting & Tabulation Machines (VTM) located in the basement of the DC Elections building after the P-LAT and secondly, go back and change anything adjusted back to the original configuration prior the PEA. There are multiple security procedures and protocols to secure the VTM and process and prevent such an event from happening.

Personnel: Personnel are trained and maintain appropriate chain-of-custody procedures when dealing with cast ballots from the time of receipt through the entire process.

Physical: The VTM are kept in the basement of the DC Elections Facility behind 3 locked doors. This area is open only to DC Elections staff. All others entering the area are under escort of the same. The area is under 24-hour camera surveillance and real-time monitoring when the building is closed.

Virtual: The VTM are networked on a closed system. In order for someone to achieve unauthorized electronic entry into the VTM, he or she would have to be physically in the Counting Room. The likelihood of which is mitigated by the physical security measures noted above. Additionally, the electronic security measures are extensive.

The VTM resides on a closed network and cannot access either the Internet nor the Douglas County Network. It is managed by a server which handles the antivirus definitions for all computers on the network. The computers require a Windows logon of which, there are various types of administrator and user accounts. The server, ballot creation, and tabulation computers are only accessible by the Elections Deputy and the Logistics & Technology Supervisor. Once inside of the computer, the Hart Voting System requires another logon which is administered by the system administrator, and then requires an encrypted USB key and another password to access any type of election data.

Finally, assuming the someone was able to successfully bypass all of the noted security measures, the VTM maintains logs of every keystroke and all activity on the system. These logs are checked regularly and automatically by system and DC Elections personnel.

Tampering with the VTM and having the activity go unnoticed is analogous to someone breaking into a bank secured under lock, key, security camera, alarms, a Quick Reaction Force, and an electronic monitoring system and robbing it immediately after an audit and then breaking back into that same bank and replacing the money, note for note, coin for coin, and spatial location for spatial location prior to the next audit. Simply given the complexity of such an event, the likelihood of occurrence is fairly low.

4. Assumption of the Importance of Secure and Transparent Elections

This is a valid and supportable assumption. Such events are critical to the foundation and continued function of American democratic republic. But secure and transparent elections cost money to administer and execute and utilizing the most efficient means of ensuring the coherency and consistency of such is the fiduciary responsibility of all public officials. The current DC Elections security processes and procedures combine to achieve this objective.

The personnel security checks in hiring and training processes and the chain-of-custody procedures provide a dependable and secure workforce. The I-LAT, P-LAT, and PEA, serve this objective by testing the VTM both immediately prior to and immediately after the actual elections. The combined effect of these activities is to produce a legal, transparent, accurate, timely, and efficient realization of Douglas County voter will.

There seems to be little or no data nor logic to support a contention that when validated water goes into one end of a pipe and validated water comes out of the other end of a pipe and someone you trust is watching the middle of that pipe that water is not what you've got.

Dr. Brett L. Mers
Doctor Management & Global Leadership
Douglas County Deputy Elections
303-503-5809 / bmers@douglas.co.us

Review of Douglas County Elections Process

Dr. Ken Horton

B.S., Mathematical Sciences, United States Air Force Academy
M.S., Statistics, Washington State University
PhD, Biostatistics, University of Colorado, Anschutz Medical Campus

June 16, 2017

Introduction

The purpose of this paper is to evaluate the Post-Election Audit (PEA) performed by the Douglas County Office of the Clerk and Recorder following each election. In a PEA, officials select 500 ballots from those cast (Rule 11.3.3(c)(1) in order to ensure that they were read by the Voting System (VS) properly. This is more than enough to ensure that the VS did not make significant errors that would result in a different election outcome. On the other hand, if the PEA is used to verify the actual outcome of the election, much larger samples would be required for small margins of victory. However, if the validity of the VS has been assured, then such a verification would be unnecessary.

A sample of 500 randomly selected ballots would be sufficient to confirm the validity of the VS counting process. Further, it would be sufficient to verify the outcome of an election, provided the election was won with at least 58% of the vote. Finally, there is no specific aspect of Douglas County's randomization process that would lead to bias in the PEA sample. For added assurance, Douglas County could implement extra randomization mechanisms.

Validity of Counting System

To assess the accuracy of the Voting System (VS), we first need to determine how many ballots would have needed to be read incorrectly to sway the result of an election. From there, using the hypergeometric distribution, we can determine the probability of randomly selecting at least one of those incorrectly read ballots in our sample of 500. It turns out this probability is quite high in most cases. This means that a Post-Election Audit (PEA) of 500 randomly selected ballots would almost certainly identify significant (i.e. election swaying) inaccuracies of the VS.

For example, let's say 50,000 votes were cast in an election and that the VS told us that a candidate garnered 55% of the votes (27,500 for and 22,500 against). Now assume the system was inaccurate and that the candidate garnered no more than half of the votes (25,000 for and 25,000 against). Under this assumption, we know that there would be at least 2,500 incorrectly read ballots. If we randomly select 500 ballots, hand check them, and make sure the system read them correctly, we would have a very high probability (>0.9999) of identifying at least one incorrect. Therefore, if we audit 500 and find no errors, we should be very confident that our system read ballots correctly.

% votes	95%	99%
51%	298	457
52%	149	228
53%	99	151
54%	74	113
55%	59	90
56%	49	75

57%	42	64
58%	36	56
59%	32	49
60%	29	44

Table 1: Sample size required to gain 95% or 99% confidence that the voting system did not make a significant (result-changing) ballot reading error. Results correspond to an election with 50,000 ballots cast.

Verification of Reported Margin of Victory

First, it should be pointed out that if the VS is determined to be valid and accurate, there should be no need to verify the margin of victory with a random sample. If we have quantified the probability of a result-changing error in our VS to be quite low, then any additional verification would be redundant [Rule 11.3.3(e)(1)(1)(3)(4)].

Assuming a result verifying audit is necessary, we need to determine what sample size we need to verify our reported results. Any sample collected will have an associated margin of error. If this margin of error contains the value 50%, we cannot rule out the possibility that our election result is wrong. Larger samples are more reflective of the overall population and provide the benefit of a smaller margin of error. Thus, our goal is to determine what sample size is necessary to minimize the probability of incorrectly concluding that our result could be wrong. This sample size turns out to be quite large.

For example, let's return to our example from the previous question. By randomness, we could select a sample of size 200, with 103 voters supporting our declared winner. Based solely on this sample, one could conclude that a 50-50 result is feasible, leading us to question the results. If we want to be 95% confident that a sample would not lead us to question our results, we would need to gather 1092 random ballots. For 99% confidence, we would need a sample of size 1556.

% votes	95%	99%
51%	27160	39554
52%	6746	9900
53%	3022	4370
54%	1672	2436
55%	1092	1556
56%	746	1090
57%	554	798
58%	414	612
59%	338	490
60%	270	394

Table 2: Sample size required to gain 95% or 99% confidence that the sample would correctly verify the results of the election, for various levels of vote share. Results correspond to an election with 50,000 ballots cast.

Randomization Process

Both of the preceding analyses assume that any sample gathered by the Douglas County Elections Office is random. Under the current sampling system, election officials collect five stacks of ballots at random. Each stack contains 100 ballots [Appendix A]. Presently, there is no specific reason to believe that ballots are stacked in any way that would bias a sample towards a particular result. A simple random sample of ballots would be difficult and time-consuming to collect, while adding no significant benefit to the statistical process.

If Douglas County would like to further ensure randomization, they could take measures to randomize stacking of ballots. For example, ballots could be divided into stacks based on last digit of house number. There is no reason to believe last digit of house number is correlated with any kind of voting behavior. This is just one way to randomize ballots. A simple random sample is not necessary; as long as there is no reason to believe certain types of voters are more likely to be included in a random sample, the current sampling methodology would be sufficient.

Appendix A

Ballots are placed into stacks or batches of 100 for accuracy of handling and are placed into those increments as received into the system. They are organized in no known order.

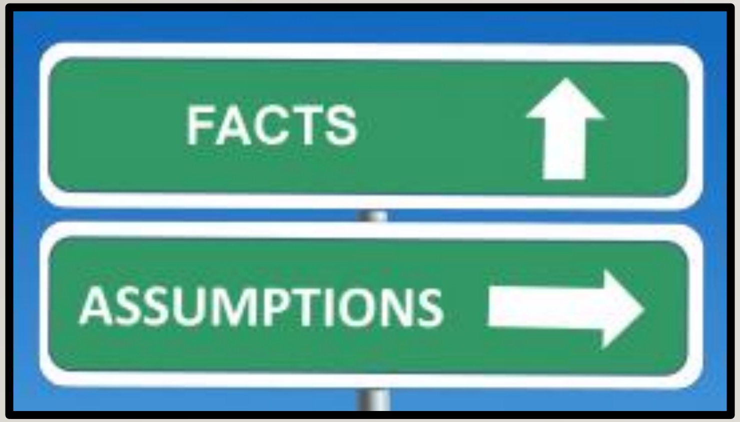
COMPARISON OF RISK LIMITING AUDIT (RLA) OPTIONS

MR. MERLIN KLOTZ
DOUGLAS COUNT CLERK & RECORDER
PRESENTED BY:
DR. BRETT L. MERS
DOUGLAS COUNTY DEPUTY ELECTIONS

DOUGLAS COUNTY COLORADO

Overview

BLUF
Assumptions of Recommended Approach
Data & Rationale for Colorado SOS Audit Protocol (CSAP)
Recommendation



. The same is true with smaller and smaller pieces.

BOTTOM LINE UP FRONT - BLUF

CRS 1-7-515 Risk Limiting Audits – Rules – Legislative Declaration – Definitions

-- (5)(b) "'Risk limiting audit' means an audit protocol that makes use of statistical methods and is designed to limit to acceptable levels the risk of certifying a preliminary election outcome that constitutes an incorrect outcome."

. The same is true with smaller and smaller pieces.

BOTTOM LINE UP FRONT - BLUF

1) The audit protocol requirement covers the initial ballot data selection, all associated security activities, ballot processing, and staff activities through and to include certification of results.

2) The current CSAP uses a statistically-based method to select ballots for review that provides a confidence level between 97% - 99% that no ballot has been miscounted.

Douglas County Colorado

ELECTION AUDIT PROTOCOL

Colorado Current Audit Protocol

Personnel, Physical, & Virtual

Ballot Creation — Ballot Proofing — LATs — Delivery of Mail Ballots VSPCs — Post-Election Audit — Election Certification

Berkley Theory

Ballot Creation — Review of X Subset of Cast Ballots — Election Certification

. The same is true with smaller and smaller pieces.

BERKLEY THEORY ASSUMPTIONS

Holographic-Like Homogeneity in the Cast-Vote Ballot Record

Homogeneity Across Colorado County Election Processes

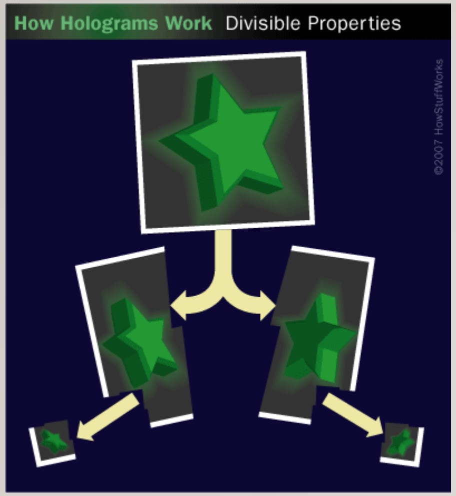
A Particular Randomness is Superior to Another

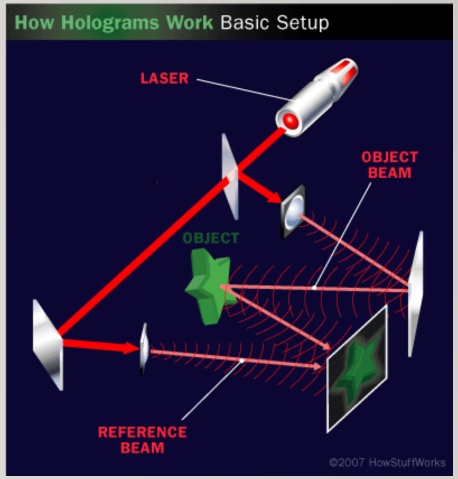
Fallacy in the CSAP as administered by Douglas County

. The same is true with smaller and smaller pieces.

BERKLEY THEORY ASSUMPTION

"Holographic-Like Homogeneity"





Order Implies Intelligence Interaction

Order Implies Purpose

Homogeneity Implies Intelligent Interaction & Purpose

Relevance of the Existence of Such Homogeneity?

e. The same is true with smaller and smaller pieces.

BERKLEY THEORY ASSUMPTION

"Homogeneity Across Colorado County Election Processes"

County Election Processes Produce the Same End

County Election Processes Use Different Means & Methods to Reach that End
-- Based on Population Differences
-- Based on Resource Constraints
-- Based on a Myriad of Other Factors

Comparison Across Counties May Provide Mis-leading Results

DOUGLAS COUNTY COLORADO

CONTEXT – EVERYTHING IS CONTEXT

SOS

Rules & Tools

Registration & Results

Requests for Election Support

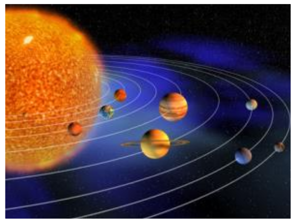
Spec Districts / Municipalities

Complete Election Support

Money & Requirements

Vendors

Equip & Supplies



Douglas County Elections

Rules, Tools, Management, & Money

County Admin

Election Execution and Records

Voter & Ballot Data

EAC

Federal Oversight & Support

Confidentiality & Secure Elections

Personal Info & Votes

Citizens

Outreach Opp's & Facility Use

School Districts

Election Information & Opp

DOUGLAS COUNTY
COLORADO

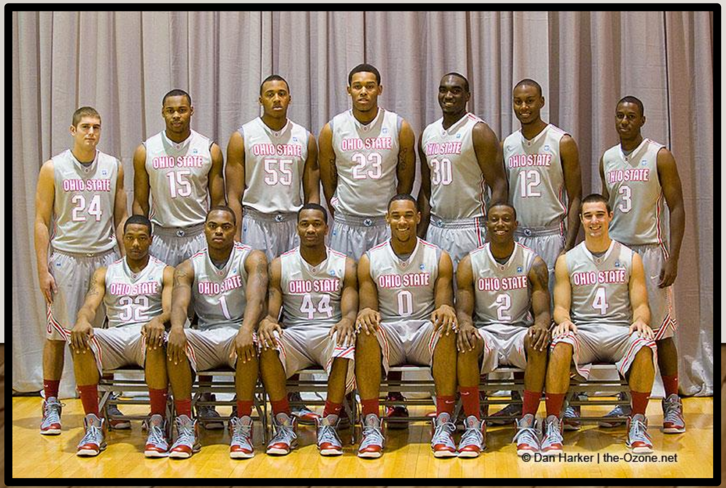
BERKLEY THEORY ASSUMPTION

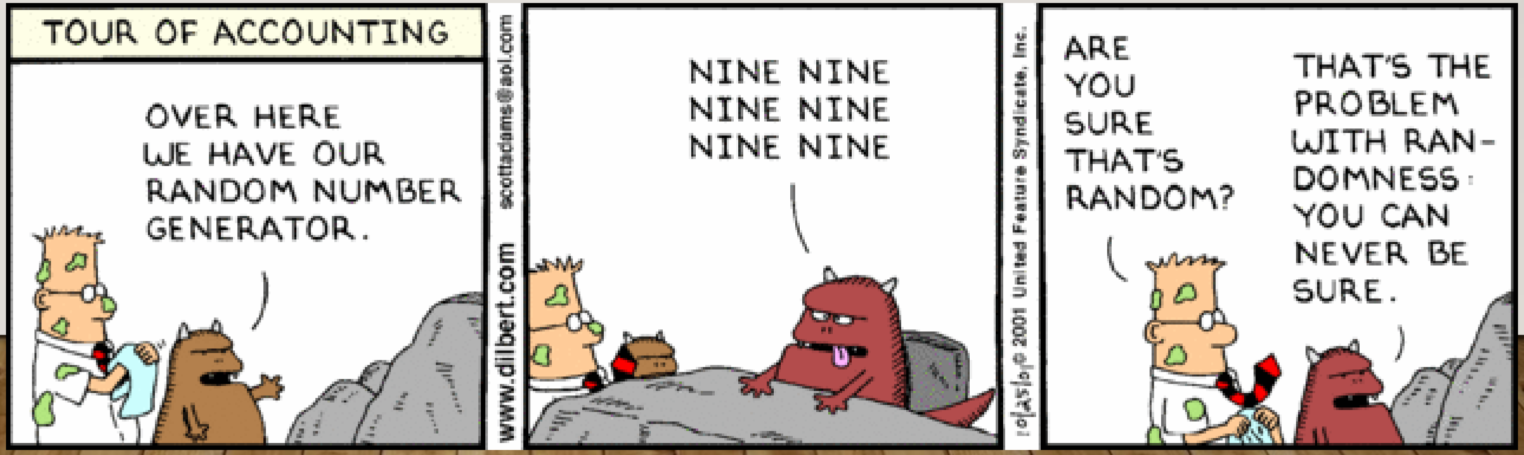
"A Particular Approach to Randomness is Superior"

What is Randomness?

- Lacking a Definite Plan, Purpose, or Pattern (Merriam-Webster.com)

- Random by One Measure – Ordered by Another





BERKLEY THEORY ASSUMPTION

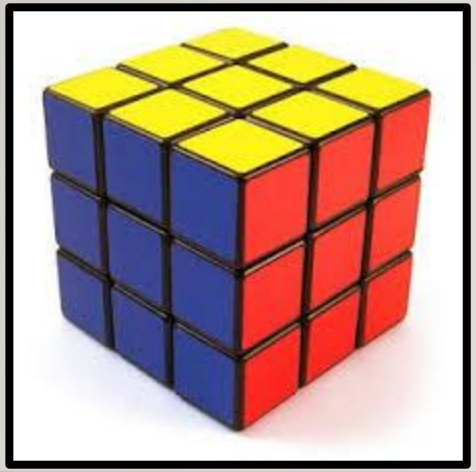
Superior Randomness?

– Lacking a Definite Plan, Purpose, or Pattern (Merriam-Webster.com)

- Random by One Measure – Ordered by Another

- Random or Ordered by Relevant Standards ?

-- Party?	Random / Ordered
-- Precinct?	Random / Ordered
-- Gender?	Random / Ordered
-- Race?	Random / Ordered
-- Zip Code?	Random / Ordered
-- Spec District?	Random / Ordered
-- UNOCAVA?	Random / Ordered
-- VSPC?	Random / Ordered
-- Mail-In?	Random / Ordered

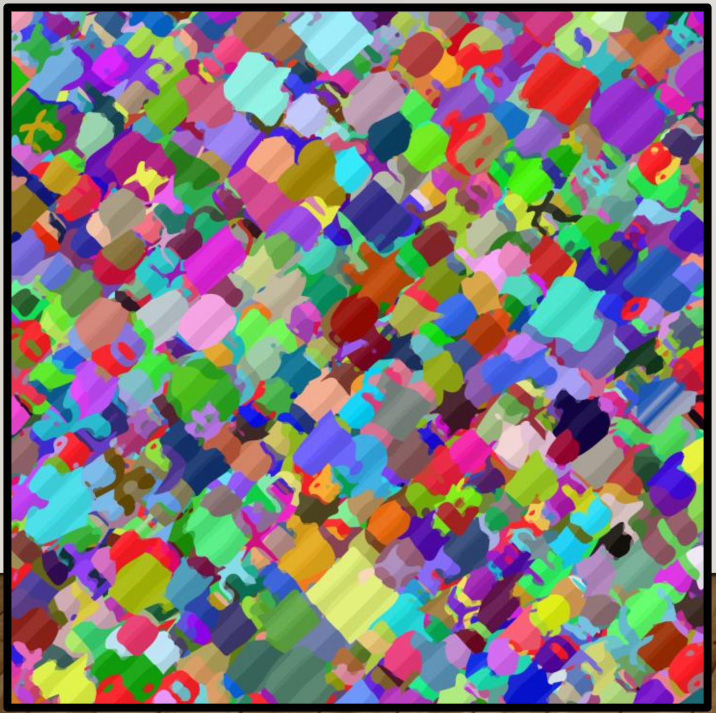


BERKLEY THEORY ASSUMPTION

Superior Randomness?

– Lacking a Definite Plan, Purpose, or Pattern (Merriam-Webster.com)

- Random by One Measure – Ordered by Another



- Relevant Standards ?

-- Party?	Random / Ordered
-- Precinct?	Random / Ordered
-- Gender?	Random / Ordered
-- Race?	Random / Ordered
-- Zip Code?	Random / Ordered
-- Spec District?	Random / Ordered
-- UNOCAVA?	Random / Ordered
-- VSPC?	Random / Ordered
-- Mail-In?	Random / Ordered

DOUGLAS COUNTY
COLORADO

BERKLEY THEORY ASSUMPTION

"Fallacy in the CSAP"

1. Internal Logic & Accuracy Test (I-LAT) – DC Elections
    - -- Randomly Marked Ballots are Noted
    - -- Then Fed Through the System
    - -- Results are Compared to Previously Noted Results
2. Public Logic & Accuracy Test (P-LAT) – Public Participation
    - -- Same Process as Noted Above
3. Election – Marked Ballots are Fed Through the System
    - -- Marked Ballots are Tabulated
4. Post Election Audit (PEA) – Canvass
    - -- Same Process as Noted 1 & 2 Above

DOUGLAS COUNTY
COLORADO

Security of the CSAP

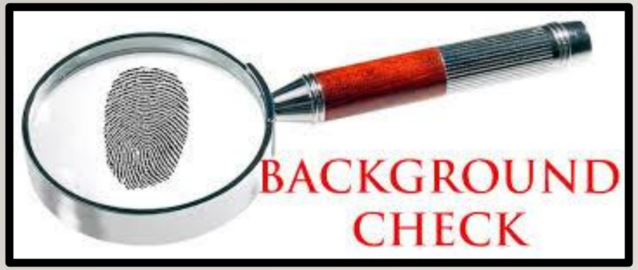
Personnel Protocol

- Background Checked
- Training

-- Chain of Custody

-- Relevant Regulatory Requirements

-- Two-Person Rule (Different Parties)



DOUGLAS COUNTY
COLORADO

Security of the CSAP

Physical Protocol
- -- In Elections Office Basement
  - --- 3 Locked Doors
  - --- 24-Hour Camera Surveillance
    - --- Real-Time Monitor When Building is Closed
  - --- Staff Escort Required

Virtual Protocol
- -- Isolated Computer Network (No Connection goes Outside of Tabulation Room)
- -- Restricted Access to all Server, Ballot Creation, and Tabulation Machines (Logistics & Technology Manager & Elections Deputy)
- -- Security Function Maintains a Key-Stroke Log

DOUGLAS COUNTY
COLORADO

Security of the CSAP

The Post Election Audit (PEA):

-- Done by Partisan Public Rep's and Monitored by DC Elections Personnel

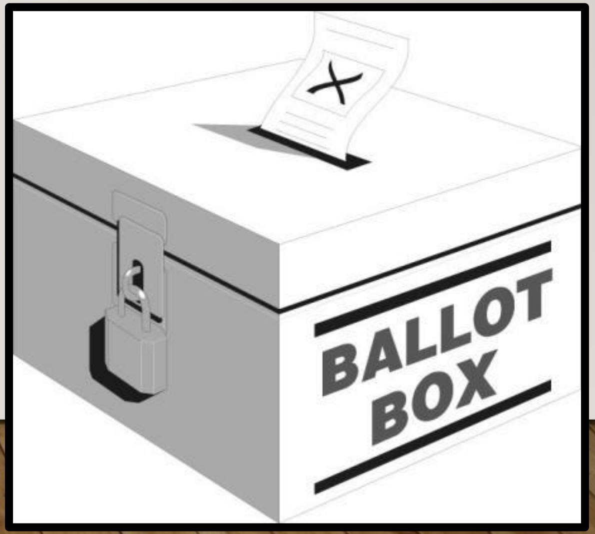
-- 500 Ballots are Randomly Selected

--- 100 Count Batches

--- Randomly Selected by Partisan Judges

--- Ballots are Noted for Results

--- Ballots are Tabulated and Results Compared

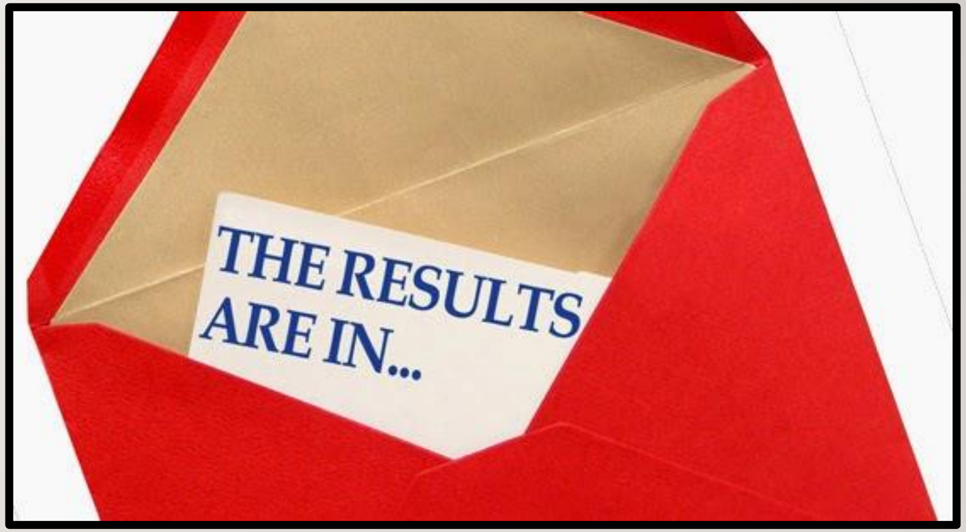


DOUGLAS COUNTY
COLORADO

Security of the CSAP

Percent of Votes Required for Percent of Confidence in No Mis-counted Ballots in a 50,000 - 200,000 Count Cast-Vote Sample

% of Votes	95%	99%
51%	298	457
52%	149	228
53%	99	151
54%	74	113
55%	59	90
56%	49	75
57%	42	64
58%	36	56
59%	32	49
60%	29	44



DOUGLAS COUNTY
COLORADO

CSAP – 2016 Elections

Percent of Votes Required for Percent of Confidence in a 50,000 - 200,000 Count Cast-Vote Sample

% of Votes	95%	99%
51%	298	457
52%	149	228
53%	99	151
54%	74	113
55%	59	90
56%	49	75
57%	42	64
58%	36	56
59%	32	49
60%	29	44

Deltas in 3 Selected 2017 Races

US Senate Votes Cast	Glenn	Bennett
177,925	107,920 61%	39,554 39%
US Presidential	**Trump**	**Clinton**
171,230	102,573 60%	68,657 40%
UC Regent	**Ganahl**	**Madden**
171,048	113,724 66%	57,324 34%

CSAP – 99% Confidence Rating of No Incorrectly Counted Ballots in Noted Races

DOUGLAS COUNTY
COLORADO

Comparison Test of Amendment 72 - 2016

Colorado SOS Audit Protocol

% of Votes	95%	99%
51%	298	457
52%	149	228
53%	99	151
54%	74	113
55%	59	90
56%	49	75
57%	42	64
58%	36	56
59%	32	49
60%	29	44

CSAP – 99% Confidence Rating of No Incorrectly Counted Votes

Berkley Theory

Amendment 72 (CONSTITUTIONAL), Vote For 1

Yes/For	89,864	48.27%
No/Against	96,317	51.73%
Cast Votes:	186,181	96.72%

Initial sample size

Contest information

Ballots cast in all contests: 192617 Smallest margin (votes): 6,453. Diluted margin: 3.35%.

Contest 1. Contest name: Amendment 72

Winners: 1

Reported votes:

Candidate 1 Name: Yes/For Votes: 89864

Candidate 2 Name: No/Against Votes: 96317

Add candidate to contest 1 | Remove last candidate from contest 1

Add contest | Remove last contest

Audit parameters

Risk limit: 5% Expected sample size: 5,368.

DOUGLAS COUNTY
COLORADO

Comparison Test of Amendment 72 - 2016

Sample Race from 2016 Election
-- Assuming Audit of 5,368 as Noted by the Berkley Theory Calculator

To Meet Statutory Timeline (Need Accomplish w/in ~3 Days)

-- Total Judges - 32
-- Total Staff - 20
-- Total Work Hours (Prep + RLA)
--- 1,187
-- Total Work Days - 2.85
-- Total Cost - $17,137.48

Sample Race from 2016 Election

Amendment 72 (CONSTITUTIONAL), Vote For 1

Yes/For	89,864	48.27%
No/Against	96,317	51.73%
Cast Votes:	186,181	96.72%

Initial sample size

Contest information

Ballots cast in all contests: 192617 Smallest margin (votes): 6,453. Diluted margin: 3.35%.

Contest 1. Contest name: Amendment 72
Winners: 1

Reported votes:

Candidate 1 Name: Yes/For Votes: 89864
Candidate 2 Name: No/Against Votes: 96317

Add candidate to contest 1 | Remove last candidate from contest 1

Add contest | Remove last contest

Audit parameters

Risk limit: 5% Expected sample size: 5,368.

CSAP – 99% Confidence Rating of No Incorrectly Counted Votes in Noted Race

DOUGLAS COUNTY
COLORADO

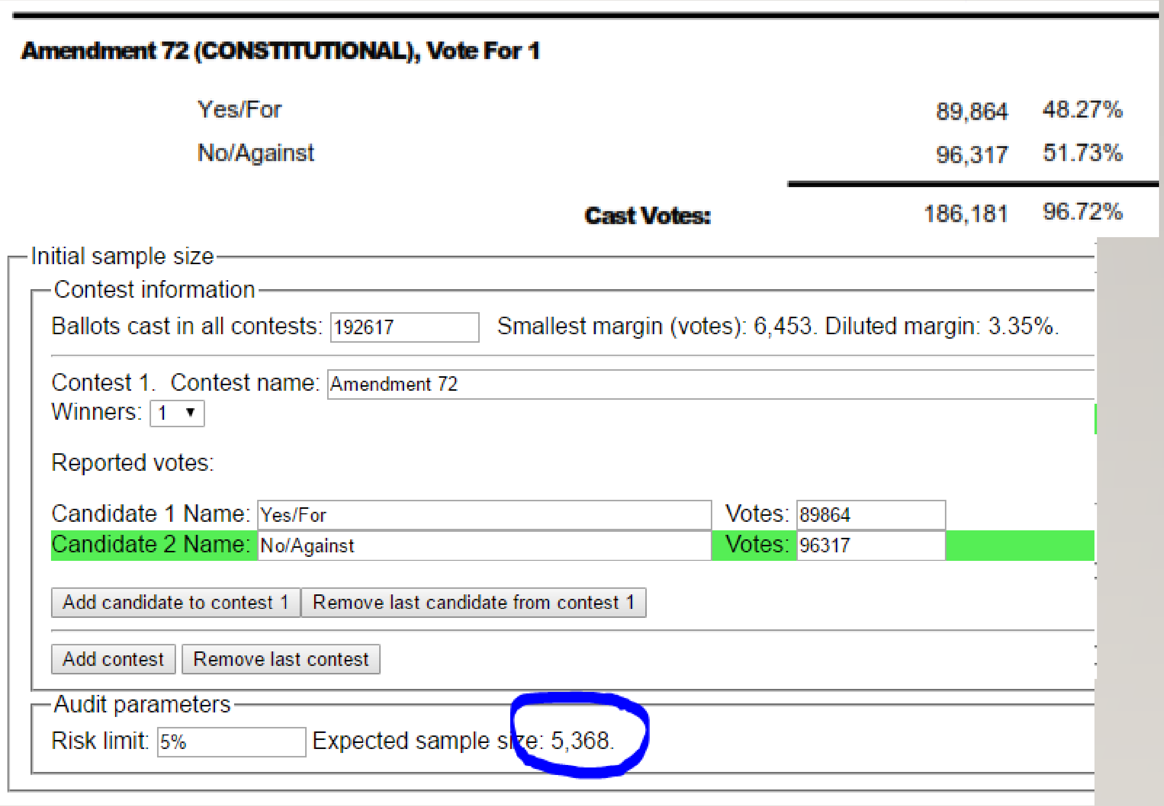
Comparison Test of Amendment 72 - 2016

Sample Race from 2016 Election
-- Assuming Audit of 5,368 Ballots as Noted by Berkley Theory Calculator

To Limit Stress on Workforce and Logistic Requirement (Likely Not Meet Statutory Timeline)

-- Total Judges - 12
-- Total Staff - 20
-- Total Work Hours (Prep + RLA)
--- 1,583
-- Total Work Days – 7.61
-- Total Cost - $22,464.66

Sample Race from 2016 Election



Amendment 72 (CONSTITUTIONAL), Vote For 1

Yes/For	89,864	48.27%
No/Against	96,317	51.73%
Cast Votes:	186,181	96.72%

Initial sample size
Contest information
Ballots cast in all contests: 192617 Smallest margin (votes): 6,453. Diluted margin: 3.35%.
Contest 1. Contest name: Amendment 72
Winners: 1
Reported votes:
Candidate 1 Name: Yes/For Votes: 89864
Candidate 2 Name: No/Against Votes: 96317
Add candidate to contest 1 Remove last candidate from contest 1
Add contest Remove last contest
Audit parameters
Risk limit: 5% Expected sample size: 5,368.

CSAP – 99% Confidence Rating of No Incorrectly Counted Votes in Noted Race

DOUGLAS COUNTY
COLORADO

Berkley Theory Not Applicable

1. Due to lack of homogeneity in county processes
   -- Colorado Counties all follow the same guidelines toward the same goals but utilize variant proprietary processes and protocols to do so
2. Due to increased labor costs
   -- In the 2016 selected race example, labor costs would have increased
3. Due to extended timelines
   -- In the 2016 selected race example, Certification may have been put off by several days
4. Due to "esoteric knowledge" concept
   -- Plato's "Philosopher-King" class of individuals possessed the esoteric knowledge to run the machinations of the state. Not applicable to governance in a free society.

DOUGLAS COUNTY
COLORADO

THEORETICAL SECURITY CHALLENGE

To Affect the Election Results You Must:

1. Manipulate the Tabulation System
2. Manipulate the Tabulation System Undetected
3. Erase all Trace of the Manipulation

DOUGLAS COUNTY
COLORADO

Security of the CSAP

Tampering with the Voter Tabulation Machine and having the activity go unnoticed is analogous to someone breaking into a bank secured under lock, key, security camera, alarms, a Quick Reaction Force, and an electronic monitoring system and robbing it immediately after an audit and then breaking back into that same bank and replacing the money, note for note, coin for coin, and spatial location for spatial location prior to the next audit.

This basement and process is more secure than the bank.

(Statement from a Castle Rock Banker after touring the DC Elections Building.)

Suggested Improvement Considerations

1. Doubling the standard number of audited ballots in the PEA
   -- Auditing 500 Ballots in the 2016 Election yielded a 99% confidence of no incorrect tabulations
   -- Auditing 1,000 ballots in a 51% victory-margin race would provide a 99.9916% confidence that no ballots are mis-counted
2. 30 days prior to election, Sheriff's Office will sweep the Elections Building to confirm no illicit electronic monitoring devices.
   -- To enhance public confidence in the process

e. The same is true with smaller and smaller pieces.

DOUGLAS COUNTY
COLORADO

BOTTOM LINE

CRS 1-7-515 Risk Limiting Audits – Rules – Legislative Declaration – Definitions

-- (5)(b) "'Risk limiting audit' means an audit protocol that makes use of statistical methods and is designed to limit to acceptable levels the risk of certifying a preliminary election outcome that constitutes an incorrect outcome."

DOUGLAS COUNTY
COLORADO

e. The same is true with smaller and smaller pieces.

BOTTOM LINE - RECOMMENDATION

1) The audit protocol requirement covers the initial ballot data selection, all associated security activities, ballot processing, and staff activities through and to include certification of results.

2) The current Douglas County Audit Protocol uses a statistically-based method to select ballots for review that provides a confidence level between 97% - 99% that no ballot has been miscounted.

Douglas County Colorado

ELECTION AUDIT PROTOCOL

Colorado Current Audit Protocol

Personnel, Physical, & Virtual

Ballot Creation

Ballot Proofing

LATs

Delivery of Mail Ballots VSPCs

Post-Election Audit

Election Certification

Berkley Theory

Ballot Creation

Review of X Subset of Cast Ballots

Election Certification

DOUGLAS COUNTY COLORADO

Conclusion

There seems to be little or no data nor logic to support a contention that when validated water goes into one end of a pipe and validated water comes out of the other end of a pipe and someone you trust is watching the middle of that pipe that water is not what you've got.

A Gentle Introduction to Risk-limiting Audits

Mark Lindeman and Philip B. Stark

***Abstract*—Risk-limiting audits provide statistical assurance that election outcomes are correct by manually examining portions of the audit trail—paper ballots or voter-verifiable paper records. We sketch two types of risk-limiting audits, *ballot-polling audits* and *comparison audits*, and give example computations. Tools to perform the computations are available at statistics.berkeley.edu/~stark/Vote/auditTools.htm.**

I. WHAT IS A RISK-LIMITING AUDIT?

A risk-limiting audit is a method to ensure that at the end of the canvass, the hardware, software, and procedures used to tally votes found the real winners. Risk-limiting audits do not guarantee that the electoral outcome is right, but they have a large chance of correcting the outcome if it is wrong. They involve manually examining portions of an *audit trail* of (generally paper) records that voters had the opportunity to verify recorded their selections accurately.

Risk-limiting audits address limitations and vulnerabilities of voting technology, including the accuracy of algorithms used to infer voter intent, configuration and programming errors, and malicious subversion. Computer software cannot be guaranteed to be perfect or secure, so voting systems should be *software-independent*: An undetected change or error in voting system software should be incapable of causing an undetectable change or error in an election outcome [Rivest and Wack, 2006, Rivest, 2008]. A well-curated audit trail provides software independence; a risk-limiting audit leverages software independence by checking the audit trail strategically.

Systems that do not produce voter-verifiable paper records, such as paperless touchscreen voting systems, cannot be audited this way. Records of cast votes printed after the voter has left do not confer software independence, because voters had no chance to verify them.

The simplest risk-limiting audit is an accurate full hand tally of a reliable audit trail: Such a count reveals the correct outcome. However, a full hand count generally wastes resources: Examining far fewer ballots often can provide strong evidence that the outcome is correct, if those ballots are chosen at random by suitable means. Hence, to keep the counting burden as low as possible, the methods described here conduct an "intelligent" incremental recount that stops when the audit provides sufficiently strong evidence that a full hand count would confirm the original (voting system) outcome. As long as the audit does not yield sufficiently strong evidence, more ballots are manually inspected, potentially progressing to a full hand tally of all the ballots. (The full hand count can be part of the audit, or a separate process.) "Sufficiently strong" is quantified by the *risk limit*, the largest chance that the audit will stop short of a full hand tally when the original outcome is in fact wrong, no matter why it is wrong, including "random" errors, voter errors, configuration errors, bugs, equipment failures, or deliberate fraud.

To appear in *IEEE Security and Privacy*.

ML: email: taxshift AT gmail DOT com. PBS: Department of Statistics, University of California, Berkeley CA, 94720-3860, USA. e-mail: stark AT stat DOT berkeley DOT edu.

We are grateful to Jennie Bretschneider, Ronald L. Rivest, and Barbara Simons for helpful comments.

Manuscript received ?? December 2011. First published ? October 2012.

Smaller risk limits entail stronger evidence that the outcome is correct: All else equal, the audit examines more ballots if the risk limit is 1% than if it is 10%. Smaller (percentage) margins require more evidence, because there is less room for error: All else equal, the audit examines more ballots if the margin is 1% than if it is 10%.

The risk limit is *not* the chance that the outcome (after auditing) is wrong. A risk-limiting audit emends the outcome if and only if it leads to a full hand tally that disagrees with the original outcome. Hence, a risk-limiting audit cannot harm correct outcomes. But if the original outcome is wrong, there is a chance the audit will not correct it. The risk limit is the *largest* such chance. If the risk limit is 10% and the outcome is wrong, there is at most a 10% chance (and typically much less) that the audit will not correct the outcome—at least a 90% chance (and typically much more) that the audit will correct the outcome.

There is an extensive literature on post-election audits; we do not summarize it here. And we omit important implementation details. Our point is merely that efficient risk-limiting audits do not require complicated calculations or in-house statistical expertise.

A. The audit trail

Risk-limiting audits involve manually interpreting the votes in portions of the audit trail. The best audit trail is voter-marked paper ballots. Voter-verifiable paper records (VVPRs) printed by voting machines are not as good. Voters might not actually inspect VVPRs. Printers can jam or run out of paper. VVPRs can be fragile and cumbersome to audit. (As noted above, paperless touchscreen voting machines do not provide a suitable audit trail.) Below, we call entries in the audit trail "ballots" regardless of how they were created.

Like a recount, a risk-limiting audit assumes there is a correct interpretation of each ballot. Rules for interpreting ballots must be established before the audit starts.

B. Ballot-level audits

States that mandate hand counting as part of audits generally require counting the votes in selected *clusters* of ballots (sometimes called "batches," but "batches" means something else below). For instance, under California law, each county counts the votes in 1% of precincts; each cluster comprises the ballots cast in one precinct.

The smaller the clusters, the less counting a risk-limiting audit requires—if the outcome is correct. (If the outcome is wrong, the audit has a large chance of counting all the votes, regardless of the size of the clusters.) A random sample of 100 individual ballots can be almost as informative as a random sample of 100 entire precincts! Hand counting is minimized when clusters consist of one ballot each, yielding "ballot-level" audits or "single-ballot" audits. See Stark [2010a] for more discussion.

Ballot-level audits save work, but finding individual ballots among millions stored in numerous boxes or bags ("batches") is challenging. It requires knowing the number of ballots in each batch (i.e., having a *manifest*, discussed below), how to locate each batch, and how to identify each ballot within each batch uniquely. Labeling each ballot helps, but is prohibited in some jurisdictions. Ballot-level auditing elevates privacy concerns. The most efficient ballot-level audits, comparison audits (explained below), require the voting system interpretation of every ballot—which no federally certified vote tabulation system reports. (See Stark and Wagner [2012].)

If the voting system does not report its interpretation of each ballot, one can audit using an unofficial system that does. *Transitive auditing* checks the unofficial system, rather than the system of record. If the two systems show different outcomes, all votes should be counted by hand. If the systems show the same outcome, a risk-limiting audit of the unofficial system checks the outcome of the system of record: Either both are right or both are wrong. If both are wrong, the risk-limiting audit has a large chance of requiring a full hand count. See, e.g., Calandrino et al. [2007], Benaloh et al. [2011].

II. Before the audit starts

Because a risk-limiting audit relies upon the audit trail, preserving the audit trail complete and intact is crucial. If a jurisdiction's procedures for protecting the audit trail are adequate in principle, ensuring compliance with those procedures (possibly as part of a comprehensive canvass or a separate *compliance audit*) can provide strong evidence that the audit trail is trustworthy. If the compliance audit does not generate convincing affirmative evidence that the ballots have not been altered and that no ballots have been added or lost, a risk-limiting audit may be mere theater [Benaloh et al., 2011, Stark and Wagner, 2012].

To sample ballots efficiently requires a *ballot manifest* that describes in detail how the ballots are organized and stored. For instance, the jurisdiction might keep cast ballots in 350 batches, labeled 1 to 350. The manifest might say "There are 71,026 ballots in 350 batches: Batch 1 has 227 ballots; batch 2 has 903 ballots; ...; and batch 350 has 114 ballots." If the jurisdiction numbers its ballots, the manifest might say, "Batch 1 contains ballots 1–227; batch 2 contains ballots 228–1,130; ...; and batch 350 contains ballots 70,913–71,026."

Auditors should verify that the number of ballots in the manifest matches the total according to the election results. It is good practice to count the ballots in the batches containing the ballots selected for audit, to check whether the manifest is accurate. If the manifest is inaccurate, the risk limit may not be correct.

III. Two kinds of simple risk-limiting audits

We present simple examples of two kinds of risk-limiting audits: *ballot-polling audits* and *comparison audits*. (Johnson [2004] makes an analogous distinction, but does not address risk-limiting audits per se.) "Simple" means that the calculations are easy, even with a pencil and paper, so observers can check the auditors' work. Tools that perform these calculations are available at statistics.berkeley.edu/~stark/Vote/auditTools.htm, the "auditTools page."

This section addresses risk-limiting audits of a vote-for-one contest. Section V discusses auditing more than one contest at once, contests with more than one winner, contests that require a super-majority, and ranked-choice voting.

A. Ballot-polling audits

Ballot-polling audits examine a random sample of ballots. When the vote shares in the sample give sufficiently strong evidence that the reported winner really won, the audit stops.

Ballot-polling audits require knowing who reportedly won, but no other data from the vote tabulation system. They are best when the vote tabulation system cannot export vote counts for individual ballots or clusters of ballots or when it is impractical to retrieve the ballots that correspond to such counts. Ballot-polling audits generally require examining more ballots than ballot-level comparison audits (described below) and the workload is disproportionately higher for contests with smaller margins—but comparison audits require much more information from the vote tabulation system, information that might not be available quickly in a useful format, if at all.

The following ballot-polling audit, which relies on Wald's sequential probability ratio test [Wald, 1945], has risk limit 10%: There is at least a 90% chance it will require a full hand count if the reported winner actually lost. It assumes that the winner's reported share s of valid votes is greater than 50%: a majority rather than a mere plurality. With small changes, it applies to contests that require a super-majority. Slightly more complicated procedures deal with winners who fall short of a majority.

1) Let s be the winner's share of the valid votes according to the vote tabulation system; this procedure requires $s > 50\%$. Let t be a positive "tolerance" small enough that when t is subtracted from the winner's vote share s, the difference is still greater than 50%. (Increasing t reduces the chance of a full hand count if the voting system outcome is correct, but increases the expected number of ballots to be counted during the audit.) Set $T = 1$.
2) Select a ballot at random from the ballots cast in the contest (see section IV). A ballot can be selected more than once; the following steps apply each time.
3) If the ballot does not show a valid vote, return to step 2.

4) If the ballot shows a valid vote for the winner, multiply T by
$$(s-t)/50\%.$$
5) If the ballot shows a valid vote for anyone else, multiply T by
$$(1-(s-t))/50\%.$$
6) If $T > 9.9$, the audit has provided strong evidence that the reported outcome is correct: Stop.
7) If $T < 0.011$, perform a full hand count to determine who won. Otherwise, return to step 2.

If the reported winner's true share of the vote is at least $s-t$, there is at most a 1% chance that this procedure will lead to a full hand count; that chance and the risk limit can be altered by adjusting the comparisons in steps 6 and 7.

As a numerical example, suppose one candidate reportedly received $s = 60\%$ of the valid votes. Set $t = 1\%$. If the reported winner really received at least $s-t = 59\%$ of the vote, there is at most a 1% chance that the procedure will lead to a (pointless) full hand count. Note that $1-(s-t) = 1-59\% = 41\%$. To audit, we repeat steps 2–7, drawing ballots at random and updating T until either $T > 9.9$ or $T < 0.011$.

The number of ballots eventually audited depends on the vote shares and on which ballots happen to be selected. If the first 14 ballots drawn all show votes for the winner,

$$\begin{aligned} T &= (59\%/50\%) \times (59\%/50\%) \times \cdots \times (59\%/50\%) \\ &= (59\%/50\%)^{14} = 10.15, \end{aligned}$$

and the audit stops.

If the reported winner's true vote share is 60%, the audit is expected to examine 120 ballots; for a 55% share, 480; and for a 52% share, 3,860: The expected workload grows quickly as the margin shrinks.

When the outcome is correct, the number of ballots the audit examines depends only weakly on the number of ballots cast, so the percentage of ballots examined in large contests can be quite small. For example, in the 2008 presidential election, 13.7 million ballots were cast in California; Barack Obama was reported to have received 61.1% of the vote. A ballot-polling audit could confirm that Obama won California at 10% risk (with $t = 1\%$) by auditing roughly 97 ballots—seven ten-thousandths of one percent of the ballots cast—if Obama really received over 61% of the votes.

The expected auditing workload in each county is proportional to the percentage of ballots cast in the county. Almost 25% of the ballots were cast in Los Angeles county, the largest of California's 58 counties. Over 75% of the ballots were cast in the largest 12 counties. The smallest 14 counties together account for less than 1% of ballots cast. So, about 24 of the 97 ballots would be from Los Angeles; 73 from the largest 12 counties, including Los Angeles; and perhaps one ballot total from the smallest 14 counties.

If the winner's share were 52% rather than 61.1%, the expected number of ballots to examine would be 3,860—far more, but still less than three hundredths of one percent of the ballots cast. Of those, Los Angeles would have expected to examine about 946, the largest 12 counties about 2,922 total, and the smallest 14 counties about 35 total. Since ballot-polling audits do not require data from the vote tabulation system, they are an immediate practical option for auditing large contests. Indeed, *all* statewide contests could be confirmed with a single ballot-polling audit expected to examine 3,860 ballots if the winners' smallest vote share was 52%. Comparison audits, described next, generally involve examining fewer ballots, but require much more from the vote tabulation system.

B. Comparison audits

Comparison audits check outcomes by comparing hand counts to voting system counts for clusters of ballots. In ballot-level comparison audits, each cluster is one ballot. Comparison audits can be thought of as having two phases: (i) Check whether the reported subtotals for every cluster of ballots sum to the contest totals for every candidate. If they do not, the reported results are inconsistent; the audit cannot proceed. (ii) Spot-check the voting system subtotals against hand counts for randomly selected clusters, to assess whether the subtotals are sufficiently accurate to determine who won. If not, the audit has a large chance of requiring a full hand count.

This section is based on the "super-simple" ballot-level risk-limiting comparison audit [Stark, 2010b]. It presumes we know how the vote tabulation system (or, for transitive audits, an unofficial system) interpreted every ballot. The audit compares a manual interpretation of ballots selected at random to the system's interpretation of those ballots, continuing until there is strong evidence that the outcome is correct—or requiring a full hand count.

Suppose the manual interpretation of a ballot disagrees with the voting system interpretation. If changing the voting system interpretation to match the manual interpretation would increase the margin(s) between the winner and every loser, the ballot has an "understatement." If the voting system interpretation of a ballot records an overvote but the manual interpretation shows a vote for the winner, the ballot has an understatement. Understatements do not call the outcome into question, because correcting them benefits the winner.

If changing the voting system interpretation to match the manual interpretation would decrease the margin between the winner and any loser, the ballot has an "overstatement" equal to the maximum number of votes by which any margin would decrease. If the voting system interpretation of a ballot records an undervote but the manual interpretation finds a vote for one of the losers, the ballot has an overstatement of one vote: The voting system interpretation overstated the margin by one vote. If the voting system interpretation of a ballot recorded a vote for the winner but the manual interpretation finds an overvote, that ballot has an overstatement of one vote.

If the voting system interprets a ballot as a vote for the winner while a manual interpretation finds a vote for one of the losers, that ballot has an overstatement of *two* votes. For voter-marked paper ballots, occasional one-vote misstatements are expected, owing to the vagaries of how voters mark their ballots: From time to time the system will interpret a light mark as an undervote or a hesitation mark as an overvote. But two-vote overstatements should be quite rare: A properly functioning voting system should not award a vote for one candidate to a different candidate.

We now present a simple rule for a risk-limiting comparison audit with risk limit 10%. The rule depends on the "diluted margin" m, the smallest reported margin (in votes), divided by the number of ballots cast. Dividing by the number of ballots, rather than by the number of valid votes, allows for the possibility that the vote tabulation system mistook an undervote or overvote for a valid vote, or vice versa. Suppose the audit has inspected n ballots. Let u_1 and o_1 be the number of 1-vote understatements and overstatements among those n ballots, respectively; similarly, let u_2 and o_2 be the number of 2-vote understatements and overstatements. The audit can stop when

$$n \geq \frac{4.8 + 1.4(o_1 + 5o_2 - 0.6u_1 - 4.4u_2)}{m}. \tag{1}$$

(This follows from equation [9] of Stark [2010b] with risk limit $\alpha = 10\%$ and $\gamma = 1.03905$, by the same conservative approximation used to derive equation [17] there, with a bit of rounding.)

Overstatements increase the required sample size and understatements decrease it, but not by equal amounts. We have more confidence in the outcome if the sample shows no misstatements than if it shows large but equal numbers of understatements and overstatements. In condition [1] a 1-vote understatement offsets 60% of a 1-vote overstatement and a 2-vote understatement offsets 88% of a 2-vote overstatement.

If the diluted margin m is 10%, each 1-vote overstatement increases the required sample size by $1.4/10\% = 14$ ballots and each 1-vote understatement decreases the required sample size by $1.4 \times 0.6/10\% = 8.4$ ballots. Each 2-vote overstatement increases the required sample size by $1.4 \times 5/10\% = 70$ ballots and each 2-vote understatement decreases the required sample size by $1.4 \times 4.4/10\% = 61.6$ ballots. For $m = 5\%$, these numbers double; for $m = 2\%$, they quintuple.

With this method, the auditor can check one ballot at a time against its voting system interpretation sequentially or check a larger number in parallel. Moreover, the auditor can decide at any point to abort the audit and require a full hand count. The risk limit will be 10% provided the audit continues either until condition [1] is satisfied or until there is a full hand count; then the hand-count outcome replaces the reported outcome.

Numerical examples might help. Suppose that 10,000 ballots were cast in a particular contest. According to the vote tabulation system, the reported winner received 4,000 votes and the runner-up received 3,500 votes. Then the diluted margin is $m = (4000 - 3500)/10000 = 5\%$. We consider sampling ballots incrementally and sampling in stages.

1) Sampling incrementally: In an incremental audit, the auditor draws a ballot at random and checks by hand whether the voting system interpretation of that ballot is right before drawing the next ballot. If there is one 1-vote understatement and no other misstatements among the first 80 ballots examined, $u_1 = 1$ and o_1, u_2, and o_2 are all zero and the audit can stop, because

$$80 \geq \frac{4.8 - 1.4 \times 0.6 \times 1}{5\%}. \tag{2}$$

If there are no overstatements or understatements among the first 96 ballots examined, u_1, o_1, u_2, and o_2 are all zero and the audit can stop, because

$$96 \geq 4.8/5\%. \tag{3}$$

2) Sampling in stages: To simplify logistics, an auditor might draw many ballots at once, then compare each to its voting system interpretation. If condition [1] is not met, the auditor draws another set of ballots and compares them to their voting system interpretations. Each set of draws and comparisons is a *stage*. (If a ballot is drawn more than once, it enters the calculations as many times as it is drawn.)

If the auditor expects errors at some rate, she can select the first-stage sample size so that the audit stops there if her expectation proves correct or pessimistic. Suppose she expects one 1-vote overstatement and one 1-vote understatement per thousand ballots (0.001 per ballot), and expects 2-vote misstatements to be negligibly rare. For a contest with a diluted margin m of at least 5%, an initial sample of $4.8/m$ ballots (rounded up) is 96 ballots or fewer. If overstatements are as infrequent as expected, there are unlikely to be any among the first 96 ballots: The audit will stop at the first stage. An initial sample of $6.2/m$ (124 ballots or fewer if the margin is at least 5%) allows the audit to stop at the first stage if it shows one 1-vote overstatement.

It can save effort to sort the sample (for instance, by precinct) before retrieving the ballots and checking their interpretation. But then all ballots drawn in the stage should be checked before determining whether to stop. Otherwise the procedure is biased in favor of ballots from precincts that are early in sorted order.

Table I gives stopping sample sizes for various diluted margins and numbers of overstatements and understatements, for 10% risk. It can help select the first-stage sample size for different expected rates of error.

IV. Random selection

Risk-limiting audits rely on random sampling. (Random samples can be augmented with "targeted" samples chosen by other means; see, e.g., Stark [2009a].) If the sample is not drawn appropriately, the risk limit will be wrong. The risk-limiting methods described above rely on drawing a random sample of ballots with replacement. This is like putting all the ballots into an enormous mixer, stirring them thoroughly, and drawing a ballot without looking. The ballot is returned to the mixer, the ballots are mixed again, and another ballot is drawn (possibly the same ballot), until the audit stops.

Public confidence requires that observers can verify the selection is fair—that all ballots are equally likely to be selected in each draw. This speaks against a number of common methods for selecting samples, including "arbitrary" selection by the election officials; drawing slips of paper, where there is little hope of confirming that each ballot is represented by exactly one slip and that the slips have been adequately mixed; using proprietary software such as Excel; or using any source of putative randomness that cannot readily be checked.

Trustworthy methods of generating random numbers often have two features: a physical source of randomness (such

	0 understatements					*1 1-vote understatement*				
diluted	# 1-vote overstatements					# 1-vote overstatements				
margin	0	1	2	3	4	0	1	2	3	4
0.2%	2400	3100	3800	4500	5200	1980	2680	3380	4080	4780
0.5%	960	1240	1520	1800	2080	792	1072	1352	1632	1912
1%	480	620	760	900	1040	396	536	676	816	956
2%	240	310	380	450	520	198	268	338	408	478
5%	96	124	152	180	208	80	108	136	164	192
10%	48	62	76	90	104	40	54	68	82	96
20%	24	31	38	45	52	20	27	34	41	48

TABLE I
EXEMPLAR SAMPLE SIZES FOR BALLOT-LEVEL COMPARISON AUDITS WITH VARIOUS DILUTED MARGINS AND VARIOUS NUMBERS OF MISSTATEMENTS IN THE SAMPLE, 10% RISK LIMIT.

as dice rolls) and inputs from multiple parties (so that even if some parties collude, any non-colluding party could foil an attempt to rig the sample). It can be efficient, effective, and transparent to use a simple mechanical method—such as rolling dice [Cordero et al., 2006]—to generate a “seed” for a well-designed *pseudo-random number generator* (PRNG). PRNGs can generate arbitrarily many “pseudo-random” numbers from a single seed. PRNG output is deterministic given the seed, but the numbers produced by good PRNGs have many of the desirable properties of random sequences. And any observer who knows the seed and the PRNG can check the output. For good PRNGs, small changes in the seed yield very different sequences, so starting with a random seed makes it effectively impossible for anyone to render the audit less effective by anticipating which ballots will be examined.

The auditTools page (described in section III) provides a good PRNG suggested by Ronald L. Rivest. It relies on the SHA-256 cryptographic hash function, which is in the public domain and has been implemented in many programming languages. That allows observers to confirm that the sequence of pseudo-random numbers is correct, given the seed.

A ballot manifest can be used to identify the particular ballots that correspond to the random (or pseudo-random) numbers in the sample. Before the audit, we use the manifest to assign a unique number to each ballot, if the ballots are not already marked uniquely. Suppose that the manifest lists 822 ballots in three batches, numbered 1 through 3; the batches contain, respectively, 230, 312, and 280 ballots. Then we can number the 230 ballots in batch 1 ballots 1 through 230; the 312 ballots in batch 2 ballots 231 through 542; and the 280 ballots in batch 3 ballots 543 through 822. Ballot 254 is the 24th ballot in batch 2. We assume that the ballots are stored in some order that remains unchanged during the audit, so that “the 24th ballot in batch 2” uniquely identifies a particular ballot.

To draw the audit sample, we generate random numbers between 1 and 822, and retrieve the corresponding ballot. If 254 is generated, we retrieve batch 2 and count into that batch to find the 24th ballot, which we audit.

V. MORE COMPLICATED SITUATIONS

We have discussed only contests where the candidate with the most votes wins. The methods can be extended to audit contests that require a supermajority, contests with more than one winner, cross-jurisdictional contests, and ranked-choice voting; and to audit a collection of contests simultaneously with a single sample.

Contests with more than one winner and collections of contests can be audited with a comparison audit based on the *maximum relative overstatement of pairwise margins* (MRO) [Stark, 2008b, 2009b], defined as follows. A *pairwise margin* is the margin in votes between any winner and any loser in a given contest. An overstatement of a pairwise margin, divided by that margin, is the *relative overstatement* of the pairwise margin. A one-vote overstatement of a wide margin casts less doubt on the outcome than a one-vote overstatement of a narrow margin; relative overstatements take this into account. The MRO is the maximum relative overstatement on each audited ballot. The arithmetic can be simplified by treating all overstatements as if they affected the smallest diluted margin. This is conservative, but if overstatements are rare, the workload remains manageable. That is the heart of the “super-simple” simultaneous audit method [Stark, 2010b].

For simultaneous audits of multiple contests, the diluted margin is the smallest reported margin in votes, divided by the total number of ballots on which at least one of the contests appears. If a contest appears on only a small fraction of ballots, it may take less work to audit it separately, so that its diluted margin considers only the ballots that contain the contest.

Auditing contests that cross jurisdictional boundaries is straightforward if all the results are available before the audit starts, and the sample can be drawn from all ballots as a pool. If the jurisdictions draw samples independently, the computations are complicated [Stark, 2008a, Higgins et al., 2011]. Auditing instant-runoff or ranked-choice (IRV/RCV) contests is a topic of research: Even computing the “margin of victory” is difficult [Magrino et al., 2011, Cary, 2011].

VI. A PRACTICAL EXAMPLE: MERCED COUNTY, CALIFORNIA

The methods described above have been used to audit elections in California, including the November 2011 election in Merced County. That audit, authorized by California’s 2010 law AB 2023 and funded by a grant from the U.S. Election Assistance Commission, was a comparison audit that used a single sample to confirm two City of Merced contests: the mayoral contest, and the (vote-for-three) councilmember contest. In the mayoral contest, which had five candidates, the voting system reported that Stan Thurston received 2,231 votes, and runner-up Bill Blake received 2,037—a margin of 194 votes, or 2.79% of valid votes cast. In the councilmember contest, the margin of decision (between the third-place and fourth-place candidates) was wider, 959 votes.

Because Merced's voting system cannot report its interpretation of individual ballots, a transitive audit was conducted: The 7,120 cast ballots were digitally scanned. A ballot manifest was prepared. Kai Wang, Ph.D. student at the University of California, San Diego, interpreted the images using software he wrote, spot-checking "difficult" cases by hand. His vote totals were slightly higher than the official totals, but gave the same winners. The margin he found for the mayoral contest was 192 votes, a diluted margin m of about 2.70%. Before the audit started, the unofficial interpretations were posted to a website so that anyone interested could verify that those interpretations did not change during the audit.

The initial sample was large enough to confirm the original results at 10% risk limit if it revealed few overstatements. The minimum sample size if there were no misstatements would be $4.8/m = 178$. The initial sample size was chosen on the assumption that the rates of one-vote overstatements and understatements would be 0.001, rounded up to the nearest whole number, and that the rates of two-vote overstatements and understatements would be negligible. That led the auditors to anticipate one 1-vote overstatement and one 1-vote understatement in the sample. Expression [1] with $o_1 = 1$ and $u_1 = 1$ yields

$$n \geq (4.8 + 1.4 \times (1 - 0.6 \times 1))/0.027 = 198.5. \qquad (4)$$

Expression [1] rounds to the nearest tenth but the auditTools page does not; the initial sample was 198 ballots. (To allow for a one-vote overstatement without any compensating one-vote understatement, the initial sample size would be 230 instead: When $o_1 = 1$ and $u_1 = o_1 = o_2 = 0$, $n \geq (4.8 + 1.4 \times 1)/0.027$, giving an initial sample size $n \geq 229.6$.)

Each of the four people present contributed two digits to a seed, which was was used with the PRNG on the auditTools page to generate 198 numbers between 1 and 7,120, the number of ballots. Auditors retrieved each of the corresponding ballots using the manifest and the lookup tool on the auditTools page. Their manual interpretation of each ballot matched Kai Wang's interpretation, so the audit stopped, transitively confirming the official winners of both contests at 10% risk limit by looking at 198 ballots.

VII. Discussion

Risk-limiting audits guarantee that if the vote tabulation system found the wrong winner, there is a large chance of a full hand count to correct the results. Providing this guarantee requires a voting system that produces a voter-verifiable paper record—an audit trail—and requires the local election official to ensure that the audit trail remains complete and accurate. Risk-limiting audits examine portions of the audit trail by hand until there is sufficiently strong evidence that a full hand count would confirm the reported result, or until there has been a full hand count.

There are two general types of risk-limiting audits: *ballot-polling audits* and *comparison audits*. Both types are most efficient when the audit checks individual ballots, *ballot-level auditing*. For both, sample size depends on the margin (or diluted margin) and the luck of the draw—the particular ballots that happen to be in the sample—but only weakly on the size of the contest. Comparison audit sample sizes also depend on the number and nature of errors in the original tally.

Ballot-polling audits require almost nothing but the audit trail and a list of reported winners. In contrast, ballot-level comparison audits require detailed information from the vote tabulation system: its interpretation of each ballot. However, ballot-level comparison audits examine fewer ballots than ballot-polling audits when the margin is small and the outcome is correct: The number grows like the reciprocal of the margin, versus the square of the reciprocal for ballot-polling audits. At 10% risk limit, assuming the vote tabulation system is perfectly accurate, the ballot-polling method we presented would be expected to examine 120 ballots if the winner's share is 60%, 480 if it is 55%, or 3,860 if it is 52%, versus 24, 48, and 120 for the comparison audit method we presented.

Unfortunately, current commercial vote tabulation systems do not report their interpretation of each ballot, so ballot-level comparison audits sometimes rely on unofficial systems, giving *transitive audits*. Ballot-polling audits may be immediately practical for large contests, because they require so little of the vote tabulation system, and the counting burden typically is spread across many jurisdictions.

These auditing methods require random samples, which must be drawn properly, in a way that precludes manipulation, and ideally in a way that the public can verify is proper. Using a high-quality public pseudo-random number generator with a "seed" generated at random by audit participants satisfies these requirements.

While the mathematics that underlie risk-limiting audits might be daunting, the calculations required to conduct the audit can be extremely simple: arithmetic that could easily be done with pencil and paper or a four-function calculator. Simplicity improves transparency and can increase public confidence by allowing anyone interested to check the calculations.

References

Benaloh, J., Jones, D., Lazarus, E., Lindeman, M., and Stark, P. (2011). SOBA: Secrecy-preserving observable ballot-level audits. In *Proceedings of the 2011 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '11)*. USENIX.

Calandrino, J., Halderman, J., and Felten, E. (2007). Machine-assisted election auditing. In *Proceedings of the 2007 USENIX/ACCURATE Electronic Voting Technology Workshop (EVT 07)*. USENIX.

Cary, D. (2011). Estimating the margin of victory for instant-runoff voting. In *Proceedings of the 2011 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '11)*. USENIX.

Cordero, A., Wagner, D., and Dill, D. (2006). The role of dice in election audits – extended abstract. In *IAVoSS Workshop On Trustworthy Elections (WOTE 2006)*.

Higgins, M., Rivest, R., and Stark, P. (2011). Sharper p-values for stratified post-election audits. *Statistics, Politics, and Policy*, 2(1).

Johnson, K. (2004). Election certification by statistical audit of voter-verified paper ballots. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=640943. Retrieved 6 March 2011.

Magrino, T., Rivest, R., Shen, E., and Wagner, D. (2011). Computing the margin of victory in IRV elections. In *Proceedings of the 2011 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '11)*. USENIX.

Rivest, R. (2008). On the notion of 'software independence' in voting systems. *Phil. Trans. R. Soc. A*, 366(1881):3759–3767.

Rivest, R. and Wack, J. (2006). On the notion of "software independence" in voting systems (draft version of july 28, 2006). Technical report, Information Technology Laboratory, National Institute of Standards and Technology. http://vote.nist.gov/SI-in-voting.pdf Retrieved April 20, 2011.

Stark, P. (2008a). Conservative statistical post-election audits. *Ann. Appl. Stat.*, 2:550–581.

Stark, P. (2008b). A sharper discrepancy measure for post-election audits. *Ann. Appl. Stat.*, 2:982–985.

Stark, P. (2009a). CAST: Canvass audits by sampling and testing. *IEEE Transactions on Information Forensics and Security, Special Issue on Electronic Voting*, 4:708–717.

Stark, P. (2009b). Efficient post-election audits of multiple contests: 2009 California tests. http://ssrn.com/abstract=1443314. 2009 Conference on Empirical Legal Studies.

Stark, P. (2010a). Risk-limiting vote-tabulation audits: The importance of cluster size. *Chance*, 23(3):9–12.

Stark, P. (2010b). Super-simple simultaneous single-ballot risk-limiting audits. In *Proceedings of the 2010 Electronic Voting Technology Workshop / Workshop on Trustworthy Elections (EVT/WOTE '10)*. USENIX. http://www.usenix.org/events/evtwote10/tech/full_papers/Stark.pdf. Retrieved April 20, 2011.

Stark, P. B. and Wagner, D. A. (2012). Evidence-based elections. *IEEE Security and Privacy*, page submitted.

Wald, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat.*, 16:117–186.

PLACE PHOTO HERE

Mark Lindeman is a political scientist whose research includes public opinion, political behavior, and election verification issues. He was an executive editor of the widely endorsed Principles and Best Practices for Post-Election Audits, and has participated extensively in conferences and other discussions focused on verification methods. He is co-author of Public Opinion (Westview).

PLACE PHOTO HERE

Philip B. Stark is Professor of Statistics, University of California, Berkeley. He served on the 2007 California Post Election Audit Standards Working Group and designed and conducted the first risk-limiting post election audits. He is working with the California and Colorado Secretaries of State on pilot risk-limiting audits. For a more complete biography, see http://statistics.berkeley.edu/~stark/bio.pdf.

Some Assumptions for the Douglas County RLA Time Study

There are several different roles in the RLA process that need to be defined. These roles may expand due to the number of ballots required to audit, which the calculator will figure for us.

1. **Random Ballot Sampler** - This role will utilize the SoS tool to distribute box, batch, and ballot numbers to the Ballot Pulling Teams for the ballot to audit. This will consist of one person.
2. **Box Custodian** - This role is to ensure that chain of custody is achieved, and is tasked with pulling the batches of ballots while ensuring that no batches get mixed up with each other. The Box Custodian can be multiple, and is flexible based on the number of Ballot Pulling Teams.
3. **Ballot Pulling Team** - The Audit Teams consist of a bi-partisan team who receive the batch/ballot required to audit from the Box Custodian, and pull the ballot for audit. There will typically be more than one team.
4. **Auditor** - This is the person who enters the result that the contest requires in to the SOS tool to determine whether more ballots need to be pulled or not.