Do Imputed Educational Histories Provide Satisfactory Results in Fertility Analysis in the West German Context?

Cordula Zabel (zabel@demogr.mpg.de)

Do Imputed Educational Histories Provide Satisfactory Results in Fertility Analysis in the West German Context?

Cordula Zabel

Abstract

This paper investigates how well imputed educational histories perform in the analysis of first birth rates in the West German context. The focus here is on the quality of estimates when only rudimentary information on the timing of education is available. In many surveys, information on respondents' educational histories is restricted to the highest level of educational attained by the time of interview and the date at which this highest degree was attained. Skeleton educational histories can be imputed simply from such rudimentary information. The German Life History Study has complete educational histories. We use these to compare estimates based on the complete histories with estimates based on corresponding imputed histories. We find that the imputed histories produce relatively reliable estimates of the effect on first-birth rates of having a university degree vs. having a vocational certificate. Estimating corresponding rates for women who have no such education proved to cause greater difficulties.

1. Introduction

Much of fertility research focuses on the influence of women's level of education. Unfortunately, complete educational histories are rarely available from survey data. When they are not, one option is to impute educational histories from more rudimentary information that is available, which might include the highest level of education attained by the time of the interview and the date when this level was reached. The aim of this paper is to determine whether such imputed educational histories can serve as satisfactory substitutes for complete educational histories when one analyzes the effect of educational attainment on entry into motherhood. In the present paper we compare results using imputed and complete educational histories for the case of western Germany, applying data from the German Life History Study.

Whether educational histories can be approximated sufficiently closely using only the date of attainment of the highest degree depends strongly on the country context. In some countries it is common to exit and re-enter the educational system and to acquire higher levels of education even after a considerable period of employment. This appears to be the case especially in the United States, where young labor market entrants often return to education on a full-time basis, as well as in Great Britain, where vocational credentials are frequently acquired on a part-time basis at the same time as one is employed (Kerckhoff 2001). The Swedish educational system, too, is very flexible, with a wide range of policies aimed at supporting students who reenter education at a later age. A very high proportion of students in Sweden resume education after already having gathered labor market experience, or take part-time courses next to employment (Hoem, Neyer, and Andersson 2006). We suspect that it will be difficult to approximate educational histories using only sparse information for countries such as Sweden, Great Britain, or the United States. Approximation of educational histories should promise to be more successful in a country like Germany where educational re-entry at later ages is more difficult. Schütze and Slowey (2002) compare institutional characteristics that facilitate or inhibit access to post-secondary education for non-traditional students, including older students, between a number of countries. They find that the institutional framework in Germany is much less conducive to educational participation by older students. Post-secondary education in Germany is less institutionally diversified and provides fewer programs that specifically meet the needs of older students. In addition, there are few alternative modes of ac-

cess for students who do not fulfill the regular entrance requirements. There is also no formal recognition of part-time student status. These findings lead to the expectation of low educational reentry rates. Kerckhoff (2001) however reports that a significant proportion of Germans who have completed their apprenticeships subsequently enroll in university education. If they make this transition right after the completion of their apprenticeship, this is still in line with the assumption we use for the imputed histories, namely that respondents are in education continuously up until they acquire their highest educational degree. Longer breaks between the apprenticeship and the start of university education could be more problematic.

Kravdal (2004) investigates the extent to which imputed educational histories bias estimates of the effect of educational level on fertility transitions in Norway. However, the research question in his study is somewhat different than in the present context. Kravdal (2004) addresses the implication of having no information on educational trajectories at all besides the highest level of education at interview. In the present context, by contrast, we assume that we also know the date when the respondent's highest level of education was attained. The data used by Kravdal (2004) is Norwegian register data for cohort 1969, the first cohort for which complete educational histories are available. The author compares results for the effect of level of education on fertility transitions using complete educational histories with results based on three different methods of imputation of educational histories. He finds that both imputed educational histories assuming a slower but more realistic progression across educational levels than officially recommended, as well as stochastically imputed educational histories produce better results than imputations assuming the officially recommended pace of progression to each educational level. Nonetheless, each of the three methods of imputation still lead to substantial deviations from the results produced when using complete educational histories. Therefore, the author recommends including items on the timing of education in future surveys in order to ensure the quality of estimates of the effect of educational level on fertility.

The problem that incomplete educational histories pose for the estimation of the effect of level of education on fertility transitions has been pointed out again by Hoem and Kreyenfeld (2006). The authors call attention to the distortions caused by anticipatory analysis, in particular when summary statistics of childlessness are presented by level of education at interview only. They devise an alternative summary statistic based on time-varying information on educational status and level. For this

purpose, complete educational histories would be ideal. However, as they point out, complete educational histories are seldom available, as is the case with the data set they use, namely the German Family and Fertility Survey. Like the first round of the Gender and Generations Surveys, the German Family and Fertility Survey provides the date of attaining the highest level of education at interview, but does not provide any information on educational trajectories before that date. Hoem and Kreyenfeld (2006) suggest a way to impute educational histories using only the highest level of education at interview, the date of attainment, the time the respondent completed school, and the starting date of the respondent's first job.

The present study uses their imputational method as a starting point. First, we will use the complete educational histories provided by the German Life History Study to model the effect of educational level as a time-varying variable on transition rates to first birth. Next, all information will be ignored other than the highest level of education attained by the time of interview, the date this level was gained, the date school education was completed, and the starting time of the respondent's first job. Using only this rudimentary information, educational histories are imputed based on the method proposed by Hoem and Kreyenfeld (2006). We then re-analyze first-birth rates using only these imputed educational histories and compare the results with those produced using complete educational histories. Our aim is to see to what extent the results are comparable.

The following (second) section describes the data set, our data preparation, and the procedure used for imputing educational histories. The third section then compares the results from the two procedures. We provide a summary in the fourth (and final) section.

2. Data and Method of Imputation of Educational Histories

The German Life History Study includes retrospective data on many different realms of the life course for a number of cohorts both from eastern and western Germany. For the present paper, we used data for the cohort born in 1964. This is the youngest western German cohort for which the data contain advanced first-birth histories by the time of the interview in 1998/99. We excluded respondents who were not born in western Germany, since our objective is to determine how well educational histories

can be imputed in the context of the West German educational system. We conducted the study for female respondents only. After excluding a small number of respondents with missing first-birth information, this left a sample size of 641 respondents, who gave birth to 471 first children during the observation period.

To compare results obtained using imputed and complete educational histories, first, complete educational histories were prepared using the detailed information provided in the survey. We used information from the vocational training record file as well as the school education record file. The former includes spells of apprenticeship and of university education as well as certificates and degrees obtained at the end of these spells. We grouped the various types of degrees and certificates into the categories 'university degree' or 'vocational degree'. We used the dates at which each degree was reached to construct a time-varying variable called "level of education" with the categories "no degree", "vocational degree", and "university degree". This variable was devised so as to provide the highest level attained at any given point in time. Thus, if a respondent attains a vocational degree after having already earned a university degree, the education variable for this respondent will continue to indicate that this respondent has a university degree.

Information on times spent enrolled in education was taken from the school education record combined with the vocational training record. Respondents were considered to be in education only if they participated full-time. The trajectories of times spent in and out of education were combined with the vocational-degree histories. In this manner, a time-varying variable combining enrollment status and vocational degree could be constructed. The variable has the categories "in education", "not in education/no degree", "not in education/vocational degree", and "not in education/university degree". The complete educational histories constructed in this manner reflect transitions in and out of education as well as changes in educational levels.

We prepared the imputed histories in part from the more complete histories described above. The highest degree level at interview and the date this degree level was attained for the first time were taken from the complete histories. For people who had either a vocational certificate or a university degree by the time of interview, this was all the information that was used for the imputed histories. For the imputed histories, respondents were assumed to be in education until the time they reached their highest level and outside of education afterwards. For people who had no degree by the time of interview, the time spent in education was imputed using the date they left

school and the starting date of the respondent's first job, as proposed by Hoem and Kreyenfeld (2006). The idea here is to take into account that the respondent may have started vocational training after leaving school, but then dropped out at some time before starting the first job. Thus, we imputed a variable giving a random point in time between leaving school and starting the first job. If the respondent was never employed, we used a random time between the time of finishing school and age 20. The respondent was then assumed to have been in education until that random time and outside education afterwards. A time-varying education variable with the same categories as the education variable generated from the complete histories ("in education", "not in education/no degree", "not in education/vocational degree", and "not in education/university degree") can thus be derived from the imputed histories as well. The difference is that the imputed histories do not take into account any transitions in and out of education except for the one transition out of education that is assumed to have taken place at the time the respondent first attained the level recorded as her highest at the time of the interview. Also, the imputed histories do not provide any information on lower degree levels gained prior to the highest educational level attained at interview.

Both the complete and the imputed education histories were merged with information from the record file on the respondents' children. That information was whether the respondent had any children at all, and if so the date of birth of the first biological child.

3. Results

The results presented in this section indicate that, in the case of western Germany, imputed educational histories are quite a good substitute for complete histories for the purpose of estimating the effect of educational level on first-birth rates. Table 1 compares estimates using complete and imputed educational histories. Complete histories take into account all transitions in and out of education as well as changes from lower to higher degree levels. Imputed educational histories on the other hand allow us to model the effect of education as a pseudo-time-varying variable even when only very rudimentary information on the respondents' educational trajectories is available. The drawback of using imputed histories, however, is that all respondents are assumed to

have made only one transition from being inside to being outside the educational system. Also, transitions from lower to higher educational levels are left out.

The two models shown in Table 1 give very similar estimates. The only sizable difference is that women with no degree are estimated to have only 11% higher risks of first birth as compared to women with a vocational certificate when we use complete histories but as much as 37% higher first birth risks when we use imputed histories. The problem is that many people do not transfer directly from school education to vocational training or to university education. Nonetheless, birth rates can be expected to be quite low during any period in between school and further education. Birth rates are also likely to be quite low during any short breaks between vocational training episodes, for example if a respondent drops out of one type of vocational training and starts a different type shortly after. In the complete histories, respondents are coded as not enrolled and having no degree during such breaks. This should substantially lower the rates of transition to first birth estimated for respondents coded as not enrolled and having no degree in the complete as compared to the imputed histories, where respondents are coded as being enrolled rather than not enrolled with no degree for these time periods.

Table1: Effect of level of education on the rate of transition to first pregnancy.

	Complete histories		**Imputed histories**	
age				
15-19	0.0035		0.0036	
20-24	0.0061		0.0060	
25-29	0.0125		0.0119	
30-34	0.0099		0.0095	
Education				
no degree	1.11		1.37	*
vocational	*1*		*1*	
university	0.67	**	0.66	**
in education	0.15	***	0.14	***

Table 2: Interaction between age and level of education

Complete histories

	enrolled	not enrolled, no degree	not enrolled, vocational degree	not enrolled, university degree
15-19	0.0005	0.0054	0.0028	—
20-24	0.0008	0.0065	0.0063	n.e.
25-29	0.0027	0.0121	0.0128	0.0070
30-34	0.0022	0.0072	0.0090	0.0096

Imputed histories

	enrolled	not enrolled, no degree	not enrolled, vocational degree	not enrolled, university degree
15-19	0.0004	0.0106	0.0025	—
20-24	0.0009	0.0079	0.0060	n.e.
25-29	0.0023	0.0114	0.0125	0.0067
30-34	0.0060	0.0070	0.0087	0.0091

The results when we interact age and level of education (Table 2) also support this explanation. The interactions show that for women with no degree, using imputed histories produces higher estimates only at young ages, which is when intervals between school education and further education are most likely to occur.

Table 3 shows what we get when we subdivide the category 'not enrolled/ no degree' into three subcategories. The first category includes respondents who have finished school education but have not yet commenced vocational training. The second category represents the period between two vocational training episodes. Finally, the third includes respondents who have completed their last vocational training episode but are still without a degree, as well as those who never enrolled in vocational training. The results show that rates of entry into motherhood are very low in the first two 'no degree' subcategories. These two subcategories are part of the 'no degree' category in the analysis using the complete education histories, but they are not in the analysis using the imputed educational histories; there they contribute to the 'in education' category. The estimate for the third subcategory (having no degree after the completion of vocational training) is practically the same as the estimate for having no degree in the imputed histories. This gives further support to the explanation for the estimate differences mentioned above. It seems that people who expect to continue their education shortly have quite low first-birth rates. Since they are included in the 'no degree' category in the complete educational histories, rates of transition to the first child are quite a bit lower for respondents with no degree when we use the complete than when we use the imputed educational histories.

Table 3: Differentiated 'no degree' categories

age		
15-19	0.0037	
20-24	0.0062	
25-29	0.0124	
30-34	0.0098	
no degree, between school and vocational training	0.51	
no degree, between two vocational training spells	0.28	
no degree, after vocational training/ no vocational training	1.36	*
vocational degree	*1*	
university degree	0.68	**
in education	0.14	***

There may however be another explanation in addition. In the imputed histories, respondents are coded as being in education on the basis of knowledge we have about their future educational enrollment. If however they become pregnant after exiting education, they may be less likely to re-enroll. Their future educational enrollment is thus a consequence of their childbearing behavior. An explanation for the difference in estimates for the 'no degree' category between the two models may thus be that the rates are flawed in the model using imputed histories as a consequence of anticipatory analysis.

Although the effect of having a university degree vs. having a vocational certificate is very similar for imputed and complete educational histories (Table 2), the interaction models (in Table 3) show that at all ages, rates of first birth are slightly underestimated for both women with vocational and university educations when we use imputed histories. An explanation may be that the imputed histories do not account for re-entries into education after a highest degree has been attained. In reality though, some respondents re-enter education after attaining their highest degree in order to acquire a further degree at the same level but in a different subject. Some have also re-enrolled in education after attaining their highest degree without completing any further degree at all. Thus, in the imputed histories, people who actually are in education will be falsely coded as being outside of education if they already have their highest degree. This could explain the slightly lower estimates of rates of transition to

the first child for women with vocational and university degrees in the imputed than in the complete histories.

4. Conclusion

The comparison of estimates using complete and imputed educational histories has shown that differences in rates of transition to first birth between women with vocational and university degrees are reproduced quite well when we use imputed educational histories. One exception is for women with no degree; for this group first-birth rates are overestimated considerably when we use imputed educational histories. An explanation may be that the imputed histories do not account for times spent outside the educational system between school and the start of further education. During this time, respondents have no vocational degree and are not participating in education. However, as they often already plan to take up further education shortly, first-birth rates are quite low during this time. The same applies to respondents who take short breaks between two spells of further education. In the imputed histories, this group of respondents is not included in the category 'not enrolled/no degree' (instead, they are in the category 'enrolled in education'). Excluding them from the category 'not enrolled/no degree' leads to an upward bias in the first-birth rate estimates for that category. A further explanation for the overestimation is that the imputation procedure actually is a form of anticipatory analysis when respondents are coded as being in education. Respondents who are not really in education, but of whom we know they will return to education, are coded as being in education in the imputed histories. Their future educational re-entry may however be a product of their present childbearing behavior. Those who do not have children may be more likely to re-enroll in education. Thus, in the imputed histories, those who will have a child may be more likely to be coded as having no degree, artificially raising the rate of transition to the first child for that educational category.

Altogether, the bias caused by using imputed histories did not turn out to be very serious in the case of western Germany. For the comparison of first-birth risks between women with completed vocational training and with a university degree, it was negligible. For countries where rates of educational reentry at later ages are sub-

stantially higher than in western Germany, it is likely that using imputed histories causes more serious biases.

Acknowledgements

References

Hoem, Jan M., and Michaela Kreyenfeld. 2006. Anticipatory analysis and its alternatives in life-course research. Part 1: Education and first childbearing. *Demographic Research* 15, no. 16: 461-484.

Hoem, Jan M., Gerda Neyer, and Gunnar Andersson. 2006. Education and Childlessness. The Relationship Between Educational Field, Educational Level, and Childlessness Among Swedish Women Born in 1955-59. *Demographic Research* 14, no. 15: 331-380.

Kerckhoff, Alan C. 2001. Education and Social Stratification Processes in Comparative Perspective. *Sociology of Education* 74. Extra Issue: Current Thought: Sociology of Education at the Dawn of the 21st Century: 3-18.

Kravdal, Øystein. 2004. An Illustration of the Problems Caused by Incomplete Education Histories in Fertility Analyses. *Demographic Research*. Special Collection 3, no. 6: 134-54.

Schütze, Hans G., and Maria Slowey. 2002. Participation and Exclusion: A comparative analysis of non-traditional students and life-long learners in higher education. *Higher Education* 44: 309-27.