

THE
CARTER CENTER



**Report on an Analysis of the Representativeness of the Second Audit Sample, and
the Correlation between Petition Signers and the Yes Vote in the Aug. 15, 2004
Presidential Recall Referendum in Venezuela**

This study was conducted by The Carter Center and confirmed by the OAS in response to a written request from Sumate presented to The Carter Center Sept. 7, 2004. Sumate asked that The Carter Center evaluate a study performed by Professors Ricardo Hausmann and Roberto Rigobon.

The Hausmann/Rigobon study states the second audit conducted Aug. 18-20 and observed by The Carter Center and the OAS was based on a sample that was not random and representative of the universe of all voting centers using voting machines in the Aug. 15, 2004, recall referendum.¹ The study further indicates that the correlation coefficient (elasticity) for the correlation between the signers and the YES votes for the sample was 10 percent higher than that for the universe. The Hausmann/Rigobon study came to these conclusions through an analysis of the exit poll data, petition signers data, and electoral results data provided by Sumate.

1 Objectives of the Carter Center Study

1. Determine the correlation between the number of signers of the presidential recall petition and the electoral results of the Aug. 15 recall referendum.
2. Compare the characteristics of the universe of voting machine results with those of the sample for the 2nd audit performed Aug. 18.
3. Determine the universe from which the sample generation program used Aug. 18 was drawn.

The scope of this study is limited to the voting centers that used voting machines during the Aug. 15, 2004, recall referendum, to respond to concerns that the electronic voting machines results were manipulated.

¹ Conclusions section, page 25: "Nuestro análisis indica que la muestra seleccionada para realizar la auditoría del 18 de agosto de 2004 no es aleatoria y representativa del centro del conjunto de centros de votación. En dicha muestra, la elasticidad de las firmas frente a los votos es 10 por ciento más alta ..."

2 Data Sources

The data used to perform this study was officially received from the CNE. The following data was used:

1. The Voters List (REP) used for the Aug. 15 recall referendum officially received from the CNE July 30, 2004.
2. The “*cuadernos de reparo*” database containing the valid signatures and the “repairable” signatures used during the *Reparos* process and officially received from the CNE.
3. The rejected signature list (signatures that could not be repaired during *Reparos*) obtained from the CNE.
4. The electoral results file of voting machines for the Aug. 15, 2004, recall referendum, per machine, officially received from the CNE Aug. 18, 2004.
5. The sample generation program for the Aug. 18 audit including the source code, the executable file, the input file with the universe and the generated sample.

3 Methodology

The four data sources were loaded into different tables² on an IBM DB2 database to facilitate processing. The following calculations were performed:

1. The number of voters per voting center was calculated from the REP, excluding foreigners³.
2. A single table of signers in the database was loaded from the *cuadernos de reparos* file and the rejected signatures file, eliminating duplicate ID card numbers.
3. The number of signers per voting center was calculated from the table of all signers by matching the ID card number in the signer’s table with the ID card number in the REP and aggregating the signers into voting centers identified by voting center on the REP table.
4. The YES and NO votes per voting center were calculated by adding the electoral results from each voting machine in that center.
5. A final results table was produced with the following columns for each voting center:
 - a. State
 - b. Municipality
 - c. Parish
 - d. Voting Center Number
 - e. Total registered voters in the voting center
 - f. Total signers registered in the voting center
 - g. Total YES votes
 - h. Total NO Votes

² A table in a relational database is a storage entity where all records have the same columns. A database can have multiple tables and allows operations between tables.

³ Foreigners are not allowed to vote for president in Venezuela, consequently they cannot recall him either.

The final results table was exported into an Excel file. The correlation coefficients were calculated using SPSS version 12.

Additionally an Excel worksheet with only the voting centers that had a mesa (voting station) audited in the Aug. 18 random sample used to perform the second audit was generated to evaluate the representativeness of the sample in the universe.

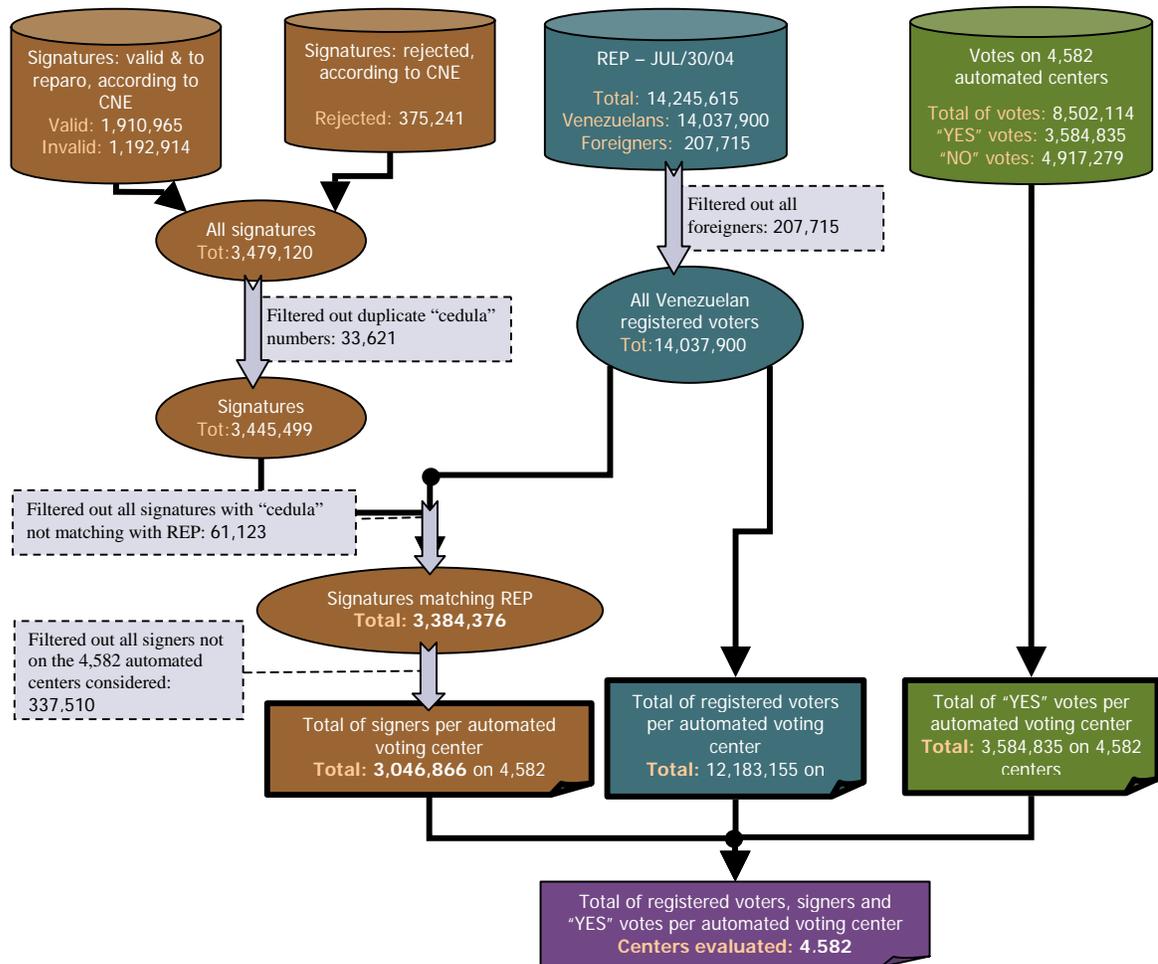


Diagram 1: Data processing flow

We calculated the distribution of differences between signer turnout and YES votes, as well as the correlation between signer turnout and YES votes, both for the universe of all voting centers and the sample drawn Aug. 18.

The total amount of signers in the presidential recall, excluding duplicates by ID card, foreigners, and those not in the voter's list, was considered for this analysis. This study is based on the assumption that all signatures were made in good faith by real voters; for this reason rejected signatures because of *acta* problems, similar handwriting, fingerprint, and other problems were included in the universe of signatures.

The sample generation source and executable program were analyzed as were the input files with the universe and the output generated sample files.

The sample generation program was run 1,020 times with different seeds using an automated testing program. The 1,020 generated samples were loaded into a database. The database was used to determine the number of times a *mesa* (voting station) was included in a sample.

4 Findings

4.1 Signer vs. YES Vote Correlation in Automated Voting Machine Universe

A very high correlation between the number of signers and the number of YES votes per center in the universe of automated voting machines has been found--a correlation coefficient of 0.988⁴. This means that in voting centers where a high signer turnout was obtained, a high YES vote also was obtained. As noted below, there were more YES votes Aug. 15 than signers of the original petition. This YES vote total is the net result of original signers, additional voters who chose YES, and signers that abstained or voted NO.

In this analysis the NO vote turnout is ignored since if a voter did not sign the recall petition, there is no way to differentiate between signer abstention and the will of the signer to oppose the recall.

In 88.9 percent of the voting centers that used voting machines there were more YES votes than signers. This result is expected because there were 537,969 more YES votes than signers: 3,046,866 signatures were collected from citizens assigned to an automated voting center and 3,584,835 YES votes were cast in those centers.

In only 11.1 percent of the voting centers were there less YES votes cast than signers, representing 29,866 votes. This figure can be interpreted as the minimum number of voters that signed the recall petition and voted NO or abstained from voting.

Chart 1 clearly shows the frequency of negative differences, where the signers are more than the YES votes, is very low compared to the positive frequencies. The chart also shows in most voting centers there were at least 100 more YES votes than signers assigned to the voting center. Positive differences indicate more YES votes than signers.

⁴ The correlation coefficient is a number between -1 and 1 that measures the relationship between two data sets that are scaled to be independent of the unit of measurement. If the coefficient is near to 1 then there is a high correlation between the two variables, if it is near 0 there is a low correlation and if it is near -1 there is an inverse correlation (as one variable increases the other decreases).

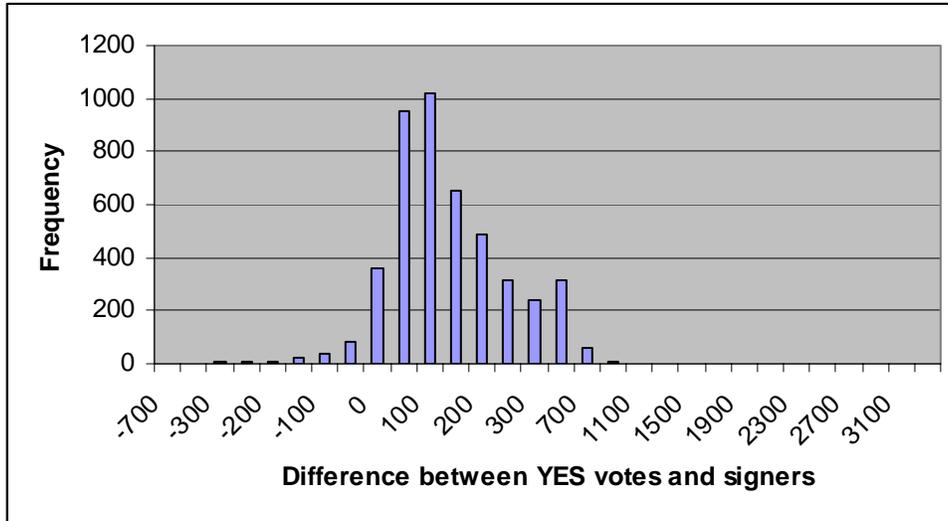


Chart 1: Difference between YES vote and signers histogram

4.2 Signer vs. YES Vote Correlation in Sample

A similarly very high correlation between the signer and the YES votes in the Aug. 18 audit sample has been found, with a correlation coefficient of 0.989.

4.3 Sample vs. Universe Comparison

The analyzed sample was drawn Aug. 18 by the CNE with the purpose to test the accuracy of the voting machines. The paper receipts (*comprobantes*) recording the votes in each of the machines of the voting tables selected in the sample were recounted by CNE auditors in the presence of Carter Center, OAS, and other international observers. The audit concluded the machines correctly tallied and transmitted the votes cast; the findings of this audit are documented in the 2nd audit report posted on the Web sites of The Carter Center and the OAS.

We note the sample also correctly represents the electoral result of all automated voting machines:

	YES Votes	% YES	NO Votes	% NO
Universe	3,584,835	42.2%	4,917,279	57.8%
Sample	145,785	41.6%	204,640	58.4%

Furthermore, the correlation between the signers and the YES votes is almost identical in the universe and in the sample. The difference between the correlations is less than 1 percent:

	Correlation Coefficient
--	-------------------------

Universe	0.988
Sample	0.989

The distribution of the difference between the YES votes and the signers per voting center, presented in Chart 2 below, also shows a very similar behavior:

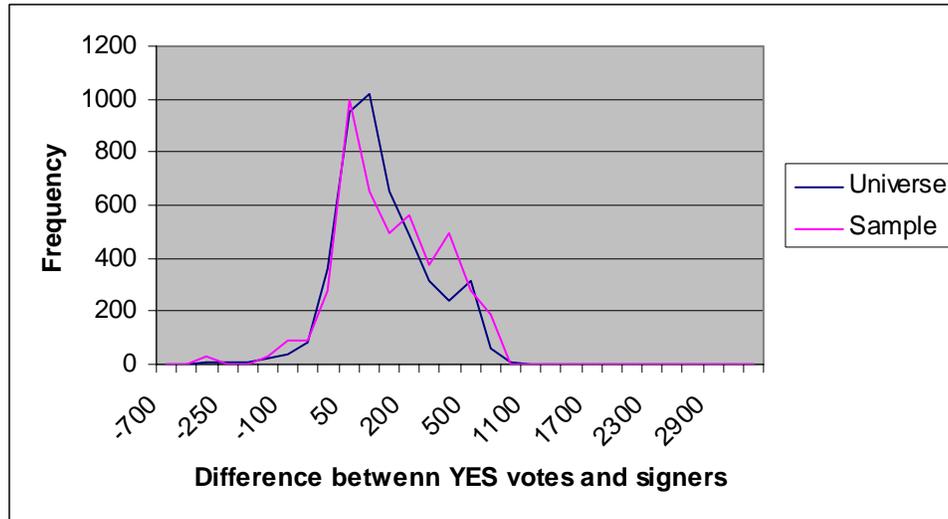


Chart 2: Distribution of difference between YES vote and signers per center

The occurrence of the differences for the sample has been projected to the universe to present the sample and the universe lines on the same scale.

4.4 Analysis of Sample Drawing Program

The CNE requested a group of university professors to develop a sample generation program for the 2nd audit. The program is written in Pascal for the Delphi environment. The program receives a 1 to 8 digit seed. The CNE delivered to the international observers the source code, the executable code, the input file, and the sample. Carter Center experts analyzed the program and concluded:

1. The program generates exactly the same sample given the same seed.
2. The program generates a different sample given a different seed.
3. The program generates a sample of voting stations (*mesas*) based on the universe of *mesas* that have voting machines.
4. The source code delivered produces the executable file delivered.
5. The input file used to generate the sample is missing only six of 8,147 voting stations (*mesas*). The input file has one missing voting center.
6. The program, when run enough times, includes each *mesa* (voting station) in the sample, and the number of times a given *mesa* is included in a sample is evenly distributed, indicating the sample generation program is random.

The sample generation program was run 1,020 times. With no exception all of the 8,141 *mesas* appeared at least 14 times in a sample. Not a single *mesa* was excluded from the sample in the test run.

Average appearance of a mesa in a sample	25.05
Standard deviation	5.52
Minimum appearance	14
Maximum appearance	40

5 Conclusions

The sample drawing program used Aug. 18 to generate the 2nd audit sample generated a random sample from the universe of all mesas (voting stations) with automated voting machines. The sample was not drawn from a group of pre-selected mesas. This sample accurately represents different properties of the universe, including the accuracy of the machines, the total YES and NO votes and the correlation between the YES votes and signer turnout.

There is a high correlation between the number of YES votes per voting center and the number of signers of the presidential recall request per voting center; the places where more signatures were collected also are the places where more YES votes were cast. There is no anomaly in the characteristics of the YES votes when compared to the presumed intention of the signers to recall the president.

The second audit showed a high accuracy of the voting machines with discrepancies of less than 0.1 percent. The sample was analyzed, and it does not have different properties than the universe. The sample generation program was analyzed as part of the 2nd audit process and again in this study. Both studies showed that the sample does not operate on a subset of the universe, thus hiding or masquerading some of the properties of the universe. Consequently the results of the 2nd audit accurately confirm the electoral results of Aug. 15.