Family Formation and the Frequency of Various Kinship Relationships

Leo A. Goodman

The University of Chicago

Nathan Keyfitz and Thomas W. Pullum

Harvard University

Received January 19, 1970

A set of age-specific rates of birth and death implies expected numbers of kin. An individual girl or woman chosen at random out of a population whose birth and death rates are specified can be expected to have a certain number of older sisters, younger sisters, nieces, cousins; expressions for these values are provided for both total kin and kin who are still living. Included also are the probabilities of living mother, grandmother, and great grandmother for girls and women of various ages. The methods are applicable to the size of the nuclear and the extended family. All formulas have been programmed and specimen numerical values are given.

If two communities have equal death rates, girls in the community of higher fertility will expect to have more sisters and aunts. On the other hand, if birth rates are equal, girls in the community of lower mortality are more likely to have a living mother or grandmother. We shall show here how to calculate expected frequencies and probabilities for these and a variety of other kinship relationships, as well as providing expressions for the mean size of the nuclear and extended families.

The results will be based on a given and fixed regime of mortality and fertility, represented by

(a) the chance that a newborn girl survives to age x, denoted l_x. This is the same as the number-surviving column of the life table. when the radix l_0 is set equal to 1. If the age-specific death rate in the age interval t to $t + dt$ is μ_t (per unit of time), then $l_x = \exp(-\int_0^x \mu_t \, dt)$. Thus, being given the age-specific death rate μ_t (for $0 \leqslant t < \infty$) is equivalent to being given the survival probability l_x (for $0 \leqslant x < \infty$).

(b) the chance that a woman of age x has a child in the next dx years (i.e., in the age interval from x to $x + dx$), denoted $m_x dx$. We will call the lowest and

1

highest ages of reproduction α and β, respectively. That is, $m_x = 0$ for $x < \alpha$ and $x > \beta$, with positive fertility restricted within the range $\alpha \leqslant x \leqslant \beta$. (We assume throughout that m_x is an integrable function of x. If we assume, in addition, that m_x is a continuous function, then we should set $m_x = 0$ for $x \leqslant \alpha$ and $x \geqslant \beta$, with $m_x > 0$ for $\alpha < x < \beta$.)

The mortality and fertility rates will be taken as fixed. We shall see what conclusions about family size and kinship relationships follow from a given regime of mortality and fertility, supposing the regime to act independently for the several individuals involved. We shall calculate how plentiful sisters, aunts, grandmothers, etc., will be under different regimes. The technique that will be developed is capable of extension to estimating the expected number of sisters, etc., in a changing regime, but we will not enter here into that problem.

Our results provide a model for the study of actual communities. If the regime of fertility and mortality is known (and if various assumptions pertaining to the regime are true), then no further empirical data are needed to ascertain the expected number of nieces a woman of 45 has or the probability that her mother is alive. In a community of high mortality orphanhood will be common, and our calculations show just how common, given the birth and death regime (and given that the various assumptions pertaining to the regime are true).

If the regime is not known and data can be gathered on the mean number of nieces of women of 45 and other ages, on the proportion of instances in which they have a living mother, and similar matters, then our formulas may be seen as equations of which the unknowns are the age-specific rates of birth and death. We leave discussion of the conditions under which they would be solvable and the method of solution for another place. Suffice it to say here that our results can be applied to two opposite sorts of problems: Finding the implications of given l_x and m_x for the expected frequencies of the various kinship relationships and inferring the l_x and m_x from the observed frequencies of the various kinship relationships.

Our argument with some modification is applicable to males as well as to females, and only for concreteness of the exposition will females be repeatedly mentioned.

1. DIRECT DESCENDANTS

If a woman of age a is talking to us at time t, there is no question that she is alive, and she may be considered to have been exposed at each moment to the fertility of the population of which she is a random member. If that fertility is $m_x \, dx$ for the age interval x to $x + dx$, then the total expected number of children she has had is the integral

$$\int_\alpha^a m_x \, dx. \qquad (1.1.a)$$

We have integrated out x, the age of the mother at childbearing between α, the lowest fertile age, and a, the present age of the woman. If m_x is the rate of bearing of daughters, then the integral gives the expected number of daughters; for an age $a > \beta$, that is, for a woman past the age of reproduction, the integral will be the quantity referred to in demography as the gross reproduction rate (which applies to the women who survive past the reproductive period). If the m_x are fertility rates for children of both sexes and $a > \beta$, then the integral is known as the total fertility rate. (In the following calculations, m_x is taken as the rate of bearing daughters.)

The probability that the daughter, born when her mother was aged x, is alive when her mother is aged a depends on survival through $a - x$ years, a probability of l_{a-x}. Hence, the probability that a daughter will be born to a woman between age x and $x + dx$ *and* that the daughter will survive to time t when the mother is a years old is $m_x \, dx \, l_{a-x}$. (We assume throughout that the probability of survival through $a - x$ years is l_{a-x}, that it is independent of the fact that the mother was still alive at age a, and that it is independent of other factors as well.) Thus, we obtain the following formula for the expected number of living daughters that a woman of age a $(a > \alpha)$ will have at time t:

$$\int_\alpha^a l_{a-x} m_x \, dx. \tag{1.1.b}$$

Brass (1953) derived (1.1.b) and used it for estimation of child mortality.

Next, we go down two generations and seek the expected number of grand-daughters alive at time t, when the woman is aged a. While the woman was aged x to $x + dx$ she would be expected to have $m_x \, dx$ daughters. Consider one of these. The chance that the daughter will live to age y and then have a daughter is $l_y m_y dy$. Then the expected number of granddaughters through the daughter born at age x of the original woman up to time t must be $\int_\alpha^{a-x} l_y m_y \, dy$. Since $m_x \, dx$ is the expected number of daughters born to the original woman at age x to $x + dx$, we obtain the following formula for the expected number of granddaughters born up to time t:

$$\int_\alpha^a \left[\int_\alpha^{a-x} l_y m_y \, dy \right] m_x \, dx. \tag{1.2.a}$$

To calculate the expected number of granddaughters who are still alive at time t, we need to enter a survivorship probability for the time from birth to time t, that is, over $a - x - y$ years from birth. This is obtained as the difference between the a years of the grandmother's present age and the $x + y$ years that passed from the grandmother's birth to the birth of her granddaughter. The

inner integrand of (1.2.a) must be multiplied by the probability l_{a-x-y} to provide the expected number of granddaughters alive at time t:

$$\int_\alpha^a \left[\int_\alpha^{a-x} l_y m_y l_{a-x-y} \, dy \right] m_x \, dx. \tag{1.2.b}$$

The argument can be carried through further generations. The expected number of births of daughters at age x to $x + dx$ is still $m_x \, dx$; the expected number of these themselves bearing daughters at age y is $l_y m_y \, dy$; the expected number of these (i.e., daughters' daughters) in turn bearing daughters at age z is $l_z m_z \, dz$. For particular x, y and z, the expected number of great-grand-daughters is

$$l_z m_z \, dz \, l_y m_y \, dy \, m_x \, dx.$$

Now the relevant range of age x is from α to a, the relevant range of age y is from α to $a - x$, and the relevant range of age z is from α to $a - x - y$. Thus, we obtain the following formula for the expected number of great-granddaughters born up to time t:

$$\int_\alpha^a \left[\int_\alpha^{a-x} \left\{ \int_\alpha^{a-x-y} l_z m_z \, dz \right\} l_y m_y \, dy \right] m_x \, dx. \tag{1.3.a}$$

For the expected number of great-granddaughters that survive to time t, we enter a factor $l_{a-x-y-z}$ in the innermost integral, and this would give (1.3.b). By rules which should now be discernible, we can obtain formulas pertaining to further generations, which we would denote as (1.4.a), (1.4.b), (1.5.a), etc.

Using a computer program described in the Appendix, we have estimated the frequencies of direct descendants using stable population assumptions for the United States (1967), Venezuela (1965), and Madagascar (1966). The input data for these countries are drawn from Keyfitz and Flieger (1971, pp. 360, 376, 312). As Table I indicates, these data sets represent three typical and contrasting

TABLE I

Fertility and Mortality Measures for Three Selected Countries

	United States 1967	Venezuela 1965	Madagascar 1966
Gross reproduction rate	1.26	3.13	3.29
Expectation of life at birth (females only)	74.22	67.70	38.48

patterns of fertility and mortality. The United States shows low mortality and low fertility; Venezuela, low mortality but high fertility; Madagascar, high mortality as well as high fertility.

We present in Tables IIa, IIb, and IIc the results for formulas 1.1.a and b, 1.2.a and b, 1.3.a and b, respectively. These give the expected numbers of female descendants ever born or still alive, by age of woman. Very little fertility occurs before age 15, so the first positive entries for granddaughters are at age 30. Similarly, since most childbearing has ended by age 45, the eventual number of granddaughters ever born should stabilize at about age 90 as the product of the the gross reproduction rate (since the woman herself has survived the childbearing period) times the net reproduction rate (since her daughters may not have survived). For the three countries under discussion, these asymptotic numbers are 1.5246, 8.8579, and 6.1852, respectively. Our tables extend only to age 85 because we lack survival rates beyond that age.

Note that Venezuela and Madagascar much exceed the United States in terms

TABLE IIa

Expected Number of Daughters Ever Born and Still Alive, by Age of Woman,
for Three Selected Countries

Age	Daughters ever born			Daughters still alive		
	United States 1967	Venezuela 1965	Madagascar 1966	United States 1967	Venezuela 1965	Madagascar 1966
15	0.0021	0.0046	0.0076	0.0021	0.0043	0.0056
20	0.1678	0.3163	0.3533	0.1639	0.2946	0.2607
25	0.5923	1.0791	1.0906	0.5784	1.0038	0.7928
30	0.9401	1.8729	1.7948	0.9172	1.7387	1.2765
35	1.1336	2.4615	2.4912	1.1043	2.2786	1.7301
40	1.2276	2.9178	2.9411	1.1932	2.6914	1.9710
45	1.2534	3.0812	3.1795	1.2143	2.8269	2.0324
50	1.2551	3.1220	3.2587	1.2105	2.8430	1.9600
55	1.2551	3.1267	3.2888	1.2028	2.8184	1.8441
60	1.2551	3.1267	3.2888	1.1916	2.7801	1.7048
65	1.2551	3.1267	3.2888	1.1752	2.7294	1.5693
70	1.2551	3.1267	3.2888	1.1511	2.6617	1.4386
75	1.2551	3.1267	3.2888	1.1166	2.5670	1.3084
80	1.2551	3.1267	3.2888	1.0672	2.4334	1.1681
85	1.2551	3.1267	3.2888	0.9964	2.2580	1.0073

TABLE IIb

Expected Number of Granddaughters Ever Born and Still Alive, by Age of Woman,
for Three Selected Countries

Age	Granddaughters ever born			Granddaughters still alive		
	United States 1967	Venezuela 1965	Madagascar 1966	United States 1967	Venezuela 1965	Madagascar 1966
30	0.0004	0.0014	0.0019	0.0003	0.0013	0.0014
35	0.0149	0.0508	0.0477	0.0146	0.0473	0.0352
40	0.0982	0.3211	0.2618	0.0959	0.2989	0.1916
45	0.3109	1.0395	0.7609	0.3036	0.9667	0.5513
50	0.6288	2.2586	1.5553	0.6136	2.0979	1.1122
55	0.9558	3.8064	2.5604	0.9319	3.5296	1.8011
60	1.2161	5.4152	3.6120	1.1839	5.0102	2.4868
65	1.3836	6.8043	4.5384	1.3442	6.2762	3.0376
70	1.4704	7.7946	5.2402	1.4242	7.1587	3.3813
75	1.5066	8.3893	5.7003	1.4528	7.6582	3.5123
80	1.5188	8.6890	5.9613	1.4550	7.8645	3.4723
85	1.5218	8.8093	6.0886	1.4441	7.8799	3.3203

TABLE IIc

Expected Number of Great-Granddaughters Ever Born and Still Alive,
by Age of Woman, for Three Selected Countries

Age	Great-granddaughters ever born			Great-granddaughters still alive		
	United States 1967	Venezuela 1965	Madagascar 1966	United States 1967	Venezuela 1965	Madagascar 1966
45	0.0000	0.0003	0.0004	0.0000	0.0003	0.0003
50	0.0014	0.0088	0.0070	0.0014	0.0082	0.0052
55	0.0128	0.0752	0.0508	0.0125	0.0700	0.0372
60	0.0599	0.3522	0.2083	0.0585	0.3277	0.1518
65	0.1821	1.1122	0.5933	0.1778	1.0342	0.4291
70	0.4035	2.6477	1.3126	0.3938	2.4597	0.9400
75	0.7054	5.1078	2.4117	0.6879	4.7395	1.7064
80	1.0333	8.3875	3.8382	1.0068	7.7709	2.6751
85	1.3282	12.1272	5.4468	1.2923	11.2130	3.7245

of descendants ever born because of their high fertility rates, but the very high mortality of Madagascar sharply reduces the number of survivors. For example, a woman of 75 under the latter regime can expect to have lost 40% of her grandchildren, but in the U.S. a woman of 75 will have lost fewer than 4%. Similar manifestations of the basic birth and death schedules will be seen to dominate all of our results.

2. Direct Progenitors

A different approach is required for the probability that a specified one of a girl's ancestors is alive. We see before us the girl of age a, taken at random from all girls aged a in a population, and seek the probability that her mother is alive.

If the mother was of age x at the time she gave birth to the girl now of age a, then the probability that the mother is alive now is l_{a+x}/l_x. Let $M_1(a)$ denote the probability that the mother is alive now, and let $W(x \mid t - a)$ denote the age distribution (at time $t - a$) of the women who gave birth to a daughter at time $t - a$. Then we find that

$$M_1(a) = \int_\alpha^\beta (l_{a+x}/l_x)\, W(x \mid t - a)\, dx.$$

To justify the use of the above formula, it is assumed that (a) the probability l_{a+x}/l_x applies to each woman of age x who gave birth to a daughter at time $t - a$, and (b) the conditional age distribution (at time $t - a$) of the women who gave birth to a daughter at time $t - a$, given that the daughter is alive at time t, is equal to $W(x \mid t - a)$.

If the population under consideration is stable, then $W(x \mid t - a)$ will not depend upon $t - a$, and in this case we can replace $W(x \mid t - a)$ by

$$W(x) = l_x m_x e^{-rx},$$

where r is the intrinsic rate of natural increase. Thus, in this case, the probability $M_1(a)$ that the mother is alive now is (Lotka, 1931)

$$M_1(a) = \int_\alpha^\beta (l_{a+x}/l_x)\, W(x)\, dx = \int_\alpha^\beta l_{a+x} m_x e^{-rx}\, dx. \tag{2.1}$$

Henry (1960) used (2.1) to estimate mortality in a population lacking registration.

If the population under consideration is not stable, then the formulas for $W(x \mid t - a)$ and $M_1(a)$ will differ from the ones in the preceding paragraph. In this case, the observed age distribution (at time $t - a$) of the women who gave birth to a daughter at time $t - a$ could be used for $W(x \mid t - a)$ in the

formula for $M_1(a)$ which was presented in the paragraph before the preceding one; or, if the observed age distribution is not known, then other kinds of demographic information might be used to estimate this age distribution. For example, letting $B(t)$ denote the number of girls born at time t, we might estimate $W(x \mid t - a)$ by

$$W(x \mid t - a) = B(t - a - x) \, l_x m_x \Big/ \int_\alpha^\beta B(t - a - x) \, l_x m_x \, dx.$$

Note that the denominator in the above expression will be equal to $B(t - a)$ if (a) we think of the B's as expected numbers, and (b) the regimes described by the l_x and m_x apply to all girls born at time $t - a - x$ (for $\alpha < x < \beta$). Of course, we are concerned here with the case where $t \geqslant a + \beta$.

The random choice of girl and the unvarying regime are essential to application of the argument. If the girl aged a were of poor health, and if the health of mothers and daughters were correlated, then she would be less likely than this average to have a living mother. Note that we have also supposed that the several generations entering each expression are independently subject to the mortality and fertility schedules assumed. The population is taken as closed to migration.

The same technique can be carried to further generations, and we next use it to find the probability that a girl aged a, drawn at random from a population of fixed regime of mortality, has a living maternal grandmother.

If the mother was of age x at the time she gave birth to the girl now of age a, then the conditional probability that the maternal grandmother is alive now is equal to the probability $M_1(a + x)$ that the mother of a woman of age $a + x$ (viz., the mother of the girl's mother) is alive now. For the girl of age a, let $M_2(a)$ denote the probability that her maternal grandmother is alive now. Then we find that

$$M_2(a) = \int_\alpha^\beta M_1(a + x) \, W(x \mid t - a) \, dx,$$

where $W(x \mid t - a)$ is the age distribution defined earlier in this section.

Similarly, for a girl of age a, let $M_3(a)$ and $M_4(a)$ denote the probabilities that her great-grandmother and great-great-grandmother, respectively, are living now (where only the direct maternal antecedents are considered in each generation). Then we find that

$$M_3(a) = \int_\alpha^\beta M_2(a + x) \, W(x \mid t - a) \, dx,$$

$$M_4(a) = \int_\alpha^\beta M_3(a + x) \, W(x \mid t - a) \, dx.$$

Similar considerations lead to the general recurrence relation

$$M_i(a) = \int_\alpha^\beta M_{i-1}(a + x)\, W(x \mid t - a)\, dx,$$

where $M_i(a)$ is the probability of survival to time t of the ith antecedent generation, or the great^{i-2}-grandmother (where only the direct maternal antecedents are considered in each generation).

If the population under consideration is stable, then $W(x \mid t - a)$ can be replaced by the quantity $W(x)$ defined earlier in this section, and the above formulas can be simplified as follows.

$$M_2(a) = \int_\alpha^\beta M_1(a + x)\, W(x)\, dx, \tag{2.2}$$

$$M_3(a) = \int_\alpha^\beta M_2(a + x)\, W(x)\, dx, \tag{2.3}$$

$$M_4(a) = \int_\alpha^\beta M_3(a + x)\, W(x)\, dx. \tag{2.4}$$

Tables IIIa and IIIb give numerical probabilities that the preceding three maternal progenitors are still alive in our illustrative stable populations. At exact age 0, the moment of birth, the probability that the mother is alive is necessarily unity. This probability falls off quite slowly in the low-mortality situations of the United States and Venezuela, so that the .50 level is reached at ages 53 and 50, respectively, in those two countries, but at age 34 in Madagascar. The .50 level for the maternal grandmother is reached almost exactly one-generation length earlier, at ages 27, 22, and 6, respectively. A woman of 25 is approximately three times as likely to have her maternal grandmother alive if she is in the United States rather than in Madagascar.

Applying formula 2.3, we have computed that at birth the chances are .5069, .3974, and .1636, respectively, that her great-grandmother (on the female line) is alive. Since each person has four great-grandmothers, if we trace back through parents and grandparents of both sexes, we can estimate the chance that *one or more* great grandmothers is alive (at the woman's birth) as $1 - (1 - .5069)^4 = .94$ in the United States and only $1 - (1 - .1636)^4 = .51$ in Madagascar. Such an estimate is inexact in that it assumes the same net maternity (or paternity) function for males as for females. However, one can infer that the high birth and death rates of traditional societies will restrict the formation of linearly extended households. Regardless of cultural norms favoring such structures, it is clear that in a situation of high mortality the range of generations will typically be small.

TABLE IIIa

Probabilities that Mother and Grandmother are Still Alive,
by Age of Woman, for Three Selected Countries

	Chance mother is alive			Chance grandmother is alive		
Age	United States 1967	Venezuela 1965	Madagascar 1966	United States 1967	Venezuela 1965	Madagascar 1966
0	1.0000	1.0000	1.0000	0.9193	0.8673	0.5999
5	0.9946	0.9895	0.9256	0.8790	0.8125	0.5190
10	0.9870	0.9756	0.8523	0.8232	0.7424	0.4329
15	0.9758	0.9569	0.7817	0.7485	0.6557	0.3438
20	0.9594	0.9317	0.7126	0.6532	0.5532	0.2564
25	0.9356	0.8971	0.6421	0.5385	0.4383	0.1773
30	0.9014	0.8502	0.5660	0.4108	0.3186	0.1121
35	0.8527	0.7879	0.4806	0.2815	0.2055	0.0639
40	0.7846	0.7073	0.3855	0.1652	0.1120	0.0321
45	0.6931	0.6083	0.2859	0.0756	0.0478	0.0136
50	0.5769	0.4921	0.1922	0.0230	0.0141	0.0044
55	0.4393	0.3591	0.1160	0.0035	0.0022	0.0008
60	0.2975	0.2226	0.0611	0.0001	0.0001	0.0000
65	0.1629	0.1029	0.0286	0.0000	0.0000	0.0000
70	0.0443	0.0259	0.0086	0.0000	0.0000	0.0000
75	0.0006	0.0005	0.0002	0.0000	0.0000	0.0000

TABLE IIIb

Probabilities that Great-Grandmother and Great-Great-Grandmother are Still Alive,
by Age of Woman, for Three Selected Countries

	Chance great-grandmother is alive			Chance great-great-grandmother is alive		
Age	United States 1967	Venezuela 1965	Madagascar 1966	United States 1967	Venezuela 1965	Madagascar 1966
0	0.5069	0.3974	0.1636	0.0493	0.0303	0.0089
5	0.3881	0.2902	0.1069	0.0205	0.0124	0.0037
10	0.2716	0.1931	0.0642	0.0068	0.0041	0.0013
15	0.1690	0.1140	0.0349	0.0017	0.0010	0.0003
20	0.0899	0.0577	0.0168	0.0003	0.0002	0.0001
25	0.0386	0.0239	0.0070	0.0000	0.0000	0.0000
30	0.0123	0.0076	0.0023	0.0000	0.0000	0.0000
35	0.0026	0.0016	0.0006	0.0000	0.0000	0.0000
40	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000

3. SISTERS

Whether a woman has younger sisters (in contrast to older sisters) depends partly on the survival of her mother past the woman's own birth. For this reason, we treat older and younger sisters separately. Our first step is to find how many older sisters a girl now of age a would be expected to have. Again, suppose the girl born when her mother was of age x. By that time her mother would be expected to have had $\int_\alpha^x m_y \, dy$ children. Then we find that the expected number of older sisters would be

$$S(a) = \int_\alpha^\beta \left[\int_\alpha^x m_y \, dy \right] W(x \mid t - a) \, dx,$$

where $W(x \mid t - a)$ was the age distribution defined in the preceding section. If the population under consideration is stable, then the above formula can be replaced by

$$\int_\alpha^\beta \left[\int_\alpha^x m_y \, dy \right] W(x) \, dx, \tag{3.1.a}$$

where $W(x)$ is the quantity defined also in the preceding section. Note that (3.1.a) does not depend upon the age a of the girl under consideration.

To find the number of older sisters now alive requires repetition of the same argument, except that we need a factor for their survival. The girl born when her mother is y years old ($y < x$) is $x - y$ years older than her sister now aged a, so the older sister must be $a + x - y$ years old at time t if she is alive. Hence, a survival factor l_{a+x-y} must appear in the inner integral:

$$\int_\alpha^\beta \left[\int_\alpha^x m_y l_{a+x-y} \, dy \right] W(x) \, dx. \tag{3.1.b}$$

Younger sisters require a slightly different arrangement of the algebra. The chance that the mother lives from age x (when she had the daughter aged a at time t who is our starting point) to age z ($z > x$) and then gives birth to a daughter in the age interval from z to $z + dz$ is $(l_z/l_x) m_z \, dz$. The expected number of these daughters for the a years of possible exposure is

$$\int_x^{a+x} (l_z/l_x) m_z \, dz = \int_0^a (l_{x+y}/l_x) m_{x+y} \, dy.$$

Thus, we find that the expected number of younger sisters would be

$$S'(a) = \int_\alpha^\beta \left[\int_0^a (l_{x+y}/l_x) m_{x+y} \, dy \right] W(x \mid t - a) \, dx,$$

and for the stable population, the above formula is replaced by

$$\int_\alpha^\beta \left[\int_0^a (l_{x+y}/l_x)\, m_{x+y}\, dy \right] W(x)\, dx. \tag{3.2.a}$$

A factor of l_{a-y} under the double integral will provide for younger sisters living to time t:

$$\int_\alpha^\beta \left[\int_0^a (l_{x+y}/l_x)\, m_{x+y} l_{a-y}\, dy \right] W(x)\, dx. \tag{3.2.b}$$

Note that the denominator l_x can cancel with the l_x in $W(x)$ to give an alternative form for (3.2.a) and (3.2.b).

We note that in the usual model of population growth, the birth rate m_x for a female of age x can depend upon the age of the female (i.e., the rate is age specific), but it does not depend upon any other characteristic of the female, and,

TABLE IV

Expected Number of Sisters ever Born and Still Alive,
by Age of Woman, for Three Selected Countries

	Sisters ever born			Sisters still alive		
Age	United States 1967	Venezuela 1965	Madagascar 1966	United States 1967	Venezuela 1965	Madagascar 1966
0	0.6103	1.3250	1.3804	0.5952	1.2281	0.9643
5	0.8860	1.9299	1.9484	0.8633	1.7852	1.3342
10	1.0837	2.4418	2.4147	1.0541	2.2518	1.6067
15	1.1902	2.7934	2.7351	1.1548	2.5648	1.7506
20	1.2351	2.9893	2.9172	1.1939	2.7275	1.7759
25	1.2485	3.0697	2.9999	1.2003	2.7761	1.7170
30	1.2508	3.0917	3.0296	1.1931	2.7617	1.6140
35	1.2509	3.0957	3.0378	1.1794	2.7186	1.4941
40	1.2509	3.0961	3.0393	1.1594	2.6558	1.3696
45	1.2509	3.0961	3.0393	1.1305	2.5704	1.2425
50	1.2509	3.0961	3.0393	1.0892	2.4556	1.1101
55	1.2509	3.0961	3.0393	1.0309	2.3040	0.9679
60	1.2509	3.0961	3.0393	0.9510	2.1094	0.8144
65	1.2509	3.0961	3.0393	0.8458	1.8680	0.6535
70	1.2509	3.0961	3.0393	0.7152	1.5801	0.4947
75	1.2509	3.0961	3.0393	0.5650	1.2545	0.3503
80	1.2509	3.0961	3.0393	0.4060	0.9147	0.2302
85	1.2509	3.0961	3.0393	0.2544	0.5980	0.1385

in particular, it does not depend upon whether the female of age x did or did not give birth during the preceding nine months. In other words, the period of pregnancy and postpartum amenorrhea is one of the phenomena that the usual model of population growth does not take into account. For the sake of simplicity, we have used the usual model of population growth throughout this paper, though we realize that a more complete model (which would take into account, among other things, the period of pregnancy) would lead to formulas that would differ somewhat from those presented here. For example, formulas 3.2.a and b would differ because a younger sister cannot be born within approximately a year of the woman's own birth. Related formulas in Sections 4, 5, and 6 would also differ.

Numerical results for this section appear in Table IV. We have combined the results for older and younger sisters and simply present the total. The number of sisters ever born is more then twice as great in Venezuela and Madagascar as in the United States. Ninety percent of a woman's eventual total is reached by age 12 in the United States, and by age 15 in both Venezuela and Madagascar, reflecting the greater variance in age at childbearing in the latter two countries.

The number of sisters still alive reaches its peak at ages 25–30 in the U.S. and Venezuela but at ages 15–20 in Madagascar. The number of sisters still alive is greater in Madagascar than in the United States until age 51, at which point the greater mortality effect dominates the greater fertility effect, so that by the highest age (85) the U.S. woman has about twice the number of living sisters.

4. NIECES

Again taking a randomly selected woman of age a, born when her mother was aged x, as our starting point, we turn now to the number of her nieces born to older sisters. To an older sister who was born when the mother was of age y, the expected number of daughters born would be $\int_\alpha^{a+x-y} l_z m_z \, dz$ (when $a+x-y>\alpha$). Thus, by introducing this factor into (3.1.a), we obtain the following formula for the expected number of nieces born to older sisters:

$$\int_\alpha^\beta \left[\int_\alpha^x \left\{ \int_\alpha^{a+x-y} l_z m_z \, dz \right\} m_y \, dy \right] W(x) \, dx. \tag{4.1.a}$$

If we are concerned with the expected number of these nieces who are alive at time t, the factor $l_{a+x-y-z}$ should be included in the innermost integral:

$$\int_\alpha^\beta \left[\int_\alpha^x \left\{ \int_\alpha^{a+x-y} l_z m_z l_{a+x-y-z} \, dz \right\} m_y \, dy \right] W(x) \, dx. \tag{4.1.b}$$

Replacing $W(x)$ by $W(x \mid t - a)$ provides for unstable populations in (4.1.a) and (4.1.b), as well as in (4.2.a) below.

What about nieces born to younger sisters of the girl of age a? To a younger sister who was born when the mother was of age w ($w > x$), the expected number of daughters born would be

$$\int_{\alpha}^{a-y} l_z m_z \, dz,$$

where $y = w - x$ is the difference in age between the girl whose age is a and her younger sister whose age is $a - y$. Thus, by introducing this factor into (3.2.a), we obtain the following formula for the expected number of nieces born to younger sisters:

$$\int_{\alpha}^{\beta} \left[\int_{0}^{a} \left\{ \int_{\alpha}^{a-y} l_z m_z \, dz \right\} (l_{x+y}/l_x) \, m_{x+y} \, dy \right] W(x) \, dx. \qquad (4.2.a)$$

Again, l_x in parentheses could be cancelled with the l_x contained in $W(x)$ to give a variant form. If we are concerned with the expected number of these nieces who are alive at time t, the factor l_{a-y-z} should be included in the innermost integral of (4.2.a) to produce (4.2.b).

Table V shows our numerical results for nieces, combining those born to older

TABLE V

Expected Number of Nieces Ever Born and Still Alive,
by Age of Woman, for Three Selected Countries

	Nieces ever born			Nieces still alive		
Age	United States 1967	Venezuela 1965	Madagascar 1966	United States 1967	Venezuela 1965	Madagascar 1966
0	0.0092	0.0536	0.0616	0.0090	0.0498	0.0446
5	0.0336	0.1712	0.1661	0.0328	0.1592	0.1198
10	0.0954	0.4434	0.3820	0.0931	0.4120	0.2735
15	0.2203	0.9659	0.7632	0.2150	0.8968	0.5417
20	0.4228	1.8105	1.3407	0.4124	1.6791	0.9415
25	0.6844	2.9637	2.0897	0.6671	2.7447	1.4468
30	0.9558	4.3023	2.9260	0.9306	3.9767	1.9885
35	1.1853	5.6378	3.7395	1.1522	5.1971	2.4802
40	1.3467	6.7923	4.4344	1.3058	6.2380	2.8500
45	1.4420	7.6580	4.9578	1.3932	6.9961	3.0622
50	1.4891	8.2197	5.3067	1.4310	7.4538	3.1206
55	1.5086	8.5342	5.5129	1.4382	7.6586	3.0562
60	1.5151	8.6847	5.6206	1.4277	7.6804	2.9086
65	1.5169	8.7451	5.6701	1.4048	7.5774	2.7114
70	1.5173	8.7648	5.6899	1.3699	7.3820	2.4864
75	1.5173	8.7700	5.6968	1.3200	7.1017	2.2433
80	1.5173	8.7710	5.6987	1.2504	6.7287	1.9848
85	1.5173	8.7712	5.6992	1.1568	6.2496	1.7121

and younger sisters. Because we have essentially convoluted two maternity functions, that of the woman's mother and that of her sisters, we obtain a distribution of "niecebearing" which has a greater variance and is less skewed than that for childbearing. In addition, the age at which half of a woman's nieces have been born is consistently about two years greater than the age at which half of her own daughters have been born. Thus, one-half of a woman's nieces are born by ages 26, 31, and 30 for the United States, Venezuela, and Madagascar, respectively. The number of nieces still alive is a maximum at ages 55–60 in the United States and Venezuela and at ages 45–50 in Madagascar, after which it declines. However, at age 85 a woman in Madagascar still has nearly 50% more nieces than a woman in the United States.

5. Maternal Aunts

The number of maternal aunts of the girl of age a depends on the number of sisters her mother had. First consider her mother's older sisters. In Section 3, $S(a)$ was used to denote the expected number of older sisters of a girl or woman of age a. If the girl of age a was born at time $t - a$ to a mother of age x (at time $t - a$), then the expected number of older sisters of the mother would be $S(a + x)$. Thus, the expected number of maternal aunts (of the girl of age a) who are older than the girl's mother (i.e., the expected number of older sisters of the mother) would be

$$\int_\alpha^\beta S(a + x)\, W(x \mid t - a)\, dx,$$

which can also be written as

$$\int_\alpha^\beta \left[\int_\alpha^\beta \left\{ \int_\alpha^y m_z\, dz \right\} W(y \mid t - a - x)\, dy \right] W(x \mid t - a)\, dx.$$

For the stable population, the above formula is replaced by

$$\int_\alpha^\beta \left[\int_\alpha^\beta \left\{ \int_\alpha^y m_z\, dz \right\} W(y)\, dy \right] W(x)\, dx. \tag{5.1.a}$$

If we are concerned with the expected number of these maternal aunts who are alive at time t, the factor $l_{a+x+y-z}$ should be included in the innermost integral, thus obtaining (5.1.b).

Let us now consider the mother's younger sisters. In Section 3, $S'(a)$ denoted the number of younger sisters of a woman of age a. If the girl of age a was born at time $t - a$ to a mother of age x (at time $t - a$), then the expected number of

younger sisters of the mother would be $S'(a + x)$. Thus, the expected number of maternal aunts of the girl of age a who are younger than the girl's mother (i.e., the expected number of younger sisters of the mother) would be

$$\int_\alpha^\beta S'(a + x) \, W(x \mid t - a) \, dx,$$

which can also be written as

$$\int_\alpha^\beta \left[\int_\alpha^\beta \left\{ \int_0^{a+x} (l_{y+w}/l_y) \, m_{y+w} \, dw \right\} W(y \mid t - a - x) \, dy \right] W(x \mid t - a) \, dx;$$

and for the stable population, the above formula is replaced by

$$\int_\alpha^\beta \left[\int_\alpha^\beta \left\{ \int_0^{a+x} (l_{y+w}/l_y) \, m_{y+w} \, dw \right\} W(y) \, dy \right] W(x) \, dx. \tag{5.2.a}$$

If we are concerned with the expected number of these maternal aunts who are alive at time t, the factor l_{a+x-w} should be included in the innermost integral to give (5.2.b).

The type of argument developed above will serve equally well for establishing numbers of maternal great-aunts or any similar kind of relationship.

Numerical results for aunts ever born and aunts still alive are shown in Table VI. First note that virtually all of a woman's aunts are born by the time she reaches age 15. More precisely, all of a woman's aunts must be born by the time she reaches age $\beta - 2\alpha$. To show this, we suppose that a woman is born in year T. Her first granddaughter will be born no earlier than year $T + 2\alpha$, and her last daughter will be born no later than year $T + \beta$. The last daughter will be the youngest aunt of the first granddaughter and will be born when this granddaughter is of age $(T + \beta) - (T + 2\alpha) = \beta - 2\alpha$, or less. (Since the value of β was not the same for all our data sets, we have computed a small number of new aunts beyond age 15 for Venezuela and Madagascar.)

Second, the number of aunts eventually expected is identical to that for sisters. This is simply because a woman's aunts are her mother's sisters, an observation which will be exploited in Section 7 of this paper.

Because so few aunts are born after the woman herself, the number of aunts still alive decreases monotonically, roughly in proportion to the probability that the mother herself is alive. Indeed, by the time a woman under the Madagascar regime is born, 45% of her aunts have already died. The age at which half of a woman's aunts (ever born) can be expected still to be living is 47, 43, and 6, in the United States, Venezuela, and Madagascar, respectively. A woman in Madagascar can expect 2.4 times as many aunts *ever born* as a woman in the United States, but due to high mortality rates, she will actually have fewer *still*

alive after age 23. By contrast, a woman in Venezuela ultimately expects 2.5 times as many aunts ever born as the U.S. woman, and by age 50 she still has 2.1 times as many remaining.

TABLE VI

Expected Number of Aunts Ever Born and Still Alive,
by Age of Woman, for Three Selected Countries

	Aunts ever born			Aunts still alive		
Age	United States 1967	Venezuela 1965	Madagascar 1966	United States 1967	Venezuela 1965	Madagascar 1966
0	1.2417	3.0400	2.9745	1.1891	2.7257	1.6379
5	1.2491	3.0804	3.0173	1.1846	2.7204	1.5462
10	1.2507	3.0929	3.0332	1.1694	2.6745	1.4315
15	1.2509	3.0957	3.0379	1.1455	2.6001	1.3059
20	1.2509	3.0961	3.0390	1.1112	2.4981	1.1739
25	1.2509	3.0961	3.0391	1.0631	2.3643	1.0357
30	1.2509	3.0961	3.0391	0.9975	2.1934	0.8910
35	1.2509	3.0961	3.0391	0.9112	1.9817	0.7417
40	1.2509	3.0961	3.0391	0.8027	1.7298	0.5923
45	1.2509	3.0961	3.0391	0.6741	1.4443	0.4502
50	1.2509	3.0961	3.0391	0.5320	1.1399	0.3231
55	1.2509	3.0961	3.0391	0.3877	0.8387	0.2174
60	1.2509	3.0961	3.0391	0.2551	0.5666	0.1361
65	1.2509	3.0961	3.0391	0.1480	0.3460	0.0785
70	1.2509	3.0961	3.0391	0.0739	0.1877	0.0413
75	1.2509	3.0961	3.0391	0.0311	0.0888	0.0196
80	1.2509	3.0961	3.0391	0.0108	0.0357	0.0083
85	1.2509	3.0961	3.0391	0.0029	0.0117	0.0030

6. Cousins

We now discuss female maternal parallel cousins (i.e., daughters of maternal aunts). We again start with a girl aged a and ask first about her girl cousins through older sisters of her mother. The probability that her mother was aged x to $x + dx$ at her birth is $W(x)\, dx$ and that her grandmother was aged y at the birth of her mother is $W(y)\, dy$. (This is for the stable case; for more general application, $W(x)$ and $W(y)$ would be replaced by $W(x \mid t-a)$ and $W(y \mid t-a-x)$, respectively.) The probability that her grandmother had an older daughter at age

z, where $z < y$, is $m_z\,dz$. The prospective number of daughters that this older sister of her mother would have had would be $l_w m_w\,dw$, added from α to the present time, which is to say to $a + x + y - z$, using again the device of counting time through the ages of the several persons involved. To find the complete answer, we need only multiply the probabilities and integrate through all possible ages x (mother's age at childbirth), y (grandmother's age at childbirth), and z (grandmother's age at birth of older sister of mother) to obtain

$$\int_\alpha^\beta \left[\int_\alpha^\beta \left\{ \int_\alpha^y \left(\int_\alpha^{a+x+y-z} l_w m_w\,dw \right) m_z\,dz \right\} W(y)\,dy \right] W(x)\,dx. \qquad (6.1.a)$$

Figure 1 may help in following the lines of descent involved in this argument.

For cousins whose mother is younger than the mother of the girl aged a, we need only replace the integral over z. We have to allow for survival of the grandmother to that age z, and to do so we replace m_z in the above expression by
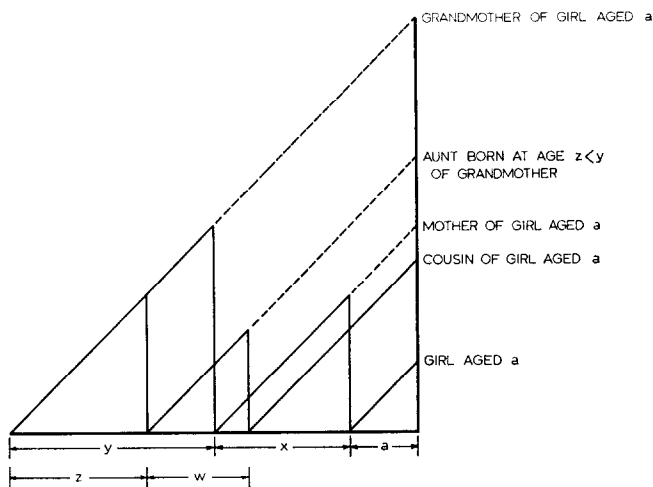

GRANDMOTHER OF GIRL AGED a

AUNT BORN AT AGE z < y
OF GRANDMOTHER

MOTHER OF GIRL AGED a

COUSIN OF GIRL AGED a

GIRL AGED a

FIG. 1.   Lexis diagram for cousin of a girl aged a through mother's elder sister.

$(l_z/l_y)\,m_z$. In addition, the limits of the corresponding integration have to be altered; instead of α to y, they become y to $a + x + y$. Hence, we have

$$\int_\alpha^\beta \left[\int_\alpha^\beta \left\{ \int_y^{a+x+y} \left(\int_\alpha^{a+x+y-z} l_w m_w\,dw \right) \frac{l_z}{l_y}\,m_z\,dz \right\} W(y)\,dy \right] W(x)\,dx, \quad (6.2.a)$$

for cousins whose mother is a younger sister of the mother of the girl aged a. The l_y in the denominator could be cancelled with the l_y contained in $W(y)$.

The sum of the two integrals would give the expected number of cousins

related through sisters of the person's mother. For the number of these that are alive, we enter in the innermost integral of each the further factor $l_{a+x+y-z-w}$ to obtain

$$\int_\alpha^\beta \left[\int_\alpha^\beta \left\{ \int_\alpha^y \left(\int_\alpha^{a+x+y-z} l_w m_w l_{a+x+y-z-w}\, dw \right) m_z\, dz \right\} W(y)\, dy \right] W(x)\, dx \quad (6.1.b)$$

for cousins through older sister of mother, and

$$\int_\alpha^\beta \left[\int_\alpha^\beta \left\{ \int_y^{a+x+y} \left(\int_\alpha^{a+x+y-z} l_w m_w l_{a+x+y-z-w}\, dw \right) \frac{l_z}{l_y} m_z\, dz \right\} W(y)\, dy \right] W(x)\, dx$$

$$(6.2.b)$$

for cousins through younger sister of mother.

Numerical results from three selected data sets appear in Table VII. Slightly

TABLE VII

Expected Number of First Cousins (Mother's Sister's Daughters)
Ever Born and Still Alive, by Age of Woman, for Three Selected Countries

	Cousins ever born			Cousins still alive		
Age	United States 1967	Venezuela 1965	Madagascar 1966	United States 1967	Venezuela 1965	Madagascar 1966
0	0.7332	3.5129	2.4824	0.7137	3.2437	1.6694
5	0.9666	4.6998	3.2129	0.9398	4.3291	2.1157
10	1.1701	5.8527	3.9064	1.1353	5.3729	2.5050
15	1.3224	6.8514	4.5006	1.2793	6.2605	2.7913
20	1.4207	7.6223	4.9601	1.3684	6.9197	2.9519
25	1.4754	8.1512	5.2811	1.4119	7.3324	2.9884
30	1.5016	8.4725	5.4833	1.4231	7.5239	2.9211
35	1.5123	8.6443	5.5981	1.4130	7.5393	2.7776
40	1.5160	8.7243	5.6566	1.3872	7.4212	2.5834
45	1.5170	8.7564	5.6833	1.3468	7.1964	2.3575
50	1.5173	8.7673	5.6940	1.2897	6.8749	2.1111
55	1.5173	8.7704	5.6978	1.2126	6.4553	1.8510
60	1.5173	8.7711	5.6990	1.1125	5.9327	1.5828
65	1.5173	8.7712	5.6993	0.9886	5.3074	1.3132
70	1.5173	8.7712	5.6993	0.8432	4.5915	1.0516
75	1.5173	8.7712	5.6993	0.6834	3.8130	0.8085
80	1.5173	8.7712	5.6993	0.5202	3.0158	0.5938
85	1.5173	8.7712	5.6993	0.3680	2.2535	0.4148

less than half of the ultimate number of cousins ever born will be reached at the moment of birth, as we observed for sisters. Using the same reasoning as for aunts, all of a woman's cousins will be born by the time she reaches age $(T + 2\beta) - (T + 2\alpha) = 2(\beta - \alpha)$, which is twice the age by which all sisters will be born. The age at which a woman has the greatest number of *living* cousins is 30–35 for the United States and Venezuela and 20–25 in Madagascar.

The eventual number of cousins ever born is identical to that for nieces ever born. We observed in Section 5 that the eventual numbers of aunts and sisters ever born are equal, and the present observation follows because cousins are simply daughters of aunts and nieces are simply daughters of sisters. The pattern of descent is made more explicit in Section 7.

7. EXPECTED EVENTUAL NUMBERS OF KIN

Every live birth can be taken as an origin or reference point for a line of descent. That is, at the point of birth a baby girl represents one mother, one maternal grandmother, etc., and also an expected number of daughters, referred to as the net reproduction rate (N), an expected number of granddaughters (N^2), etc. In this section we shall find that if one knows the numerical values of (a) the net reproduction rate and (b) the expected eventual number of sisters, then for a stable population we can also easily compute the expected eventual number of each possible category of kin.

By using the term "expected eventual number", we imply only that a continuing count is kept of persons born into specified kin roles, such as "aunt" or "niece". We do not imply that the woman (for whom we shall employ the anthropological reference "ego") who defines the structure lives to or beyond any specific age. All births and deaths occur according to the fixed schedules described previously. The quantities obtained do not refer to any specific age of ego and are not even conditional upon her still being alive.

Figure 2 shows how the expected eventual frequencies are generated. The movement down each column, from one generation to the next, simply involves multiplying the preceding number by N, the net reproduction rate. The equivalence of the expected number of sisters, aunts, great-aunts, etc., may require justification. Note that ego has one and only one mother, and this woman (ego's mother) would expect eventually to have just as many sisters as would ego, in a stable population. And from ego's point of view, her mother's sisters are aunts. Similarly for great-aunts, etc.

The next table (Table VIII) gives the numerical quantities for the cells of Fig. 2 for the three sets of data used earlier in this paper. The quantities for sisters, aunts, nieces, and cousins agree with those cited earlier if the woman has survived to the oldest ages. For example, according to the rate schedules for 1965

and 1967, a Venezuelan woman will expect 31.5 times as many third cousins (removed zero times) as a United States woman and 12.7 times as many great-granddaughters.

GENERATIONS REMOVED FROM EGO	MAIN LINE	SIBLING LINE	FIRST COUSINS	SECOND COUSINS	THIRD COUSINS
-3	$1*$ (8)*	(16)	Great-great-aunts: S		
-2	$1*$ (4)*	(8)	Great-aunts:S	S X N (32)	
-1	Mother: $1*$ (2)*	(4)	Aunts: S	S X N (16)	S X N^2 (64)
0	Ego: $1*$ (1)*	Sisters:S (2)	S X N (8)	S X N^2 (32)	S X N^3 (128)
1	Daughters: (2) N	Nieces: (4) S X N	S X N^2 (16)	S X N^3 (64)	S X N^4 (256)
2	N^2 (4)	S X N^2 (8)	S X N^3 (32)	S X N^4 (128)	S X N^5 (512)
3	N^3 (8)	S X N^3 (16)	S X N^4 (64)	S X N^5 (256)	S X N^6 (1024)

FIG. 2. Eventual expected number of kin in a stable population. S = Eventual expected number of sisters, N = net reproduction rate. Number in parentheses is approximate multiplier in a two-sex model, "*" indicates number is exact.

TABLE VIII

Quantities Appearing in Fig. 2 Computed for Three Countries

	United States, 1967	Venezuela, 1965	Madagascar, 1966
S	1.2509	3.0961	3.0393
$S \times N$	1.5173	8.7712	5.6992
$S \times N^2$	1.8404	24.8488	10.6871
$S \times N^3$	2.2324	70.3966	20.0404
$S \times N^4$	2.7079	199.4335	37.5797
$S \times N^5$	3.2846	564.9951	70.4694
$S \times N^6$	3.9842	1600.6311	132.1442
N	1.2130	2.8330	1.8752
N^2	1.4714	8.0259	3.5164
N^3	1.7848	22.7374	6.5940

All of the quantities computed to this point have been restricted to females, related through females. If the sex of the relationship is ignored, and if the sex of the respondent is left open, then the expected number of kin increases rapidly with the distance of the relationship. The factor by which each frequency increases is approximately a power of two, depending on the sex ratio at various ages. Each cell of Fig. 2 ontains a number in parentheses which shows the increase in frequency as "mother" is generalized to "parent", "sister" to "sibling," "daughter" to "child," etc. Note that under the generalization, the expected eventual number of *brothers and sisters* is only about half the number of *uncles and aunts*, etc.

8. MEAN SIZE OF HOUSEHOLD

Let us consider households consisting of a nuclear family of husband, wife, and unmarried children (if any). Suppose that all children live with their parents until they themselves marry, and that to the data consisting of l_x and m_x used throughout this paper we add $P(a)$, the proportion of females who are married among all females of age a. We seek the proportion of married women in the total female population; the reciprocal of this proportion will be the mean number of females per household. (We do not count as separate those households in which the wife has died; their numbers are implicitly taken as part of some other household with living wife.)

Let $f(a \mid t)$ denote the age distribution of the female population at time t. Then the proportion of married women in the female population is

$$\int_0^\omega P(a) f(a \mid t)\, da,$$

where ω is the highest age to which anyone lives. For a stable population, we note that $f(a \mid t)$ can be replaced by

$$f(a) = l_a e^{-ra} \Big/ \int_0^\omega l_x e^{-rx}\, dx.$$

In this case, the mean number of females per household would be

$$1 \Big/ \int_0^\omega P(a) f(a)\, da. \tag{8.1}$$

Expressions for the sex ratio are provided by Goodman (1953 and 1968). To the degree to which populations consist of about equal numbers of men and

women, the mean number of members of the nuclear family, male as well as female, may be found by doubling the above expression, thus obtaining

$$2 \Big/ \int_0^\omega P(a) f(a) \, da. \tag{8.2}$$

Formula (8.2) is applicable when (a) each married female forms a household; (b) each unmarried female is a member of the household of some married female (her mother or foster mother); (c) each male is a member of the household of some married female (his mother, foster mother, or wife). The implicit supposition is that illegitimate children also are in the household of some married woman. Modifications of (8.2) are required when conditions (a), (b), or (c) are not met, but we shall not go into these details here.

Consider now the extended family rather than the nuclear family, where (a) each married female forms a household if her mother is dead; (b) each married female is a member of her mother's household if her mother is alive; (c) each unmarried female is a member of the household of some married female (her mother or foster mother); (d) each male is a member of the household of some married female (his mother, foster mother, or wife). In this case, the mean size of household is given by

$$2 \Big/ \int_0^\omega \left[1 - \int_\alpha^\beta (l_{x+a}/l_x) \, W(x \mid t - a) \, dx \right] P(a) f(a \mid t) \, da,$$

where $W(x \mid t - a)$ is the age distribution defined in the earlier sections herein. The quantity in brackets in the above formula is the probability that the mother of a female of age a is dead. We assume here that this probability does not depend upon whether the female of age a is married. If this assumption is not met, modifications in the formula are required, but we shall not go into them here. For a stable population, the above formula for the mean size of household of the extended family is replaced by

$$2 \Big/ \int_0^\omega \left[1 - \int_\alpha^\beta (l_{x+a}/l_x) \, W(x) \, dx \right] P(a) f(a) \, da, \tag{8.3}$$

with $W(x)$ and $f(a)$ calculated from the formulas given earlier.

For some related (but different) results on this topic, we refer the reader to Coale (1965) and Burch (1970). They provide expressions for the mean size of the nuclear and the extended family under assumptions somewhat more restrictive than ours. They assume that all females marry, that all marriages take place exactly at the average age of marriage, and that all births take place at the average age of childbearing. If these assumptions are violated (as they are in fact), the formulas used in the earlier literature can give incorrect results, but the formulas introduced herein will continue to be applicable.

SUMMARY

This article introduced some simple techniques that can be used to ascertain the expected frequencies or probabilities pertaining to various kinship relations. For the following six categories, these expected frequencies or probabilities are specific for the age of the woman: (1) Daughters and other descendants; (2) mother and other progenitors; (3) sisters; (4) nieces; (5) aunts; (6) cousins. Under 1, 3–6, the expected number ever alive and survivors still alive were shown as formulas a and b, respectively. The method is applicable to more distant relatives and, with some modification, to the male as well as to the female line. In Section 7 we showed that the expected eventual frequency for any possible kinship relation in a stable population depends only upon the net reproduction rate and the expected eventual number of sisters. In Section 8 we gave formulas for the expected size of nuclear and extended families under simple assumptions of family formation.

Numerical approximations to the integrals were computed using as input only age-specific rates of birth and death (see Appendix). Computations for Section 8 would require, in addition, the proportion of married females in each age interval.

APPENDIX: NUMERICAL EVALUATION OF INTEGRALS

Formula 2.1, an example of a single integral, was given as

$$M_1(a) = \int_\alpha^\beta l_{a+x} m_x e^{-rx} \, dx. \tag{A.1}$$

Ordinarily, we cannot evaluate the l_x and m_x functions for arbitrary values of x, since data are usually collected for five-year age intervals. Thus, when x is a multiple of 5, we have the life table function

$$_5L_y = \int_y^{y+5} l_x \, dx \tag{A.2}$$

and observed five-year fertility rates

$$_5F_y = \int_y^{y+5} m_x k_x \, dx \Big/ \int_y^{y+5} k_x \, dx \tag{A.3}$$

where k_x is proportional to the observed age distribution.

We enlarge the range of integration somewhat by redefining α to be the largest multiple of 5 that is less than or equal to the lowest age of reproduction, and β

to be the smallest multiple of 5 that is greater than or equal to the highest age of reproduction. Then, if y is restricted to multiples of 5, we can rewrite (A.1) as

$$M_1(a) = \sum_{y=\alpha}^{\beta-5} \int_y^{y+5} l_{a+x} m_x e^{-rx} \, dx. \tag{A.4}$$

Since each term in the integrand is a continuous function, we can apply the fundamental theorem of the integral calculus to get, for $y = \alpha, \alpha + 5, ..., \beta - 5$,

$$\int_y^{y+5} l_{a+x} m_x e^{-rx} \, dx = m_{\bar{y}} e^{-r\bar{y}} \int_y^{y+5} l_{a+x} \, dx$$

$$= m_{\bar{y}} e^{-r\bar{y}} \, {}_5L_{a+y}, \tag{A.5}$$

where $\bar{y}$ is some number between y and $y + 5$, i.e., $y \leqslant \bar{y} \leqslant y + 5$. We then assumed, first, that $\bar{y} \doteq y + 5/2$, and second that $m_{\bar{y}} \doteq m_{y+5/2} \doteq {}_5F_y$. The second assumption will be true if m_x is constant between y and $y + 5$, but it can be true under other conditions as well, depending on the age distribution within the interval. We thus employed the following approximation for $M_1(a)$:

$$M_1(a) = \sum_{y=\alpha}^{\beta-5} e^{-r(y+5/2)} \, {}_5L_{a+y} \, {}_5F_y. \tag{A.6}$$

The other single integrals (1.1.a and b) were approximated analogously.

We shall indicate our numerical approach to double integrals using a general form rather than a specific formula. Each double integral in this paper has the following form:

$$S = \int_\alpha^\beta \left(\int_{b_1(x)}^{b_2(x)} f(x, y) \, dy \right) g(x) \, dx,$$

where b_1 and b_2 are the lower and upper boundaries of integration of the inner integral and are themselves functions of x. The inner integral can then be summarized by a new symbol,

$$I(x) \equiv \int_{b_1(x)}^{b_2(x)} f(x, y) \, dy,$$

a function of x, so that

$$S = \int_\alpha^\beta I(x) g(x) \, dx = \sum_{z=\alpha}^{\beta-5} \int_z^{z+5} I(x) g(x) \, dx,$$

where α, β, and z are all multiples of 5.

Again applying the fundamental theorem of the integral calculus, since $I(x)$ is continuous,

$$S = \sum_{z=\alpha}^{\beta-5} I(\bar{z}) \int_z^{z+5} g(x)\, dx$$

for some $\bar{z}$, $z \leqslant \bar{z} \leqslant z + 5$. Our data are such that $I(x)$ can only be evaluated for arguments which are multiples of 5, so we approximate $I(\bar{z})$ with $\frac{1}{2}[I(z)+I(z+5)]$. This approximation will be exact if $I(x)$ is constant between z and $z + 5$, but it can also be exact (or more or less accurate) depending on the configuration of the component functions.

We have shown how to reduce double integrals to single integrals, which have already been discussed; reduction of triple integrals, etc., is analogous and need not be detailed.

There are two obvious directions for improving the quality of these approximations. One could assume that within five-year intervals l_x and m_x are linear, quadratic, etc. Alternatively, the $_5L_x$ and $_5F_x$ functions could be graduated to single years (or less) of age and the five-year units of the preceding sums could be changed to single-year units.

The second approach is simpler in principle but would have involved much computer time. We did compute the probability of a living grandmother (formula 2.2) for the data of Venezuela, 1965, using five-year and single-year age intervals, to estimate the order of approximation for five-year summations. Table A shows our findings to three decimal places. For each age, it appears

TABLE A

Comparison of $M_2(a)$ as Calculated in Five-Year and Single-Year Age Intervals, Venezuela 1965, Females

	(2.2) Probability of living grandmother calculated in	
Girl of exact age a	Five-year age groups	Single years of age
0	0.867	0.869
5	0.812	0.814
10	0.742	0.745
15	0.656	0.659
20	0.553	0.556
25	0.438	0.442
30	0.319	0.322
35	0.206	0.209
40	0.112	0.114
45	0.048	0.049
50	0.014	0.015
55	0.002	0.002
60	0.000	0.000

that five-year computations slightly underestimate the true probability of a living grandmother by an amount which never exceeds .004 in absolute value. The discrepancy would presumably be larger for our triple and quadruple integrals, and less for the single integrals.

The computer program which yielded the numerical results in Tables I–VII, and which can be applied to any of the formulas in Sections 1–6, was written in FORTRAN IV. It is available upon request.

REFERENCES

BRASS, W. 1953. The derivation of fertility and reproduction rates from restricted data on reproductive histories, *Population Studies* **7**, 137–166.

BURCH, T. K. 1970. Some demographic determinants of average household size: An analytic approach, *Demography* **7**, 61–70.

COALE, A. J. 1965. Appendix: Estimates of average size of household, *in* "Aspects of the Analysis of Family Structure" (A. J. Coale, Ed.), pp. 64–69, Office of Population Research, Princeton, NJ.

GOODMAN, L. A. 1953. Population growth of the sexes, *Biometrics* **9**, 212–225.

GOODMAN, L. A. 1968. Stochastic models for the population growth of the sexes, *Biometrika* **55**, 469–487.

HENRY, L. 1960. Mesure indirecte de la mortalité des adultes, *Population* 15e année, 457–465.

KEYFITZ, N., AND W. FLIEGER. 1971. "Population: Facts and Methods of Demography," W. H. Freeman and Company, San Francisco.

LOTKA, A. J. 1907. Relation between birth rates and death rates, *Science* N.S. **26**, 21–22.

LOTKA, A. J. 1931. Orphanhood in relation to demographic factors: A study in population analysis, *Metron* **9**, 37–109.