Research Article

An Illustration of the Problems Caused by Incomplete Education Histories in Fertility Analyses

Øystein Kravdal

This special collection is in honor of Jan M. Hoem on his 65th birthday.
The authors presented their papers at a working party at the Max Planck
Institute for Demographic Research in Rostock, Germany in April 2004.
The collection is edited by Gunnar Andersson and Gerda Neyer.

Table of Contents

Research Article

An Illustration of the Problems Caused by Incomplete Education Histories in Fertility Analyses

Øystein Kravdal [1]

Abstract

When assessing the importance of education for fertility, one should ideally use complete education histories. Unfortunately, such data are often not available. It is illustrated here, using register data for Norwegian women born in 1969, that inclusion of educational level at the latest age observed (28), rather than at the current age, can give substantially biased education effect estimates. It is also illustrated that imputation of education for earlier ages may lead to wrong conclusions. A simple imputation of educational level and enrolment based on the assumption that everyone passes through the educational system with the officially stipulated progress gives particularly misleading results. Somewhat better estimates are obtained when a slower progress more in accordance with reality is assumed, or when educational level and enrolment are imputed stochastically on the basis of distributions calculated from real data. Obviously, one should be very careful when faced with incomplete education histories, and try to make use of relevant information from other sources about the actual educational careers.

---

[1]   Department of Economics, University of Oslo. P.O. Box 1095 Blindern, Norway.
Telephone: (47) 22855158 Fax: (47) 22855035E-mail: okravdal@econ.uio.no

1. Introduction

Although a large number of studies have addressed the relationship between education and fertility, both in developing and developed countries, the research area is far from exhausted. For example, Hoem and Hoem (1989) noticed the lack of a negative education effect on second and third births in Sweden in the 1970s, and more recent analyses (e.g. Hoem 1996, Kravdal 2001) suggest that college education in Nordic countries now has a rather weak negative effect, if any, on fertility, except for the delayed entry into parenthood associated with school enrolment (Note 1). This result is theoretically challenging and politically interesting, and should stimulate further inquiries.

Unfortunately, many surveys and other kinds of individual data that may otherwise be well-suited for studies of fertility in developed countries only include the women's education at the time when the data were collected, which is partly a consequence of their previous childbearing. Put differently, it is an endogenous variable. It would be more relevant to consider how their birth rate, at any given age, is influenced by their educational level and enrolment status at that age, or some time earlier (and ideally their plans for further education, which are rarely available in any data) (Note 2). This would typically require education histories that cover all the years since the start of the reproductive period for the individuals under investigation. Even if the intention were merely to find out, for example, how education affects third births after 1990, educational data for earlier years would be welcome. This is because there is good reason to estimate a model for all parity transitions simultaneously, with a common unobserved factor, rather than analyzing only third births with a follow-up starting at the time of second birth or in 1990, whichever comes last.

Of course, there is also much that remains to be known about the importance of education in poorer countries. One particularly rich data source especially designed for fertility analyses, the DHS surveys, only includes information about the educational level at the time of interview, and this is also the case for other surveys. However, this limitation is less problematic in these societies, where education is often completed well in advance of the start of the reproductive period (Note 3).

In this paper, I illustrate empirically how one may be misled by considering education at the time of data collection, rather than current education. The illustration is based on register data with complete birth histories through 1997 for all Norwegian women born in 1969, and information on their educational level and enrolment for the years 1985-1997. This is the oldest cohort in the Norwegian register data system for which the education histories are complete. I compare estimates from birth rate models that include the educational level in 1997, which is the latest year of observation, and models that include the educational level for each year 1985-1997.

In addition, I show the limitations of some simple techniques one might try to use in an attempt to circumvent the problem arising from lack of data on education before the last year of observation. More precisely, I check how the estimated effects of current educational level and enrolment are changed when the data for 1985-1994 are ignored and the educational level and enrolment for these years are imputed instead, according to three different algorithms. In addition to the empirical explorations, I illustrate the endogeneity problem and the implications of one type of imputation, theoretically.

This is a purely methodological contribution. The estimates are of little substantive interest, because the cohort in focus had barely reached the median age at first birth by 1997, and only 9% had had their third child (as compared to 31% of women in the cohorts from the mid-1950s, when measured at age 40).

2. The education histories in the Norwegian register data

The register data on which this study is based include all women born 1936-81 who have received a Norwegian personal identification number, i.e. who have lived in Norway some time after 1960. Birth histories through 1997 are nearly complete for these women. Furthermore, education histories and date of death or last emigration, if any, are included.

The education histories are derived from the annual education statistics files produced by Statistics Norway. These histories provide information about the highest educational level attained as of the 1st of October for the years 1980-1982 and 1985-1997, and whether the individuals were enrolled in school at those dates. This means that, for women who were born in 1969 or later, and who were 16 or younger in 1985, we have largely complete education histories. (Norwegians born before 1991 started school in August the year they turned 7, unless there were special reasons for an earlier or later start. Thus, almost everyone had completed their compulsory lower secondary education, 9th grade, in June the year they were 16 (Note 4). There are missing values for one or more years for some of the women, primarily due to immigration. This will be further described below.

After compulsory education, there are two main educational career tracks. One possibility is to take a vocational education that lasts 1-3 years. After one year, the student is recorded as having '10 years of schooling' ('medium secondary education'), and after 2-3 years '11-12 years of schooling' ('upper secondary education') is recorded. The other possibility is to take a theoretical ('general') secondary education that qualifies the student for entry into college. When completed, it is also recorded as 11-12 years of education ('upper secondary education'), but no medium secondary

education is recorded for the intermediate years. Special arrangements are needed to take a theoretical upper secondary education in less than three years. It is not uncommon to take a theoretical upper secondary education after a vocational education.

After a theoretical upper secondary education, 3-4 years is the officially stipulated length of enrolment to receive a Bachelor's degree, or the equivalent. In principle, a clever or eager student might be able to pass the necessary examinations earlier; a slower progress is far more common, however. This level is recorded as '15-16 years of schooling'. One may also take a shorter post-secondary education ('low college education'), which would require 1-2 years of schooling at that level, or '13-14 years of schooling' in total. Some students are recorded as passing through this level on their way to a Bachelor's degree; others are not. This depends on how the undergraduate studies are organized.

Two years of graduate-level courses (possibly including thesis work) are meant to be sufficient to proceed from a Bachelor's to a Master's degree, or the equivalent. It is registered as '17-18 years of schooling'. However, it is, in principle, possible to accomplish this more quickly, and one may also proceed directly to a Master's degree without first taking a Bachelor's degree. An organized PhD program typically takes 3-4 years. Such a degree is normally obtained at higher ages than those that are considered in this study. Just how common the different educational careers are is briefly described below, and is based on data from the 1969 cohort.

In the Norwegian registration system, only these 'milestones' (10, 11-12, 13-14, 15-16, and 17-18 years of education, plus the PhD level) are recorded. People are registered with their highest level of education attained so far, according to these categories, until they satisfy all requirements at a higher level. For example, a student may have been enrolled in a Bachelor program for 4 years, and successfully completed a large number of undergraduate courses, without being registered as having more than theoretical upper secondary education, corresponding to 11-12 years of schooling. One may also continue taking undergraduate courses for several years after having earned a Bachelor's degree, without ever taking a higher degree (perhaps with the intention of earning a Bachelor's degree in another field, or for pure pleasure, without an eye to formal qualifications). This will show up as a long period of enrolment, but no change in the highest educational level attained. Such 'unproductive' enrolment is fairly common at all educational levels.

3. Brief description of educational careers among women born in 1969

The very few women (0.7%) in the 1969 cohort who were recorded with a higher level than a lower secondary education in October 1985 were dropped from the analysis. More importantly, the approximately 24% with unknown education for at least one of the years 1985-1997 were omitted. This left a sample of 29,740 women. (Half of the excluded women were foreign born. Among those who were born in Norway and had a registered educational level in 1985, only 5% had missing values a later year, largely because of emigration or temporary residence abroad.) (Note 5)

Among these 29,740 women, only 7% had not attained more than compulsory education by 1997; that is at age 28. The proportions with medium or upper secondary education at that age were 14% and 40%, respectively. About 70% of those who ever took a medium secondary education had reached this level within one year after compulsory school. Only 20% of them went on to a higher level.

About half of the women who ever took an upper secondary education were 19 years old at that time, and the large majority of these 19-year olds had not been recorded as passing through a medium secondary education. About 70% of all women with an upper secondary education eventually reached a higher educational level.

The majority of those who were recorded with a low college education attained this level after one year at college. The most common pattern among those who took even more post-secondary education was to spend another three years at college to take a Bachelor's degree, and perhaps three years beyond that to take a Master's degree (although two should be sufficient, in principle).

In total, 27% of the 1969 cohort had taken at least a Bachelor's degree by age 28. About half of them had skipped the low college education level, according to the registration. On average, women had been enrolled for 4 years at college to take their Bachelor's degree, but they were on average 5 years older at that time than when they took their upper secondary education. This illustrates that interruptions are common.

Many women had taken further education after having first reached the level they were recorded with at age 28. For example, about three-quarters of the women with an upper secondary or low college education had been enrolled for at least one year after having attained those levels.

4. An example of imputation for years prior to 1995 for the 1969 cohort

Presently, I will illustrate three different ways to impute educational level and enrolment. In practice, imputation is, of course, only relevant when data is incomplete, and the imputation will somehow be based on common knowledge or information from other data sources. However, because the current intention is to show the strength and weaknesses of the different techniques, I impute values for a sample for which complete education histories do exist. Starting with the observed educational level and enrolment in 1995, at age 26, I impute backwards for the years 1985-1994 for women in the 1969 cohort. (Imputation starts in 1995 rather than in 1997 since the procedure uses information about subsequent changes in educational level to impute enrolment after last previous exam.)

One type of imputation is deterministic, in the sense that all women with a given educational level in 1995 are assigned the same educational career prior to that. The other imputation is stochastic, because educational level and enrolment for all earlier years are drawn from given distributions.

4.1 Deterministic imputations

Quick progress

One simple way to impute educational level and activity is to let everyone pass through the educational system according to the officially stipulated progress, without any interruptions, and with the most common sequencing. In this study, I illustrate this approach by assuming that all women recorded with at least an upper secondary education at age 26 have taken their upper secondary education at age 19, without passing through the medium secondary level. Furthermore, I assume that they have taken a low college education (if any) at age 20, a Bachelor's degree (if any) at age 23, and a Master's degree (if any) at age 26 (i.e. a slower progress than stipulated from Bachelor to Master, but more in accordance with common practice). I also assume, for simplicity, that everyone with a higher college education has passed through a low college education, although that is far from the case, as explained above. Those who are recorded with a medium secondary education at age 26 are assumed to have taken that education at age 17.

The women are assumed not to be enrolled beyond the age when they reached the highest level they were recorded with at age 26. There is one exception: If a woman attained a higher educational level between age 26 and 28, but without a corresponding

number of years enrolled in this age interval, I add enrolment immediately after the year when the highest level as of age 26 was attained.

The results of this imputation are shown in Table 1. As expected, the imputed enrolment (proportion enrolled) at low ages is higher than the real. The average level of education is somewhat lower, however, because a substantial number of women actually pass through the medium secondary level. At ages 20-22, the imputed enrolment is instead lower than the real, and the figures are extremely low at ages 23-25.

Table 1:   *Real and imputed average educational level $(E_a)^1$ and enrolment $(W_a)^2$, by age, for the 1969 cohort.*

	Real data		Imputed data, deterministic, quick		Imputed data, deterministic, slow		Imputed data, stochastic	
	E_a	W_a	E_a	W_a	E_a	W_a	E_a	W_a
Age								
16	9.00	0.89	9.00	0.93	9.00	1.00	9.00	0.89
17	9.41	0.78	9.15	0.78	9.15	0.93	9.41	0.78
18	9.83	0.67	9.15	0.77	9.15	0.78	9.82	0.67
19	10.76	0.37	11.08	0.37	9.15	0.77	10.76	0.37
20	11.04	0.39	11.79	0.27	11.08	0.77	11.04	0.40
21	11.27	0.38	11.79	0.24	11.08	0.36	11.26	0.38
22	11.46	0.36	11.79	0.22	11.79	0.35	11.46	0.36
23	11.67	0.33	12.23	0.04	11.79	0.24	11.68	0.33
24	11.91	0.28	12.23	0.03	11.79	0.23	11.91	0.28
25	12.11	0.22	12.23	0.04	12.23	0.22	12.12	0.23
26[3]	12.29	0.18						
27[3]	12.42	0.15						
28[3]	12.53	0.12						

[1] Defined as average of the number of years in school stipulated as necessary to reach the levels in consideration. These number of years are 9, 10, 11.5, 13.5, 15.5 and 17.5. The last 4 values are mid-points of the intervals 11-12, 13-14, 15-16, and 17-18, respectively.

[2] Average over enrolment, which is yes or no (1 or 0).

[3] Not imputed at these ages.

Slow progress

The assumptions above do not fit well with reality. People spend somewhat longer time at school, on average, to attain a given educational level. Besides, there are interruptions, and many continue to be enrolled after having reached their highest

educational level (as judged by data at age 28). There is no easy way to incorporate interruptions. I instead assume a one-year longer enrolment up to upper secondary education, as well as to low college education. In addition, all women who did not reach a higher level after age 26 are assigned one additional year of enrolment after the year when the highest level, as of age 26, was attained.

The results are shown in Table 1. The imputed enrolment is generally higher than in the quick-progress imputation, and far too high at low ages. However, the imputed values of enrolment are quite close to the real ones at ages 22-25. On the whole, both average educational level and enrolment are closer to reality, according to the slow-progress approach, as compared to the quick-progress approach.

Using a try-and-error strategy, one might have fine-tuned the assumptions about progress from one level of education to the next, and eventually managed to obtain a rather good fit. Given the complexity described above, there are many alternatives that can be tried. Assumptions leading to a good fit could then be used in situations with a real lack of data, where imputation, indeed, would be necessary.

4.2 Stochastic imputation

Drawing from a distribution, rather than assigning the same behavior to everyone in a particular educational group, can make another kind of imputation. More precisely, one might use distributions over educational level and enrolment status, conditional upon age, educational level and enrolment status the following year, calculated from real data.

In this study, I calculate distributions for the 1969 cohort, and use these to impute backwards from 1995 (age 26) for the same cohort. The distributions over educational level and enrolment should then be the same for the real and imputed data, and this is confirmed in Table 1 (within a margin of 0.01). However, the values of enrolment and educational level are not assigned to the correct women. For example, at age 20, 30% of the women are assigned a wrong educational level, three-quarters of whom are assigned a level that is immediately below or above the correct one (not shown in tables). At that age, 38% are assigned a wrong enrolment status. Half of them were actually enrolled, but turned up as not enrolled in the imputation, and the situation was opposite for the other half. The proportions that are 'mis-classified' are lower at higher ages, when the educational activity is generally lower. At age 25, 11% are assigned a wrong educational level, and 21% a wrong enrolment status. These wrong assignments have, of course, consequences for the estimated effects of education on fertility.

5. Theoretical illustration of the endogeneity problem and the consequences of stochastic imputation

Before turning to the estimation of birth-rate models, I provide a theoretical illustration of the consequences of considering a future rather than current education, and of using a stochastic imputation. Let us consider one period in which women may bear children. Their educational level at the start or end of the period is either low (E) or high (E*). Those with low education at the start of the period may be enrolled in school. Thus, there are three possibilities at that time, denoted as E0, E1 and E*0 (where 1 signals enrolment, and 0 non enrolment). Moreover, let us assume (i) that there are N women in each of these three groups initially, (ii) that the probability of a birth is 0.5 for those who are not enrolled, regardless of educational level, (iii) that the probability of a birth is 0.2 for the enrolled, (iv) that a birth among the enrolled makes further educational advances impossible, and (v) that all others who are enrolled proceed to a higher level within the period under consideration. This can be set up in the following table: (Note 6)

Group	Number of women	Educational level time 1	Enrolment time 1	Number of births	Educational level time 2
1	N	E	0	N· 0.5	E
2	N· 0.8	E	1	0	E*
3	N· 0.2	E	1	N· 0.2	E
4	N	E*	0	N· 0.5	E*

Consequences of considering future rather than current education

Although it is assumed that educational level has no effect (the birth probability is 0.5 both in group 1 and 4), there is a difference in fertility between the women who have level E at time 2, and those who have level E*. (The birth probability is N· 0.7/N· 1.2 in the former group, while it is N· 0.5/N· 1.8 in the latter.) Put differently, if the women are grouped according to education at time 2, we see the lowest fertility among those with a high educational level (E*). This is a result of the impact that births have on the educational achievements, as well as the lower fertility among the initially enrolled. If the women in group 3 had also reached level E*, and if this group had been as large as group 2 (because of a general birth probability of 0.5), then the women ending at E* would have had the same fertility as those ending at E.

Consequences of imputation

With the set-up described above, a deterministic extrapolation would be rather meaningless. The most common status at time 1 for those ending at educational level E is E0, and the most common status for those ending at E* is E*0. Assigning these two values to all those ending at E and E*, respectively, would wipe out the possibility of being enrolled. It is more interesting to illustrate the implications of a stochastic imputation. The probabilities from which this would be drawn would be as follows: Among the women ending at E, 0.2/1.2 had E1 at time 1, and 1/1.2 had E0. Among those ending at E*, 0.8/1.8 had E1 at time 1, and 1/1.8 had E*0. If these proportions are used for each of the four groups, they are split up as follows from the table below: (For example, 1/1.2 of those in group 1, who had level E at time 2 and E0 at time 1, are assigned E0 at time 1 and called group 1a, whereas 0.2/1.2 are assigned E1 and called group 1b.)

Group	Number of women		Education/ enrolment time 1 *Real*	Education/ enrolment time1 *Imputed*	Number of births		Educational level time 2
1a	N·	1 /1.2	E 0	E 0	N· 0.5·	1/1.2	E
1b	N·	0.2/1.2	E 0	E 1	N· 0.5·	0.2/1.2	E
2a	N· 0.8·	0.8/1.8	E 1	E 1	0·	0.8/1.8	E*
2b	N· 0.8·	1/1.8	E 1	E*0	0·	1/1.8	E*
3a	N· 0.2·	1/1.2	E 1	E 0	N· 0.2·	1/1.2	E
3b	N· 0.2·	0.2/1.2	E 1	E 1	N· 0.2·	0.2/1.2	E
4a	N·	0.8/1.8	E*0	E 1	N· 0.5·	0.8/1.8	E*
4b	N·	1/1.8	E*0	E*0	N· 0.5·	1/1.8	E*

When the imputed values are used, the effect of enrolment corresponds to the difference in fertility between groups 1b, 2a, 3b, and 4a on one hand, and groups 1a and 3a on the other. This gives an effect of enrolment that is 0.34/0.58, rather than the 0.2/0.5 with real data. The ratio of these two estimates is 1.44. However, the bias in the enrolment effect must not always be in this direction. For example, if the real effect of enrolment had been set to 0.35/0.5 instead of 0.2/0.5, then the corresponding ratio would have been 0.94. In other words, the negative effect of enrolment would have been somewhat *sharper* than with the real data.

The effect of the imputed educational level is 0.28/0.58 (=0.48), rather than the real absence of effect (=1). Such a negative bias would be seen for any choice of effect

of real enrolment and educational level, regardless of whether the Ns are set to different values for the original three groups (Note 7). However, one cannot generalize much beyond this simple situation of three states at the start of the period, two at the end, and termination of enrolment in case of childbirth.

6. Comparison of education effects in birth rate models based on different kinds of educational data

6.1 The birth rate model

The first-, second- and third-birth rates are modeled jointly, with a common unobserved heterogeneity factor (Note 8). I ignore fourth and higher-order births, which are generally quite rare in Norway, and in particular for the 1969 cohort before age 28. The estimation is done in aML (Lillard and Panis 2000). The women are followed from January, the year they turn 17. They are censored at a first emigration, death, or the end of 1998, which is the last date covered by the data.

The first-birth rate is assumed to depend on current age (specified as a spline function) and educational level and enrolment in the preceding calendar year (to reflect that births are a result of behavior or decisions made some time earlier). Second- and third-birth rates are, in addition, assumed to depend on the current duration since the last previous birth (also specified as a spline function).

In mathematical terms, the specifications are as follows:

$$\log h^{(1)}(a,x,w) = \beta_0^{(1)} + \boldsymbol{\beta}_1^{(1)} \mathbf{A}(a,v_1,v_2) + \beta_2^{(1)} x_1 + \beta_3^{(1)} x_2 + \beta_4^{(1)} x_3 + \beta_5^{(1)} w + \delta$$

$$\log h^{(2)}(a,x,w,d) = \beta_0^{(2)} + \boldsymbol{\beta}_1^{(2)} \mathbf{A}(a, v_2) + \beta_2^{(2)} x_1 + \beta_3^{(2)} x_2 + \beta_4^{(2)} x_3 + \beta_5^{(2)} w \\ + \boldsymbol{\beta}_6^{(2)} \mathbf{D}(d,z_1,z_2,z_3,z_4) + \delta$$

$$\log h^{(3)}(a,x,w,d) = \beta_0^{(3)} + \boldsymbol{\beta}_1^{(3)} \mathbf{A}(a, v_2) + \beta_2^{(3)} x_1 + \beta_3^{(3)} x_2 + \beta_4^{(3)} x_3 + \beta_5^{(3)} w \\ + \boldsymbol{\beta}_6^{(3)} \mathbf{D}(d,z_1,z_2,z_3,z_4) + \delta$$

where h is a birth rate, and $^{(1)}$, $^{(2)}$ and $^{(3)}$ are symbols for first, second and third births, respectively.

In these equations, β_0 is a constant, and $\mathbf{A}(a,v_1,v_2)$ is a piecewise linear spline transformation of age, with nodes at $v_1=20$ years and $v_2=25$ years. It is defined as a column vector whose transpose is

$\mathbf{A}^t = (\min[a,v_1], \max[0,\min[a-v_1, v_2-v_1]], \max[0,a-v_2])$.

$\boldsymbol{\beta}_1$ is the corresponding row vector of effects. Defining compulsory education (9 years) as the reference category, x_1 is 1 if the woman has 10 years of schooling (otherwise 0), x_2 is 1 if she has 11-12 years of schooling, and x_3 is 1 if she has 13 or more years of schooling. The corresponding effects are $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$. The enrolment status is w (1 if enrolled, otherwise 0), and the corresponding effect is $\boldsymbol{\beta}_5$.

Also $\mathbf{A}$ (a, v_2), which is included for second and third births, is an age spline, with one node at $v_2 = 25$ years, whereas $\mathbf{D}$ (d, z_1,z_2,z_3,z_4) is a duration spline with four nodes at $z_1 = 2$ years, $z_2 = 4$ years, $z_3 = 6$ years, and $z_4 = 8$ years.

It is assumed that δ is independently drawn from a normal distribution with zero mean and a variance to be estimated. There are 8 support points that approximate this distribution (which appears to be sufficient, because 12 points gave the same results).

For comparison, some models are estimated separately for each parity transition. This corresponds to omitting δ.

6.2 Estimates

Actual educational level and enrolment

Effects of real educational level and enrolment are shown in Table 2 (Model 1). As always, enrolment tends to reduce birth rates. The effect becomes less sharp as the birth order increases. Upper secondary education and college education reduce first-birth rates and increase second-birth rates, whereas effects on third-birth rates are not significant.

If enrolment is excluded from the model, effects of educational level change, but not in a consistent manner (Model 2). For first births, the effects of upper secondary education and college education become more negative, reflecting that the women at these levels are over-represented among the enrolled. For third births, there is a change in the opposite direction (significant only for college education).

When only the educational level in 1997 is included, education effects are generally much more negative (Model 3), just as shown in the theoretical illustration above (Note 9). This substantial difference illustrates the seriousness of the endogeneity problem in this particular type of analysis.

In addition, as seen also in Kravdal (2001), effects of educational level are generally more positive or less negative in models estimated separately for each parity transition (Models 4 and 5).

Table 2: *Estimated effects of current enrolment and educational level in joint models, unless otherwise stated, for first, second and third births, for women in Norway born in 1969, according to real or imputed data.*

	Real data					Imputed data, deterministic quick	Imputed data, deterministic slow	Imputed data, stochastic
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
		Enrolment ignored		Separate for each transition	Separate for each transition			
			Based on education in 1997		Based on education in 1997			
First birth								
Compulsory[1]	0	0	0	0	0	0	0	0
Medium secondary	0.03	0.09	-1.06*	0.06	-0.19*	0.02	0.24*	-0.03
Upper secondary	-0.63*	-0.74*	-2.21*	-0.34*	-0.61*	-0.73*	-0.50*	-0.71*
College	-0.85*	-1.12*	-2.19*	-0.38*	-1.31*	-1.31*	-0.67*	-1.04*
Enrolled	-1.14*			-1.12*		-1.26*	-0.94*	-0.62*
Not enrolled[1]	0			0		0	0	0
Second birth								
Compulsory[1]	0	0	0	0	0	0	0	0
Medium secondary	0.11*	0.12*	-0.09	0.10*	0.04	0.06	0.14*	0.01
Upper secondary	0.19*	0.17*	-0.27*	0.32*	0.10*	-0.06	0.01	0.02
College	0.33*	0.26*	-0.60*	0.61*	0.11*	-0.13*	0.11*	0.11*
Enrolled	-1.02*			-0.94*		-0.86*	-0.68*	-0.50*
Not enrolled[1]	0			0		0	0	0
Third birth								
Compulsory[1]	0	0	0	0	0	0	0	0
Medium secondary	0.01	0.08	-0.13	0.02	-0.03	-0.03	0.03	-0.07
Upper secondary	-0.06	0.03	-0.37*	0.13*	-0.01	-0.26*	-0.26*	-0.19*
College	0.16	0.36*	-0.66*	0.57*	0.31*	-0.25*	-0.10	0.03
Enrolled	-0.71*			-0.61*		-0.50*	-0.50*	-0.35*
Not enrolled[1]	0			0		0	0	0

[1] Reference category
* significant at the 0.05 level

Deterministically imputed education, quick progress

The deterministic imputation of educational level and enrolment based on a rather quick career track gives very different results from those of the real data (compare Model 6 with Model 1). Above all, the effects of college education are (more) negative. When we only distinguish between significantly positive effects, significantly negative effects, and all others, 4 out of the $3 \cdot 3 = 9$ effects of educational level are correct. Effects of enrolment are not very different from those based on real data.

Deterministically imputed education, slow progress

When I assume, instead, a slower progress between educational levels and some additional enrolment at the end of the school career, effects become more similar to those from the model based on real data (compare Model 7 with Model 1). For example, the positive effect of college education on second-birth rates now appears, and there is no longer a negative effect for third births, just as with real data. However, the effects of secondary education are no more correct than those based on the other deterministic imputation. In total, 6 out of the 9 effects of educational level have the correct sign. Effects of enrolment are generally weaker when these imputed data are used, as compared to when the real data are used.

Stochastically imputed education

In the theoretical discussion above, it was shown that a stochastic imputation might give sharper as well as weaker effects of enrolment. With these data from the 1969 cohort, enrolment effects turn out to be generally less sharp with this kind of imputation than in the model based on real data (compare Model 8 with Model 1). They are also less sharp than in the models based on deterministically imputed education. The effects of college education are more negative or less positive than those appearing with real data (in accordance with the theoretical illustration), but the same conclusions about significance can be drawn. Just as with the deterministic imputation where a slow progress was assumed, the stochastic imputation produces 6 out of 9 education effects with a correct sign.

7. Summary and conclusion

Some surveys cover complete education histories that allow the current educational level and enrolment to be included in birth rate models (e.g. the Norwegian Family and Occupation Survey of 1988; see Statistics Norway 1991), but most data used in fertility research do not. In the absence of such information, one might consider using the educational level at the end of the observation period. However, that level may be different from those at earlier and more relevant ages, and the difference may even be a result of childbearing. As shown here, the estimated effect of such a variable may deviate very much from that of current education.

Another possibility would be to impute data on education histories from the information on the level at a higher age. After all, some educational careers are much more plausible than others. One might assume, for example, that everyone proceeds through the educational system in accordance with official recommendations, with little interruption. Unfortunately, this may give values of educational level and enrolment that are quite different from the real ones, and the estimated effects in birth rate models may also be very different.

In the 1969 cohort that I used as an example, assumptions of a slower educational progress than that given by official recommendations produced a better fit, both in terms of educational distributions and estimated effects. With that approach, one would draw correct conclusions about the significance of the parity-specific effects of college education. However, the effects of secondary education were not much closer to those from real data than those based on the other deterministic imputation.

Another alternative might be to calculate distributions of educational level and enrolment, conditional upon age and values of these variables for the next year, and use them for a stochastic backward imputation. These distributions must be taken from other data for which educational histories are more adequate. In an analysis of Norwegian register data, it would be a good idea to turn to the younger cohorts. If people's educational behavior has not changed over the relevant periods, the distributions from younger cohorts will be equal to the true ones, and the average imputed educational level and enrolment would fit with reality. However, the educational characteristics would be assigned to the wrong women, and estimated effects on fertility would be biased. For the 1969 cohort, at age 16-26, this method performed just as well as the deterministic imputation, based on the assumption of a slow progress. The signs of the effects of college education on birth rates were correct, but only half of those of medium or upper secondary education were as well. Enrolment effects were generally weaker than in the models based on real data. In these stochastic imputations, distributions calculated from real data for the same cohort were used, because of the special purpose of the study. It remains to be seen whether imputation,

based on distributions from another cohort, perhaps for a longer age interval, would produce much poorer estimates (Note 10).

Generally, the appropriateness of all these methods depends on whether childbearing exerts a strong effect on subsequent educational activity. It is possible that this effect is relatively weak in Norway, because of various support schemes for students. If so, imputation might be even more problematic for other countries.

The Norwegian registration system has a peculiarity: Some people are, for example, registered as passing through a 'low college education' on their way to the equivalent of a Bachelor's degree, while others are not. There is a similar ambiguity both at higher and lower educational levels. The deterministic imputation requires a choice to be made, and it has been assumed that everyone passes through the low college education. Thus, while the imputed value, say, two years after completed upper secondary education, is low college education, many of these young adults are actually registered with no more than upper secondary education as their highest level at that time. On the other hand, this 'wrong' imputed level may actually be a more relevant representation, because some of the factors that educational level is likely to operate through may depend more on the number of years at school than on whether specific degrees have been taken.

The bottom line is that imputation of education and enrolment is problematic. If imputations are made, one should first gather as much information as possible, from other cohorts in the data or from other sources, about how people's educational careers actually proceed. Moreover, one should check the implications of the different assumptions for the estimates. However, there will always be uncertainty about the quality of the results. The obvious lesson to be learnt is that it would indeed be advantageous to include data on the timing of educational activities in future surveys or other data used to explore the education-fertility relationship.

8. Acknowledgements

Notes

1. The less negative effect may partly reflect that the classic opportunity cost argument has lost much of its relevance in recent years in the Nordic countries. Parents purchase childcare, at a price relatively independent of incomes, rather than letting one adult, usually the mother, stay at home. In addition, one may speculate whether there originally was a gap between the educational groups in terms of childbearing preferences (given childbearing costs and income) and efficiency of contraceptive use, and that this gap has narrowed and perhaps eventually closed (see Kravdal 2001 for further discussion).

2. There are many reasons to expect both educational level and enrolment to have an influence. Because these two variables are also correlated, the estimated effect of one of them may be sensitive to the inclusion of the other. Blossfeld and Huinink (1991) found, in a study from Germany, that the effects of educational level were considerably weaker once enrolment was taken into account. The importance of including enrolment in the models, and not only educational level, is only very briefly addressed in this paper.

3. Effects of primary schooling would probably not be substantially biased, because this level of education is typically reached before age 15. Generally, education effects on second and higher-order births may also be largely correct, as few women would probably take (more) education after having become mothers, regardless of further childbearing. However, the effects of secondary education on first births may be severely biased, and especially in countries where pregnant young girls are expelled from school.

4. For those born in 1991 or later, compulsory school started at age 6, but lasts 10 years, so June the year they are 16 will still be the end of compulsory schooling.

5. It might well be that the foreign born and other excluded women would have displayed other education effects, for example, due to less compatibility between motherhood and employment. The impact of imputation might also be different, because of another pattern of school tracks, more severe educational consequences of childbearing, or for other reasons. However, that should not be much of a concern in this study, which is only an illustration of a methodological problem.

6. This table also serves as an illustration of the need to control for enrolment: If enrolment were ignored, we would get a positive effect of high education at time 1. This is because the groups with low education (E) include some who are enrolled, and who contribute to bringing the average fertility in those groups down. (The

contribution is large in this particular example, where it is assumed that an equal number are enrolled and not enrolled at this educational level).

7. This is because the education effect corresponds to the ratio between the fertility in 2b and 4b and that in 1a and 3a. This ratio would have been 1, as in the real data, had it not been for 2b and 3a. Fertility in 2b is 0 and contributes to push fertility associated with E* down, and fertility in 3a is 1 and contributes to push fertility associated with E up. Thus, the education effect is negatively biased.

8. The motive for the joint-model approach is explained in Kravdal (2001, 2002), where it was also shown that this approach did not give the positive effects of educational level (measured at the end of the reproductive period) that were found in separate models for each parity transition. Alternatively, the unobserved heterogeneity might be taken into account in a separate-model approach by including age at previous birth relative to the average at this parity transition for the woman's educational category (Hoem 1996, Hoem et al. 2001).

9. In contrast to this, Kravdal (2001) found that effects of college education at the end of the observation period were non-negative for second- and third-births in the cohorts from the 1950s. However, the women in that study were followed up to age 40, not 28.

10. The possibility that there are 'holes' in the education histories, rather than a series of missing values up to a certain age, has not been discussed in this paper. The Norwegian register data actually have this structure, as the education histories cover 1980-1982 and 1985 onwards. A stochastic imputation can, of course, be made for 1983 and 1984, based on distributions conditional on age and characteristics immediately before and after the two-year period (for example from younger cohorts). Generally, the errors introduced when 'filling in' such 'holes' will be less pronounced than when the imputation is anchored only in the years after the period with missing values.

References

Blossfeld, H.-P., and Huinink, J., 1991. "Human capital investments or norms of role transition? How women's schooling and career affect the process of family formation." *American Journal of Sociology* 97: 143-168.

Hoem, B., 1996. "The social meaning of education for third-birth fertility: A methodological note on the need to sometimes re-specify an intermediate variable." *Yearbook of Population Research in Finland* 33: 333-339.

Hoem, B., and Hoem, J.M., 1989. "The impact of female employment on second and third births in modern Sweden." *Population Studies* 43: 47-67.

Hoem, J.M., Prskawetz, A., and Neyer, G., 2001. "Autonomy or conservative adjustment? The effect of public policies and educational attainment on third births in Austria, 1975-96." *Population Studies* 55: 249-261.

Kravdal, Ø., 2001. "The high fertility of college educated women in Norway: An artefact of the separate modelling of each parity transition." *Demographic Research* 5: 185-216. Available http://www.demographic-research.org/Volumes/Vol5/6.

Kravdal, Ø., 2002. "Is the previously reported increase in second- and higher-order birth rates in Norway and Sweden from the mid-1970s real or a result of inadequate estimation methods?" *Demographic Research* 6: 239-262. Available http://www.demographic-research.org/Volumes/Vol6/9.

Lillard, L., and Panis, C.W.A., 2000. *aML Multilevel Multiprocess Statistical Software. Release 1.0.* EconWare, Los Angeles, California.

Statistics Norway, 1991. *Family and Occupation Survey 1988*. NOS B 959. Oslo-Kongsvinger.