Statistical Methods for Detecting Anomalous Voting Patterns: A Case Study

Naval Research Laboratory

Washington, DC 20375-5320

NRL/MR/5500--11-9358

Statistical Methods for Detecting Anomalous Voting Patterns: A Case Study

Roger Hillson

Adversarial Modeling and Exploitation (AMX) Office
Information Technology Division

September 23, 2011

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*
23-09-2011

2. REPORT TYPE
Memorandum Report

3. DATES COVERED *(From - To)*
February 2010 – August 2010

4. TITLE AND SUBTITLE
Statistical Methods for Detecting Anomalous Voting Patterns: A Case Study

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Roger Hillson

5d. PROJECT NUMBER
55-8844-31

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375

8. PERFORMING ORGANIZATION REPORT NUMBER
NRL/MR/5500--11-9358

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR / MONITOR'S ACRONYM(S)

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Both observational and laboratory studies suggest that human beings find it difficult to fabricate truly random numbers. Any sufficiently large set of nominally random number sequences, irrespective of its source, may be analyzed to determine if the criterion of randomness is satisfied. Statistical methods have, therefore, been applied forensically to detect anomalies in accounting data, scientific data sets, and voting data. As a case study, we apply methods developed by Beber and Scacco to analyze polling station counts in Helmand province for the four leading candidates in the 2009 Afghanistan presidential election. In this election, there were allegations of extensive electoral fraud, and nineteen percent of all votes cast were ultimately discarded. The pattern of statistical anomalies present in the raw polling station data is consistent with the findings of the post-election audit conducted by the Afghanistan Electoral Complaints Commission. The strengths and weaknesses of statistical methods for anomaly detection are discussed, as are possible motives for perpetuating voter fraud in elections that pose little uncertainty in projected outcomes.

15. SUBJECT TERMS
Random numbers, Statistical methods, Anomaly detection, Voting patterns, Electoral fraud

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	Unclassified Unlimited	25	Roger Hillson **19b. TELEPHONE NUMBER** *(include area code)* (202) 404-7332

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.

Table of contents

List of Tables

List of Figures

1) Introduction

On 20 August 2009, the presidential election was held in Afghanistan. The outcome was marred by extensive allegations of electoral fraud. These allegations were divided into 6 categories: (1) polling irregularities, (2) counting irregularities, (3) tally center results, (4) access [restriction or denial], (5) missing materials, and (6) undue influence [6, p 32]. Following completion of the auditing process, the Afghanistan Electoral Complaints Commission (ECC) concluded that "at least 1,391 polling stations had entirely fraudulent voting, and at least 452 polling centers and 23 entire districts had fraud in at least half of their polling stations [6, p 9; 11, 12, 25]." Nineteen percent of all votes cast were ultimately discarded from the final certified presidential results. Following these adjudications, no presidential candidate received more than 50% of the total vote. Under Afghan law, the winning candidate must receive a simple majority of the votes. Thus, a run-off was mandated between the top two presidential candidates - Hamid Karzai (49.67%) and Abdullah Abdullah (30.59%). Mr. Abdullah elected to withdraw prior to the run-off, leaving Mr. Karzai to be declared the winner [6, pp 37-38].

Investigating allegations of voter fraud is a difficult and time-consuming process. Statistical methods have been developed which can augment - but probably not replace - this necessary task. A significant advantage of these techniques is that they do not require the participation of on-site observers or investigators. Access to the tabulated voting tallies alone is usually sufficient to carry out the necessary analysis.

Beber and Scacco [4] have developed one such model for analyzing voting tallies. Their methods exploit the apparent difficulties human subjects experience when attempting to fabricate statistically random numbers. For the 2009 Afghanistan presidential election, Weidmann [26] applied this model to map anomalous patterns of voting at the province level. The discrepancies identified appear to be correlated with the patterns of voting fraud reported by the ECC. In the analysis to be described, we use several tests suggested by Beber and Scacco to analyze the raw voting data collected within a single Afghan province - Helmand. More specifically, we analyzed selected statistical signatures for the four leading presidential candidates, and attempt to identify any biased patterns that are present. One limitation is that the proposed model is invalid for small raw vote counts, and the latter were reported for many local candidates entered in the presidential election.

Manuscript approved July 26, 2011.

Our study was guided by two questions:

(1) Are there significant differences between the statistical signatures of the leading candidates?

(2) Did the subsequent vote certification process rectify the observed biases?

2) Statistical Models for Anomaly Analysis

a) The Beber-Scacco Model

In both controlled laboratory and uncontrolled non-laboratory settings, human beings show distinct biases when asked to generate 'random' numbers. In some experiments, subjects favored numbers that terminate in low digits over choices that would result in a uniform distribution of the terminal digits. In addition, human subjects tend to underestimate the probability that any two consecutive penultimate (next to last) and terminal digits will be identical when generating a sequence of random numbers. If the digits from zero to nine are uniformly distributed, the probability of two consecutive random digits being identical is 10%. Based on the experimental psychology literature, Beber and Scacco concluded:

> *The literature thus suggests three key findings: (1) digits are not selected with equal frequency, (2) repetition is avoided, and (3) serial sequences are preferred.* (4, p. 8)

Given these findings as motivation, the investigators developed a model for analyzing the distribution of the last and penultimate digits for raw vote counts. The purpose of the model is to identify biases in the distribution of the vote count that might suggest manipulation or fabrication of the data. The authors state:

> *We first show that last digits will occur with equal frequency for a large class of theoretical distributions, and we argue that non-fraudulent electoral returns are likely drawn from such a distribution.*(4, p. 2). [See Figure 1.]

For the above statement to be true, three propositions must be satisfied. (See 4, Appendix). An important exception is that the third proposition - and therefore the assumption that the last digits of the vote tallies are uniformly distributed - is invalid if there is a "non-trivial probability of observing less than eighteen votes for a unit of interest" [4, p.5]. This can and did happen for many of the minor candidates in the 2009 Afghanistan presidential election. The implications of deviation from the model's assumptions will be discussed subsequently. An additional implication is that this model is only applicable to ballot box level data (i.e. raw vote counts). The assumption that the ultimate and penultimate (next to last) digits are uniformly and independently distributed will no longer be true if the data is averaged, or otherwise aggregated.

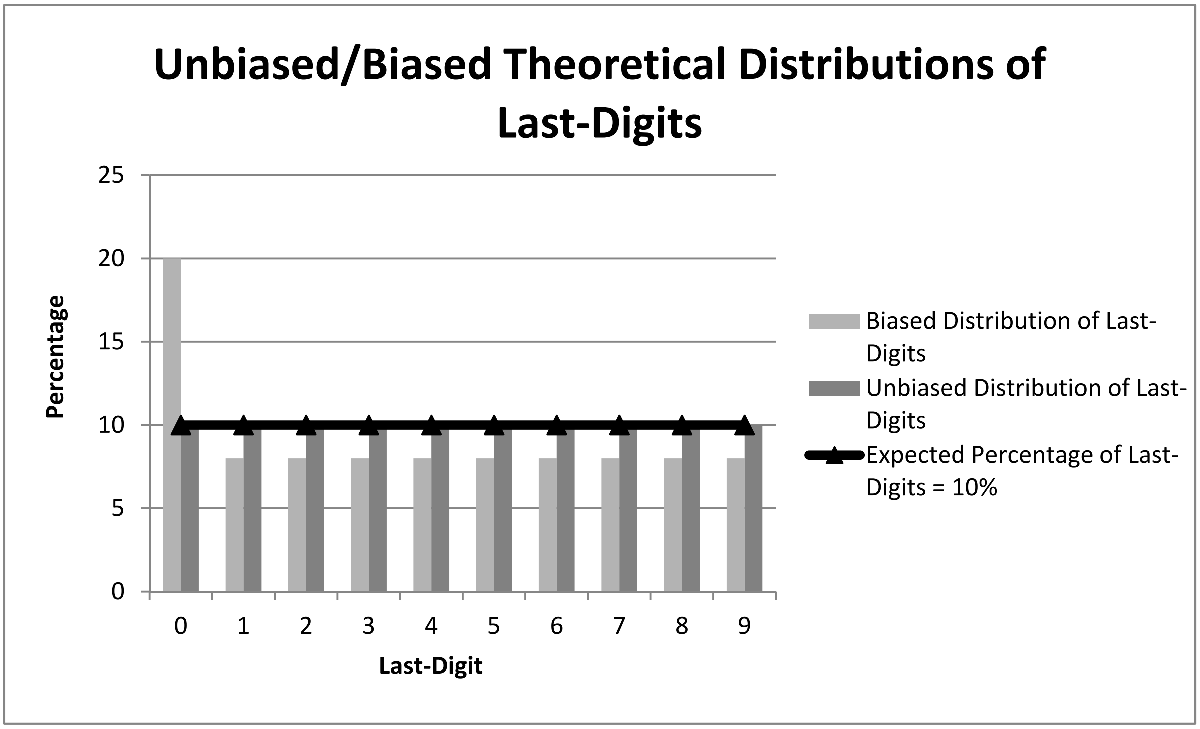

Figure 1 - This chart shows the expected uniform unbiased distribution of random last digits for vote tallies. Also shown is a hypothetical distribution in which the percentage of terminal '0's (zeros) significantly exceed the expected value of 10%.

Given that three propositions are satisfied, the authors prove that the expected number of repetitions $\hat{R}ep$ (i.e. consecutive draws of the same last digit) is:

Equation 1 - $$\hat{R}ep = N - 1/b$$

Let N=2, and b = 10 [the ten digits from 0 to 9]. The probability that the terminal and penultimate digits will be identical is therefore (2-1)/10 = .1

Beber and Scacco have applied their model to the analysis of two elections. In the case of the Swedish parliamentary election [2002], the assumption of uniformly-distributed last-digits was supported. In the case of the Nigerian presidential election [2003], in which wide-spread fraud was documented, this assumption was significantly violated. In the frequencies of the last digits, zero was substantially overrepresented [4, p. 13]. Interestingly, on the same return sheets, zero is also overrepresented in counts of the last digit of the number of *registered* voters (not the vote count), and the last digit of the number of votes received by the incumbent People's Democratic Party (PDP). Leaders of the PDP were particularly flagrant in terms of packing the ballot boxes, and directing voters on how to complete their ballots. In a large number of wards, the authors also note that digit pairs are adjacent much more frequently than expected, based on the assumed uniform distribution of the digits.

In Weidmann's analysis of Afghanistan voting patterns [26], overrepresentation of the terminal digit '0' was the most common finding. Other biases, such as the overrepresentation of the digit '5,' were also noted.

b) Other Forensic Applications of Probability Theory

Given the apparent difficulty humans have in fabricating truly random numbers, investigators have also used probability models for analyzing other types of data. Given any sufficiently large set of nominally random number sequences, irrespective of their source, appropriate statistical tests may be used to determine if the criterion of randomness is satisfied.

O'Kelly [18] asked qualified physicians to fabricate scores on a standard scale for depression for subjects at three different sites. This data was then interleaved with authentic data from an additional 18 sites. A statistician who examined the data for unusual means and correlations succeeded in identifying one of the three target sites correctly, while incorrectly flagging a second legitimate site. Although this study was only partially successful, it does suggest the utility of probability theory for helping to identify fraudulent data.

Mosimann et al. [14, 24] review the application of the analysis of terminal digits to a variety of scientific data sets. In all four cases, manipulation or fabrication of data was strongly suspected. These studies were conducted by the Office of Research Integrity, DHHS (United States Department of Health and Human Services). The first two cases involved the analysis of published data - falsified scintillation counts for a series of radioassays, and partially falsified data enumerated for bacterial cultures. The third case discusses the analysis of the distribution of muscle fiber action potentials (i.e. muscle fiber "firings"), and the fourth analyzes blood flow measurements purportedly made on the hind limb of rats for two different experimental conditions. In each of these four cases, the last digit analysis helped to confirm the partial fabrication of reported data.

In an intriguing paper, Nigrini [17] discusses the significance of analyzing the distribution of *initial* digits for detecting anomalies in financial data. Benford's Law asserts that for many classes of financial and demographic data, the *first* digit is more likely to be low [1, 2, or 3] than high [8, 9, 10]. Specifically, Benford's law asserts that the probability that the first digit D_1 will equal a digit d_1 between 1 and 9 is:

Equation 2 - $$P(D_1 = d_1) = \log 10(1 + 1/d_i) \; for \; d_1 \hat{i} \{1,2,..,9\}$$

Nigrini demonstrates that a series of 23 fraudulent checks (17, p. 82) violate the above observation. Although the perpetrator attempted to randomize the amount of each check, the initial digits for the amounts were typically 8 or 9, and occasionally 7. In some cases, this approach can be generalized further by examining the distribution of values for the first two digits of selected financial variables.

c) Biases in Random Number Generation by Human Subjects

i) Overview

Many studies have reported biases when human subjects are asked to either construct random sequences of binary or decimal numbers - or conversely, to differentiate between random and non-random sequences. These findings, however, are more ambiguous than the above summary would indicate [4]. Vague or inaccurate instructions, as well as a lack of real-time feedback, can both exacerbate the biases observed in the human generation of "random" sequences of numbers. We should also note that a machine-generated sequence of numbers may be judged as *subjectively* random, yet still be nonrandom in a statistical sense [21].

As an example of study which clearly demonstrated the human biases, Boland and Hutchinson [5] asked subjects to generate random sequences of 25 digits from the set of ten digits [0, 1, 2, … , 8, 9]. They found that the digit '0' was substantially underreported, and that the incidence of repeated digits was lower than predicted for truly random sequences.

Many other studies have investigated the generation of binary sequences of numbers or tokens. In a review of the literature, Nickerson [16] reported that the most common finding is that subjects asked to generate random *binary* sequences produce sequences for which the frequency of alteration (i.e. the frequency of transitions from '1' to '0' and vice versa) is greater than that predicted by probability theory. When asked to identify random sequences of binary numbers, subjects also tend to select sequences for which the frequency of alteration is higher than would be produced by a truly random process.

It is almost certainly erroneous to conclude that people are inherently incapable of generating or recognizing random sequences of numbers. This is an important point, because of its bearing on the question as to whether knowledgeable individuals could successfully fabricate data that could evade detection by the type of statistical test describe herein. For one, biases in the generation of random sequences may sometimes reflect imprecise direction on the part of the experimenter. Given vague or ambiguous direction, subjects may apply private criteria for generating random sequences, and this criterion may differ substantially from the expectations of the experimenter. Nickerson concluded:

> *"I have argued in this article that efforts to produce or perceive randomness must be judged in light of the specifics of the tasks as understood by the people who perform them. When instructions are vague, it is difficult to know what to make of whatever people do. I believe that ambiguous or imprecise instructions to participants have been*

> *factors in a sufficiently high percentage of experiments on randomness production or perception that the results, in the aggregate, do not constitute a compelling case for the conclusion that people generally have faulty conceptions of randomness. (p. 353).*

ii) Random Number Generation with Feedback

More so, given active feedback during the course of an experiment, subjects are capable of generating sequences that are nearly random. Two examples are given below.

Rapoport and Budesco [20] analyzed the behavior of students playing a two-person competitive game. Each student was asked to pick a black card or a red card; one student won if the cards matched, the other won if they didn't. The first group of players was provided feedback on each turn with respect to the outcome, prior to making their next selection (Condition D, for dyad). The second group of players was asked to list their 150 selections in advance i.e. the sequence or black and red cards (Condition S, for single). As a control, a third group of students selected the color of the card to be played by mentally simulating a coin toss 150 times (Condition R, for randomization). It might appear that Condition R and S used identical protocols, because in both cases, the students were asked to generate a sequence of 150 choices *a priori*. However, members of the S group were asked to generate a viable sequence of choices that would subsequently be tested against an opponent, whereas students in the R group were simply instructed to treat each choice as an independent simulated coin toss.

The sequences were analyzed with respect to runs (i.e. consecutive cards of the same value) and sequential dependencies for the conditional probabilities between choices. The sequences generated for Condition D were more nearly random than the sequences generated in the R group. The sequences generated in Condition S were intermediate. Because each player viewed the other's player's choice at each turn, his opponent was given the opportunity to strategically exploit any *nonrandom* aspects of the sequence of card choices presented. This protocol enforced the generation of effectively random choice sequences of choices by both players. The outcome supported the investigator's hypothesis that "*The hypothesis is that in this situation [Condition D] people would display randomlike behavior not because they are instructed to do so but because this is the best and most effective way of achieving a desired goal* [20, p. 360]."

Neuringer [15] also provided feedback for his subjects, but in the context of generating random sequences that were compared with the output from a random number generator (RNG). The subjects were asked to create random binary sequences ('1's and '2's) of 100 numbers. Following a baseline session, in which the subject-generated sequences were demonstrably non-random, the subjects were provided with statistical feedback after each 100 number sequence had been generated. Based on the five statistical measures provided, all subjects learn to generate sequences that were indistinguishable from those produced by a random number generator. The sequences were not completely random, as could be demonstrated by using additional statistical tests, but the latter were not used to provide feedback to the subjects during the experiment.

3) Data Sources

The models that we will apply require unaggregated (i.e. "raw") vote counts. In order to understand the subsequent data analysis, it is necessary to briefly review the protocols for the prior Afghanistan presidential election. Afghanistan has 34 *provinces*; each province is divided into *districts*. There are usually multiple *Polling Centers* within each district, and each Polling Center will usually have multiple *Polling Stations* - usually several for males, one for females, and sometimes a station for Kuchi (Afghan Pashtun nomads) (Figure 2). Within the Helmand Province, 404 polling stations were eventually certified, and 144 polling stations were uncertified, for a total of 548 polling stations. The tallies reported for the polling stations are at the level of the ballot box.

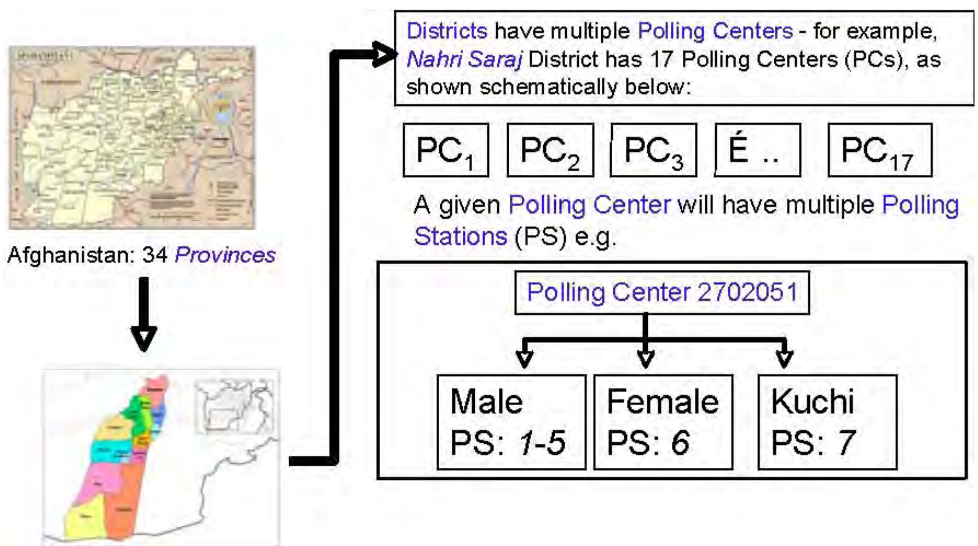

Figure 2 - This chart illustrates the relationship between polling centers and polling stations. Polling Center 2702051 is located in the Nahri Saraj district of Helmand Province.

All statistical analysis ["last-digit analysis"] was conducted at the polling station level. Using the aggregated vote counts at the polling center level would invalidate the statistical model, as discussed earlier. Tabulated lists of the unofficial vote tallies were obtained for 16 September 2009 [10]; this list included all reported vote counts, including vote counts for polling stations that were eventually disqualified [9]. The final certified presidential results were also obtained, and the former was partitioned into certified and uncertified vote counts.

Obviously, the number of "last digits" that can be tabulated for each candidate in Helmand cannot exceed the number of polling stations in a given category (i.e. for Helmand, Certified: 404, Uncertified: 144, Total: 548. See Table 1.) Data were analyzed for the top-four candidates in Helmand (Hamed Karzai, Dr. Abdullah Abdullah, Ashraf Ghani Ahmadzai, and Ramazan

Bashardost). These four individuals accounted for 92.54% of the vote in Helmand. As Table 2 illustrates, the percentage of polling stations reporting one or more votes for each candidate varied from about 44% for Ramazan Bashardost to over 86% for Hamed Karzai. (If a candidate receives zero votes at a given location, these statistical methods cannot be applied.)

	Polling Stations in Helmand
Certified	*404*
Uncertified	*144*
Total	*548*

Table 1 - Total Count of CERTIFIED/UNCERTIFIED polling stations in Helmand.

Top-Four candidates	**Ashraf Ghani Ahmadzai**	**Abdullah Abdullah**	**Hamed Karzai**	**Ramazan Bashardost**
# CERTIFIED Polling Stations	**404**	**404**	**404**	**404**
# CERTIFIED vote counts > 0	***251***	***254***	***350***	***181***
% CERTIFIED non-zero votes out of possible maximum	***62.13%***	***62.87%***	***86.63%***	***44.80%***

Table 2 - The percentage of certified polling stations reporting one or more votes for each candidate ranged between 44.8% and 86.63%.

4) Observed Biases

(a) Low-Digit Bias:

There is a greater bias towards higher counts than expected for the lower digits of '1', '2', and '3.' This bias occurs because candidates other than Karzai often received low vote counts of 1, 2, or

3 votes at many different polling station locations (Figure 3). This is unsurprising, given Karzai's dominance in the total vote counts (see Table 2 preceding, and Tables 3 and 4 on page 12).

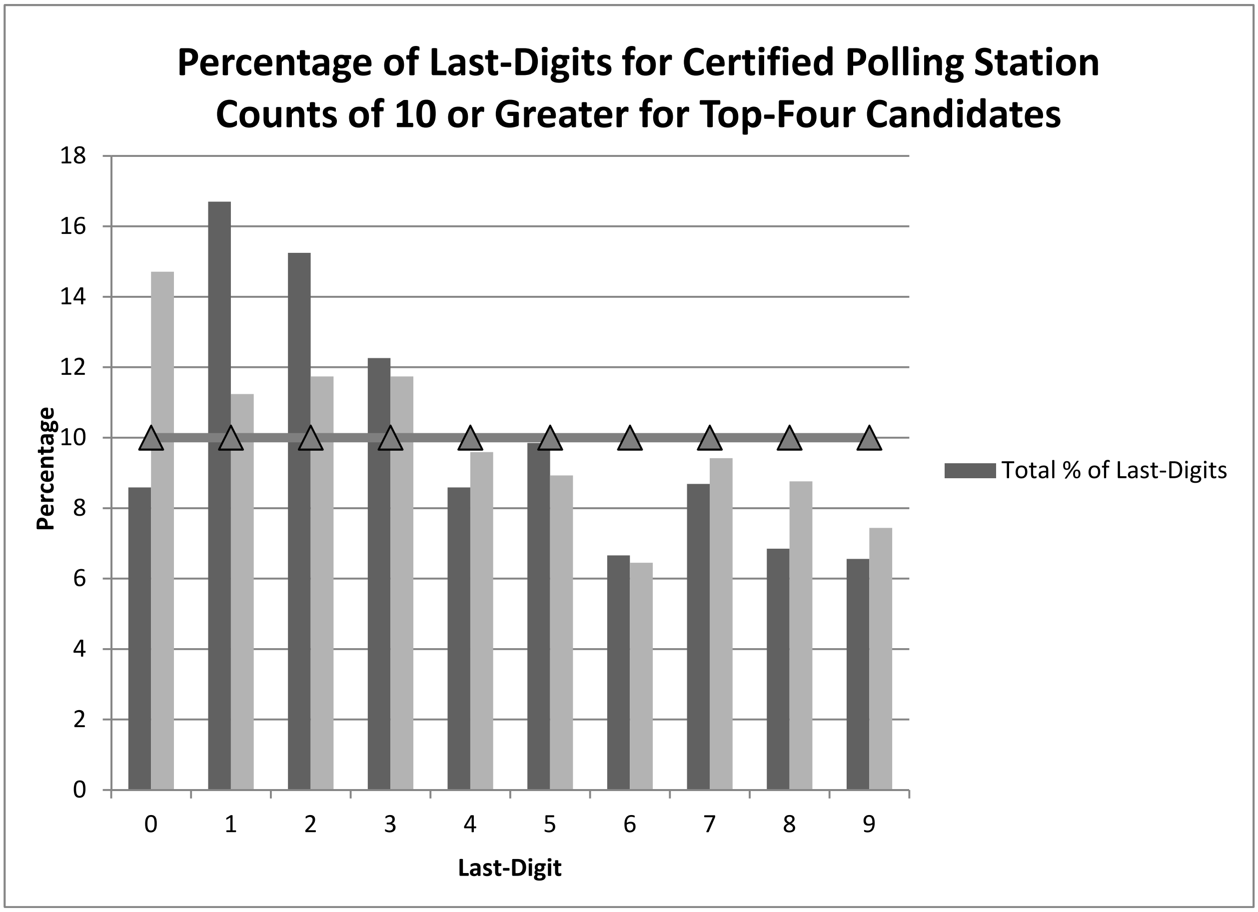

Figure 3 - "Low-Digit" Bias (Add the rest of the key. E.g. Expected Value)

As shown in Figure 3, when vote counts greater than 10 are tabulated, this bias is largely eliminated. (In addition, vote counts of less than 18 violate the assumptions of the Beber-Scacco model.) There is still an apparent under-representation of the digit 6.

(b) Zero-Digit Bias:

Karzai's raw vote counts contained a disproportionate number of values that were rounded off to zero, e.g. 50, 600, etc. None of the distributions for the other three candidates displayed a comparable bias. In Karzai's certified vote counts, only 12% of the last-digits terminated in zero, but in the uncertified counts, almost 35% did (Figure 4).

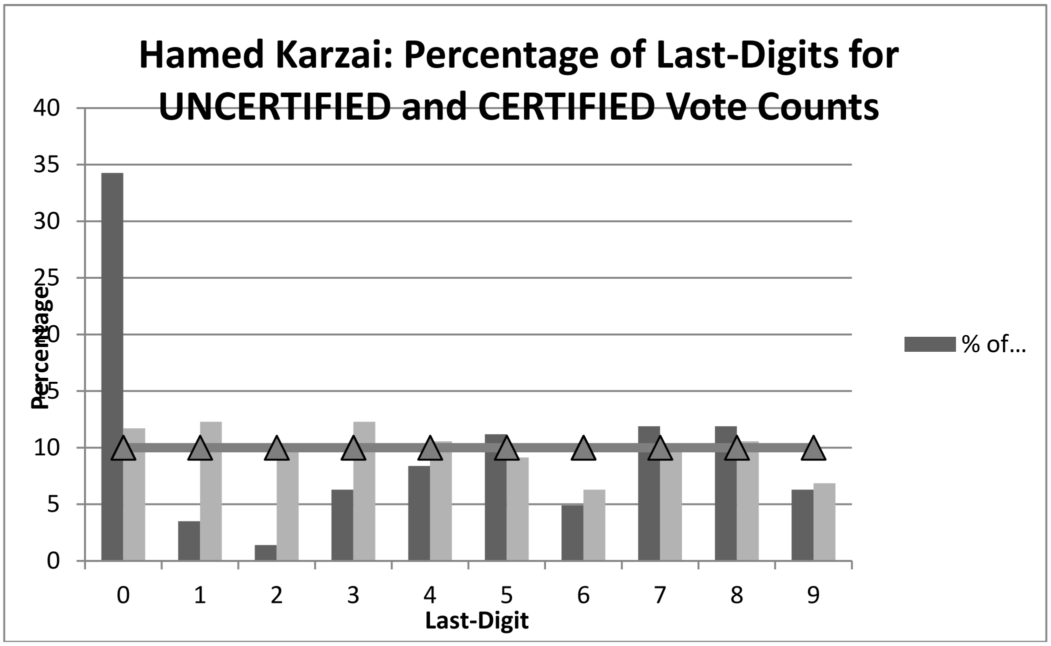

Figure 4 - Percentage of Last-Digits for both Certified and Uncertified Vote Counts (Add the rest of the key. E.g. Expected Value)

A Chi-squared Goodness-of-Fit-Test was conducted for the last-digit distribution of the *certified* vote counts for Karzai. The null hypothesis is that the samples were drawn from a uniform underlying distribution. The Chi-square value is 13.48, with 9 degrees of freedom. The two-tailed p value is .1418; at a significance level of p = .05, the null hypothesis can no longer be rejected. Because the sample sizes are small [27, see p. 822], the test was also replicated using Fisher's exact test for small sample sizes [13], with comparable results. More colloquially, the distribution of the last digits for Karzai's certified votes cannot be statistically differentiated from the hypothetical unbiased distribution illustrated in Figure 1.

In contrast, Figure 5 on the following page shows a comparable distribution for Abdullah. The digit zero is underrepresented for both the certified and uncertified last-digit analysis. The digits '1', '2', and '3' are overrepresented partially due to the low-digit bias. This is apparent in Figure 6. The distribution of the last-digits for vote counts of 10 or greater more nearly approximate the expected value of 10.

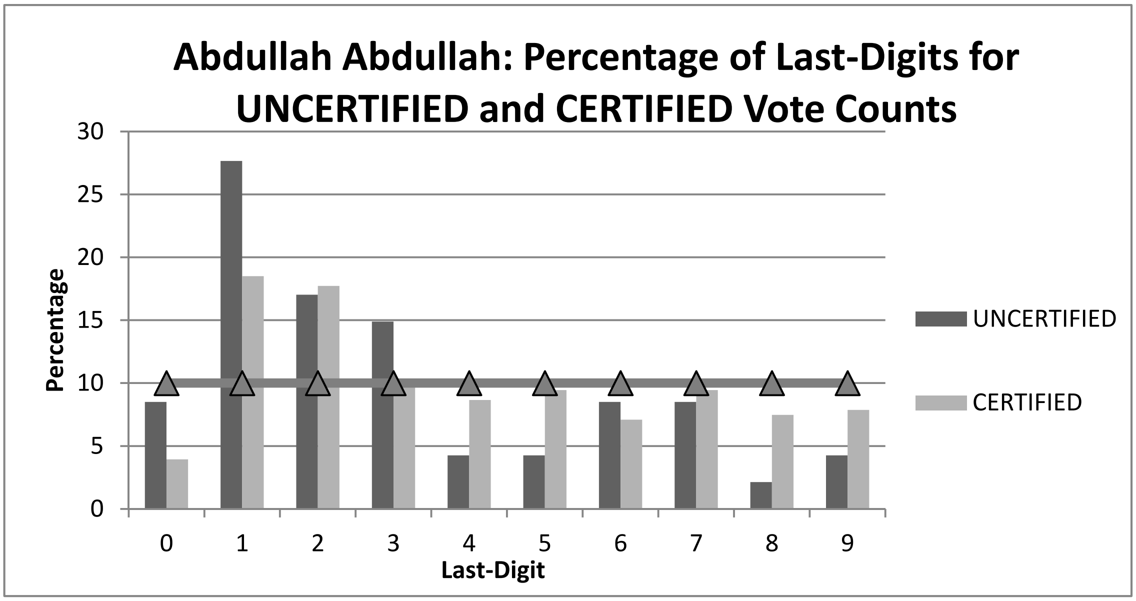

Figure 5 - Percentage of last-digits for Abdullah CERTIFIED and UNCERTIFIED vote counts

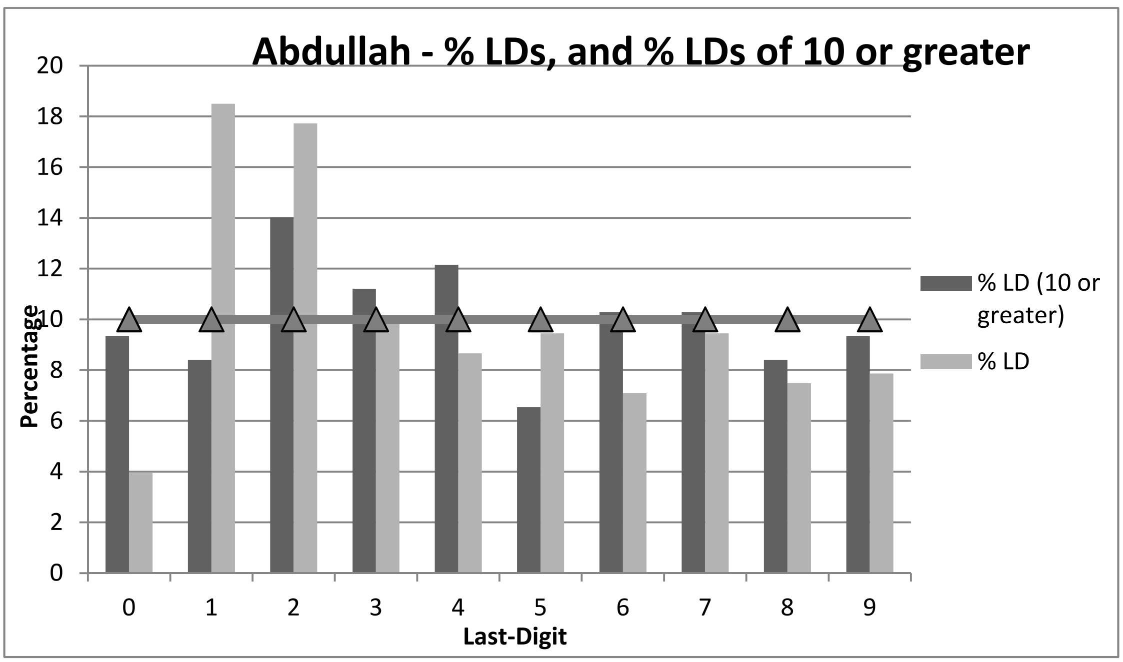

Figure 6 - Percentage of last-digits of Abdullah vote counts (1) for all last digits (2) for all last-digits greater than 10

	Ashraf Ghani Ahmadzai	Abdullah Abdullah	Hamed Karzai	Ramazan Bashardost	Total Count
0	3	4	49	2	58
1	9	13	5	2	29
2	6	8	2	5	21
3	4	7	9	2	22
4	7	2	12	3	24
5	3	2	16	4	25
6	6	4	7	3	20
7	1	4	17	1	23
8	1	1	17	2	21
9	1	2	9	2	14
Non-Zero Counts	41	47	143	26	257
Total Votes	676	716	65602	679	58380

Table 3 - Last-Digit Counts for UNCERTIFIED Vote Counts Only

	Ashraf Ghani Ahmadzai	Abdullah Abdullah	Hamed Karzai	Ramazan Bashardost	Total Count
0	29	10	41	9	89
1	45	47	43	38	173
2	45	45	36	32	158
3	36	25	43	23	127
4	15	22	37	15	89
5	26	24	32	20	102
6	19	18	22	10	69
7	14	24	35	17	90
8	10	19	37	5	71
9	12	20	24	12	68
Non-Zero Counts	251	254	350	181	1036
Total Votes	4643	3468	47271	2998	58380

Table 4 - Last-Digit Counts for CERTIFIED Vote Counts Only

(c) The '600' Anomaly

Some reported vote counts were so anomalous that we initially considered the possibility that clerical errors had occurred during the tabulation process. If not, these counts would appear to provide prima facie evidence of electoral fraud at selected polling stations, a conclusion consistent with the subsequent findings of the ECC. A singular example is that exactly 600 votes were allegedly cast for Hamed Karzai at 24 different polling stations, in 13 different polling centers in Helmand. No votes were recorded for any other candidate at these stations. The results from these polling stations were disqualified in final certified vote counts.

Location	Polling Center	Polling Station	Ashraf Ghani Ahmadzai	Abdullah Abdullah	Hamed Karzai	Ramazan Bashardost	Other Candidates
PROVINCIAL CENTER 1	2701003	7	0	0	600	0	0
PROVINCIAL CENTER 1	2701016	6	0	0	600	0	0
PROVINCIAL CENTER 1	2701017	4	0	0	600	0	0
PROVINCIAL CENTER 1	2701029	1	0	0	600	0	0
NAHRE SARAJ	2702034	1	0	0	600	0	0
...	2702034	5	0	0	600	0	0
NAHRE SARAJ	2702043	3	0	0	600	0	0
PROVINCIAL CENTER 1	2702047	1	0	0	600	0	0
...	2702047	2	0	0	600	0	0
NAHRE SARAJ	2702051	4	0	0	600	0	0
NAD ALI	2703078	3	0	0	600	0	0
NAWEI BARUKZAI	2704089	1	0	0	600	0	0
...	2704089	2	0	0	600	0	0
NAWEI BARUKZAI	2704107	1	0	0	600	0	0
...	2704107	2	0	0	600	0	0
...	2704107	3	0	0	600	0	0
...	2704107	4	0	0	600	0	0
MOOSA QALAH	2709178	1	0	0	600	0	0
...	2709178	2	0	0	600	0	0
...	2709178	3	0	0	600	0	0
...	2709178	4	0	0	600	0	0
MOOSA QALAH	2709185	1	0	0	600	0	0
...	2709185	2	0	0	600	0	0
...	2709185	3	0	0	600	0	0

Table 5 - List of 24 polling stations reporting 600 votes for Hamed Karzai

The Electoral Complaint Commission (ECC) adjudicated each of the above cases (ECC, Guardian article). Findings for several polling stations are listed below, although the investigation summaries quoted are no longer available on-line ("Afghanistan election investigation," Excel spreadsheet downloaded 8 DEC 2009 from ECC):

> *2701017 (PS 4): 100% uniform markings; no used materials in box; all votes for one candidate (votes at other stations of center distributed among at least five candidates)*
>
> *2704107 (PS 4): Forms not in box; 100% uniform markings; 100% of ballots never folded.*

2702051 (PS 1): 100% uniform markings; 100% of ballots never folded.

2702051 (PS 4): 100% uniform markings; 2 seals missing; reconciliation form matches.

Although this example is particularly egregious, these anomalies were also captured by the statistical approach described in this paper. The digit '0' was overrepresented as a terminal digit in the raw tabulations for Karzai, and the terminal double-digit zero-zero ('00') was also overrepresented.

Obviously, a statistical approach cannot capture all types of potentially fraudulent behavior. It was noted (25) that *"some polling stations showed competing vote-rigging. At Faselah compound in Zarghoon Shaher, Paktika, all the votes showed identical markings, none of the ballots was folded, and all 600 votes went to one candidate – but they were recorded as votes for someone else."* The deletion of votes for a particular candidate, or the deliberate reassignment of votes cast for one candidate as votes for a different candidate, cannot be detected by statistical methods alone.

(d) The Double-Digit Distribution

Beber and Scacco [4] also examined the distribution of the two terminal (penultimate and ultimate) digits of the reported voting counts. If the assumptions of the model are satisfied, the probability that the last two digits will be identical is 10%.

Given the data at hand, the relatively low vote counts create a difficulty. As an example, there are 350 vote counts greater than zero reported for Karzai for the certified counts (see Table 4). (i.e. 350 is the number of polling stations at which Karzai received one or more votes; this number should not be confused with the total *sum* of the vote count of 47271.) The expected value for the total number of doubled terminal digits is therefore 0.10 X 350 = 35. Since there are ten digits, the expected number of doubled terminal digits for each of the ten digits (i.e. the possible doubled terminal digits are '00', '11', '22', '33', '44', '55', '66', '77', '88', '99') is only 3.5. This violates the assumption of proposition three [4] that the expected values for any count will be at least 18. Even given this limitation, it is still instructive to examine the distribution of terminal double-digits for Karzai.

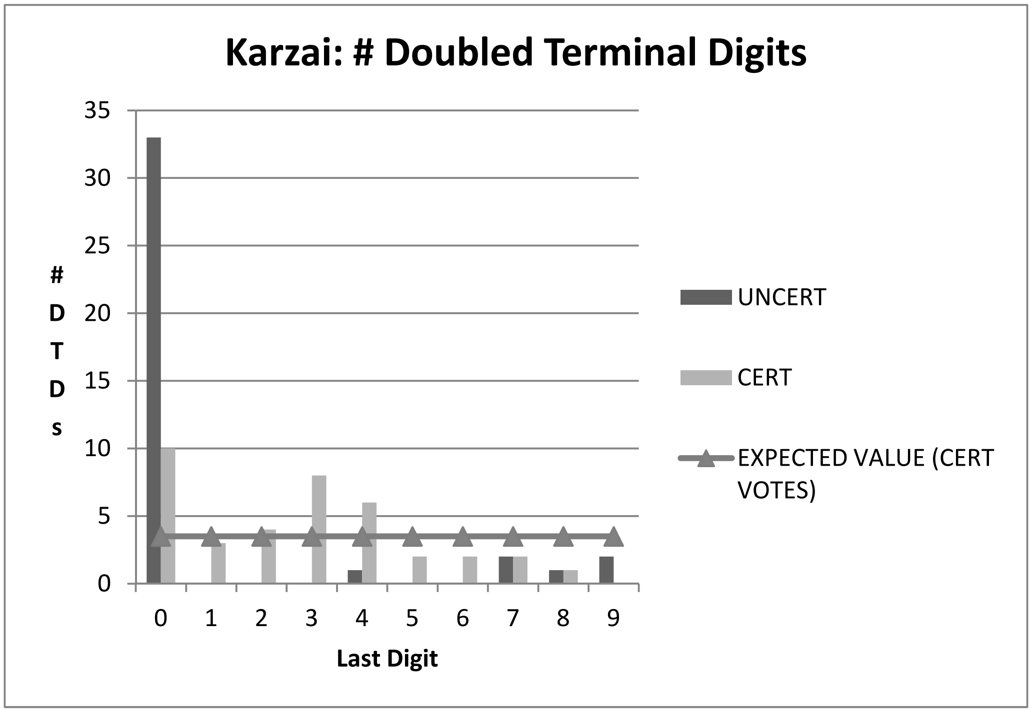

Figure 7 - Karzai terminal double-digit frequencies for UNCERTIFIED and CERTIFIED vote counts.

As can be seen in Figure 7, the frequency of terminal double-digits much more closely approximates the expected value of 3.5 in the *certified* data than in the *uncertified* data. The number of terminal double-digit zeros (i.e. '0-0') has been reduced from 33 to 10.

In the certified data, there were a total of 38 doubled terminal digits, and 312 single terminal digits. Given expected proportions of 1:9, the chi-square was computed to be .198, with p = 0.656 for a two-tailed test. The null hypothesis that the total number of doubled digits deviates from the expected value of 3.5 cannot be rejected at a significance level of 0.05.

5) Conclusions

Two different types of biases were observed: a "low-digit" bias for the four top candidates (i.e. the lower-digits were over-represented as the terminal digits of the raw voting counts), and a "zero-digit" terminal-digit bias for Hamed Karzai alone. The first was shown to be a statistical artifact; the second, a real effect that was eliminated in the certified vote counts. This suggests that the zero-digit statistical anomaly observed in Karzai's data was both real and significant. This conclusion was verified by examining the anomalous reporting of exactly 600 votes for Hamed Karzai at 24 different polling stations. More so, this statistical bias was absent in the final certified vote counts, because these aberrant tallies were discarded by the ECC. The vote certification process was not based on statistical methods, but our findings appear to be consistent with the outcome of the IEC certification process. The analysis of statistical voting anomalies may be useful as an independent adjunct method for monitoring and enforcing legitimate electoral processes.

The findings that we have presented are agnostic with respect to the motivation and mechanism of the voting irregularities identified. Given the lead possessed by Mr. Karzai going into the election [1], the adverse publicity garnered would seem to outweigh any possible advantage that might be gained through widespread election fraud [8]. Simpser [22, see also 23] has constructed a database of worldwide country level elections for executive office [1990-2007]. In 22 percent of the cases electoral corruption was wide spread. In forty percent of the cases where electoral corruption was widespread, the voting margins for the winners exceeded 40% [23, p. 9]. Excessive margins of victory occur frequently in corrupt elections. Simpser identified at least five possible motivations for large-scale electoral corruption:

1. *Uncertainty about the outcome*
2. *Low marginal cost of obtaining corrupt votes*
3. *High stakes for retaining power*
4. *Local demonstrations of loyalty and effectiveness - ability to 'deliver the vote'*
5. *To influence future voter behavior and participation*

To the above list, I would add two additional possible motivations:

6. *Raw demonstration of local power/authority*
7. *A "false-flag" demonstration intended to embarrass the inevitable winner*

Although any of these are possible factors in a given election, Simpser favors influencing future voter behavior and participation as a major factor in the majority of cases. His hypothesis is that overwhelming victories by one party may suppress future voter turnout by the opposition - particularly if there are mechanisms that can reward or punish voters as a function of their choices [22, p. 20]. This argument is advanced through the analysis of a formal multi-equilibrium voting model that incorporates these assumptions. In Simpser's model, multiple equilibrium may exist for the same set of model parameterizations, given different model inputs. For example, there could be one equilibrium for a scenario in which a low opposition turnout creates a high probability of an incumbent victory, leading to suppressed turn out of the opposition at the next election – and a subsequent incumbent victory. Conversely, there may be another model equilibrium for a high opposition turnout that creates a low probability of an incumbent victory, while increasing the probability for a high opposition turnout – and opposition victory - in the next election as well. [22, p. 21].)

Given the data at hand, there is no way to test which of the above factors - if any - played a significant role in the 2009 Afghanistan presidential election. We present these considerations only to demonstrate that an "excessively" corrupt election may provide long-term benefits to the winner, and these benefits may be perceived as justifying the risk and adverse publicity so engendered.

If knowledge of the statistical techniques described herein becomes widespread, an interesting question is whether knowledgeable individuals could exploit this information to modify or fabricate electoral counts in a way that evades statistical detection. With proper training and/or feedback, human subjects are capable of constructing sequences of numbers that are effectively random [15]. Sets of random numbers in the desired range of values could also easily be distributed to members of a group in advance, or in real-time by cell phone or texting. The question as to whether such strategies could be realistically implemented under the time pressures imposed by a vote count, and in the presence of neutral observers, remains an open question. The statistical biases described here, such as non-uniform distribution of ultimate digits, are manifest at the district or country level, but not at the level of a single polling station. A high level of centralized planning would be required in order to successfully fabricate *apparently* unbiased distributions of voting tallies at the district or county level.

Acknowledgements

The Joint IED (Improvised Explosive Device) Defeat Organization (JIEDDO) funded this study, and the author is pleased to acknowledge their support. The author would also like to acknowledge the invaluable assistance of Ruth Willis, Principal Investigator; Denise Royle, for assisting in data collection, Joel Alejandre, who helped partition the polling station data into certified and uncertified data sets, and Carol Chang, for assisting with aspects of the statistical analysis. All errors of omission or commission are solely those of the author.

References

1. Abrashi, Fisnik. "Karzai Leads in Polls Going Into Afghan Presidential Election (14 August 2009)." *Associated Press.* http://www.huffingtonpost.com/2009/08/14/karzai-leads-in-polls-goi_n_259640.html
2. Afghanistan Election Data. http://afghanistanelectiondata.org/.
3. Afghanistan Election Data document. CSV file containing polling station counts for top four candidates - both certified and uncertified. Derived from IEC report "Final Uncertified Presidential Results (16 September 2009)."
4. Beber, Bernd and Alexandra Scacco. "What the numbers say: a digit-based test for election fraud using new data from Nigeria." Working paper available from http://www.columbia.edu/~bhb2102/files/Beber_Scacco_ElectionFraud.pdf.
5. Boland, Philip J. and Kevin Hutchinson. "Student selection of random digits." *Statistician*, 49(4): 519-529, 2000.
6. ECC (Electoral Complaint Commission) Report. "Electoral Complaints Commission Final Report 2009 Presidential and Provincial Council Elections (April 2010)." Available from http://www.ecc.org.af/en/ Last accessed 27 June 2010.
7. Gibbons, Jean Dickinson and Subhabrata Chakraborti. *Nonparametric Statistical Inference: Fourth Edition.* Marcel Dekker, Inc., 2003.
8. Green, Matthew. "Karzai under fire over poll reform." *Financial Times*, 18 April 2010, http://www.ft.com/cms/s/0/2a327796-4b05-11df-a7ff-00144feab49a.html.
9. IEC (Independent Election Commission). "Final Certified Presidential Results (2009)." Available from http://www.iec.org.af/results/Index.html or http://afghanistanelectiondata.org/about/election-data . Last accessed 27 June 2010.
10. IEC (Independent Election Commission). "Final Uncertified Presidential Results (16 September 2009)." Available from http://afghanistanelectiondata.org/about/election-data. Last accessed 27 June 2010. (No longer available from http://www.iec.org.)
11. IEC (Independent Election Commission). "20091021_PollingStations_Disqualified_ByECCBasedOnComplaints." Available from http://www.iec.org.af/results/Index.html. Last accessed 27 June 2010.
12. IEC (Independent Election Commission). ""20091021_PollingStations_QuarintinedAndDisqualified." Available from http://www.iec.org.af/results/Index.html. Last accessed 27 June 2010.
13. McDonald, J.H. *Handbook of Biological Statistics (2nd Ed.)*. Sparky Publishing House, Baltimore, MD. pp. 70-75, 2009.
14. Mosimann, James E., John E. Dahlberg, Nancy M. Davidian, John W. Krueger. "Terminal Digits and the Examination of Questioned Data." In *Investigating Research Integrity: Proceedings of the First ORI Research Conference on Research Integrity* (Nicholas H. Steneck and Mary D. Scheetz, Editors). US Dept of Health and Human Services, pp. 269-278, 2002.
15. Neuringer, Alan. "Can People Behave "Randomly?": The Role of Feedback." *Journal of Experimental Psychology: General*. 115: 62-65, 1986.

16. Nickerson, Raymond S. "The perception and production of randomness." *Psychological Review*, 109(2): 330-357, 2002.
17. Nigrini, Mark J. "I've got your number: How a mathematical phenomenon can help CPAs uncover fraud and other irregularities." *Journal of Accountancy*, 187(5), 1999.
18. O'Kelly, Michael. "Using statistical techniques to detect fraud: a test case." *Pharmaceutical Statistics*, 3: 237-246, 2004.
19. Python-Statlib. Software and documentation available from http://code.google.com/p/python-statlib/.
20. Rapaport, Anon., David V. Budescu. "Generation of random series in two-person strictly competitive games." Journal of Experimental Psychology: General, 121: 352-363,1992.
21. Sanderson, Yasmine B. "Effective generation of subjectively random binary sequences." *Advances in Applied Mathematics*, 43: 1–11, 2009.
22. Simpser, Alberto. "Cheating big: on the logic of electoral corruption in developing countries (18 May 2008)." Available from http://home.uchicago.edu/~asimpser/Research.html.
23. Simpser, Alberto. "The Manipulation of Mass Elections: Corruption, Expectations, and Turnout." Presented at the *2004 Midwest Political Science Association Meeting*, Chicago, IL.
24. Walter, Charles F. and Edward P. Richards, III. "Using Data Digits to Identify Fabricated Data." *IEEE Engineering in Medicine and Biology*, 20(4): 96-100, 2001.
25. Weaver, Mathew. "A catalogue of errors that cost Karzai victory (19 October 2009)." *Guardian.* http://www.guardian.co.uk/world/2009/oct/19/afghanistan-election-karzai-fraud
26. Weidmann, Nils. "Mapping election fraud in Afghanistan." http://nils.weidmann.ws/projects/afghanelection/.
27. Winkler, Robert L. *Statistics: Probability, Inference, and Decision. Second Edition.* Holt, Rinehart, and Winston, 1975.