Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1 · 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 TECHNICAL REPORT 2012-004
AUGUST 2012

Episode Data from the Russian Education and Employment Survey

Michaela Kreyenfeld (kreyenfeld@demogr.mpg.de)
Dirk Konietzka

For additional material see www.demogr.mpg.de/tr/

Episode Data from the Russian Education and Employment Survey

Michaela Kreyenfeld und Dirk Konietzka

Abstract: This paper documents how to generate consistent education, employment, fertility and residential careers from the Russian Employment and Education Survey (EES). We describe how we imputed missing or inaccurate data. Furthermore, we estimate a first birth model to demonstrate how this data could be used for event history modelling. A sensitivity analysis shows that the results are very stable in respect to the imputation of missing dates. Data is organized in spell or episode format. For manipulating the data, we used the software STATA 10.0.

Keywords: Life course data, education, employment, fertility

1 Introduction

This paper is a documentation of how we generated employment, education and residential careers from the Russian Education and Employment Survey (EES). The EES is a subsample of the Russian Generations and Gender Survey (GGS).[1] It provides longitudinal information on respondents' employment, education, fertility and residential histories. It is a retrospective survey and it includes biographical information since January of the year the respondents turned age 17 (for details, see Independent Institute for Social Policy 2005). The virtue of the data set is the accuracy with which it surveys employment and educational careers. For example, it allows for multiple activities, such as second jobs.

In this paper, we provide an overview of how the raw EES-data can be transferred into episode format so that they can be used for event history modelling. The paper is organized as follows. In the following section (part 2), we show how we generated the careers and how we imputed missing or implausible data. Part 3 describes how we joined the various careers into a single data set. Part 4 provides an example of how to use this data for event history modelling. In that part, we also conduct a sensitivity analysis. The data is assembled in spell or episode format. For data manipulation, we used STATA 10.0. The command syntax is available with the Technical Report. The code may be used by other users by citing this document. However, we hold no responsibility for errors that may have arisen during the coding procedures.

2 Generating Educational, Employment and Fertility Careers

In this section, we explain the principle of how we manipulated the raw data. We draw on the educational career to explain the data set-up. For generating the other careers, we proceeded in a similar fashion. We generated each career separately and finally integrated them in a single data set. We opted for this step-by-step procedure,

[1] The data may be merged by a personal identifier with the GGS-data. Due to data protection reasons, the personal identifier is not available in the scientific use file of the EES data. However, a data file that contains EES and GGS data is available upon request.

because it makes it easier to detect errors and miscoding in the data. Altogether, we generated the following seven careers (or biographies) from the raw data:

(1) The educational biography

(2) The main job biography

(3) The second job biography

(4) The other activity biography

(5) The unemployment benefit biography

(6) The residential biography

(7) The fertility biography

2.1 The main data set-up

Table 1 provides a data extract from the EES data. It shows the main educational career of the first respondent in the data set (with the personal identification number bid 101009). On the basis of this respondent, we show how we manipulated the raw data. Here, we selected only the variables related to the respondent's education to simplify matters. As the table depicts, the respondent started his or her first education in January 1971. This first activity ends in June 1972. There is one more educational episode (from September 1972 until June 1976). The remaining rows in this data extract provide only missing information since they refer to other activities, such as employment.

Table 1: Educational career in the raw data

bid	o9s_m	o9s_y	o9k_m	o9k_y	o11	bh4y
101009	january	1971			secondary	1954
101009						1954
101009			june	1972		1954
101009						1954
101009	september	1972			Higher	1954
101009			june	1976		1954
101009						1954
101009						1954
101009						1954
101009						1954
101009						1954
101009						1954
101009						1954
101009						1954

Notes: bid: Personal identification number; o9s_m: Month start of education; o9s_y: Year start of education; o9k_m: Month end of education; o9s_y: Year end of education; o11: Type of education; bh4y: Year of birth

In a first step, we selected all rows which contain information on the start and ending of an education spell. Furthermore, we joined the beginning and the end of an educational episode into one row.

Table 2: Educational career after joining beginning and end of episode

bid	o9s_m	o9s_y	o9k_m	o9k_y	o11	bh4y
101009	January	1971	June	1972	Secondary	1954
101009	September	1972	June	1976	Higher	1954

Notes: bid: Personal identification number; o9s_m: Month start of education; o9s_y: Year start of education; o9k_m: Month end of education; o9s_y: Year end of education; o11: Type of education; bh4y: Year of birth

In the next step, we generated two new variables. BEGIN denotes the beginning of an episode, measured in months since January of the year the respondent turned age 17. END denotes the end of an episode, also measured in months since January of the year the respondent turned age 17 (see Table 3).

Table 3: Educational career after using respondent's age as process time

bid	BEGIN	END	o11
101009	1	18	Secondary
101009	21	66	Higher

Notes: bid: Personal identification number; BEGIN: Start of the episode in person months since January of the year the respondent turned age 17; END: End of the episode in person months since January of the year the respondent turned age 17; o11: Type of education

In the next step, we subtracted one month from BEGIN (see Table 4). This procedure is consistent with the idea that respondents mean the period January 1st until June 30th when they report that they were employed from January until June. It also fits the notion that a new job or a new education is usually started at the beginning of a month. However, for some cases ending month and starting date of the subsequent episode were the same. For such cases, we assumed that a change of activity occurred in the middle of the month.

Table 4: Educational career after subtracting one month from beginning of episode

Bid	BEGIN	END	o11
101009	0	18	Secondary
101009	20	66	Higher

Notes: bid: Personal identification number; BEGIN: Start of the episode in person months since January of the year the respondent turned age 17; END: End of the episode in person months since January of the year the respondent turned age 17; o11: Type of education

In a final step, we filled the gaps within the educational career (see Table 5). For filling the gaps, we used the Stata-Ado-File "Spellsort" (Walke and Kreyenfeld 2006).

Table 5: Educational career after filling the gaps

Bid	BEGIN	END	o11
101009	0	18	Secondary
101009	18	20	-1
101009	20	66	higher:

Notes: bid: Personal identification number; BEGIN: Start of the episode in person months since January of the year the respondent turned age 17; END: End of the episode in person months since January of the year the respondent turned age 17; o11: Type of education

There are some respondents who did not report any educational episode. All educational variables were set to missing (-1) for these respondents. Unfortunately, it is not possible to distinguish between different types of missing information. We are therefore not able to differentiate respondents who *never received* any education from respondents who *failed to report* their educational career (non-response). This applies to all other careers as well.

For some episodes, only starting and ending times, but no further information was given. In order to identify episodes from the educational career in the final data set, we generated a variable called EDU which is equal to 1 for all educational episodes (and -1 otherwise). For the other careers, we generated similar variables. An exception is the fertility career. Only the dates of childbirth, but no further information on children is provided. For respondents who did not report a date of childbirth, we had to assume that these respondents are childless. As mentioned before, more accurate fertility histories may be derived from the Generations and Gender Survey which may be linked to the EES data.

2.2 How cases with missing or inaccurate information on month of occurrence of event were treated

There are some cases, where respondents did not report the exact date of occurrence of an event, but only the quarter or the half of the year. This issue is exemplified in Table 6 which displays the frequency count of the month of the beginning of the educational episode. We recoded the inaccurate monthly information based on a random procedure. For example, respondents who report that they started a new educational spell in the first quarter of the year were assigned a random month variable that has the values one, two or three. Respondents who report that they started a new educational spell in the second quarter of the year were assigned a random variable that has the values four, five or six.

This procedure can produce overlaps when ending times are larger than subsequent starting times. We adjusted the ending times in all cases where starting and ending times did not join. Episodes in the educational career which have been recoded have been flagged out by a variable called IMP_EDU. IMP_EDU equals 1 for “beginning recoded”, 2 for “end recoded”, 3 for “beginning and end recoded” and 4 for “end adjusted”.

Table 6: Distribution of the month of beginning of educational episode

Month	Frequency	Percent
january	5,953	50.56
february	124	1.05
march	108	0.92
april	77	0.65
may	102	0.87
june	284	2.41
july	231	1.96
august	512	4.35
september	3,837	32.59
october	261	2.22
november	130	1.1
december	120	1.02
1st quarter	6	0.05
2nd quarter	5	0.04
3rd quarter	19	0.16
4th quarter	6	0.05
Total	11,775	100

2.3 Miscellaneous

2.3.1 Unclear starting or ending dates

The data structure of the EES is designed in a way that the starting date is given in the first row, the ending date in the subsequent row (see Table 1). There are some cases that violate this order. This is exemplified for a respondent in Table 7. This person entered his/her second educational period in his/her life in September 1999. Before the end of this episode is reported, a new educational episode is given. There is also a case where the first row is an ending date. Furthermore, there are cases where the last entry is a starting date. We recoded the data and flagged out cases by a variable called ODD_EDU.

Table 7: Case with “odd” educational career

bid	SPELL	o9s_m	o9s_y	o9k_m	o9k_y	o11
925155	1	January	1999			Secondary
925155	2			June	1999	
925155	3	September	1999			Higher
925155	4	September	2001			Higher
925155	5			3rd quarter	2001	
925155	6			December	2003	

Notes: bid: Personal identification number; SPELL: Row number; o9s_m: Month start of education; o9s_y: Year start of education; o9k_m: Month end of education; o9s_y: Year end of education; bh4y: Year of birth; o11: Type of education

2.3.2 Peculiarities in the educational career

There are 10 such cases (as described in Section 2.3.1) in the education career (out of 6.455 cases in the whole data set). Apart from this, we did not notice any peculiarities in the educational career.

2.3.3 Peculiarities in the main employment career

There were 60 such odd cases in the main employment career. Apart from cases as described in section 2.3.1, we did not notice any peculiarities in the main employment career.

2.3.4 Peculiarities in the second job career

There were 7 such odd cases in the second job career. Apart from cases as described in section 2.3.1, we did not notice any peculiarities in the second job career.

2.3.5 Peculiarities in the other activity career

In the other activity career, there are 64 cases as described in Section 2.3.1. There was one more case where the starting time was larger than the ending time (see Table 8). We deleted the erroneous episode and flagged out the case by the variable ODD_OTH.

In general, we transferred all variables that were available into episode format. There is/was only one exception: we omitted the (string) variable "r32s" which refers to the other activity career from the data set.

Table 8: Case with peculiar other activity career

bid	r31s_m	r31s_y	r31k_m	r31k_y
4507010	November	1983		
4507010			May	1985
4507010	July	2005		
4507010			April	2005

Notes: bid: personal identification number; z31s_m: month start of other activity; z31s_y: year start of other activity; z31k_m: month end of other activity; z31k_y: year end of other activity

2.3.6 Peculiarities in the unemployment benefit career

Periods in which respondents received unemployment benefits have been organized in a slightly different way than the other careers. These periods were surveyed within the "other activity career". For unemployment periods, the starting and ending dates are stored in the same row (see Table 9). We did not observe any peculiar cases in the unemployment career.

Table 9: Selected cases in the unemployment career

bid	r31sp_m	r31sp_y	r31kp_m	r31kp_y	r32s_c
102002	november	2004	february	2005	br (unem
102008	2nd quar	2003	april	2004	br (unem

Notes: bid: personal identification number; z31sp_m: month start of unemployment; z31sp_y: year start of unemployment; z31kp_m: month end of unemployment; z31kp_y: year end of unemployment; r32s_c: type of occupation

2.3.7 Peculiarities in the residential career

In the residential careers, only starting dates were reported. We assumed that a respondent resided in a certain district until he/she reports the next starting date for another district. Similar to the other careers, we assumed that the change occurred at the beginning of the month. Since only starting dates were reported, the last episode was open. For this reason, we generated a censoring date. As the date of censoring, we used the last reported year and month that was given as an ending time in the employment, education or "other activity" career. There were 14 cases, where the last reported date was in the residential career. We censored the case at this date. For three cases, there were two similar dates of one move. We flagged out these cases by the variable ODD_MIG.

Table 10 shows a data extract from the residential career. Table 11 displays the career after manipulating the data.

Table 10: Data extract from the residential career

bid	m4_y	m4_m	m5	bh4y	CENSORY	CENSORM
111109	1966	january	ukraina	1949	2005	4
111109	1976	july	primorsk	1949	2005	4
111109	1983	december	leningra	1949	2005	4
111109	1993	october	ta zhe,	1949	2005	4

Notes: bid: Personal identification number; m4_y: Year start of residing in a new area; m4_m: Month start of residing in a new area; m5: Region code; bh4y: Year of birth; CENSORY: Year of censoring; CENSORM: Month of censoring

Table 11: Data extract from the residential career after manipulating the data

bid	BEGIN	END	m5
111109	0	126	ukraina
111109	126	215	primorsk
111109	215	333	leningra
111109	333	472	ta zhe,

Notes: bid: Personal identification number; BEGIN: Start of the episode in person months since January of the year the respondent turned age 17; END: End of the episode in person months since January of the year the respondent turned age 17; m5: Region code

In a final step, we generated a cleaned variable that denotes the place of residence (m5new). We created it in a way that all labels of this variable provide the English name of the region. Furthermore, if a person reported "the same" (*ta zhe*), we use information on region that was provided in the previous spell of the respondent.

2.3.8 Peculiarities in the fertility career

In the EES, there is one variable for the year a child was born and another one for the month of childbirth. The data set-up is exemplified in Table 12. Based on the year and month of childbirth, we generated the fertility career, as shown in Table 12 and 13. We generated a new variable called PARITY which denotes the parity of he respondent. We also generated a censoring date similar to the censoring date we have used in the residential career (see section 2.3.7).

In the EES, only the date of birth is recorded. No further information on children is given; neither if the birth is a twin or triplet birth. This means that for higher order births, one is not able to determine the number of children of a respondent. Also births that occur before age 17 are not recorded in the data set. (Merging the EES-data to the GGS would provide more accurate fertility histories).

Table 12: Data extract from the fertility career

bid	y1	Y3	bh4y	CENSORY	CENSORM
111109	1969	March	1949	2005	4
111109	1974	July	1949	2005	4
111109	1977	May	1949	2005	4

Notes: bid: Personal identification number; y1: year; y3: month of childbirth; bh4y: Year of birth of respondent; CENSORY: Year of censoring; CENSORM: Month of censoring

Table 13: Data extract from the fertility career after manipulating the data

bid	BEGIN	END	PARITY
111109	0	39	0
111109	39	103	1
111109	103	137	2
111109	137	473	3

Notes: bid: Personal identification number; BEGIN: Start of the episode in person months since January of the year the respondent had turned age 17; END: End of the episode in person months since January of the year the respondent had turned age 17; PARITY: Number of births

2.4 Merging the careers into a single episode data set

In a final step, we merged the seven careers into a single episode data set. For joining the data, we used the Stata-Ado-file „Spelljoin" (Walke und Kreyenfeld 2006). Table 14 and 15 exemplify how we proceeded based on the first respondent in the data set (with the bid 101009). Tables 14a to 14d show the separate biographies. For improved readability, we only selected four files and only one control variable from each file. Table 15 finally shows the complete history. The final data set is saved as EES01_clean.dta. Before saving, we also "cleaned" the labels of the variables. The final data set includes 6,455 respondents (just like the original EES file).

Table 14a: Data extract from the educational career

bid	BEGIN	END	o11
101009	0	18	secondar
101009	18	20	-1
101009	20	66	higher:

Table 14b: Data extract from the main employment career

bid	BEGIN	END	z15
101009	0	67	-1
101009	67	155	professi
101009	155	236	workers
101009	236	412	professi

Table 14c: Data extract from the residential career

bid	BEGIN	END	m5ew
101009	0	20	Leningrad Oblast
101009	20	67	Saint Petersburg
101009	67	412	Leningrad Oblast

Table 14d: Data extract from the fertility career

bid	BEGIN	END	PARITY
101009	0	171	0
101009	171	412	1

Table 15: Data extract from the final data set

bid	BEGIN	END	o11	z15	m5new	PARITY
101009	0	18	secondar	-1	Leningrad Oblast	0
101009	18	20	-1	-1	Leningrad Oblast	0
101009	20	66	higher:	-1	Saint Petersburg	0
101009	66	67	-1	-1	Saint Petersburg	0
101009	67	155	-1	professi	Leningrad Oblast	0
101009	155	171	-1	workers	Leningrad Oblast	0
101009	171	236	-1	workers	Leningrad Oblast	1
101009	236	412	-1	professi	Leningrad Oblast	1

3 An Application: Employment and Transition to First Birth

3.1 Descriptive statistics of employment pattern

In this part, we demonstrate how to use the EES data to estimate a first birth model. To simplify matters, we only use two covariates: cohort and activity status. Table 16 provides the distribution of the cohort variable which we used for the analysis. Since this variable is a time-constant variable, we show the distribution by respondents (instead of person-months).

Table 16: Descriptive statistics of time-constant covariates

	Males	Females
Cohort		
1948-1954	13.7%	17.9%
1955-1959	14.3%	16.6%
1960-1964	14.5%	16.3%
1965-1969	13.9%	12.0%
1970-1974	13.3%	12.9%
1975-1979	12.6%	11.5%
1980-1988	17.6%	12.8%
Total	100%	100%
Number of cases	2,460	3,995

In a next step, we generated a new variable which unifies the different activities into one variable. This new variable distinguishes between (1) periods of full-time education or periods of military service (2) periods of employment (3) periods of unemployment (4) periods of other activities and (5) periods of missing information. If there is an overlap of activities, we assigned a hierarchy (employment over education, education over unemployment and unemployment over other activity). Table 17 displays the distribution of the person months for this variable. The cases have not been censored at first birth. Therefore, the distribution gives the distribution of the activity status until time of interview.

Table 17: Descriptive statistics of time constant covariates, person-months

	Males	Females	Total
In Education	103760	122959	226719
Employment	458786	799112	1257897
Unemployment	18919	16410	35329
Other activity	25845	85890	111734
Missing	11019	65055	76074
Total person months	618328	1089424	1707752

3.2 Descriptive statistics of first birth pattern[2]

In this section, we describe the first birth pattern with the EES data. Altogether, we observe 5,063 first births. Figure 2a and Figure 2b displays Kaplan-Meier-survival curves by sex and cohort. Against the background of the societal transformation in Russia, one would expect a strong postponement of childbirth among younger cohorts. Particularly the cohorts who entered their reproductive period in the 1990s should show a pronounced differing childbearing pattern. However, the survival curves do not fully support this hypothesis.

Males of the cohorts 1948-1974 show a rather homogeneous pattern of an early first birth. The median age at first birth for males of these cohorts is age 25. Only the cohorts born 1975 or later have their first child much later. The median age at first birth for these cohorts is age 27.5 and therefore more than two years higher than for the younger cohorts.

For females, the first birth pattern is not as homogenous as for the older male cohorts. It might be best described as: Postponement, acceleration and postponement. The cohorts 1948-1964 had their first birth relatively late. The median age at first birth for these cohorts was age 24. The cohorts 1965-1974 were slightly younger at first birth. The cohorts born 1980 or later have their children at later ages again.

2 It should be noted that the EES is not designed for stand-alone analysis of fertility patterns. For this purpose the data should be merged with the Russian Generations and Gender Study. This section is proving fertility estimates for demonstrating purposes only.

Figure 2a: Kaplan-Meier-survival curves to first child, males

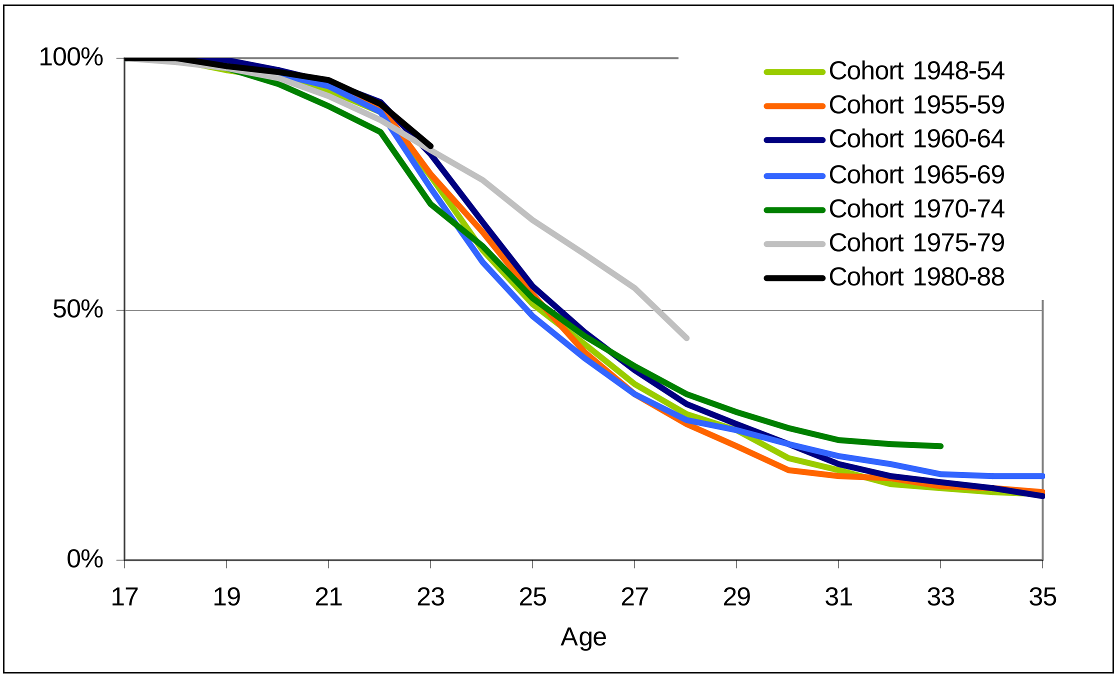
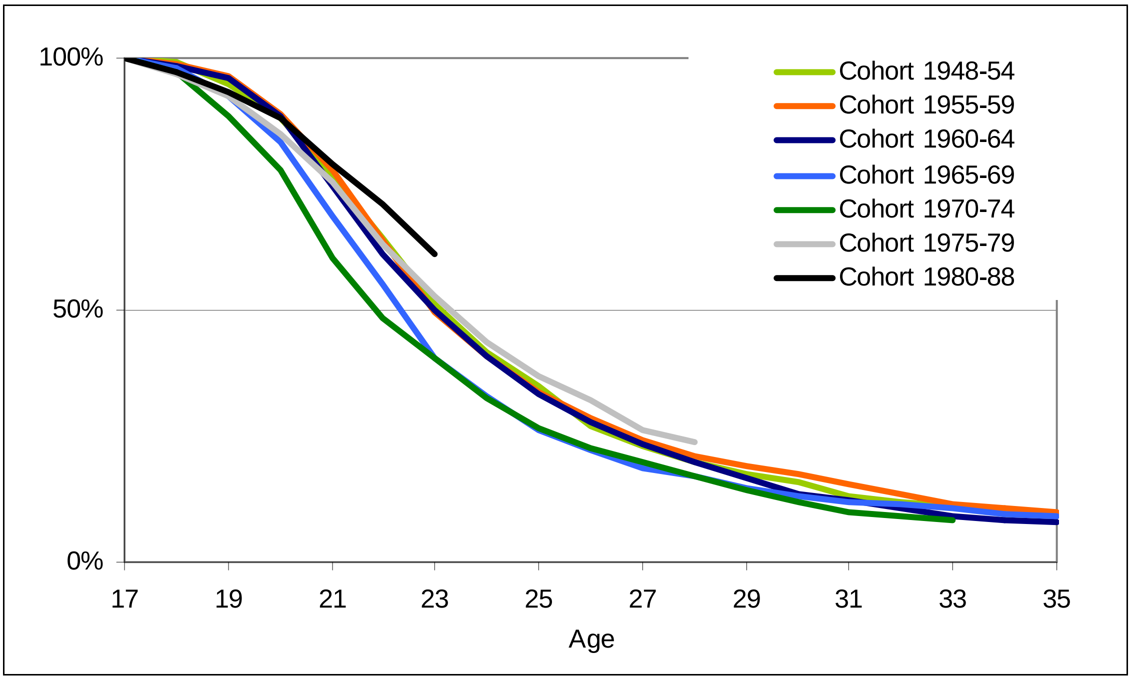


Figure 2b: Kaplan-Meier-survival curves to first child, females



3.3 Results from event history model

In this section, we estimate a first birth model. For this analysis, we have backdated the date of childbirth by nine months. By this procedure, we make sure that we measure the employment status before pregnancy. This is particularly of importance for females since they often reduce their labour market participation after pregnancy and childbirth. However, we are unable to take into account that women reduce their labour market activities in anticipation of a pregnancy. Because of backdating the date of childbirth by nine months, we are unable to consider births that have happened until September of the year the respondent turned age 17. We furthermore censored the cases at age 45. This leaves a sample of 5,003 births.

Table 18 displays the results from a piecewise constant model. The only control variables in the model are age (process time), cohort and activity status. The age profile is bell-shaped. For females, the highest intensity of childbirth is at ages 20-24, for males at ages 25-29. The results of the cohort variable support the previous analysis with the Kaplan-Meier survival curves. The activity status shows the expected pattern. To be in education cuts first birth risks roughly in half (compared to being employed). Unemployment and other types of activities reduce transition rates significantly for the male population. For females, unemployment has a positive but insignificant impact on first birth risks. Other activities significantly accelerate childbearing for females.

Table 18: Transition to first birth, relative risks, piecewise constant model

	Males	Females
Age		
Age 17-19	0.21 ***	0.88 ***
Age 20-24	1.01	1.55 ***
Age 25-29	1	1
Age 30-34	0.43 ***	0.64 ***
Age 35-39	0.25 ***	0.28 ***
Age 40-45	0.10 ***	0.07 ***
Activity status		
In Education	0.32 ***	0.41 ***
Employment	1	1
Unemployment	0.54 ***	1.23
Other activity	0.27 ***	1.43 ***
Missing	0.62 ***	0.66 ***
Cohort		
1948-1954	1	1
1955-1959	0.94	0.97
1960-1964	0.96	1.03
1965-1969	0.97	1.16 **
1970-1974	0.85 **	1.25 ***
1975-1979	0.61 ***	0.92
1980-1988	0.59 ***	0.77 ***

Notes: ***: $p \leq 0.01$ **: $0.01 \leq p \leq 0.05$ *: $0.05 \leq p \leq 0.10$.

3.4 Sensitivity analysis

In a final step, we investigated how stable the results are with respect to imputed cases. For this reason, we estimated three different kinds of models. The first model only includes age, cohort and activity status (as the model in section 3.3). The second model controls also for whether dates have been imputed or the career includes odd cases. In the third model, we omitted all odd cases or cases with imputed dates. We only conducted this sensitivity investigation for male respondents. The results are displayed in Table 19. The Table shows that the imputation of dates and/or the omission of “odd” cases from the analysis have no major impact on the results.

Table 19: Transition to first birth, relative risks, piecewise constant model, sensitivity analysis, males

	Model 1: Model including age, activity status and cohort	Model 2: Controlled for odd and omitted cases	Model 3: Odd and imputed cases omitted
Age			
Age 17-19	0.21 ***	0.21 ***	0.21***
Age 20-24	1.01	1.00	0.98
Age 25-29	1	1	1
Age 30-34	0.43 ***	0.43 ***	0.42***
Age 35-39	0.25 ***	0.25 ***	0.23***
Age 40-45	0.10 ***	0.11 ***	0.08***
Activity status			
In Education	0.32 ***	0.31 ***	0.31***
Employment	1	1	1
Unemployment	0.54 ***	0.54 ***	0.52***
Other activity	0.27 ***	0.28 ***	0.29***
Missing	0.62 ***	0.63 ***	0.62***
Cohort			
1948-1954	1	1	1
1955-1959	0.94	0.94	0.87
1960-1964	0.96	0.96	0.92
1965-1969	0.97	0.96	0.89
1970-1974	0.85 **	0.84 *	0.79**
1975-1979	0.61 ***	0.62 ***	0.60***
1980-1988	0.59 ***	0.59 ***	0.55***
Dates imputed			
Education		0.00	
Job date		1.03	
Other activity		0.00	
Odd cases			
Education		0.54	
Job date		1.10	
Other activity		0.89	

Notes: ***: $p \leq 0.01$ **: $0.01 \leq p \leq 0.05$ *: $0.05 \leq p \leq 0.10$.

4 Concluding Remarks

This paper documented how the Russian Education and Employment Survey (EES) had been cleaned so that it can be used for event history modelling. Our overall assessment is that the data is of fairly good quality. There were only few episodes with inaccurate starting and ending dates. We noticed only very few “odd” cases. The exclusion of these cases in our first birth analysis did not change the results much. Nevertheless, there are also some pitfalls of this data. A particular drawback is that we are often not able to distinguish non-response from “does not apply”. This particularly pertains to the fertility career, where it was impossible to distinguish childless respondents from those who failed to report their fertility career. Such drawback in the fertility career can be ironed out by combining the data with the Gender and Generation Survey (GGS). For the employment and educational career, it is not possible to double-check with the GGS data in the same manner.

5 Acknowledgements

The authors would like to thank Sunnee Billingsley and Rainer Walke for valuable comments on code and paper. We thank Alina Pelikh for the recoding of the regional information.

6 References

Independent Institute for Social Policy (2005): Education and Employment Survey on Russia - Technical Report, Moscow 2005.

Walke, Rainer/ Kreyenfeld, Michaela (2006): Spelljoin. A Stata-.ado file to join two spell data sets. Max Planck Institut für demografische Forschung Rostock.

Walke, Rainer/ Kreyenfeld, Michaela (2006): Spellsort. A Stata-.ado file to join two spell data sets. Max Planck Institut für demografische Forschung Rostock.

6.1 Variable overview

Bid	Respondent's id
SPELL	Number of episode
BEGIN	Start of episode, since January of year respondent turned 17
END	End of episode, since January of year respondent turned 17
Sex	Sex of respondent
bh4y	Year of birth of respondent
EDU	EDU: Episode from educational career
co10	EDU: Education interrupted?
o11	EDU: Type of education
o12	EDU: Diploma received?
o13	EDU: Full, part-time study?
o9ck	EDU: Main activity? at end
o9cs	EDU: Main activity? at start
JOB	JOB: Episode from main employment career
z15	JOB: Occupation
z16	JOB: Type of employee
z17	JOB: Ownership
z18	JOB: Industry
z19	JOB: Position, at end
z19s	JOB: Position, at start
z20	JOB: Working hours/week, at end
z20s	JOB: Working hours/week, at start
z21	JOB: Working schedule/week, at end
z21s	JOB: Working schedule/week, at start
JOB2	JOB2: Episode from 2nd job career
d23	2nd JOB: Occupation
d24	2nd JOB: Type of employee
d25	2nd JOB: Ownership
d26	2nd JOB: Industry
d27	2nd JOB: Position, at end
d27s	2nd JOB: Position, at start
d28	2nd JOB: Working hours/week, at end
d28s	2nd JOB: Working hours/week, at start
d29	2nd JOB: Working schedule/week, at end
d29s	2nd JOB: Working schedule/week, at start
OTHER	OTHER: Episode from other activity career
r31cs	OTHER: Main activity at start
r31ck	OTHER: Main activity at end
r32s_c	OTHER: Type of leave
ob_code	OTHER: Type of non-employment
UNEMP	UNEMP: Unemployment benefits received
UNEMP_r31cs	UNEMP: Unemployment benefits received, main activity
UNEMP_r31s_c	UNEMP: Occupation type, (unemployed)

MIG	MIG: Episodes from residential career
m5	MIG: Region code
m6	MIG: Type of settlement
m7	MIG: Distance
m8	MIG: Reason
m5new	MIG: Region code (cleaned)
PARITY	Number of births
IMP_EDU	Dates educational career imputed
IMP_JOB	Dates main employment career imputed
IMP_OTH	Dates other activity career imputed
IMP_SEC	Dates second job career imputed
IMP_UNE	Dates unemployment career imputed
IMP_RES	Dates residential career imputed
IMP_FER	Dates fertility career imputed
ODD_EDU	Odd educational career
ODD_JOB	Odd main employment career
ODD_OTH	Odd other activity career
ODD_SEC	Odd second job career
ODD_MIG	Odd residential career