Rehmann

March 20, 2023

State of Michigan Bureau of Elections
and Board of State Canvassers

Re: Recommendations for Random Sampling when Canvassing Signatures

We have been engaged by the Bureau of Elections to provide consulting services regarding random sampling when canvassing signatures. In connection with that engagement, we have conducted interviews with various Bureau staff, reviewed the State's current approach to random sampling, and have modeled the effects of various proposed revisions to the approach in order to make recommendations to the Board of State Canvassers for systemic improvements. Our goal was first to determine whether the State's current methodology remained theoretically sound, and if so, to identify ways to modify it without fundamentally altering it. The first priority is maintaining the objectivity and reliability of the canvassing process. A secondary priority is improving its efficiency. This report contains the results of our procedures, and our recommendations for the Board.

Overview and Basis for Current Methodology

Since the 1980's, the Bureau's procedure for canvassing petitions on behalf of the Board involves random sampling, based upon a Department of State Report titled "Random Sample Signature Canvassing in Michigan," drafted in 1979 (and later reaffirmed in 1990) with assistance from the Department of Statistics and Probability at Michigan State University. The Board of State Canvassers adopted this method of random sampling in the early 1980's. As a result of the Board's adoption of the methodology used in the report and the mathematical formulas presented within it, a software tool was developed in the early 1980's that is still in use today.

The Bureau uses random sampling when canvassing statewide candidates and ballot initiatives because the volume of signatures submitted makes it impractical to validate 100% of the population. Despite advances in technology, the comparison of individual signatures to the official voter registration logs still requires a subjective evaluation by a trained eye. Human signatures change over time for a variety of reasons, and the Bureau has guidelines on what constitutes acceptable deviations from the official signature on file. At present, there is not a cost-effective technological solution to replace the manual review that currently takes place. Accordingly, some degree of sampling is needed, as it would be too costly and time consuming to review every signature by the statutory deadlines for completing canvassing procedures.

The 1979 report lays out a sound theoretical model for statistical sampling. While somewhat technical in its content, the methods and approaches described in that report remain consistent with modern sampling guidance found in publications of the American Institute of Certified Public Accountants (AICPA) and employed in the audits of governmental financial statements and federal grant programs. Essentially, that report proposes selecting sample sizes that provide at least a 90% chance that the decisions reached to certify or deny a petition will be correct.

State of Michigan Bureau of Elections
  and Board of State Canvassers
March 20, 2023
Page 2 of 6

Rather than relying on complex statistical formulae, we set out to validate the State's current approach by designing a series of simulations. Using the framework of Microsoft Excel and its integrated Visual Basic for Applications programming language, we built multiple spreadsheets to model the results of drawing samples of various sizes and comparing the results of those samples to the known actual contents of the sample data used. This allowed us to demonstrate near-real world results through thousands of repetitions.

In the balance of this report, we will describe each aspect of the Bureau's current sampling methodology, and provide our recommendations for modifying or retaining those elements.

Sampling Risk, Confidence Level, and Margin of Error

Any time a sample is selected, there is a risk that the sample will not be representative of the population from which it was drawn. Auditors refer to this as *sampling risk*. There are several ways to reduce sampling risk, but it can never be fully eliminated. In general, larger samples are more reliable than smaller ones.

When deciding how large of a sample to draw, it is important to have a goal in mind. Since sampling risk cannot be reduced to zero, some degree of risk must be accepted. The Board's current guidance calls for a 90% "accuracy requirement". In modern statistical literature, this is often referred to as the *confidence level*. That is, sample sizes are designed to limit the chances that relying on a sample will result in the wrong decision (certify vs. deny) only 10% of the time. Today, auditors would refer to a 90% confidence level as a "high degree" of confidence.

Once a sample size has been selected, and signatures verified, the results are then projected to the entire population. For example, in a petition with 500,000 signatures, a sample of 1,000 signature might be drawn. Based on the verification procedures performed, the Bureau might determine that 800 of those signatures (80%) are valid and 200 are invalid. Based on that, one could project that since 80% of the sample was valid that 400,000 signatures (80% of the population) would likewise be valid. However, because of sampling risk, there will always be some *margin of error*. That is, the actual number of valid signatures may range between 388,000 and 412,000. That range of potential results (±12,000 signatures) can be expressed as a percentage (e.g., ±2.4%) and used as another way of articulating the goal in selecting an appropriate sample size.

The 1979 report set a theoretical goal of 90% confidence with a ±1.0% margin of error, which it expressed as the "(.01, 10%) rule". As the 1979 report points out, depending on number of signatures submitted in excess of the required number, following this rule could result in sample sizes ranging from 333 to 4109. However, in practice, the Board-approved procedure is to initially select a sample of 500 signatures, which has a nominal margin of error of approximately ±4.3%. Depending on the results of that initial sample, the Bureau has then been selecting an additional 500-2,000 signatures. Selecting a pre-determined number of signatures to sample results in a more predictable time investment, and typically still provides an acceptable confidence level and margin of error.

State of Michigan Bureau of Elections
  and Board of State Canvassers
March 20, 2023
Page 3 of 6

Recommendation: the Board's current goal of a 90% confidence level remains appropriate. However, the acceptable margin of error should be set to ±2.5%. This is because the sample sizes needed to achieve a ±1.0% margin of error vastly exceed what is being used today, and would be impractical to achieve. Essentially, this change would be aligning the stated goal with something closer to current practice. Of course, when the number of submitted signatures exceeds the required number of signatures by a large margin, much higher overall confidence levels can be achieved.

Organizing and Counting Petition Sheets

Signatures are collected on petition sheets containing up to 10-12 signatures per sheet. The Bureau's current practice for counting the number of signatures submitted is as follows:

1. Sort the petition sheets into batches based on the number of signatures each contains.
2. Count the number of sheets in each batch.
3. Multiply the number of sheets in each batch by the number of signatures on each sheet in that batch.
4. Sum the total of all the batches.

Once complete, current practice is then to randomize (i.e., shuffle) the sheets before stamping each with a unique serial number. This shuffling of the sheets was intended to both better randomize the samples drawn and to decrease the likelihood of "misses"—sample selections made by the computer that do not exist (such as a selection of line 8 from a petition sheet with only 4 signatures).

Recommendations: With minor modifications to the sampling process, some significant efficiencies can be achieved here. First, there is no statistical advantage to drawing a truly random selection from a population of petition sheets in a sequential or randomized order. Second, the software could be so designed as to avoid making selections that do not exist. By specifying the number of petition sheets containing each number of signatures, the software could better select only signatures that exist, while still maintaining complete randomness.

Finally, with the move toward scanning and digitizing petition sheets, the need to hand count sheets is eliminated. The page count of the final scanned PDF can serve that purpose with greater speed and accuracy than hand-counting the petition sheets.

Identifying Samples by Signature Number

Current practice is for the computer software to select a specific line number on a specific sheet. If that line number is blank (or crossed out), then the selection is considered to be a "miss", and must be re-drawn. Due to the way the sheets are randomized, there is a significant likelihood of misses.

Recommendation: As described in the previous section, we do not believe there is any practical value in shuffling the sheets. If they remain in order of the number of signatures they contain, the software can select signature numbers that are less than or equal to the number of signatures on the sheet. However, the current nomenclature of selecting a signature on *line* 7 would need to be replaced with selecting the 7th *signature*, even if that appears on a different line number. For example, a

State of Michigan Bureau of Elections
  and Board of State Canvassers
March 20, 2023
Page 4 of 6

petition sheet with 10 signature lines may only have 9 signatures filled in, with line 7 having been skipped over for whatever reason. By referring to signatures by their ordinal position, rather than their specific line number, this potential issue can be resolved. This should result in substantially fewer misses, which in turn leads to a faster sample selection process.

Face Review

Current practice is for Bureau staff to perform an initial "face review" on 100% of the petition sheets before sampling begins. The purpose of this review is to weed out any wholly invalid sheets before turning any attention to the specific signatures on that sheet. As might be expected, this is an extremely time intensive process. Based on anecdotal results from several petitions processed in 2022, it appears that approximately 3% of the petition sheets in any particular petition drive may fail face review. However, it requires looking at the entire population of petition sheets (often numbering over 100,000 sheets) to make this determination.

Once wholly invalid sheets have been removed from the population through the face review process, the required percentage of valid signatures increases, accordingly. However, the statistical likelihood of a drawn sample being representative of the population remains unchanged.

To demonstrate this fact, we built a simulation model that drew samples of various sizes (ranging from 500 to 2500) and compared the results with and without a face review being conducted. Our modeling showed no appreciable difference in the results. That is, the confidence level and margin of error were identical in each case, and the differences in results varied by statistically insignificant amounts (sometimes higher and sometimes lower, but typically varying by less than 0.1%).

Recommendation: because the face review process is labor-intensive, and does not improve the reliability of the sampling results, we recommend it be discontinued as part of the sampling process.

One-Stage vs. Two-Stage Sampling

The 1979 report outlined two sampling plans, and suggested that while both offered advantages and disadvantages, that either was acceptable to use in practice. Plan A (one-stage sampling) resulted in larger sample sizes, but required only a single draw. Plan B (two-stage sampling) allowed for a smaller initial draw, with the *potential* need for a second (and possibly much larger) draw depending on the results of the initial sample.

In practice, the Bureau has mostly been using Plan B, initially drawing a sample of approximately 500 (for practical reasons, this was often somewhat larger to allow for potential misses in the sample selection), and then determining whether the results warranted additional sampling. The reason a second sample might be needed is that a sample of only 500 signatures has a fairly large margin of error at a 90% confidence level. According to the existing software used, this margin of error is approximately ±4.3%, though in our modeling, we found it to be closer to ±3.2%. Regardless, that is a larger margin of error than the Board would likely be willing to accept in most cases. It would be acceptable, however, when a sample reveals that the projected number of valid signatures differs from the requirement by (for example) more than 5.0%. In that case, the margin of error is

State of Michigan Bureau of Elections
  and Board of State Canvassers
March 20, 2023
Page 5 of 6

not at all problematic. However, if the projected number of valid signatures is short of the requirement by only 3.0%, this margin of error would make the results of the initial sample inconclusive.

Recommendation: Given the significant administrative time involved in selecting and drawing a sample, there are practical benefits to making a single larger sample immediately from the beginning. While drawing a larger sample will certainly require more time than a smaller one, a one-stage approach represents a more predictable time commitment, and the "extra" time required by the larger initial sample sizes will be offset by the other time savings achieved through eliminating the face review and not shuffling the petition sheets after counting.

Sample Sizes

As mentioned previously, the 1979 report theoretically called for sample sizes ranging from 333 to 4109, while past practice resulted in samples ranging from 500 to 2500.

Recommendation: Understanding that some degree of risk will remain, regardless of the sample size selected, and considering the previous recommendation of a 90% confidence level and a margin of error of ±2.5%, we recommend a one-stage sample size of 1,000 for all state-wide ballot initiatives, and 750 for all state-wide candidate petitions (if sampling is used for state-wide candidate petitions). In our modeling, these sample sizes consistently yielded results of ±2.3%, which exceeds the stated target.

While larger sample sizes could be used to reduce that margin of error further, the incremental value diminishes rapidly, while the time cost increases significantly. In our judgment, these recommended sample sizes strike an appropriate balance between cost and benefit.

Software Recommendations

Some of the recommendations contained in this report could be implemented immediately, with no change in software. Others (such as making selections that consider the number of signatures on each petition sheet to reduce misses) would require an upgrade to the system.

Practically speaking, an "upgrade" is likely infeasible, as the Bureau's current software program was developed in the early 1980s using a platform that is no longer supported by modern PCs. Accordingly, a new system will need to be developed.

As part of our proposal to the Bureau of Elections, Rehmann outlined one possible way to accomplish this using Visual Basic for Applications (VBA), which is included in Microsoft Excel. This would provide a simple and transparent programming platform to make random selections in the future. Of course, other methods could also be employed by a different vendor, should that be desired.

State of Michigan Bureau of Elections
  and Board of State Canvassers
March 20, 2023
Page 6 of 6

Summary / Conclusion

The current methodology used by the Bureau of Elections to draw random samples remains fundamentally sound. However, opportunities exist to improve its efficiency without sacrificing accuracy.

We appreciate this opportunity to be of service to the Bureau of Elections and Board of State Canvassers. If you have any questions, please do not hesitate to contact us.

Very truly yours,

Rehmann Robson LLC