Published in final edited form as:
J Occup Environ Med. 2016 June ; 58(6): 631–635. doi:10.1097/JOM.0000000000000740.

A Validation Method to Determine Missing Years of Birth in a Cohort Study of Shipyard Workers Using Social Security Number

Citadel J. Cabasag, M.S.,
University of California, Irvine

Argyrios Ziogas, Ph.D.,
University of California, Irvine

Merna Shehata, B.S., and
University of California, Irvine

Hoda Anton-Culver, Ph.D.
University of California, Irvine

Occupational cohorts are especially useful for etiologic studies when employment records, union records, and work-related exposure records are available. The objective of this study is to examine the relationship between occupational exposures and disease outcomes associated with employment in a shipyard cohort. Employment records used to construct historical occupational cohorts are not always compiled for the purpose of utilizing these data in epidemiological studies. A study to assess the effect of completeness of data may thus be necessary. Methods used to address the problems associated with incomplete data in historical occupational cohorts includes linkage with public records and databases, such as motor vehicle records, phone records, cancer registries, and state death files.[1] Methodological issues to compare the effect of the missing observations to the non-missing observations of a given variable have been previously studied including methods to impute missing data.[2–4] In this paper, we present findings on a large historical occupational cohort of naval shipyard workers where we imputed dates of birth and evaluated a method to predict missing birth year and age variable based on Social Security Numbers (SSN).[5]

Starting in 1936 workers in most occupations were required to have SSNs when entering the workforce.[6] The method developed by Block, et al. (1983) predicts the year of birth of workers using information from these SSNs. Multiple studies have applied this prediction method to estimate missing age.[7–11] The method has also been utilized to help impute nativity and year of immigration to the United States when this information is unknown.[12–33] Block et al. assessed the application of the method using the Florida phosphate workers employed between 1950 and 1979. Johnson et al. (1986) also performed a validation study with 1,000 meat-cutter union members prior to applying Block's method to members with missing age in their cohort.[10, 11] Although Block's method has been used

Correspondence: Hoda Anton-Culver, Ph.D., Tel: (949) 824-7401, Fax: (949) 824-7443, hantoncu@uci.edu.

Conflict of Interest:
All authors have no conflict of interest.

in several study populations, a validation study using a large retrospective cohort has not yet been performed.

We applied the method developed by Block et al. to predict the years of birth of workers in the Long Beach Naval Shipyard (LBNS) cohort to evaluate the feasibility and external validity of this prediction method. To our knowledge, this is the first study to examine the generalizability of Block's method for predicting the years of birth using a large occupational retrospective cohort. Following our examination of the external validity (validation study), we modified Block's method to improve the accuracy of the predicted data.

METHODS

Study population

Our study included a sample of a large LBNS cohort, which was previously studied by Anton-Culver et al.[34, 35] The LBNS employed over 41,000 workers at the shipyard between 1942 and 1997. In the current study, all available employment rosters were utilized, which contained 13,924 shipyard workers employed between 1978 and 1985. The 13,924 were included in the current study because of the availability of their data in the shipyard records and we were granted access to their specific jobs and shops through the LBNS employment rosters. We will refer to this group of LBNS workers as the LBNS-Validation Cohort (LBNS-VC). The data of the LBNS-VC included last name, first name/initial, SSN, shop number and job title, as well as other occupational data. The information on the cohort was derived from both the LBNS employment roster and the LBNS employment cards. Linkage to both the California cancer registry and death statistics master files up to 2013 were carried out on the LBNS-VC to obtain additional information such as year of cancer diagnosis, type of cancer, year of death, and cause of death.

Of the 13,924 workers in the LBNS-VC, 6,980 (50.1%) had known years of birth through the original shipyard personnel data, and from the results of the linkage to the California cancer registry and death files. Meanwhile, 6,944 workers did not have data on age or year of birth but had data on their SSN. In order to study the health effects associated with occupational exposures at the shipyard, it was necessary to obtain, as complete as possible, either the age or the date of birth of the workers. In LBNS workers with missing age, prior to applying Block's prediction method, we used the data from the LBNS-VC with known years of birth to validate Block's method. Block's method was modified to improve the accuracy before the method was used to impute the birth years of LBNS workers with missing age and unknown year of birth.

Block's prediction method to impute years of birth

The method described by Block et al. (1983) uses the SSN to predict the year of birth, which is structured into three different segments: geographic area number (first three-digits), group number (middle two-digits), and serial number (last four-digits).[5, 36, 37] For each worker, the SSN year of issue was determined using information derived from the Social Security Administration. Block et al. combined data from the Social Security Administration and extrapolated data to construct a matrix for SSN years of issue from 1937 to 1978. The matrix contains the geographic area number, the group number, and the corresponding SSN years-of-issue. A simple calculation of the difference between the SSN year-of-issue and the age when SSN was issued produced the predicted year of birth.

Validation using the 6,980 workers with known years of birth from the LBNS-VC

We used the 6,980 group of workers from the LBNS-VC with known years of birth to evaluate the external validity of Block's method. Of the 6,980 workers we excluded 248 who were either >64 years old during the first year of the study or were hired at the age of >64 years between 1978 and 1985. The data on the remaining 6,732 workers were used for the validation analysis.

We applied the method and used the SSN year-of-issue matrix (included in the publication by Block et al. (1983)) to the 6,732 workers, and imputed the SSN year-of-issue for 4,909 workers. Workers whose SSN year-of-issue were not located using the matrix (n=1,823) were not included in the subsequent calculations for the prediction of the year of birth. The true year of birth was known for the 4,909 workers in this validation study. We calculated the predicted year of birth using the same distribution of the median age at SSN used by Block et al. (1983) (Table 1) and the estimated SSN year-of-issue for each of the LBNS worker. We also performed additional analysis using the distribution of the median age of SSN issuance based on the LBNS-VC population (Table 1). Pearson's correlation coefficient (percent agreement) was calculated to assess the linear relationship between the predicted years of birth and the observed years of birth. A two-tailed $P < 0.05$ is considered as a statistically significant correlation between the predicted and the observed birth years.

The predicted years of birth of the 4,909 workers calculated using Block's method with the median age distribution originally used by Block et al. had an 88.8% agreement ($P < 0.0001$, R^2=0.789) with the workers' observed years of birth, while the percent agreement using the median age distribution from the LBNS-VC was 89.2% ($P < 0.0001$, R^2=0.796) (Table 2). In the original article, Block et al. reported an agreement of 91.0% ($P < 0.0001$, $R^2 = 0.820$).

Modification of Block's prediction method

In order to improve Block's method, we performed modifications on the original method. For our analysis, we used the median age of SSN distribution derived from the LBNS-VC. For the first modification (Modification I), we separated the 4,909 workers into three different groups based on each worker's predicted year of birth: <1920, 1920–1930, >1930. An additional four years were added to the birth year of workers with predicted years of birth <1920, while two years were added to the group with predicted years of birth from 1920 to 1930. These two values are the median difference (years) between the predicted and the observed years of birth for the corresponding groups of the LBNS-VC. No adjustments were added to the >1930 group.

In the second modification (Modification II), the sample was not separated instead we excluded 208 workers who had SSN issued in U.S. territories (SSNs with area number 580 – 586), such as Guam, Virgin Islands, Philippines, American Samoa and Puerto Rico (Table 3). For this modified version of Block's method, the predicted year of birth was in 93.1%

agreement ($P < 0.0001$, $R^2 = 0.867$) with the observed year of birth (Table 2). To assess whether excluding workers with SSN in U.S. territories will improve the correlations in Modification I, we combined Modification I and Modification II. Both modifications combined produced a very similar result when using Modification II alone (Pearson's r = 93.2%, $P < 0.0001$, $R^2 = 0.869$). Overall, Modification II produced one of the highest percent agreements between the predicted and the observed years of birth.

Applying the modified Block's method to the 6,944 workers with missing years of birth

In the confirmation study for Modification II, we used the LBNS-Alumni sub-cohort which contained the 6,944 shipyard workers who initially had unknown years of birth and who were later followed-up using the AlumniFinder (Accudata Integrated Marketing, Fort Myers, Florida) (Figure 1). Of the 6,944 workers, 1,850 were not found using the AlumniFinder, while 5,094 workers were identified and information regarding their age and year of birth were subsequently obtained. These 5,094 workers were used to confirm our results from the validation study of Block's method with Modification II in the LBNS-VC.

Workers who were >64 years old during the first year of the study or workers hired at the age of >64 years between 1978 and 1985 were excluded from the sample (n=5,052). Out of these 5,052 workers, there were 1,725 whose SSN year-of-issue could not be determined using the matrix produced by Block et al. As specified from Modification II, workers in the sample were separated according to the geographic location where the SSN was issued. This yielded 3,094 workers who had U.S. state-issued SSNs and 233 workers who had U.S. territories-issued SSNs (Table 3). To assess Block's method with Modification II, the predicted years of birth were calculated in workers who had both estimated SSN years-of-issue and SSNs that were issued from U.S. states (n=3,094). All data analyses in this study were generated using SAS software, version 9.3 (SAS Institute, Inc., Cary, North Carolina).

Data Safeguards

This study was approved by the Institutional Review Board of the University of California, Irvine (HS # 2013-9428). The data are stored in protected file servers managed by the Security Team of the University of California, Irvine Health Affairs Information Services. All data analyses were performed using secure and protected workstations within the Department of Epidemiology facility and managed by the University of California, Irvine Health Affairs Information Systems. Access to the data is restricted to study personnel only.

RESULTS

Validation analysis of Block's method with modifications using LBNS-VC

Results from the modifications applied to Block's method are shown in Table 2. The percent agreement between the predicted years of birth and the workers' observed years of birth in the LBNS-VC for Modification I was similar to the result using Block's method without modifications (Pearson's r = 89.1%, $P < 0.0001$, $R^2 = 0.796$ vs. Pearson's r = 89.2%, $P < 0.0001$, $R^2 = 0.793$). The percent agreement in Modification II (Pearson's r = 93.1%, $P < 0.0001$, $R^2 = 0.867$) was higher than the percent agreement using the unmodified Block's method. As consistently observed in both plots (Figure 2A & 2B), the prediction method is highly variable in workers born prior to 1940, whereas the variability around the prediction line becomes smaller in workers born after 1940. We also observed a cluster of outliers above the prediction line (Figure 2A), which was not present in the plot for Modification II (Figure 2B). In addition, the relationship between the predicted and observed years of birth contained empty spaces in the x-axis during the early 1920s and 1930s. This result indicates that the method was unable to predict the birth years of individuals born in those years.[5] The cross tabulation of the SSN year-of-issue and the median year of birth is not a smooth continuous line, which caused the observed pattern.[5]

We examined the accuracy of Block's method with Modification II by analyzing the difference in the number of years between the predicted and observed years of birth grouped by the years of SSN issuance based on the type of Social Security Administration documentation/form utilized for the SSN year-of-issue (Table 4). The method accurately predicted the birth years for 916 (19.0%) workers. More than half of the workers (n=2,415) had 1–2 years difference between the predicted and observed years of birth. The majority of the workers in both groups had their SSN issued after 1950. An opposite trend from the previous groups was observed in workers with 3–5 years difference between the predicted and observed years of birth. The majority of the workers in this group had SSN issued prior to 1948. Lastly, there were 460 (9.80%) workers out of 4,701 with >5 years difference between the predicted and observed years of birth.

Analysis to confirm modification of Block's method in a second subset of LBNS workers

In the LBNS-Alumni sub-cohort, the predicted years of birth using Block's method with Modification II was in 92.6% agreement ($P < 0.0001$, $R^2 = 0.857$) with the workers' observed years of birth (Table 2). Block's method with the modification had a higher percent agreement compared to the unmodified method (Pearson's r=89.1%, $P < 0.0001$, $R^2 = 0.794$).

DISCUSSIONS

Our results show that the age prediction method we describe in this paper yielded superior results than the original method by Block et al. (1983). Our data show that the correlation between the predicted and observed years of birth increased when workers with U.S. territory-issued SSN are removed from the analysis. Block et al. showed that Railroad Board SSNs were highly inaccurate in predicting the year of births. However, they did not examine the accuracy of U.S. territory-issued SSNs. In our validation study, 4.20% of the workers in the LBNS-VC have a U.S. territory-issued SSNs. The majority of the U.S. territory-issued SSNs produced incorrect predicted years of birth values, and more than half had predicted years of birth with >5-years difference from the true value.

Overall, the accuracy in estimating the years of birth in our validation study was worse for older workers relative to younger workers. One source of error originates from the method, which uses the matrix (see Method section above) to determine the SSN year-of-issue for each worker. Another source of error is due to the inaccurate age estimates when the SSN was issued.[5] The age at which individuals obtained their SSN varied widely in the first few years of the Social Security Administration.[37] After the majority of the U.S. workforce obtained their SSNs, the typical age at which SSN is obtained became less variable.[6, 38] Although the distribution of SSN age-of-issue in the population is generally more predictable several years after the introduction of the SSN program, discrepancies existed in certain subgroups, such as adult immigrants entering the workforce.

The matrix described by Block et al. did not include SSNs obtained after 1978 and therefore the information in the matrix may also not accurately characterize younger workers who most likely have obtained their SSN in later years of the program. Finally, in 1991 the majority of U.S. states allowed the parent of a newborn to request an SSN as part of the state's birth registration process.[37] In addition, in June 2011 the Social Security Administration changed the method of assigning SSN numbers reffered to as "randomization."[39] Thus, the ability of Block's method to accurately predict the year of birth is limited to the SSN year-of-issuance.

The large population used in our validation study represents multiple groups of workers in variety of occupations at the shipyard. Although the majority of the LBNS employees were blue-collar workers, a portion of the cohort held professional degrees, which were mostly in engineering and architecture. Our study addresses the concern of the applicability of the prediction method to other populations including professional groups. We also performed subsequent analysis using "AlumniFinder" methodology to verify the results from the LBNS-VC. The result from this analysis is consistent with the initial findings (Table 2).

In conclusion, this study and other similar prediction methods are valuable in epidemiological studies where important data elements are missing such as age while SSN is more complete. It can be useful to researchers in the calculation of age-specific incidence and mortality rates. The method presented here was applied to occupational cohorts who are primarily white male workers. Future applications in non-occupational populations as well as in predominantly female populations are warranted to determine if this method is equally applicable in other populations.

Acknowledgments

This publication was supported by the Pilot Project Research Training Program of the Southern California NIOSH Education and Research Center, Grant Agreement Number T42 OH0084 from the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of CDC.

Source of Funding:

Citadel Jungco Cabasag is currently funded through the Ruth L. Kirschstein National Research Service Award (F31) fellowship from the National Institute of Health.

References

1. Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. Int J Epidemiol. 2002; 31:1246–52. [PubMed: 12540730]
2. Williams, R. Missing Data Part 1: Overview, Traditional Methods 2015. Available at: https://www3.nd.edu/~rwilliam/stats2/l12.pdf. [Accessed June 17, 2015]
3. Meray N, Reitsma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. J Clin Epidemiol. 2007; 60:883–91. [PubMed: 17689804]

4. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009; 338:b2393. [PubMed: 19564179]
5. Block G, Matanoski GM, Seltser RS. A method for estimating year of birth using social security number. Am J Epidemiol. 1983; 118:377–95. [PubMed: 6613981]
6. Tacker HR. Notes and Brief Reports: Social Security Numbers Issued, 1937–71. Social Security Bulletin. 1972:30–32.
7. Lamm SH, Levine MS, Starr JA, Tirey SL. Analysis of excess lung cancer risk in short-term employees. Am J Epidemiol. 1988; 127:1202–9. [PubMed: 3369419]
8. Johnson ES, Shorter C, Rider B, Jiles R. Mortality from cancer and other diseases in poultry slaughtering/processing plants. Int J Epidemiol. 1997; 26:1142–50. [PubMed: 9447392]
9. McDonald JA, Potter NU. Re: "A method for estimating year of birth using Social Security number". Am J Epidemiol. 1991; 134:1475. [PubMed: 1776625]
10. Johnson ES, Fischman HR, Matanoski GM, Diamond E. Occurrence of cancer in women in the meat industry. Br J Ind Med. 1986; 43:597–604. [PubMed: 3756110]
11. Johnson ES, Fischman HR, Matanoski GM, Diamond E. Cancer mortality among white males in the meat industry. J Occup Med. 1986; 28:23–32. [PubMed: 3485192]
12. Gomez SL, Quach T, Horn-Ross PL, et al. Hidden breast cancer disparities in Asian women: disaggregating incidence rates by ethnicity and migrant status. Am J Public Health. 2010; 100(Suppl 1):S125–31. [PubMed: 20147696]
13. Lauderdale DS, Jacobsen SJ, Furner SE, Levy PS, Brody JA, Goldberg J. Hip Fracture Incidence among Elderly Asian-American Populations. Am J Epidemiol. 1997; 146:502–509. [PubMed: 9290511]
14. Shimizu H, Ross R, Bernstein L, Yatani R, Henderson B, Mack T. Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County. Br J Cancer. 1991; 63:963–966. [PubMed: 2069852]
15. Chang ET, Yang J, Alfaro-Velcamp T, So SK, Glaser SL, Gomez SL. Disparities in liver cancer incidence by nativity, acculturation, and socioeconomic status in California Hispanics and Asians. Cancer Epidemiol Biomarkers Prev. 2010; 19:3106–18. [PubMed: 20940276]
16. Keegan TH, John EM, Fish KM, Alfaro-Velcamp T, Clarke CA, Gomez SL. Breast cancer incidence patterns among California Hispanic women: differences by nativity and residence in an enclave. Cancer Epidemiol Biomarkers Prev. 2010; 19:1208–18. [PubMed: 20447917]
17. Mack TM, Walker A, Mack W, Bernstein L. Cancer in Hispanics in Los Angeles County. Natl Cancer Inst Monogr. 1985; 69:99–104. [PubMed: 3834353]
18. Gomez SL, Clarke CA, Shema SJ, Chang ET, Keegan TH, Glaser SL. Disparities in breast cancer survival among Asian women by ethnicity and immigrant status: a population-based study. Am J Public Health. 2010; 100:861–9. [PubMed: 20299648]
19. Telli ML, Chang ET, Kurian AW, et al. Asian ethnicity and breast cancer subtypes: a study from the California Cancer Registry. Breast Cancer Res Treat. 2011; 127:471–8. [PubMed: 20957431]
20. Mack TM, Floderus B. Malignant melanoma risk by nativity, place of residence at diagnosis, and age at migration. Cancer Causes and Control. 1991; 2:401–411. [PubMed: 1764565]
21. Shimizu H, Mack TM, Ross RK, Henderson BE. Cancer of the gastrointestinal tract among Japanese and white immigrants in Los Angeles County. J Natl Cancer Inst. 1987; 78:223–8. [PubMed: 3468285]
22. Lauderdale DS, Jacobsen SJ, Furner SE, Levy PS, Brody JA, Goldberg J. Hip fracture incidence among elderly Asian-American populations. Am J Epidemiol. 1997; 146:502–9. [PubMed: 9290511]
23. Shimizu H, Ross RK, Bernstein L, Yatani R, Henderson BE, Mack TM. Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County. Br J Cancer. 1991; 63:963–6. [PubMed: 2069852]
24. Gomez N, Guendelman S, Harley KG, Gomez SL. Nativity and neighborhood characteristics and cervical cancer stage at diagnosis and survival outcomes among Hispanic women in California. Am J Public Health. 2015; 105:538–45. [PubMed: 25602869]

25. Howe HL, Alo CJ, Lumpkin JR, Qualls RY, Lehnherr M. Cancer incidence and age at northern migration of African Americans in Illinois, 1986–1991. Ethn Health. 1997; 2:209–21. [PubMed: 9426985]
26. Schupp CW, Press DJ, Gomez SL. Immigration factors and prostate cancer survival among Hispanic men in California: does neighborhood matter? Cancer. 2014; 120:1401–8. [PubMed: 24477988]
27. Horn-Ross PL, Chang ET, Clarke CA, et al. Nativity and papillary thyroid cancer incidence rates among Hispanic women in California. Cancer. 2012; 118:216–22. [PubMed: 21692062]
28. Chang ET, Shema SJ, Wakelee HA, Clarke CA, Gomez SL. Uncovering disparities in survival after non-small-cell lung cancer among Asian/Pacific Islander ethnic populations in California. Cancer Epidemiol Biomarkers Prev. 2009; 18:2248–55. [PubMed: 19622719]
29. Horn-Ross PL, McClure LA, Chang ET, et al. Papillary thyroid cancer incidence rates vary significantly by birthplace in Asian American women. Cancer Causes Control. 2011; 22:479–85. [PubMed: 21207130]
30. Patel MI, Schupp CW, Gomez SL, Chang ET, Wakelee HA. How do social factors explain outcomes in non-small-cell lung cancer among Hispanics in California? Explaining the Hispanic paradox. J Clin Oncol. 2013; 31:3572–8. [PubMed: 23960183]
31. Chang ET, Gomez SL, Fish K, et al. Gastric cancer incidence among Hispanics in California: patterns by time, nativity, and neighborhood characteristics. Cancer Epidemiol Biomarkers Prev. 2012; 21:709–19. [PubMed: 22374991]
32. Raz DJ, Gomez SL, Chang ET, et al. Epidemiology of non-small cell lung cancer in Asian Americans: incidence patterns among six subgroups by nativity. J Thorac Oncol. 2008; 3:1391–7. [PubMed: 19057262]
33. Nasseri K, Moulton LH. Patterns of death in the first and second generation immigrants from selected Middle Eastern countries in California. J Immigr Minor Health. 2011; 13:361–70. [PubMed: 19621261]
34. Anton-Culver H, Culver BD, Kurosaki T. An epidemiologic study of asbestos-related chest x-ray changes to identify work areas of high risk in a shipyard population. Appl Ind Hyg. 1989; 4:110–118.
35. Anton-Culver H, Culver BD, Kurosaki T. Immune response in shipyard workers with x ray abnormalities consistent with asbestos exposure. Br J Ind Med. 1988; 45:464–8. [PubMed: 3260800]
36. Barron E, Bamberger F. Meaning of the Social Security Number. Social Security Bulletin. 1982:29–30.
37. Puckett C. The Story of the Social Security Number. Social Security Bulletin. 2009; 69:55–74. [PubMed: 19697506]
38. Long WS. Social Security Numbers Issued: A 20-Year Review. Social Security Bulletin. 1993; 56:83–86.
39. Social Security Number Randomization: Social Secuirty Administration.

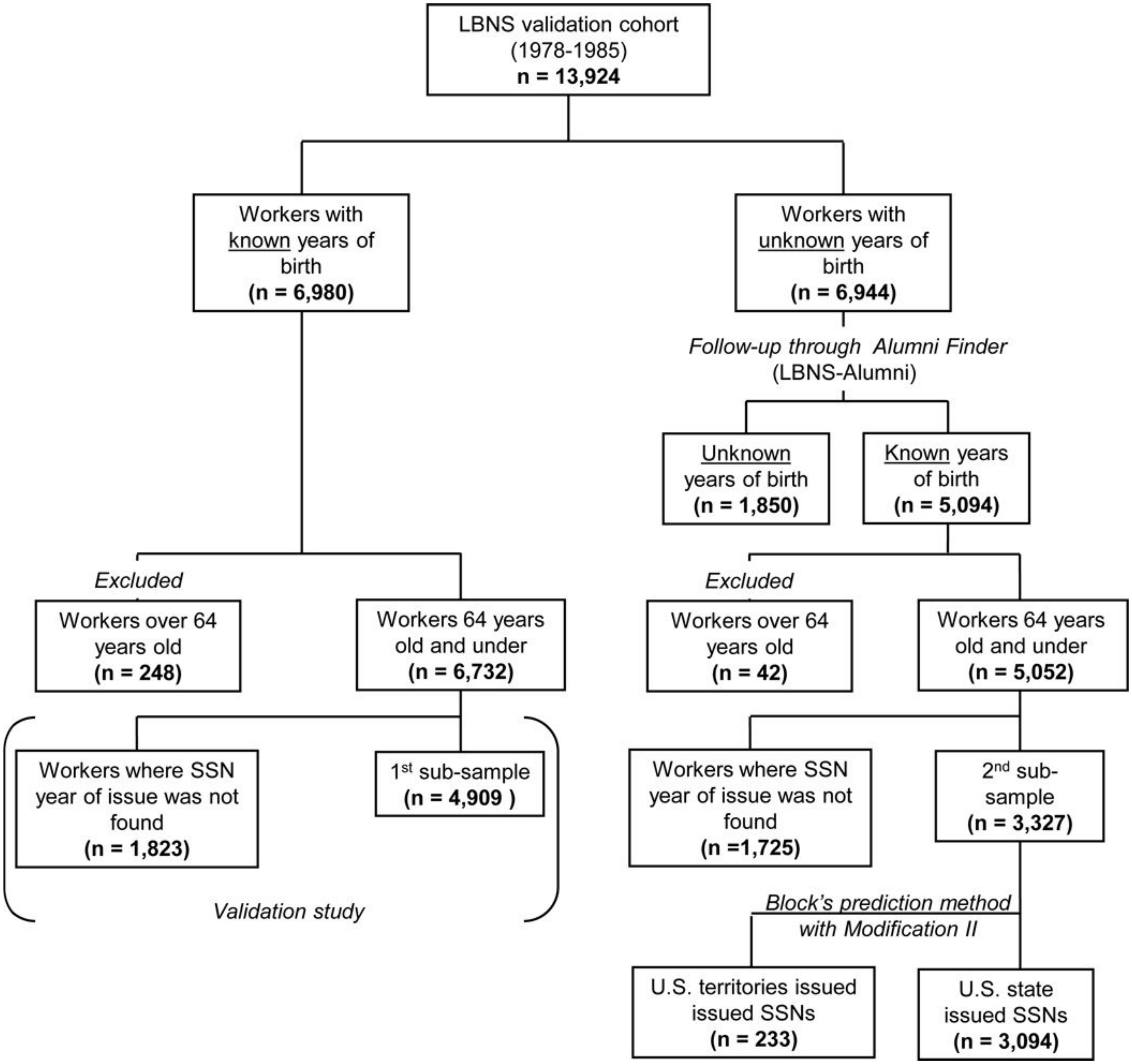

Figure 1.
Outline of the Long Beach Naval Shipyard cohort (1978–1985) validation study of Block's method to predict years of birth using Social Security Numbers. LBNS, Long Beach Naval Shipyard; RRB, Railroad Board; SSN, Social Security Number.

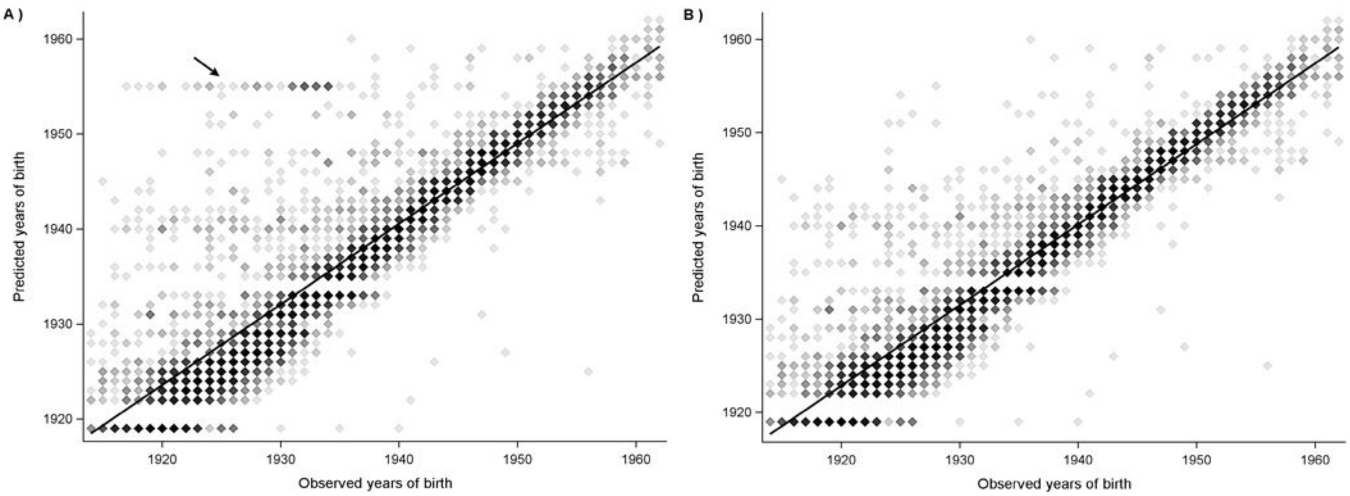

Figure 2.

Relationships between predicted and observed years of birth in LBNS-VC workers (n = 6,980). A) unmodified Block's method, B) Block's method with modification II. Arrow is cluster of workers with U.S. territories-issued SSN

Table 1

Distribution of the Social Secuirty Number Median Age of Issue

Year obtained SSN	Median age used by Block et al. (1983)[a]	Median age from LBNS-VC
1937	26	18
1938	20	19
1939–1950	17	17
1951–1959	16	16
1960–1962	15	16
1963–1974	14	16
1975–1978	13	16

Abbreviations: LBNS-VC, Long Beach Naval Shipyard validation cohort; SSN, Social Security Number.

[a]Source: Block G, Matanoski GM, Seltser RS. A method for estimating year of birth using social security number. *Am J Epidemiol* 1983;118(3): 377–95.

Table 2

Comparison of Modifications of Block's Prediction Method Using the Long Beach Naval Shipyard Cohort (1978–1985)

Sample	Modifications	N	Pearson's *r* [a] (% agreement)	R^2	*P* value[b]
LBNS-VC					
	unmodified	4,909	89.2	0.796	< 0.0001
	I	4,909	89.1	0.793	< 0.0001
	II	4,701	93.1	0.867	< 0.0001
LBNS-Alumni					
	unmodified	3,327	89.1	0.794	< 0.0001
	II	3,094	92.6	0.857	< 0.0001

Abbrivations: LBNS, Long Beach Naval Shipyard; LBNS-VC, Long Beach Naval Shipyard validation cohort; N, frequency.

[a] Pearson's correlation coefficient, expressed in %.

[b] Statistically significant if $P < 0.05$ for correlation between predicted years of birth and observed years of birth.

Table 3

U.S. Territories- and States-issued Social Security Number in the Long Beach Naval Shipyard Cohorts, 1978–1985

Sample	Location of SSN issuance	Frequency
LBNS-VC		
	U.S. territories	208
	U.S. states	4,701
	Total	**4,909**
LBNS-Alumni		
	U.S. territories	233
	U.S. states	3,094
	Total	**3,327**

Abbreviations: LBNS, Long Beach Naval Shipyard; LBNS-VC, Long Beach Naval Shipyard validation cohort; SSN, Social Security Number.

Table 4

Distribution of Predicted Year of Birth in the LBNS-VC With Varying Degree of Accuracy After Excluding U.S. Territories-issued SSNs, by SSN Year of Issue

SSN year of issuance	0 years		2 years		3–5 years		>5 years		Total
	Frequency	%	Frequency	%	Frequency	%	Frequency	%	
1936–1947	253	27.6	843	34.9	466	51.2	135	29.3	1,697
1948–1950	84	9.2	274	11.3	108	11.9	58	12.6	524
1951+	579	63.2	1,298	53.7	336	36.9	267	58.0	2,480
Total[a]	916	19.5	2,415	51.4	910	19.3	460	9.8	4,701

Abbreviations: LBNS-VC, Long Beach Naval Shipyard validation cohort; SSN, Social Security Number.

[a]% total are row percents.