Submission to the JSCEM Inquiry into the 2022 Federal Election

A/Prof. Vanessa Teague
Thinking Cybersecurity
and the Australian National University

[redacted]

Prof. Philip B. Stark
University of California, Berkeley, USA

Dr. Michelle Blom
University of Melbourne

Prof. Peter J. Stuckey
Monash University

Dr. Damjan Vukcevic
University of Melbourne

September 28, 2022

Summary: Australian Senate ballots are cast on paper, then digitized and electronically counted. The accuracy of the digitization process is therefore crucially important for the correctness of the election outcome. This year the AEC was required to sample ballot papers to verify that they had been accurately digitized. This represents a substantial improvement on prior years' election conduct.

The audit provided some evidence of an accurate election outcome in 5 states and 2 territories, because the measured error rate was low compared with the closeness of the result. It also identified a possible indication of a serious problem in Victoria, where the election was close and the measured error rate relatively large. Both these cases show that the *Electoral Legislation Amendment (Assurance of Senate Counting) Bill 2021* was a success. In this submission we recommend the next steps for examining and correcting the problems that were identified in Victoria, and explain some important improvements to be made in future audits.

We conclude the submission with an explanation of why voting over the Internet is not a genuine way of improving access to voting, and suggest alternative uses of technology in voting that have better security, privacy and verifiability.

Who we are: We are cryptographers, statisticians and election integrity researchers. We have previously written a more detailed explanation of how audits of the Australian Senate digitisation process should be conducted [BSS+22].

We would be happy to discuss any of these issues with the committee.

Contents

1. Introduction

This year for the first time, the AEC was obliged under the *Electoral Act 1918* to conduct an audit of the Senate ballot papers to verify that they had been accurately digitised.

The audit was a great success.

- In 5 states and 2 territories, the audit measured an error rate well below the level that could have affected the outcome.[1]
- In Victoria, the audit identified a serious problem: four ballot papers had no corresponding image in the image database. Although a similar incident occurred in NSW, the Victorian case is much more serious because the estimated number of such errors is large enough to alter the election outcome.

It is very important to realise that both of these findings represent a successful audit: the first group because they provide (some) evidence that the election outcome is correct, and the second because it identified a serious problem that might otherwise have gone undetected.

There were six incidents in which the auditors retrieved a ballot paper and were unable to find a matching image in the database. (In one other case, there was a matching image but it was not a good likeness.) One of the six seems to have an obvious physical explanation—an extra ballot was probably stuck behind another. The other five incidents, two from NSW and three from Victoria, remain unexplained.

Unfortunately, the AEC and its partners do not seem to have fully appreciated the seriousness of the problem, perhaps not realising how close the Victorian Senate race was (see Table 2). It would have been better if the public had been immediately notified, as we were in 2013 when a ballot box went missing in Western Australia. A thorough investigation could then have commenced immediately. However, it is not too late. There are many good reasons to re-examine the ballot papers to fully understand the problem.

- If the problem was caused by a software error, it can be identified and corrected before the next election.
- If the problem was caused by a mechanical failure (for example, a malfunctioning scanner), that equipment can be removed from service.
- If the problem affected ballots randomly, then a larger sample might show that there was no net political impact. Showing this publicly, after a thorough investigation, is important for earning public trust in the election outcome.

Obviously, there is a chance that further investigation will show that the errors possibly did affect the election result. We do not assert that this is likely. If that is the case, it is still better to thoroughly examine the evidence than to ignore the problem.

It is good when problems are noticed—now some further work needs to be done to understand the causes and the necessary corrections. We recommend a more thorough and direct investigation be conducted, by people with access to both the Senate ballot papers and the AEC image and preferences databases.

[1] Because of methodological errors in selecting the sample, the measured rates are not reliable estimates of the true error rates—see Appendix A.1 for an explanation of the changes needed for next time.

In this submission we describe the problem and its implications as accurately as we can, based on documents obtained via Freedom of Information requests. We first explain the documents we used and the basics of the audit process. (The AEC refers to it as an "assurance process" but the term 'post-election audit' is more consistent with international usage.) Section 2 describes the problem, then Section 3 outlines the steps recommended for a full technical investigation of the issue in 2022. Appendix A identifies some more general shortcomings in the conduct of the audit, and makes recommendations for improving the process in future. Section 4.1 addresses the Inquiry's term of reference related to encouraging increased electoral participation. We explain that voting by Internet is not a genuine method of enfranchisement because there is currently no effective solution for keeping votes private while allowing people to verify their votes. Section 4.2 sketches some possible alternative uses of technology that might improve access without undermining integrity.

1.1. Documents we used and their terminology

Our analysis is based on the following documents, which we list in the order they became available.

The AEC Process Description [Com22] This was released before the election and is rather confusingly titled "Methodology for the ballot paper sampling process for Senate election 2022," though it summarises the audit process, not the statistical sampling methodology.

The ABS Audit methodology [oS22] This was released in reponse to our Freedom of Information request after the election.[2]

The Electoral Commissioner's Senate ballot paper sampling outcomes statement [Rog22] This mentions that for 7 ballot papers, "the nominated image did not exactly reflect the physical ballot paper," but the Commissioner does not seem to have been aware that for 6 ballot papers, no nominated image could be found.

Axiom's audit report [Ass22] *External statistical assurance of the Senate ballot papers Detailed report and statement of outcomes (draft).* This was released in response to our Freedom of Information request,[3] together with error examples.[4]

1.2. The Senate scanning and digitisation process

The Senate ballot process has several discrete steps relevant to the audit. First, ballots are sorted by first preference in the polling place. The first-preference tallies are manually counted. Next, ballot papers are brought to a central counting station and scanned to produce a digital image of the ballot paper. This digital image is then converted into digital preferences by a combination of optical character recognition

[2] https://www.righttoknow.org.au/request/8988/response/26588/attach/8/Relevant%20Document%20ABS%20Advice%20to%20AEC%20on%20sampling%20methodology.pdf
[3] https://www.righttoknow.org.au/request/9125/response/27277/attach/8/22FE%20Senate%20assurance%20detailed%20report%20and%20assurance%20outcomes%20DRAFT%20A1916202.pdf
[4] https://www.righttoknow.org.au/request/9125/response/27277/attach/9/22FE%20Senate%20collation%20of%20Stage%202%20assurance%20exceptions%20DRAFT.pdf

and human data entry. Finally, the digital preferences are electronically counted—this is not part of the audit.

This means that, once ballots are received at a counting centre, two kinds of error are possible: the image might differ from the ballot paper, or the digital preferences might not be a correct interpretation of the image.

1.3. The audit process

The auditors sampled slightly over 10,000 ballots and checked each of them for two kinds of errors:

A Stage 1 error is a discrepancy between the ballot paper and the scanned image;

A Stage 2 error is a wrong interpretation of a ballot paper into digital preferences.[5]

2. Audit outcomes

The audit identified 7 errors in Stage 1 and 33 errors in Stage 2. Both were distributed unevenly across the states: smaller states tended to have fewer errors (even as a percentage), and Victoria had the largest error counts for both types.

In estimating the overall error rate, Axiom's analysis attributed the Stage 1 errors proportionally across all states, rather than directly to the states and territories where they occurred (as was done for Stage 2 errors). There is no reason to do this.[6] Confusingly, this attributes a much smaller share of Stage 1 errors to Victoria than were actually observed, substantially understating the overall Victorian error rate.

Adding Stage 1 and 2 errors in each state and territory where they appeared, gives a more accurate version of Axiom's Table 4, shown here as Table 1.

Stage 2 error rates are approximately as expected. Axiom's audit report included a separate detailed report characterising all 33 observed errors, which shows that overall the errors did not advantage or disadvantage any particularly party.

Stage 1 errors, in contrast, are not expected: the ABS methodology document states that none were detected in 2019 and none are expected in 2022. Unfortunately, there is no report detailing the Stage 1 errors observed in either NSW or Victoria.

2.1. Bias, confidence and misinformation

Unfortunately, "bias" has very different meanings in the scientific literature and more general speech. When the Electoral Commissioner states, "The independent assurance process found there was no suggestion of any political or logical bias in the exceptions," it is important to remember that "political bias" and "mathematical or statistical bias" (our interpretation of what was meant by "logical bias") are different.

[5] Unfortunately the ABS Methodology document uses a slightly different definition of Stage 2 error, but we use Axiom's terminology here. We assume that Axiom counted a Stage 2 error only when there was no Stage 1 error, that is, when the image seemed to be an accurate representation of the ballot paper.

[6] Axiom's report states that this is based on the ABS's recommended methodology, but the ABS methodology document makes no such recommendation. It states that the expected number of exceptions of this kind is zero. (It does, on p.4, explain that a weighted average is necessary for correctly computing an overall *national* error rate, but that is irrelevant to state-by-state results.)

Table 1: **Errors detected by the audit, with Stage 1 and 2 errors added by state & territory.** This is equivalent to Axiom's Table 4, but with Stage 1 errors listed where they occurred rather than attributing them equally everywhere.

State /Ter.	Total ballots	Sample size	Stage 1 Errors	Stage 2 Errors	Total Errors	(%) Errors	Extrapolated Errors
NSW	4,996,110	1,735	2	7	9	0.52%	25,916
VIC	3,960,958	1,400	4	8	12	0.86%	33,951
QLD	3,111,034	1,310	0	2	2	0.15%	4,750
WA	1,571,899	1,285	0	3	3	0.23%	3,670
SA	1,162,472	1,245	0	5	5	0.40%	4,669
TAS	372,973	1,135	0	5	5	0.44%	1,643
ACT	290,308	1,028	0	2	2	0.19%	565
NT	106,907	965	1	1	2	0.21%	222

To illustrate the difference, suppose a political poll sampled only three voters. It is reasonably likely that they would all express the same voting intention.[7] Should that occur, it would not mean that the poll was biased in an intentional way, nor that the voters really were unanimous—small samples often give inaccurate results. This example is not "biased" in the statistical sense, but the preferences in the sample are not an accurate estimate of the preferences among the electorate because the sample is so small. The only way to gain a more accurate estimate is to poll more voters.[8]

It is highly likely that the three unexplained exceptions detected in Victoria all affected the same first preference (candidate or party) because they all occurred in the same packet, and ballots are sorted by first preference before they are packed. This does not mean that all the omissions of ballot papers *throughout the state* affected the same first preference—the sample is too small to give any confidence. The only way to gain a more accurate estimate is to take a larger, genuinely random sample of ballot papers.

The audit was undermined to some extent by the decision not to sample the ballot papers randomly. This seems to be a consequence of poor communication between the ABS, the AEC and Axiom. Although the ABS Methodology document does not explicitly recommend sampling *randomly* chosen packets, it is implicit in the notion of proper statistical sampling—the confidence intervals in their document are meaningless unless the sample is drawn randomly. Unfortunately, Axiom does not seem to have understood the importance of random sampling, instead choosing the first packet from selected ballot paper transport containers (BPTCs). Hence, the confidence bounds in Table 4 of their report are meaningless, as are all their

[7] For example, if the true fractions were 50% Blue, 40% Red and 10% Green, then the probability that three random voters are unanimous would be about 20% ($0.5^3 + 0.4^3 + 0.1^3 \approx 0.2$).

[8] An example of a poll that *is* biased in a statistical sense would be one that samples in a non-representative way, for instance by being more likely to survey people from large cities rather than in regional areas, and then not taking this into account in the analysis. Such biases can be avoided by ensuring that every voter is equally likely to be in the poll.

statements of the form "The AEC can be 95% confident that there are between a and b exceptions per 1000 ballot papers." No such confidence can be inferred from this non-random sampling method. Because the inspected ballot papers were not selected randomly, the measured error rates provide little evidence about the true error rates. See Appendix A.1 for more explanation of this issue. For the purpose of the rest of this submission, we shall assume—optimistically—that the measured error rates are exactly equal to the true error rates, despite the faulty sampling design. Future audits should select ballot papers at random using a well-defined mechanism.

2.2. Could the observed error rate alter the election outcome?

A well-conducted audit would generally confirm the election outcome if it found evidence that the error rate was below the electoral margin. In this section we consider this question for the 2022 Australian Senate outcomes.

In general it is computationally difficult to calculate electoral margins for Senate elections. We have therefore developed heuristics to look for small changes that are *sufficient* to alter the results. The main ideas are described in [TC], which contains links to the source code we used to do these calculations. The heuristics find small changes that are *sufficient* to change the election result. However, it may not be *necessary* to change so many ballots—it is possible that better solutions may exist, and that more sophisticated search algorithms such as those given by [BCST20] may find ways of altering outcomes that do not involve changing as many votes.

Our results are in Table 2. For each state, we list the smallest vote change we found that can change who won at least one seat (usually the last seat allocated through the preference distribution process). In both territories and most states, the smallest outcome-changing modifications our algorithm found involved changing more than 1% of the votes. In WA and SA, a change of a little less than 1% suffices to change the electoral outcome. Victoria is by far the closest: altering just 0.24% of votes is sufficient to change one seat.

We will refer to these values as the "apparent margin" as a reminder that there may be smaller margins that our algorithm did not detect. However, for the purposes of the audit it seems that these numbers are the most useful basis available for assessing whether the measured error rates are acceptably low. If the true error rate is smaller than the apparent margin, this is the best evidence we can expect of an accurate election outcome; if it is larger, this is an indication of a possible problem that deserves further investigation.

Table 3 compares the measured error rates from the audit with the apparent margins, for each state and territory. In all except Vic, the measured error rate is well below the apparent margin. This provides some evidence that the election results are correct; the evidence would be far stronger had the sample been drawn using an appropriate statistical method.

In Victoria, the measured error rate is substantially higher than the margin. Indeed, the Stage 1 error rate alone is comparable to the margin. Because the sample was not selected randomly, and because only a small number of incidents were detected, the indication of the problem may vary greatly from its true extent, in terms of both the party or parties affected and the overall size of the problem. This could go either way: the problem may be either more or less serious than it appears based on these data. It is necessary to examine a larger, properly selected, sample to assess whether the election result is correct.

Table 2: **Vote changes that can change Senate outcomes for 2022 Australian Senate results.** Two kinds of changes are shown. 'Allow 1st-prefs' lets first-preferences be different. In 'No 1st-prefs', first-preferences are assumed to be perfectly recorded. In some cases, this results in a slightly larger number of vote changes, in others it makes no difference, and in some cases it precludes any solution—this is indicated with '×'.

		Allow 1st-prefs		**No 1st-prefs**		
State /Ter.	**Formal votes**	**Change**	**(%)**	**Change**	**(%)**	**Effect**
ACT	285,217	14,137	4.96%	×	×	+Seselja (LP) −Pocock (Ind)
NSW	4,800,722	57,340	1.19%	57,340	1.19%	+McCulloch (ON) −Molan (LP)
NT	103,617	11,412	11.01%	×	×	+Anlezark (Grn) −McCarthy (ALP)
QLD	3,013,868	54,810	1.82%	59,428	1.97%	+Stoker (LP) −Chisolm (ALP) (1st-prefs) −Hanson (ON) (no 1st-prefs)
SA	1,128,524	9,306	0.82%	12,558	1.11%	+Gill (ALP) −Liddle (LP)
TAS	361,048	11,697	3.24%	×	×	+Mav (ON) −Tyrrell (JL)
VIC	3,821,539	9,341	0.24%	9,341	0.24%	+Pickering (ON) −Babet (UAP)
WA	1,526,123	11,745	0.77%	×	×	+Filing (ON) −Payman (ALP)

Table 3: **Comparison between error estimates from the audit and margin estimates from the voting data.** In every state except Vic, the error rate seems well below the number of errors sufficient to change the outcome.

State /Ter.	**Total ballots**	**Error estimate (%)**	**Sufficient changes (%)**
NSW	4,996,110	0.52%	1.19%
VIC	**3,960,958**	**0.86%**	**0.24%**
QLD	3,111,034	0.15%	1.82%
WA	1,571,899	0.23%	0.77%
SA	1,162,472	0.40%	0.82%
TAS	372,973	0.44%	3.24%
ACT	290,308	0.19%	4.96%
NT	106,907	0.21%	11.01%

Had the papers been selected at random using appropriate statistical sampling, the uncertainty in the measured rates as estimates of the true error rates could be assessed rigorously, and the evidence the measured rates provide about whether the true error rates are below the apparent margins could be quantified.

3. Proposals for further investigations

In this section we describe recommended investigations for this year's data, to be performed before the ballot papers are destroyed. The intention of this examination is to understand the causes and implications of the problem in Victoria, with the goal of improving the process for future elections.

Since the unexplained exceptions (Stage 1 errors) consist of discrepancies between the ballot papers and the ballot image database, there is no way to assess the extent of the problem other than by examining more ballot papers.

- Sample more ballot papers (at random, using appropriate statistical sampling) and verify that there are matching images in the database. Carefully record the exceptional cases where no matching image can be found.
- Unlike Axiom's audit, the ballots should be selected *randomly*. It would be reasonable to follow ABS's advice and select *packets* randomly; the confidence interval calculation must take into account how the sampling is performed (See Section A.1).

The other recommendations are quicker checks that can be automated. These may help to give some clues about the nature of the issue, but do not provide a complete explanation on their own.

- Verify the digital signatures on the whole set of ballot images.
- Count the duplicate images. There ought to be none, but our understanding of Axiom's report is that the two omitted ballots in NSW were compensated for by two pairs of duplicate images. These should be easy to find.
- For each duplicate image, retrieve the batch of ballot papers and check for papers with no matching image.
- Record the nature of all duplicated images and omitted ballot papers.
- If possible, try to access the data in the form it was retrieved from the scanner, to understand whether the image set was free of duplicates at that stage.

At the end of this process, an informal investigation should be able to identify the root cause of the problem.

- Look for consistent *political* patterns: are some parties systematically advantaged or disadvantaged by the errors?
- Look for consistent *physical or procedural* patterns: are there some scanners, or some human operators, or some patterns in the timing, that seem to have contributed more errors, particularly Stage 1 errors?

The detailed results of this investigation should be promptly communicated to the public. This builds trust by either showing evidence that the problems are sufficiently small, or showing honesty in acknowledging a serious problem.

4. Technology in voting: why some approaches truly improve enfranchisement and others do not

We have written extensively about voting technology in previous submissions, and will not repeat the details here. The following summary and suggestions are adapted from our recent submissions to an inquiry by the Parliament of NSW.

4.1. Internet voting

Apart from the obvious privacy implications, the key concept in electronic elections is *verifiability*, the opportunity to check whether an announced election outcome is accurate. Plain paper voting in a polling place can be verified by observers and scrutineers; computerised voting is much harder to observe, because watching the screen gives scrutineers no real evidence of what the computer is doing with the votes. Another key concept is *evidence-based elections*, the principle that it is not enough for elections to find the true winners: they must also provide convincing evidence that they did [SW12, AS20]. Anyone can claim that their system is secure and protects voter privacy, but elections should *demonstrate* that the result accurately reflects the choice of the people, to the satisfaction of scrutineers, disappointed candidates and their supporters, and members of the public.

Australia's largest Internet voting project, the NSW iVote system, has suffered from numerous serious security and privacy problems [HT15, CEET17, HLPT20]. iVote's security problems are not unusual among Internet voting systems. Independent studies of similar systems have found similar vulnerabilities [WWIH12, SFD+14]. A recent Russian e-voting system was shown to be using cryptographic primitives that could be compromised in just 20 minutes.[9]

In 2021 during Local Government Elections, iVote suffered substantial downtime that disenfranchised thousands of voters and led the NSW Supreme Court to annul the election results in three local councils. Based on our analysis of NSWEC data, 33 other council outcomes may also have been affected.[10]

The voting rights of marginalised populations of voters are not satisfied by offering them a system that does not adequately protect their privacy, is not reliable, and does not give them evidence that their vote is accurately recorded and counted.

We have not conducted a detailed analysis of the AEC's phone voting solution, because no detailed technical information is publicly available. However, the same questions about privacy, reliability and verifiability should be asked.

4.2. More transparent and trustworthy alternatives

This section briefly outlines alternative uses of technology in elections that are much more consistent with a secret ballot and verifiable election outcomes. In all cases, the details matter, and we are not necessarily advocating any of these without further careful examination of the pros and cons and publication of the particular implementations contemplated. Our point is simply that there are numerous other ways of meeting the needs of voters, without sacrificing election integrity. Some of these approaches might work well, if they were designed with transparency, verifiable outcomes, and ballot secrecy as primary goals.

[9] https://members.loria.fr/PGaudry/moscow/

[10] https://github.com/AndrewConway/ConcreteSTV/blob/main/reports/NSWLGE2021Report.pdf

Electronic voting in a polling place with a voter-verifiable paper record. Voters with disabilities could use a computer in a polling place as an aid to fill in their paper ballot. The computer could then print out a human-readable paper record of the vote. Although not all voters with disabilities are able to check this printout directly, many could, and the overall opportunity to verify the result would be much better than nothing. These printouts could then be included in the normal scrutineered counting process. In the USA these are called *Electronic Ballot Marking Devices.*

(Note that there is significant controversy in the United States over the use of these systems for general voters, which we are not advocating here—we are considering an alternative to human assistance for voters who would otherwise need to dictate their vote to a person.)

Delivering candidate information electronically; returning a paper vote by mail. Another alternative for those who miss the postal voting deadline for reasons beyond their control (such as a COVID-19 diagnosis) could be the opportunity to print out a ballot at home and return it by mail. This is clearly a last resort, but compares favourably to Internet voting because at least voters could check whether their official ballot accurately reflects their preferences. In the USA, this is called "remote accessible vote-by-mail."

Secure drop boxes to reduce dependence on the post. Many US jurisdictions provide special secure drop-boxes for ballot return, so that voters do not need to rely on the postal service. In general, "smart" drop boxes that include video surveillance and log each ballot as it is deposited are preferable; boxes should be emptied at least daily; and appropriate chain-of-custody protocols should be followed in emptying the drop boxes and transporting voted ballots.

No-excuse early voting. This would allow people to vote early without a specific reason, hence reducing the pressure on polling day and increasing each person's opportunity to vote while well.

5. Conclusion

This year's audit of the Senate scanning and counting process represents a substantial improvement in Australian electoral processes. While there were serious deficiencies in how the sample was drawn, the measured error rates suggest that the results are correct in most states and both territories. The identification of a possibly serious problem in one state (Victoria) should also be considered a success. Of course, like any new process, there are imperfections in the first run which need improvement for future years. In this submission we have made specific suggestions for investigating this year's problem in Victoria, and general recommendations for improving future audits, including the use of appropriate statistical sampling.

Any other use of technology in electoral processes should be subject to the same principles: privacy of the votes, combined with an opportunity to verify the accuracy of the process. We have explained why Internet voting does not currently meet this requirement, and suggested some other approaches for improving access without undermining integrity.

References

[AS20] A.W. Appel and P.B. Stark. Evidence-based elections: Create a meaningful paper trail, then audit. *Georgetown Law Technology Review*, 4.2:523–541, 2020. https://georgetownlawtechreview.org/wp-content/uploads/2020/07/4.2-p523-541-Appel-Stark.pdf.

[Ass22] Axiom Associates. External statistical assurance of the senate ballot papers: Detailed report and statement of outcomes (draft), July 2022. https://www.righttoknow.org.au/request/9125/response/27277/attach/8/22FE%20Senate%20assurance%20detailed%20report%20and%20assurance%20outcomes%20DRAFT%20A1916202.pdf.

[BCST20] Michelle Blom, Andrew Conway, Peter J Stuckey, and Vanessa J Teague. Did that lost ballot box cost me a seat? Computing manipulations of STV elections. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13235–13240, 2020. https://web.archive.org/web/202011031729 32id_/https://aaai.org/ojs/index.php/AAAI/article/download/7029/6883.

[BSS+22] Michelle Blom, Philip B Stark, Peter J Stuckey, Vanessa Teague, and Damjan Vukcevic. Assessing the accuracy of the Australian Senate count: Key steps for a rigorous and transparent audit. *arXiv preprint arXiv:2205.14634*, 2022. https://arxiv.org/abs/2205.14634.

[CEET17] Chris Culnane, Mark Eldridge, Aleksander Essex, and Vanessa Teague. Trust implications of DDoS protection in online elections. In *International Joint Conference on Electronic Voting*, pages 127–145. Springer, 2017.

[Com22] Australian Electoral Commission. Methodology for the ballot paper sampling process for Senate election 2022, May 2022. https://www.aec.gov.au/About_AEC/cea-notices/files/2022/s273AC-senate-assurance-methodology-fe2022.pdf.

[HLPT20] Thomas Haines, Sarah Jamie Lewis, Olivier Pereira, and Vanessa Teague. How not to prove your election outcome. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 644–660. IEEE, 2020.

[HT15] J Alex Halderman and Vanessa Teague. The New South Wales iVote system: Security failures and verification flaws in a live online election. In *International conference on e-voting and identity*, pages 35–53. Springer, 2015.

[oS22] Australian Bureau of Statistics. ABS advice to AEC on sampling methodology, 2022. https://www.righttoknow.org.au/request/8988/response/26588/attach/8/Relevant%20Document%20ABS%20Advice%20to%20AEC%20on%20sampling%20methodology.pdf.

[Rog22] Tom Rogers. Senate ballot paper sampling outcomes statement, 7 July 2022. https://www.aec.gov.au/Voting/counting/files/section-237AC(7)-statement.pdf.

[SFD+14] Drew Springall, Travis Finkenauer, Zakir Durumeric, Jason Kitcat, Harri Hursti, Margaret MacAlpine, and J Alex Halderman. Security analysis of the Estonian internet voting system. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 703–715. ACM, 2014.

[SW12] Philip B. Stark and David A. Wagner. Evidence-based elections. *IEEE Security and Privacy*, 10:33–41, 2012.

[TC] Vanessa Teague and Andrew Conway. iVote issues: Assessment of potential impacts on the 2021 NSW local government elections. https://github.com/AndrewConway/ConcreteSTV/blob/main/reports/NSWLGE2021Report.pdf.

[WWIH12] Scott Wolchok, Eric Wustrow, Dawn Isabel, and J Alex Halderman. Attacking the Washington, DC internet voting system. In *International Conference on Financial Cryptography and Data Security*, pages 114–128. Springer, 2012.

A. Improving the methodology for future audits

We have already mentioned the importance of *random* sampling for future audits. This section explains its importance and lists some other improvements for future audits.

A.1. Sampling and confidence intervals

The legislation requires a "statistically significant sample." This is not a clearly defined term. That said, we cannot think of any reasonable interpretation that could be met by a sampling method that can never sample anything but the first packet in each ballot paper transport container (BPTC).

The confidence intervals presented in Axiom's report were calculated using an online calculator on the ABS website.[11] However, that calculator is only valid for simple random samples (a warning to that effect is written on the website). The sample that Axiom drew is *not* a simple random sample of ballots. It was a *systematic cluster sample* that drew the first packet from each ballot paper transport container (BPTC).

The easiest way to understand why this is a problem is by analogy with a financial audit: suppose an auditing firm tried to estimate the extent of expenses fraud, but was widely known for checking receipts only for expenses incurred on Mondays. They might detect a very low apparent rate of cheating, but no confidence about the overall rate can be inferred: people who know how the audit will be conducted will simply tell the truth about their Monday expenses, then cheat between Tuesday and Friday. Even accidental omissions might not be accurately represented by this kind of sampling: if people are more likely to accidentally forget to keep receipts because they are tired or drunk on Fridays, then checking only Monday's receipts will not detect this. Hence a sample of only the first day of the week might result in a very inaccurate estimate, for both accidental and deliberate failures.

[11] https://www.abs.gov.au/websitedbs/d3310114.nsf/home/sample+size+calculator

The same argument applies to Axiom's choice to sample only the first packet of any BPTC. Moreover, even if packets were selected at random, the fact that the sample drew entire *clusters* of papers (packets rather than individual ballot papers) would need to be taken into account in calculating uncertainties and confidence intervals. Sampling clusters of papers selected at random provides much weaker statistical evidence than sampling the same number of papers selected individually at random. Axiom's calculations were inappropriate and misleading for both reasons.

A.2. Summary and other issues

This section summarises process improvements the AEC should implement in future audits. Items 4 and 5 below refer to a detail of Axiom's audit: they counted an error only if a team of AEC officials unanimously considered it to be so.

1. Draw the random sample in a verifiable way.
2. Calculate confidence intervals using a method consistent with how the sample was drawn (e.g., take into account whether individual papers or packets of papers are sampled).
3. Notify the public honestly and promptly if a problem appears.
4. Record non-unanimous detected mistakes even if they are not counted in the final error rate.
5. Ensure separation between the AEC officials who determine the final error judgement and those who could have made the original determination on the ballot.

The AEC's process sketch arguably did not meet the legislative requirement to publish their audit methodology in advance. It has been very difficult for us, a team of experts in the field, to understand how the audit was conducted. Most of the necessary information was contained in documents that we received only after waiting on Freedom of Information requests after the election.

Finally, and perhaps most concerningly, Axiom seems to have simply not understood the primary purpose of the audit. Paragraphs 1.10 and 1.11 of their report give the impression that they think that detection of added or deleted ballot records is not their responsibility:

> The approach to assurance testing was not designed to detect whether there were electronic data records held by the AEC that were not associated with a physical ballot paper or whether there were multiple electronic data records associated with a single physical ballot paper.

However, this is *exactly* what the audit was supposed to check for: discrepancies between the ballot papers and the corresponding images and digitized preferences.