Correlations Between Frequencies of Kin*

Thomas W. Pullum
>
> Population Research Center
> The University of Texas
> Austin, TX 78712

Douglas A. Wolf
>
> The Urban Institute
> 2100 M Street, NW
> Washington, DC 20037

Recent years have seen the development of formal and microsimulation models of the structure and dynamics of kin networks. These models generally assume uncorrelated fertility within and across generations. Several sets of real data, however, show positive correlations between the frequencies of various categories of kin. This paper uses formal models to calculate the correlations that will exist between certain categories of kin even if mothers and daughters have independent fertility. Mechanisms by which fertility might be transmitted from mothers to their daughters are considered and the implications for kin correlations are evaluated.

The past few years have seen the development of demographic methods and data pertaining to the numbers of grandchildren, siblings, and other kinds of kin in addition to children, who form the category of greatest interest. Some of the formal models concern the numbers of kin who would be be implied in a stable population, in which birth and death schedules are fixed (Goodman, Keyfitz, and Pullum 1974, 1975; Le Bras 1973). In a stable population it is possible to estimate analytically the mean numbers of kin of specific types who ever would have been born, or still would be living, according to the age of a reference individual known as "ego". Other procedures allow estimation of the distribution of the numbers who ever will be born into specific kin categories, by using the frequency distribution of daughters (Pullum 1982; Waugh 1981).

[Editor's Note: The journal has a policy of not reviewing manuscripts from members of its editorial board. This article was submitted and reviewed under the editorship of Gordon F. De Jong. Pullum, now an associate editor, was not on the editorial board at the time. Avery M. Guest, the current editor, is honoring the decision to publish that was made by De Jong.]

Computer simulation methods allow the estimation of frequencies of kin with greater flexibility in assumptions than is possible with formal or mathematical approaches (Bongaarts 1987). The numbers of immediate kin have been projected or simulated under a range of assumptions by Hammel, Wachter, and McDaniel (1979), Reeves (1987), and Smith (1987). Finally, data have been collected in recent years that permit the validation of the other procedures—in addition, of course, to other valuable uses. A variety of approaches to these themes are described in Bongaarts, Burch, and Wachter (1987), including a chapter by Pullum (1987).

Such work, including the present paper, has two main objectives. The first is to articulate the population dynamics of kinship—to specify the implications of different patterns of fertility and mortality for the network of kin. The second objective is less theoretical, and concerns the actual availability of kin for social contact and support, particularly in the older ages and in contexts where reproduction is near replacement levels (Wolf 1988). Formal modeling and simulation are important for this purpose as well as for the first, because of a general inadequacy of empirical data and because of interest in projections into the future.

A closely related issue concerns intergenerational transmission of fertility patterns, an object of research for many years (Berent 1953; Hendershot 1969; Huestis and Maxwell 1932; Johnson and Stokes 1976; Kantner and Potter 1954; Pearson and Lee 1899). A recent contribution by Anderton et al. (1987) includes a summary of the previous literature. A positive correlation between the fertility levels of mothers and those of their daughters would imply that the fertility levels of more distant relatives are also correlated. If such correlations exist, then some individuals will be located in relatively sparse networks of kin, and others in relatively dense networks, with a greater range in the overall density than if no such correlations existed.

This paper will bring together formal models, computer simulation, and actual data in an effort to answer questions about correlations between the frequencies of kin of different types. Except as noted, we will deal solely with one-sex models. Thus the reference individual will be female; children will be daughters; granddaughters will be the daughters of daughters; siblings will be sisters; and so on. When necessary, we will describe limitations due to this simplification. The mathematical results to be presented here will be new in this context, but would be considered elementary within the general theory of branching processes.

Some Relevant Issues

It will be helpful to describe some possible ambiguities and to propose how to resolve them. Suppose, for specificity, that we asked the following question in relation to a formal model, a simulation, or a body of data: "Is there a correlation between the number of sisters and the number of daughters?" How is such a question to be understood?

It should be clear that the units of analysis are individual females whose sisters and daughters have been tallied. The two counts, however, will depend on the age of the reference individual and on whether they are defined in terms of numbers ever born or numbers still alive. The number of kin ever born is a nondecreasing function of age, and the number still alive will tend to increase and then to decrease.

Age has another complicating property: it affects the probability that the woman herself is still alive. If she dies early, before bearing any daughters or even before all her sisters are born, how should she contribute to the correlation?

These issues are illustrated in Table 1, which shows several alternative concepts of kin correlations, by age, in a simulated one-sex population of approximately 100,000 women.

Table 1. Mean Numbers of Daughters, Granddaughters, and Sisters, and Correlations among Them, by Age of Ego, in a Simulated Stationary Population with Independence of Reproduction

Age Group	Daughters X	Grand-daughters Y	Sisters Z	Correlations	
				rxy	rxz
A. Living Kin					
0–4	0.00	0.00	0.63	0.00	0.00
5–9	0.00	0.00	0.85	0.00	0.00
10–14	0.00	0.00	0.94	0.00	0.00
15–19	0.05	0.00	0.99	0.00	−0.02
20–24	0.33	0.00	1.00	0.00	0.01
25–29	0.65	0.00	0.98	0.00	0.01
30–34	0.86	0.00	0.98	0.02	−0.01
35–39	0.98	0.02	1.02	0.12	−0.01
40–44	1.00	0.13	1.00	0.33	0.01
45–49	0.99	0.32	0.97	0.46	−0.02
50–54	0.99	0.56	0.92	0.58	0.01
55–59	0.99	0.78	0.88	0.66	0.00
60–64	0.98	0.91	0.85	0.68	0.01
65–69	0.98	0.95	0.78	0.70	0.02
70–74	0.97	0.99	0.69	0.70	0.01
75–79	0.94	1.01	0.59	0.68	−0.01
80–84	0.91	0.96	0.46	0.69	0.00
85–89	0.90	1.01	0.34	0.64	−0.02
90+	0.75	0.89	0.16	0.63	−0.04
B. Kin Ever Born					
0–4	0.00	0.00	0.63	0.00	0.00
5–9	0.00	0.00	0.86	0.00	0.00
10–14	0.00	0.00	0.95	0.00	0.00
15–19	0.05	0.00	1.00	0.00	−0.02
20–24	0.34	0.00	1.02	0.00	0.01
25–29	0.66	0.00	1.00	0.00	0.01
30–34	0.87	0.00	1.00	0.02	−0.01
35–39	0.99	0.02	1.05	0.12	−0.01
40–44	1.02	0.13	1.04	0.33	0.01
45–49	0.01	0.33	1.02	0.46	−0.03
50–54	1.01	0.57	0.99	0.58	0.00
55–59	1.02	0.79	0.99	0.65	−0.00
60–64	1.01	0.92	1.00	0.68	0.01
65–69	1.02	0.97	1.00	0.70	0.02
70–74	1.02	1.01	1.00	0.71	0.01
75–79	1.02	1.03	1.00	0.70	−0.01
80–84	0.99	0.98	0.99	0.70	0.00
85–89	1.04	1.04	1.02	0.69	−0.06
90+	0.95	0.93	0.96	0.72	−0.05

We generated the population using computer simulation techniques described in Wolf (1988). The assumptions underlying the simulation are the same in all important respects as those made in the analytic model developed by Goodman et al. (1974); simulation, however, permits the full frequency distribution of kin, by age, to be determined. In the simulated population, all births and deaths are independent, and are determined by birth and death rates that are fixed over time. The net reproduction rate in the simulated population is exactly 1.

The table shows the lifetime profile of the correlation between the numbers of daughters (X) and granddaughters (Y) and between the numbers of daughters and sisters (Z). The effects of mortality are illustrated by including figures both for living kin and for ever-born kin. Considering sisters and daughters, we see a sequence of correlations that are only trivially different from 0, with no apparent pattern.

In the case of daughters and granddaughters, the correlation is 0 until ages 35 to 39, when granddaughters first begin to appear, and rises over the next seven age groups to its maximum of approximately 0.71. For living daughters and granddaughters, the correlation declines somewhat at older ages as the number of living daughters decreases but the number of living granddaughters does not. This decline, of course, does not appear for daughters and granddaughters ever born.

In our view, the optimal way to define the counts of kin is in terms of genealogies, which can be modeled, simulated, or described with real data (although not easily with data from a sample survey). The genealogy for each reference individual, ego, is constructed outward and extends far enough to include all of her kin ultimately born, regardless of their survival. It is convenient to assume that the data extend far past the birth of ego, regardless of how long she survives; just how far will depend upon the categories of kin of interest. Kin will be counted if they were ever born into a specified relationship to ego, regardless of how long they survive. It is not necessary for ego and the kin ever to have been alive simultaneously. For example, ego's sisters will consist of all her mother's daughters, apart from ego herself, even if the sister died before ego was born or was born after ego's death.

This definition of kin, in terms of a list of individuals ultimately born into a complete genealogy, is attractive because it eliminates the need to choose any of the variously objectionable and complex controls for age and survivorship. It corresponds to the formulation developed by Pullum (1982), which gives the full frequency distributions of all categories of kin ever born under the assumptions of fixed and homogeneous rates. Under the definition adopted, we need not consider age when discussing the magnitude of any correlation. In the simulated population in Table 1, the correlation between ultimate numbers of sisters and daughters is 0; between ultimate numbers of daughters and granddaughters, it is 0.71 (an estimate of a parameter to be calculated below).

Figure 1, designed by Atkins (1974) and used by Pullum (1982), provides a convenient representation of the links between ego and her kin. The arrows along the top show the line of direct descent to ego; each category contains exactly one progenitor. The vertical arrows in the left-hand column show the line of direct descent from ego. The other vertical arrows show lines of descent to and from the siblings of ego's progenitors, i.e., her collateral kin. Kin categories will be identified by their row (i) and column (j) in this figure.

In this paper we will consider two alternative sets of assumptions about the pattern of reproductivity, and for each set we will provide formulas and numerical estimates of certain correlations among kin. Then we will compare these estimates with some observed correlations in real human populations.
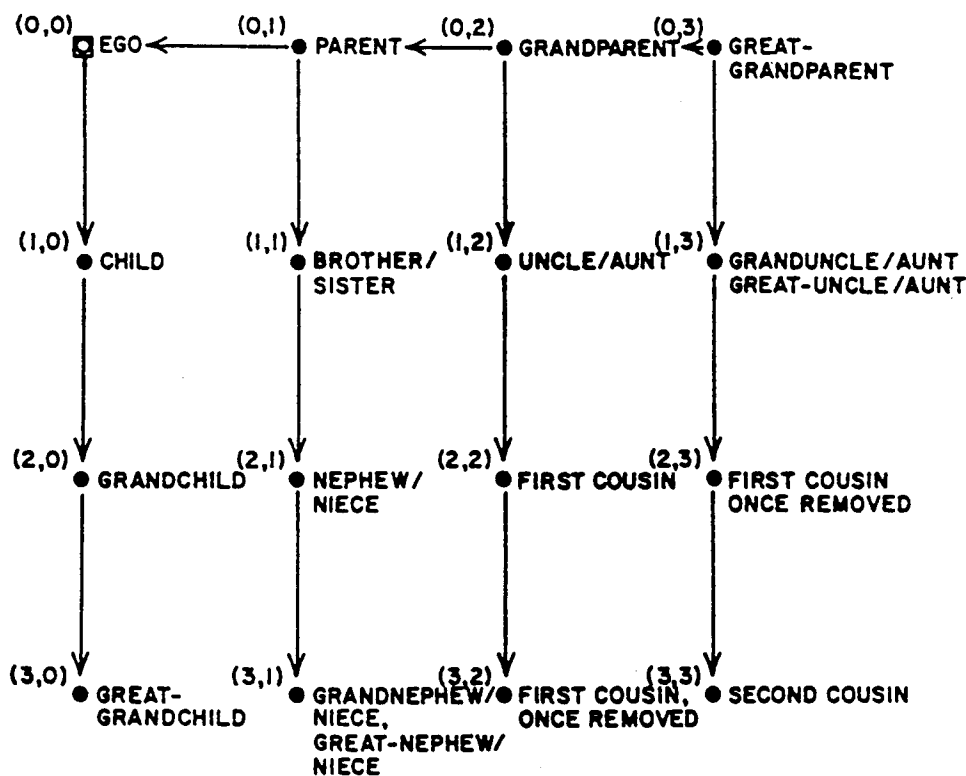
Figure 1. Two-dimensional Lattice of Kinship Categories *(i,j)* for Ego at (0,0). Design from Atkins (1974)

Model 1: Homogeneity and Independence

Let X denote a random variable, the number of daughters, and let x be a specific value of X (limited to nonnegative integers). Define

$$f_x = Pr(X = x)$$

to be the probability, defined at birth, that an individual female (ego) eventually will give birth to exactly x daughters. The set of these probabilities will be referred to as ego's *reproductive regime*.

The first model assumes that all women have the same reproductive regime and that the probabilities for any woman are independent of the probabilities and behavior of any other woman, particularly her mother. These are the assumptions used in earlier demographic modeling of kinship and in the simulation presented in Table 1.

From these probabilities and assumptions it is possible to obtain the moments, and in fact the complete probability distribution, for any kin category (i,j) (see Pullum 1982). For these purposes it is helpful to employ the probability generating function for daughters, defined by

$$f(s) = \Sigma f_x s^x.$$

The mean of the distribution of X is the eventual expected number of daughters, better known as the net reproduction rate. This is simply

$$N = E(X) = \Sigma x f_x = f'(1),$$

the first derivative of f(s) with respect to s, evaluated at $s = 1$. The corresponding second derivative, divided by N, that is,

$$E[X(X-1)]/E(X) = f''(1)/N,$$

is also useful and will be labeled S because it can be shown to be the eventual expected number of ego's sisters (see Pullum 1982). Higher moments can be found from higher-order derivatives. As an example of a specific probability, the chance that ego will have no granddaughters is the chance that she will have x daughters and that each of them will have no daughters,

$$\Sigma f_x (f_0)^x,$$

and so on.

An assumption of stability or long-term constancy of probabilities is not required. It is necessary only to assume homogeneity and independence within those generations which link ego to some other specified kin category or categories. For example, the estimated correlation between daughters and granddaughters will require only that the assumptions are valid for all women in ego's generation and in her daughter's generation. The estimated correlation between sisters and daughters will require the assumptions only for those women in ego's generation and her mother's generation.

Correlations between Daughters and Granddaughters

The first two categories to be correlated are daughters and granddaughters. Over the universe of all possible egos, define two random variables: X, the number of daughters; and Y, the number of granddaughters. By definition the correlation between these two variables is

$$r = Cov(X,Y)/ \sqrt{Var(X)Var(Y)}.$$

An elegant simplification of this formula can be obtained without any reference to generating functions. The assumption of independence can be interpreted to mean that

$$E(Y|X = x) = cx$$

for each value x of the random variable x and for some constant c. That is, each daughter of ego will be expected to have c daughters herself, no matter how many daughters ego had. Multiplying both sides by $Pr(X = x)$ and adding over x will give

$$E(Y) = cE(X),$$

that is,

$$c = E(Y)/E(X).$$

Next note that

$$E(XY|X = x) = xE(Y|X = x) = cx^2$$

for each value of x. Again, multiplying both sides by $Pr(X = x)$ and adding over x will give

$$E(XY) = cE(X^2) = E(Y)E(X^2)/E(X).$$

Substituting this value of E(XY) into

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

will yield the result

$$r = [E(Y)/E(X)][\sqrt{Var(X)/Var(Y)}]$$

or, equivalently,

$$r = [\sqrt{Var(X)}/E(X)]/[\sqrt{Var(Y)}/E(Y)].$$

That is, the correlation between X and Y is the coefficient of variation of X divided by the coefficient of variation of Y.

Although this result has been phrased in terms of daughters and granddaughters, it will apply whenever the kin denoted Y are the daughters of the kin denoted X and the assumption of independence holds. It will also apply in a two-sex situation where the kin Y are the children of the kin X.

Now assume that the daughters have the same reproductive regime as the mothers; that is, that homogeneity as well as independence is present across generations. Then from Pullum (1982) we have, for daughters,

$$E(X) = N,$$
$$E(X^2) = N(S + 1), \text{ and}$$
$$Var(X) = N(S + 1 - N),$$

obtained from the first and second derivatives of f(s). For granddaughters, using the first and second moments of $[f(s)]^2$, we have

$$E(Y) = N^2, \ E(Y^2) = N^2(NS + S + 1), \text{ and}$$
$$Var(Y) = N^2(N + 1)(S + 1 - N).$$

Note that $S + 1$ can be interpreted as the mean number of daughters in a family from the daughter's perspective, so $S + 1 - N$ is the difference between the mean from the daughter's perspective and the mean from the mother's perspective. If there is any dispersion in reproductivity, this difference will be positive.

Substituting all components into the previous formula for the correlation and simplifying, we obtain the striking result

$$r = \sqrt{N/(N + 1)}.$$

For example, in a stationary population the correlation will be

$$\sqrt{1/2} = .71;$$

half of the variation in the number of granddaughters is "explained" by the number of daughters. If the population were doubling in a generation, the correlation would be even higher,

$$\sqrt{2/3} = .82.$$

Because most human populations have a current or recent level of reproduction in this range, it is possible to say that the correlation between daughters and granddaughters will be in the range of .7 to .8 even in the absence of any tendency for high or low fertility to run in families. Any such tendency would serve to raise this correlation to an even higher level.

Other Direct Descendants

Consider next the correlation between direct descendants in generation i and in the subsequent generation i + 1. The case considered above was for generations 1 and 2. Let random variables X and Y refer to the numbers of kin in each of the two generations.

The general formulas for means and variances, drawn from the basic literature on branching processes (see Harris 1963), are

$$E(X) = N^i,$$
$$E(X^2) = N^i[1 + S(N^i - 1)/(N - 1)],$$
$$Var(X) = N^i[(N^i - 1)/(N - 1)](S + 1 - N),$$
$$E(Y) = N^{i+1}, \text{ and}$$
$$Var(Y) = N^{i+1}[(N^{i+1} - 1)/(N - 1)](S + 1 - N).$$

After substitution, the correlation simplifies to

$$r = \sqrt{N(N^i - 1)/(N^{i+1} - 1)} = \sqrt{(N^{i+1} - N)/(N^{i+1} - 1)}.$$

The correlation rapidly approaches unity as i increases, regardless of the value of N. The special case of stationarity can be evaluated by first factoring a term N-1 out of both the numerator and the denominator and then setting $N = 1$, to produce the strikingly simple result

$$r = \sqrt{i/(i + 1)}.$$

This result is easily generalized to the correlation between any two categories of direct descendants, i_1 and i_2, with $i_1 < i_2$. This will simply be the *product* of the correlations between i_1 and $i_1 + 1$, between $i_1 + 1$ and $i_1 + 2, \ldots,$ and between $i_2 - 1$ and i_2, because the partial correlation between any two categories, controlling for an intermediate category, will be 0 under the assumptions of the process. That is, the correlation between the numbers of descendants of orders i_1 and i_2 will be

$$r = \sqrt{N^{i2}(N^{i1} - 1)/N^{i1}(N^{i2} - 1)},$$

and if $N = 1$, then

$$r = \sqrt{i_1/i_2}.$$

Sisters and Daughters

A correlation of great potential interest is that between the number of sisters and the number of daughters. As we have described the process, however, there is no sense in which the number of daughters is conditional on the number of sisters, or vice versa. If a woman has no daughters, then she must have no granddaughters; but if she has no sisters, we know nothing about how many daughters she has, or vice versa. In short, the correlation between sisters and daughters is 0. If it is observed empirically to be nonzero, there is evidence of some form of inheritance.

Other Categories

More generally, the numbers of kin in categories that are displayed in different columns of Figure 1 will be uncorrelated. The top category in each column contains one and only one individual, and there are no pathways of descent to link categories in different columns which are below the top row.

Within any column, however, descent will produce correlations, just as in the first column of direct descendants. As noted in Pullum (1982), for i>0 and j>0 the expected values and variances will depend upon i but not upon j; that is, they vary by row but not by column.

Within any arbitrary column j, j$>$0, let X and Y refer to the number of individuals in two successive categories, i and i$+$1 respectively. By the same reasoning as we followed before, the correlation between the two counts is

$$r = [E(Y)/E(X)][\sqrt{Var(X)/Var(Y)}].$$

The ratio of expected numbers will be

$$E(Y)/E(X) = N.$$

Thus, when we use general formulas in the literature on branching processes for the variance (given, for example, in Pullum 1982), it is readily possible to calculate r. The correlation between categories in rows i_1 and i_2, $i_1 < i_2$ (within a given column j) again can be obtained by multiplying together the intermediate correlations. Empirical correlations will be attenuated insofar as individuals may occupy multiple kin categories relative to ego, as will occur when marriages between cousins are allowed.

Extension to Two Sexes

Adjustments easily can be made to these formulas to encompass both male and female kin. For example, if the probability-generating function referred to all children, not only to daughters, then the correlation between children and grandchildren would be obtained by replacing N in the formula

$$r = \sqrt{N/(N+1)}$$

by the expected eventual number of children, rather than that of daughters. That is, N would be replaced by $N(1+s)$, where s is the sex ratio at birth, the number of male births divided by the number of female births. Thus the correlation between children and grandchildren is

$$r = \sqrt{N(1+s)/[N(1+s)+1]}.$$

If we use $s \doteq 1.04$, this correlation will be .82 when N$=$1 and .90 when N$=$2. The correlation between children and grandchildren is substantially larger than that between daughters and daughters' daughters.

The only other quantities required for the other correlations in this section are S and Var(Z). If sisters are replaced by siblings, then the two-sex formulas can be developed readily. This step simply requires that the assumptions of homogeneity and independence be extended to include males, and that the probability of having another child depends only on the number of children, not on their gender. Under these assumptions, the numbers of sons and of daughters (born to an ego of either gender) will be independent of each other. Therefore the kth moment of the distribution of children will be $(1+s)^k$ times the kth moment of the distribution of daughters. Hence the expected number of siblings will be $(1+s)S+s$, which should be used in place of S. The variance in the number of siblings will be $(1+s)^2 Var(Z)$, which should be used in place of Var(Z). Similar links between single-sex and two-sex correlations could be developed for the remaining models in this paper.

Model 2: Mixtures of Reproductive Regimes

So far we have considered populations in which all reproduction is governed by identical probabilities, independent within genealogies. Now consider a model in which correlations between sisters and daughters can arise. Various assumptions can lead to the

intergenerational transmission of fertility. In the assumption to be considered here, an individual's reproductive regime is assigned to her at the point of her birth on the basis of her mother's eventual completed parity. According to the arguments developed in Anderton et al. (1987), it is plausible that heterogeneity results from the size of the family of socialization.

A specification of this socialization effect might be as follows: a woman's reproductive regime will increase monotonically, as indexed by her expected fertility, according to the eventual number of her sisters. Possible refinements could be imagined, depending upon the mechanisms that are postulated to underlie the transmission.

The extension to two sexes will be problematic if the chance that a woman will have x children depends not only on the number of her own siblings, but also on the number of her husband's siblings. Direct application of the present model to two sexes will require an assumption that the socialization effect is female-dominant; that is, that the reproductive regime is affected by only the female partner's sibship size.

Say that the regime which applies to a specific woman is determined completely by her eventual number of sisters, z ($z = 0, 1, \ldots$), taken to define her reproductive class. All women in class z are subject to a set of probabilities of having x daughters themselves ($x = 0, 1, \ldots$), labeled f_{zx}. These probabilities are assumed to be fixed from one generation to the next.

Indexing by ego's eventual number of sisters, z, is completely equivalent to indexing by ego's mother's eventual number of daughters, $z + 1$. Because ego's mother had at least one daughter (ego), the latter index would start at 1 rather than at 0.

Define N_{1z} to be the expected completed parity of women who have z sisters, i.e.

$$N_{1z} = \Sigma x f_{zx} .$$

It is helpful to define the expected value of the squared parity of such women to be

$$N_{2z} = \Sigma x^2 f_{zx}.$$

It is also necessary to specify the proportion of the women who are in class z in ego's generation, labeled p_z. The sum Σp_z must equal 1.

If the conditional probabilities f_{zx} are known, together with the distribution p_z from some starting generation, it is possible to calculate recursively the probabilities for all categories of kin. Eventually, the unconditional probabilities f_x and the distribution of women across the classes will become stable; i.e., they will not change from one generation to the next. We will not require stability or explicate these links.

Daughters and Granddaughters

The probability that ego has z sisters is p_z, in which case she expects N_{1z} daughters. Therefore the unconditional expected number of daughters is

$$E(X) = \Sigma p_z N_{1z}.$$

Similarly, the expected number of daughters squared is

$$E(X^2) = \Sigma p_z N_{2z}.$$

Also note that the probability that ego will have no daughters is $p_z f_{z0}$.

Again let Y represent the number of granddaughters. If ego has z sisters, then she has probability f_{zx} of having x daughters. Each of these x daughters will be in class $x - 1$, and herself will expect $N_1, x - 1$ daughters, furnishing ego with an expected

$$E(Y|X=x) = xN_{1,x-1}$$

granddaughters. Weighting by the proportions of women in generation 0 who have z sisters, the unconditional expected number of granddaughters will be

$$E(Y) = \sum_z p_z \sum_x f_{zx}(xN_{1,x-1}).$$

The chance of having no granddaughters is the chance that each of ego's x daughters (each of whom has x-1 sisters) will have no daughters:

$$\sum_z p_z \sum_x f_{zx} (f_{x-1,0})^x .$$

It is more difficult to calculate the expectation of granddaughters squared. The probability generating-function of the number of granddaughters, *given* that ego has exactly x daughters, is

$$h_x(s) = [f_{x-1}(s)]^x,$$

where $f_{x-1}(s)$ is the probability-generating function of the number of daughters born to women with $x-1$ sisters. Also

$$E(Y^2|X=x) = E(Y|X=x) + E[Y(Y-1)|X=x];$$

that is, $E(Y^2|X=x)$ will be the sum of the first and second derivatives of $h_x(s)$ evaluated at $s=1$. This calculation leads to

$$E(Y^2|X=x) = xN_{2,x-1} + x(x-1)N_{1,x-1}{}^2.$$

Weighting by the probabilities that a woman with z sisters will have x daughters, we obtain

$$E(Y^2) = \sum_z p_z \sum_x f_{zx}[xN_{2,x-1} + x(x-1)N_{1,x-1}{}^2].$$

To obtain the expected value of the product of daughters and granddaughters, note that if ego has x daughters, the expected product of daughters and granddaughters will be x times $xN_{1,x-1}$, or

$$E(XY|X=x) = x^2N_{1,x-1}.$$

Therefore

$$E(XY) = \sum_z p_z \sum_x f_{zx}(x^2N_{1,x-1}).$$

The correlation is then calculated in the usual way, from

$$Var(X) = E(X^2) - E(X)^2,$$
$$Var(Y) = E(Y^2) - E(Y)^2,$$
$$Cov(X,Y) = E(XY) - E(X)E(Y), \text{ and}$$
$$r = Cov(X,Y) / \sqrt{Var(X)\ Var(Y)}.$$

When the reproductive regime does not depend upon sibship size (i.e. $f_{zx} = f_x$ for all z), it can be shown that all of these formulas, and those given below for sisters, are consistent with those for Model 1.

Sisters and Daughters

The probability that ego has z sisters was defined earlier to be p_z. Therefore the expected number of sisters is

$$E(Z) = \sum z\ p_z$$

and the expected number of sisters squared is

$$E(Z^2) = \Sigma z^2\, p_z.$$

The remaining quantity required for the correlation between sisters and daughters is $E(ZX)$. Because

$$\Pr(Z = z \text{ and } X = x) = \Pr(X = x | Z = z)\Pr(Z = z) = f_{zx}\, p_z,$$

it follows that

$$E(ZX) = \sum_{z,x} zx\, f_{zx}\, p_z = \sum_z z\, p_z \sum_x x f_{zx} = \sum_z z\, p_z\, N_{1z}\,.$$

The correlation is then calculated from

$$Var(Z) = E(Z^2) - E(Z)^2,$$
$$Var(X) = E(X^2) - E(X)^2,$$
$$Cov(Z,X) = E(ZX) - E(Z)E(X), \text{ and}$$
$$r = Cov(Z,X) / \sqrt{Var(Z)\, Var(X)}.$$

Observe that in this model, this correlation is determined wholly by the following quantities: 1) the expected parity of females with z sisters, N_{1z}; 2) the expected squared parity of females with z sisters, N_{2z}; and 3) the probability that ego has z sisters, namely p_z. Thus it is not necessary to specify—at least in order to compute the correlation—the full distribution of conditional probabilities f_{zx}. By contrast, the correlation between daughters and granddaughters requires detailed specification of the reproductive regimes because $E(Y)$, $E(Y^2)$, and $E(XY)$ involve links among three successive generations.

Numerical Estimates Under Model 2

To estimate the value that these correlations can take plausibly in empirical settings, assume that each of the reproductive regimes $\{f_{zx}\}$ has a Poisson distribution, with parameter and mean N_{1z}. Also assume that the link between the actual number of ego's sisters and the expected number of her daughters takes the simple form of the linear relationship

$$N_{1z} = a + bz,$$

where $b \geq 0$. That is, an increase of one sister will lead to an average increase of b daughters, but will affect the entire distribution of daughters by altering the parameter of the Poisson distribution. The correlation between daughters and granddaughters, and between sisters and daughters, can be calculated with alternative values of the slope b to express the socialization effect.

The parameter a is constrained because in ego's generation,

$$\Sigma p_z N_{1z} = \Sigma p_z(a + bz) = a\Sigma p_z + b\, \Sigma p_z z = a + b\, \Sigma p_z z.$$

Now $\Sigma p_z N_{1z}$ is the overall mean number of daughters born to ego's generation (i.e. N), and $\Sigma p_z z$ is the overall mean number of sisters in ego's generation (i.e. S). Therefore

$$a = N - bS,$$

so an alternative statement of the linking equation is

$$N_{1z} = N + b(z - S).$$

Within each class the reproductive regime is assumed to be Poisson with the implication that if an ego in class z has (on average) N_{1z} daughters, then her daughters will have (on

average) N_{1z} sisters. This use of Poisson distributions does not imply that in ego's generation the pooled expected numbers of daughters (N) and of sisters (S) will be equal. A mixture of Poisson distributions will not itself be Poisson.

Empirical values of N, S, and the distribution $\{p_z\}$ come from a distribution of probabilities used by Pullum (1982), describing a real cohort of women in the United States in 1978 with expected fertility at a replacement level. Referring to the probability of having x daughters as f_x, we take

$$p_z = \Sigma(z+1)f_z/ \Sigma zf_z = \Sigma(z+1)f_z$$

because

$$\Sigma zf_z = N = 1.$$

The mean number of sisters is S = .835. These probabilities are given in Table 2. The distribution of daughters is more concentrated than a Poisson, but the departure is not great; it can be calculated that a shift of less than 6% of the population would give a perfect fit to a Poisson. In a Poisson distribution with mean 1.000, the probability of no daughters would be .368, somewhat more than the probability .336 in the illustrative distribution.

We will make no effort to reconcile this empirical distribution with the class-specific reproductive regimes $\{f_{zx}\}$. Our purpose is to estimate numerically the sensitivity of the correlations to changes in the value of the coefficient b; we do not propose to generate the stable distributions that would result over the long term if reproductivity remained constant.

Table 3 gives the mean number of daughters for women whose number of sisters ranges from $z = 1$ through $z = 6$ when b ranges from 0.0 to 0.5 in increments of 0.1, implied by the equation

$$N_{1z} = N + b(z - S).$$

As b increases, the mean number of children to women socialized in large families is shifted upward, but if the population is not to increase, the mean fertility of only daughters—that is, of those women with no sisters—must decline. These are the only women from below-average sibships and therefore the only ones who can balance the fertility of those from above-average sibships. This point deserves emphasis. A model of inheritance which proposed that the mean number of children in the second generation was intermediate between the overall mean and the number in the first generation would require that the

Table 2. Illustrative Probability Distributions of Daughters and Sisters in a Stationary Population. (Expected fertility of women aged 25-29 in the June 1978 Current Population Survey, converted to females only)

x or z	Probability of x daughters	Probability of z sisters
0	.336	.395
1	.395	.426
2	.213	.137
3	.046	.032
4	.008	.007
5	.001	.002
6	.000	.000
Mean	1.000	.835
Variance	.815	

Source: U.S. Bureau of the Census 1979, p. 24.

Table 3.   Illustration of Model 2; Mean Number of Daughters for Women with
Specified Numbers of Sisters, with $N = 1$ and $S = .835$

	Number of Sisters					
Slope b	0	1	2	3	4	5
0.0	1.00	1.00	1.00	1.00	1.00	1.00
0.1	0.92	1.02	1.12	1.22	1.32	1.42
0.2	0.83	1.03	1.23	1.43	1.63	1.83
0.3	0.75	1.05	1.35	1.65	1.95	2.25
0.4	0.67	1.07	1.47	1.87	2.27	2.67
0.5	0.58	1.08	1.58	2.08	2.58	3.08

population was increasing. A model will be compatible with stationarity only if women with no siblings tend disproportionately to be childless. In a low-fertility population, any inheritance of fertility will act to increase the level of childlessness and the proportion of women with no kin in their old age.

Using the equations given earlier and the illustrative stationary regime, we can calculate the correlation between daughters and granddaughters, and between sisters and daughters, implied by different values of b. The results are presented in Table 4. Although the mean number of daughters is constrained to be 1.000 and the mean number of sisters is constrained to be .835, Table 4 shows a slight increase in the mean number of granddaughters as b increases, due to small departures of the distribution $\{p_z\}$ from the stable distribution implied by the regimes $\{f_{zx}\}$. Table 4 also gives the expected proportions of women who will have no daughters or no granddaughters for different levels of inheritance. The probability of no daughters depends on the Poisson approximation and therefore differs from .336, the proportion with no daughters shown in Table 2.

Perhaps the most important conclusion to be drawn from Table 4 is that substantial changes in the correlation between sisters and daughters, indexed either by the slope b or by r_{xz}, will have relatively little impact on the correlation between daughters and granddaughters or on the probability of having no daughters. As r_{xz} ranges between .00 and an implausibly high level of .39, r_{xy} increases only by .09, rising from .71 to .80, and the probability of having no daughters increases by only .03. The chance of having no granddaughters is increased by about .08.

As a final observation in this section, Model 2 assumes that ego's reproductive regime depends stochasticically upon an observable characteristic of ego's mother—namely, her eventual parity. In this model, daughters from large families would tend to have large

Table 4.   Consequences of Reproductive Socialization for the Numbers of Daughters X,
Granddaughters Y, and Sisters Z. (For all values of b, $E(X) = N = 1.000$ and
$E(Z) = S = .835$)

Slope b	r_{xy}	r_{xz}	$E(Y)$	$Pr(X = 0)$	$Pr(Y = 0)$
0.0	.71	.00	1.00	.37	.53
0.1	.74	.09	1.02	.37	.54
0.2	.76	.17	1.04	.37	.56
0.3	.78	.25	1.07	.38	.57
0.4	.79	.32	1.10	.39	.59
0.5	.80	.39	1.16	.40	.61

families themselves and daughters with no siblings would tend to be childless. It would be difficult, however, to distinguish the outcome of such a process from an alternative in which the reproductive regime was determined by an unobservable characteristic of the mother, such as her fecundity, which could be transmitted genetically rather than socially.

Some Empirical Correlations

We now present several empirical correlations for selected pairs of kin types and from several countries. As noted earlier, several studies have considered transmission of fertility patterns across generations. Generally the results were conveyed in the form of coefficients from a multiple regression (e.g., in Danziger and Newman 1989 or Hodge and Ogawa 1986) or comparisons of group means (e.g., in Anderton et al. 1987) rather than in the form of a zero-order, product-moment correlation coefficient. Here we present correlation coefficients and are not restricted to sibling/offspring kin pairs.

The correlations presented here are subject to the limitations and qualifications discussed at the outset. They are gathered from surviving members of a population and (with few exceptions) pertain to counts of currently living kin. The figures cannot be viewed as evidence for or against any particular theory about population dynamics, but rather are to be regarded as illustrative and possibly suggestive facts to guide further investigations.

Figures are presented from five surveys: the Hungarian Microcensus of 1984 (Hungarian CSO 1986), the U.S. Supplement on Aging to the National Health Interview Survey of 1984 (Kovar 1986), Canada's 1985 General Social Survey, the 1981-82 German Life History Survey (Tuma and Huininck 1987), and the 1974 Israeli Labor Mobility Survey (Danziger and Newman 1989). In the first four of these surveys, it is possible to group ego by age. The types of kin considered include children, grandchildren, siblings, sisters, and daughters.

The empirical correlations are presented in Table 5, which gives first the correlations between numbers of children and numbers of grandchildren for several age groups of older Hungarian women. In the age groups under 70 the acquisition of grandchildren is undoubtedly incomplete; in the oldest groups, problems due to selective mortality may be operating. Even so, the observed correlations fall within a fairly narrow range.

In Model 1, which assumed no inheritance, the correlation between children X and grandchildren Y was given as

$$r = [E(Y)/E(X)][\sqrt{Var(X)/Var(Y)}].$$

Table 6 compares the observed correlations in Panel A of Table 5 with those implied by the means and variances of the numbers of children and grandchildren if there were no inheritance. For the first three age groups in particular, the correspondence is extremely close. After age group 60-64 the deviation is negative; that is, the observed correlation is less than the expected correlation. This anomaly is probably due to cumulative mortality in the children's generation, and points to the role of age and survivorship in the data. Because of these confounding factors, we will not proceed to an actual test of independence. (Even if a test were appropriate for these data, the deviations—though small—would be statistically significant because of the large sample sizes.)

Direct evidence on the correlation between numbers of sisters and daughters is provided for the United States in Panel B of Table 5. All the age groups of ego are past childbearing, but the contaminating effects of mortality (in both generations) remain. The correlations (with one peculiar exception, for the small group of women aged 90 +) are distinctly positive.

Much larger correlations between mother's and ego's fertility are revealed by the

Table 5. Selected Empirical Correlations between Kin Types from Various Countries ˙

A. Results from Hungarian Microcensus of 1984; Older Women; Children and Grandchildren

Age Group	Mean Number of Children	Variance Children	Mean Number of Grand-children	Variance Grand-children	Correlation Coefficient	N
55–59	1.9307	1.7531	2.5572	5.8281	0.7431	7444
60–64	1.9870	2.0949	2.9568	7.1700	0.8070	7239
65–69	1.9442	2.3018	3.0685	8.3768	0.8255	3725
70–74	1.9007	2.5138	3.0863	9.0399	0.8337	5259
75–79	1.8434	2.6997	3.0874	9.8986	0.8349	3467
80–84	1.8339	2.7609	3.2100	10.7975	0.8039	1957
85–89	1.6773	2.3658	2.9902	10.2723	0.7859	815
90 +	1.6449	1.9300	3.2757	11.7231	0.7043	214

B. Results from U.S. SOA File of 1984; Older Women; Sisters and Daughters

Age Group	Mean Number of Sisters	Variance Sisters	Mean Number of Daughters	Variance Daughters	Correlation Coefficient	N
55–59	1.7015	3.6942	1.5103	1.7923	0.0916	1262
60–64	1.6394	2.7817	1.3991	1.8984	0.0246	1207
65–69	1.5717	2.4859	1.2398	1.6568	0.0686	2139
70–74	1.4638	2.1810	1.1811	1.8583	0.0738	1805
75–79	1.2357	1.9569	1.1485	1.8329	0.0825	1396
80–84	1.0265	1.5542	1.0609	1.9516	0.0710	839
85–89	0.8375	1.9440	1.1615	1.5760	0.0886	420
90 +	0.5460	0.8312	1.3664	2.2896	−0.1346	150

C. Results from the 1985 Canadian GSS; Older Women; Siblings and Children

Age Group	Mean Number of Siblings	Variance Siblings	Mean Number of Children	Variance Children	Correlation Coefficient	N
55–59	4.1469	10.9157	3.2030	3.6694	0.3138	315
60–64	3.6726	9.0147	3.2887	5.7285	0.3266	311
65–69	3.5634	9.2108	3.2593	5.6054	0.2407	401
70–74	3.4654	7.8613	2.8560	5.7521	0.1874	507
75–79	3.3790	9.5679	2.3742	4.8865	−0.0578	434
80–84	2.6147	5.4356	2.6742	5.7102	0.0943	362
85 +	1.8872	4.3967	2.5524	5.7265	0.1663	357

D. Results from 1981-1982 German Life History Survey; Women by Birth Cohort; Siblings and Children

Birth Cohort	Mean Number of Siblings	Variance Siblings	Mean Number of Children	Variance Children	Correlation Coefficient	N
1949–51	2.117	3.516	1.185	1.105	0.1519	729
1939–41	2.563	2.618	1.932	1.683	0.2127	727
1929–31	2.720	5.298	2.229	2.347	0.1552	709

(Continued)

Table 5. *(Continued)*

E. Results from 1974 Israeli Labour Mobility Survey; Currently-Married Women 35 and Older; Siblings and Children

Mean Number of Siblings	Variance Siblings	Mean Number of Children	Variance Children	Correlation Coefficient	N
3.768	6.062	3.428	7.089	0.286	708

Canadian, German, and Israeli data that complete Table 5. For these three countries, only the Canadian women and the earliest cohort of German women are beyond childbearing age. On the other hand, whereas the Hungarian, U.S., and Canadian data pertain only to currently living kin, the German and Israeli data correlate the number of children born to ego with the number of siblings ever born to ego's mother (i.e. ego's aunts and uncles). The former are relatively uninfluenced by mortality, which may be the reason for the high observed correlations. The correlation betwen ego's sisters and her aunts is not identical with the correlation between her sisters and her daughters and has not been discussed explicitly in this paper, except to note that a non-zero correlation would suggest some kind of reproductive inheritance.

Conclusion

One's pool of available kin constitutes a resource that influences other dimensions of life such as household composition and the provision of health and social support services. It is useful to know how changes in childbearing and survivorship are likely to affect the numbers of kin going beyond children. We have attempted here to describe the dynamics of kinship networks by using the notion of eventual expected numbers, self-contained analytic tools, and simple numerical assumptions. These methods have clarified the correspondences between frequencies in different kin categories. We have identified which categories will be correlated even in the absence of reproductive inheritance, and which will be correlated only if such inheritance exists.

The analytic approach has established relatively narrow ranges within which some

Table 6. Comparison of Observed Correlations (r) between Children and Grandchildren from the Hungarian Microcensus of 1984 and Correlations Expected (r_e) in Absence of Inheritance

Age Group	Observed r	Expected r_e	$r - r_e$
55–59	0.7431	0.7264	0.0167
60–64	0.8070	0.8044	0.0026
65–69	0.8255	0.8273	−0.0018
70–74	0.8337	0.8563	−0.0226
75–79	0.8349	0.8747	−0.0398
80–84	0.8039	0.8851	−0.0812
85–89	0.7859	0.8555	−0.0696
90 +	0.7043	0.8080	−0.1037

correlations must lie. In particular, the correlation between daughters and granddaughters will tend to increase if the population is increasing, or if inheritance of reproductivity is present (through alternative mechanisms), but it will rarely be observed outside a range from about .71 to about .83 and is not a sensitive index of whether inheritance actually occurs. Reproductive inheritance will increase the proportion of kin who have no daughters or no granddaughters, but numerical analysis suggests that a high level of inheritance would be required for this source of heterogeneity to increase the level of childlessness by even a few percentage points.

It could be very difficult to determine whether a correlation between sisters and daughters was produced by inheritance (broadly interpreted) of factors that affect fertility, or by socialization into a particular family size. Others (particularly Anderton et al. 1987) have concentrated on correlations between the fertility of mothers and of their daughters, and have speculated on the mechanisms by which observed correlations arise.

Empirical findings suggest the presence of correlations between kin counts in several countries. The interpretation of these data is affected by the roles of age and survivorship, but there is some evidence of inheritance. We hope that future data sets will permit more specific judgments about reproductive continuities and the importance of socialization.

References

Anderton, D.L., N.O. Tsuya, L.L. Bean, and G.P. Mineau. 1987. "Intergenerational Transmission of Relative Fertility and Life Course Patterns." *Demography* 24: 467–80.

Atkins, J.R. 1974. "On the Fundamental Consanguineal Numbers and Their Structural Basis." *American Ethnologist* 1: 1–31.

Bongaarts, J. 1987. "The Projection of Family Composition over the Life Course with Family Status Life Tables." Pp. 189–212 in *Family Demography: Methods and their Application*, edited by J. Bongaarts, T. Burch and K. Wachter. Oxford:Oxford University Press.

Bongaarts, J., T. Burch, and K. Wachter (eds.). 1987. *Family Demography: Methods and their Application*. Oxford: Oxford University Press.

Berent, J. 1953. "Relationship between Family Sizes of Two Successive Generations." *Milbank Memorial Fund Quarterly Bulletin* 31: 39–50.

Broberg, P. 1987. "Sibling Dependencies in Branching Populations: Effects on Growth, Composition and Extinction Probability." Department of Mathematics, Gothenburg University.

Danziger, L. and S. Newman. 1989. "Intergenerational Effects on Fertility: Theory and Evidence from Israel." *Journal of Population Economics* 2: 25–38.

Duncan, O.D., R. Freedman, J.M. Coble, and D.P. Slesinger. 1965. "Marital Fertility and Size of Family of Orientation." *Demography* 2: 508–15.

Goodman, L.A., N. Keyfitz, and T.W. Pullum. 1974. "Family Formation and the Frequency of Various Kinship Relationships." *Theoretical Population Biology* 5: 1–27.

————. 1975. "Addendum." *Theoretical Population Biology 8: 376–81*.

Hammel, E.A., K.W. Wachter, and C.K. McDaniel. 1979. "The Kin of the Aged in the Year 2000." *The Elderly of the Future*, edited by E.S. Morgan.

Harris, T.E. 1963. *The Theory of Branching Processes. Englewood Cliffs, NJ: Addison-Wesley*.

Hendershot, G.E. 1969. "Familial Satisfaction, Birth Order, and Fertility Values." *Journal of Marriage and the Family* 31: 27–33.

Hodge, R.W. and N. Ogawa. 1986. *Siblings and Family Size From Generation to Generation*. Tokyo: Population Research Institute, Nihon University.

Huestis, R.R. and A. Maxwell. 1932. "Does Family Size Run in Families?" *Journal of Heredity* 23: 77–79.

Hungarian Central Statistical Office. 1986. "Data of the 1984 Hungarian Microcensus." Budapest.

Johnson, J. and C.S. Stokes. 1976. "Family Size in Successive Generations: The Effects of Birth Order, Intergenerational Change in Life Styles, and Familial Satisfaction." *Demography* 13: 175–87.

Kantner, J.F. and R.G. Potter. 1954. "Social and Psychological Factors Affecting Fertility: XXIV. The Relationship of Family Size in Two Successive Generations." *Milbank Memorial Fund Quarterly* 32: 294–311.

Kovar, M.G. 1986. *Aging in the Eighties, Age 65 Years and Over and Living Alone, Contacts with Family, Friends, and Neighbors*. Hyattsville, MD: Public Health Service.

Langford, C.M. and C. Wilson. 1985. "Is There a Connection Between a Woman's Fecundity and That of her Mother"? *Journal of Biosocial Science* 17: 437–443.

Le Bras, H. 1973. "Parents, Grands-parents, et Bisaieux." *Population* 28: 9–38.

Pearson, K. and A. Lee. 1899. "On the Inheritance of Fertility in Mankind." *Philosophical Transactions of the Royal Society of London, Series A* 192: 282–330.

Preston, S.H. 1976. "Family Sizes of Children and Family Sizes of Women." *Demography* 13: 105–14.

Pullum, T.W. 1982. "The Eventual Frequencies of Kin in a Stable Population." *Demography* 19: 549–65.

———. 1987. "Some Mathematical Models of Kinship and the Family." In *Family Demography: Methods and their Application*, edited by J. Bongaarts, T. Burch and K. Wachter. Oxford: Oxford University Press.

Reeves, J.H. 1987. "Projection of Number of Kin." In *Family Demography: Methods and their Application*, edited by J. Bongaarts, T. Burch and K. Wachter. Oxford: Oxford University Press.

Smith, J.E. 1987. "The Computer Simulation of Kin Sets and Kin Counts." In *Family Demography: Methods and their Application*, edited by J. Bongaarts, T. Burch and K. Wachter. Oxford: Oxford University Press.

Tuma, N.B. and J. Huininck. 1987. "Post-war Fertility Patterns in the Federal Republic of Germany." In *Applications of Event History Analysis in Life Course Research*, edited by K. U. Mayer and N. B. Tuma. Berlin: Max-Planck-Institut fur Bildungsforschung.

U.S. Bureau of the Census. 1979. "Fertility of American Women: June, 1978." Washington, DC: U.S. Government Printing Office.

Waugh, W.A. O'N. 1981. "Application of the Galton-Watson Process to the Kin Number Problem." *Advances in Applied Probability* 13: 631–49.

Wolf, D.A. 1988. "Kinship and Family Support in Aging Societies." In *Economic and Social Implications of Population Aging*. New York: United Nations Department of International Economics and Social Affairs.