AGE HEAPING IN THE UNITED STATES CENSUS: 1880–1950*

MELVIN ZELNIK

ALTHOUGH the problem of age misreporting in the census (and other social surveys) has long been recognized,[1] few techniques have been developed for estimating and correcting the errors involved. Those techniques which are available for demographic analysis are, for the most part, concerned with digital preferences or five year age distributions;[2] as such they are inadequate for determining preferences or avoidances for individual years of age.

The method employed in this paper is an attempt to establish the size of error for each year of age, relative to its neighbors, for the native white male and native white female populations enumerated in the United States censuses from 1880 to 1950.[3] A concept of central importance is that of age ratio.

[1] See Young, Allyn A.: Age, *Supplementary Analysis and Derivative Tables, Twelfth Census of the United States,* U. S. Bureau of the Census, 1900, pp. 130–174. In the remainder of the paper the term age heaping will be used rather than age misreporting. By age heaping the author means the recognized phenomenon of people reporting themselves at an age other than, but close to, their true age, as for example, the preference for ages ending in 0 and 5. Defined in this manner, age heaping can be considered a major type of age misreporting.

[2] Department of Social Affairs, Population Branch, United Nations, "Accuracy Tests for Census Age Distributions Tabulated in Five-Year and Ten-Year Groups," *Population Bulletin,* No. 2, October 1952: pp. 59–79; and "The Accuracy of Quality of Basic Data for Population Estimates, Chapter 3 *in Methods of Appraisal of Quality of Basic Data for Population Estimates,* ST/SOA, Series A, (Population Studies) No. 23.

[3] U. S. Census Office, 1880: *Statistics of the Population of the United States,* Table 20, pp. 548–550; *idem,* 1890, *Report on Population of the United States,* Part II, Table 1, pp. 2–5; *idem,* 1900, *Population,* Part II, Table 1, pp. 2–5; U. S. Bureau of the Census: Population, General Report and Analysis, Table 29, *in 1910 Census of Population,* Vol. I, pp. 310–313; *idem,* Population, General Report and Analytical Tables, Table 9, *in 1920 Census of Population,* Vol. II, pp. 162–165; *idem,* General Report, Statistics by Subjects, Table 21, *in 1930 Census of Population,* Vol. II, pp. 595–596; *idem,* Characteristics by Age, Part I (U. S. Sum-

As used in this paper, age ratio is defined as the ratio resulting from the number of persons at any age divided by the average of the ten adjacent ages, five on each side.[4]

It is important to make explicit the fact that no attempt is here being made to determine the amount or degree of under (or over) enumeration of the entire population or at any age.[5] Even a census of complete enumeration (and therefore no under or over enumeration at any age) could contain errors in the age distribution due to the preference of people for certain ages and the avoidance of others. The problem may be re-phrased in this manner: errors in the number of people reported at any age in a census are composed of two parts, 1) the error resulting from the under or over enumeration of that age and 2) the error resulting from the tendency of people to avoid or select certain ages other than their "true" age. The method developed in this paper is concerned only with the second source of error.[6]

An age distribution of a closed population enumerated at any point in time is a consequence of four factors: 1) the number of births in each year preceding the count; 2) the number of deaths in each birth cohort from time of birth to time of

mary), Table 3, *in 1940 Census of Population,* Vol. IV, p. 13; *idem,* Characteristics of the Population, Part I (U. S. Summary), Table 94, *in 1950 Census of Population,* Vol. II, pp. 1–165.

The study is restricted to the white population because the method used was found to be inapplicable to the nonwhites (see footnote 16). 1880 was picked as the starting point because of the poor quality of the 1870 census and because censuses prior to that date do not contain the necessary single year age distributions.

[4] This term has been introduced and defined at this point to eliminate the possibilities of confusion resulting from the use of a similar term that frequently has another meaning.

[5] See Coale, Ansley J., The Population of the United States in 1950 Classified by Age, Sex, and Color—A Revision of Census Figures. *Journal of the American Statistical Association,* March 1955, Vol. 50, No. 269: pp. 16–54.

[6] This is to some degree an overstatement. The method being used is designed to discover relative error but it is impossible to distinguish its source. A low age ratio may result from an under enumeration of that age or from an avoidance of that age. Under enumeration, however, covers a range of ages and to this extent its influence is mitigated by the technique being used.

This method is also inadequate for the detection of gross errors in age misreporting, i.e., when the selected age is not close to the true age. Some of this would undoubtedly be detected but cannot be adequately corrected for, as the method is based on the assumption that avoided or preferred ages are adjacent or close to the true age (see footnote 1). It seems unlikely that any noticeable amount of misreporting would be the result of this form of age preference.

enumeration; 3) age heaping; and 4) under and/or over enumeration of the population. (As previously mentioned, the fourth factor is outside the scope of the method being employed.)[7] It follows that age ratios calculated on the basis of a given age distribution will also be affected, in varying degrees, by these factors. If the provisional assumption is made that, over a "short span of time" the number of births does not deviate from a linear trend and that deaths do not cause a marked deviation from linearity, the age ratios will approximate unity; where deviations do result, they will be the result of age heaping.[8]

On the basis of this rationale, age ratios were calculated for the native white males and native white females enumerated in each of the eight censuses from 1880 to 1950.[9] A ten year age-interval was used in the calculation of these ratios because: 1) it was considered short enough to approximate a straight line; 2) long enough to reduce the effect of small yearly fluctuations from the trend; and 3) the denominator for each age was composed of a series of ages ending in all other digits.[10]

The effect of mortality on the age ratios was estimated by the use of life tables. Age ratios were calculated for the "populations"[11] of different life tables widely separated in time.[12] In

[7] A closed population is here being approximated as the techniques are being applied to the native white populations. Emigration is of negligible importance except for the males in the 1950 census, owing to the size of the military forces overseas. In the hope of increasing the readability of this paper, technical details have been put in appendices at the end. For the way in which the 1950 overseas population was reallocated, and adjustments made in the 1930 and 1940 figures to provide necessary single year age distributions of comparable populations, see Appendix A.

[8] The assumption of the linearity of births and the effects of mortality will be examined and corrected for in subsequent sections of the paper. In addition, see Appendix B.

[9] See footnote 3 and Appendix A.

[10] More accurately, the denominator for each age was composed of a series of ages in which eight digits appeared once and one digit twice. See Appendix C for the manner in which adjustments were made to age ratios affected by the duplication of a digit.

[11] i.e., the $_1L_x$ columns of the life tables.

[12] For the females: U. S. Bureau of the Census, *United States Life Tables 1890, 1901, 1910 and 1901–1910*, prepared by Glover, James A., Table 21, pp. 92–93; *idem, United States Life Tables and Actuarial Tables 1939–1941*, prepared by Greville, Thomas N. E., Table 6, pp. 36–37.

each series deviations from unity were of negligible proportions in the years of low mortality (i.e., ages 10–50); more noticeable deviations appeared at the earlier and later ages.

These life table age ratios were taken as evidence of the need for adjustment of the population age ratios in so far as they indicated the effect of mortality on the linearity of the enumerated populations. The same correction factor was applied to all population age ratios for a specific age, as calculated from each of the eight censuses. This seemed justifiable in view of the small differences existing between the two series of life table age ratios for the same year of age. The single correction, for each age, was arrived at by averaging the life table age ratios for corresponding years of age. Where this average deviated from unity by less than .005, no correction was applied; where the average deviated from unity by .005 or more, the population age ratios were corrected by dividing through by this factor. Corrections were of minor influence except at the extreme ages. The native white female life table adjusted age ratios, and the correction factor applied to each age, are shown in Table 1.[13]

The ideal situation for estimating the validity of the assumption of linearity of births (and for correcting where invalid) would be an annual series of birth statistics for the total United States beginning about 1850. Unfortunately, these figures are lacking, birth statistics for the total United States not having been collected until 1933, when the birth registration area first included all states.[14] The method that has been used to correct for the provisional assumption of the linearity of births depends instead on the identification of large and

For the males: *United States Life Tables 1890, 1901, 1910 and 1901–1910*, Table 19, pp. 90–91 and *United States Life Tables and Actuarial Tables 1939–1941*, Table 5, pp. 34–35.

[13] Although both sexes were treated separately, but by identical methods, tables and discussion are restricted mainly to the females. Final results are shown for the males.

[14] An attempt was made to use the birth statistics available for a few states as far back as the mid-1800's. For the manner in which these figures were handled and the reasons why an alternative method proved necessary, see Appendix D.

Table 1. Age ratios* for census enumerated native white females, ages 5–85, 1880–1950, adjusted by life table correction factor.

AGE	CENSUS								LIFE TABLE CORRECTION FACTOR
	1880	1890	1900	1910	1920	1930	1940	1950	
5	1.041	1.047	1.019	1.013	1.037	1.049	1.006	.943	.992
6	1.062	1.093	1.043	1.025	1.026	1.035	.936	.989	1.000
7	.994	.988	1.020	.983	1.005	1.004	.951	1.059	1.000
8	1.016	1.007	1.031	.980	1.013	1.068	.997	.986	1.000
9	.955	.932	.979	.947	.973	1.025	.983	.951	1.000
10	1.057	1.012	1.032	.986	1.022	1.011	1.020	.953	1.000
11	.939	.894	.968	.928	.995	.968	.979	.945	1.000
12	1.083	1.051	.990	1.029	1.057	1.028	1.041	.992	1.000
13	.976	.949	.965	.976	.985	.978	1.010	.973	1.000
14	.969	1.032	.990	1.022	1.002	1.010	1.007	.956	1.000
15	.847	.968	.995	.967	.938	.982	1.006	.962	1.000
16	.933	1.077	1.030	1.060	1.017	1.023	1.031	.930	1.000
17	.916	.962	.992	.998	.963	.990	.987	.945	1.000
18	1.156	1.116	1.022	1.089	1.009	1.043	1.086	1.002	1.000
19	1.018	.943	.967	.967	.971	.995	1.025	1.004	1.000
20	1.146	1.072	1.021	1.035	.969	1.008	1.000	1.002	1.000
21	.961	.958	.946	.963	.963	1.004	1.008	1.015	1.000
22	1.102	1.070	1.027	1.012	1.017	1.030	.988	.999	1.000
23	1.012	1.008	1.017	1.008	1.028	1.006	.992	.996	1.000
24	1.028	.989	1.052	1.022	1.040	1.009	1.010	1.001	1.000
25	1.065	.984	1.062	1.053	1.056	1.013	1.025	1.065	1.000
26	.961	.911	.981	1.012	1.026	.985	1.009	1.025	1.000
27	.866	.877	.984	.931	.981	.932	.988	.994	1.000
28	1.040	1.102	1.015	1.061	1.055	.987	1.028	1.052	1.000
29	.814	.896	.934	.893	.930	.985	.995	1.012	1.000
30	1.310	1.358	1.155	1.158	1.135	1.127	1.078	1.064	1.000
31	.711	.786	.875	.798	.856	.846	.902	.943	1.000
32	.992	1.038	1.008	1.033	1.010	1.023	1.064	1.020	1.000
33	.903	.947	.959	.925	.926	.945	.965	.970	1.000
34	.896	.928	.936	.984	.959	.978	1.000	.993	1.000
35	1.227	1.138	.985	1.116	1.111	1.096	1.036	1.052	1.000
36	1.004	.979	.893	.995	1.000	1.010	.970	1.010	1.000
37	.888	.880	.919	.926	.911	.960	.917	.980	1.000
38	1.074	1.067	1.104	1.121	1.105	1.115	1.000	1.031	1.000
39	.886	.868	1.071	.952	.972	.961	1.033	.998	1.000
40	1.436	1.350	1.207	1.251	1.175	1.174	1.148	1.108	1.000
41	.684	.731	.886	.789	.779	.815	.833	.907	1.000
42	.999	1.008	1.030	1.081	1.079	1.079	1.073	1.088	1.000
43	.882	.891	.954	.920	.947	.924	.961	.987	1.000
44	.920	.895	.970	.877	.928	.921	.941	.931	1.000
45	1.233	1.239	1.070	1.043	1.137	1.113	1.060	1.023	1.000
46	.904	.951	.908	.830	.909	.934	.964	.939	1.000
47	.890	.896	.932	.886	.946	.924	.993	.939	1.000
48	1.023	1.047	.989	1.134	1.053	1.064	1.063	.976	1.000
49	.873	.861	.972	1.040	.987	.965	.970	1.057	1.000
50	1.492	1.459	1.204	1.335	1.293	1.230	1.175	1.163	1.000
51	.714	.728	.880	.825	.813	.795	.864	.859	1.000
52	1.039	1.012	.991	1.078	1.064	1.051	1.053	1.025	1.000
53	.901	.918	.920	.925	.928	.946	.932	.963	1.000
54	.982	1.013	.980	1.014	.943	1.039	1.005	1.008	1.000

Table 1 (*Continued*)

AGE	CENSUS								LIFE TABLE CORRECTION FACTOR
	1880	1890	1900	1910	1920	1930	1940	1950	
55	1.103	1.042	1.081	1.020	.934	1.035	1.010	1.002	1.000
56	1.000	1.034	1.000	.972	.905	.962	.975	.987	1.000
57	.815	.881	.933	.865	.885	.925	.931	.996	1.000
58	.967	.964	.948	.984	1.089	1.034	1.046	1.039	1.005
59	.841	.782	.959	.912	1.015	.953	.973	.975	1.005
60	1.512	1.436	1.193	1.241	1.282	1.250	1.124	1.103	1.006
61	.721	.705	.819	.767	.816	.798	.808	.872	1.007
62	1.006	.970	.975	.986	1.019	1.012	.967	.963	1.008
63	.955	1.013	1.012	.976	1.028	.993	.986	.957	1.008
64	.936	.971	.985	.967	.957	.918	1.017	.958	1.009
65	1.213	1.240	1.064	1.196	1.136	1.088	1.202	1.211	1.009
66	.886	.918	.978	.924	.860	.798	.904	.952	1.010
67	.880	.898	.967	.917	.884	.878	.966	.966	1.010
68	1.007	.970	.932	.998	.966	1.076	1.005	.999	1.011
69	.886	.803	.932	.938	.956	1.045	.968	.940	1.011
70	1.332	1.352	1.112	1.138	1.122	1.191	1.138	1.065	1.012
71	.728	.744	.816	.754	.791	.844	.842	.841	1.012
72	.990	1.019	1.000	1.019	1.002	1.047	1.020	1.012	1.012
73	.943	.946	.943	.958	.931	.960	.966	.988	1.012
74	.958	.944	.973	.965	.960	.945	.914	.999	1.011
75	1.120	1.108	1.080	1.074	1.155	1.076	.975	1.085	1.009
76	.972	.948	.941	.999	.960	.925	.863	.960	1.006
77	.846	.871	.929	.897	.902	.852	.856	.913	1.000
78	.931	1.017	.938	.942	.935	.920	1.037	.980	1.000
79	.906	.815	.926	.883	.950	.916	1.037	.949	.991
80	1.340	1.228	1.092	1.037	1.005	1.044	1.126	1.106	.982
81	.712	.792	.887	.773	.790	.824	.897	.880	.972
82	.917	.972	.950	.986	.976	.960	.995	.999	.960
83	.897	.970	.924	.969	1.005	.950	.983	.989	.946
84	1.068	1.084	.976	1.048	1.024	.996	.985	.971	.930
85	.954	1.004	.923	.976	.944	.979	.932	.895	.912
Average Deviation	.118	.106	.058	.076	.070	.066	.051	.045	

$$* \text{Age Ratio} = \frac{\text{Number at any age}}{\text{Average of ten adjacent ages}}$$

Sources: The age ratios have been calculated from the United States censuses 1880–1950 (see footnote 3). The life table adjustment factor is the average of the life table age ratios calculated from the $_1L_x$ values of the 1901 and 1939–1941 life tables (see footnote 12).

small birth cohorts (relative to their neighbors) at points in time some years after birth. More explicitly, cohort size was estimated from its effect on the age ratios calculated from the enumerated populations in a series of censuses.

Table 1 shows a decrease in the average deviation (the mean of the absolute deviations from unity) for each census, if 1900 is ignored.[15] When the age ratios for each year of age are

[15] This is true of the males also, where the values are 1880: .107; 1890: .097;

AGE RATIOS & FIRST TREND LINE

AGE 5

AGE 15

AGE 25

AGE 35

AGE 45

AGE 55

AGE 65

AGE 75

AGE 85
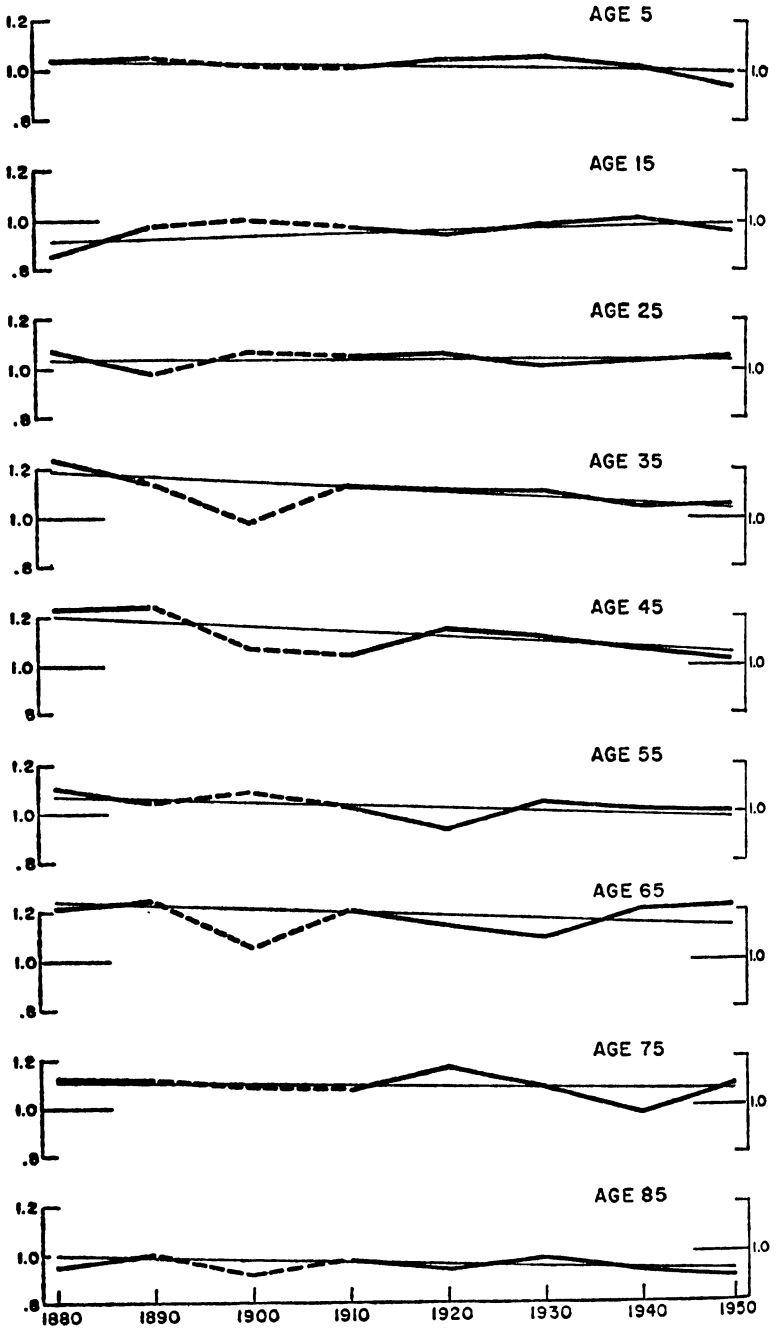
Fig. 1. Trend lines fitted to age ratios of enumerated native white females, censuses of 1880–1950: ages ending in 5. Source: Table 1.

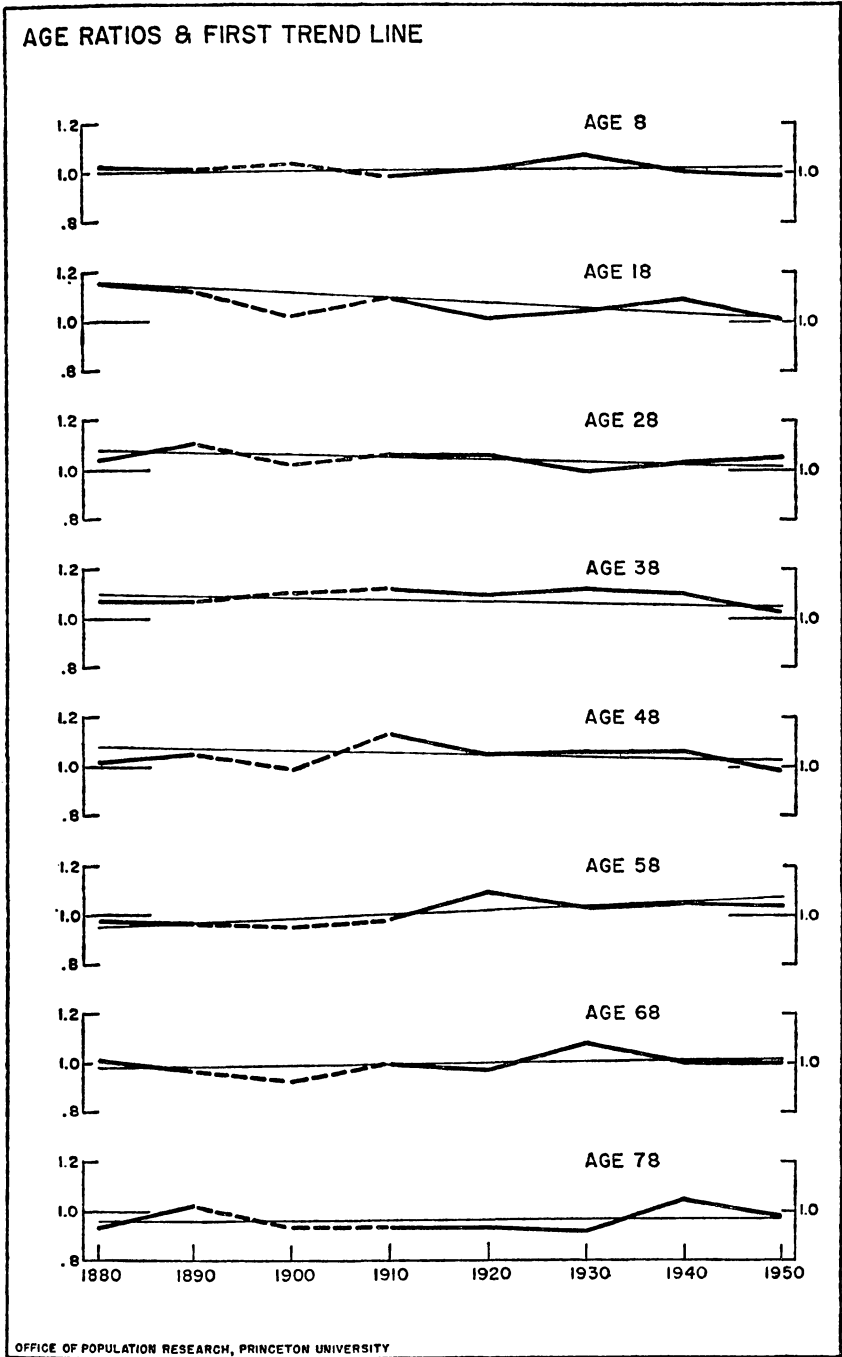AGE RATIOS & FIRST TREND LINE

AGE 8

AGE 18

AGE 28

AGE 38

AGE 48

AGE 58

AGE 68

AGE 78

Fig. 2. Trend lines fitted to age ratios of enumerated native white females, censuses of 1880–1950: ages ending in 8. Source: Table 1.

plotted, they appear to be approaching unity in a linear fash-
ion,[16] with the exception in most instances of 1900 (see Fig-
ures 1 and 2).[17]

It was therefore assumed, on the basis of this evidence, that
the age ratios for each age were, through time, approaching
unity in a linear fashion (again, with the noted exception of
1900) with deviations from a trend line due to the original
size of the birth cohort and to random fluctuations of age
heaping about the trend line. For each age a trend line was
established by taking the average, for one point, of the 1880,
1890, and 1910 values and for the other point the values of
1920, 1930, 1940, and 1950.[18] These average age ratios were
centered on 1893 1/3 and 1935 respectively; intermediate
censal values were estimated by simple interpolation while
values for censuses beyond these points were arrived at by
simple extrapolation. (See Figures 1 and 2).

The true size of a birth cohort relative to its neighbors re-
mains very nearly constant through time (except perhaps for
certain ages of males through decimation of a cohort by war
losses). A cohort of unusual size should appear as a deviation
from the trend line each time the cohort is enumerated. Thus
the birth cohort of 1864 can easily be traced through time by
its appearance as a deviation from eight separate trend lines
(representing eight ages) from 1880 when it was age 15 to 1940
when it was age 75 (Figure 1). If the trend lines actually rep-

1900: .054; 1910: .066; 1920: .061; 1930: .057; 1940: .046; and 1950: .038. The
1900 census asked not only "age at last birthday" but also "date of birth," the only
time (prior to 1960) this information has been requested in a United States census.
All available evidence seems to suggest a higher degree of accuracy in the 1900 age
distributions resulting from the inclusion of this question. See Young, *op. cit.*

[16] It was at this point that it became apparent that the method used to adjust
the native white populations would not be applicable to the nonwhites. The life
table adjusted age ratios for them are clearly curvilinear; this meant that much
more sophisticated methods would be needed, assuming that the problem was even
solveable. No attempt has been made to work out an alternative technique which
would be applicable to the nonwhites.

[17] To avoid repetition of figures which are basically similar, it was decided to
include figures for some ages only, for the purpose of illustration.

[18] Since 1900 obviously does not fit the trend of age heaping, it would be in-
correct to include it for the establishment of the trend line. This also resulted in
treating the figures for this year slightly differently; see below.

resent the degree of age heaping, the deviations from the trend lines reflect the size of the birth cohorts. The size of the deviations does not remain constant because of random elements and consequently the average of the deviations was considered to be the closest approximation to the size of the birth cohort.[19]

The deviations of each birth cohort from the trend lines as it moves through time, and is therefore enumerated at an age approximately 10 years older in each census,[20] were totaled and averaged; this average figure may be considered the deviation of the age ratio from the trend line, due to cohort size. Because of the non-consistency of the 1900 census with the general assumption of linearity in age heaping, the deviations of the cohorts enumerated in 1900 were not included in estimating the average size of the cohort.

The trend lines themselves, however, were influenced by the size of the cohort. In other words, the age ratios used to determine the trend lines reflect both age heaping and the relative size of the birth cohort. An estimate of cohort size based on these trend lines will not be a true measure of actual cohort size but may be considered a first approximation to it. Consequently, a correction factor was derived, reflecting the influence of the cohort on the trend lines, and applied to the average cohort deviation.[21] This corrected cohort deviation (which is the first approximation to the relative size of the birth cohort) was used to adjust the age ratios, taking into account the differences in census date.

Since the age ratios have been corrected by the size of the cohort (or at least a first approximation to it), they may be considered a "more pure" measure of age heaping, and trend lines based on these corrected age ratios a more accurate rep-

[19] To the degree that random fluctuations are actually present, they tend to cancel one another.

[20] A major source of complication arises from the fact that the censuses have not always been taken as of the same date. This means that we are not dealing with the same birth cohort as it moves through time and is enumerated at each census. For the differences in census dates and the way in which this problem was handled, see Appendix E.

[21] See Appendix F for the derivation of this correction factor.

resentation of the linear trend in this phenomenon. This also implies that the differences between the original age ratios (as corrected for mortality) and trend lines established on the basis of the cohort adjusted age ratios will provide a more accurate measure of the size of the cohort.

Following this line of reasoning, trend lines were calculated from the cohort adjusted age ratios in the same manner as the first trend lines.[22] These trend lines, following the original assumption of linearity in age heaping, are the degree of age heaping for each year of age at each census, with the exception of 1900.[23]

If this assumption is valid, then the age ratios, corrected by adjustment for cohort size from the second approximation trend lines, should agree very closely with the trend line values. The "closeness of fit" between these two measures is shown in Figures 3 and 4. While not perfect, it can be seen that there is a high degree of improvement (see Figures 1 and 2) and a relatively good agreement.[24]

Theoretically the relative size of each birth cohort should be the same for both sexes, i.e., the correlation between the relative size of birth cohorts as estimated for males and females should be 1.00. A scatter diagram of these two series (Figure 5) shows that the correlation, while not unity, is extremely high and when calculated turns out to be 0.97. The deviations present can possibly be attributed to random elements present in age heaping, to the effects of under enumeration of one sex as compared to the other, or to the effects of war losses on the males.

Figure 6 shows these estimates of the relative size of the birth cohorts plotted against time. This diagram illustrates mainly

---

[22] Since every cohort had to be enumerated at least twice to estimate its size, it was not possible to adjust the age ratios for ages 5–14 in 1950 and 75–85 in 1880. The second trend lines for ages 5–14 were thus arrived at by averaging the 1920, 1930, and 1940 figures, centering it on 1930; for ages 75–85, the left side of the trend line was the average of 1890 and 1910, centered on 1900. For all other ages, the procedure used was the same as in determining the first trend lines.

[23] See Appendix B.

[24] See Appendix G for discussion of ages 62–68, where it has been suggested a change has taken place in age heaping as a result of social security legislation.
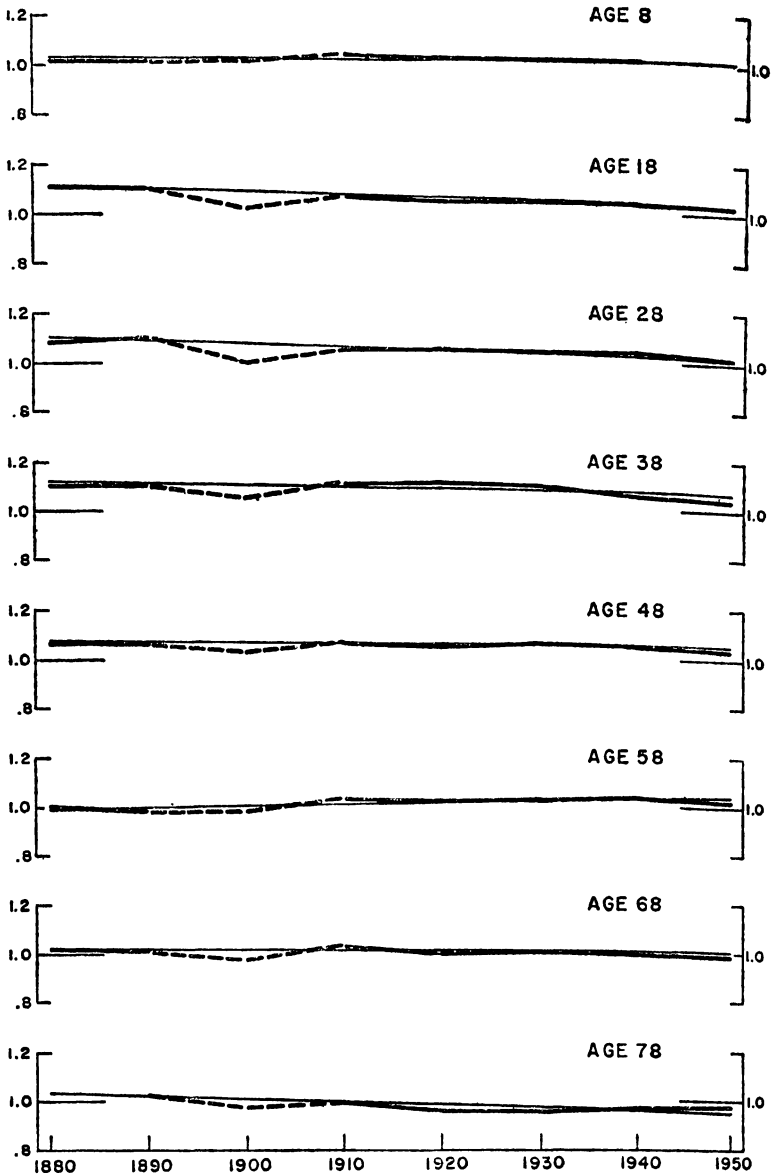
AGE RATIOS & SECOND TREND LINE
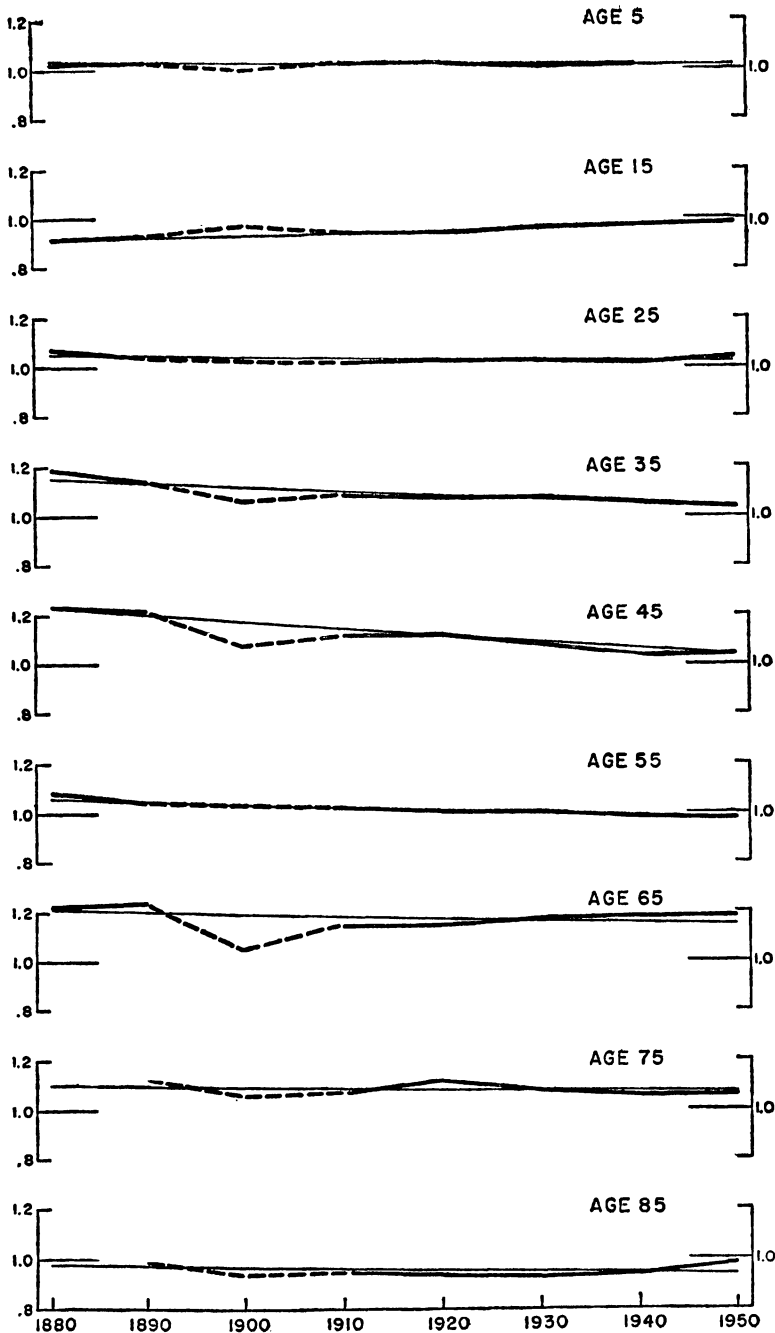
Fig. 3. Modified trend lines and age ratios* adjusted for cohort size, native white females, 1880–1950: ages ending in 5.

* The trend lines have been fitted to the age ratios after the first correction for cohort size, while the age ratios shown are based on the second correction for cohort size; see text, p. 550.

Fig. 4. Modified trend lines and age ratios* adjusted for cohort size, native white females, 1880–1950: ages ending in 8.
* The trend lines have been fitted to the age ratios after the first correction for cohort size, while the age ratios shown are based on the second correction for cohort size; see text, p. 550.
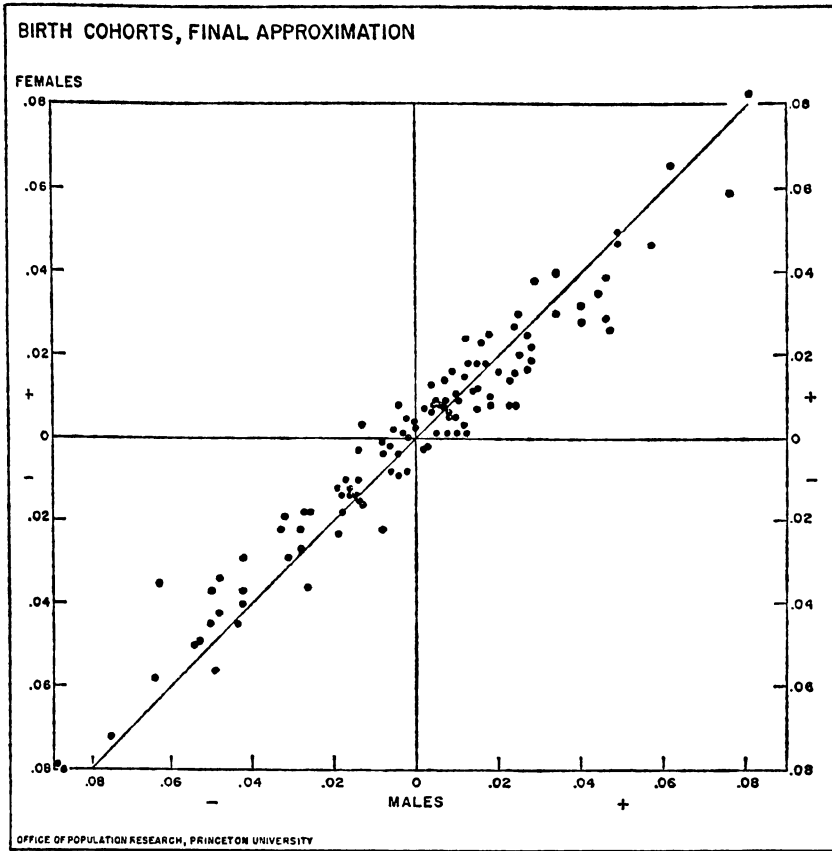
Fig. 5. Scattergram of average relative size of birth cohorts for white males and females, final approximation, 1820–1834.

two points: 1) as mentioned, the close correlation between the male and female series; 2) the large dips and rises in the births at periods of time when they would be "expected" (e.g., the hollow during the Civil War and the small size of the 1930 cohorts).[25]

These estimates of birth cohorts, while providing some confirmation of the method used and some indication of the relative size of each cohort, cannot be used to estimate the absolute size of the yearly birth cohorts. They are "relative figures" and therefore are affected by their neighbors; a cohort "bounded"

[25] The series actually extended back to the birth cohort of April 1805–March 1806. The earliest years are not shown simply to reduce the size of the figure. The correlation for the years eliminated is hardly lower than for all the years shown.
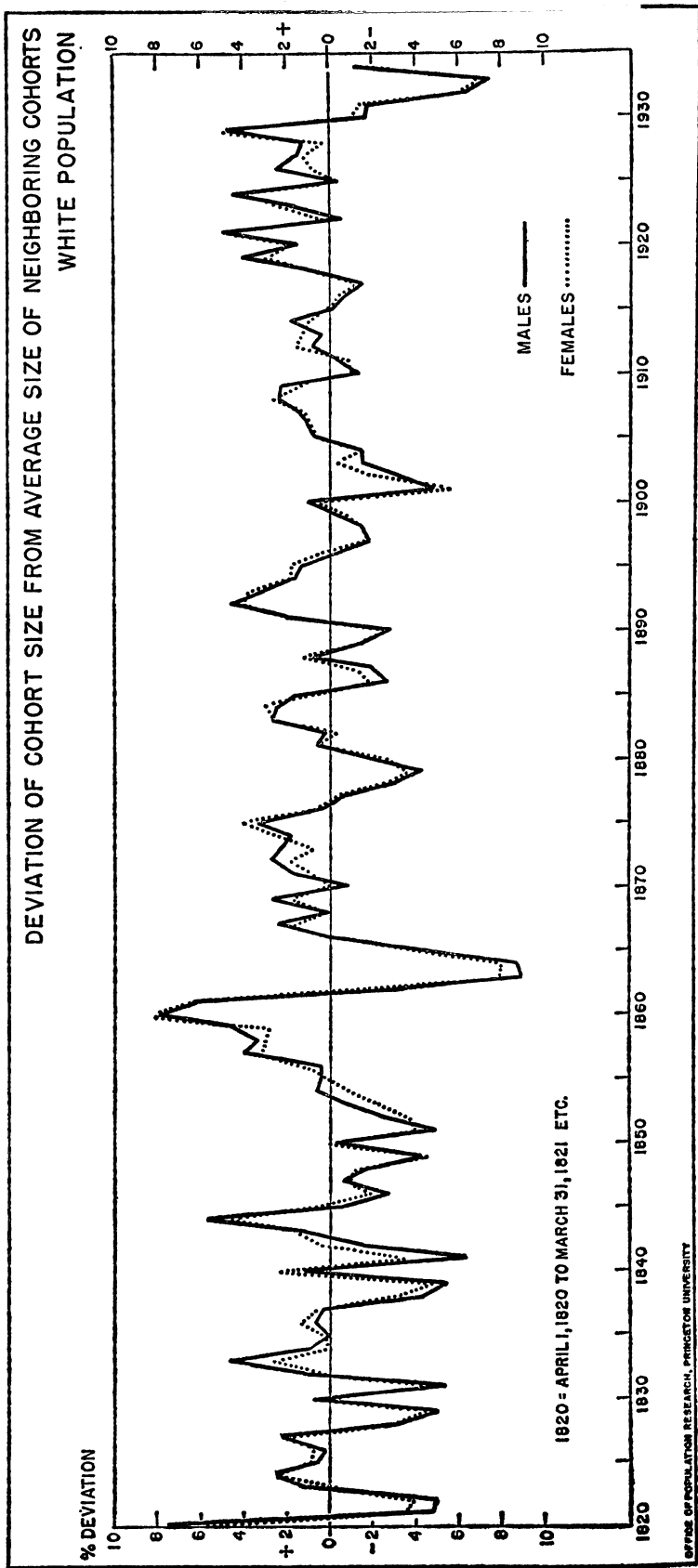
Fig. 6. Percentage deviation of white male and female birth cohorts from average size of ten neighboring birth cohorts, 1820–1934.

Table 2. Age heaping adjustment factors, native white females, ages 5–85, 1880–1950.*

AGE	CENSUS							
	1880	1890	1900	1910	1920	1930	1940	1950
5	1.027	1.025	1.001	1.022	1.020	1.018	1.016	1.014
6	1.055	1.046	1.007	1.028	1.019	1.010	1.001	.992
7	.980	.985	.984	.994	.998	1.003	1.007	1.012
8	1.021	1.019	1.010	1.015	1.014	1.012	1.010	1.008
9	.935	.946	.998	.968	.978	.989	1.000	1.011
10	1.032	1.024	1.044	1.008	1.000	1.000[1]	1.000[1]	1.000[1]
11	.931	.939	.965	.955	.963	.971	.979	.987
12	1.062	1.056	.998	1.043	1.037	1.031	1.024	1.018
13	.960	.966	.984	.977	.982	.987	.993	.998
14	1.007	1.007	.988	1.007	1.008	1.008	1.008	1.008
15	.932	.940	.972	.954	.961	.968	.976	.983
16	1.033	1.030	1.004	1.022	1.019	1.015	1.012	1.008
17	.946	.954	.991	.970	.978	.986	.995	1.003
18	1.097	1.086	1.015	1.062	1.050	1.039	1.027	1.015
19	.949	.957	.986	.973	.981	.989	.997	1.005
20	1.090	1.070	1.050	1.031	1.011	1.000[1]	1.000[1]	1.000[1]
21	.928	.940	.973	.963	.975	.986	.998	1.010
22	1.059	1.050	1.030	1.030	1.020	1.010	1.000	.990
23	1.011	1.009	1.009	1.005	1.004	1.002	1.000	.998
24	1.020	1.018	1.016	1.014	1.013	1.011	1.009	1.007
25	1.046	1.042	1.031	1.035	1.032	1.028	1.024	1.021
26	.986	.987	.973	.991	.993	.995	.996	.998
27	.920	.929	.970	.947	.956	.965	.974	.983
28	1.086	1.076	1.004	1.054	1.043	1.032	1.022	1.011
29	.828	.856	.941	.914	.941	.968	.995	1.022
30	1.295	1.258	1.127	1.184	1.146	1.107	1.068	1.030
31	.747	.774	.879	.827	.853	.879	.907	.932
32	1.009	1.014	.995	1.022	1.027	1.032	1.036	1.040
33	.918	.925	.962	.940	.947	.954	.961	.968
34	.907	.920	.972	.945	.958	.970	.983	.996
35	1.139	1.124	1.052	1.095	1.081	1.067	1.052	1.038
36	.985	.985	.976	.985	.984	.984	.984	.984
37	.895	.903	.955	.919	.927	.935	.943	.951
38	1.105	1.099	1.048	1.086	1.080	1.074	1.068	1.061
39	.891	.914	.992	.955	.976	.996	1.017	1.038
40	1.370	1.331	1.152	1.250	1.212	1.169	1.127	1.085
41	.752	.769	.867	.803	.820	.837	.854	.871
42	1.012	1.023	.998	1.047	1.058	1.070	1.082	1.094
43	.896	.906	.948	.928	.939	.950	.960	.971
44	.923	.923	.970	.923	.923	.923	.923	.923
45	1.207	1.181	1.069	1.129	1.103	1.076	1.050	1.024
46	.901	.906	.935	.914	.919	.924	.928	.932
47	.879	.892	.969	.917	.929	.941	.952	.962
48	1.060	1.058	1.027	1.052	1.050	1.047	1.044	1.041
49	.883	.907	.980	.950	.972	.994	1.015	1.037
50	1.436	1.395	1.217	1.310	1.267	1.227	1.182	1.137
51	.769	.785	.908	.815	.830	.845	.860	.875
52	1.024	1.029	1.000	1.038	1.042	1.047	1.051	1.056
53	.904	.913	.942	.931	.940	.949	.958	.967
54	.997	.997	.978	.997	.996	.996	.996	.996

Table 2 (*Continued*)

AGE	CENSUS							
	1880	1890	1900	1910	1920	1930	1940	1950
55	1.056	1.045	1.037	1.023	1.013	1.002	.991	.980
56	1.014	1.004	.982	.984	.974	.964	.954	.944
57	.852	.865	.935	.892	.906	.919	.932	.944
58	.994	.999	.978	1.010	1.015	1.021	1.026	1.032
59	.819	.848	.950	.907	.935	.963	.991	1.019
60	1.416	1.372	1.208	1.282	1.236	1.192	1.144	1.096
61	.742	.758	.863	.790	.806	.822	.838	.854
62	.996	.994	.978	.990	.988	.986	.985	.983
63	.989	.990	.999	.992	.993	.994	.995	.996
64	.961	.962	.983	.964	.965	.966	.967	.968
65	1.194	1.186	1.057	1.170	1.162	1.154	1.146	1.138
66	.917	.915	.960	.912	.910	.908	.906	.905
67	.890	.901	.966	.919	.928	.937	.946	.955
68	1.021	1.017	.976	1.010	1.006	1.003	.999	.996
69	.833	.858	.942	.910	.934	.959	.983	1.007
70	1.267	1.240	1.133	1.188	1.160	1.132	1.104	1.076
71	.722	.744	.860	.787	.809	.831	.852	.874
72	.996	.998	.989	1.004	1.006	1.009	1.012	1.014
73	.940	.944	.941	.953	.958	.962	.967	.971
74	.960	.959	.968	.957	.956	.955	.954	.953
75	1.094	1.090	1.055	1.081	1.076	1.072	1.068	1.063
76	.978	.974	.946	.965	.960	.956	.951	.947
77	.877	.881	.965	.891	.897	.902	.906	.911
78	1.023	1.014	.975	.994	.984	.974	.964	.954
79	.819	.842	.895	.889	.914	.936	.959	.981
80	1.166	1.148	1.054	1.114	1.097	1.080	1.063	1.046
81	.760	.778	.885	.814	.832	.850	.867	.885
82	.955	.959	.946	.966	.969	.973	.977	.980
83	.955	.960	.946	.971	.977	.982	.987	.993
84	1.082	1.068	.986	1.039	1.024	1.010	.996	.981
85	.982	.977	.942	.966	.960	.955	.950	.944

* These figures were used as divisors to correct the census enumerated populations for age heaping. The figures for 1880, 1890, and 1910–1950 represent the adjusted second approximation trend lines, the derivation of which is described in the text and Appendix B. The figures for 1900 are the life table adjusted age ratios corrected for cohort size.

¹ These age ratios have not been allowed to go below unity by assumption; see pp. 558 and 563.

by smaller sized cohorts appears larger than it actually should.

The second approximation trend line values, or the age heaping adjustment factors, were used to correct the enumerated populations. In the case of 1900, the original age ratios as corrected by the final approximation of cohort size were used to adjust the enumerated population.[26] This procedure was followed because of the afore-mentioned deviation of 1900 from

---

[26] Technically speaking, the adjustment factors are the second approximation trend line values corrected by the adjustment described in Appendix B.

Table 3. Age heaping adjustment factors, native white males, ages 5–85,
1880–1950.*

AGE	CENSUS							
	1880	1890	1900	1910	1920	1930	1940	1950
5	1.021	1.021	1.003	1.021	1.020	1.020	1.020	1.020
6	1.041	1.035	1.010	1.022	1.016	1.010	1.004	.997
7	.977	.981	.980	.990	.995	.999	1.004	1.008
8	1.009	1.009	1.003	1.009	1.008	1.008	1.008	1.008
9	.937	.949	1.003	.972	.984	.996	1.007	1.019
10	1.053	1.040	1.049	1.015	1.003	1.000[1]	1.000[1]	1.000[1]
11	.932	.938	.965	.950	.957	.963	.969	.976
12	1.080	1.073	1.007	1.058	1.051	1.044	1.037	1.030
13	.968	.972	.994	.979	.983	.986	.990	.994
14	1.040	1.036	.997	1.029	1.025	1.022	1.018	1.014
15	.932	.940	.975	.954	.961	.968	.976	.983
16	1.004	1.007	.999	1.013	1.015	1.018	1.021	1.023
17	.948	.959	.988	.980	.991	1.002	1.013	1.023
18	1.022	1.022	.991	1.022	1.022	1.022	1.022	1.022
19	.970	.975	.970	.984	.988	.993	.997	1.002
20	.962	.960	1.013	.954	.951	.949	.946	.943
21	1.058	1.050	1.032	1.036	1.029	1.022	1.014	1.007
22	1.034	1.027	1.025	1.013	1.005	.998	.991	.984
23	1.037	1.029	1.004	1.013	1.004	.996	.988	.980
24	1.011	1.010	1.010	1.008	1.007	1.006	1.005	1.004
25	1.000	.999	1.011	.997	.996	.996	.995	.994
26	.978	.980	.964	.986	.988	.991	.994	.996
27	.958	.964	.975	.977	.983	.989	.996	1.002
28	1.094	1.081	1.013	1.056	1.043	1.031	1.018	1.005
29	.873	.899	.958	.946	.969	.993	1.016	1.040
30	1.296	1.253	1.144	1.165	1.119	1.073	1.027	.981
31	.779	.803	.877	.850	.873	.898	.921	.943
32	.985	.990	.986	1.001	1.006	1.012	1.017	1.022
33	.927	.934	.960	.949	.956	.963	.970	.978
34	.910	.926	.968	.959	.975	.991	1.007	1.023
35	1.161	1.142	1.067	1.104	1.086	1.067	1.048	1.029
36	.948	.954	.968	.967	.973	.979	.986	.992
37	.880	.897	.954	.926	.941	.955	.969	.984
38	1.097	1.088	1.057	1.070	1.061	1.052	1.043	1.034
39	.928	.945	.994	.979	.996	1.014	1.031	1.048
40	1.345	1.303	1.160	1.221	1.177	1.133	1.089	1.045
41	.791	.807	.874	.839	.855	.871	.887	.905
42	1.003	1.015	.991	1.040	1.053	1.066	1.078	1.091
43	.878	.895	.929	.924	.938	.952	.967	.981
44	.887	.892	.960	.903	.907	.912	.916	.921
45	1.232	1.202	1.098	1.140	1.110	1.079	1.049	1.018
46	.907	.915	.930	.932	.940	.948	.956	.964
47	.881	.899	.958	.930	.945	.960	.976	.991
48	1.023	1.022	1.010	1.018	1.016	1.014	1.013	1.011
49	.899	.921	.976	.964	.986	1.007	1.029	1.050
50	1.372	1.332	1.216	1.252	1.213	1.171	1.129	1.086
51	.817	.828	.937	.851	.862	.873	.884	.897
52	1.076	1.075	1.042	1.073	1.072	1.071	1.070	1.069
53	.959	.962	.971	.969	.973	.977	.980	.984
54	.968	.973	.970	.982	.986	.991	.996	1.000

Table 3 (*Continued*)

AGE	CENSUS							
	1880	1890	1900	1910	1920	1930	1940	1950
55	1.036	1.025	1.017	1.004	.993	.982	.971	.960
56	1.001	.996	.965	.985	.979	.974	.968	.963
57	.879	.889	.925	.909	.918	.927	.936	.945
58	.968	.972	.968	.981	.986	.990	.995	.999
59	.820	.851	.944	.912	.941	.969	.998	1.027
60	1.320	1.284	1.170	1.212	1.173	1.134	1.095	1.057
61	.787	.799	.879	.821	.833	.844	.855	.866
62	1.030	1.025	.999	1.016	1.012	1.007	1.003	.998
63	1.028	1.026	1.021	1.022	1.021	1.019	1.017	1.015
64	.950	.956	.985	.967	.972	.978	.983	.988
65	1.149	1.143	1.028	1.130	1.124	1.118	1.112	1.105
66	.919	.915	.951	.908	.905	.901	.897	.894
67	.895	.906	.992	.930	.942	.953	.965	.977
68	1.022	1.014	.979	.996	.986	.978	.968	.960
69	.878	.899	.942	.937	.956	.975	.994	1.013
70	1.199	1.174	1.096	1.123	1.098	1.073	1.048	1.022
71	.782	.800	.891	.836	.853	.871	.889	.908
72	1.014	1.017	1.001	1.022	1.025	1.028	1.031	1.033
73	.958	.962	.977	.971	.976	.980	.985	.989
74	.963	.966	.971	.971	.974	.977	.979	.982
75	1.032	1.032	1.018	1.030	1.029	1.028	1.027	1.026
76	.962	.960	.934	.954	.951	.949	.946	.943
77	.864	.873	.987	.892	.903	.912	.921	.930
78	1.041	1.029	1.003	1.004	.991	.978	.966	.953
79	.858	.879	.911	.924	.944	.965	.986	1.006
80	1.142	1.120	1.008	1.075	1.052	1.030	1.007	.985
81	.808	.822	.914	.851	.865	.879	.895	.908
82	.975	.974	.959	.972	.971	.970	.969	.968
83	.943	.951	.941	.968	.976	.984	.992	1.000
84	1.023	1.021	.977	1.015	1.013	1.010	1.007	1.004
85	.926	.925	.920	.924	.923	.922	.921	.920

* These figures were used as divisors to correct the census enumerated populations for age heaping. The figures for 1880, 1890, and 1910–1950 represent the adjusted second approximation trend lines, the derivation of which is described in the text and Appendix B. The figures for 1900 are the life table adjusted age ratios corrected for cohort size.
1 These age ratios have not been allowed to go below unity by assumption; see pp. 558 and 563.

the trend in age heaping. These adjustment factors, for both sexes, are shown in Tables 2–3.

Figures 7–14 show the enumerated and adjusted native white female populations for each of the eight censuses from 1880 to 1950. The adjusted populations, while showing much smoother age distributions, retain genuine differences in cohort size.

The age heaping adjustment factors just described resulted in an irregular pattern in the adjusted 1930, 1940, and 1950
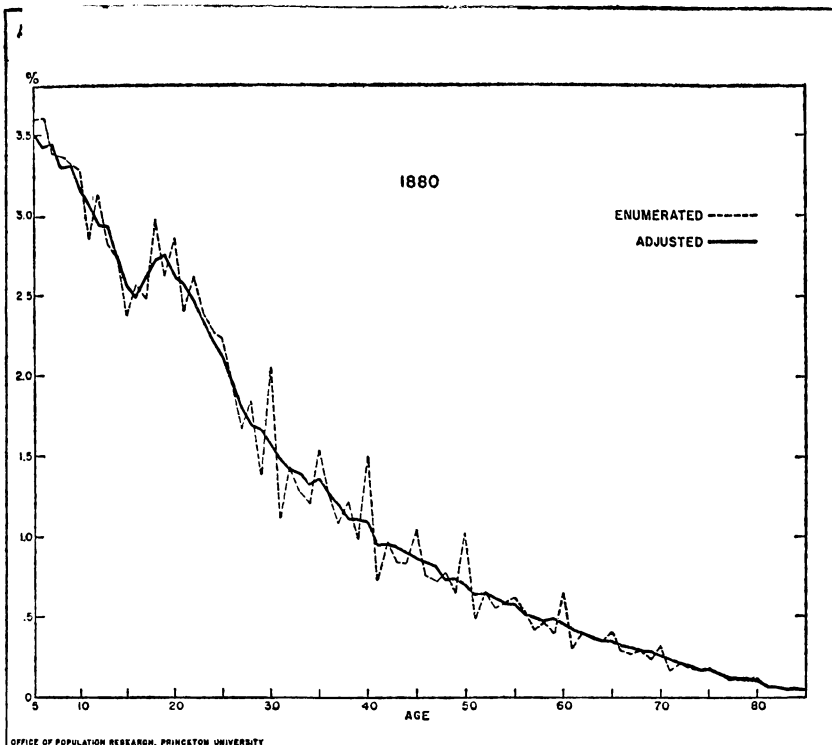
Fig. 7. Percentage distribution of native white females, ages 5–85, as enumerated by the census and as adjusted for age heaping: 1880. Sources: Adjusted population derived by use of correction factors shown in Table 2; census enumerated population, footnote 3.
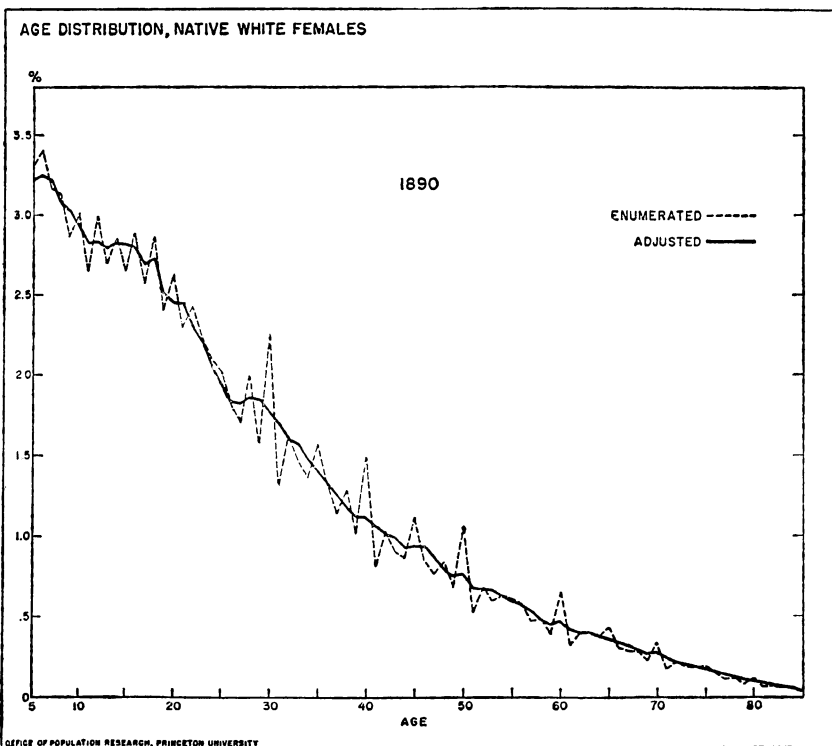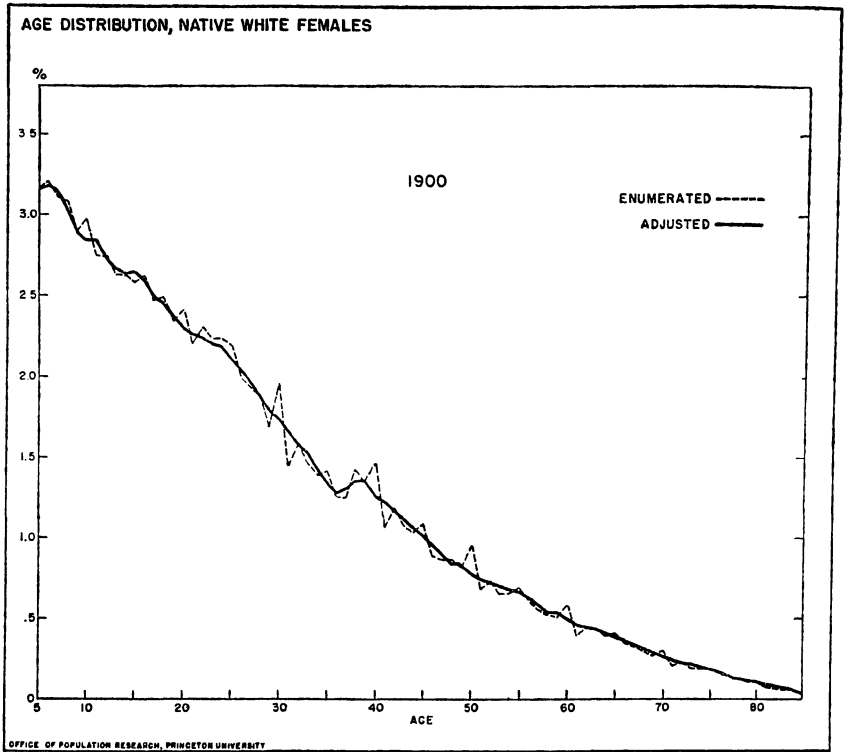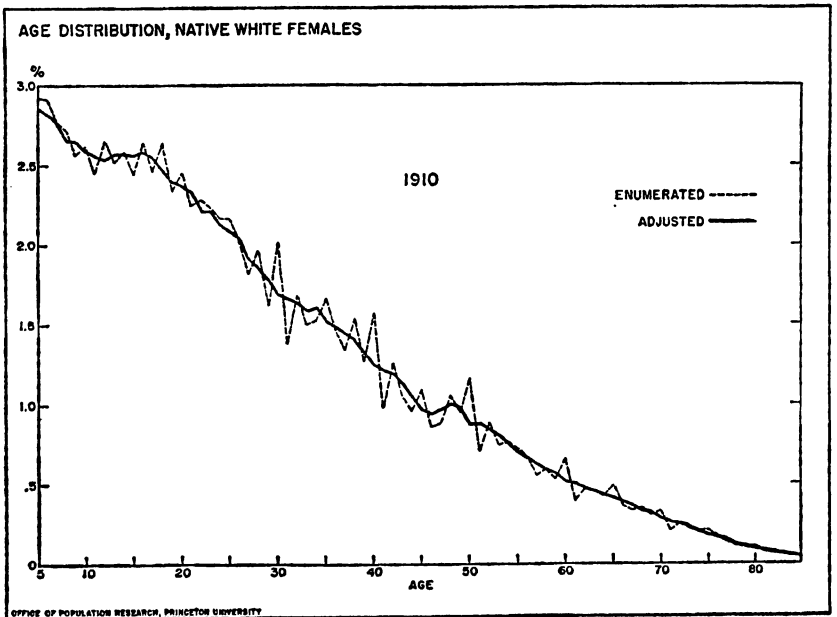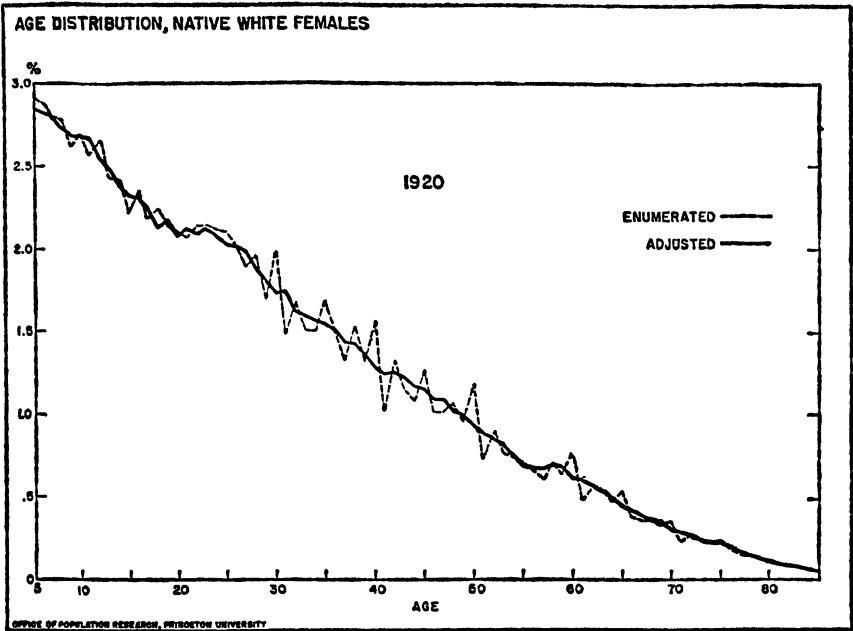

Fig. 8. Percentage distribution of native white females, ages 5–85, as enumerated by the census and as adjusted for age heaping: 1890. Sources: Adjusted population derived by use of correction factors shown in Table 2; census enumerated population, footnote 3.

Fig. 9. Percentage distribution of native white females, ages 5–85, as enumerated by the census and as adjusted for age heaping: 1900. Sources: Adjusted population derived by use of correction factors shown in Table 2; census enumerated population, footnote 3.
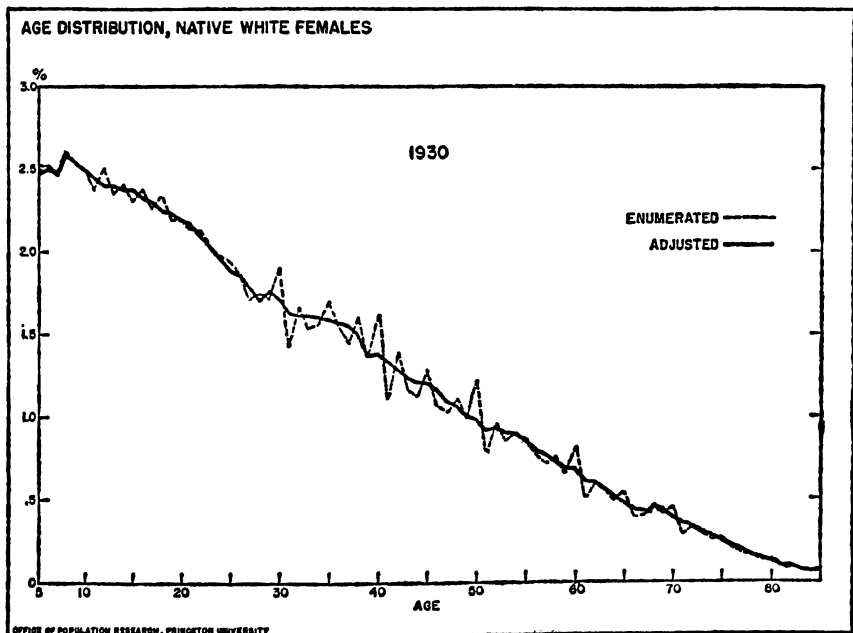

Fig. 10. Percentage distribution of native white females, ages 5–85, as enumerated by the census and as adjusted for age heaping: 1910. Sources: Adjusted population derived by use of correction factors shown in Table 2; census enumerated population, footnote 3.
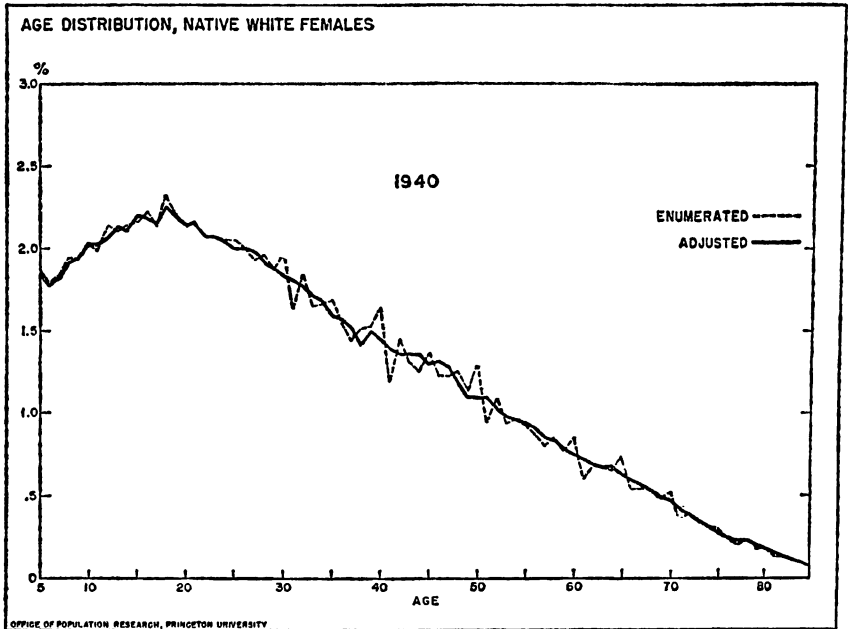
Fig. 11. Percentage distribution of native white females, ages 5–85, as enumerated by the census and as adjusted for age heaping: 1920. Sources: Adjusted population derived by use of correction factors shown in Table 2; census enumerated population, footnote 3.

Fig. 12. Percentage distribution of native white females, ages 5–85, as enumerated by the census and as adjusted for age heaping: 1930. Sources: Adjusted population derived by use of correction factors shown in Table 2; census enumerated population, footnote 3.
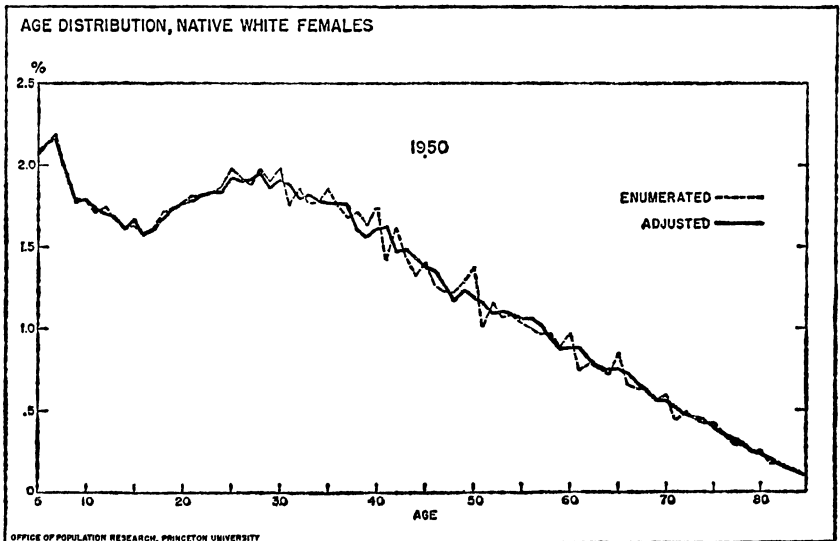
AGE DISTRIBUTION, NATIVE WHITE FEMALES

1940

ENUMERATED —————
ADJUSTED ————————

OFFICE OF POPULATION RESEARCH, PRINCETON UNIVERSITY

Fig. 13. Percentage distribution of native white females, ages 5–85 as enumerated by the census and as adjusted for age heaping: 1940. Sources: Adjusted population derived by use of correction factors shown in Table 2; census enumerated population, footnote 3.

AGE DISTRIBUTION, NATIVE WHITE FEMALES

1950

ENUMERATED —————
ADJUSTED ————————

OFFICE OF POPULATION RESEARCH, PRINCETON UNIVERSITY

Fig. 14. Percentage distribution of native white females, ages 5–85, as enumerated by the census and as adjusted for age heaping: 1950. Sources: Adjusted population derived by use of correction factors shown in Table 2; census enumerated population, footnote 3.

values at ages 10 and 20. This irregularity seemed to be traceable to the original assumption of a linear trend in age heaping. It should be obvious that while a linear trend may be valid for a series of censuses, it cannot continue indefinitely through time, for if such were the case all preferred ages would eventually become avoided and vice versa. In other words, if age heaping for age 40, for example, has diminished in a linear trend, a continuation of this trend would lead to age 40 becoming increasingly avoided, a conclusion which is highly doubtful. It appears that the assumption of linearity in age heaping led to the improbable result of avoidance of young ages ending in 0 in 1930, 1940, and 1950. What is probably more accurate for these ages is a leveling off of the age heaping rather than a continuation of a linear trend which leads to avoidance. To the degree that this conjecture is correct, the assumption of linearity resulted in an improper correction being applied to the enumerated populations at these ages.

On the basis of this reasoning,[27] the assumption was made that the age heaping adjustment factors for ages 10 and 20 in 1930, 1940, and 1950 should not be allowed to go below unity. A correction factor of 1.000, therefore, was used to adjust the enumerated populations at these ages. The same assumption was applied to the males but only for age 10 since the age heaping trend line for age 20 was continuously below unity.[28] The adjusted populations resulting from the use of this correction factor are included in Figures 12–14.

In Table 4, the per cent of age heaping for each sex in 1950 is shown for each terminal digit of age. In general, those ages which are preferred by males are preferred by females, while

---

[27] Further evidence confirming this reasoning resulted from the birth estimates which were subsequently generated from the age heaping adjusted populations.

[28] If the trend of age heaping for ages 10 and 20 in 1930, 1940, and 1950 has been "leveling off" and approaching unity asymptotically, rather than linearly, then the "true" correction factor should be slightly higher than 1.000. The use of 1.000, however, prevents an over correction of the enumerated populations and serves as a useful approximation of what the lower limit of the age heaping correction factor should have been. The difference between the "true" correction factor and the modified value used, 1.000, is probably negligible.

Table 4. Per cent of over and under age-preference of native white males and native white females, ages 5–85, 1950.

AGE	MALE	FEMALE	AGE	MALE	FEMALE
	TERMINAL DIGIT 5			TERMINAL DIGIT 6	
5	2.0	1.4	6	−0.3	−0.8
15	−1.7	−1.7	16	2.3	0.8
25	−0.6	2.1	26	−0.4	−0.2
35	2.9	3.8	36	−0.8	−1.6
45	1.8	2.4	46	−3.6	−6.8
55	−4.0	−2.0	56	−3.7	−5.6
65	10.5	13.8	66	−10.6	−9.5
75	2.6	6.3	76	−5.7	−5.3
85	−8.0	5.6			
	TERMINAL DIGIT 7			TERMINAL DIGIT 8	
7	0.8	1.2	8	0.8	0.8
17	2.3	0.3	18	2.2	1.5
27	0.2	−1.7	28	0.5	1.1
37	−1.6	−4.9	38	3.4	6.1
47	−0.9	−3.8	48	1.1	4.1
57	−5.5	−5.6	58	−0.1	3.2
67	−2.3	−4.5	68	−4.0	−0.4
77	−7.0	−8.9	78	−4.7	−4.6
	TERMINAL DIGIT 9			TERMINAL DIGIT 0	
9	1.9	1.1	10	0.0	0.0
19	0.2	0.5	20	−5.7	0.0
29	4.0	2.2	30	−1.9	3.0
39	4.8	3.8	40	4.5	8.5
49	5.0	3.7	50	8.6	13.7
59	2.7	1.9	60	5.7	9.6
69	1.3	0.7	70	2.2	7.6
79	0.6	−1.9	80	−1.5	4.6
	TERMINAL DIGIT 1			TERMINAL DIGIT 2	
11	−2.4	−1.3	12	3.0	1.8
21	0.7	1.0	22	−1.4	−1.0
31	−5.7	−6.8	32	2.2	4.0
41	−9.5	−12.9	42	9.1	9.4
51	−10.3	−12.5	52	6.9	5.6
61	−13.4	−14.6	62	−0.2	−1.7
71	−9.2	−12.6	72	3.3	1.4
81	−9.2	−11.5	82	−3.2	−2.0
	TERMINAL DIGIT 3			TERMINAL DIGIT 4	
13	−0.6	−0.2	14	1.4	0.8
23	−2.0	−0.2	24	0.4	0.7
33	−2.2	−3.2	34	2.3	−0.4
43	−1.9	−2.9	44	−7.9	−7.7
53	−1.6	−3.3	54.	0.0	−0.4
63	1.5	−0.4	64	−1.2	−3.2
73	−1.1	−2.9	74	−1.8	−4.7
83	0.0	−0.7	84	−8.0	−1.9

Sources: Table 2 for the females and Table 3 for the males.

those ages avoided by males are also avoided by females. In most cases, however, the degree of avoidance or preference is higher for females than for males.

Table 4 also points up the possible dangers involved in discussing digital preferences, thereby obscuring the range of preference and avoidance for specific ages rather than digits. It has been common to discuss age heaping as being over in ages ending in 5, 0, and even numbers, with ages ending in odd numbers being avoided. Table 4 shows this to be an over-generalization. For most digits there is heaping in both directions. Ages ending in 4 and 6 are generally avoided, probably the result of being adjacent to ages ending in 5. At the same time, some of the ages ending in odd numbers (other than 5) are preferred.

APPENDIX A

CORRECTION OF THE 1930, 1940, AND 1950 CENSUSES

In the 1930 census, Mexicans were not included as white, as they previously and subsequently were, but were classified as nonwhite, under "other races." There was a total of 409,672 native male Mexicans and 395,482 native female Mexicans, shown only by a five year age distribution. (These figures do not include those reported at "age unknown.")[1] Single year estimates were arrived at by redistributing these numbers in the same proportion within each five year age group as the native white of foreign or mixed parentage. These figures were then added to the figures given for native whites, the sum representing a total native white population comparable to other censuses.

Single year age distributions for the native white male and native white female populations are given in the 1940 census only for 35 years of age and over; under 35 years the classification is by five year age groups. There is, however, a single year age distribution, by sex, for the total white population.[2] The numbers of foreign white up to age 29 were relatively small; only in the 30–34 years of age group did they amount to a significant proportion (7.5 per cent for the males

[1] U.S. Bureau of the Census, *1930 Census of Population, op. cit.,* Table 15, p. 586.
[2] U.S. Bureau of the Census, *1940 Census of Population, op. cit.,* Table 2, p. 9.

and 7.9 per cent for the females). It therefore seemed justifiable to redistribute the 0–35 native white population to single years by the per cent distribution of total whites within each five year age group.

In the 1950 census, a total of 337,290 native white males were reported abroad in 1950.[3] Although this represents only 0.5 percent of the total native white male population, there is a fairly high degree of concentration in the 20–34 years of age group. It therefore seemed advisable to make some adjustment to the figures shown for the native white males, which are only for continental United States.[4] No adjustment was made for the native white females since the number abroad represented only 0.1 per cent of the total, with a much smoother distribution over the entire age range.

The native white males abroad in the age ranges 0–14, and 40 and over, were allocated in the same proportions as the continental native white males. For ages 15–39, a single year age distribution was arrived at by using the same percentage age distribution of white males in the armed forces stationed in continental United States. This latter figure was determined by subtracting the civilian labor force from the total labor force, the difference representing the males in the armed forces.[5]

APPENDIX B

Calculation of True Correction Factor for Age Heaping

It can easily be demonstrated that if a series of numbers is linear, the ratio of the mid-point of the series divided by the average of the ten adjacent numbers will be unity. If the assumption is made that the correct age distribution is a linear trend, then $\frac{n+r}{n}$ would be the true index of age heaping, where n represents the average of the number at the ten ages adjacent to any age, and r the net number of persons misreporting themselves at that age. This however ignores the fact that the r persons added to the numerator are "drawn out" of other ages. If a preferred age "draws" on people from within the

[3] U.S. Bureau of the Census, *1950 Census of Population, op cit.,* Table 35, p. 1-87.

[4] *Ibid.,* Table 94, pp. 1–165.

[5] U.S. Bureau of the Census: Employment and Personal Characteristics, Part 1, Chapter A in *U.S. Bureau of Population: 1950,* Vol. IV (Special Reports), Table 1, p. 1A–22.

span used to determine the denominator, then the effect is to decrease the denominator by .10r; i.e., the actual ratio is $\dfrac{n+r}{n-\dfrac{r}{10}}$.

If it is assumed that persons incorrectly reporting a given age typically have an age within five years of the one reported, the above argument shows that an "age ratio" is not an appropriate correction factor to remove (by division) the effect of age heaping. It can be shown that $CF = \dfrac{11AR}{10+AR}$ where AR represents the age ratio and CF the true correction factor. If $AR = \dfrac{n+r}{n-\dfrac{r}{10}}$ and $CF = \dfrac{n+r}{n}$

then $AR = \dfrac{n+r}{n} \times \dfrac{1}{1-\dfrac{r}{10n}}$

$$AR = CF \times \dfrac{1}{1-\dfrac{r}{10n}}$$

$$CF = AR\left(1-\dfrac{r}{10n}\right)$$

$$CF = AR\left(1-\dfrac{CF}{10}+\dfrac{1}{10}\right)$$

$$CF = AR\left(\dfrac{11}{10}-\dfrac{CF}{10}\right)$$

$$CF = AR\left(\dfrac{11}{10}\right)-AR\left(\dfrac{CF}{10}\right)$$

$$CF + AR\left(\dfrac{CF}{10}\right) = AR\left(\dfrac{11}{10}\right)$$

$$CF\left(1+\dfrac{AR}{10}\right) = AR\left(\dfrac{11}{10}\right)$$

$$CF = \dfrac{\dfrac{11}{10}AR}{\dfrac{10+AR}{10}}$$

$$CF = \dfrac{11AR}{10+AR}$$

This adjustment was applied to the age heaping correction factors

(i.e., the second approximation trend line values) and is incorporated in the figures shown in Tables 2 and 3.

APPENDIX C

ADJUSTMENT OF AGE RATIOS FOR CERTAIN AGES

Age ratios for ages ending in 5 were calculated with a denominator containing two ages ending in 0. Since ages ending in 0 appear to be heavily preferred ages, at least for age 30 and on, this would tend to depress the age ratios for those ages ending in 5. The same type of factor concerns the age ratios for those ages ending in 4 and 6 (where the former has a double 9 in the denominator and the latter a double 1—in both cases a digit which is generally thought to be avoided)[6] except that these ratios would be inflated. Since the over preference for 0 seems to be largely at the expense of the two adacent ages (those ending in and 9 and 1), the average of the three would tend to approximate the "true" figure for each age.

On the basis of this reasoning, age ratios for ages ending in 4, 5 and 6 were calculated slightly differently from those of other ages. The ages ending in 9, 0 and 1 were averaged with this average figure being included in the denominator the appropriate number of times, i.e., instead of the original numbers representing the 0 and 1 ages (in the case of an age ratio for an age ending in 5), twice the average of the ages ending in 9, 0 and 1 was used. This correction is of negligible proportions and is included in the figures shown in Table 1. Because of the small size of this correction, however, it did not seem warranted to correct the age ratios for ages ending in 0, which have a double 5 in the denominator (a digit which is less preferred than 0), or for ages ending in 1, 2, 3, 7, 8 and 9.

APPENDIX D

USE OF BIRTH REGISTRATION DATA IN ELIMINATING COHORT SIZE FROM AGE RATIOS

There is a long history of birth registration for a small number of states. An attempt was made to use these on the provisional assump-

---

[6] However, see Table 4 where, in 1950, ages ending in 9 are shown as preferred rather than avoided.

tion that they were a representative sample of fluctuations in the fertility behavior of the entire country. The states used and the periods covered were Connecticut and Massachusetts, 1853–1888; Connecticut, Massachusetts and New Hampshire, 1889–1896; Connecticut, Massachusetts, New Hampshire and Maine, 1897–1919; the ten original states of the Birth Registration area, 1920–1937; from 1933 on, birth statistics for the total United States. As these births were registered by calender year, it was first necessary to convert them to census-year births, that is, to June 1–May 31 births and to April 1–March 31 births. These census-year births were then transformed into birth ratios by the same method used to calculate the age ratios—the ratio of one year to the average of its ten adjacent neighbors. The birth ratios were then used in an effort to eliminate the cohort effect from the age ratios.

It should be clear from the method of calculating the ratios that it is not the absolute numbers which are of importance but the relative numbers. A proportional increase in the numbers for all ages, or births for all years, will in no way affect the ratios. Furthermore, where a series of births (regardless of what per cent it may be of the total absolute figures) is distributed in the same proportions as the total figures, the birth ratios are the same.

At the same time it should be obvious that the sudden "inclusion" of more states (in calculating the birth ratios) would result in meaningless figures. The "splicing" was therefore accomplished by treating each series of states separately until 5 years had elapsed. Thus, New Hampshire birth statistics, first available in 1884, meant that birth ratios up to 1888 were based on Connecticut and Massachusetts figures only. The inclusion of New Hampshire in 1884 allowed for calculation of the 1889 birth ratio using Connecticut, Massachusetts and New Hampshire data. This procedure was followed throughout.

Certain difficulties inhered in the use of this "spliced" series of birth ratios. First, it is extremely doubtful that the fertility behavior of the states used was actually representative of the fertility behavior of the total population of the United States. Secondly, the degree of under registration in Massachusetts ranged from 18 per cent in 1850 to 3.3 per cent in 1890.[7] Although there is no evidence, it is highly probable that the birth registration completeness changed in a

[7] Gutman, Robert: The Birth Statistics of Massachusetts During the Nineteenth Century. *Population Studies*, July 1956, Vol. X, No. 1: Table 2, p. 76.

similar fashion for the other states. It is exactly this changing rate of completeness which, while not influencing the number of people reported in a census, would adversely affect the birth ratios. On the basis of these considerations, this method of adjusting the age ratios for cohort size was dropped for the method described in the text.

APPENDIX E

EFFECT OF DIFFERENCES IN CENSUS DATE AND QUESTION ON AGE

The 1950, 1940, and 1930 censuses were taken as of April 1 of the censal year. The 1910 census was taken as of April 15 but has been treated in this paper as if it it had been taken on April 1, the difference of 14 days being considered insignificant. The 1920 census was effective as of January 1, while the date of the 1900 and 1880 censuses was June 1. For all of these censuses, the question pertaining to age was "age last birthday" (and in 1900 the additional question of "date of birth," as previously mentioned). The differences in census date mean that we are not always dealing with the same birth cohort separated by ten year intervals. Thus, those reporting themselves at age $x + 10$ in the 1910 census are not the survivors of the group of people reporting themselves at age x in the 1900 census. The effect of the question "age last birthday" is to make the mean age of persons reporting themselves at any age $x + .5$ rather than x.

The 1890 census was taken as of June 1. The age question for this census however was "age at nearest birthday" instead of the more usual one "age at last birthday." This different question had the effect (if accurately answered) of centering the mean age of persons reporting at any age at x and not $x + .5$ (in addition to "moving up" the reporting age by one year). The birth cohort born in 1885 was reported as age 5 in the 1890 census but the same (or approximately the same) birth cohort was reported as of age 14 in 1900.

The differences in the questions pertaining to age and in the census dates, and thereby reflecting different birth cohorts in advancing from one census to the next, necessitated rough approximations to arrive at the same birth cohort. In determining the age heaping adjustment factors, and the relative size of the birth cohorts (see Figure 6), an attempt was made to make all measurements as of April 1,

the most common of the census dates. For example, of the cohort born April 1, 1874–March 31, 1875, 5/6 were age 5 in 1880 and 1/6 age 6; in 1890 1/3 were age 15 while 2/3 were age 16; in 1900 5/6 were age 25 and 1/6 age 26; all were age 34 in 1910; in 1920 1/4 were age 43 and 3/4 age 44; in 1930 all were age 54, in 1940 age 64, and in 1950 age 74. In measuring the deviation of an age ratio from its various trend lines it was not possible to take account of these differences.

It was however possible to allow for the differences in correcting the age ratios for the effect of cohort size. This was done by reallocating the cohorts by the proper weights—those already given. This method may have affected the precision of the estimate of relative cohort size; it does not however diminish in any sense the correlation between the series for males and females since identical procedures were used for both sexes. Further, since the trend line of age heaping is continuous, it represents the degree of age heaping at each point in time.

APPENDIX F

ESTIMATION OF THE EFFECT OF COHORT SIZE ON THE TREND LINES

If the cohort size is known to be d then one end of each trend line is raised by $\dfrac{d}{n_i}$ where n_i represents the number of age ratios used in establishing that end of the trend line.

Let n_i = number of age ratios determining position of one end of trend line

N = number of times a cohort is observed

d' = estimated size of cohort as determined by the average of the deviations from the several original trend lines

d = true size of cohort = $d' + \dfrac{d}{n_i}$

then $\dfrac{\left(\dfrac{d}{n_i}\right)}{N} = \dfrac{\text{sum of effects when affected}}{\text{number of times observed}}$ = the error in estimating the cohort size due to the effect of this cohort on each of its several trend lines. If this error is rd, then the estimate $d' = (1 - r)d$ or $d = \dfrac{d'}{(1 - r)}$.

APPENDIX G

THE EFFECT OF SOCIAL SECURITY LEGISLATION ON THE PATTERNS OF AGE HEAPING

There has been some speculation as to the effect of old age assistance legislation enacted in the 1930's on patterns of age heaping.[8] According to this view, ages 60–64 were relatively understated in 1940 and 1950, whereas ages 65–69 were heavily overstated with the former reporting themselves at the latter ages so as to be eligible to receive old age benefits.

A test of this "hypothesis" was made by treating ages 62–68 in a slightly different way (that is, in addition to the way in which all ages, including these, were treated). Since the change is supposed to have occurred in 1940 and 1950, age heaping for these ages in these two censuses would not follow the linear trend originally assumed to exist. The pattern would instead be curvilinear. For these ages therefore, the second end of each trend line was determined by averaging the values of 1910, 1920, and 1930 and centering it on 1920. Values after this date were determined by extrapolation.

The average relative size of each cohort was then determined in the same manner as explained in the text except that the deviations for ages 62–68 in 1940 and 1950 were not included in the estimation. In this respect these ages were treated in a fashion similar to the way in which 1900 was treated.

After adjusting the age ratios for cohort size and plotting the cohort adjusted values, certain features were revealed which threw some doubt on the supposed differences of these ages in 1940 and 1950. First, the "closeness of fit" between these values and their trend lines was not as good as had been the case when these ages had not been treated differently from the other ages.

Secondly, for some ages, the divergence from the linear trend was as large, or almost as large, in 1930 as in 1940 and 1950. Since the old age assistance legislation had not been enacted until the middle 1930's, it cannot explain this. Thirdly, for certain ages, notably 64 and 68 (for both sexes), the cohort adjusted age ratios went in a direction opposite to what one would expect—age 64 showed an increase in heaping while age 68 showed a decrease. This result also

---

[8] See e.g., Coale, Ansley J., *op. cit.*, p. 20, and U.S. Bureau of the Census, *1940 Census of Population, op. cit.*, p. 3.

tends to contradict the hypothesis of changing patterns resulting from the legislation.

It also appears that different estimates of age heaping could have been obtained for any age, on the basis of shortened trend lines. In view of these considerations, it was decided to use the results obtained originally and to ignore the supposed change in the age heaping of these ages in the 1940 and 1950 censuses. This test is not intended as a denial of the hypothesis that there was a change in the patterns of age heaping resulting from the legislation mentioned. It does however suggest that the simple cause and effect relationship posited is not obvious when dealing with single year age distributions rather than 5 year groups.