Biology and Society (1984), Vol. 1, 75-84

RANDOM ISONYMY AND SURNAME DISTRIBUTION IN JAPAN

NORIKAZU YASUDA

Division of Genetics, National Institute of Radiological Sciences, Chiba 260, Japan

and

NARUYA SAITOU

Department of Anthropology, Faculty of Science, The University of Tokyo, Tokyo 113, Japan

Abstract

Random isonymy in Japan was calculated to be 0.0022 from three insurance files, one customer's list from department stores and two directories of *Who's Who*. Examination of surname distributions ranked in frequency revealed that the first half was sufficient to calculate 95 per cent of the total random isonymy. The remaining half (the least frequent surnames) was distributed in a log-log linear fashion. For these findings the naming processes that occurred at the Meiji Restoration in Japan in the late nineteenth century seem to be responsible.

Introduction

It was Crow and Mange in 1965[1] who first indicated how the degree of inbreeding could be calculated from the frequency of marriage between persons of the same name (isonymy), although others, e.g. Darwin (1875)[2] and Kamizaki (1954)[3] had earlier recognized the parallelism between paternally inherited genetic characters and names, and many have applied the method since [4,5,6,7,8]. An advantage in the use of surnames is that, by a rather simple method, one can distinguish random and nonrandom components of inbreeding in the population. This paper is concerned with the estimate of random isonymy in Japan. Preliminary results have been communicated elsewhere[9].

Materials

Attempts to develop computer software for processing Chinese characters (*kanji*) in Japan required a survey of name lists. Tanaka [10,11,12] published statistical summaries of two insurance files, file A with 715,815 persons and file B with 2,647,911 individuals, sampled from all over Japan and compiled from names coded in Japanese characters (*katakana*). From the most frequent sixty names, random isonyms in Japan were calculated as 0.00184 for file A and 0.00212 for file B[9]. Not only is the complete name list of file A now available, but also an additional file C has been provided to us by Mr. Y. Tanaka[13]. File C consists of 1,077,322 customers of department stores in Kanto district, in which Tokyo is located, and is compiled from names written in Chinese characters, the ordinary expression of Japanese surnames. A difficulty in counting the total number of Japanese surnames is that Chinese symbols may be read phonetically several ways, and Japanese letters can be translated into several Chinese characters. Ordinary Japanese names contain two or three Chinese letters[9]. Comparison of two different systems of compilation for

Table 1

Random Isonymy (I) and Surname Distribution in Japan*

Source		*File A*	*File C*	*Taki A*	*Taki B*	*Taki C*	*Taki D*
100I		0.238	0.222	0.177	0.232	0.229	0.210
50I	K	8	9	15	10	15	11
	N(%)	65,991 (9.50)	105,086 (9.82)	6,098 (11.36)	6,493 (10.39)	440 (12.85)	420 (10.39)
60I	K	13	13	24	15	24	21
	N (%)	91,904 (13.23)	134,658 (12.59)	8,192 (15.26)	8,432 (13.49)	577 (16.85)	578 (14.94)
70I	K	22	24	40	27	46	48
	N (%)	120,959 (17.42)	187,100 (17.49)	10,962 (20.42)	11,761 (18.82)	825 (24.09)	888 (22.96)
80I	K	47	52	75	57	104	117
	N (%)	174,419 (25.11)	270,547 (25.29)	15,167 (28.25)	16,940 (27.11)	1,224 (35.75)	1,368 (35.37)
90I	K	113	125	177	133	299	340
	N (%)	259,462 (37.36)	403,618 (37.73)	22,441 (41.80)	24,923 (39.88)	1,950 (56.95)	2,172 (56.15)
95I	K	216	238	309	246	511	600
	N (%)	335,310 (48.28)	522,309 (48.83)	27,930 (52.03)	31,942 (51.11)	2,374 (69.33)	2,692 (69.60)
99I	K	687	717	1,142	768	—	—
	N (%)	472,612 (68.05)	728,535 (68.11)	39,973 (74.46)	44,890 (71.83)		
Total	K	25,550	30,635	7,827	8,020	1,561	1,776
	N	694,553	1,069,710	53,682	62,492	3,424	3,868
	unknown	21,262	3,806	—	—	—	—

* Names are ranked in decreasing frequency.

K = number of names. N = number of persons. I = random isonymy.

names allowed the examination of possible biases in the calculation of random isonymy in Japan. In addition four name lists, taken from Takiguchi [14], were examined, as also in the later part of the analysis were three previously published name lists [7,8].

Results

Random isonymy. Calculated random isonyms (i.e. the probability that two names chosen at random are the same) are 0.00238 for file A and 0.00222 for file C. An estimate (0.00184) from the most frequent sixty names from file A, which accounted for only 28 per cent of all persons, was 77 per cent of the total random isonymy calculated from the total list. This led us to study the relation between the proportion of random isonymy and that of the total population. Names were ranked in decreasing frequency, i.e. by the number of persons who have each name. From the known random isonymy of the total population (I) which could be examined, the number of surnames (K) that accounted for a given proportion (p) of the random isonymy of the total population and the sum of the persons (N) they represent were calculated. This was done for the populations represented in file A and file C as well as for four surname rankings of Takiguchi [14] (Taki A, Taki B, Taki C and Taki D). The results are given in Table 1. Because of the small sample sizes in Taki C and Taki D, tabulated numbers, N and K, do not always exactly correspond to the p-values of random isonymy, but are given the nearest numbers.

The large files A and C show that ten per cent of persons with the most frequent names (approximately) account for 50 per cent of random isonymy (I); 26 per cent of persons with *ca* 50 names account for 80 per cent; and 50 per cent with *ca* 140 names for about 95 per cent. In this they are supported by the Taki A and Taki B lists though less closely at the higher percentages, while the two small Taki lists vary quite markedly. Figures 1 and 2 depict for files A and C, respectively, the cumulative proportion of total random isonymy and the cumulative frequency of persons against the number of surnames ranked in decreasing order. The agreement between the two files is remarkable.

Surname distribution. Looking at the ranked name lists in the reverse direction, the least frequent surnames ranking up to $m_o = 499$, yielded the relation:

$$\log_{10} k = a - b(\log_{10} m) \quad \text{for } 1 < m < m_o \qquad (1)$$

where k is the number of different names and m indicates the number of persons for a particular surname. For the remaining surnames, which are most frequent names and each occurs in many individuals, obviously $k = 1$ or

$$\log_{10} k = 0. \qquad \text{for } m_o < m \qquad (2)$$

Figures 3 and 4 show the number of surnames by the number of persons carrying them, plotted on log-log scales for file A and file C respectively. There appear to be similar patterns shown by plots of the surname distributions in Takiguchi's four lists, as well as those from two small cities, Ohdate and Mine [7], and from a farm village, Uto [8], with two straight lines meeting each

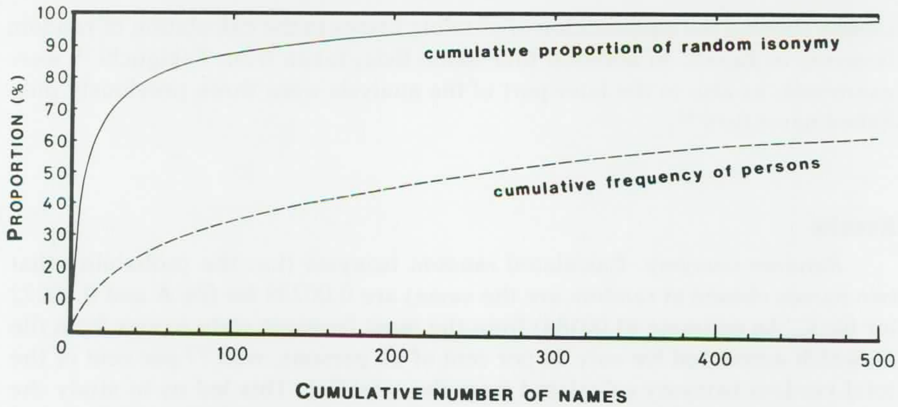

Figure 1. Cumulative distributions of random isonymy and persons in Japan (file A).

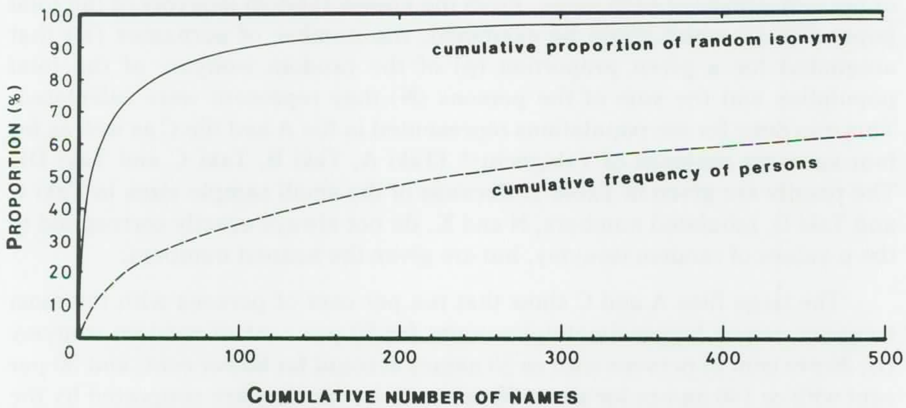

Figure 2. Cumulative distributions of random isonymy and persons in the Kanto district, Japan (file C).

other at an obtuse angle. The last three sets of data include surname distributions of three antecedent generations, parental (P), grandparental (GP) and great-grandparental (GGP) generations. All materials are consistent with the relations (1) and (2) although populations are different from each other in the threshold 'm_o'.

Regression analysis showed the slope parameters 'b' to be between 1.5-1.7 in all files except in two of Takiguchi's lists (Taki C and D) and those from Mine city, which yielded a slightly steeper slope (b = 2.0-2.4). Obviously, the threshold 'm_o' is chosen somewhat arbitrarily by inspection of the graph, but nonetheless the consistency of the slope is remarkable (Table 2). Examination of three antecedent generations in the local surname lists indicated clearly that the name distributions have been very stable indeed during the last three generations, or probably since 1875.

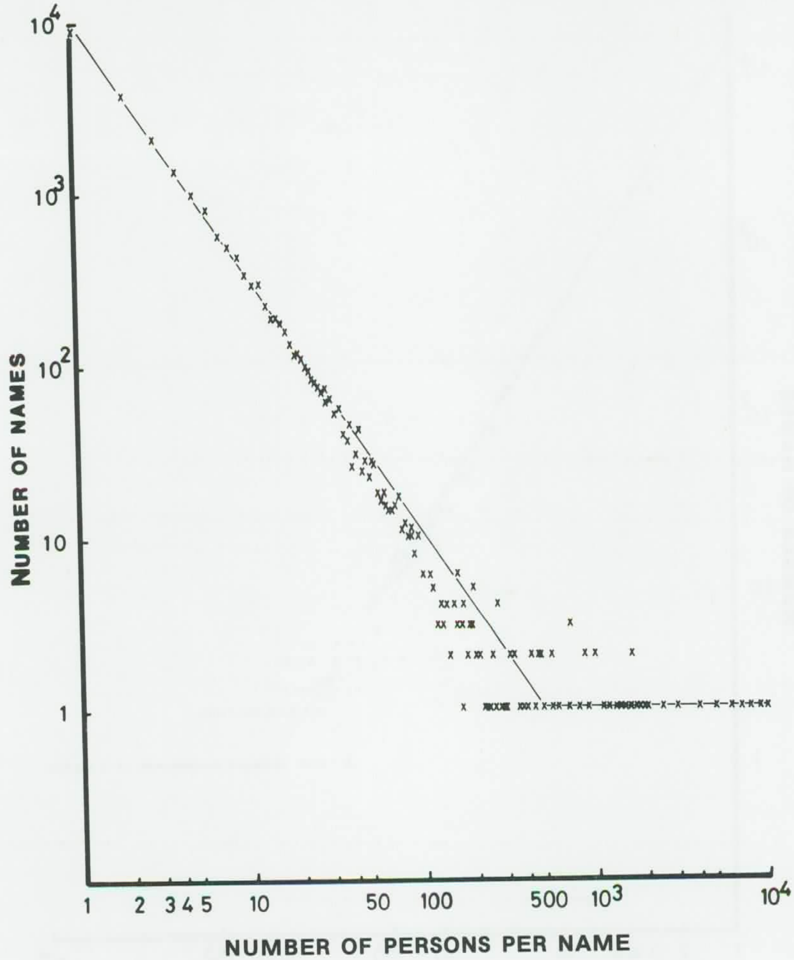

Figure 3. Distribution of surnames in Japan (file A).

Discussion

Random isonyms, 0.00238, 0.00222, 0.00177 and 0.00232, calculated from file A, file C, Taki A and Taki C, respectively, could be taken as representative of all Japan. The insurance file (file A) and the department store file (file C) might have contained individuals from the same family, but it is currently not possible to detect any bias this causes in calculating random isonymy. Yet, it is satisfying that the two files yielded very similar figures despite the fact that they were compiled in completely different ways. Taki A was from *Who's Who in Medicine* and Taki B was from *Who's Who in Japan* about forty years ago. Nonetheless, the four estimates from these different sources are remarkably consistent. Including an estimate from file B, previously calculated from the first sixty highly ranked names[9], a simple average of these five estimates becomes 0.0022; this is taken as representing random isonymy in all Japan. That is, on average two random pairs per thousand throughout Japan share the same surname.

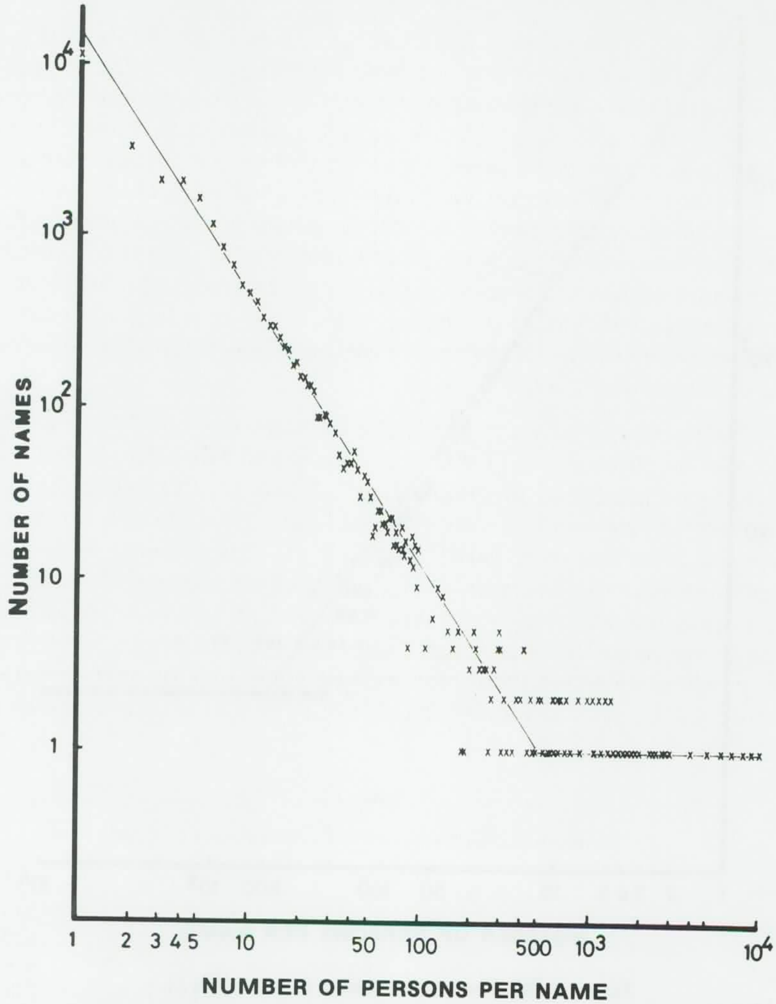

Figure 4. Distribution of surnames in the Kanto district, Japan (file C).

Studies of file A and file C indicate that the first half of the population whose surnames are ranked in decreasing order by the number of persons with the same name is sufficient for the calculation of random isonymy for 95 per cent of the total random isonymy, and could be calculated from the first 48.28 and 48.83 per cent of persons (Table 1). This estimate comes from the first phase of what appears to be an historic distribution, with a point of rotation at a threshold. The second phase, relating to the least frequent names, may be used for the study of human migration or dynamics of surname transmission in populations. In files A and C the proportions of population from the least frequent surnames to the threshold were respectively 49.65 and 42.82 per cent. The extreme of the two phases is seen in Takiguchi's lists as well as in three local lists. The threshold m_o, however, may disappear when the sample is small, as occurred in the surname distribution from the parental generation in Mine city (Tables 2 and 3).

This biphasic distribution of Japanese surnames might be the reflection of naming processes at the time of the Meiji Restoration. In 1875 the government proclaimed that a name was mandatory for everyone. Before that time, only the governing classes could use names officially. To meet the new requirements, there were two main ways to provide a name: the creation of a unique surname, or the resurrection of the name of an ancient ancestor. The former would increase the number of different names and would act as a force of diversification, while the latter would tend in the opposite direction. Counterbalance of the two forces might have generated the relation (1) in the least frequent name distribution. The remainder, the most frequent names of persons, would be largely polyphyletic, each name having multiple origins at the time of the naming process.

The differences in the slope parameter between most of the series (b = 1.5-1.7), and the three others (Taki C and Taki D and Mine city, b = 2.0-2.4) may perhaps be attributed to sample size[14]. However, this seems unlikely, for data from Ohdate and Uto showed b = 1.5-1.7 even though the sample sizes are much smaller than Taki C and Taki D. This disparity is probably due to either human movement or heterogeneity of the material. To Mine city, with its mining industry, many workers come and go and so contribute rare surnames to the name list. Taki C and Taki D are from directories of university alumni, in whose selection there may well be some bias by surname distribution from all Japan. The larger the value of b, the more unique names are.

Table 2

Distribution of Surnames with Least Frequency in Japan

Source	a	b	m_o	K^1	N^1(%)	K^1/N^1
file A	3.16	1.52	499	25,307	344,866 (49.65)	0.07
file C	3.29	1.56	499	10,725	458,518 (42.82)	0.02
Taki A	3.06	1.76	107	3,551	39,522 (73.42)	0.09
Taki B	2.83	1.54	178	3,798	48,693 (77.63)	0.08
Taki C	3.13	2.21	21	1,041	3,102 (89.98)	0.34
Taki D	3.31	2.41	20	1,164	3,445 (88.55)	0.34
Ohdate						
Generation I	2.32	1.72	29	702	2,386 (51.47)	0.29
,, II	2.04	1.50	33	550	1,921 (68.63)	0.29
,, III	1.93	1.51	26	378	1,133 (78.29)	0.33
Mine						
Generation I	2.76	2.03	23	1,207	2,832 (83.88)	0.43
,, II	2.82	2.07	18	854	1,775 (89.69)	0.48
,, III	2.47	2.17	16	543	1,005 (100.00)	0.54
Uto						
Generation I	2.29	1.65	26	545	1,944 (66.22)	0.28
,, II	2.26	1.60	24	477	1,690 (74.84)	0.28
,, III	2.28	1.70	16	324	960 (77.00)	0.34

a and b are regression coefficients in the equation (1).
m_o = threshold. N^1 = number of persons with least frequent names.
K^1 = number of names among N^1 persons.
I = Great-grandparental, II = Grandparental, III = Parental generation.

Table 3

Random Isonymy (I) and Surname Distributions at Three Locales in Japan*

Source		Ohdate III	Ohdate II	Ohdate I	Mine III	Mine II	Mine I	Uto III	Uto II	Uto I
100I		1.10	1.20	1.20	0.40	0.36	0.34	0.89	0.77	0.79
50I	K	5	4	3	13	15	16	8	9	9
	N (%)	224 (15.8)	414 (15.1)	587 (12.8)	150 (14.9)	307 (15.7)	517 (15.4)	229 (18.7)	412 (18.6)	588 (18.6)
60I	K	8	6	5	27	21	22	12	13	13
	N (%)	307 (21.7)	537 (19.5)	811 (17.7)	242 (24.1)	385 (19.7)	644 (19.2)	297 (24.2)	536 (24.1)	697 (24.1)
70I	K	12	10	9	42	34	33	21	21	19
	N (%)	392 (27.7)	714 (26.0)	1,124 (24.6)	317 (31.5)	528 (27.0)	847 (25.3)	415 (33.9)	713 (32.0)	896 (31.0)
80I	K	23	20	17	60	64	53	37	35	32
	N (%)	585 (41.4)	1,051 (38.2)	1,607 (35.1)	389 (38.7)	744 (38.0)	1,128 (33.7)	569 (46.4)	944 (42.4)	1,192 (41.2)
90I	K	36	31	30	162	117	122	62	65	60
	N (%)	747 (52.8)	1,338 (48.6)	2,159 (47.2)	624 (61.1)	978 (49.9)	1,637 (48.9)	743 (60.0)	1,271 (57.1)	1,612 (55.7)
95I	K	52	56	52	—	301	283	92	100	99
	N (%)	869 (61.5)	1,675 (60.9)	2,680 (58.5)		1,399 (71.4)	2,211 (66.0)	871 (71.0)	1,525 (68.5)	1.981 (68.5)
99I	K	171	175	147	—	—	471	179	185	205
	N (%)	1,200 (84.9)	2,279 (82.8)	3,591 (78.4)			2,587 (77.3)	1,071 (87.4)	1,847 (83.0)	2,439 (84.3)
total	K	385	304	733	543	861	1,233	334	490	556
	N	1,414	2,751	4,519	1,005	1,959	3,349	1,226	2,226	2,892
	unknown	14	105	1,132	31	113	795	16	258	062

* Names are ranked in decreasing frequency.

K = Number of names. N = Number of persons. I = Random isonymy.

I = Great-grandparental, II = Grandparental, III = Parental generation.

Table 4

A Trend in Frequency of Blood-Related Marriages in Japan[15]

Period	p %
1912 - 1925	22.5
1926 - 1930	16.1
1931 - 1935	13.1
1936 - 1940	13.8
1941 - 1945	12.3
1946 - 1950	14.4
1951 - 1955	10.5
1956 - 1960	6.97
1961 - 1965	4.74
1966 - 1968	2.63
1969 - 1972	3.20

p = proportion of blood related marriages

Active immigration and emigration would increase the number of different surnames and, as a result, there would be more persons with unique surnames in the population. Therefore human movements at the local level might be one of the causes of steeper slopes.

Human movement may also be assessed from the average number of names per individual, in those persons with the least frequent names; by K^1/N^1, where N^1 is the number of persons with the least frequent names and K^1 is the number of names among them. As shown in Table 2, this ratio seems to be a good indicator of human movements when comparisons were made between populations. The three rankings, file A, Taki A and Taki B, yielded 0.07m 0.09 and 0.08 respectively, might represent all Japan; the smaller value of 0.02 from file C is perhaps due to limitation of the area studied to the Kanto district. Mine city is known as a centre of active immigration, and Taki C and Taki D are similarly heterogenous, being directories of university alumni. Three localities, Ohdate, Mine and Uto, as shown by the different generation lists, were stable, but movement seems to have increased in the latest generation, consistent with a recent trend of increasing human movement in Japan.

Finally, calculated random isonymy, 0.0022, certainly does not necessarily reflect genetic inbreeding, but is to be regarded as a basic variable for studying the evolution of surnames in Japan since 1875. Secular decrease in the frequency of consanguineous marriages in Japan has occurred (Table 4), a trend already revealed by isonymy studies[7,8]; however, those studies attributed the decreasing tendency to reduction in nonrandom inbreeding. Therefore, the comparison of isonyms at different times or localities would be of little use in attempts at genetic interpretation.

Acknowledgements

We are indebted to Mr. Yasuhito Tanaka who generously provided us with the complete listings of surnames, file A and file C. We also thank Mr. Kenjiro Fukuhisa and Mr. Sadao Shibata for assistance in running computer programmes on an ACOS-700S installed at the National Institute of Radiological Sciences. The expert secretarial assistance of Mrs. Hiroko Ito is also deeply appreciated.

References

1. Crow, J. F. and Mange, A. P. (1965). Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugenics Quarterly*, 12, 199-203.
2. Darwin, G. (1875). Marriages between first cousins in England and their effects. *Journal of the Royal Statistical Society* (*Series B*), 38, 153-184.
3. Kamizaki, M. (1954). Konin dosei-ritsu (Frequency of isonymous marriages). *Seibutsu Tokeigaku Zassi*, 2, 292-298. (In Japanese).
4. Ellis, W. S. and Starmer, W. T. (1978). Inbreeding as measured by isonymy, pedigrees and population size in Torbel, Switzerland. *American Journal of Human Genetics*, 30, 366-376.
5. Kuchemann, C. F., Lasker, G. W. and Smith, D. I. (1979). Historical changes in the coefficient of relationship by isonymy among the population of the Otmoor villages. *Human Biology*, 51, 63-77.
6. Roberts, D. F. and Rawling, C. P. (1974). Secular trends in genetic structure: an isonymic analysis of Northumberland parish records. *Annuals of Human Biology*, 1, 395-410.
7. Yasuda, N. and Furusho, T. (1971a). Random and nonrandom inbreeding revealed from isonymy study. I. Small cities of Japan. *American Journal of Human Genetics*, 23, 303-316.
8. Yasuda, N. and Furusho, T. (1971b). Random and nonrandom inbreeding revealed from isonymy study. II. A group of farm villages in Japan. *Japanese Journal of Human Genetics*, 15, 231-240.
9. Yasuda, N. (1983). Studies of isonymy and inbreeding in Japan. *Human Biology*, 55, 263-276.
10. Tanaka, Y. (1972). Nippon-jin no sei to na no tokei (Statistics on Japanese first and family names). *Gengo Saikatsu*, 254, 72-79. (In Japanese).
11. Tanaka, Y. (1973). Nippon-jin no sei to na ni tsukawareru kanji (Chinese letters in use of first and family names in Japan). *Gengo Seikatsu*, 267, 90-96. (In Japanese).
12. Tanaka, Y. (1975). Kana to Kanji conversion software for person's names. *Joho Shori*, 16, 230-238. (In Japanese).
13. Nippon Univac. (1975). *Japanese Surnames*. (Unpublished material).
14. Takiguchi, R. (1943). Hojin sei-si no tokeigaku-teki kenkyu (A statistical study of Japanese surnames). *Minzoky Eisei*, 11, 241-245. (In Japanese).
15. Jinko Mondai Singikai (1974). *Nippon-jinko no doko* (A Population Trend in Japan). Tokyo: The Printing Bureau, Department of Treasury, P. 279. (In Japanese).