Population Studies, **46** (1992), 507–522
Printed in Great Britain

Migration, Marriage, and Mortality: Correcting Sources of Bias in English Family Reconstitutions*

STEVEN RUGGLES†

Evaluations of the reliability of family reconstitution methods have stressed the potential for migration to bias the results. Family reconstitution is the process of linking together historical parish records of baptisms, marriages, and burials; it yields a set of demographic life-histories from which rates can be calculated. People who moved between parishes scattered their demographic life-histories across the countryside. Since these life-histories cannot usually be re-assembled, they must be excluded from most demographic analyses.

Most of the concern about the effects of the exclusion of migrants has focused on the question whether demographic behaviour of migrants and non-migrants was similar, or not.[1] It has been less commonly noted that migration can bias estimates of such measures as mean age at marriage and life expectancy, even if age-specific demographic rates of migrants and non-migrants were identical.

I first became aware of this problem some years ago, in the course of analysing the demography of eighteenth-century Yale graduates. The data consisted of virtually complete information about the mortality of both migrants and non-migrants, and I found to my initial surprise that the life expectancies of migrant Yale graduates were substantially higher than those of non-migrants.[2] On reflection, the reason was obvious: the longer the graduates lived, the greater the odds that they would eventually migrate.

The same mechanism applies in family reconstitutions. Consider the case of age at marriage. The English family reconstitutions limit the analysis of marriage age to those

* My thanks to Jim Oeppen, who was willing to discuss every point in excruciating detail. Table 5 was prepared by Jim Oeppen and Rosalind Davies. I am also grateful to Bob McCaa, Roger Schofield, Nancy Shoemaker, David Weir, and Tony Wrigley for their helpful comments and suggestions.

† Department of History, University of Minnesota, USA.

[1] T. H. Hollingsworth, *Historical Demography* (Ithaca, New York, 1969), pp. 181–196; R. S. Schofield, 'Historical demography: some possibilities and some limitations', *Transactions of the Royal Historical Society*, 5th ser, 21 (1971), pp. 119–132; 'Representativeness and family reconstitution', *Annales de démographie historique* 1972 (Paris, 1972), pp. 121–125; Jacques Dupâquier, 'Problèmes de représentativité dans les études fondées sur la reconstitution des familles', *Annales de démographie historique* 1972 (Paris, 1972), pp. 82–91; Dezso Danyi, 'La Migration et les méthodes nominatives: l'exemple hongrois', *Annales de démographie historique* 1972 (Paris, 1972), pp. 69–82; T. H. Hollingsworth *et al.*, 'Discussion: representativeness and family reconstitution', *Annales de démographie historique* 1972 (Paris, 1972), pp. 127–146; D. Levine, 'The reliability of parochial registration and the representativeness of family reconstitution', *Population Studies*, **30** (1976), pp. 107–122; J. Knodel and E. Shorter, 'The reliability of family reconstitution data in a German village', *Annales de démographie historique* 1976 (Paris, 1976), pp. 115–153; S. Akerman, 'An evaluation of the family reconstitution technique', *Scandinavian Economic History Review*, **25** (1977); S. L. Norton, 'The vital question: are reconstituted families representative of the general population?', in B. Dyke and W. T. Morrill (eds), *Genealogical Demography* (New York, 1980); J. Rogers, 'Family reconstitution: new information or misinformation?' Reports From the Family History Group 7 (1988), Department of History, Uppsala University; John Knodel, *Demographic Behavior in the Past: A Study of Fourteen German Village Populations in the Eighteenth and Nineteenth Centuries* (Cambridge, 1988), pp. 461–502.

[2] The analysis of Yale graduates was based on the remarkable volumes produced by Franklin Bowditch Dexter, *Biographical Sketches of the Graduates of Yale College* (New Haven, 1885–1912).

who were married in their place of birth.[3] This is necessary because without both the baptismal and the marriage record it is usually impossible to determine age at marriage. The odds of migrating before marriage are greater for those who marry late than for those who marry early, just because they are at risk of migrating for longer. Thus, late marriages tend to take place after migration and are systematically excluded from analysis. The bias in mortality is analogous: The chance of migrating before death is highest for those who survive longest, so early deaths are overrepresented among the observed deaths. This censoring of late marriages and deaths can lead to substantial underestimates of age at marriage and life expectancy in family reconstitution studies.

In this paper I illustrate the downward bias in marriage age and life expectancy in family reconstitution data under pre-industrial English demographic conditions, and suggest methods for correcting the errors. In addition, I evaluate the procedures proposed by Alain Blum for the estimation of adult mortality.

Before proceeding, I should note that migration is not the only reason for failure to link baptisms, marriages, and burials in family reconstitutions. Faulty record-keeping, changes of name, and errors introduced in the process of reconstitution all contribute to the problem.[4] The particular reason for failure to link demographic events is relatively unimportant in the present context; all these problems could potentially censor observations. For the sake of convenience, I shall refer to all disappearances from the population as migration. One can think of it as migration out of the parish records, if not out of the parish.

MICRO-SIMULATION AND MIGRATION

To illustrate the bias resulting from migration-censoring and to test alternative solutions to the problem, I used a demographic micro-simulation model. Micro-simulation is an ideal laboratory for experiments of this sort. It provides complete demographic life-histories for an hypothetical population that shares the demographic behaviour of an observed population. Within a micro-simulation model, migration histories can be assigned randomly, in accordance with an assumed set of age-specific migration probabilities. We can compare the age at marriage and life expectancy of the population as a whole with that of the sub-population that would be included in a family reconstitution study. Since migration is assigned independently of all other demographic behaviour, age-specific marriage rates and age-specific death rates in the migrant population are the same as in the non-migrant population. Any systematic differences between the general population and the reconstitutioned sub-population can, therefore, be attributed to migration-censoring.

I should stress, however, that in real populations there are hidden interactions between migration and other demographic events, and since we lack data on such interactions, the micro-simulation cannot take these into account. Thus, the magnitude of migration-censoring within the model may differ from that in a real population. For this reason, micro-simulation cannot be directly used to correct the bias in family

[3] In the French reconstitutions, this restriction is not always necessary, because information on age can typically be obtained from death records. Therefore, the analysis of age at marriage can include persons who either were born locally or died locally, and this will tend to reduce the effects of migration censoring. See L. Henry, *Techniques d'analyse en démographie historique* (Paris, 1980), pp. 113–115.

[4] Failures to link can also occur because of ambiguity; for example, when two or more women's births can plausibly be linked to a marriage record, the reconstituter is instructed to make no link at all under the reconstitution rules employed by the Cambridge Group. On the other hand, in some reconstitution studies marriage indexes or searches of the registers surrounding parishes have been used to identify demographic events for migrants, so that some migrants are included in family reconstitutions, although the number of cases is apparently small.

reconstitutions; rather, it should be viewed as an experimental tool for developing and testing solutions to the problem.

I used the MOMSIM micro-simulation model, which has been documented elsewhere.[5] The demographic assumptions of the model, given in Table 1, were designed to approximate conditions in England during the first quarter of the eighteenth century, as described in Wrigley and Schofield's *Population History*.[6] Age-specific migration probabilities were calculated from data gathered by David Souden on the timing of migration for nearly 10,000 witnesses in English ecclesiastical courts between 1601 and 1707.[7] These probabilities are given in Table 2.

Table 1. *Assumed demographic characteristics for simulations*

Mean age at first marriage (women)	26.0
Mean age at first marriage (men)	27.1
Percentage of females who never marry	12.3
Women's life expectancy at 20	36.6
Men's life expectancy at 20	37.3
Total fertility	4.4
Mean generation length	31.4

Table 2. *Assumed annual migration probabilities*

Ages	Males	Females
16–20	0.0910	0.0684
21–25	0.0772	0.1040
26–30	0.0594	0.0730
31–35	0.0380	0.0522
36–40	0.0262	0.0280
41–45	0.0180	0.0248
46–50	0.0180	0.0198
51–60	0.0128	0.0162
61+	0.0114	0.0128

Source: see note 7.

The migration experience of witnesses in ecclesiastical courts, of course, may not have been representative of the general population. However, the migration probabilities in Table 2 yield overall migration frequencies that fall within the general range observed for English reconstitutions of the period. In Table 3 the level of migration in the family reconstitutions is compared with that produced by the micro-simulation. These figures are expressed in terms of Family Reconstitution Forms (FRFs). In family reconstitution studies, a new FRF is created whenever a marriage takes place, or whenever there is a birth that cannot be linked to an existing FRF. Thus, if all parishes were reconstituted, a couple would generate a new FRF when they married, and an additional FRF each time they moved, and then had a child in the new location.

[5] S. Ruggles, *Prolonged Connections*: *The Rise of the Extended Family in Nineteenth Century England and America* (Madison, Wisconsin, 1987), pp. 84–105, 156–184. Some refinements of the remarriage procedure have been incorporated in the present version of the model.

[6] E. A. Wrigley and R. S. Schofield, *The Population History of England, 1541–1871*: *A Reconstruction* (Cambridge, Mass.: 1981).

[7] D. Souden, 'Pre-industrial English Local Migration Fields', Ph.D. Dissertation, Cambridge University, 1981, pp. 91, 114. Souden did not provide data on migration for persons under the age of 16; in the absence of evidence, I made the arbitrary assumption that annual migration rates for children came to one per cent.

Table 3. *Measures of migration in 14 English family reconstitutions and simulated populations*

(A) Percentage migrating before and after FRFs in English reconstitutions

	All parishes	Highest	Lowest
Females			
Before FRF	79.2	86.7	71.4
After FRF	56.3	68.7	45.2
Males			
Before FRF	73.3	85.4	56.1
After FRF	55.1	70.7	40.1

(B) Percentage migrating before and after FRFs in micro-simulations

	Medium migration	High migration	Low migration
Females			
Before FRF	73.6	81.2	59.5
After FRF	54.6	63.6	41.9
Males			
Before FRF	74.1	83.4	59.2
After FRF	47.1	56.2	35.4

In Table 3 A we show data from 14 English family reconstitutions, tabulated by David Souden.[8] The row labelled 'Migrated before FRF' shows the percentage of FRFs in which the wife or husband is not linked to a baptism record; the row labelled 'Migrated after FRF' shows the percentage not linked to a burial record. In the first column the combined figures for all reconstitutions are shown, while the second and third columns show the maximum and minimum, respectively. These data suggest remarkably high migration rates: about three-fourths of the population migrated before their FRF, and in addition over half migrated afterwards.

Table 3 B provides comparable figures for the simulated population. The left column, the 'medium migration' model, is based on the unadjusted migration probabilities from Table 2. The 'high' and 'low' migration models were constructed by the simple expedient of inflating or deflating the age-specific migration probabilities by 35 per cent.

In general, the migration figures derived from the ecclesiastical court records are slightly lower than those found in the family reconstitutions. This tendency is especially pronounced for migration before FRF among women, and after FRF among men. Of course, in the simulated population there are no failures to link on account of name changes or faulty record keeping.[9] On the whole, the level of migration in the simulation is surprisingly close to that of the family reconstitutions; taken as a group, the low, medium, and high simulated migration models provide a plausible approximation of the extent of migration in the 14 English parishes.

[8] *Ibid.* pp. 191–192. These are not ideal measures of migration, since the denominator (the number of FRFs) as well as the numerator is affected by migration; the more migration in a population, the more FRFs there will be, other things being equal. Moreover, the timing of the creation of FRFs is sensitive to a variety of factors, including the level of migration. For a more subtle approach to the analysis of migration in family reconstitutions, see Equation (3) below.

[9] Errors in the frequency of migration could also result from interaction between marriage and migration in the real population; see the discussion below. Moreover, some additional understatement of migration was introduced in the calculation of migration probabilities. Souden expressed these as the percentage of persons who had arrived within the previous five years, and I simply divided by five to obtain annual estimates. This ignores the possibility that people migrated more than once during the previous five years, and thus understates the migration probabilities.

MORTALITY, MIGRATION AND AGE AT MARRIAGE

To assess the effects of censoring on the mean age at first marriage I began with a simulated population that consisted of about 50,000 married couples. Since micro-simulation works by randomly assigning characteristics in accordance with assumed probability distributions, it is necessary to generate large populations to reduce random error. I used the same population for the high, medium, and low migration models, so that any differences between the models can be attributed entirely to migration.

The first row of Table 4 shows the overall mean age at first marriage in the population. Since these figures are based on the entire population, they are unaffected by migration censoring. They are, however, affected by mortality censoring. Some people who would have eventually married die before they get the chance, and the odds of dying first are greatest for those who would have married late. In the twentieth-century industrial societies, the effect of mortality censoring on age at marriage is trivial, since mortality rates in the years when nuptiality is at its peak are very low. Under the conditions of high mortality in pre-industrial Europe, however, mortality had a significant effect on age at marriage. If we wish to compare the reconstitution results to the usual mortality-free measures of age at first marriage – such as the singulate mean age at marriage (SMAM), or marriage ages calculated from vital statistics by means of a nuptiality table – we should correct for mortality-censoring as well as for migration-censoring.

The easiest way to correct for mortality-censoring is to limit the analysis to persons who survive to an age beyond which first marriages occur rarely. This strategy is followed in the second row of Table 4, which is restricted to the population who survive to age 50. Because of this restriction, the second row is equivalent to a mean age at marriage calculated from a nuptiality table, and is, therefore, directly comparable to most government statistics on age at marriage and to SMAMs. The effect of eliminating mortality censoring is to raise the mean age at marriage by just over half a year.

The next three rows of Table 4 are restricted to the population that would typically be used for an analysis of age at marriage in an English family reconstitution study.[10] These are the cases in which birth and marriage occurred in the same parish. The results are subject to both migration-censoring and mortality-censoring, and the effects are dramatic. Compared with the mortality-free measures of age at first marriage given in the previous row, the reconstitution estimates for women are understated by 2.6 years in the low-migration model, and rise to 4.5 years in the high-migration model. The range of error for men is 2.0–3.7 years. Thus, if there were no interaction between migration and marriage, reconstitutions restricted to marriages of the locally-born would underestimate age at marriage by at least two years, and probably more.

The simplest way of correcting the biases is to restrict the analysis to women who are observed locally at age 50, or older. By eliminating those who migrated or died during the marriage years, we eliminate the possibility of censoring. In the absence of information on the number of never-married eventual migrants who remain in the population at each age, the evidence on the marriages of those who leave the parish before age 50 cannot be evaluated, because we do not know the population at risk of

[10] The French studies are less restrictive; see note 3. The English reconstitutions generally control for a third sort of censoring of marriage age, beyond mortality and migration censoring. If marriages are classified according to marriage cohort, marriages that occur near the starting date of a family reconstitution study will be biased downwards: the baptisms of those who marry late are likely to have occurred before the starting date, and thus will be excluded. To correct for this, the first 50 years of each reconstitution is excluded from the analysis of age at marriage; Wrigley and Schofield, *op. cit.* in fn. 6, p. 255. An analogous problem at the end of the reconstitution will occur if the marriages are classified by birth cohort.

Table 4. *Mean ages at first marriage in simulated populations*

	Females	*N*	Males	*N*
All first marriages	26.0	48,700	27.1	52,785
First marriages of persons surviving beyond age 50	26.7	33,106	27.6	36,268
First marriages occurring in parish of baptism				
High-migration model	22.2	15,825	23.9	14,332
Medium-migration model	23.1	20,370	24.8	20,100
Low-migration model	24.1	27,506	25.6	28,809
First marriages of persons with an event in parish of birth beyond age 50				
High-migration model	26.7	1,435	27.5	2,212
Medium-migration model	26.6	3,491	27.6	4,873
Low-migration model	26.6	7,981	27.6	10,134

marriage at each age.[11] Therefore, such cases must be excluded from the analysis. This approach was adopted for the last three rows on Table 4, and it works quite well: regardless of migration level, the mean age at marriage falls within one-tenth of a year of the mortality-free estimates of age at first marriage given in the second row of the table. The small differences that remain are caused by purely stochastic effects.

But there is a cost to using this method. In high-migration, high-mortality populations, few people remain alive in the parish of their birth until age 50. In the high-migration model shown here, for example, about 90 per cent of the cases that would ordinarily be included in a family reconstitution study had to be eliminated to obtain the unbiased estimates, and even in the low-migration model more than two-thirds of the cases were thrown out. Since conventional English reconstitutions are typically based on the 20–30 per cent of marriages that are linked to baptism records, the additional restrictions proposed here mean that in most cases the analysis must be based on only about five to ten per cent of all marriages. For small reconstitutions, therefore, the approach may not be feasible.[12]

When this restriction to correct migration-censoring is applied to real reconstitution data, the apparent extent of bias appears to be less than is predicted by the micro-simulation runs. In Table 5, tabulated for me by Jim Oeppen and Rosalind Davies of the Cambridge Group for the History of Population and Social Structure, measures of age

[11] If we knew the number of never-married eventual migrants who remained in the population at each age, we could construct a nuptiality table and make use of the available information on the marriages of that group (see note 12). Later in this paper, I describe a method for estimating the frequency of remaining eventual migrants at each age, but the technique is based entirely on information about the ever-married population. Given the available data, it is impossible to construct age-specific migration estimates for the never-married; since there was probably an interaction between marriage and migration (discussed below), the estimates for the ever-married population cannot be assumed to apply to the never-married.

[12] It is possible to get a few more cases by constructing a nuptiality table. A nuptiality table begins with a set of age-specific marriage rates. The rates should be based on the population born in the parish who either are known to die in the parish, or are known to be present in the parish at age 50 or beyond. This approach adds significantly to the available cases, since it includes information from persons who died locally before reaching age 50. The rates are constructed by calculating the proportion of living never-married persons who marry at each age between 15 and 50. The nuptiality table can then be constructed by standard procedures. The appropriate method is described in H. S. Shryock and J. S. Siegel, *The Methods and Materials of Demography*, condensed edition (New York, 1976), pp. 338–340. Note that if the estimates of marriage age are to be free of mortality effects, the denominators of the rates should include those persons who never marry but who die locally between the ages of 15 and 50. If the evidence on the deaths of never-married persons who die locally is considered unreliable, then the rates can be adjusted for mortality by means of a model life table. The life table used for this purpose need not be precise, since the effects of mortality—censoring are modest.

Table 5. *Mean ages at first marriage of women in family reconstitutions*

Parish	Mean age at first marriage: Measure A (all marriages with baptism)	Mean age at first marriage: Measure B (women with event at 50+)	dif.	Number of women	Percentage of women included: Measure A	Percentage of women included: Measure B
Alcester	27.07	29.90	2.83	2138	16.18	2.95
Aldenham	25.23	26.48	1.25	2428	16.27	3.25
Ash	24.84	25.09	0.25	2709	20.63	4.17
Austry	27.59	28.50	0.91	575	17.57	2.26
Banbury	25.73	27.51	1.78	7367	18.05	2.93
Birstall	24.49	25.33	0.84	8429	27.36	6.44
Bottesford	26.29	27.06	0.77	2749	18.48	4.18
Bridford	26.37	28.84	2.47	615	18.21	3.74
Colyton	27.73	30.69	2.96	3321	13.25	2.71
Dawlish	26.01	27.12	1.11	1265	28.93	5.77
Earsdon	25.79	26.66	0.87	589	24.11	5.94
Gainsborough	25.51	26.90	1.39	8148	20.27	3.41
Gedling	26.03	27.14	1.11	2305	24.73	6.64
Gt. Oakley	23.74	26.52	2.78	440	12.05	1.14
Hartland	28.25	29.67	1.42	1646	31.83	8.93
Ipplepen	26.16	26.57	0.41	350	23.14	5.43
Lowestoft	25.13	26.27	1.14	2280	15.26	3.55
March	25.45	31.60	6.24	2662	10.26	0.30
Methley	25.98	26.14	0.16	2085	19.47	2.49
Mor. Bishop	26.05	26.18	0.13	2607	30.26	10.74
Odiham	25.10	25.72	0.62	4423	17.16	4.75
Reigate	25.17	27.19	2.02	1932	18.53	1.66
Shepshed	25.67	26.14	0.47	2585	25.76	5.76
Southill	24.84	25.25	0.41	3103	12.73	1.77
Terling	24.29	24.07	−0.22	1583	13.27	2.15
Willingham	24.71	27.45	2.74	889	10.12	1.69
Combined parishes	25.59	26.93	1.34	69223	19.90	4.16

at marriage and migration are shown for 26 English parishes. The first column (Measure A) shows the conventional measure of age at first marriage for women: it is simply the average age at first marriage for all first marriages linked to a baptism. The second column (Measure B) gives the age at marriage for those women with an event recorded beyond age 50.[13] The latter measure is effectively unaffected by either migration or mortality censoring. The third column gives the difference between the conventional measure and the unbiased measure. In practice, the degree of error ranges from −0.2 years to 6.2 years, and for the combined parishes the error is 1.3 years. This is a substantial bias, but it is considerably smaller than the error amounting to 3.5 years suggested by the micro-simulation for mortality- and migration-censoring under medium migration.

There are several possible explanations for this discrepancy. Part of the difference may result from differences in the age pattern of migration or marriage between the micro-simulation and the reconstitution parishes.[14] Alternatively, it is possible that age-specific

[13] The reconstitutions were restricted to the period of their greatest reliability, according to criteria developed by Wrigley. In addition, to enhance temporal comparability between Measures A and B, the final 50 years of each reconstitution were excluded from the analysis.

[14] In practice, censoring of marriage age proves to be fairly insensitive to the age pattern of migration, as long as the absolute level of migration before marriage is held constant; in fact, even a flat distribution of age at migration – in which the probability of migrating in a given year is the same for every individual in the population – reduces the censoring effect on age at marriage by only 22 per cent.

marriage rates for persons who marry before migrating are actually lower than those for non-migrants; in the real world, migrants in general might marry later than non-migrants. But the most plausible explanation, it seems to me, is an interaction between age at migration and age at marriage. If age at first migration were entirely determined by age at first marriage and the probability of migration at marriage was, therefore, unrelated to age at marriage, there would be no censoring effect. Experimentation with the model shows that if about half of first migrations are made to depend on first marriages, then the degree of censoring predicted by the model conforms closely to that observed in the family reconstitutions.[15]

The fourth column of Table 5 shows the number of women born in each parish, and the last two columns give the percentages of these births used for each measure. Overall, the conventional measure allows use of 19.9 per cent of female births, and if we restrict the analysis to women with an event at age 50 or over, we can only use 4.2 per cent of the births. These last two columns are also indicators of the extent of mortality and migration in each parish. Together, these two variables can explain 49 per cent of the variation across the 26 parishes in the difference between Measures A and B, and this relationship is significant at the 0.005 level. This reinforces the interpretation that the conventional measure of age at marriage is biased downwards because of migration and mortality-censoring.

MIGRATION AND MORTALITY: BLUM'S METHOD

The problem of migration-censoring of age at marriage is easy to address, because we can use events following marriage to establish who is at risk of marrying locally. In the case of mortality, however, there are no subsequent events, so correcting the problem is more difficult.

Several historical demographers have proposed strategies to correct for the effects of migration on estimates of adult mortality. The most subtle of these techniques was suggested by Alain Blum.[16] In this section I evaluate Blum's approach, and in the following I develop some alternative measures.

Because of data limitations, reconstitution analysis of adult mortality is focused on the ever-married population. In the case of the English reconstitutions, where age at death is not usually recorded on burial records, the analyses are further restricted to the locally-born population. Thus, adult mortality can be assessed only for the tiny sub-set of the population that was baptized, married, and buried in a single parish.

[15] Peter Laslett and John Harrison have reported that in Clayworth and Cogenhoe 7.1 and 3.0 per cent, respectively, of in-migration resulted from a marriage. However, these figures refer to all migrations, not just first migrations. Moreover, the proportions would doubtless be higher for the never-married. Thus it is possible that 50 per cent of first migrations of previously never-married persons may have been precipitated by a marriage. The probability of this occurring, however, is probably associated with age at marriage, which may diminish its effect on migration censoring. P. Laslett, *Family Life and Illicit Love in Earlier Generations* (Cambridge, 1977), pp. 70, 98.

[16] Alain Blum, 'An estimate of local adult mortality based on family cards', *Population*, **44** (1989), English Selection no. 1, pp. 23–38. Other approaches to the problem include L. Henry, *Techniques*; Jacques Dupâquier, 'Refléxion sur la mortalité du passé: mesure de la mortalité des adultes d'après les fiches de famille', *Annales de Démographie Historique* (Paris, 1978), pp. 31–48; E. A. Wrigley, 'Mortality in pre-industrial England: the example of Colyton, Devon, over three centuries', *Daedalus*, **97** (1968), pp. 546–580. In addition to Blum's approach, I implemented Wrigley's 'High Mortality' and 'Low Mortality' estimates. Unlike Blum's approach, Wrigley's method is quite sensitive to the general level of mortality in the population. The simulation suggests that under the high-mortality conditions of the early eighteenth century, the high and low estimates will bracket true mortality at all migration levels, but actual mortality is much closer to the low-mortality estimate than to the high one. Wrigley's mid-point of the estimates, therefore, understates life expectancy at marriage by between two and five years. Under the lower mortality prevailing by the late nineteenth century, Wrigley's high and low mortality estimates both produce life expectancies that are too low at all three levels of migration.

If no correction is made, age-specific mortality rates for this sub-population will be biased upwards, since the denominators will exclude all eventual migrants who remain in the parish at a given age. The problem is illustrated in Table 6, in which life expectancy

Table 6. *Life expectancies for simulated populations, by level of migration*

	Females				Males			
		Married non-migrant				Married non-migrant		
Age	All Females	High	Medium	Low	All Males	High	Medium	Low
20	36.6	21.4	26.1	30.5	37.3	26.0	29.4	32.5
30	30.5	24.1	25.7	27.4	30.6	25.4	26.7	28.4
40	24.4	21.2	21.7	22.7	24.1	21.1	21.7	22.7
50	18.2	16.2	16.5	17.2	18.2	16.6	17.0	17.5
60	12.5	11.4	11.6	11.8	12.6	12.0	12.2	12.5
70	7.8	7.3	7.4	7.5	8.2	7.7	7.8	7.9

by age for the entire simulated population is compared with that of the ever-married non-migrant population. The life expectancies at age 20 for ever-married non-migrants are between five and 15 years lower than those for the population as a whole.

What is needed to correct the bias is a method for adding in the exposure of migrants between their age at marriage and their age at migration. Eventual migrants should be considered part of the non-migrant population at risk, because if they had died before they migrated they would have been non-migrants.

Blum has proposed two methods for estimating the age of migrants at their departure, which provide the basis for 'minimum' and 'maximum' estimates of mortality. Under the maximum mortality estimate, migrants are counted as resident in the parish from their marriage until the date of their last recorded demographic event, whether that event is the birth of a child, widowhood, or remarriage. This strategy is equivalent to assuming that migration occurred at the moment of the last recorded event. Thus, the exposure to the risk of death is understated, and the technique provides an estimate of maximum mortality (that is, minimum life expectancy). As is shown in Table 7, the method works well. For the high-migration model, the maximum-mortality life expectancies are consistently about ten per cent below those for the overall population: assuming low migration, the figures are understated by about five per cent.

Blum's minimum mortality estimate is a little more complex. Instead of assuming that people migrate immediately after their last observed event in the community, he assumes that they migrate immediately before the first event which occurs after they have departed. Of course, the first event after migration cannot be directly observed, so it is imputed. The imputation of events is carried out by what the U.S. Census Bureau calls a 'hot deck' procedure. For eventual migrants who have their last recorded child at a given age, a donor is located among the non-migrant population who has a child of the same birth order at the closest possible age. The donor's interval to the birth of their next child is then assigned to the migrant. If the donor has no subsequent children, their interval to death is assigned instead. If the eventual migrant has no children recorded, an interval from marriage to the birth of the first child is allocated. The eventual migrants are then counted as part of the population at risk of dying from the date of their marriage until the date of their imputed event following migration.

The results of this procedure are given in Table 8. In general, the life expectancies generated by Blum's minimum-mortality procedure come out quite close to the life

Table 7. *Life expectancies according to Blum's maximum mortality procedure simulated populations, by level of migration*

	Females				Males			
		Married non-migrants (Blum's maximum)				Married non-migrants (Blum's maximum)		
Age	All females	High	Medium	Low	All males	High	Medium	Low
20	36.6	33.0	33.7	34.9	37.3	34.0	34.8	35.7
30	30.5	27.6	28.2	29.0	30.6	28.1	28.7	29.5
40	24.4	22.4	22.6	23.3	24.1	22.2	22.5	23.3
50	18.2	16.7	16.9	17.4	18.2	17.1	17.3	17.7
60	12.5	11.6	11.7	12.0	12.6	12.2	12.3	12.6
70	7.8	7.3	7.5	7.6	8.2	7.8	7.9	8.0

Table 8. *Life expectancies according to Blum's minimum mortality procedure simulated populations, by level of migration*

	Females				Males			
		Married non-migrants (Blum's minimum)				Married non-migrants (Blum's minimum)		
Age	All females	High	Medium	Low	All males	High	Medium	Low
20	36.6	37.9	36.9	36.6	37.3	37.6	37.0	37.1
30	30.5	31.9	31.0	30.4	30.6	31.2	30.6	30.7
40	24.4	26.0	25.0	24.5	24.1	24.9	24.2	24.2
50	18.2	19.9	19.0	18.5	18.2	19.4	18.7	18.5
60	12.5	14.5	13.6	12.9	12.6	14.1	13.5	13.2
70	7.8	9.7	9.0	8.3	8.2	9.3	8.8	8.5

Table 9. *Mean intervals to next event and to migration simulated populations, by migration level*

	Females			Males		
	High	Medium	Low	High	Medium	Low
Mean interval between last event before migration and first event after migration	11.0	12.6	14.3	12.7	14.0	15.0
Blum's imputed mean interval	4.0	4.7	5.1	4.5	4.8	5.2
Mean interval to migration	3.4	4.1	5.1	4.0	4.9	5.6

expectancies for the population as a whole. For medium and low-migration males, life expectancies at age 20 are actually higher in the general population than under the minimum-mortality rules. If Blum's method was working as expected, this should never occur: the minimum-mortality estimates should always produce a higher life expectancy than in the general population.

The explanation for the false minima can be seen in Table 9. The top row of the table shows the actual mean number of years to the next event, which for this purpose is limited to the birth of a subsequent child, or to death. The second row shows the mean number of years imputed by Blum's method. In all cases, the true interval to next event is much greater than the interval imputed by the hot-deck procedure.

There are two reasons why Blum's imputation procedure underestimates intervals to the next event. Intervals between events can vary from less than a year to more than 70 years. In a real population, the odds of migrating in a given interval are proportional to the length of the interval; people are far less likely to migrate between two events separated by a year, than between two events separated by 70 years. In Blum's method, all intervals are equal, so the odds of migrating in a given interval are unrelated to the length of the interval. The imputation procedure, therefore, assigns too many short intervals, and not enough long ones. The problem is compounded when the next event is a death; Blum's method assigns a death donated from the non-migrant population, and as we have seen, the ages at death for non-migrants are much lower than those for migrants.

The result is that Blum's minimum understates intervals to the next event following migration by almost two-thirds. Under high-migration conditions, however, people generally migrate soon after their last observed event, as is shown in the bottom row of Table 9. As long as Blum's imputed interval to next event exceeds the actual interval to migration, the method works. But under lower-migration conditions – such as those in the French reconstitutions – the method can produce false estimates of minimum mortality. In the simulation runs, this occurs for medium- and low-migration males.

The magnitude of this problem is small, and if we were only interested in overall life expectancy at a particular age, Blum's minimum would be an acceptable technique. But the estimated life expectancies mask more severe biases in age-specific death rates: overestimated death rates at younger ages are compensated by substantial underestimates at older ages. After the childbearing years, all imputed intervals are intervals to death, and these intervals tend to be much longer than intervals to actual migration. Thus, Blum's minimum overstates the age-specific death rates between ages 20 and 40 by as much as 20 per cent, and understates the death rates at older ages by a similar amount. This distortion of the age pattern of mortality is a substantial liability, and justifies a search for alternate approaches.

MIGRATION AND MORTALITY: NEW METHODS

If we are willing to assume that demographic behaviour between marriage and migration is similar for eventual migrants and non-migrants there is a very simple and effective solution to the problem of migration censoring. It is debatable, however, whether this basic assumption is tenable. Therefore, in addition to describing the simple solution, I will also present a slightly more complex and more conservative method for estimating minimum mortality.

The simple method works by directly estimating the number of ever-married eventual migrants present in the population at each age. Assuming that demographic behaviour of non-migrants and eventual migrants is similar, the ratio of demographic events to population size can be considered the same for both groups. Thus, we can estimate the size of the eventual migrant population present at a given age from the number of demographic events they experience. For this purpose, demographic events are considered to be childbirth, widowhood, and remarriage. The number of ever-married eventual migrants remaining at a given age is estimated as

$$n_a = \frac{P_a}{E_a} e_a, \qquad (1)$$

where P_a is the living ever-married non-migrant population aged a (that is, persons with a baptism record, marriage record before age a, and a burial record after age a); E_a is the number of events that occur at age a to those persons (events are childbirth, widowhood, and remarriage); e_a is the number of events to ever-married eventual migrants at age a (that is, to persons with baptism and marriage records before age a, but no burial record).

The denominator for the calculation of age-specific death rates is then $n_a + P_a$, the population of ever-married non-migrants plus the estimated population of ever-married eventual migrants still present.

The calculation of n_a for a simulation of a moderate-sized family reconstitution is given in Table 10. The final column shown in the table, p_a, is the eventual number of

Table 10. *Calculation of eventual married migrants present at each age*

a	P_a	E_a	P_a/E_a	e_a	n_a	(p_a)
20	199	114	1.746	311	543	547
21	261	157	1.662	379	630	594
22	309	164	1.881	345	649	644
23	370	192	1.927	377	727	736
24	452	196	2.304	353	813	763
25	502	231	2.171	330	716	760
26	563	266	2.115	329	696	746
27	617	266	2.318	300	695	745
28	680	265	2.566	295	757	720
29	719	268	2.683	247	663	692
30	770	266	2.895	232	672	665
31	833	280	2.973	223	663	633
32	873	267	3.270	188	615	615
33	869	245	3.547	156	553	574
34	876	216	4.053	143	580	540
35	888	210	4.229	112	474	498
36	897	211	4.249	115	489	449
37	895	174	5.144	98	504	422
38	885	166	5.328	71	378	394
39	880	187	4.706	58	273	370
40	875	153	5.719	48	275	340
41	863	146	5.908	44	260	314
42	848	133	6.376	52	332	296
43	830	119	6.971	31	216	283
44	812	106	7.660	27	207	268
45	795	120	6.621	27	179	250
46	777	97	8.005	19	152	225
47	760	80	9.494	16	152	208
48	742	80	9.269	11	102	192
49	722	62	11.645	11	128	174
50	700	79	8.861	11	97	163
51	685	50	13.690	4	55	141
52	666	42	15.857	8	127	134
53	644	38	16.947	9	153	119
54	622	27	23.037	12	276	110
55	598	29	20.621	4	82	106
56	577	32	18.031	12	216	95
57	558	35	15.943	3	48	89
58	538	34	15.809	1	16	80
59	521	31	16.790	3	50	70
60	503	28	17.964	6	108	63
61	483	13	37.115	3	111	59
62	471	21	22.429	1	22	50
63	451	25	18.020	2	36	46
64	427	24	17.792	2	36	42
65	406	19	21.368	3	64	41

migrants who actually remain. These figures, of course, would not be available in a real family reconstitution. They differ from the estimated eventual number of migrants remaining (n_a) because of random variation.

The smaller the population, of course, the greater the random error. Since all demographic events are used, however, the method works well for small populations. Accuracy is generally improved if five-year moving averages are used for the number of events when the population of married non-migrants is less than 500. If this step is taken, the technique yields life expectancies that are usually correct within half a year, even where the total population of married non-migrants fall below 300. The results of the procedure for the standard high-, medium- and low-migration populations are given in Table 11. These estimates are unbiased, and generally fall within 0.2 years of the truth.

This method is not very sensitive to violations of its basic assumption, but significant differences between the demographic behaviour of ever-married eventual migrants and non-migrants will bias the results. Suppose, for example, that births, widowhoods, and remarriages among eventual migrants were consistently 20 per cent lower than among non-migrants. Under the medium-migration model for males, this would lead to life expectancy being understated by just less than one year at age 20, and by about one-third of a year at age 40. If ever-married eventual migrants experienced on average half the demographic events of non-migrants, estimated mortality could exceed Blum's maximum mortality figures.

Although such large differences between the demographic behaviour of migrants and non-migrants seem unlikely, it would be useful to have an upper bound estimate of life expectancy that can be considered reliable. I therefore propose a new minimum estimate of mortality that avoids the major difficulties of Blum's minimum.

As we have seen, under the high-migration conditions characteristic of pre-industrial England, people would be expected to migrate quite soon after their last recorded event. In general, the rate of out-migration for persons for whom no subsequent events are recorded in their home parish would be considerably higher than that for the entire group of eventual migrants. The proposed minimum estimate of mortality assumes that people depart after their last recorded event at the same rate as all eventual migrants. By applying a rate of out-migration known to be too low, we exaggerate the exposure of eventual migrants, and thus obtain a minimum estimate of mortality.

The first task, then, is to estimate the timing of out-migration for all eventual migrants. The proportion of eventual migrants who remain in the parish at age a is estimated as

$$m_a = \frac{e_a/p_t}{E_a/P_a}, \tag{2}$$

where e_a is the total number of events at age a among all eventual migrants (that is, events occurring to persons with a linked baptism but no linked burial; valid events are marriage, childbirth, widowhood, and remarriage); p_t is the total number of baptisms of persons without a linked burial record; P_a is the number of living non-migrants aged a (persons with a linked baptism and a burial record after age a); E_a is the total number of events experienced by those persons.

E_a/P_a is the expected mean number of events experienced by persons aged a, and e_a/p_t is the actual mean number of events to migrants. The ratio of the two yields the estimated proportion of migrants ever born who are still present.

Note that unlike the previous method, this estimate is based on the total population, not just the married population. This is necessary, because we want to estimate the timing of departure, so we have to get rid of additions to the population through

Table 11. *Life expectancies by estimation of eventual migrant population simulated populations, by level of migration*

	Females				Males			
		Married non-migrants (simple method)				Married non-migrants (simple method)		
Age	All females	High	Medium	Low	All males	High	Medium	Low
20	36.6	36.8	36.6	36.7	37.3	37.0	37.0	37.2
30	30.5	30.4	30.4	30.4	30.6	30.4	30.4	30.7
40	24.4	24.6	24.3	24.3	24.1	23.9	23.9	24.1
50	18.2	18.2	18.0	18.2	18.2	18.2	18.2	18.3
60	12.5	12.3	12.3	12.4	12.6	13.0	12.9	13.0
70	7.8	7.6	7.8	7.9	8.2	8.3	8.3	8.2

marriage. Because we are dealing with the entire population, we can count first marriage as an event at well as child-birth, remarriage, and widowhood.

The next step is to tabulate the number of eventual migrants who have their last observed event at age a. Finally, for each age n from $a+1$ to the maximum age in the population, calculate

$$n'_a = \frac{m_n}{m_a} l_a, \tag{3}$$

where m_n is the proportion of migrants remaining at age n, m_a is the proportion remaining at age a, and l_a is the number of eventual migrants with a last event at age a. In effect, we are estimating the number of migrants with their last event at age a who would still remain in the population at each age following age a, assuming the rate of out-migration for persons who have already experienced their last event is the same as the rate estimated for all eventual migrants. These numbers must then be added to the denominators of Blum's maximum estimate. The resulting life expectancies are given in Table 12.

Table 12. *Life expectancies by alternate minimum mortality procedure simulated populations, by level of migration*

	Females				Males			
		Married non-migrants (alternate minimum)				Married non-migrants (alternate minimum)		
Age	All females	High	Medium	Low	All males	High	Medium	Low
20	36.6	39.8	38.8	38.2	37.3	39.2	38.6	38.2
30	30.5	33.6	32.8	32.1	30.6	32.8	32.2	31.8
40	24.4	27.0	26.2	25.7	24.1	26.1	25.4	25.1
50	18.2	19.9	19.4	19.2	18.2	19.8	19.4	19.1
60	12.5	13.3	13.2	13.2	12.6	14.1	13.8	13.6
70	7.8	8.1	8.3	8.5	8.2	9.0	8.9	8.6

One of the attractive features of this approach is that it is insensitive to differences between the demographic behaviour of non-migrants and eventual migrants, as long as such differences are reasonably consistent across age. Thus, for example, if eventual migrants experienced half the average number of events that non-migrants did at every age, the results would be unaffected. This is because only the change in the proportion of migrants remaining is used; the absolute proportion of migrants who remain is irrelevant. Moreover, experimentation with the micro-simulation revealed that even if dramatic differences in the age-patterns of demographic events between migrants and non-migrants are introduced – such as using the fertility rates of a late twentieth-century population for the migrants – it is almost impossible to get a false minimum. The sole exception is at the very oldest ages where the fit is quite close to begin with, and there are few non-migrants who remain alive; in these circumstances, a false minimum is possible simply because of random variation.

A side-benefit of this approach is that it produces as a by-product estimates of the proportion of the eventual-migrant population remaining at each age [Equation (2)]. These figures can easily be converted into age-specific rates of first migration, which in turn can provide the basis for a wide variety of migration measures. This should significantly increase the power of family reconstitution as a tool for the analysis of geographical mobility.

DISCUSSION

What are the implications of all this for family reconstitutions? The most significant consequence of correcting the censoring biases is probably the upward revision of age at marriage.[17] If the increased age at marriage were incorporated in Wrigley and Schofield's analysis of English population history, it might necessitate additional revisions in order to maintain internal consistency. Raising marriage age by 1.3 years might imply an implausibly low rate of celibacy in some periods, if all the other variables, such as marital fertility and migration, were held constant.[18] The effects of the refinement of adult mortality estimation are more difficult to pin down, but it is likely that the family reconstitutions will show moderately longer life expectancies if the methods I have proposed are adopted. In addition, if there were long-run changes in migration behaviour, we may be forced to modify our interpretation of trends in mortality and age at marriage.[19]

I have not touched on the issue of the representativeness of the non-migrant population. The demographic measures obtained from family reconstitution can be made unbiased, but they will still only apply to a tiny minority of the parish populations. For the analysis of uncensored ages at marriage, for example, the reconstitutions will have to exclude some 94 per cent of FRFs. We have no way of knowing whether the reconstitutable minority was typical of the general population.

[17] Statistics of marital fertility could also be affected by migration-censoring. Family reconstitution studies ordinarily restrict their fertility analyses to that sub-set of marriages for which the date of termination of marriage is known. As long as we are concerned strictly with age-specific rates, this restriction should take care of the main problems of censoring. However, since both the ages of women at the outset of marital unions and at the termination of those unions will be biased downwards in family reconstitutions, some measures of fertility may be distorted. The safest course would be to restrict analysis to marriages which remained intact and local until the wife reached age 45.

[18] See David Weir, 'Rather never than late: celibacy and age at marriage in English cohort fertility, 1541–1871', *Journal of Family History*, **9** (1984), pp. 340–354; Roger Schofield, 'English marriage patterns revisited', *Journal of Family History*, **10** (1985), pp. 2–20; Wrigley and Schofield, *op. cit.* in fn. 6. The increase in ages at marriage would also imply greater limitations of the potential frequency of multi-generational living arrangements; see Ruggles, *op cit.* in fn. 5, Appendix D.

[19] David Souden has argued that the extent of migration diminished during the eighteenth century, which would have led to a reduction of the effects of migration censoring. When the parishes used by Souden are combined, however, there appears to have been relatively little change; if anything, the evidence points to a slight increase in migration between the middle of the seventeenth and the middle of the eighteenth century, except for migration before FRF among women. The figures below are comparable to those in Table 3, but they have been broken into thirty-year cohorts. They are calculated from Souden, *op. cit.* in fn. 7, pp. 194–197.

This does not mean that the results of these studies are worthless. Far from it: family reconstitution can provide unbiased demographic estimates for a clearly defined sub-set of the population. In this respect, family reconstitution can be compared to the study of the demography of Yale graduates or Hollingsworth's studies of the demography of the British peerage.[20] Any unbiased data on demographic behaviour before the statistical era are valuable, no matter which sub-set of the population they refer to. And if reconstitutions are limited to the analysis of a small percentage of the population, that is still probably a larger section of society than any alternative source of individual-level data can provide.

What must be avoided is the implication that the non-migrants were representative of the population as a whole. Those who stayed behind were almost certainly of different economic status and birth order than those who moved.[21] They are missing a key aspect of demographic behaviour – migration – that occurred as frequently as marriage itself during this period. In the absence of firm evidence to the contrary, we should assume that they were untypical of the general population with respect to other aspects of demographic behaviour as well.

If we recognize that family reconstitution yields unbiased estimates only for a small sub-set of the population that is known to be atypical in some respects, then it makes sense to calculate all demographic measures for the sub-set of families in which there was no migration before age 50. This will ensure that although the population may not be representative, it will at least be clearly defined, and its various different demographic characteristics will be consistent with one another. For those measures which allow the use of a substantially greater section of the population – such as child-spacing and infant mortality – this core population can be compared with the larger population; this will give some indication of the representativeness of the completed families.

In sum, then, family reconstitution studies should pay close attention to the population at risk and attempt to correct problems of migration-censoring. Moreover, even when their demographic estimates are free from such biases, they should be seen for what they are: studies of a small sub-group of a population, which may or may not resemble the population as a whole.

Percentage migrating before and after FRF by cohort:
English Family Reconstitutions

	Females		Males	
Cohort	Before	After	Before	After
1601	78.5	57.4	72.3	53.0
1631	78.2	54.8	70.4	53.4
1661	79.7	52.7	70.7	48.9
1691	79.7	55.1	72.5	52.4
1721	79.1	56.1	75.4	58.9
1751	74.1	58.7	75.3	61.5

[20] T. H. Hollingsworth, 'The demography of the British peerage', Supplement to *Population Studies*, **18**, 2 (1964).

[21] For evidence on economic and occupational biases in family reconstitution populations, see Dupâquier, *loc. cit.* in fn. , p. 86; Knodel, *op. cit.* in fn. 1, pp. 469–475; Danyi, *loc. cit.* in fn. 1, pp. 76–77.