
Social Security Number Validation

September 2, 2010

Quality Control Systems Corporation • www.quality-control.us

Simulated Performance of a Technique for Social Security Number Validation in Customer Identification Programs



– White Paper –

Social Security Number Validation

September 2, 2010

Quality Control Systems Corporation • www.quality-control.us

Introduction

All banks in the United States must have Bank Secrecy Act/Anti-Money Laundering (BSA/AML) compliance procedures. These procedures include a Customer Identification Program (CIP) requiring a taxpayer ID. Yet private institutions have no means to directly verify with the Social Security Administration (SSA) the issuance of a Social Security Number to a specific, living person outside the context of employment or without signing a remarkably stringent agreement with SSA, specifying record retention policies and compliance reviews.

Fortunately, a simpler, effective solution does exist. Social Security Numbers (SSNs) can be validated against a list or table of numbers that were actually issued on a date that is consistent with a person's age. This simple technique of Social Security Number validation is highly recommended for Customer Identification Programs, particularly at the time of account inception. On a per-transaction basis, Social Security Number validation is remarkably cheap because it requires nothing more than a computer-driven table-lookup. Under several reasonable models of fraudulent use, it is also remarkably effective when combined with data for age or birth-year and an accurate SSN issuance history.

Methods and Materials

We used Monte Carlo simulations in which Social Security Numbers, ages, and birth-year were randomly generated under three different domains or protocols. Each of these protocols might be reasonably assumed to mimic a process by which such numbers, ages, and dates are fraudulently constructed. These "fraudulent" data were then validated against a table of all area and group numbers known to have been issued and their dates of issuance. The combination of SSN and birth-year was counted as "detected" if the number is known never to have been issued. Detections were also counted for numbers known to have been validly issued, but which were only issued before the claimed year of birth. This method takes advantage of the fact that, while Social

Security Numbers can be issued to persons at any age of life, these numbers can only be issued to persons actually alive on the dates of issuance.

The historical issuance data used in these simulations is based on proprietary information that extends to 1936, with annual data for issuance beginning in 1952. On a few occasions, these data have proven to be more reliable than those published by the Social Security Administration (Reference 1).

Under Protocol 1, one thousand area and group numbers were chosen at random, but without duplication, using all digits zero through nine in every position. In addition an age was randomly generated between 0 and 100, which also yields a year of birth. For Protocol 2, these ages were randomly generated in such a way as to closely match the age distribution of the estimated resident population of the United States on November 1, 2006 (Reference 2). In Protocol 3, randomly generated ages were chosen to closely match the age distribution (15 through 75) of the male population of the United States counted in the decennial census in prisons on April 1, 2000 (Reference 3). All of the simulations were performed in early 2008.

Results

Table 1 shows the detection rates achieved for the ten simulations under Protocol 1. The average detection rate of 70% is markedly better than what could be achieved without the histories of the dates of issuance of each area and group number – about 45%. The lower detection rate is based on the fact that 44,586 valid combinations are known to have been issued through January 2, 2008 (Reference 4).

Table 1. Detection Rates for Random Ages

Simulation	Simulation Size	Detection Rate
1	1000	0.72
2	1000	0.69
3	1000	0.68
4	1000	0.69
5	1000	0.72
6	1000	0.72

Simulation	Simulation Size	Detection Rate
7	1000	0.68
8	1000	0.69
9	1000	0.70
10	1000	0.70

The detection rates of the ten simulations performed according to Protocol 2 are shown in Table 2. The average detection rate in these simulations is 75%. All of these simulations resulted in a detection rate equal to or better than would have been the case if the ages 0 through 100 had been chosen uniformly at random, rather than (randomly) to match the population age structure. This is because the actual, resident U.S. population does not have the same numbers of persons in every single age and because the average age of the population is 36.

Table 2. Detection Rates for Ages Matching General Population

Simulation	Simulation Size	Detection Rate
1	1000	0.75
2	1000	0.77
3	1000	0.75
4	1000	0.76
5	1000	0.75
6	1000	0.76
7	1000	0.76
8	1000	0.72
9	1000	0.73
10	1000	0.75

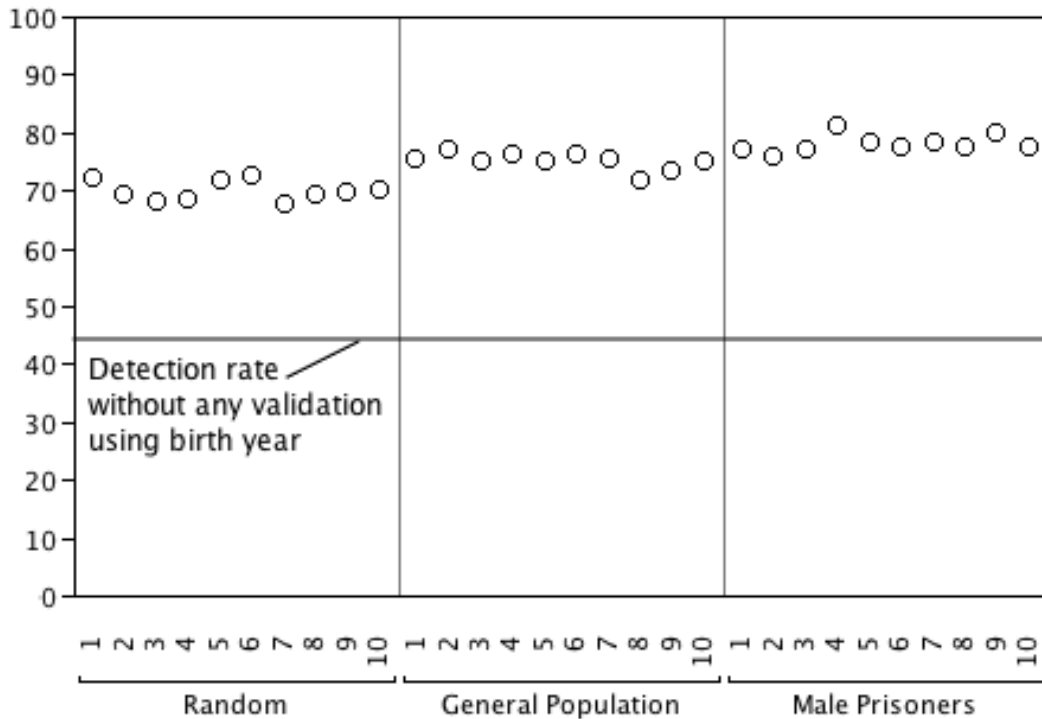
Table 3 presents the results of the ten simulations run under Protocol 3. In this case, the youngest population structure among the three simulation domains (estimated average age, 33) results in the best rates of detection. The average detection rate for Protocol 3 is 78%.

Table 3. Detection Rates for Ages Matching the Male Prison Population

Simulation	Simulation Size	Detection Rate
1	1000	0.77
2	1000	0.76
3	1000	0.77
4	1000	0.81
5	1000	0.78
6	1000	0.78
7	1000	0.78
8	1000	0.77
9	1000	0.80
10	1000	0.77

Figure 1 summarizes the detection rates from each simulation under each of the three protocols. The reference line in the figure shows the expected detection rate that would be achieved by simple number validation, without using information for SSN issuance histories. As the figure shows, the issuance histories greatly improve the detection rate.

Figure 1. Percentage of “Fraudulent” SSNs Detected



Discussion

It is unlikely that very many perpetrators of fraud with Social Security Numbers are using random number generators to construct the numbers. According to the Social Security Administration, the most abused SSN in history (078-05-1120) was claimed inappropriately by more than 40,000 persons using a number that was not randomly chosen (Reference 5). Certainly some perpetrators depend on the use of SSNs they know to have actually been issued and that are still valid. Detection methods other than simple validation are necessary to find these.

Even so, we know from the results of our own use of Social Security Number validation that many perpetrators do simply “make up” the numbers they use for fraudulent purposes. While some of these fraudulent numbers have a lower detection probability than the random numbers generated for this study, we know from our twenty years of practice in this field that many perpetrators do not succeed nearly as well and are easy to catch.

Conclusion

Social Security Number validation based on known dates of issuance is both inexpensive and effective. Because of this cost-benefit ratio, SSN validation should be routinely used for Customer Identification Programs. Please feel free to email us for further details, addressed to:

inquiry@quality-control.us

Subject: SSN Validation

References

- 1) Alice K. Whitfield, "HighGroup Listing of SSNs," July 18, 2003, *The Risks Digest*, Volume 22: Issue 81, Uniform Resource Locator: <http://catless.ncl.ac.uk/risks/22.81.html#subj13>, accessed January 18, 2008.
- 2) U. S. Census Bureau, Monthly Postcensal Resident Population for the 2000s, Uniform Resource Locator: <http://www.census.gov/popest/national/asrh/files/NC-EST2006-ALLDATA-R-File22.dat>, accessed January 14, 2008.
- 3) U. S. Census Bureau, Population in Group Quarters by Type, Sex and Age, for the United States: 1990 and 2000 (PHC-T-26), Uniform Resource Locator: <http://www.census.gov/population/cen2000/phc-t26/tab01.pdf>, accessed January 14, 2008.
- 4) Social Security Administration, Social Security Number Verification Service (SSNVS), Uniform Resource Locator: <http://www.ssa.gov/employer/highgroup.txt>, accessed January 8, 2008.
- 5) Social Security Administration, Social Security Cards Issued by Woolworth, Uniform Resource Locator: <http://www.socialsecurity.gov/history/ssn/misused.html>, accessed January 14, 2008.