THEORETICAL POPULATION BIOLOGY 8, 376–381 (1975)

Addendum

Family Formation and the Frequency of Various Kinship Relationships*

LEO A. GOODMAN,

University of Chicago

NATHAN KEYFITZ,

Harvard University

AND

THOMAS W. PULLUM

University of California, Davis

Received June, 1975

1. INTRODUCTION

In this addendum, we shall consider in more detail an assumption arising in the application of our formulas for expected kin other than direct descendants and direct progenitors. The derivation of these formulas involved, in part, consideration of the joint probability that a woman would give birth to daughters at each of two specified times. In applying these formulas, the joint probability $m(x, y)$ that a woman exposed to both age x and age y will have a daughter at age x and another at age y was assumed to be equal to the product of the two marginal probabilities. Our notation, which indicated that a woman's fertility depended solely on her age, also indicated (implicitly) an independence from any other events associated with her, such as another birth. This assumption is applied when a marginal maternity function is inserted in formulas that call for a conditional one. Our formulas are correct as they presently stand when a conditional maternity function is used. In this addendum, we comment briefly

* *Theor. Pop. Biol.* 5, 1–27. This Addendum was stimulated by correspondence and discussion with J. Hobcraft, J. Menken, T. Schelling, R. Solow, K. Wachter, and others.

upon the application of these formulas in which a conditional maternity function is replaced by a marginal one. That is, we comment upon the application of the assumption of independence described above.

More specifically, consider formula (3.1.a) in our paper. In the exposition preceding this formula, we considered the situation where the girl (ego) was born when her mother was age x. Thus, the maternity function m_y in the formula denotes the conditional maternity function for the mother at age y, given that she gave birth to a girl (ego) when she was age x. Formula (3.1.a) is correct as it presently stands when the conditional maternity function is used. (The more explicit conditional notation $m_{y|x}$ might be preferable here.) In the application of this formula, if the conditional maternity function is replaced by a marginal one, then the assumption of independence noted above is being applied.

We shall comment briefly on two aspects of this assumption. First, we note some of the ways in which the assumption may be inappropriate. Second, we attempt to evaluate the extent of the possible error obtained when the assumption is applied to the formulas for expected sisters and we suggest ways of reducing this error.

2. Validity of the Assumption

As noted on page 13 of our article, it is biologically impossible for two births to be less than approximately 12 months apart, because of the period of sterility following each delivery and the period of gestation prior to the next one. Thus, the bivariate maternity function $m(x, y)$ will have the value zero within (approximately) the strip $| x - y | < 12$ months. An analogous statement could be made about the conditional function. Within this strip the assumed factorability of $m(x, y)$ is contradicted.

Second, the conditional maternity function applies only to those women who had a child when age x, say, and by that circumstance these women are known to be fertile. We can be certain that the maternity function for women who will eventually have at least one child is different from the function for all women.

Third, there may be other relevant kinds of heterogeneity. That is, there may be subpopulations with their own distinctive maternity functions. (Of course, under the model of the stable population, such a circumstance would require that all subpopulations have the same intrinsic rate of growth and that each subpopulation be a fixed proportion of the total.) However, one can show that to have an impact, heterogeneity must involve the *shape* of the maternity function. Variation between subpopulations according to a *nonzero* multiplier (such as the GRR), which is independent of age will not, of itself, invalidate the assumption. Thus, for example, heterogeneity in completed family size will only affect the assumption if different family sizes are associated with maternity functions of different shapes. There would be an effect if, for example, the population was stratified according to age at marriage and the onset of reproduction, as might occur across ethnic and socioeconomic subgroups.

Fourth, it is possible for the assumption to be invalidated even if the population is homogeneous. For example, if there is autocorrelation in the birth sequence, which would result from intentionally regulated spacing in human populations (but occurs naturally in some animal populations, such as the black bear, which ordinarily mates only in alternate years), then the assumption will be incorrect.

3. Evaluation of the Possible Error

We shall now attempt to evaluate the possible error obtained when the assumption under discussion is applied to our formulas. Attention will be restricted to formulas for the expected number of sisters. Even more specifically, we shall focus on a single measure: the eventual number of sisters a woman would have if there were no mortality prior to the last age of childbearing.

Let us ignore age and mortality. It is clear that the eventual expected number of sisters could be calculated directly if we knew the distribution of completed sororities, i.e., of the eventual number of daughters born to a woman. We would proceed as follows. Define f_i to be the fraction of women who have i daughters. The expected number of daughters to a randomly selected woman would then be $G = \sum if_i$ and the variance in the number of daughters would be $\sigma^2 = \sum i^2 f_i - G^2$.

The eventual expected number (S) of *sisters* to a randomly selected woman would depend on the chance if_i/G that she was a member of a sorority of size i, in which case she would have $i - 1$ sisters, so that $S = \sum (i - 1)\, if_i/G$. It is then easily shown that $S = G - 1 + \sigma^2/G$. That is, we need only the mean and variance of the distribution of completed sororities to know the eventual expected number of sisters, and to know it exactly.

By contrast, when our formulas are applied to calculate the eventual expected total number of sisters (adding older sisters (3.1.a) and younger sisters (3.2.a) as given in the previous paper) in the case where the independence assumption is invoked, we find that the expected number of sisters cannot exceed the gross reproduction rate G. In particular, in an hypothetical population with no mortality, and under the independence assumption, application of our formulas yields the expected number $S = G$. From the preceding paragraph, we see that this will be the correct result only if $\sigma^2 = G$, as is the case when the completed sororities have a Poisson distribution, although these equalities may obtain more generally. (We emphasize that G, S, and σ^2 all pertain to an hypothetical population with no mortality prior to the last age of childbearing and that S is the number of sisters *ever born.*)

Thus, we have a quantity for which an exact calculation is possible (when G and σ^2 refer to an hypothetical population as noted above) and to which our formulas can be applied, obtaining an approximation when the independence assumption is invoked. This quantity, the eventual expected number of sisters when mortality is ignored, is of little intrinsic interest; in general, mortality is not negligible. Our interest lies in the departure of the approximation (G) from the exact value ($G - 1 + \sigma^2/G$), as a relatively simple measure of the accuracy of the approximation for sisters (when mortality is ignored). Some (although not all) of any discrepancy can be attributed to the assumption under discussion.

We now attempt to estimate numerically the level of this departure, using observed frequency distributions of the number of children ever born to women who have survived the childbearing ages. Because of their availability and accuracy, our data come from a half-century in England and Wales when "the strategy of family formation changed" as "fertility fell so drastically" (Wrigley, 1969, p. 199). (Our formulas were developed for stable populations. However, if formulas (3.1.a) and (3.2.a) are added together, the sum will not depend upon $W(x)$ so long as the l_x in these formulas are ignored and the numerical value of a is sufficiently large. These two conditions are equivalent to our assumption of "no mortality" and our reference to the "eventual number.")

As occurs with nearly all published completed family size distributions, these data are available for ever-married women only and with no distinction between sons and daughters. We must first adjust our formulas for data of this nature. Let p be the proportion of women in a cohort who marry by the last reproductive age and let m_1 and m_2 be the first and second moments of the distribution of children ever born (by the last reproductive age). Assume that for all births there is the same binomial probability that the child will be a girl (see Pullum, 1975) and, for simplicity, that this probability is 0.5. Then it can be shown that $G = pm_1/2$ and $-1 + \sigma^2/G = (m_2/m_1 - pm_1 + 1)/2$.

Table I is based on data presented by Wrigley (1969, p. 198) but originally described by Glass and Grebenik (1954, p. 87). Our estimates of p come from Cox (1970, p. 341). The table shows that for the considerable range in family size distributions that these data represent, and for a highly heterogeneous society, the approximation would yield numerical values that are low by as much as one full sister. In terms of relative error, the approximation is inaccurate for these data by an amount ranging from 29 to 36%. These figures clearly indicate that the assumption of independence can have a substantial effect on the quantity we have singled out.

We have shown that if mortality during the childbearing years is negligible, then the eventual expected number of sisters can be obtained exactly by adding a correction, $-1 + \sigma^2/G$, to the earlier approximation, namely, G (here G and σ^2

TABLE I

Stötistics derived from the completed distributions of children ever born for five marriage cohorts, England and Wales.

	Year of Marriage				
	1870–1879	1890–1899	1900–1909	1915	1925
p	0.86	0.84	0.84	0.84	0.84
p^*	0.21	0.24	0.25	0.29	0.30
m_1	5.82	4.30	3.34	2.50	2.21
m_2	46.74	28.10	17.83	10.64	8.54
1st approximation: G	2.50	1.81	1.40	1.05	0.92
2nd approximation: $\theta/2$	3.16	2.36	1.82	1.39	1.20
Exact value: $G - 1 + \sigma^2/G$	3.52	2.78	2.16	1.63	1.43
Discrepancy of first approximation: $-1 + \sigma^2/G$	1.02	0.97	0.76	0.58	0.51
Relative error of first approximation	0.29	0.35	0.35	0.36	0.36
Relative error of second approximation	0.10	0.15	0.16	0.15	0.16

refer to an hypothetical population as noted above). If σ^2 is not known, then some improvement in the approximation generally can be made so long as the proportion of childless women is known. We noted earlier that (a) ego is evidence of a fertile mother, that fertile women will have a different maternity function than all women combined, and that (b) our approximation G will be correct, i.e., will equal S, if completed families have a Poisson distribution.

We may combine these observations by hypothesizing two categories of women. The first, a proportion p^* of the total, consists of women with one or more children, distributed according to a truncated Poisson with parameter Θ. The second, a proportion $1 - p^*$ of the total, consists of women who have no children at all, for any reason, including celibacy. We have intentionally referred to children here, rather than daughters, because the proportion childless is usually easier to obtain than the proportion daughterless.

Again assume, for convenience only, that exactly half of all children are daughters. It is then easily shown that the eventual expected number of sisters will be $\Theta/2$, where Θ is the solution to the equation

$$\frac{\Theta}{1 - e^{-\Theta}} = \frac{2G}{p^*}.$$

That is, we would "correct" G, our initial approximation, by multiplying it by $(1 - e^{-\Theta})/p^*$, where Θ is the root of the preceding equation.

We have made this calculation for the British data described earlier; p^* is equal to p less the proportion of all women who were married but childless. Table I presents our results and shows that for these data, this second approximation has a relative error of only 10 to 16%.

If the level of mortality is not negligible and/or if one is interested in the size of the sorority before it reaches its maximum, then one must use age-specific maternity functions.

Earlier, we considered the situation in which the discrepancy between the conditional maternity function and the marginal maternity function is due to heterogeneity. That is, there may be two or more stable subpopulations with distinct maternity functions of different shapes; for example, the marginal function may result from a mixture of two or more Poisson distributions. As mentioned earlier, the assumption of stability requires that all groups have the same intrinsic growth rate and that each group be a fixed proportion of the total. If (a) these distinct maternity functions are known and (b) the proportion of women who follow any given one of these functions is known, then our formulas can be modified in a straightforward way. The expected number of sisters would then be simply a weighted average of the expectation within each subpopulation. Similarly for nieces, aunts, and cousins. These weighted forms will always yield a larger expected value than the unweighted forms appearing in our paper.

In conclusion, we reiterate that the formulas in our paper are correct as they stand so long as the maternity function is understood to be conditional upon a known birth. Lacking conditional maternity functions, we illustrated the formulas in our paper using marginal functions. If there is in fact a dependence between childbearing at one age and childbearing at another age, which cannot be incorporated by adjustments of the sort described above, then the conditional functions can be estimated by using birth histories and by shifting one's orientation from period to cohort input data. More detailed work is required to give a firm judgment of the value of the extra effort.

References

Cox, Peter R. 1970. "Demography," Cambridge University Press.

Glass, D. V. and Grebenik, E. 1954. "The Trend and Pattern of Fertility in Great Britain: A Report on the Family Census of 1946," Papers of the Royal Commission on Population, vol. VI, H.M.S.O., London.

Pullum, Thomas W. 1975. On the analysis of family continuation probabilities which are composition-specific, *J. Math. Biosci.*, **24**, 93–105.

Wrigley, E. A. 1969. "Population and History," Weidenfeld and Nicolson, London.