DON KARL ROWNEY AND EDWARD G. STOCKWELL

The Russian Census of 1897: Some Observations on the Age Data

The most common Russian population records of the eighteenth and early nineteenth centuries—the *revizii*—were the product of the state's effort to keep track of the population primarily for tax purposes. The narrowness of this approach to documenting the size and distribution of the population—particularly the absence of socioeconomic data—gradually led to replacement of the *revizii* by more comprehensive statistics, including the census. Unlike the *revizii*, the census of 1897[1] was to be a statement of population size and characteristics on a specific date of record, a "single-day" census (*odnodnevnaia perepis'*).[2] In addition, the census collected relatively broad data on the population, including items ranging from age, sex, and place of birth to items such as class, literacy and schooling, employment, and so forth. Finally, it was the aim of the census to collect and publish these data for the entire population of the empire regardless of social class, tax status, or place of residence.

The conception, development, and prosecution of the census as a project are attributed to P. P. Semenov Tian'-Shan'skii and A. I. Troinitskii, both civil servants in the Central Statistical Committee of the Ministry of Internal Affairs. Semenov was a man of broad scientific interests, trained in the physical and life sciences, who became interested in statistics after establishing his reputation in geography, botany, and natural history.[3] In 1864 he was appointed director of the Central Statistical Committee.[4] Semenov held this position for sixteen years, during which time he was in a position to dominate the organizational development of the committee and to influence the development of its scientific standards. It was during Semenov's tenure as director that some of

the most important statistical series were begun.[5] Semenov's influence on the collection of socioeconomic data and their statistical analysis was enhanced by his role, after 1875, as chairman of the Statistical Council, an interministerial body which in the 1880s and 1890s became an important source of policy on data gathering. .

Troinitskii seems likely to have been a more successful bureaucrat than professional statistician. Although he was at one point a vice-president of the International Statistical Institute, his training was at the Alexander Lyceum, one of the elite schools established for the purpose of training bureaucrat-generalists, not specialists.[6] Troinitskii's career began in 1862 and he spent almost all of his first twenty years of service in the provinces: first on the staff of governors, then as a vice-governor, and finally as governor (of Viatka and Riazan').[7] It was not until 1883 that he succeeded Semenov as director of the Central Statistical Committee of the Ministry of Internal Affairs; in 1897 he became president of the Statistical Council, again following Semenov's path.[8] In these two positions Troinitskii had ultimate responsibility for a large number of statistical collections, analyses, and publications, just as Semenov had before him. Although these were mainly of an economic nature, the ministry also published collections of demographic data.[9] Owing to the very long tenure of both Semenov and Troinitskii in the statistical administration, and probably because of their important senior positions elsewhere in government (they were both appointed senators, for example), each seems likely to have exercised considerable influence on what were the formative years of large-scale data gathering and statistical analysis in Russia.

The degree to which Troinitskii was responsible for the original proposal for a national census is not clear. Most of the contemporary literature assigns this role to Semenov, not Troinitskii.[10] What is clear, however, is that the effort was viewed by Troinitskii and his colleagues in the ministry as a bold, massive achievement.[11] The census was published in groups of volumes, or folios, each of which dealt with one geographic unit, such as a major city, a province, or European Russia. Troinitskii's introductions (*predisloviia*) to these volumes changed somewhat according to the variations in subject matter of individual volumes and, perhaps, because the publication process required some ten years. As reflected in Troinitskii's introductions, principal responsibility for execution of the census was attributable to the Central Statistical Committee of which he was director. In the area of execution of the design, the credit was shared with

5. For example, the *Statisticheskii vremenik rossiiskoi imperii* (St. Petersburg, 1867–84).

6. *Spisok vysshikh chinov tsentral'nykh ustanovlenii Ministerstva vnutrennikh del,* part 1 (St. Petersburg, 1905), "Statisticheskii Sovet: Troinitskii, Nikolai Aleksandrovich. . . ."

7. Ibid.

8. Ibid.

9. There are directories, or bibliographies, of Ministry of Internal Affairs publications in this area. See, for example, *Spisok izdanii Tsentral'nago statisticheskago komiteta* (St. Petersburg, 1914).

10. St-k, "Vseobshchaia perepis'," p. 52. Similar credit is given to Semenov by Shokal'-skii in his article in the Brokgauz-Efron encyclopedia (*Entsiklopedicheskii slovar'*, 29:436). Semenov himself discussed the census in "Kharakternye vyvody iz pervoi vseobshchei pere-pisi," *Izvestiia imperatorskago rossiiskago geograficheskago obshchestva* (St. Petersburg, 1897), cited by Shokal'skii (Brokgauz and Efron, *Entsiklopedicheskii slovar'*, 29:436).

11. See, for example, Adrianov, *Ministerstvo vnutrennikh del*, vol. 1, pp. 223 ff.

the Main Census Commission (*Glavnaia perepisnaia kommissiia*) and a special expanded staff of the Statistical Committee.

As noted above, the content of Troinitskii's introductions changed somewhat. One constant, however, was his continual emphasis upon what a great undertaking—and achievement—the census was. Thus, he stated with pride that the first Russian census was an exceptionally ambitious event of its kind, covering one-sixth of the earth's surface and providing socioeconomic as well as demographic data on a huge population.[12] With equal consistency (and perhaps pride) Troinitskii explained that data storage and analysis had been achieved using automatic data processing equipment provided by Herman Hollerith, the technical genius for what later became the IBM Corporation. What none of Troinitskii's introductions mentioned however, was that, nearly ten years after the single-day collection of 1897, basic analysis and publication of the census were still in progress. Whatever the Hollerith machines achieved, they did not stimulate prompt reporting of the 1897 census. Naturally, the glacial pace of analysis and publication were sources of irritation to those who reviewed or evaluated the census in the Russian press. But this was only one of many points selected for criticism; and if modern historians find it possible to use population data drawn from the census of 1897 uncritically it is only by ignoring the many statisticians and demographers who have gone on record to complain about important aspects of the census. While it is not the purpose of this essay to repeat or even to summarize these critiques, it is worth calling the reader's attention to them.[13]

A common explanation for perceived and suspected inadequacies of the enumerations and the first census was the presumed organizational inadequacies and inefficiencies of collection and analysis processes.[14] This is understandable

12. See, for example, Troinitskii, "Predislovie," *Pervaia vseobshchaia perepis'*, vol. 37: *Gorod S.-Peterburg*, book 1. Evidently his pride was justified. The 1897 census was certainly the largest undertaking of its kind in Europe up to that time. Moreover, the Russian effort preceded a similar effort in China by several years. Ping-ti Ho states that a Directorate of Statistics was created in 1908 and that it worked out a plan for a modern-style census. Ho concludes: "Because of immediate political exigency, this six-year plan was 'completed' in four. Theoretically, therefore, China had taken her first modern census by 1911" (Ping-ti Ho, *Studies on the Population of China, 1368–1953* [Cambridge, 1953], p. 73).

13. In addition to the previously mentioned articles by Semenov and St-k, these include A. Lositskii, "Etiudy o naselenii Rossii po perepisi 1897 goda," *Mir bozhii*, 1905, no. 8, pp. 224–44; and V. G. Mikhailovskii, "Fakty i tsifry iz russkoi deistvitel'nosti: Naselenie Rossii po pervoi vseobshchei perepisi," *Novoe slovo*, June 1897, cited by A. Rashin, in *Naselenie Rossii za 100 let (1811–1913 gg.): Statisticheskie ocherki* (Moscow, 1956), p. 20. A short article in English was published in 1897 (F. Volkhovsky, "The Census," *Free Russia*, 8, no. 7 [1897]: 50–52).

Population data, from their inception, evidently have not only been subject to error—and thus prone to belie their apparent numerical exactness—but the error has been obvious enough for many who would use the data to notice it. A recently published work on eighteenth-century Russian urban history, for example, has argued that estimates of the Russian urban population in the eighteenth century are off by as much as half, and that these errors are more the result of misinterpretation of the enumeration data than of internal inaccuracy (see Gilbert Rozman, "Comparative Approaches to Urbanization: Russia, 1750–1800," in Michael F. Hamm, ed., *The City in Russian History* [Lexington, Ky., 1976], pp. 73–79).

14. Articles in both *Russkoe ekonomicheskoe obozrenie* and *Mir bozhii* emphasized inadequacies in the interpretative or analytic quality of the work, but Volkhovsky's note in

in light of the small amount of technical information made available by the Central Statistical Committee regarding the details of collection and analysis of either the *revizii* or the census. Troinitskii, for example, promised to provide details in his first introduction to the census.[15] But in 1900 critics were still complaining that no such information had been provided, rendering detailed evaluation of the census impossible.[16] Later criticisms of the census focused on both internal inconsistencies and contradictions between the census data and other sources, such as the *revizii* and the data published in the *Statistical Year-book* after 1897.[17] In the first instance, observations fixed on disproportionately large—or small—summary figures,[18] and overselection (that is, selecting beyond what one would normally expect) for certain ages or for ages ending in certain figures.[19] But observations of inconsistencies with other sources have been much more common, on the whole, than critiques of internal validity. Perhaps this simply reflects the fact that most of the interest in population data necessitates time series rather than extensive information on single years.[20]

Nevertheless, by taking the data as they stand in published form, it is possible to make some systematic observations about their accuracy. It is also possible to draw certain conclusions from these observations regarding the source of the observed inaccuracies. The balance of this study offers an evaluation of the accuracy of the age-distribution data of the census (that is, the number of individuals in different age categories) together with an explanation of the causes of the observed errors. The conclusion is that there are substantial in-accuracies in the 1897 census as far as age records are concerned. It can also be concluded that, whatever the degree of competence of the Central Statistical Com-mittee, the source of the errors observed was mainly the subject population, which was less amenable to control by the Central Statistical Committee, and not the government agents.

The procedure adopted for evaluating the accuracy of reported ages is a statistical one called "blending," ordinarily used by demographers to measure the degree of preference for, or avoidance of, certain ages. If, for example, there had been a tendency for any reason to record the age 30 or 40 instead of the exact ages 28, 29 or 41, 42, this procedure is designed to identify and characterize the resulting inaccuracies. Developed by Robert J. Myers,[21] the procedure creates

---

Free Russia heavily emphasized descriptions of the collection process that were unlikely to be productive of accurate results (Volkhovsky, "The Census," pp. 50–51).

15. Troinitskii, "Predislovie," *Pervaia vseobshchaia perepis'*, vol. 1: *Arkhangel'skaia guberniia,* book 1.

16. St-k, "Vseobshchaia perepis'," pp. 54–55.

17. *Statisticheskii ezhegodnik Rossii* or *Ezhegodnik Rossii* (St. Petersburg, 1904–15), published by the Ministry of Internal Affairs and not to be confused with *Ezhegodnik* of the Ministry of Finance or other, privately published, yearbooks.

18. See, for example, Lositskii, "Etiudy," pp. 226 ff.; and St-k, "Vseobshchaia perepis'," pp. 57 ff.

19. St-k, "Vseobshchaia perepis'," pp. 59–60.

20. See, for example, A. Rashin, *Naselenie Rossii,* pp. 20–21; and Robert A. Lewis, Richard H. Rowland, and Ralph S. Clem, *Nationality and Population Change in Russia and the USSR: An Evaluation of the Census Data, 1897–1970* (New York, 1976), especially chapter 2.

21. Robert J. Myers, "Errors and Bias in the Reporting of Ages in Census Data," *Trans-actions of the Actuarial Society of America,* no. 104 (October 1940), p. 395.

a statistically ideal or "blended" population which is essentially a weighted sum of the number of persons reporting ages ending in each of the ten terminal digits. In the statistically ideal population, where there are no systematic irregularities in the reporting of age, the "blended" sum at each terminal digit should be approximately equal to 10 percent of the total "blended" population. If the sum at any given digit exceeds 10 percent of the total "blended" population it indicates overselection of ages ending in that digit (digit preference). Conversely, a negative deviation (or a sum that is less than 10 percent of the "blended" total) indicates underselection of ages ending in that digit (digit avoidance). An overall measure of the extent to which there is digit preference and/or avoidance in a census age distribution is the "index of preference," which is obtained as one-half of the absolute sum of the deviations for each of the ten terminal digits.

Stockwell has calculated indexes of preference for several countries based on data taken from the *Demographic Yearbook, 1962.*[22] To illustrate, a list of countries organized from lowest to highest index is reproduced here:

Sweden, 1960	0.4
Netherlands Antilles, 1960	1.0
Ryukyu Islands, 1960	1.3
Rhodesia and Nyassaland, 1956	1.4
Republic of Korea, 1955	1.4
Taiwan, 1956	1.6
Monaco, 1961	1.8
Bermuda, 1960	1.8
Western Samoa, 1961	2.0
Hong Kong, 1961	2.0
St. Pierre and Miquelo, 1962	2.2
Thailand, 1960	2.2
Singapore, 1957	2.2
Macau, 1960	2.7
Malta and Gozo, 1957	2.7
Turks and Caicos Islands, 1960	2.8
Bulgaria, 1956	3.1
Puerto Rico, 1960	3.5
Rumania, 1956	3.6
Cayman.Islands, 1960	4.0
Jamaica, 1960	5.4
Uganda, 1959	6.5
Fiji Islands, 1956	6.5
Seychelles, 1960	6.7
Federation of Malaya, 1957	8.2
Tanganyika, 1957	8.8
Philippines, 1960	10.1
Mexico, 1960	13.3
Ghana, 1960	15.7
Turkey, 1960	22.3

22. Edward G. Stockwell, "Digit Preference and Avoidance in the 1960 Census of Mexico," *Estadistica: Journal of the Inter-American Statistical Institute* (September 1965), pp. 440–41.

Even a cursory examination of the order of this ranking suggests some tentative hypotheses about the source of error. To some extent, a higher index (that is, indication of increasingly inaccurate reporting) may be associated with the degree to which a country is unindustrialized or at an early stage of industrial development. Additional evidence of this relationship has been provided by Stockwell and Jerry W. Wicks who argue that, today, inaccuracy of age reporting or the tendency for age data to "heap" on certain preferred digits is characteristic of non-Western societies.[23] Using data from the *Demographic Yearbook, 1971*,[24] Stockwell and Wicks calculated indexes of preference for each of the sixty-four countries reporting single year of age data. Countries with an index of less than 1.0 were as follows:

> Australia, 1970
> Denmark, 1969
> England and Wales, 1966
> Finland, 1969
> Iceland, 1970
> Isle of Man, 1971
> Luxembourg, 1970
> Netherlands, 1970
> New Zealand, 1969
> Northern Ireland, 1966
> Norway, 1970
> Poland, 1970
> Scotland, 1966
> Sweden, 1970
> Switzerland, 1970
> United States, 1970[25]

In this same vein—and to illustrate further the patterns which the indexes tend to follow—it can be shown that indexes for a given country may show systematic variation over time. Thus, indexes for the United States from 1880 to 1960,[26] inclusive, are as follows:

> 1880 .... 10.4
> 1890 .... 7.8
> 1900 .... 4.7

23. Edward G. Stockwell and Jerry W. Wicks, "Age Heaping in Recent National Censuses," *Social Biology*, 21, no. 2 (Summer 1974): 163–67, and "Age Heaping in Recent National Censuses: An Addendum," *Social Biology*, 22, no. 3 (Fall 1975): 279–81. It should also be noted that these studies offer an explanation for overselecting certain "convenient" digits which are defined as "multiples of the divisors of the base of the number system." This explanation, verified by the data presented on thirty-seven national censuses, is based on interpretations offered earlier by Stanley H. Turner, "Patterns of Heaping in the Reporting of Numerical Data," *Proceedings of the Social Statistics Section of the American Statistical Association* (Washington, D.C., 1958), pp. 248–51.

24. United Nations, *Demographic Yearbook, 1971* (New York, 1972).

25. Stockwell and Wicks, "Age Heaping in Recent National Censuses," p. 164.

26. *U.S. Census of Population: 1960. Detailed Characteristics: United States Summary, Bureau of the Census*, PC (1) -1D (Washington, D.C., 1963), p. xii.

Table 1. *Terminal Digit Deviations and Indexes of Preferences, by Sex, for the Total and Literate Population: Russian Empire, 1897*

	Terminal Digit Deviations			
Terminal Digits	Total Population		Literate Population	
	Male	Female	Male	Female
0	+7.3	+11.8	+2.7	+5.8
1	−3.3	−4.6	−2.5	−3.5
2	−0.2	−1.1	+0.3	−0.1
3	−1.2	−2.0	−0.3	−0.9
4	−2.3	−2.6	−1.1	−1.2
5	+3.7	+4.8	+3.2	+2.1
6	−0.7	−1.3	0.0	0.0
7	−0.8	−1.3	−0.3	−0.6
8	+0.3	0.0	+0.3	+0.7
9	−3.1	−3.7	−2.2	−2.3
Sum	22.9	33.2	12.9	17.2
Index of Preference	11.5	16.6	6.5	8.6

Combined index of preference, total population in 1897: 13.9

1910	5.6
1920	4.5
1930	4.3
1940	3.0
1950	2.2
1960	0.8

The pattern suggests a systematic reduction of error as a function of one or more factors such as increased urbanization, industrialization, gross domestic product per capita, and literacy. There are undoubtedly other reasonable explanations as well.

Finally, it should be noted that, in addition to suggesting some of the social and economic characteristics associated with high or low indexes, these data also suggest what a "high" or "low" index is. In the preceding list of countries, for example, an index of 1 or less may be regarded as exceptionally small and an index of 10 or more as relatively high. This range should help place the Russian data from 1897 in some perspective.

Table 1 presents the range and pattern of deviation of indexes calculated from the census of 1897. It will be noted that the index of preference for the total population was 13.9, high by comparison with that of the United States even in 1880, and quite high by world standards today. It will also be noted that the range of indexes is considerable—from 16.6 for all females to 6.5 for literate males. Moreover, looking at deviation patterns by digit, it is evident that certain digits (0 and 5) were preferred systematically, while others were avoided just as systematically (1 and 9). The tendency of illiterate females to overselect or underselect certain digits is especially pronounced. This pattern

Table 2. *Age Ratios Indicating Preference for Specific Ages: Russian Empire, Females, 1897*

Ages	Terminal Digits									
	8	9	0	1	2	3	4	5	6	7
8-17	1.031	.895	1.073	.813	1.155	.942	.967	.959	1.054	1.036
18-27	1.134	.883	1.374	.638	.971	.897	.851	1.394	.905	.926
28-37	1.084	.562	2.244	.465	.859	.765	.678	1.479	.933	.870
38-47	1.053	.565	2.497	.448	.859	.722	.673	1.570	.754	.714
48-57	.979	.510	3.163	.419	.779	.665	.705	1.505	.845	.638
58-67	.763	.407	4.074	.381	.698	.648	.515	1.536	.648	.714
68-77	.648	.373	4.606	.331	.598	.496	.401	1.460	.627	.514

Source: Troinitskii, *Pervaia vseobshchaia perepis'*.

Table 3. *Indexes of Preferences, by Sex and Literacy Status, for Regions: Russian Empire, 1897*

Regions	Total Population		Literate Population	
	Male	Female	Male	Female
All regions	11.5	16.6	6.5	8.6
European Russia	9.3	15.3	5.7	8.6
Poland	7.3	13.3	6.0	9.1
Caucasus	23.3	30.7	9.9	12.7
Siberia	9.0	12.5	5.6	5.6
Central Asia	18.3	26.1	8.7	9.4
Other	24.1	14.8	22.4	4.0

Source: Troinitskii, *Pervaia vseobshchaia perepis'.*

emerges more clearly in a comparison of age ratios for all women in the Russian Empire in 1897 (table 2). An age ratio is the ratio of the number of individuals enumerated at a given age, say 40, to the arithmetic mean of the number enumerated at the five ages above and below the given age, the 35–39 and 41–45 age ranges.[27] Where there are no distortions or tendencies to "heap" on the given age, the ratio should be roughly equal to 1. In other words, the difference between the total enumerated for any given age and the mean of those enumerated in the five preceding and the five following ages should be so small that when one divides one by the other, the result should be very close to unity.

The basic pattern evident in table 2 is that ages terminating in 0 or 5 (such as 10, 20, 30, 40 or 25, 35, 45) are most heavily selected except for teenagers for whom ages such as 12, 16, and 17 were reported more frequently. The main point to be drawn from this table, however, is that digit preference/avoidance is in part a function of age (note that the magnitude of overselection of preferred digits such as "0" and the corresponding underselection of digits such as "1" and "9" increases with age).

A partial explanation of the selection patterns may be found by studying table 3. Clearly these data argue that literacy—either the ability to read an identity or birth record or to record it in the first place—plays an important role in determining the tendency to inaccurate reporting of ages. Clearly regional differences exist as well: for example, European Russia, Siberia, and Poland yielded relatively lower indexes than all regions combined; on the other hand, Central Asia and the Caucasus were relatively high. Nevertheless, the literacy characteristic is still sustained as a crucial factor, as a comparison of literacy rates among these areas will show; that is, European Russia, Siberia, and Poland were each more distinctly literate than Central Asia, the Caucasus, or the empire as a whole. For example, in the whole empire 79 percent of the population

27. Symbolically:

$$\frac{N}{\frac{(N-1)+(N+1)}{x}}$$

Where: N = total enumerated at any specified age;
$N-1$ = sum of enumerated at five preceding ages;
$N+1$ = sum of enumerated at five succeeding ages;
$x = 10$

Table 4. *Indexes of Preferences of Urban and Rural Populations, by Sex and
Literacy Status, for Regions: Russian Empire, 1897*

Regions	Total Population		Literate Population	
	Male	Female	Male	Female
Empire Total	11.5	16.6	6.5	8.6
		URBAN		
All regions	6.9	12.5	5.2	6.7
European Russia	5.7	11.5	4.1	6.5
Poland	8.6	10.6	7.8	7.7
Caucasus	15.4	23.5	10.8	9.5
Siberia	7.8	9.4	6.0	5.0
Central Asia	15.9	25.0	12.2	8.9
		RURAL		
All regions	12.6	17.3	6.9	9.9
European Russia	11.5	16.1	6.9	10.0
Poland	9.3	14.2	6.1	9.8
Caucasus	25.0	31.8	11.0	15.2
Siberia	9.6	12.8	6.2	6.1
Central Asia	18.9	26.2	9.8	10.2

Source: Troinitskii, *Pervaia vseobshchaia perepis'.*

was illiterate, in European Russia the proportion was 77 percent, and in Poland
69 percent. By contrast, the rate in the Caucasus was 87 percent and in Central
Asia 95 percent. Although additional factors were unquestionably important in
accounting for the fluctuation of indexes, it seems clear that literacy played a
central role. Similar observations may be made about the differences among
indexes calculated for men and women: in the Russian Empire 71 percent of
males were illiterate, while for females the figure rose to 87 percent.

Of course, there is little doubt that the effectiveness—that is, the accuracy
and completeness—of data gathering and data analysis for the census varied
from region to region. It seems reasonable that trained or trainable census per-
sonnel could be more easily found in urban areas than in the countryside. Varia-
tions of this kind, if they could be shown to have been independent of literacy
rates in the population, might more clearly fix responsibility for inaccuracies on
the statistical bureaucracy. An attempt to deal with this issue is made in table 4
where indexes of preference by sex are presented for the whole empire, for cities,
and for rural areas as defined by the census. These data suggest that illiteracy
is still highly significant among those factors accounting for inaccuracy in age
reporting. Highest indexes are found among the entire population in both the
cities and the countryside. The highest indexes of all are found in both the urban
and rural regions of the Caucasus and Central Asia, areas exceptionally high in
illiteracy. Nevertheless, these data also suggest that illiteracy—or the inability
to read and take records—does not account for all of the skewing of the census
age data. Even among the literate populations indexes of preference are rather
high on the whole by comparison with the index for the United States in 1900
(4.7). Moreover, among the literate populations we now see reversals of patterns
observed earlier. The rate for males in rural Poland is lower than for urban
males in the same region. Similarly, males in rural Central Asia had a lower

index than did their urban counterparts. A reversal among the rates for certain females is also evident: females in urban Poland, the Caucasus, and Central Asia have lower indexes than males, as do both rural and urban literate females in Siberia. The greatest overall differences, however, appear between total populations in urban areas as compared with their counterparts in the countryside. Males in all rural regions had an index of 12.6, while urban males had an index of only 6.9. All males and females in rural European Russia were indexed at 11.5 and 16.1 respectively, while in the urban regions the scores were 5.7 and 11.5 respectively. The urban rates were still high in absolute terms, partly because of the large component of illiterates. Nevertheless, the reduction of indexes in the cities as compared to the countryside seems quite pronounced.

To summarize, one might say that literacy seems to have been a crucial factor in explaining inaccuracies in reporting age in the 1897 census.[28] The first census of the Russian Empire was, as Troinitskii asserted, an audacious undertaking in a country at once so large and so lacking in the technical base necessary for its successful execution. From the time of its publication, the census has been subjected to much criticism of the published data, their collection, and their analysis. One should not, however, underrate the significance of the 1897 census. It was an impressive achievement in its own time and it certainly remains an important document for students of Russian and Soviet history. Nevertheless, any scholar who makes use of the data should do so only with an awareness of their limitations and biases.

Many critiques in the past have focused attention on the Central Statistical Committee as the main source of error in the census data. Although the evidence presented to sustain these criticisms is often superficial, it is perhaps obvious that in such a massive and unprecedented undertaking the relatively inexperienced committee was bound to make mistakes, even very large ones.[29] After all, this was the first census of its scope—not only in Russia but anywhere. Even so, the analysis presented here shows that systematic biases were introduced into the 1897 census not simply by inadequately trained or manipulative officials, but by the population itself. Finally, it should be emphasized that the systematic errors in age reporting discussed here limit the value of the census only in cases where single year of age data are needed (such as in actuarial analysis). Even in this case all is not lost, because techniques are available for "smoothing" the observed irregularities.[30] For most purposes serious errors in age reporting can be minimized by aggregating the data into five-year or even ten-year age groups. Such broader groupings are generally adequate for most historical, social, or demographic studies.

28. Volkhovsky states that "in towns, where, as a rule, it was expected that everybody would put the necessary information about himself in writing personally, the illiterate portion of the population had recourse to scribes for whom they had to pay" (Volkhovsky, "The Census," pp. 50–51). If this is so, it may offer a partial explanation for improved accuracy of age reporting in the cities: in cases where records were, in fact, available they could be read, if not by the subject, then by a scribe.

29. This observation was made by critics of the census in its own time (see St-k, "Vseobshchaia perepis'," p. 52).

30. See, for example, Morton D. Miller, *Elements of Graduation* (Chicago, 1946).