



Demographic Research a free, expedited, online journal
of peer-reviewed research and commentary
in the population sciences published by the
Max Planck Institute for Demographic Research
Konrad-Zuse Str. 1, D-18057 Rostock · GERMANY
www.demographic-research.org

DEMOGRAPHIC RESEARCH

**VOLUME 21, ARTICLE 6, PAGES 135-176
PUBLISHED 11 AUGUST 2009**

<http://www.demographic-research.org/Volumes/Vol21/6/>

DOI: 10.4054/DemRes.2009.21.6

Research Article

**Do imputed education histories provide
satisfactory results in fertility analysis in the
western German context?**

Cordula Zabel

© 2009 Cordula Zabel.

This open-access work is published under the terms of the Creative Commons Attribution NonCommercial License 2.0 Germany, which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/by-nc/2.0/de/>

Table of Contents

1	Introduction	136
2	Potential sources of bias when using imputed histories	137
3	Data and method of imputation of education histories	144
4	Results	148
4.1	Vocational degrees: why are the estimates so close in the imputed compared to the original histories?	149
4.2	University degrees: differences between imputed and original histories	155
4.3	Differences in estimates for school degrees between the imputed and original histories	156
4.4	Imputation without knowledge of the date the highest degree was obtained	159
5	Conclusion	161
6	Acknowledgements	163
	References	164
	Appendix	166

Do imputed education histories provide satisfactory results in fertility analysis in the western German context?

Cordula Zabel¹

Abstract

In many surveys, information on respondents' education histories is restricted to the level and sometimes the date they attained their highest degree. We compare estimates of education effects on first birth transitions using imputed histories based on this rudimentary information with estimates drawing on complete histories, using the German Life History Study. We find that imputed histories produce relatively reliable estimates for most but not all education categories, especially when information on the date the highest degree was attained is available. We investigate possible explanations for these findings and indicate contexts in which biases may be stronger.

¹ Institute for Employment Research (IAB), Regensburger Straße 104, D- 90478 Nuremberg, Germany. Tel.: +49 (0)911-179-5956. Fax: +49 (0)911-179-5912. E-mail: Cordula.Zabel@iab.de.

1. Introduction

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

A very similar question has been investigated for the case of Norway. Kravdal (2004) addresses the implications of having no information on educational trajectories at all other than the highest level of education at interview, and finds that each of three methods of imputation lead to substantial deviations from the results produced when using complete education histories. In the present study, we investigate whether a little more information, that is, information on the date the respondent attained her highest level of education, can improve imputation and lead to satisfactory results, at least in the western German context.

The problem that incomplete education histories pose for the estimation of the effect of level of education on fertility transitions has also been pointed out by Hoem and Kreyenfeld (2006). The authors call attention to 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 education histories would be ideal. However, as they point out, complete education 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 Generations and Gender 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 education histories using this basic information. The analyses in the present study are based on their method of imputation.

The German Life History Study, from which we use data for our empirical analyses, is exceptional in that it does include respondents' complete education histories. First, we use these complete education histories to model the effect of educational level as a time-varying variable on transition rates to first birth. Next, all information is ignored other than the highest level of education attained by the time of

interview and the date this level was reached. Using only this rudimentary information, we conduct four different imputations based on alternative assumptions about how the education question is phrased in surveys that do not include complete histories. For additional comparison, we also impute histories using only the highest degree at interview, without any information on the date this degree was attained. We then re-analyze first birth risks using only these imputed education histories and compare the results with those produced using complete education histories. Our aim is to see to what extent the results are comparable.

The following section discusses possible problems that imputation may provide for fertility analysis. The third section describes the data set, our data preparation, and the procedure used for imputing education histories. The fourth section then presents the results of our model for the impact of education on rates of transition to first birth using complete as compared to imputed histories. We provide a summary in the fifth, and final, section.

2. Potential sources of bias when using imputed histories

The basic method of imputation used in this study is to consider respondents to be continuously enrolled in education up until the time they attained their highest degree. Education histories could be imputed perfectly in this manner if everyone followed very regular educational pathways. If everyone completed school² at a normative age, transferred directly from school to vocational or university education, and without interrupting their vocational or university education completed exactly one degree, imputed histories would correspond precisely to actual education histories.

However, educational trajectories will always be irregular to some extent. The diversity of education histories, and thus also whether education histories can be approximated sufficiently closely using only the date of attainment of the highest degree, or even only the highest degree itself, depends strongly on the country context. In some countries it is common to exit and reenter 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

² The term 'school' used here refers to secondary education, not to post-secondary education. This corresponds to the terminology most often used in the German context. In the remainder of this paper, secondary education will continue to be referred to as 'school,' since this is shorter. Vocational and university education on the other hand will be referred to as 'post-secondary education.'

one is employed (Kerckhoff 2001). The Swedish educational system is similarly 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 while employed (Hoem, Neyer, and Andersson 2006). We suspect that it would be difficult to approximate education histories using only sparse information for countries such as Sweden, Great Britain, or the United States.

The reason why imputation may be difficult in these countries is that there can be a great deal of variation in the timing of educational reentry. It may be very difficult to infer at what point in time people resume education exactly, or whether they transferred directly from school to post-secondary education. In countries with more flexible education systems, more students may also reenroll to acquire a further degree in a different subject but at the same level as before. This can also cause difficulties for imputation, since surveys generally only collect one date for respondents' highest degree. If, following our imputation method, we assume that people are enrolled in education up until the time they obtain their highest degree, substantial parts of respondents' activity histories are likely to be misrepresented. For instance, people will be coded as being enrolled in education for long periods of time during which they are actually employed or pursuing some other activity before obtaining their highest degree. Another concern is that they will be imputed as not enrolled after they obtained their highest degree, even if they do subsequently continue or reenter education.

Confusing time spent enrolled and not enrolled in education is especially problematic in the context of fertility analysis. In many studies, educational enrollment has been shown to have an extremely strong negative impact on rates of transition to first birth (e.g. in Rindfuss, Morgan, and Swicegood (1988), Blossfeld and Huinink (1991), Kreyenfeld (2008), Vikat (2004), and Edwards (2002)). Using imputed education histories may therefore become problematic if a lot of people continue their education without completing a further degree, or obtain a lower level degree after having already attained their highest degree, or complete a further degree at the same level as their previously completed highest degree. Since imputations are based on respondents' highest degree alone, any time of enrollment after they completed their highest degree is not registered. Respondents are instead imputed as not enrolled and holding their highest degree. However, as mentioned above, risks of first birth tend to be substantially lower for women who are in reality currently enrolled. Therefore, risks of first birth would be underestimated for non-enrollment categories when using the imputed histories. To some extent, this will be a problem in the context of any country, as some people will always continue their studies for some time without completing a further degree. However, it is likely to be especially relevant in countries that are supportive of students wishing to obtain a degree at the same level but in a different

subject than before, whether immediately after completing their first degree, or sometime later in the life course. If a survey only collects the first date respondents obtained their highest degree, imputations based on this information will misrepresent time respondents in reality spent enrolled after this date to obtain a second degree at the same level.

Gaps in respondents' education histories before they obtain their highest degree can be problematic for imputation as well. For the imputations, we assume that respondents were enrolled continuously up until they received their highest degree. Therefore, mis-imputations occur whenever respondents return to education and obtain a higher level degree at a later point in time after having interrupted education for some time. Gaps between education spells during which they held lower level degrees and were not enrolled will then be misrepresented as continuous enrollment. It may be easier for childless women than for women who had a child sometime during their lower degree spell to reenter education. Then, it will occur more frequently for women who remained childless that time they actually spent holding a lower level degree and not enrolled in education is imputed as time spent continuously enrolled in education.

This is illustrated in Figures 1 and 2. Figure 1 shows the case of a person who returned to education at a later point in time. First, this person was enrolled in education and by the end of her first education spell had acquired a lower level degree. She then spent some time not enrolled in education, during which she may have been employed, for example. During this time, she was holding her lower level degree. She then returned to education and gained a higher level degree. Her last spell is again spent not enrolled and holding a higher level degree. In the imputed education history, represented by the dotted line, she is assumed to have been enrolled continuously up until the time she gained her highest degree. The time during which in reality she was not enrolled and was holding a lower level degree is highlighted in red in the imputed history, which misrepresents this spell.

Figure 2 illustrates the case of a person who does not return to education. This person is also first enrolled in education and by the end of her education spell has acquired a lower level degree. Then, in her second spell, she is not enrolled and is holding a lower level degree, just like in the case illustrated in Figure 1. During this spell, she has a first child. Because combining education and childcare is difficult in her particular country context, she decides not to reenroll in education. Thus, she continues to hold a lower level degree up until the time of interview. In this case, the imputed history matches the original history, since the respondent actually was enrolled continuously for the entire time before she gained her highest degree.

Given that motherhood reduces the propensity of returning to education, imputation is more likely to be correct (in this particular respect) for people who had a child (Figure 2) than for those who did not have a child (Figure 1). For people who did

not have a child, time actually spent holding a lower level degree is more likely to be lost in the imputed histories (Figure 1). In effect then, exposure time belonging to people who did not have a child is removed from the 'lower degree' category in the imputed histories. This of course raises the estimated risk of first birth for the 'lower degree' category when using the imputed, as compared to the original, histories.

The problem thus is that imputations of the education variable, which is our predictor variable, depend on educational attainment at a later point in time, which in turn may have already been influenced by our dependent variable, the transition to the first child. As described above, we would expect the consequence to be an overestimation of transition rates to the first child for the status 'lower level degree, not enrolled,' because people who did not have a child are selectively taken out of this status. This problem is similar to the problem discussed formally by Hoem and Funck Jensen (1982). The authors analyze how bias is caused in fertility analysis when respondents are considered to be holding their highest educational degree throughout the entire risk period. In this study, we will empirically investigate the consequences of imputation when respondents are not considered to be holding their highest degree the whole time, but only from the date they obtained their highest degree onwards. Previous to this date, however, we consider them to be enrolled continuously. We will investigate to what extent this procedure still entails bias.

Figure 1: Reenrollment at a later point in time: imputed and original education histories

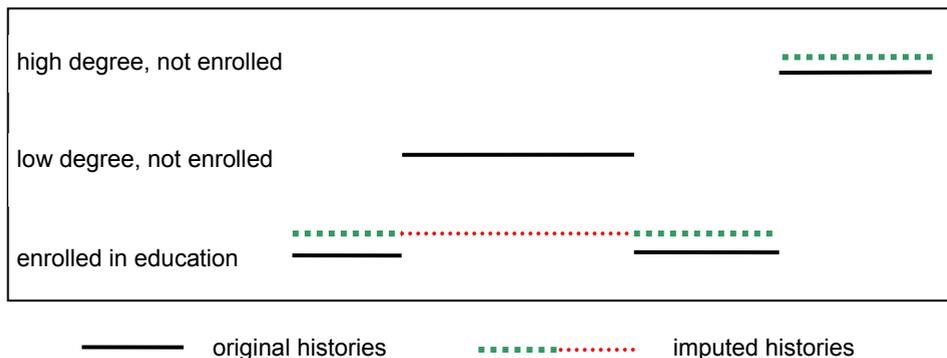
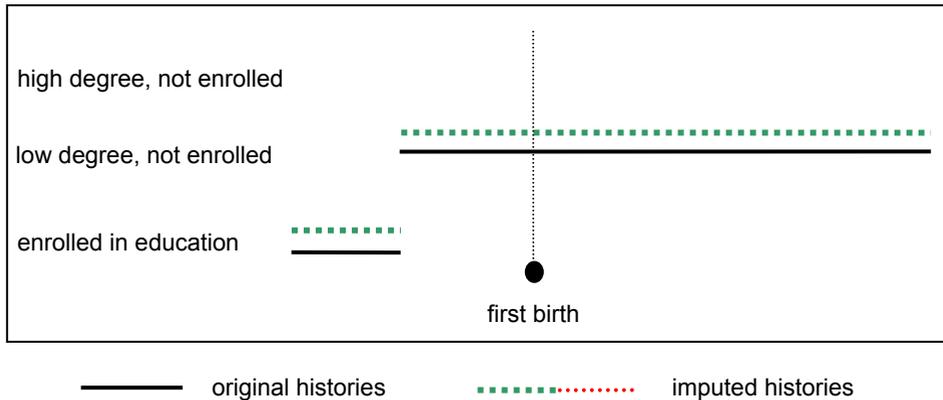


Figure 2: No reenrollment: imputed and original education histories

The same type of problem may occur if people obtain several different educational degrees at their highest level in sequence but the survey only asks for the last date they obtained their highest degree. If people obtain one degree, then work for some period of time, and then reenroll in education in order to obtain an additional degree at the same level (though perhaps in a different subject than before), the same selectivity problems described above may occur. If the information used for imputation refers to the date the respondent last obtained a degree at their highest level, time they spent outside of educational enrollment between spells would be misrepresented as a time of educational enrollment in the imputed histories. If reenrollment is more likely for those who did not have a child after obtaining their first degree, similar distortions to the ones described above are likely to occur.

Further problems may arise in contexts where it is common to obtain additional educational degrees on a part-time basis while employed or pursuing another activity. If, following our method of imputation, we use the date respondents received their highest degree to impute their duration of educational enrollment, the problem is that we are not accounting for differences between part- and full-time students. Educational enrollment may not have been respondents' main activity, and fertility rates could differ strongly between full-time students and students who are simultaneously employed.

For the present context, the question is to what extent these imputation problems might apply to Germany. Findings from previous research indicate important obstacles to educational reentry at later ages in Germany. If it is uncommon to resume education later in the life course, that would make imputation less problematic. Schütze and

Slowey (2002), for example, 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 comparatively inconducive 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 than in many other countries. In addition, there are few alternative modes of access for students who do not fulfill the regular entrance requirements. There is also no formal recognition of part-time student status.

Table 1 shows the distribution of post-secondary degree combinations for the cohort studied in the empirical analyses in this paper. Here we can see that 56% of the women in cohort 1964 had exactly one basic vocational degree, while 15% had two or more of this type of degree. Altogether, at least 27% of the female members of this cohort had more than one degree (given that a basic vocational degree is a necessary precondition to qualify for a master craftswoman's or technician's degree). This is in line with findings by Lauterbach and Weil (2008) who also report that a high proportion of members of an adjacent cohort, born in 1966/67, take up a second post-secondary education after their first. Thus, the extent to which respondents hold multiple degrees does not seem to be negligible. This could imply some difficulties for imputing education histories on the basis of information on the highest degree alone.

However, there seems to be some evidence that in Germany, second vocational degrees are often obtained shortly after the first. Jacob (2005) reports that rates of transition to second vocational training episodes are especially high immediately upon completing a first vocational degree. Furthermore, among people who are already working, rates of transition to second vocational training spells are highest at only short employment durations. This kind of a pattern is very likely to indicate that gaps between first and second training episodes are quite small, which would make imputation less problematic (although special cases are conceivable where long average gap lengths are compatible with these findings as well). As long as we know the last date the respondent received a degree, generally short gaps would mean that we can quite safely apply our method of imputation and consider respondents to be enrolled up to that date. Only short durations of time between training spells would then be misrepresented as a consequence of imputation.

Table 1: Post-secondary degrees attained by time of interview: degree combinations^a. Women born in West Germany in 1964, interviews in 1998/ 99

no degree	10%
basic vocational degrees only	70%
<i>number of basic vocational degrees</i>	
1	56%
>1	15%
master craftswoman's/ technician's degree	8%
university/ college degree only	9%
vocational ^b and college/ university degree	4%
	100%
	N=634

Source: GLHS, cohort 1964

(sample used for empirical analyses, excluding cases with missing values for education)

Notes: ^a 'Basic vocational degree' is our translation for 'Ausbildungsabschluß,' 'master craftswoman's/ technician's degree'

translates as 'Meister' or 'Techniker,' and 'university or college degree' refers to 'Hochschulabschluss' or

'Fachhochschulabschluss.' A master craftswoman's degree is a higher level vocational degree giving permission to train apprentices and, at least in the past, also necessary to head an enterprise specializing in one's vocation.

^b basic vocational or master craftswoman's/ technician's degree

As can be seen in Table 1, it appears that a substantial proportion of women in cohort 1964 hold multiple degrees at the same level. Therefore, for the western German context, it might make a difference whether in a survey, respondents are asked for the first date or the last date they received a degree at a given level. If, as we suspect, gaps between training spells are small, then it would be better to ask for the last date the respondent received her highest degree. Our method of imputation would then consider respondents to be continuously enrolled until the last time they received a degree at their highest level. If gaps between training spells are small compared to the length of the last spell, the extent of misrepresentation would be smaller using the last than the first date the respondent received a degree at her highest level. To investigate whether, for the purpose of imputing education histories, it would make a difference for survey questionnaires to ask for the first or the last date respondents received a degree at their highest level, separate imputations simulating each case are conducted for the empirical analyses.

Another way to ensure that we get the last date a respondent received a degree is to provide more differentiated degree levels in the survey questionnaire. In a simple list of degree categories, a basic vocational degree and a master craftswoman's degree, for example, would both just be considered to be a vocational degree. If, however, the survey explicitly differentiates between basic vocational degrees and master

craftswoman's degrees, or between a standard university degree and a doctoral degree, (or, in different contexts, also between a bachelor's and a master's degree), we would be more likely to get the last date the respondent ever received a degree by asking for their highest degree than if the survey only included simple, undifferentiated degree categories. For the empirical analyses, separate imputations are also performed assuming that the questionnaire uses a simple vs. a more differentiated list of degree categories.

3. Data and method of imputation of education histories

For the empirical analyses, data from the project "Education, Training, and Occupation: Life Courses of the 1964 and 1971 Birth Cohorts in West Germany" is used³. The data includes retrospective information on many different realms of the life course. We used data for the cohort born in 1964, since first birth histories for this cohort were sufficiently advanced by the time of the interview in 1998/99. We excluded respondents who were not born in West Germany, since our objective is to determine how well education histories can be imputed in the context of the West German education 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.

Our method of analysis is event-history analysis, where the dependent variable is the rate of transition to first birth. The date of first birth was backdated by nine months, to better account for the influence of educational status at the point in time the decision to have a first child was made. First, complete education 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. Next, several different imputations were conducted based on different assumptions about information available in surveys lacking complete education histories.

The specification of the education variable chosen differentiates between enrollment in school education, vocational training, and university education. The idea here is to account for the social context connected with the type of educational institution people are attending. For those not enrolled in education, several different levels of educational attainment were included, encompassing both school degrees as

³ The project "Education, Training, and Occupation: Life Courses of the 1964 and 1971 Birth Cohorts in West Germany" is part of the German Life History Study (GLHS), conducted by the Max Planck Institute for Human Development, Center for Sociology and the Study of the Life Course (Berlin) in cooperation with the Institute for Employment Research (IAB, Nürnberg). For documentation, see Hillmert et al. (2004).

well as post-secondary degrees⁴. A person's current degree refers to the highest degree the person has attained up to that point in time. Post-secondary degrees are generally considered to be higher than school degrees. Table 2 shows estimation results for this model specification using both imputed and complete histories. These results will be discussed in the next section.

As mentioned earlier, the imputations conducted here are based on the method presented by Hoem and Kreyenfeld (2006). They develop a method of imputation using the respondent's 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. However, for the results to be more generally applicable, in contrast to Hoem and Kreyenfeld (2006), we do not use the date the respondent finished school, unless a school degree was the highest degree ever attained, nor do we use the starting date of the respondent's first job, as this information cannot be expected to be generally available in data from surveys focusing on fertility histories. The general method of imputation is to consider respondents to be continuously enrolled in education up until the point in time at which they attain their highest degree. The remainder of this section explains how this was done in detail.

We perform five different imputations based on different assumptions about the type of information available from surveys that do not include complete education histories. When asking for the highest level of education a respondent has attained by the time of interview and the date this level was attained, a questionnaire generally will not specify whether to give the first or the last date a degree at this level was attained. However, as pointed out in the previous section, this can make a difference. Therefore, education histories are imputed alternately assuming that the questionnaire asked for the first or the last date the highest degree was attained. This was done by taking from the complete histories either the first or the last date a respondent received her highest degree, and imputing the histories from there.

As pointed out in the previous section as well, it can make a difference for models of the effect of education on first birth risks whether the questionnaire uses simple or more differentiated education categories. Thus, different imputations are performed assuming each case. In addition, a further imputation of education histories is undertaken assuming that no information on the date a respondent attained her highest degree is surveyed at all.

⁴ 'Lower secondary degree' is our translation for 'Hauptschulabschluß,' 'advanced lower secondary degree' refers to 'mittlere Reife,' upper secondary degree to 'Hochschulreife' or 'Fachhochschulreife,' 'vocational degree' to 'Ausbildungsabschluß,' and 'university or college degree' refers to 'Hochschulabschluß' or 'Fachhochschulabschluß.' An upper secondary school degree is generally required to enroll at a university.

Table 2: Effect of education on risk of first birth. Results using original education histories compared to results using education histories imputed by different methods

	original histories	imputed histories						without information on date of highest degree at interview was attained
		simple or differentiated degree classification			without information on date of highest degree at interview was attained			
		simple	simple	differentiated	simple	simple	differentiated	
		first date or last date highest degree at interview was attained			highest degree at interview was attained			
	first	last	first	last	first	last		
age (absolute risks)								
15-19	0.0036	0.0037	0.0040	0.0038	0.0040	0.0040	0.0038	0.0038
20-24	0.0061	0.0059	0.0066	0.0060	0.0066	0.0066	0.0053	0.0053
25-29	0.0125	0.0120	0.0130	0.0123	0.0130	0.0130	0.0115	0.0115
30-34	0.0099	0.0097	0.0100	0.0099	0.0100	0.0100	0.0095	0.0095
education (relative risks)								
no degree	0.55	0.65	0.59	0.63	0.59	0.59	0.67	0.67
school degrees								
lower sec. degree	1.52 *	1.98 ***	1.80 ***	1.92 ***	1.81 ***	1.81 ***	2.02 ***	2.02 ***
adv. lower sec. degree	1.53	1.67	1.51	1.62	1.52	1.52	1.69	1.69
upper sec. degree	0.63	0.73	0.67	0.71	0.68	0.68	0.84	0.84
post-sec. degrees								
vocational degree	1	1	1	1	1	1	1	1
univ./coll degree	0.68 **	0.65 **	0.65 **	0.64 **	0.64 **	0.64 **	0.63 ***	0.63 ***
enrolled in education								
in school	0.09 ***	0.08 ***	0.07 ***	0.08 ***	0.07 ***	0.07 ***	0.06 ***	0.06 ***
in vocational training	0.20 ***	0.18 ***	0.21 ***	0.18 ***	0.21 ***	0.21 ***	0.19 ***	0.19 ***
in university	0.14 ***	0.17 ***	0.20 ***	0.22 ***	0.20 ***	0.20 ***	0.12 ***	0.12 ***
missing	0.58	0.17 *	0.15 *	0.16 *	0.15 *	0.15 *	0.17 *	0.17 *

Notes: ***: $p < 0.01$; **: $0.01 \leq p < 0.05$; *: $0.05 \leq p < 0.1$

Source: GLHS, cohort 1964

Sample: women born in West Germany

In the first four imputations, if a respondent's highest degree was a school degree, she was assumed to be attending school up to that date. Since respondents may have started vocational training or university education, but dropped out without receiving a degree, we imputed a short duration of post-secondary education. If the school degree was an (advanced) lower secondary degree, the respondent was assumed to have attended vocational training for one year, without completing this training or receiving a vocational degree. Thus, after this time, she was considered to be not enrolled in education but to be holding an (advanced) lower secondary school degree up until the time of interview. If the school degree was an upper secondary degree, the respondent was considered to have attended university for two years, likewise without receiving a degree, so that after this time and up until the date of interview, the respondent was considered to hold an upper secondary school degree.

If the respondent's highest degree was a vocational degree, she was first considered to be enrolled in school, then in vocational training up until the date she attained this degree, after which time she was considered to be holding a vocational degree and not to be enrolled in education. To impute the date the respondent switches from school to vocational training, we took into account the respondent's year and month of birth. If the respondent was born between January and September, she was considered to have attended school until July the year she turned 16. If she was born between October and December, she was considered to have attended school until July the year she turned 17. This broadly corresponds to the duration of school attendance for people with an advanced lower secondary degree, the most common type of school degree among those with a vocational degree. Generally, an advanced lower secondary degree can be obtained after ten years of school (although there is some variation here across time and federal states). Children generally begin school the year they turn six if they were born before a specific month, which also varies across time and federal states. However, the most common cut-off month is October.

If the respondent's highest degree was a university degree, she was considered to have attended school up until July the year she turned 19 or 20, depending on her birth month. Generally, an upper secondary degree, which is the most common school degree among those with a university degree, can be obtained after 13 years of school. The respondent is considered to switch directly from school to university education, and to remain in university education up until the date she received her university degree. Then, up until the date of interview, she is considered to be holding a university degree and not to be enrolled.

If the respondent has no degree, she is considered to have attended school up until July the year she turned 15 or 16, which corresponds to nine years of school, generally the legal minimum duration of school attendance. She is then considered to have taken part in some type of vocational training program for one year, without, however,

attaining any degree. After this time, up until the time of interview, she is considered to have no degree and not to be enrolled.

The above description applied to the first four imputations, where we assume that we have information on the date the respondent received her highest degree. For the fifth and last imputation, however, we assume that the survey contains no information on the date the respondent received her degree, and only provides information on the level of that degree. Here then, in addition to imputing the date the respondent switched from school to vocational or university education, we also need to impute the date the respondent received her vocational or university degree. We imputed the date the respondent received a vocational degree as July, three years after completing school. Vocational training can take between two and four years, so three years was chosen as an intermediate duration. The date a respondent received a university degree was imputed as July, six years after finishing school. Six years is a realistic duration for many subjects, although study durations can vary greatly in either direction.

If the respondent's highest degree was a school degree, in the fifth imputation where we do not know the date the respondent received this degree, we need to differentiate between lower secondary school degrees, advanced lower secondary degrees, and upper secondary degrees. For lower secondary degrees, we assume the respondent attended school up to ninth grade, which corresponds to July the year she turned 15 or 16, depending on her birth month. For advanced lower secondary degrees, we impute July the year the respondent turned 16 or 17, and for upper secondary degrees, July the year the respondent turned 19 or 20.

The next section compares estimation results for the effect of education on risks of first birth using the complete and imputed histories, and discusses explanations for deviations.

4. Results

The results in Table 2 show that the first four imputations generally produce estimates that are quite close to the ones we obtain using the original histories. In terms of estimates for the baseline, deviations from the original histories are quite small in each of the four cases. Estimates of the effect of having a vocational as compared to a university degree are likewise very close to the estimate produced using the original histories. The strongest deviation between the results obtained from the imputed as compared to the original histories, however, occurs for the relative risk of first birth for having a lower secondary vs. a vocational degree. The original histories produce an estimate of 1.52 for the relative risk of first birth for having a lower secondary degree as compared to the reference category. The first four imputed histories, on the other hand,

give estimates ranging between 1.80 and 1.98⁵. The remaining education categories generally encompass only a small number of events⁶. The results for these categories probably should not be over-interpreted. Even for these though, deviations appear to be quite small. So, the question is why the estimates derived from the imputed histories are so close to the ones given by the original histories (with the exception of the effect of having a lower secondary vs. a vocational degree)?

4.1 Vocational degrees: why are the estimates so close in the imputed compared to the original histories?

Estimation results for the category ‘vocational degree’ in principle could be particularly strongly affected by using imputed instead of original histories. This is because many people gained another degree after having obtained a vocational degree first. This applies to 27% of the respondents⁷ (Table 1). Altogether, 15% have more than one basic vocational degree, 8% have a master craftswoman’s or technician’s degree in addition to their basic degree, and 4% have both a university and a vocational degree. Thus, it is of particular interest to see how exposure time and events are allocated into and away from the category ‘vocational degree’ in the imputed histories as a consequence of the respondents obtaining subsequent degrees at the same or a higher level, and because of gaps between spells of educational enrollment.

The tables provided in the appendix map exactly to which education categories exposure time and events from the original histories are allocated in the imputed histories. Tables A1a - A1c do this for the imputation that was performed assuming that respondents were asked for the first date they received their highest degree⁸. In Table A1a in the appendix we can see that 99% of the exposure time from the original

⁵ Table A4 in the appendix gives standard errors for the estimates shown in Table 2. Here we can see that the standard error for the estimate affected most by imputation, the estimate for the relative risk of first birth for those with a lower secondary school degree as compared to a vocational degree, increases when using the imputed histories. However, the estimates of this relative risk not only remain significant, they are significant on a higher level in the imputed as compared to the original histories. This is because the size of the estimate itself increases strongly enough to offset the increase in the standard error.

⁶ The reason these categories were included was mainly to avoid grouping them together with other categories for which we do have a sufficient number of events. This way, effects for vocational, university, or lower secondary degrees, for example, could be kept as clear-cut and generally comparable as possible.

⁷ given that the 4% of respondents with both a vocational and a university degree obtained their vocational degree first.

⁸ The tables provided in the appendix only give cross-tabulations of exposure time and events for the first, second, and fifth imputation. Cross-tabulations for the third and fourth imputation were omitted in order to save space. General patterns of allocation of exposure time and events between the original histories and imputed histories can be demonstrated sufficiently well on the basis of the three examples shown here.

histories during which respondents held a vocational degree was correctly allocated to the category 'vocational degree' in the imputed histories (Table A1a). However, additional exposure time is also allocated to this category in the imputed histories as can be seen in Table A1b. Only 93% of the exposure time allocated to the category 'vocational degree' in the imputed histories actually stems from this category in the original histories, while 6% originates from the category 'in vocational training' in the original histories. In this imputation, we are assuming that we know only the first date the respondent obtained the highest degree they held by the time of interview. As described in the methods section, respondents are assumed to be enrolled in education up to this date and then to be holding this degree without being enrolled in education up until the date of interview. Thus, if respondents reenter education after the first time they received a degree at their highest level, this reenrollment is not registered in the imputed histories. This appears to be reflected in the 6% of the imputed exposure time for 'vocational degree' during which respondents actually were enrolled in vocational training in the original histories.

The other way around, for the imputation using the last date the respondents received their highest degree (Tables A2a - A2c), not all of the exposure time from the category 'vocational degree' in the original histories is correctly allocated to that category in the imputed histories. Of the exposure time that respondents actually spent holding a vocational degree and not enrolled in education, 10% is instead allocated to educational enrollment in the imputed histories (Table A2a). In this imputation, we are assuming that we only know the last date the respondents received their highest degrees, and we are imputing them to be enrolled in education up to that date. Therefore, whenever there are gaps between education spells, original time spent holding a vocational degree in these gaps is mis-imputed as time spent enrolled in education in the imputed histories. By contrast, for this imputation, 99% of the imputed time spent holding a vocational degree actually originates from that category in the original histories (Table A2b).

Thus, the extent of misallocation of exposure time into (first imputation) or away from (second imputation) the category 'vocational degree' is comparable and not very large. In the first imputation, we only have the first date the respondents received their highest degree. As we have seen from Table A1b in the appendix, time that is actually spent enrolled in education after educational reentry is imputed as time spent holding a vocational degree without being enrolled in education. This will tend to lead to an underestimation of risks of first birth for the category 'vocational degree,' because time spent enrolled in education, during which first birth risks are typically very low, is added to the 'vocational degree' category in the imputed histories. This is reflected in the disproportionately low number of events (Table A1c) added to the category 'vocational degree' as compared to the amount of additional exposure time (Table

A1b). In turn, the estimation results in Table 2 for the imputations that use the first date the highest degree was attained give lower risks of first birth for the category ‘vocational degree’ than do the results using the original histories. The category ‘vocational degree’ is the reference category here, so we can compare risks of first birth for this category just by looking at the baseline. The estimates for the baseline are generally slightly lower in the first and the third imputation, which use the first date the highest degree was attained, than in the original histories, except for the youngest age group⁹ (Table 2).

By contrast, when the last date the respondents attained their highest degree level is used for the imputations, this should in principle lead to an over-estimation of the risk of first birth for the category ‘vocational degree.’ This is because, as described in section 2 on potential sources of bias, people who do not have a child after they receive their first vocational degree may be more likely to go on to obtain a second vocational degree. Then, they will be selectively taken out of the category ‘vocational degree’ in imputations using the last date a vocational degree was attained (and instead imputed to be enrolled in education in the gap between the two spells of educational enrollment). Comparing Tables A2a and A2c, it seems that this is reflected in the disproportionately small number of events that are removed from the category ‘vocational degree’ as compared to the 10% of exposure time that is removed from this category in the imputed histories. This appears indeed to result in an overestimation of the risk of first birth for the category ‘vocational degree’ in the model estimation results for the second and fourth imputations in Table 2, which use the last date the highest degree was attained. Again comparing the baseline estimates for these imputations to those for the original histories (since ‘vocational degree’ is the reference category), we can see that the risk of first birth is slightly overestimated.

But, altogether, deviations from the original baseline estimate (which gives the risk of first birth for the reference category ‘vocational degree’) are not very large in any of the first four imputations. In the section on potential biases caused by imputation, it was suggested that although it appears to be quite common for members of this cohort to hold multiple degrees, gaps between spells of educational enrollment may be quite small. It is likely that the estimates would become more severely distorted for imputations using the last date the respondents received their highest degree if more people tended to obtain a second degree much later in their life course and gaps between education spells grew larger. Imputations using the last date respondents received their highest degree would then produce results that more strongly overestimate the risk of first birth for the category ‘vocational degree.’ However, this

⁹ This is likely to be due to the fact that not many people have completed their first vocational degree and already started on their second before the age of 19, and that birth risks at this age generally are very low.

would only be the case so long as it remains uncommon for people to reenter vocational training after having children. Alternatively, in the future, educational trajectories might change in such a way that people commonly gain additional vocational degrees later in their life course even after having children. If reentry into vocational training thus becomes independent of motherhood, false allocation of exposure time away from the category ‘vocational degree’ into the category ‘in vocational training’ would then no longer occur disproportionately often for the childless, and the risk of first birth for the category ‘vocational degree’ would no longer be overestimated in imputed histories using the last date the highest degree was attained. Given that scenario though, a different problem would arise when using the last date the highest degree was attained for the imputations. People who are actually not in vocational training would generally be imputed to be in vocational training for long periods of time if they receive their last degree comparatively late in the life course. Since birth risks are likely to be relatively high in the periods when respondents are in reality not in vocational training but represented to be in vocational training in the imputed histories, this would bias the estimate of the risk of first birth upwards for the category ‘in vocational training.’

Thus it seems that it would be a better idea to use the first date the highest degree was obtained than the last date if a very large proportion of people reenter vocational training comparatively late in their life course. Using the first date would then involve only a comparatively small amount of mis-imputation if durations of time spent in vocational training after having received one’s highest degree for the first time are short compared to the length of the gap between training spells. If, however, people repeatedly reenter vocational training very often throughout their life course and the amount of time spent in vocational training is altogether very large, but there are still fertility differences between people in education and not in education, then estimation results based on imputations using either the first or the last date the highest degree was obtained are likely to be biased. In that case, it seems imputed histories could not be a satisfactory substitute for original histories, and it would always be best to collect information on complete education histories in surveys intended to provide data for fertility analyses.

It seems that in the cohort analyzed in this study, not many people (only 4%) obtained a university degree after having already obtained a vocational degree. If it were more common to receive a university degree after a vocational degree and there were large gaps between vocational training and university education spells, that would also make imputation more problematic. The university degree, since it is the highest degree, would always be the one for which the date is recorded, independent of whether first or last highest degree dates are used for the imputations. It may be more common for childless women with a vocational degree than mothers to take up university education. Then, exposure time originating from the ‘vocational degree’ category would

be more frequently misallocated to the ‘in university education’ category for the childless. This would again bias the estimate for ‘vocational degree’ upward, since those who are childless would be selectively taken out of this category in the imputed histories. This happens to only a very small extent in the current imputations, as can be seen for instance in Table A1a in the appendix. Only 1% of the exposure time originally spent holding a vocational degree is imputed as time spent enrolled in university education. This is because, as already mentioned above, in the cohort analyzed here, it is very uncommon to begin university education after having already obtained a vocational degree. This is not very likely to change in the future, since an upper secondary school degree is generally required to enroll at universities, which is an obstacle for many people. In other countries though, it may be more common to take up university education later in the life course after already having held a lower level degree for some time. In those contexts, imputations would be more problematic.

A similar problem might turn up in Germany, however, if more women obtain a master craftswoman’s or technician’s degree after their basic vocational degree. In the cohort analyzed here, only 8% had a master craftswoman’s or technician’s degree. If in later cohorts, it becomes more common for women to obtain these higher level vocational degrees, and if they often do so later in the life course, imputation might become more problematic for reasons parallel to those described above for obtaining university degrees after vocational degrees. In the models estimated in this study, those with master craftswoman’s or technician’s degrees were not distinguished from those with basic vocational degrees. It might nonetheless make a difference for the estimates whether this differentiation is explicitly made in the questionnaire or not. If people are asked to give the first date they obtained their highest degree, and master craftswoman’s degrees are explicitly listed in the questionnaire, those who have a master craftswoman’s degree will choose that category (instead of just ‘vocational degree’) as their highest degree. Even if they are asked to give the first date they obtained their highest degree, they will give the date they obtained their master craftswoman’s degree instead of the date they obtained their basic vocational degree. The third and the fourth imputations simulate that case, assuming the questionnaire included differentiated degree categories (Table 2). Here we can see that in the third imputation, that assumes the questionnaire asked for the first date the respondent received her highest degree but provides differentiated degree categories, estimates for the baseline are between those for the first and second imputation. Using the first date the highest degree was obtained works towards underestimating the baseline like in the first imputation. But, using the date respondents obtained a master craftswoman’s degree (for respondents who have that degree) instead of the date they first obtained their basic vocational degree, works toward overestimating the baseline. Since only few respondents in this sample actually had a master craftswoman’s degree, it appears that the result was altogether still a slight

underestimation of the baseline. If in later cohorts more women obtain master craftswomen's degrees, using the date they received that degree might lead to a major overestimation of the risk of first birth for the category 'vocational degree,' since people who do not have a child and go on to gain a master craftswoman's degree are selectively taken out of the 'vocational degree' category in the imputed histories as delineated above.

To summarize, it seems one reason deviations in the estimates for the vocational degree category in the imputed, as compared to the original, histories are so small is that gaps between successive vocational training spells are not very large. Although 23% of the people in this sample obtained more than one vocational degree (a supplementary master craftswoman's degree or an additional basic vocational degree), gaps between vocational training spells were not large enough to lead to extensive misallocation of exposure time. Not many people seem to obtain second vocational degrees very late in their life course. In addition, in this cohort, it was very uncommon to obtain a university degree after a vocational degree (only 4% of the respondents did so). In countries where access to university education is less restrictive, imputations using only the highest educational degree might lead to greater distortions in estimates of first birth risks for lower educational degrees that people held before obtaining their university degree. In addition, only a comparatively small proportion of this cohort held a master craftswoman's degree in addition to a basic vocational degree (8%). Distortions of estimates based on imputed histories could grow larger for later cohorts if more women go on to obtain this higher level vocational degree.¹⁰

Reentries into or continuations of education after respondents already obtained their highest degree also did not prove to cause much bias when using the imputed histories. The greater the amount of time respondents spend enrolled in education after they are considered to have already obtained their highest degree, the greater the extent of downward bias to be expected for those highest degree categories. Particularly for the imputation that used the first date respondents obtained their highest degree, a

¹⁰ A further problem for imputations is that not everyone is enrolled in vocational training full-time before obtaining a master craftswoman's degree. Many are enrolled part-time while employed. If in later cohorts more women go on to obtain a master craftswoman's degree, this too would make our imputation method problematic, since we always assume that people are enrolled in education full-time up until the point they obtain their highest degree.

The original histories used here cannot give much insight as to the extent of part-time enrollment among those aspiring to a master craftswoman's degree. This is because educational enrollment was recorded somewhat imprecisely for this group. For some who obtained a master craftswoman's degree, enrollment was recorded, while for others, it was not clear whether they were enrolled full-time or not and when their training spell began. We considered those for whom we had no information on enrollment not to be enrolled in our preparations of the original histories. This is not likely to have strongly affected the results, since the group of people who obtained a master craftswoman's degree is altogether quite small, and the imprecision only applies to a fraction of them.

greater downward bias for the ‘vocational degree’ category might have been expected. Even though it is quite common in Germany to have more than one vocational degree at the same level, vocational training spells are generally only two to four years long, making up only a small proportion of respondents’ total exposure time after they obtained their first vocational degree. Respondents’ average number of vocational training spells would have to be quite high in order to substantially raise the extent of misimputation. It was also quite uncommon to have an incomplete university education spell after having completed a vocational training spell. In countries where university entry regulations are more flexible, this source of underestimation of first birth risks for lower level degree categories may be more relevant.

4.2 University degrees: differences between imputed and original histories

Estimates for the category ‘university degree’ are also quite similar in the imputed and the original histories (Table 2). Looking at Table A1b in the appendix, however, we can see that only 85% of the exposure time imputed for the university degree category actually stems from that category in the original histories. It seems that some people enroll in vocational training after getting their university degree. This is not registered in the imputed histories, since the university degree is the highest degree. But this does not seem to cause any distortions, since a nearly proportionate number of births are allocated to the ‘university degree’ category along with the extra exposure time. However, the overall number of events involved is too small to draw any general conclusions. It is impossible to say on the basis of this data whether first birth risks really are just as high for people enrolled in vocational training after completing their university degree as they are for people who have completed their university degrees and do not subsequently enroll in vocational training.

In contexts where first birth risks actually are lower for those who reenroll, there might be some potential for distortion if it is very common for people to enroll in vocational training after having already gaining a university degree. Then, this enrollment in vocational training would not be registered in the imputed histories because the university degree is already counted as the highest degree. Another possible source for distortions in other contexts or for future cohorts would be if many more people went on to gain degrees at a higher level than a general university degree, for example a doctoral degree, or if many people obtained a second university degree after their first.

For the cohort studied here, international bachelor’s and master’s degrees had not yet been introduced in Germany. This has only occurred very recently. Therefore, for this cohort, university students generally only received one degree; there were no

different levels of university degrees. However, wherever there is a differentiation between master's and bachelor's degrees, this may provide problems for imputations. If questionnaires only ask for respondents' highest degree, this would mean we would often only have the date they obtained their master's degree. Then, we might risk overestimating first birth risks for the 'bachelor's degree' category. This corresponds to the problems found for cases when we only have the last date respondents received a vocational degree, described above. If enrollment in studies preparing for a master's degree is more common for those who have not already become mothers, overestimations of risks of first birth for the bachelor's degree category could likewise occur.

4.3 Differences in estimates for school degrees between the imputed and original histories

As mentioned earlier, the greatest deviation in the estimates using the imputed compared to the original histories applies to the category 'lower secondary degree.' This is a school degree that is generally received after the ninth or tenth grade. For respondents who do not subsequently go on to obtain a vocational degree, this is their highest degree. In the imputed histories, they are the only ones who contribute exposure time to the category 'lower secondary degree.' In the original histories though, other respondents also contribute exposure time to this category. Respondents who do not go directly from school to vocational training, but wait for some time before starting their vocational training also contribute exposure time to the 'lower secondary degree' category during the gap between school and vocational training. As we can see in Table A1a, of the exposure time originating from the 'lower secondary degree' category, only 75% is correctly allocated to that category in the imputed histories. People who held that degree in the gap between school and vocational training or before reenrolling in school were imputed to have been in vocational training or in school the whole time. However, if people have a child after finishing school, this is likely to affect the probability of their enrolling in vocational training. Those who do not have a child may be more likely to enroll. If this is the case, people who do not have a child during their 'lower secondary degree' spell would be selectively taken out of the 'lower secondary degree' category in the imputed histories, since they are more likely to go on to vocational training and to be imputed as having been enrolled the entire time after finishing school. For them, the gap during which they held a lower secondary degree would be neglected. Indeed, the number of births removed from the 'lower secondary degree' category is disproportionately low compared to the amount of exposure time allocated away from that category in the imputed histories (Table A1a and Table A1c).

In addition, further births are allocated into this category from the categories ‘in vocational training’ and ‘in school.’ The number of births added to this category is disproportionately high compared to the amount of additional exposure time (Table A1b and Table A1c). Thus, it seems that some people dropped out of vocational training or further school education after becoming pregnant without completing their vocational degree or higher level school degree¹¹. The lower secondary school degree then remains their highest degree, and the time during which they were enrolled in vocational training or higher level school education is not registered in the imputed histories. They are therefore represented as having become pregnant while they were not enrolled and were holding a lower secondary school degree, although in reality, they became pregnant while enrolled in vocational training or higher level school education. These two factors – people who did not have children being selectively removed from the ‘lower secondary degree’ category, and people who did have children being added to this category – seem to have caused the overestimation of first birth risks for the category ‘lower secondary degree’ in the imputed histories¹².

The size of the bias may even be underestimated here. This is because the original histories may give slightly biased results themselves. In the form they were distributed, the original histories were pre-prepared so as to close small gaps of up to four months between school spells or between school and subsequent activity spells (like vocational training spells) (Hillmert et al. 2004, part IV, p.24). Just like in the imputed histories, this has the potential to lead to an overestimation of first birth risks for school degree categories. Thus, differences between the original and imputed histories only arise if the gap between school and vocational training is more than four months long. If small gaps of up to four months had not been filled in the original data, then the difference in estimates between the imputed and original histories would likely have been even greater¹³.

Although the estimate for the category ‘lower secondary degree’ was quite strongly biased in the imputed histories, not much bias was found for the other types of

¹¹ As described in the methods section, the date of first birth was backdated by nine months. Thus, strictly speaking, the dependent variable is always the rate of transition to first pregnancy. However, we have often used the term ‘first birth’ since this is more customary.

¹² In the case of the ‘lower secondary degree’ category, the imputed amount of exposure time and events does not differ between the first and second type of imputation shown in the appendix. This is because it is generally not possible to receive the same school degree twice, thus there is no difference between the first and last date the highest degree was obtained.

¹³ The data preparation procedure used for the originally distributed data closes small gaps of up to four months between school and vocational training spells by extending the length of the school spells. In the imputations used for the present analyses, gaps between school and vocational training spells are effectively closed by extending the length of the vocational training spell backwards up to the point in time school education is assumed to have ended. Both procedures, however, in the same way lead to an overestimation of first birth risks for school degree categories.

school degrees, that is, for the categories ‘advanced lower secondary degree’ and ‘upper secondary degree’ (Table 2). As we can see in the appendix, this is the case despite a large extent of misallocation of exposure time for these categories. Only 49% and 33%, respectively, of the original exposure time is correctly allocated to these school degree categories. People who gain an upper secondary school degree, on the basis of which they can apply for university, especially tend to take long breaks before enrolling at a university (or in vocational training). They therefore in reality spend a lot of time holding an upper secondary school degree without being enrolled, which in the imputed histories is represented as a time of enrollment in university education (or vocational training), for all of those who do in the end receive a university (or vocational) degree. Distortions of the model estimates for the school degree categories ‘advanced lower secondary degree’ and ‘upper secondary degree’ do not seem to be dramatic. However, the number of events for these categories is too small to draw any general conclusions. Larger sample sizes could show whether the uptake of post-secondary education is influenced by motherhood for these school degree categories or not, and whether or not imputation would therefore lead to distortions.

A detail of the imputation procedure used here may have also contributed to misallocation of exposure time. As described in the methods section, for people without a post-secondary degree, a short period of post-secondary education was imputed nonetheless. The idea was that people who had not completed a post-secondary degree may still have begun post-secondary education, but have dropped out without completing a degree. Only a short period of one or two years, depending on respondents’ type of school degree, was imputed to account for this. This may have still overstated average lengths of enrollment in post-secondary education for respondents without a post-secondary degree though, since many may have never begun any post-secondary education at all. A reason that some respondents never entered post-secondary education may be that they had a child. Exposure time which these respondents in reality spent holding a school degree is allocated away from the school degree categories as a consequence of this imputation procedure. This is likely to bias the first birth estimates for school degree categories downward when using the imputed histories. This source of mis-imputation will thus work in the opposite direction to the source of mis-imputation discussed above, which worked towards overestimating first birth risks for school degree categories. If lengths of enrollment in post-secondary education for those without any post-secondary degree are indeed overstated by the imputation procedure used here, and if a method could be found to more realistically represent these interrupted phases of post-secondary education in the imputed histories, then, in sum, the overestimation of first birth risks for school degree categories would be even greater. This is because the first source of mis-imputation, which led to an

overestimation, would no longer be counteracted by this second source of imprecision in the imputed histories.

Altogether, although the bias for the ‘lower secondary school degree’ category is the greatest of all categories examined here, the total number of events involved is moderate. Teenage pregnancy rates are not very high in Germany, not many people have children at a very young age before beginning or while still in vocational training. In other contexts, this may play a stronger part. Whether estimates of first birth rates for school degree categories will be biased when using imputed histories depends on whether people actually do drop out of vocational training or decide never to start vocational training if they become pregnant, leaving them with their school degree as their highest educational degree. In some countries, parenthood and vocational training may be compatible. This would lead to a lower bias when using imputed histories, even if pregnancy rates among those in vocational training or among those who have not yet begun vocational training are high. A similar situation may arise in contexts where it is common to drop out of school after becoming pregnant. In Germany, this does not seem to be a major issue, but in other countries it may lead to an overestimation of the risk of first birth for the ‘no degree’ category when using imputed histories.

4.4 Imputation without knowledge of the date the highest degree was obtained

The fifth and last model estimate shown in Table 2 uses education histories that were imputed without making use of any information on the date a respondent’s highest degree was obtained. Here, we simulated the case where the questionnaire asks for the respondent’s highest degree only, and does not additionally ask for the date this degree was obtained. Using this imputation, the estimates for the baseline diverge more strongly from the estimates using the original histories than is the case for the first four model estimates based on imputed histories that do make use of information on the date the highest degree was obtained (Table 2). For the last imputation, it was assumed that respondents obtained their degrees after standard durations of enrollment, as described in the methods section. Since many respondents are likely to have taken longer than a given standard duration of time to obtain their degree, or did not enroll in vocational or university education straight after school, or obtained more than one degree, the procedure used for the last imputation is very likely to extensively miscode exposure time that was actually spent in vocational or university education as exposure time spent holding a vocational or university degree. As we can see in Table A3a, this is indeed the case. Of the time originally spent enrolled in vocational training, 34% is imputed as time spent holding a vocational degree and not enrolled in education. Seen from a different perspective, of the imputed exposure time for the category ‘vocational

degree,' 13% stems from the 'vocational training' category in the original histories (Table A3b). This is substantially more than in either of the first two imputations, where 6% and less than 0.5% (rounded to 0%), respectively, of the imputed exposure time for the category 'vocational degree' stems from the 'vocational training' category in the original histories (Tables A1b and A2b). Since risks of first birth during vocational training are generally very low, a disproportionately small number of births is allocated to the 'vocational degree' category from the 'vocational training' category along with the additional exposure time (Table A3c). This results in an underestimation of risks of first birth for the 'vocational degree' category, which, as we have seen, is reflected in the low baseline estimates in the last model shown in Table 2.

Likewise, a substantial amount – 21% – of the exposure time allocated to the category 'university degree' in the last imputation was in reality spent in university education (Table A3b). Here, too, the corresponding proportion of events allocated to the 'university degree' category stemming from the 'in university education' category is much smaller (Table A3c). Thus, the risk of first birth for those with university degrees is underestimated as well when using data from the last imputation (Table 2).

Assuming that respondents took longer to obtain their highest degrees than the standard durations used for the last imputation might improve the estimates to some extent. Kravdal (2004) found that assuming more realistic, longer durations up until respondents obtained their degrees rather than normative durations for obtaining each type of degree gave improved, though still not satisfactory, estimates in the case of Norway. However, using longer durations of time could on the other hand have the effect of mis-imputing time that people actually spent holding a degree as time spent in education or vocational training. This would then likely result in an overestimation of first birth risks for the 'in vocational training' or 'in university education' categories in the imputed histories. If there is a lot of individual variation in study durations, any assumed standard duration is likely to lead to extensive mis-imputation.

In any case, it seems that knowledge of the date respondents obtained their highest degrees clearly improves imputations. The first four imputations all give estimates of first birth risks that are closer to the original estimates than the fifth imputation, for which no information on the date respondents obtained their highest degree is used. Thus, when constructing questionnaires, even if time constraints do not allow surveying complete education histories, it appears it would still be worthwhile to include a question on the date the highest degree was obtained.

5. Conclusion

This study has investigated whether imputed education histories can act as satisfactory substitutes for complete education histories when analyzing first birth transitions, in the case of western Germany. Our results do not indicate very extensive distortions of model estimates when using imputed histories. This appears to be related to the special nature of educational trajectories in Germany. Nonetheless, small to moderate distortions did occur, and examining how exposure time and events from the original histories were allocated in the imputed histories has enabled us to identify some general patterns of misallocations.

One finding is that imputations seem to become problematic whenever there are breaks between education spells during which people hold lower level degrees for some amount of time before reenrolling or continuing their studies to obtain a higher level degree. The greatest distortion of this nature that we found was for the category ‘lower secondary degree’ (‘Hauptschulabschluß’). The imputed histories gave upwardly biased results for this category. This is because childless women were more likely to continue their studies and attain a post-secondary degree. Any gap during which they held a lower secondary degree before starting their post-secondary education was misrepresented as a consequence of imputation.

A slight upward bias was also found for the category ‘vocational degree’ for one specific type of imputation. For that imputation, we simulated a case where the questionnaire asks only for the last date the respondent ever obtained a degree at their highest level. A distinction between the first and last date respondents received their highest degree can be relevant in the German context, because it is common to have more than one vocational degree at the same level. However, the bias did not turn out to be very large. It seems that not enough people obtained a second degree late enough in their life course to lead to a serious extent of misallocation of exposure time in the imputed histories.

Thus, gaps between education spells are one source of difficulties for imputation. A further general pattern of misallocation of exposure time in the imputed histories is related to cases where people drop out of education upon having a child. In these cases, births that in reality took place while enrolled in education will be shifted to a lower degree category in the imputed histories. This will also bias estimates of first birth risks for lower educational degrees upwards. For the cohort studied here, this was shown to occur for people enrolled in vocational training who had previously completed a lower secondary degree (‘Hauptschulabschluß’). This factor further contributed to the overestimation of first birth risks for the ‘lower secondary degree’ category when using the imputed histories.

The types of bias named above will only be relevant if motherhood actually does have a negative impact on progression to higher level degrees. In contexts where motherhood and participation in education are very well compatible and motherhood has no impact on subsequent educational attainment, then the types of bias described above will not occur.

We have also found some evidence that first birth risks will be underestimated for an education category if this category represents the highest level of education for many respondents, but respondents nonetheless tend to frequently continue their studies or reenroll in education after receiving this degree. As mentioned above, it is relatively common in Germany to obtain a second vocational degree at the same level as the first. If a questionnaire only asks for the first date the respondent ever obtained a degree at their highest level, this appears to lead to a slight underestimation of first birth risks for the category 'vocational degree' when imputations are based on this first date. This is because respondents who subsequently enroll in their second vocational training spell are mis-imputed as holding a vocational degree and not enrolled. Since risks of first birth are typically very low while enrolled in education, this leads to a downward bias for the 'vocational degree' category.

For further comparison, we also imputed education histories without using any information on the date the respondents obtained their highest degree. For this imputation, we used standard ages for the completion of respondents' highest degrees. Estimates based on this imputation gave more strongly biased results than the estimates based on imputations that did make use of the date the highest degree was obtained. Thus, it seems that when it is not possible to collect complete education histories in surveys, collecting at least the date respondents obtained their highest degree and not only the level can improve the quality of imputations.

Altogether, when we did make use of the date the highest degree was obtained, imputation of education histories did not cause much bias, with the exception of first birth risks for the category 'lower secondary degree' ('Hauptschulabschluß'). Estimates based on imputed histories may become more strongly biased for future cohorts than for the cohort of 1964 studied here if it becomes more common to obtain higher level degrees after having already completed a lower level degree some time ago. This might occur for Germany if more women obtain a master craftswoman's degree (a higher level vocational degree) sometime later in their life course. Also, if people begin to take breaks between completing a bachelor's degree and beginning their studies for a master's degree, this might also make imputations more problematic. International bachelor's and master's degrees have only very recently been introduced in Germany. However, imputations of education histories in countries where they are already well established may be more problematic. In countries where it is generally more common

to return to education later in the life course, imputation may be much more problematic than was shown to be the case for the western German cohort studied here.

6. Acknowledgements

I would like to thank Jan Hoem and Michaela Kreyenfeld for giving me the idea to study the quality of imputed education histories for fertility analysis when the education completion date is known, and for valuable comments and advice. I would also like to thank the two anonymous reviewers for very helpful comments and advice.

References

- Blossfeld, H.-P. and Huinink, J. (1991). Human Capital Investments or Norms of Role Transitions? How Women's Schooling and Career Affect the Process of Family Formation. *American Journal of Sociology* 97(1): 143-168. doi:10.1086/229743.
- Edwards, M.E. (2002). Education and Occupations. Reexamining the Conventional Wisdom About Later First Births Among American Mothers. *Sociological Forum* 17(3): 423-443. doi:10.1023/A:1019679023616.
- Hillmert, S., Künster, R., Spengemann, P., and Mayer, K.U. (2004). Projekt „Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland.“ Dokumentation. Berlin: Max Planck Institute for Human Development (Materialien aus der Bildungsforschung; 78).
- Hoem, J.M. and Funck Jensen, U. (1982). Multistate Life Table Methodology. A Probabilistic Critique. In: Land, K.C. and Rogers, A. (eds.). *Multidimensional Mathematical Demography*. New York: Academic Press.
- Hoem, J.M. and Kreyenfeld, M. (2006). Anticipatory analysis and its alternatives in life-course research. Part 1. Education and first childbearing. *Demographic Research* 15(16): 461-484. doi:10.4054/DemRes.2006.15.16.
- Hoem, J.M., Neyer, G., and Andersson, G. (2006). Education and childlessness. The relationship between educational field, educational level, and childlessness among Swedish women born in 1955-59. *Demographic Research* 14(15): 331-380. doi:10.4054/DemRes.2006.14.15.
- Jacob, M. (2005). Ausbildungen nach der Ausbildung – Zur Attraktivität des Dualen Systems in komplexen Ausbildungsverläufen. In: Gaubitsch, R. and Sturm, R. (eds.). *Beruf und Beruflichkeit*. Vienna: Arbeitsmarktservice Österreich (AMS report; 46). Communicatio – Kommunikations- und PublikationsgmbH: 24-41.
- Kerckhoff, A.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. doi:10.2307/2673250.
- Kravdal, Ø. (2004). An Illustration of the Problems Caused by Incomplete Education Histories in Fertility Analyses. *Demographic Research* SC 3(6): 135-154. doi:10.4054/DemRes.2004.S3.6.

- Kreyenfeld, M. (2008). Ökonomische Unsicherheit und der Aufschub der Familiengründung. In: Szydlik, M. (ed.). *Flexibilisierung. Folgen für Arbeit und Familie*. Wiesbaden: VS Verlag für Sozialwissenschaften: 232-254.
- Lauterbach, W. and Weil, M. (2008). Mehrfachausbildungen und die Folgen für die Erwerbstätigkeit. Oder. Wer ist am erfolgreichsten? In: Szydlik, M. (ed.). *Flexibilisierung. Folgen für Arbeit und Familie*. Wiesbaden. VS Verlag für Sozialwissenschaften: 68-92.
- Rindfuss, R.R., Morgan, S.P., and Swicegood, C.G. (1988). *First Births in America. Changes in the Timing of Parenthood*. Berkeley: University of California Press.
- Schütze, H.G. and Slowey, M. (2002). Participation and exclusion. A comparative analysis of non-traditional students and life-long learners in higher education. *Higher Education* 44(3-4): 309-327. doi:10.1023/A:1019898114335.
- Vikat, A. (2004). Women's Labor Force Attachment and Childbearing in Finland. *Demographic Research* SC 3(8): 177-212. doi:10.4054/DemRes.2004.S3.8.

Appendix

Table A1a: Cross-tabulation of exposure time in original by imputed education variable. Imputation assumption: questionnaire asked for the *first time* respondent attained highest degree, and used a simple categorization of educational degrees. Row percentages

original	imputed									total original exp. time by education	
	no degree	low. sec. degree	adv. lower sec. degr.	upper sec. degree	vocational degree	univ./ college degree	in school	in voc. training	In university		missing
no degree	84%						1%	15%			1212
lower sec. degree		75%					3%	22%			2440
adv. lower sec. degr.			49%				8%	37%	4%	2%	1301
upper sec. degree				33%			1%	27%	39%	0%	1054
vocational degree					99%		0%		1%		43235
univ./ college degr.						100%					4310
in school	0%	0%					65%	33%	1%	0%	17922
in voc. training		0%	1%	0%	13%	3%	8%	71%	3%	0%	19157
in university				4%	5%	2%	2%	7%	80%	0%	6347
missing					25%	9%	3%	8%	0%	55%	990
											97968

Table A1b: Cross-tabulation of exposure time in original by imputed education variable. Imputation assumption: questionnaire asked for the *first time* respondent attained highest degree, and used a simple categorization of educational degrees. Column percentages

original	imputed										
	no degree	low. sec. degree	adv. lower sec. degr.	upper sec. degree	vocational degree	univ./ college degree	in school	in voc. training	In university	missing	
no degree	94%						0%	1%			
lower sec. degree		95%					1%	2%			
adv. lower sec. degr.			84%				1%	2%	1%	4%	
upper sec. degree				59%			0%	1%	6%	0%	
vocational degree					93%		0%		4%		
univ./ college degr.						85%					
in school	6%	1%					85%	27%	4%	9%	
in voc. training		4%	16%	1%	6%	11%	12%	63%	9%	9%	
in university				40%	1%	3%	1%	2%	77%	4%	
missing					1%	2%	0%	0%	0%	74%	
total imputed exp. time by education	1084	1931	759	584	46042	5088	13624	21512	6604	740	97968

Table A1c: Cross-tabulation of events in original by imputed education variable. Imputation assumption: questionnaire asked for the *first time* respondent attained highest degree, and used a simple categorization of educational degrees

original	imputed									
	no degree	low. sec. degree	adv. lower sec. degr.	upper sec. degree	vocational degree	univ./ college degree	in school	in voc. training	In university	missing
no degree	5									
lower sec. degree		21						1		
adv. lower sec. degr.			8				1	2		
upper sec. degree				3				1	1	
vocational degree					359				1	
univ./ college degr.						32				
in school		1					3	2		
in voc. training		2			4	3		9		
in university				1					6	
missing					3			1		1

Table A2a: Cross-tabulation of exposure time in original by imputed education variable. Imputation assumption: questionnaire asked for the *last time* respondent attained highest degree, and used a simple categorization of educational degrees. Row percentages

original	imputed										total original exp. time by education
	no degree	low. sec. degree	adv. lower sec. degr.	upper sec. degree	vocational degree	univ./ college degree	in school	in voc. training	In university	missing	
no degree	84%						1%	15%			1212
lower sec. degree		75%					3%	22%			2440
adv. lower sec. degr.			49%				8%	37%	4%	2%	1301
upper sec. degree				33%			1%	27%	39%	0%	1054
vocational degree					90%		0%	9%	1%		43235
univ./ college degr.						92%			8%		4310
in school	0%	0%					65%	33%	1%	0%	17922
in voc. training		0%	1%	0%	1%	1%	8%	84%	5%	0%	19157
in university				4%	4%	1%	2%	8%	81%	0%	6347
missing					14%	9%	3%	19%	0%	55%	990
											97968

Table A2b: Cross-tabulation of exposure time in original by imputed education variable. Imputation assumption: questionnaire asked for the *last time* respondent attained highest degree, and used a simple categorization of educational degrees. Column percentages

original	imputed										
	no degree	low. sec. degree	adv. lower sec. degr.	upper sec. degree	vocational degree	univ./ college degree	in school	in voc. training	In university	missing	
no degree	94%						0%	1%			
lower sec. degree		95%					1%	2%			
adv. lower sec. degr.			84%				1%	2%	1%	4%	
upper sec. degree				59%			0%	1%	6%	0%	
vocational degree					99%		0%	15%	4%		
univ./ college degr.						93%			5%		
in school	6%	1%					85%	21%	3%	9%	
in voc. training		4%	16%	1%	0%	4%	12%	57%	13%	9%	
in university				40%	1%	1%	1%	2%	69%	4%	
missing					0%	2%	0%	1%	0%	74%	
total imputed exp. time by education	1084	1931	759	584	39336	4265	13639	28203	7427	740	97968

Table A2c: Cross-tabulation of events in original by imputed education variable. Imputation assumption: questionnaire asked for the *last time* respondent attained highest degree, and used a simple categorization of educational degrees

original	imputed									
	no degree	low. sec. degree	adv. lower sec. degr.	upper sec. degree	vocational degree	univ./ college degree	in school	in voc. training	In university	missing
no degree	5									
lower sec. degree		21						1		
adv. lower sec. degr.			8				1	2		
upper sec. degree				3				1	1	
vocational degree					347			12	1	
univ./ college degr.						31			1	
in school		1					3	2		
in voc. training		2			1			12	3	
in university				1					6	
missing					3			1		1

Table A3a: Cross-tabulation of exposure time in original by imputed education variable. Imputation assumption: no information given on date highest degree was attained. Row percentages

original	imputed									total original exp. time by education	
	no degree	low. sec. degree	adv. lower sec. degr.	upper sec. degree	vocational degree	univ./ college degree	in school	in voc. training	In university		missing
no degree	84%				11%		1%	4%			1212
lower sec. degree		79%			3%		3%	14%	1%		2440
adv. lower sec. degr.			50%	3%	10%		6%	25%	4%	2%	1301
upper sec. degree				34%	25%	7%	1%	2%	31%	0%	1054
vocational degree					94%	0%	0%	5%	1%		43235
univ./ college degr.						89%			11%		4310
in school	0%	1%	0%	0%	2%		63%	31%	2%	0%	17922
in voc. training		1%	1%	0%	34%	3%	8%	50%	3%	0%	19157
in university				5%	11%	19%	2%	1%	62%	0%	6347
missing					23%	9%	3%	10%	0%	55%	990
											97968

Table A3b: Cross-tabulation of exposure time in original by imputed education variable. Imputation assumption: *no information given on date* highest degree was attained. Column percentages

original	imputed										
	no degree	low. sec. degree	adv. lower sec. degr.	upper sec. degree	vocational degree	univ./ college degree	in school	in voc. training	In university	missing	
no degree	94%				0%		0%	0%			
lower sec. degree		91%			0%		1%	2%	0%		
adv. lower sec. degr.			81%	5%	0%		1%	2%	1%	4%	
upper sec. degree				52%	1%	1%	0%	0%	5%	0%	
vocational degree					83%	0%	0%	12%	4%		
univ./ college degr.						67%			8%		
in school	6%	4%	0%	0%	1%		85%	31%	5%	9%	
in voc. training		5%	18%	1%	13%	9%	12%	52%	11%	9%	
in university				42%	1%	21%	1%	0%	65%	4%	
missing					0%	1%	0%	1%	0%	74%	
total imputed exp. time by education	1084	2126	799	700	49259	5696	13291	18266	6007	740	97968

Table A3c: Cross-tabulation of events in original by imputed education variable. Imputation assumption: no information given on date highest degree was attained

original	imputed									
	no degree	low. sec. degree	adv. lower sec. degr.	upper sec. degree	vocational degree	univ./ college degree	in school	in voc. training	In university	missing
no degree	5									
lower sec. degree		22								
adv. lower sec. degr.			8		1			1	1	
upper sec. degree				3	1	1				
vocational degree					354			5	1	
univ./ college degr.						32				
in school		1			1		2	2		
in voc. training		2			8	3	1	4		
in university				2		3			2	
missing					3			1		1

