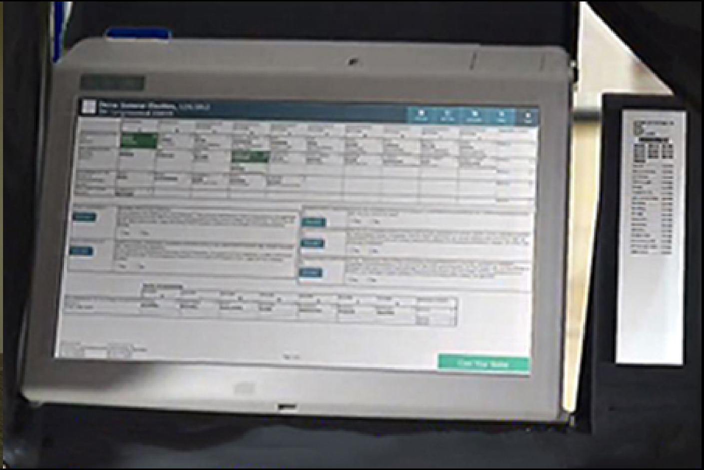
# Testing Cannot Tell Whether Ballot-Marking Devices Alter Election Outcomes

Institute for Social Research
Center for Political Studies
University of Michigan

Philip B. Stark and Ran Xie

30 September 2020

University of California, Berkeley




00000111
ELECTION: GENERAL ELECTION
ELECTION DATE: 11/08/2016
PRECINCT: 740-050 SPG 15 050

U.S. PRESIDENT AND U.S. VICE PRESI-REP DONALD J. TRUMP
U.S. SENATE--------------------REP SENATOR JOHN BOOZMAN
U.S. CONGRESS DISTRICT 03-REP CONGRESSMAN STEVE WOMACK
STATE REPRESENTATIVE DISTRICT 89 DIS-REP JEFF WILLIAMS
COUNTY JUDGE------------------------REP JOSEPH K. WOOD
COUNTY ASSESSOR---------------REP ASSESSOR RUSSELL HILL
CIRCUIT CLERK---------REP CIRCUIT CLERK KYLE SYLVESTER
CONSTABLE DISTRICT 01 DISTRI-REP CONSTABLE JOHN DUGGAR
MAYOR SPRINGDALE----------------------MAYOR DOUG SPROUSE
CLERK/TREASURER SPRINGDALE-----------NO SELECTION MADE
ALDERMAN WARD 1 POSITION 2 SPRINGDALE---------MIKE FORD
ALDERMAN WARD 2 POSITION 2 SPRINGDALE-------RAY DOTSON
ALDERMAN WARD 3 POSITION 2 SPRINGDALE------JEFF WATSON
ALDERMAN WARD 4 POSITION 2 SPRINGDALE-----JEREMY LYNCH
ISSUE NO. 1-------------------------------FOR ISSUE NO. 1
ISSUE NO. 2-------------------------------FOR ISSUE NO. 2
ISSUE NO. 3----------------------------NO SELECTION MADE
ISSUE NO. 4-------------------------------FOR ISSUE NO. 4
ISSUE NO. 5-------------------------------FOR ISSUE NO. 5
ISSUE NO. 6-------------------------------FOR ISSUE NO. 6
ISSUE NO. 7-------------------------------FOR ISSUE NO. 7
UNOPPOSED CANDIDATES-----------------NO SELECTION MADE

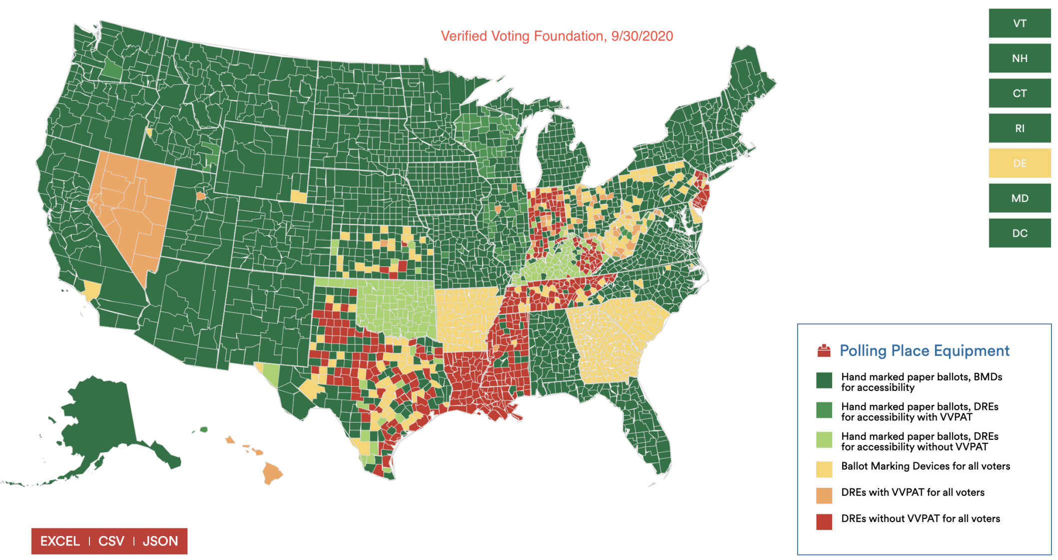
# IMAGECAST® X | BMD


**The ImageCast® X can be configured as a Ballot Marking Device (BMD), which is paired with a commercially available, compact laser printer that prints a summary of the voters selections.** No votes are stored on the ImageCast® X when in the BMD configuration.

**Get in touch**
1.866.654.VOTE (8683)
sales@dominionvoting.com
www.dominionvoting.com

DOMINION VOTING


DEMONSTRATION BALLOT
VSAP Demonstration Election
VSAP
Voting Solutions for All People
County of Los Angeles


Verified Voting Foundation, 9/30/2020
VT
NH
CT
RI
DE
MD
DC
Polling Place Equipment
Hand marked paper ballots, BMDs for accessibility
Hand marked paper ballots, DREs for accessibility with VVPAT
Hand marked paper ballots, DREs for accessibility without VVPAT
Ballot Marking Devices for all voters
DREs with VVPAT for all voters
DREs without VVPAT for all voters
EXCEL | CSV | JSON

# Can Voters Detect Malicious Manipulation of Ballot Marking Devices?

Matthew Bernhard, Allison McDonald, Henry Meng, Jensen Hwa, Nakul Bajaj*, Kevin Chang, J. Alex Halderman

University of Michigan *The Harker School

***Abstract*—Ballot marking devices (BMDs) allow voters to select candidates on a computer kiosk, which prints a paper ballot that the voter can review before inserting it into a scanner to be tabulated. Unlike paperless voting machines, BMDs provide voters an opportunity to verify an auditable physical record of their choices, and a growing number of U.S. jurisdictions are adopting them for all voters. However, the security of BMDs depends on how reliably voters notice and correct any adversarially induced errors on their printed ballots. In order to measure voters' error detection abilities, we conducted a large study (N = 241) in a realistic polling place setting using real voting machines that we modified to introduce an error into each printout. Without intervention, only 40% of participants reviewed their printed ballots at all, and only 6.6% told a poll worker something was wrong. We also find that carefully designed interventions can improve verification performance. Verbally instructing voters to review the printouts and providing a written slate of candidates for whom to vote both significantly increased review and reporting rates—although the improvements may not be large enough to provide strong security in close elections, especially when BMDs are used by all voters. Based on these findings, we make several evidence-based recommendations to help better defend BMD-based elections.**

However, BMDs do not eliminate the risk of vote-stealing attacks. Malware could infect the ballot scanners and change the electronic tallies—although this could be detected by rigorously auditing the paper ballots [50]—or it could infect the BMDs themselves and alter what gets printed on the ballots. This latter variety of cheating cannot be detected by a post-election audit, since the paper trail itself would be wrong, and it cannot be ruled out by pre-election or parallel testing [51]. Instead, BMD security relies on voters themselves detecting such an attack. This type of human-in-the-loop security is necessary in many systems where detection and prevention of security hazards cannot be automated [18]. However, as several commentators have recently pointed out [7], [20], [51], its effectiveness in the context of BMDs has not been established.

Whether such a misprinting attack would succeed without detection is highly sensitive to how well voters verify their printed ballots. Every voter who notices that their ballot is misprinted and asks to correct it *both* adds to the evidence that there is a problem *and* requires the attacker to change an additional ballot in order to overcome the margin of victory.


Fig. 3: *Warning Signage.* One of the interventions we tested was placing a sign above the scanner that instructed voters to verify their ballots. Signage was not an effective intervention.

| Experiment | *N* | Were observed examining ballot | Reported error on exit survey | Reported error to poll worker |
|---|---|---|---|---|
| *Without interventions:* | | | | |
| **E1**: Regular ballots | 31 | 41.9% | 6.5% | 6.5% |
| **E2**: Summary ballots | 31 | 32.3% | 6.5% | 6.5% |
| **E3**: Deselection only | 29 | 44.8% | 10.3% | 6.9% |
| Subtotal/Mean | 91 | 39.7% | 7.8% | 6.6% |
| *With interventions:* | | | | |
| **E4**: Signage | 30 | 13.3% | 3.3% | 6.7% |
| **E5**: Script variant 1 | 30 | 46.7% | 13.3% | 6.7% |
| **E6**: Script variant 2 | 25 | 92.0% | 16.0% | 16.0% |
| **E7**: Script variant 3 | 31 | 38.7% | 19.4% | 12.9% |
| **E8**: Slate with script variant 2 | 13 | 100.0% | 38.5% | 38.5% |
| **E9**: Slate with script variant 3 | 21 | 95.2% | 71.4% | 85.7% |
| Subtotal/Mean | 150 | 64.3% | 24.0% | 27.8% |

# Voter Verification of BMD Ballots Is a Two-Part Question: Can They? Mostly, They Can. Do They? Mostly, They Don't

Philip Kortum, Michael D. Byrne, and Julie Whitmore

Rice University, Houston, Texas

**ABSTRACT**

The question of whether or not voters actually verify ballots produced by ballot marking devices (BMDs) is presently the subject of some controversy. Recent studies (e.g., Bernhard, et al. 2020) suggest the verification rate is low. What is not clear from previous research is whether this is more a result of voters being unable to do so accurately or whether this is because voters simply choose not to attempt verification in the first place. In order to understand this problem, we conducted an experiment in which 108 participants participated in a mock election where the BMD displayed the voters' true choices, but then changed a subset of those choices on the printed ballot. The design of the printed ballot, the length of the ballot, the number of changes that were made to the ballot, the location of those changes, and the instructions provided to the voters were manipulated as part of the experiment. Results indicated that of those voters who chose to examine the printed ballot, 76% detected anomalies, indicating that voters *can* reliably detect errors on their ballot if they will simply review it. This suggests that administrative remedies, rather than attempts to alter fundamental human perceptual capabilities, could be employed to encourage voters to check their ballots, which could prove as an effective countermeasure.

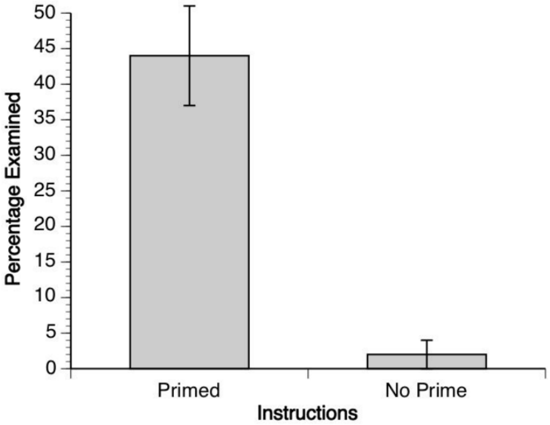



Figure 4. Percentage of voters who examined their ballot as a function of whether or not they were primed to do so by instructions and the poll worker. Error bars represent one standard error of the mean.

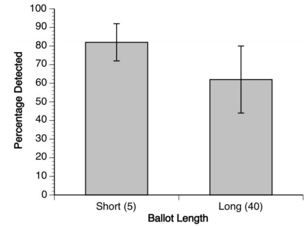



Figure 6. Percentage of voters who detected one or more anomalies as a function of the length of the ballot. “Short” ballots had 5 contests, “long” ballots had 40 contests. Error bars represent one standard error of the mean.

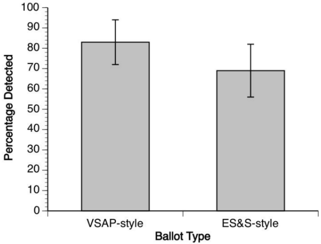



Figure 7. Percentage of voters who detected one or more anomalies as a function of which style of printout they received. Error bars represent one standard error of the mean.

# Northampton officials unanimously vote 'no confidence' in ExpressVote XL voting machine

Emily Previti/PA Post

DECEMBER 20, 2019 | 9:51 AM

EASTON – Northampton County Election Commissioners unanimously supported a "vote of no confidence" in the county's new voting machines after vendor Election Security & Software presented findings Thursday night from an investigation into tabulation errors and other problems when the system debuted.

The incorrect tallies in last month's election were linked to races with cross-filed candidates and straight-ticket ballots cast by voters. Cross-filed candidates are ones seeking an office on more than one party line, while

SIGN IN NPR SHOP DONATE NOW

NEWS ARTS & LIFE MUSIC SHOWS & PODCASTS SEARCH

NATIONAL

# Lawsuits Over Voting Machines In Pennsylvania

January 19, 2020 · 8:02 AM ET
Heard on Weekend Edition Sunday

EMILY PREVITI

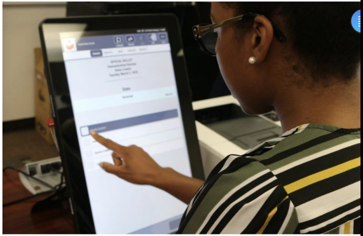
FROM witf

3-Minute Listen

+ PLAYLIST

After an Election Day meltdown last year, two lawsuits in Pennsylvania could result in the state decertifying a popular voting machine ahead of of the 2020 elections.

# Error discovered on Georgia touchscreens in US Senate race



GEORGIA POLITICS | 1 hour ago

By Mark Niesse, The Atlanta Journal-Constitution

**Election officials working to correct issue before early voting begins Oct. 12**



Georgia election officials said Saturday they discovered a programming error on the state's voting touchscreens that caused a row of candidates in the 20-candidate U.S. Senate special election to disappear at times when flipping back and forth between screens.

Washington Monthly

# Donald Trump's Favorite Voting Machines

**Ballot-marking devices in key swing states could give him the perfect excuse to contest the election**

by Art Levine September 23, 2020 POLITICS


D-Suite 5.5 EAC
DOMINION VOTING
DELL
D-Suite 5.5 EAC
DOMINION VOTING
DELL
ICC01
Tabulator 950


Activation
Codes Export
People
Photos
S
Settings
Skype
Spotify Music
Sticky Notes
T
Tips
V
Video Editor
Voice Recorder
W
Weather
Windows Accessories
Windows Administrative Tools
Windows Defender Security Center
Windows Ease of Access
Wednesday
2
See all your mail in one place
And more...
Mail
Xbox Con...
Photos
My Office
OneNote
FRIENDS
Microsoft
Solitaire Collection
Explore
Microsoft Store
Microsoft Edge
ASPHALT
STREETSTORM
Seekers Notes...
Weather
Spotify Music
Skype
Microsoft News
Fitbit Coach
ACG Player

## Why test BMDs?

- BMDs can print votes that differ from those confirmed onscreen or through audio interface.
- “voter-verifiability” & ability to spoil ballot don’t solve the problem.
- Voters don’t check much, and can’t check perfectly
- In effect, BMDs make the paper trail hackable: undermine audits

## Georgia Attorney General Contest, 2018

Official results:

```
CHRIS CARR (I) (REP) 51.30%  1,981,563
CHARLIE BAILEY (DEM) 48.70%  1,880,807
```

margin: 2.6%

ballots cast: 3,949,905

votes cast in Fulton County: 415,524

| voting method | detection rate[1] | hack rate[2] | do-over rate[3] | Fulton share[4] |
|---|---|---|---|---|
| all BMD | 6.6% | 0.014 | <0.001 | 374 |
| | 20% | 0.016 | 0.003 | 1325 |
| | 76% | 0.053 | 0.040 | 16,782 |
| 50% BMD | 6.6% | 0.027 | 0.002 | 374 |
| | 20% | 0.032 | 0.006 | 1325 |
| | 76% | 0.106 | 0.081 | 16,782 |
| 5% BMD | 6.6% | 0.273 | 0.018 | 374 |
| | 20% | 0.319 | 0.064 | 1325 |
| | 76% | 1 | 0.808 | 16,782 |

[1] Rate at which voters who use BMDs notice printout errors and request a new chance to mark a ballot.
[2] Error rate in BMD printouts sufficient to change the reported winner.
[3] Among voters who use BMDs, the fraction who request a fresh chance to mark a ballot.
[4] If the errors were spread evenly across counties, the number of do-over requests in Fulton County.

## The BMD security model is broken

- BMDs make voters responsible for BMD security
- but BMDs don't give voters the tools they need to do that job
- no way for voter to prove BMD misbehaved
- LEO can't tell whether voter's complaint is BMD malfunction, voter error, or "wolf"
- error or malfeasance could change a large percentage of votes without raising an alarm

# Claimed benefits of BMDs

- prevent overvotes
- warn of undervotes
- eliminate ambiguous marks

# But . . .

- Assume BMDs function correctly!
  - Many recent examples of failures, including Georgia, Northampton PA, Los Angeles CA
- PCOS can also protect against undervotes and overvotes—required by VVSG 1.0

# Can we establish that BMDs worked in a given election?

- need to know errors didn't change any outcomes
- 3 approaches proposed:
  - pre-election logic and accuracy (L&A) testing
  - "passive" testing
  - "live" or "parallel" testing
- this research: none of these can work in practice

## How much testing is enough?

- depends on the size of the problem deemed "material."
- sensible threshold: "enough to change the reported winner of one or more contests"
- many contests are decided by less than 1%
- margin in the 2016 U.S. presidential election was 0.22% in MI, 0.37% in RI, 0.72% in PA, and 0.76% in WI

# Auditing as an adversarial game

- Mallory seeks to alter the outcome of one or more contests in an election.
  - M does not want to be detected.
  - M knows the testing strategy
  - M knows the state history of each machine
  - M has a good model of voter behavior
- Pat seeks to ensure that any BMD problem that alters one or more outcomes will be detected.
  - P must obey the law and protect voter privacy.
  - P does not know which contest(s) M will attack nor M's strategy.

## Jurisdiction sizes, contest sizes, margins

Important contests have sizes ranging from dozens of eligible voters to millions of eligible voters.

- median turnout in the 3017 U.S. counties in 2018 was 2,980 voters,
- less than 43,000 voters for more than 2/3 of jurisdictions
- In 73% of states, more than 50% of counties have fewer than 30,000 active voters.
- In 92% of states, >50% of counties have fewer than 100,000 active voters.
- in 2019, 317 U.S. cities had populations of 100,000 or more, out of over 19,500 incorporated places
  - if 80% of the population is of voting age & turnout is 55%, contests for elected officials in 98% of incorporated places involve fewer than $44,000$ voters.
- 2019 median population of U.S. incorporated areas is 725: ~50% of the 19,500 incorporated places have turnout $<=$320 voters.

## 2018 turnout by county

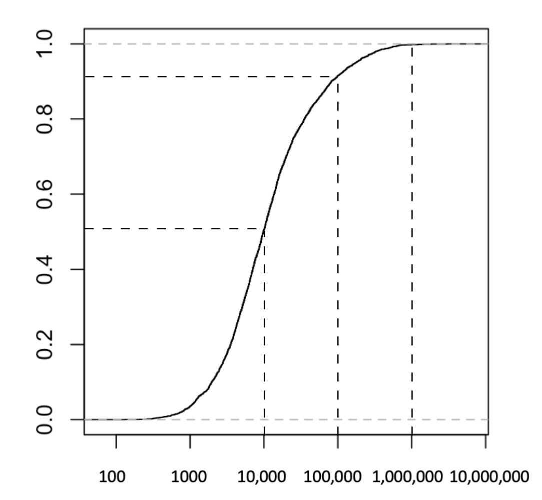



**Figure 1:** Total participation on election day per jurisdiction in 3073 counties in 2018 [@EAVS2018]. Counties ordered from small to large, plotted against total voter turnout.

# 2018 median turnout by jurisdiction

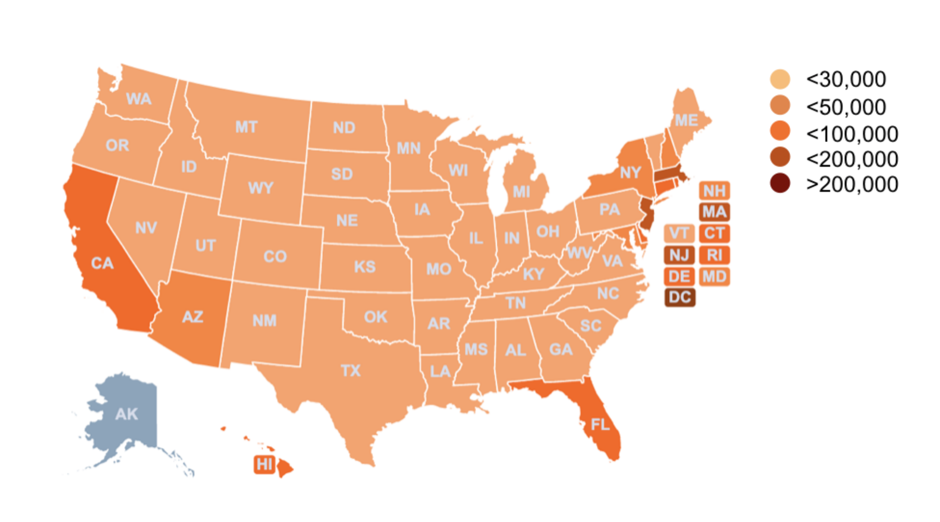



**Figure 2:** Heat map of median 2018 turnout by jurisdiction in the 50 U.S. states and Washington, DC. [@EAVS2018]

## Mallory's strategy space: alter any collection of transactions

- time of day the transaction starts
- the time since the previous voter finished using the BMD (a measure of how busy the polling place is)
- the number of voting transactions before the current transaction
- the voter's sequence of selections in each contest, including undervotes, before going to the next selection
- the number of times the voter changes selections in each contest in the first pass through the ballot, and what the voter changed the selection from and to, etc.
- the amount of time the voter takes to make each selection before taking another action (e.g., going to the next contest)
- whether the voter looks every page of candidates in a contest

- how much time (if any) the voter takes to review selections, which selections the voter changes, etc.
- whether the voter receives an inactivity warning during voting
- what part of each onscreen voting target the voter touches
- BMD settings, including font size, language, whether the audio interface is used, volume setting, tempo setting, whether voter pauses the audio, whether voter "rewinds," and whether the voter uses audio only or synchronized audio/video
- whether voter uses sip-and-puff interface

## Possible voting transactions

| Parameter | optimistic | more realistic |
|---|---|---|
| Contests | 3 | 20 |
| Candidates per Contest | 2 | 4 |
| Languages | 2 | 13 |
| Time of day | 10 | 20 |
| Number of previous voters | 5 | 10 |
| Undervotes | $2^3$ | $2^{20}$ |
| Changed selections | $2^3$ | $2^{20}$ |
| Review | 2 | 2 |
| Time per selection | 2 | $5^{20}$ |

| Parameter | optimistic | more realistic |
|---|---|---|
| Contrast/saturation | - | 4 |
| Font Size | 2 | 4 |
| Audio Use | 2 | 2 |
| Audio tempo | - | 4 |
| Volume | 5 | 10 |
| Audio pause | - | $2^{20}$ |
| Audio + video | - | 2 |
| Inactivity warning | 2 | $2^{20}$ |
| Total combinations | $6.14 \times 10^6$ | $1.2 \times 10^{47}$ |

## Pat's strategy space

- Monitor voter behavior, e.g., spoiled ballot rates

## Pat’s strategy space

- Monitor voter behavior, e.g., spoiled ballot rates
- Try to catch a malfunction by using the BMD before, during, or after an election

## Randomness is key

- if Pat’s tests are predictable, Mallory can just change other transactions (passive testing doesn’t solve this)

## Randomness is key

- if Pat's tests are predictable, Mallory can just change other transactions (passive testing doesn't solve this)
- can't just set aside machines: Dieselgate

## Randomness is key

- if Pat's tests are predictable, Mallory can just change other transactions (passive testing doesn't solve this)
- can't just set aside machines: Dieselgate
- uniform random sampling is doomed

# Randomness is key

- if Pat's tests are predictable, Mallory can just change other transactions (passive testing doesn't solve this)
- can't just set aside machines: Dieselgate
- uniform random sampling is doomed
- "ideal" sampling would mimic voter behavior

## Randomness is key

- if Pat's tests are predictable, Mallory can just change other transactions (passive testing doesn't solve this)
- can't just set aside machines: Dieselgate
- uniform random sampling is doomed
- "ideal" sampling would mimic voter behavior
- examine "oracle bounds" and "learning" distribution of transactions

## How many votes must be altered to alter the outcome?

- Altering votes on 1% of transactions in a jurisdiction can change the margin of contests that are not jurisdiction-wide by far more than 2%.

## How many votes must be altered to alter the outcome?

- Altering votes on 1% of transactions in a jurisdiction can change the margin of contests that are not jurisdiction-wide by far more than 2%.
- If a contest is on 10% of ballots & undervote rate in the contest is 30%, altering votes on 1% of transactions can change margin in that particular contest by 29%.

# Passive testing

- rely on voters to test
- use spoiled ballot rate to signal a possible problem
- need to set alarm threshold to balance false alarms and missed problems
- may depend on things that vary from election to election:
  - number of contests on the ballot
  - whether the contests have complex voting rules
  - ballot layout
  - voter demographics
  - turnout
  - familiarity w voting technology
  - . . .

## Setting the threshold

- need to know something about the distribution of spoiled ballots when BMDs malfunction to control the false negative rate
- depends on the number of transactions Mallory alters, which voters are affected, which contests are affected, etc.
- Pat won't know any of those things

# Hypothetical example

- spoiled ballots follow Poisson distribution with known rate, absent hacking, and different known rate, given hacking. (Optimistic!)
- 7% or 25% of voters will notice errors and spoil their ballots
- contest margins of 1%–5% and false positive and false negative rates of 5% and 1%.

# 5% false negative & false positive rate

| margin | detection rate | 0.5% base rate | 1% base rate | 1.5% base rate |
|---|---|---|---|---|
| 1% | 7% | 451,411 | 893,176 | 1,334,897 |
| | 25% | 37,334 | 71,911 | 106,627 |
| 2% | 7% | 115,150 | 225,706 | 336,160 |
| | 25% | 9,919 | 18,667 | 27,325 |
| 3% | 7% | 52,310 | 101,382 | 150,471 |
| | 25% | 4,651 | 8,588 | 12,445 |
| 4% | 7% | 30,000 | 57,575 | 85,227 |
| | 25% | 2,788 | 4,960 | 7,144 |
| 5% | 7% | 19,573 | 37,245 | 54,932 |
| | 25% | 1,838 | 3,274 | 4,689 |

# 1% false negative & false positive rate

| margin | detection rate | 0.5% base rate | 1% base rate | 1.5% base rate |
|---|---|---|---|---|
| 1% | 7% | 908,590 | 1,792,330 | 2,675,912 |
| | 25% | 76,077 | 145,501 | 214,845 |
| 2% | 7% | 233,261 | 454,295 | 675,242 |
| | 25% | 20,624 | 38,039 | 55,442 |
| 3% | 7% | 106,411 | 204,651 | 302,864 |
| | 25% | 9,870 | 17,674 | 25,359 |
| 4% | 7% | 61,385 | 116,631 | 171,908 |
| | 25% | 5,971 | 10,312 | 14,681 |
| 5% | 7% | 40,156 | 75,671 | 110,989 |
| | 25% | 4,036 | 6,849 | 9,650 |

## Sanity check

- 41 of California's 58 counties had fewer than 100,000 voters in the 2018 midterm election

## Sanity check

- 41 of California's 58 counties had fewer than 100,000 voters in the 2018 midterm election
- 33 had fewer than 100,000 voters in the 2016 Presidential primary election

## Sanity check

- 41 of California's 58 counties had fewer than 100,000 voters in the 2018 midterm election
- 33 had fewer than 100,000 voters in the 2016 Presidential primary election
- passive testing could not protect contests with margins of 3% or smaller.

## Sanity check

- 41 of California's 58 counties had fewer than 100,000 voters in the 2018 midterm election
- 33 had fewer than 100,000 voters in the 2016 Presidential primary election
- passive testing could not protect contests with margins of 3% or smaller.
- In many California counties, turnout is so small even in statewide contests that there would be no way to detect problems through spoilage rates reliably without high rate of false alarms.

## Sanity check

- 41 of California's 58 counties had fewer than 100,000 voters in the 2018 midterm election
- 33 had fewer than 100,000 voters in the 2016 Presidential primary election
- passive testing could not protect contests with margins of 3% or smaller.
- In many California counties, turnout is so small even in statewide contests that there would be no way to detect problems through spoilage rates reliably without high rate of false alarms.
- If turnout is roughly 50%, contests in jurisdictions with fewer than 60,000 voters (which includes 23 of California's 58 counties could not in principle limit chance of false positives & of false negatives to 5% for margins below 4%—even under these optimistic assumptions and simplifications.

# Targeting the attack

- assumed all voters are equally likely to detect discrepancies
- Mallory has access to each BMD's settings, state history, etc.
- can select whose votes to alter, inferring voter characteristics from BMD settings and the voters' interaction with the BMD.
- can target voters less likely to notice problems (&perhaps less likely to be believed if they report malfunctions)

## Voters with visual impairments

- ~0.8% of the U.S. population is legally blind; approximately 2% of Americans age 16 to 64 have a visual impairment.
- Current BMDs do not provide voters with visual impairments a way to check whether the printout matches their selections
- If 2% of voters have a visual impairment that prevents them from checking the printout and Mallory only alters votes when the voter uses the audio interface or large fonts, Mallory might be able to change the outcomes of contests with jurisdiction-wide margins of 4% or more wo increasing the spoiled ballot rate.

## Voters with motor impairments

- Some BMDs let voters print & cast a ballot without looking at it, e.g. ES&S ExpressVote® with "Autocast,"
- Voters who use this feature have no opportunity to check whether the printout matches their selections nor to spoil the ballot if there is a discrepancy.
- Mallory can change every vote cast using this feature without increasing the spoiled ballot rate.

# Voters who use languages other than English

- Federal law requires some jurisdictions to provide ballots in languages other than English.
- In 2013, ~26% of voters in Los Angeles County spoke a language other than English at home
- If a substantial percentage of voters use foreign-language ballots and are unlikely to check the English-language printout, Mallory could change the outcome of contests with large margins without increasing the spoiled ballot rate noticeably.

## Voters in a hurry, et al.

- Mallory can monitor how long it takes voters to make their selections, whether they change selections, how long they review the summary screen, etc.

## Voters in a hurry, et al.

- Mallory can monitor how long it takes voters to make their selections, whether they change selections, how long they review the summary screen, etc.
- A voter who spends little time reviewing selections onscreen also might be unlikely to review the printout carefully.

# Voters in a hurry, et al.

- Mallory can monitor how long it takes voters to make their selections, whether they change selections, how long they review the summary screen, etc.
- A voter who spends little time reviewing selections onscreen also might be unlikely to review the printout carefully.
- If a voter takes a very long time to mark a ballot or changes selections repeatedly, might be a sign that the voter finds voting difficult or confusing; such voters might also be unlikely to notice errors in the printout.

# FUD attacks on passive testing

- passive testing using the spoiled ballot rate does not produce direct evidence of malfeasance or malfunction
- does not identify which ballots and which contests, if any, have errors
- does not provide any evidence about whether the errors, if any, changed outcomes
- opens the door to a simple, legal way to undermine elections: ask voters to spoil ballots.

## Oracle bounds: "shoulder surfing"

- suppose Pat could ask an oracle whether a particular BMD printout had an error (equivalently, suppose Pat can watch over the shoulder of selected voters as they use the BMD)

## Oracle bounds: "shoulder surfing"

- suppose Pat could ask an oracle whether a particular BMD printout had an error (equivalently, suppose Pat can watch over the shoulder of selected voters as they use the BMD)
- contest w 2980 voters (2018 median jurisdiction turnout). Mallory alters 15 transactions. Could chance contest outcome by 1% or more.

## Oracle bounds: "shoulder surfing"

- suppose Pat could ask an oracle whether a particular BMD printout had an error (equivalently, suppose Pat can watch over the shoulder of selected voters as they use the BMD)
- contest w 2980 voters (2018 median jurisdiction turnout). Mallory alters 15 transactions. Could chance contest outcome by 1% or more.
- Pat would need to spy on $n = 540$ voters, about 18%. Involves testing each BMD several times per hour.

## Oracle bounds: "shoulder surfing"

- suppose Pat could ask an oracle whether a particular BMD printout had an error (equivalently, suppose Pat can watch over the shoulder of selected voters as they use the BMD)
- contest w 2980 voters (2018 median jurisdiction turnout). Mallory alters 15 transactions. Could chance contest outcome by 1% or more.
- Pat would need to spy on $n = 540$ voters, about 18%. Involves testing each BMD several times per hour.
- for once-an-hour testing per machine to give 95% chance of catching problem, need $>$6,580 voters in the contest, almost triple the median turnout in jurisdictions across the U.S., and roughly 20 times the median number of active voters in incorporated areas in the U.S.

# Modeling voter behavior

- Pat can't really shoulder surf: needs to model voter behavior

# Modeling voter behavior

- Pat can't really shoulder surf: needs to model voter behavior
- requires complete monitoring of surprisingly many voters

## Minimax lower bounds

Pat draws an IID training sample of $n$ transactions from $P$

| Confidence | Test Limit | Altered Votes | Bound (millions) |
|---|---|---|---|
| 99% | Inf | 0.5% | 3.73 |
| 99% | Inf | 1% | 3.46 |
| 99% | Inf | 3% | 2.61 |
| 99% | Inf | 5% | 2.04 |
| 95% | Inf | 0.5% | 1.65 |
| 95% | Inf | 1% | 1.57 |
| 95% | Inf | 3% | 1.29 |
| 95% | Inf | 5% | 1.08 |

| Confidence | Test Limit | Altered Votes | Bound (millions) |
|---|---|---|---|
| 99% | 2000 | 0.5% | 3.87 |
| 99% | 2000 | 1% | 3.58 |
| 99% | 2000 | 3% | 2.69 |
| 99% | 2000 | 5% | 2.09 |
| 95% | 2000 | 0.5% | 1.67 |
| 95% | 2000 | 1% | 1.59 |
| 95% | 2000 | 3% | 1.31 |
| 95% | 2000 | 5% | 1.10 |

# Complications and frustrations

- the only remedy is a new election

## Complications and frustrations

- the only remedy is a new election
- margins are not known when testing happens

## Complications and frustrations

- the only remedy is a new election
- margins are not known when testing happens
- tests have uncertainty {#sec:uncertain}

# Complications and frustrations

- the only remedy is a new election
- margins are not known when testing happens
- tests have uncertainty {#sec:uncertain}
- requires new systems, extra hardware, additional staff, training w

## Complications and frustrations

- the only remedy is a new election
- margins are not known when testing happens
- tests have uncertainty {#sec:uncertain}
- requires new systems, extra hardware, additional staff, training w
- BMDs will still pose special risks of disenfranchising some voters