
Social Security Number Validation

May 6, 2011

Copyright © 2011 by [Quality Control Systems Corp.](#) All rights reserved.

The Projected Effectiveness of Social Security Number Validation through 2026



– White Paper –

Social Security Number Validation

May 6, 2011

Copyright © 2011 by [Quality Control Systems Corp.](#) All rights reserved.

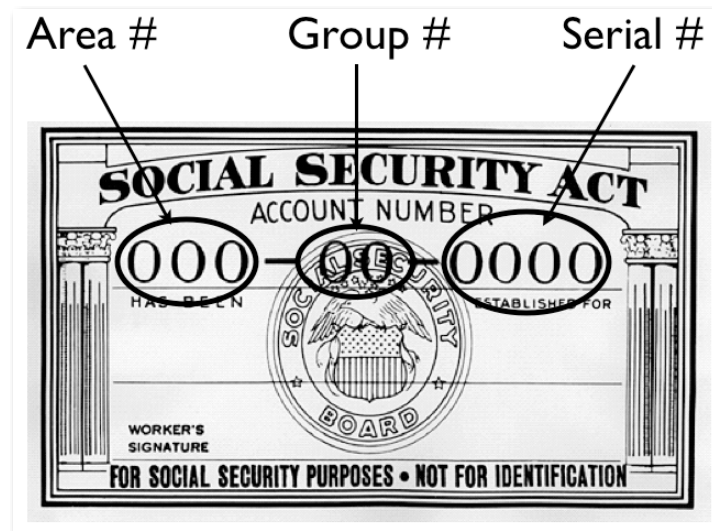
Summary

Social Security Number validation is an inexpensive, best practice solution to identify potentially fraudulent or erroneous Social Security Numbers (SSNs). Our demographic projections show that SSN validation, based on known dates of issuance, will remain a remarkably effective method to detect mistaken and fictitious SSNs for the foreseeable future. When birth dates and usage histories are utilized in SSN validation of area and group numbers, very high levels of detection will continue through 2026 – well after the implementation of randomized assignment by the Social Security Administration. Because of its cost-benefit ratio, SSN validation should be routinely used when SSNs are used in systems involving personal identification.

Introduction

The domain space of Social Security Numbers covering area and group number combinations (000-00 through 999-99), years of birth of living persons who have ever had an SSN (1896-2011), as well as years of SSN issuance (1936-2011) is very large: 874,000,000 potential combinations. (See Figure 1.) However, historical issuance data for each area and group number combination are publicly available that can be used to dramatically narrow the valid possibilities, especially when claimed birth dates and/or dates of issuance are known or can be approximated (see [Fraud Detection and Identity Validation with SecureID™](#)).

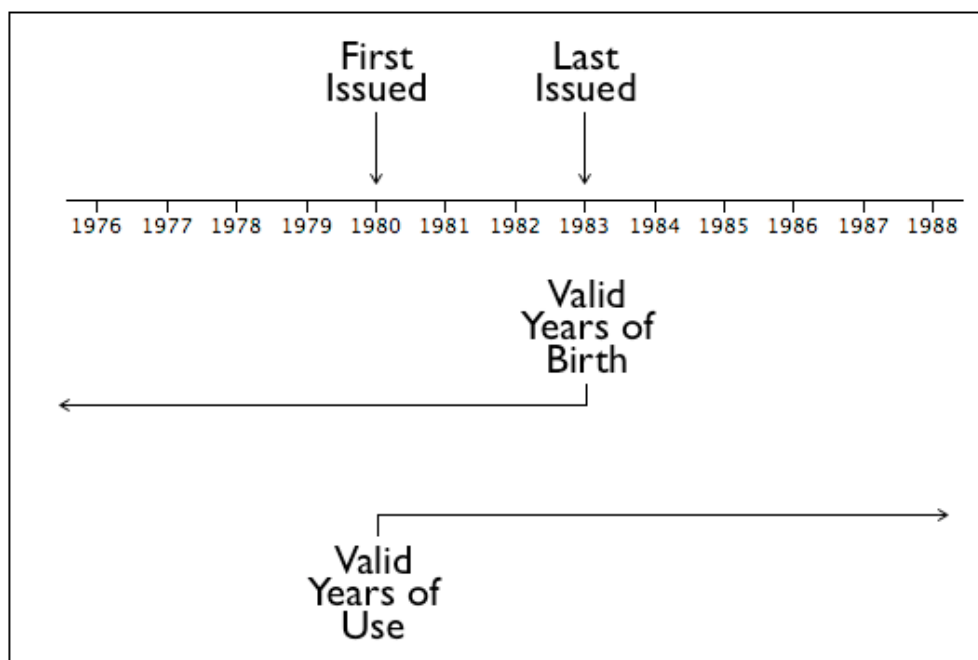
Figure 1. The Three Parts of a Social Security Number (SSN)



Fraudulent and erroneous transactions using SSNs are easy to detect for claimed SSNs that: 1) have never been issued by the Social Security Administration (SSA), or 2) were issued by the SSA only before the user's claimed date of birth, or 3) were issued only after the claimed date of issuance or earliest known use the number, or 4) are known to have been issued to persons who are deceased. SSNs failing these same four tests also indicate simple data entry errors and misremembered numbers. Such problems can be easily and inexpensively corrected before they result in readily preventable, expensive, adverse consequences. This is especially true when validation is performed at account inception.

An example of the relationships between years of issuance, birth dates, and SSN usage are shown in Figure 2 for all persons using the specific area and group number combination 037-50. Similar relationships exist for every combination of area and group number that have been issued over time by the SSA. Effective SSN validation depends on a database that accurately reflects this issuance history for all possible combinations, extending to the beginning of the Social Security program.

Figure 2. The Relationships between Years of Issuance, Birth Dates, and Usage for All Persons Using Area and Group Number, 037-50.



The new policy of randomized assignment of SSNs by the Social Security Administration introduces some important changes in the decades-long practice of SSN validation. This policy will make a large number of new combinations of area and group number simultaneously available for assignment in 2011. Fortunately, a simple counter-measure based on the “usage history” of a claimed SSN can minimize the negative impact that the new policy could have on established fraud detection and data quality control programs. This usage history can be as simple as the answer to the question, “When did you obtain your Social Security Number?”

When testing SSNs in existing records systems, the earliest date on which the SSN appears in the records can be used (for example, the date of account inception). Our analysis shows that using information about claimed SSN issuance, as well as the claimed date of birth, significantly enhances the effectiveness of validation for the foreseeable future.

Methods and Materials

We used Monte Carlo simulations to analyze the effectiveness of SSN validation to identify problematic SSNs in a population whose age structure matches the projected resident population of the United States through the year 2026 (Protocol 1). We also used Monte Carlo simulations to identify such SSNs for the cohort of persons born in 1987 and residing (or expected to reside) in the United States in each successive calendar year from 1988 through 2026 (Protocol 2).

For Protocol 1, we chose random ages and random dates of issuance matched to each of 100,000 area and group number combinations (000-00 through 999-99). The random ages were generated such that they matched the projected age structure of the resident United States population in the years 2011, 2016, 2021, and 2026 [Reference 1]. Dates of claimed issuance were also randomly chosen, but in a uniform distribution starting with the date of birth (or the beginning date of the Social Security program, 1936, if that is later than the year of birth) through the date applicable to the projection.

For Protocol 2, we used the count of live births in the United States in 1987 [Reference 2], historical estimates of the population residing in the U. S. on January 1st for the years 1988 through 2009 [References 3, 4, 5] and population projections for the years 2010 through 2026 [Reference 1] to define the cohort. We subtracted deaths from this cohort in every successive year derived from life tables [1987-1998, Reference 6] or actual counts [1999-2007, Reference 7] of deaths in each calendar year by age in order to derive estimates by year for net migration. After 2007, we estimated deaths for all ages 21 and above using the latest available life table by single year of age, 2006 [Reference 8].

According to our comparisons of the size of this cohort in each calendar year with data reported by the Social Security Administration in their 1993 report, “Social Security Numbers Issued: A 20-Year Review [Reference 9],” the issuance of Social Security Numbers to infants and children who are residents of the United States is practically universal by age 5. We apportioned SSN issuance by calendar year and single year of age to the Protocol 2 cohort through age 4 based on these data. We lacked appropriate data to estimate SSN issuance for persons aged 5 through 15 and, therefore, ascribed no SSNs issued for persons in these ages. For persons in the cohort aged 16 and above, we estimated SSNs issued by age by applying the applicable labor force participation rate based on age and calendar year to the cohort without assigned SSNs to that date [Reference 10].

The random generation of birth dates and SSN issuance dates matched to SSNs reasonably characterizes a process by which fraudulent or erroneous SSNs may be created or recorded. The first protocol was chosen to demonstrate the effectiveness of validation based on birth dates matching the population age structure which changes over time, but with issuance dates that were chosen uniformly at random. The second protocol is especially well suited to the detection of simple errors in the reporting or recording of SSNs in an actual birth cohort, using dates of SSN issuance that are representative of that cohort.

The randomly generated birth and issuance dates were used to validate all 100,000 combinations of area and group numbers based on a table linking these combinations with their known issuance status and dates of issuance. The area and group number combination was counted as “detected” if the number has never 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 or after the claimed date of usage. This validation 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 (see Figure 2). Following the introduction of randomized assignment of SSNs, SSNs from the assignable pool can only have been issued to persons actually alive on the potential first date of issuance in 2011. In addition to validation using birth dates, issuance histories are extremely valuable. This is because, for a particular person, an SSN can only be claimed after the known date that corresponding SSNs were actually first issued (or were potentially issued following the introduction of randomization) and after the claimed date of birth (again, see Figure 2).

The historical issuance data used for 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 – see Reference 11). Beginning in 2011, randomized assignment causes all group and area number combinations not previously assigned and not restricted from use by the SSA to be considered in these simulations as potentially in legitimate use in 2011. In order to prepare these simulations in advance of the introduction of randomized assignment in June, 2011, we have treated the published “High Group List” for March, 2011 [see Reference 12] as though it were the last such list that will have been issued before the beginning of randomized assignment.

Results

Through March, 2011, 46,357 of 100,000 5-digit area/group combinations (000-00 through 999-99) have actually been issued by the SSA. Our documentation is ambiguous regarding the issuance of an addition 198 combinations for area numbers 580 and 586. There were 53,445 area/group combinations that had never been issued. Without any information about birth dates or claimed issuance

histories, randomly created SSNs before the advent of randomized assignment according to Protocol 1 would be detected on the basis of this information about 53% of the time. Following the implementation of random assignment of new SSNs, the detection rate plunges to about 11% if the SSN is not supplemented by additional information regarding a claimed date of birth or date of SSN issuance.

In contrast, Figure 3 shows the results of adding information about the year of birth and the year of SSN issuance on the detection rates of 100,000 false SSNs following the introduction of randomized assignment in 2011 as well as three projected calendar years using the methodology discussed above under Protocol 1. The detection rates for all four years – ranging from 66-84% – compare very favorably to the 11% rate that would be achieved without the supplemental information for age and claimed issuance.

Figure 3. Percentage of 100,000 Fictitious SSNs Detected by Projected Calendar Year After the Introduction of Randomized Assignment of SSNs – Protocol 1

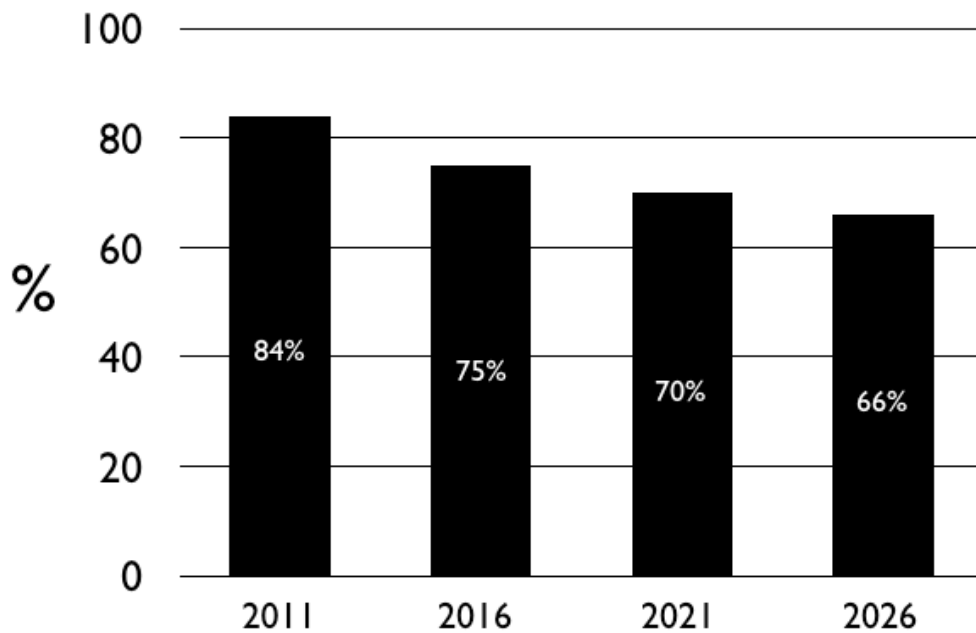
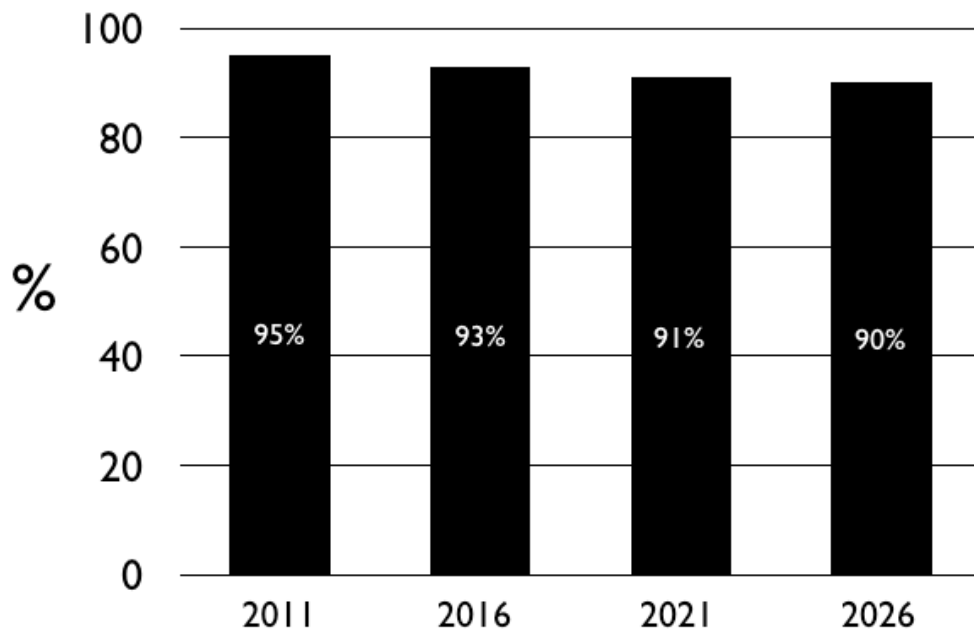


Figure 4 shows how information about the year of SSN issuance for a cohort born in 1987 and residing in the United States in successive years contributes to the detection rates from validation. The detection rates shown apply to the periods after the introduction of randomized assignment in 2011 as well as three projected calendar years using the methodology discussed above under Protocol 2. The detection rates range from 90-95%. These detection rates are uniformly higher than the comparable rates achieved in the simulations based on Protocol 1, partly because they were randomly chosen in such a way as to match the hypothetical distribution of issuance by age that is based on historical data.

Figure 4. Percentage of 100,000 Fictitious SSNs Detected by Projected Calendar Year After the Introduction of Randomized Assignment of SSNs – Protocol 2



Because the simulations are so large, all two-way comparisons of the percentages described in this section are statistically significant based on Fisher's 2-tailed exact test well below the traditional level of 0.05.

Discussion

Monte Carlo simulations are useful models to study how the effectiveness of SSN validation can change over time. For this purpose, random number generation provides a standard yardstick, even though it is unlikely that very many perpetrators of fraud are using such methods to base false SSNs, or birth dates, or the dates on which to claim SSNs were issued. Certainly some perpetrators use only SSNs they know to have actually been issued, consistent with claimed birth and issuance dates, 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. Just as importantly, detecting fraudulent or erroneous SSNs through validation frequently provides crucial leads for the discovery of larger, undetected patterns of fraud or error.

Conclusion

Enterprises that validate Social Security Numbers will find that identity fraud and error detection rates are markedly enhanced when SSN usage and birth dates are applied in SSN validation. Even in the wake of the Social Security Administration's implementation of randomized assignment, our analyses show very high detection rates are achievable through 2026.

Additional, related information is available on our website through these links:

[Fraud Detection and Identity Validation with SecureID™](#) – “The best place to deploy your anti-fraud defense is to catch the offender at the door...”

[The American Financial System’s Identity Problem](#) – “The United States relies on the Social Security Number as its de facto national identification number...”

[Randomized Assignment of Social Security Numbers](#) – “The randomized assignment of Social Security Numbers by the Social Security Administration beginning in June, 2011 introduces some important changes in the decades-long practice of SSN validation by many businesses and organizations.”

[“Exaggerated” Deaths in SSA’s Death Master File](#) – “... Since the 1930s, deaths of social security number holders have been reported to the Social Security Administration (SSA).”

Contact Information

For more information about SecureID™, call Quality Control System’s Sales Director, Betsy Love at True North International, 917-939-9409. Betsy’s email address is:

betsyl@truenorthintl.com

References

- 1) U. S. Census Bureau, National Population Projections, Released 2008 (Based on Census 2000), Uniform Resource Locator:
<http://www.census.gov/population/www/projections/files/nation/download/NP2008_D1.xls>, accessed April 16, 2011.
- 2) U. S. Census Bureau, Uniform Resource Locator:
<<http://www.census.gov/statab/USA98/dd-vs.txt>>, accessed April 20, 2011.
- 3) U. S. Census Bureau, National Estimates, Quarterly Population Estimates, 1980 to 1990, Uniform Resource Locators:
<<http://www.census.gov/popest/archives/1980s/e8788rqi.zip>>,
<<http://www.census.gov/popest/archives/1980s/e8889rqi.zip>>,
<<http://www.census.gov/popest/archives/1980s/e8990rqi.zip>>, accessed April 16, 2011.
- 4) U. S. Census Bureau, National Intercensal Estimates (1990-2000): All Months, Uniform Resource Locator:
<<http://www.census.gov/popest/archives/EST90INTERCENSAL/US-EST90INT-07/US-EST90INT-07.csv>>, accessed April 16, 2011.
- 5) U. S. Census Bureau, Monthly Postcensal Resident Population for the 2000s, Uniform Resource Locators:
<<http://www.census.gov/popest/national/asrh/files/NC-EST2008-ALLD-ATA-R-File03.csv>>,
<<http://www.census.gov/popest/national/asrh/files/NC-EST2008-ALLD-ATA-R-File05.csv>>,
<<http://www.census.gov/popest/national/asrh/files/NC-EST2008-ALLD>>

ATA-R-File07.csv>, <<http://www.census.gov/popest/national/asrh/files/NC-EST2008-ALLD-ATA-R-File07.csv>>, <<http://www.census.gov/popest/national/asrh/files/NC-EST2008-ALLD-ATA-R-File09.csv>>, <<http://www.census.gov/popest/national/asrh/files/NC-EST2008-ALLD-ATA-R-File11.csv>>, <<http://www.census.gov/popest/national/asrh/files/NC-EST2008-ALLD-ATA-R-File13.csv>>, <<http://www.census.gov/popest/national/asrh/files/NC-EST2008-ALLD-ATA-R-File15.csv>>, <<http://www.census.gov/popest/national/asrh/files/NC-EST2008-ALLD-ATA-R-File17.csv>>, <<http://www.census.gov/popest/national/asrh/files/NC-EST2008-ALLD-ATA-R-File19.csv>>, accessed April 16, 2011.

6) Centers for Disease Control and Prevention, Publications and Information Products, Life Tables, Uniform Resource Locators:

<http://www.cdc.gov/nchs/data/lifetables/life87_2acc.pdf>, <http://www.cdc.gov/nchs/data/lifetables/life88_2acc.pdf>, <http://www.cdc.gov/nchs/data/lifetables/life89_2acc.pdf>, <http://www.cdc.gov/nchs/data/lifetables/life90_2acc.pdf>, <http://www.cdc.gov/nchs/data/lifetables/life91_2acc.pdf>, <http://www.cdc.gov/nchs/data/lifetables/life92_2.pdf>, <http://www.cdc.gov/nchs/data/lifetables/life93_2.pdf>, <http://www.cdc.gov/nchs/data/lifetables/life94_2.pdf>, <http://www.cdc.gov/nchs/data/lifetables/life95_2.pdf>, <http://www.cdc.gov/nchs/data/nvsr/nvsr47/nvs47_13.pdf>, <http://www.cdc.gov/nchs/data/nvsr/nvsr47/nvs47_28.pdf>.

<http://www.cdc.gov/nchs/data/nvsr/nvsr48/nvs48_18.pdf>, accessed April 20, 2011.

- 7) Centers for Disease Control and Prevention, National Vital Statistics System, GMWK310, Deaths by Single Years of Age, Race, and Sex: United States, 1999-2007, Uniform Resource Locators:

<http://www.cdc.gov/nchs/data/statab/VS00199_TABLE310.pdf>,
<<http://www.cdc.gov/nchs/data/statab/wktbl310.pdf>>,
<http://www.cdc.gov/nchs/data/statab/mortfinal2001_work310.pdf>,
<http://www.cdc.gov/nchs/data/dvs/MortFinal2002_Work310.pdf>,
<http://www.cdc.gov/nchs/data/dvs/Mortfinal2003_worktable310.pdf>,
<http://www.cdc.gov/nchs/data/dvs/mortfinal2004_worktbl310.pdf>,
<http://www.cdc.gov/nchs/data/dvs/mortfinal2005_worktable_310.pdf>,
<http://www.cdc.gov/nchs/data/dvs/MortFinal2006_Worktable310.pdf>,
<http://www.cdc.gov/nchs/data/dvs/MortFinal2007_Worktable310.pdf>,
accessed April 20, 2011.

- 8) Centers for Disease Control and Prevention, Publications and Information Products, Life Tables, Uniform Resource Locator:

<http://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58_21.pdf>, accessed April 20, 2011.

- 9) Social Security Administration, "Social Security Numbers Issued: A 20-Year Review," Social Security Bulletin, Vol 56, No. 1, Spring 1993, p. 86, Uniform Resource Locator:

<<http://www.ssa.gov/policy/docs/ssb/v56n1/v56n1p83.pdf>>, accessed April 26, 2011.

- 10) U. S. Department of Labor, Bureau of Labor Statistics, Employment Projections, Uniform Resource Locator:
<http://www.bls.gov/emp/ep_table_303.pdf>, accessed April 16, 2011.
- 11) 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 April 28, 2011.
- 12) Social Security Administration, High Group List and Other Ways to Determine if an SSN is Valid, Uniform Resource Locator:
<<http://www.ssa.gov/employer/ssnvhighgroup.htm>>, accessed April 28, 2011.