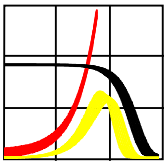

Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Doberaner Strasse 114 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
http://www.demogr.mpg.de

MPIDR WORKING PAPER WP 2001-031
OCTOBER 2001

Small effects of selective migration and selective survival in retrospective studies of fertility

Gunnar Andersson (andersson@demogr.mpg.de)
Boris Sobolev (bs9@post.queensu.ca)

This working paper has been approved for release by: Jan M. Hoem (hoem@demogr.mpg.de)
Head of the Laboratory of Contemporary European Fertility and Family Dynamics.

Max Planck Institute for Demographic Research,
Doberaner Straße 114,
D-180 57 Rostock, Germany.
E-mail to authors: andersson@demogr.mpg.de, bs9@post.queensu.ca.

GA/BS, 6 October 2001

Small effects of selective migration and selective survival in retrospective studies of fertility

by

Gunnar Andersson* and Boris Sobolev**

Abstract. In this paper, we assess the accuracy of fertility estimates that are based on the retrospective information that can be derived from an existing cross-sectional population. Swedish population registers contain the information on childbearing of all people ever living in Sweden and thus allow us to avoid any problems of selectivity by virtue of survival or of out-migration when we estimate fertility measures for previous calendar periods. We calculate two types of fertility rates for each year in 1961-1999: (i) rates that are based on the population that were living in Sweden at the end of 1999 and (ii) rates that also include the information on people who had died or emigrated before the turn of the century. We find that the omission of information on emigrated and deceased individuals, as the situation would be in any demographic survey, most often have negligible effects on our fertility measures. However, first-birth rates of immigrants gradually become more biased as we move back in time from 1999 so that they increasingly tend to over-estimate the actual fertility of that population.

* Max Planck Institute for Demographic Research, Rostock, Germany
** Queen's University, Kingston, Canada

1. Introduction

The purpose of this paper is to provide an examination of the magnitude of the bias that may appear in any fertility estimates that are based on the retrospective information on childbearing gathered at a fixed point in time. Many studies of human fertility are based on survey data that typically are collected by asking respondents about their previous histories of childbearing and of other related behaviors. Generally, such information is considered reliable since the birth of a child is such an important event in people's lives that respondents at least will report it accurately. Normally, researchers only raise some doubts about the accuracy of men's reports on childbearing since they sometimes are found to underreport the existence of children who are fathered outside any stable union of theirs (Rendall et al., 1999; Greene and Biddlecom, 2000). However, even if we restrict ourselves to the very reliable histories of childbearing as reported by women, we may be faced with some problems if we try to estimate measures of fertility of the population of a certain area for periods preceding the survey date.

A bias in estimates may arise if the cross-sectional population of that area has had a different fertility behavior than people who previously lived there but had left it at the time of data collection. The literature on the fertility of migrants, for example, suggests that long-distance migrants often tend to display a pattern of relatively low fertility before a migration and instead an elevated fertility shortly after it (Goldstein and Goldstein, 1981; Ford, 1990; Alders, 2000; Andersson, 2001). Such a pattern arises if childless people are more prone to migrate than parents are and if family formation and childbearing typically occur after a long-distance migration. Since previous out-migrants from an area not show up in a survey that is based on the cross-sectional population of that area, their segments of potential low-fertility behavior will be absent in the survey data while instead the high-fertility behavior of newly arrived in-migrants in the area is covered properly. If there are similar selection effects in reproductive histories by the virtue of survival of women we might as well be faced with a bias arising from the omission of individuals who have died before the data were collected. Doblhammer (2000) shows that childless women have a slightly higher mortality at ages above 50 than mothers have, which suggests that such selection effects indeed might appear. Again, the omission of data on deceased individuals from any sample then results in an over-estimation of the previous fertility

level in the area since the persons who are left out are suspected to have had a somewhat lower fertility than the surviving population. Nevertheless, any effects of that kind must be very small since the relationship between reproductive behavior and mortality is quite weak.

Normally, it is very difficult to grasp the existence of any selection effects of the kind we discussed above. In the present examination, however, we are indeed able to provide evidence of the existence and magnitude of such effects by using a data set that contains information on the childbearing histories of an existing cross-sectional population and, in addition, the corresponding information on people who previously had lived in the area under investigation but have died or out-migrated. For this purpose, we use population-register data of Sweden, which cover the childbearing, mortality, and migration of all women who have ever lived in that country in 1961-1999. Since data on persons who no longer live in Sweden are saved in the register records, we are able to perform a calculation of fertility measures over the period 1961-1999, as they would have appeared in a prospective study on fertility starting in 1961. As an experiment, we also choose to exclude all information that refer to people who no longer lived in Sweden at the end of 1999, as the situation would have been if we had conducted a retrospective survey at that time. By comparing fertility estimates that are based on (i) the prospective study design and (ii) the retrospective design, we are perfectly able to examine if the latter type of study produces fertility measures that are different from those stemming from the complete information of the prospective study. If any bias appears, we expect it to become more important as we move back in time from our simulated survey date of the last day of December 1999 and we report the relative magnitude of any such bias.

2. Data and methods

Our data stem from the Swedish population-registration system, which with the help of a unique identifying code of each individual ever living in Sweden and an efficient coverage of all vital events occurring in that country provides very reliable information on the demographic histories of people there. Our extract of data contains information on childbearing, mortality, and migration of all women born in Sweden in 1925 and later (who were either registered in the census of 1960 or born after that

census) as well as the corresponding information on women born abroad in 1925 and later who have ever lived in Sweden in 1961-1999[1]. The data cover their full childbearing histories until a death, an emigration, or 31 December 1999, which ever comes first.

Our study population is presented in Table 1, which gives the total number of women by three very broad country-groups of origin. The vast majority of women are of course born in Sweden but the data also contain large number of immigrants - of whom 40 percent stem from the neighboring Nordic countries. The two mid columns of Table 1 report the number of women in our study population who emigrated from Sweden or died in 1961-1999 so that they no longer lived there at the end of 1999. The immigrant population is relatively young so the exclusion of immigrants by the cause of mortality is fairly unimportant. Instead, we note that large numbers of immigrants have again emigrated from Sweden. Almost a third of immigrants from the neighboring Nordic countries and more than a fifth of immigrants from non-Nordic countries had left Sweden at the end of 1999. This is not particularly remarkable since return migration is a typical feature of all types of migratory streams. Nevertheless, it points to the need of having access to longitudinal information on immigrants as well as emigrants if one wants to have a full picture of the demographic behavior of any mobile population in a country.

Table 1: Number of women ever living in Sweden in 1961-1999 in our data, by country group of birth, and the number who no longer lived in Sweden by the end of 1999 because of own out-migration or death.

Country of birth	study population	died before year 2000	emigrated before 2000	population in Dec. 1999	percent left
Sweden	2,973,000	117,000	69,000	2,787,000	94 %
Other Nordic	197,000	8,000	59,000	129,000	66 %
Non-Nordic	313,000	6,000	64,000	242,000	77 %

We use our data in order to calculate relative risks of childbearing by calendar year in 1961-1999 for women at different parities. In our event-history models, we

[1] The data on immigrants only cover women who migrated to Sweden while being younger than 35 years old. This age limit reduces the problem of possible omission of children who never joined their mother to live in Sweden from the birth records of immigrant women while we anyway keep the vast majority of these women in our study population since most of them arrived at young ages.

control for the effect of the age of a woman and the age of any youngest child of hers. We estimate separate models of first-birth risks for women at ages 16-26 years and women at ages 30-45 years since we know that trends in childbearing have been quite different for childless women at the younger and the older age brackets (Andersson, 1999). We present separate sets of parity-specific fertility measures for women born in Sweden and for women born abroad.

We calculate our fertility rates in two rounds. First, we use the full information of all women available in our data for our calculations. Secondly, we exclude women who had died or emigrated from Sweden before the turn of the century. This gives us a data set with information on the childbearing of the cross-sectional population of 31 December 1999 - like the one we would get if we had conducted a survey at that time. We use this latter data with retrospective-type information only in order to calculate the same sets of fertility rates by calendar year as we produce in our first round of calculation. Finally, we relate the fertility measures of the second round to those of the first in order to see whether we can find any systematic deviation in risk patterns. We report the relative deviation in fertility rates at various time horizons from our simulated interview date in order to see how far back in time one typically can rely on retrospectively reported data without facing any serious problems of bias in fertility estimates of different groups of women. We use the Genmod module of SAS in order to calculate our fertility measures. For a further description of our data and the type of models we estimate, see Anderson (1999).

3. Results of our experiment

As an introduction, we present the relative risks of childbearing by calendar year for childless younger women, childless women at ages 30-45, one-child mothers, and two-child mothers, all born in Sweden, with a separate curve for each category of women in Figure 1. The risks are based on the full information on childbearing that is available in our register data. Our fertility measures are given on a relative scale for each group of women separately so we get a good picture of changes over time in the propensity to give birth but get no information on differences in fertility levels between the different categories of women. Evidently, fertility in Sweden has fluctuated considerably during our study period and important turning points appear

in 1964, 1977, 1984, 1990, and 1997. We do not intend to discuss the background to these developments in this presentation but refer instead to Hoem and Hoem (1996) and Andersson (1999) for a more detailed discussion of patterns in childbearing in Sweden during our study period. Trends in childbearing of foreign-born women in Sweden very much resemble those of the Swedish-born population even though their fertility in many cases stands at a higher level than that of the Swedish-born (not shown here). For a description of patterns in childbearing of the immigrant population in Sweden, see Andersson (2001).

In Table 2, we present the main results of our investigation, which amounts to the comparison of childbearing risks of Swedish-born women as being calculated from our two designs of data. We report the relative deviation in estimated risks for the "retrospective" study as compared to those of the "prospective" study for different calendar years prior to 1999, i.e., at different time horizons from our simulated data collection. Separate columns give the results for the various parity and age groups we examine.

Table 2: Relative bias in retrospectively collected fertility data by time from data collection and parity (and age) group of women (in percent). Comparison of childbearing rates of women born in Sweden: rates from retrospectively collected data related to rates from full data.

years from study		*first births, 16-26 years*	*first births, 30-45 years*	*second births*	*third births*
1	(1998)	0	0	0	0
2	(1997)	1	0	0	0
3	(1996)	1	1	0	0
5	(1994)	2	1	0	0
7	(1992)	2	1	0	0
10	(1989)	2	1	1	0
15	(1984)	2	2	1	0
20	(1979)	2	3	1	0
25	(1974)	2	5	1	0
30	(1969)	3	6	2	1
35	(1964)	3	7	2	0

Evidently, a retrospective gathering of data results in a minor overestimation of fertility measures as we move back in time from the year when the data were collected. However, in most cases these effects are of no importance at all. For younger childless women we only get a bias of around 2 percent when we move some

five years back in time and we do not get a bias higher than 3 percent even if we move several decades back in time. When we estimate fertility measures for mothers, we find that the bias from any selection due to survival or emigration is virtually non-existent. The only case where a bias really appears is when we estimate fertility rates for childless women at ages above 30, but this bias only turns out to be visible if we move some 20-25 years back in time. In order to check whether the bias in first-birth rates of older childless women arises from selective mortality or from selective migration, we re-estimate our "retrospective" models leaving out only one group of absent (deceased or emigrant) individuals at a time while keeping the information on the others in our data (not shown here). Such an exercise reveals that the bias in fertility estimates of older women almost entirely stems from differential mortality by the motherhood status of the elderly.

In conclusion, the general picture from our experiment is that the effects of selectivity by virtue of survival or of emigration is quite unimportant when we estimate fertility measures from retrospectively collected data for a local population. The only bias we found appeared when we estimated fertility measures for women who were childless in their 30s or 40s some 20 years back in time. In this case, differential mortality by changes in motherhood status at the older ages caused a bias in our fertility estimates. However this category of women is seldom the target of many conventional fertility studies so our finding should not cause too much of a worry for researchers who work with retrospective data.

In Table 3, we proceed and present the results of the corresponding examination of data for foreign-born women in Sweden, with results given for fertility estimates of immigrant women from the non-Nordic countries. As this is a much more mobile group of people than the native population is, we might perhaps expect more effects of a selection into the remaining cross-sectional population of December 1999, and this is indeed what we find. Retrospective first-birth rates of foreign-born women increasingly tend to overestimate the childbearing of the immigrant population in Sweden as we move back in time from 1999. The effects become visible already a few years prior to the date of our data collection and our fertility measures overestimate the true first-birth fertility by some 10 percent at 15-20 years prior to the simulated survey date. By contrast, if we only study the childbearing behavior of immigrant mothers, we find that the retrospectively collected data cover the childbearing dynamics very well. We assume that the bias we find for the childless women

mainly is due to differential emigration by motherhood status and we confirm this hypothesis by estimating models where we leave out only the emigrated women from our data while keeping the deceased ones (not shown here).

Table 3: Relative bias in retrospectively collected fertility data by time from data collection and parity (and age) group of women (in percent). Comparison of childbearing rates of foreign-born women from non-Nordic countries: rates from retrospectively collected data related to rates from full data.

years from study		*first births, 16-26*	*first births, 30-45*	*second births*	*third births*
1	(1998)	1	1	0	1
2	(1997)	2	2	1	0
3	(1996)	3	4	1	1
5	(1994)	4	9	1	1
7	(1992)	6	8	2	2
10	(1989)	7	4	2	2
15	(1984)	9	8	1	0
20	(1979)	10	10	0	4
25	(1974)	11	10	-1	1

In sum, the results of the second part of our experiment were a bit more discouraging than those of our examination of fertility estimates for the native population were. Evidently, the propensity of childless immigrants to re-emigrate is higher than the corresponding propensity of immigrant mothers and this selectivity in migration behavior causes a bias in any first-birth estimates that only are based on the remaining immigrant population of an area. The omission of substantial numbers of childless emigrants from our data results in an overestimation of the fertility of the immigrant population in Sweden. However, if we avoid stretching our fertility analyses too far back in time we can also avoid any unacceptable overestimation of the fertility of the immigrant population.

4. Conclusions

In our examination, we have managed to get a clear picture of the reliability of fertility measures that are based on retrospectively collected data when it comes to their ability to describe the childbearing of a population of a given geographical area in calendar periods prior to the data collection. We used population-register data of

Sweden in order to simulate a collection of data at a given point in time. We compared the fertility estimates from such a retrospective data collection to fertility rates that also pick up the childbearing behavior of people who have left the area under investigation before the time of data collection. Our results are rather encouraging in that they demonstrate that the omission of emigrated or deceased individuals from the data rarely results in more than a minor overestimation of fertility rates in periods before the data collection. However the reliability of retrospectively collected data mainly holds when we describe the behavior of a population with moderate or low levels of out-migration. If we focus on the mobile immigrant population, we actually face some problems of selectivity in the data that only contain information on the immigrants who did not again leave Sweden. Most immigrant populations display relatively high levels of return migration, so any demographic estimate of such a population easily risks to be affected by various types of selective out-migration. To minimize such problems, we recommend that retrospectively collected data on the childbearing behavior of immigrants mainly should be analyzed for relatively short time horizons before the data collection.

Acknowledgments

We want to express our gratitude to the late Britta Hoem for carefully having organized the raw data that we have used in our study. We thank Statistics Sweden for allowing us to use the data. In addition, we are grateful for comments received at a seminar at the Max Planck Institute of Demographic Research in Rostock.

References

Alders, M., 2000. Cohort fertility of migrant women in the Netherlands. Paper at the BSPS-NVD-URU Conference, 31 August – 1 September 2000, Utrecht, The Netherlands.

Andersson, G., 1999. Childbearing trends in Sweden 1961-1997. *European Journal of Population* 15: 1-24.

Andersson, G., 2001. Childbearing patterns of foreign-born women in Sweden. *MPIDR Working Paper*, WP 2001-011. Max Planck Institute for Demographic Research, Rostock.

Doblhammer, G., 2000. Reproductive history and mortality later in life: A comparative study of England & Wales and Austria. *Population Studies* 54: 169-176.

Ford, K., 1990. Duration of residence in the United States and the fertility of US immigrants. *International Migration Review* 24: 34-68.

Goldstein, S., and Goldstein, A., 1981. The impact of migration on fertility: an 'own children' analysis for Thailand. *Population Studies* 35: 265-284.

Greene, M., and Biddlecom, A., 2000. Absent and problematic men: Demographic accounts of male reproductive roles. *Population and Development Review* 26: 81-115.

Hoem, B., and Hoem, J., 1996. Sweden's family policies and roller-coaster fertility. *Jinko Mondai Kenkyu* (Journal of Population Problems) 52: 1-22.

Rendall, M., Clarke, L., Peters, E., Ranjit, N., and Verropoulou, G., 1999. Incomplete reporting of men's fertility in the United States and Britain: A research note. *Demography* 36: 135-144.

Figure 1: Relative risk of childbearing by calendar year, Swedish-born women in Sweden 1961-1999,
standardized for age of woman and of any youngest child.