Deliverable D21

Report on age profiles of family and fertility events

Work Package 5
Fertility and Living Arrangements

Authors:

Roberto Impicciatore & Francesco C. Billari
with the assistance of Agnese Vitali

"Carlo F. Dondena" Centre for Research on Social Dynamics, Università Bocconi
Milan, Italy

Contents

Milan, May 2007

Abstract

The present report constitutes Deliverable 21 of the project MicMac: Bridging the micro-macro gap in population forecasting. The objective of MicMac is to develop a methodology that offers a bridge between aggregate projections of cohorts (Mac) and projections of the life courses of individual cohort members (Mic). The project is supported by the European Commission within the 6th Framework Programme[1].

This report is part of the Work Package 5 "Fertility and living arrangements". The aim of this WP, which focuses on transitions between various living arrangements and on fertility behaviours, is the estimation of data to be included as input in the MicMac model. This report deals with the methodological background of the estimation of age profiles starting from survey data. A method for the estimation of age profiles for the relevant transitions is presented, with a documented implementation in the R software package—the common language of the MicMac project. Presentation of statistical models on rates with covariates is also included.

The method is applied to micro-level data from the Household survey "Famiglie e Soggetti Sociali", conducted in Italy at the end of 2003 and from the Fertility and Family Survey conducted in the Netherlands during the same year. Data preparation for this specific application is also discussed in the reports. The methods developed here can be applied in general to a setting where micro-level data on transitions are available from a large-scale representative survey.

---

[1] More information on MicMac is available on the website www.micmac-projections.org.

.

1. Introduction

The new methodology for population forecasting that is being developed within the MicMac project does not only take into account macro demographic changes but also life course trajectories (see, e.g., Willekens, 2005; van der Gaag et al., 2006). In this framework, the life course is viewed as a sequence of states and events; each event marks a transition from one state to another state. The study of a single transition is based on the estimation of a transition rate (from the original state to the destination state). From the literature on living arrangements and fertility (see, e.g., Billari et al., 2005) we know that transition rates vary with age. Indeed, such variation with age has been traditionally exploited in demographic forecasting.

The present report proposes a general method for the calculation of age profiles for the main transitions experienced by individuals, as far as living arrangement and fertility behaviours are concerned. Our analytical strategy starts from micro data, on the assumption that such data have to be collected allowing for a longitudinal (at least in a retrospective sense) reconstruction of the life course. This means that our method requires data that allow to reconstruct the biography of the individual and, for this purpose, all the dates of the more important events must be collected. Longitudinal data can be obtained both from retrospective surveys and from panel surveys. Here, we only refer to the former case. *Retrospective surveys* concerning family and fertility behaviours are rather available for most European countries (e.g., from the Fertility and Family Survey project, from the Generations and Gender Project and from other data collection ventures based on National Statistical Offices). Retrospective surveys permit to collect a wide range of information relating to the past experience of individuals with limited costs. On the other hand, the attention is limited uniquely to the survivors since we do not have information about deaths nor about emigrated individuals. However, this feature is not necessarily a disadvantage in methodological terms since we consistently reduce the number of events that drives the individual out of the observed sample over time, simplifying the calculation of transition rates. In the literature on demographic microsimulation, biographic information collected in retrospective surveys has often been used (see, e.g., Wachter et al., 1998).

The method can be segmented in four steps. The first step consists in the specification of the required micro-data for the evaluation of the age profiles for all the transitions relating to living arrangement and fertility presented in the Deliverable D2 "Report on data input requirement of MAC" (de Beer et al., 2006). In Section 2 we indicate all the information that the starting data files should contain in order to apply the method of calculation.
The second step relates to data preparation and includes the computation of ages at various events and the definition of the window of observation that we will consider for each individual.
The third step consists in the specification of the transition-specific data matrix: within the window of ages, the events experienced and the time spent in a specific state contribute to the calculation of events and time of exposure for all the individuals in the sample. Starting from a data file where each record is a set of dates and status variables relating to the *i*-th individual, we obtain a matrix for each transition where the single row refers to a specific age x and containing the number of events and time of exposure. Second and third step are discussed in Section 3.
In the last step, presented in Section 4, we analyze how an observed set of events and time of exposure can be modelled by a smoother function, obtaining age profiles. In particular, we develop GAM (Generalized Additive Model) that permits to evaluate the smoothed age profile as the transition rate baseline and, at the same time, to evaluate the effects of a vector of covariates as multiplicative changes from the baseline.
Finally, in Section 5 and 6 the method is applied to Italy and the Netherlands and age profiles for the most important events relating to family and fertility are shown.

Fig. 1 Organization of the method.

First step Data requirement (specification of the required micro-data) ⇩ *Second step* Data preparation (computation of ages; definition of the observation window) ⇩ *Third step* Transition-specific data matrix ⇩ *Fourth step* GAM models and relative risks
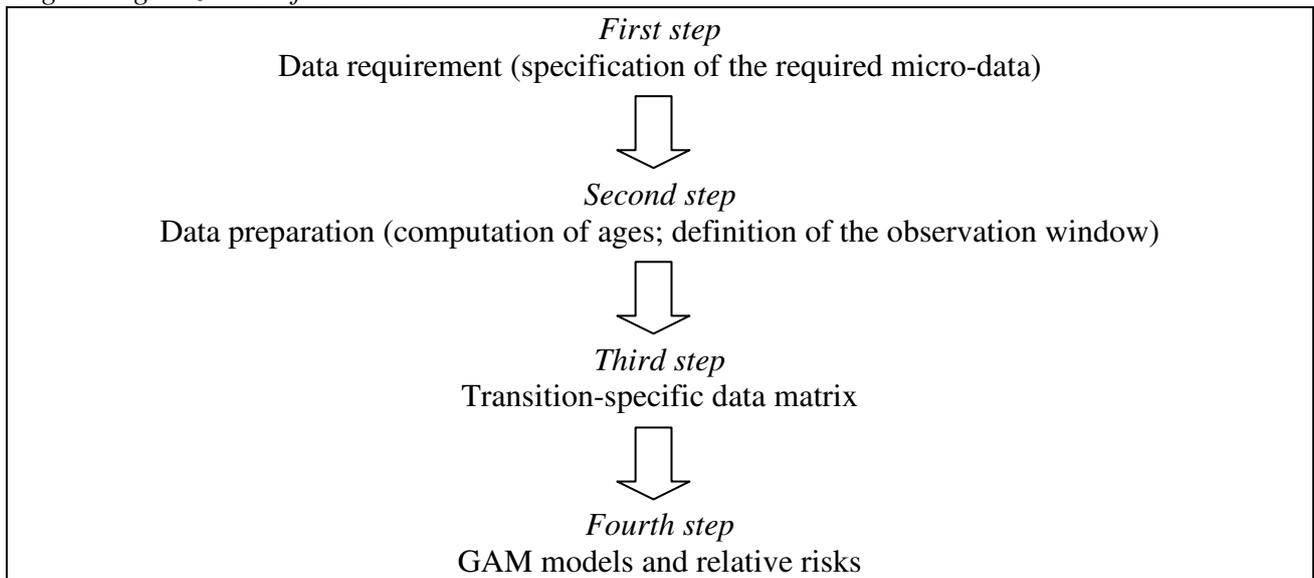
The data used in this report are the following:
- the ISTAT survey called "Famiglia e soggetti sociali" (FFS-IT) conducted in Italy at the end of 2003
- the Fertility and Family survey for Netherlands (FFS-NL) conducted between February and May of 2003.

The characteristics of the information contained in these datasets require some specific adaptation to the general indications given in Section 2. Besides, some transitions could not be considered because of missing information in the data set (for example, in both surveys there are no questions concerning the return in the parental home after a temporary exit) or because some transitions are very rare, thus implying a too small number of cases for the method to be correctly applied. For each transition that we are able to study, we report all the needed information for the analysis and, hence, we examine more deeply what has been discussed in Section 2. The resulting age profiles are plotted and included in this report.

All the passages of the method are developed in R, a very suitable statistical free and open source software. We include the whole R syntax we developed in Appendix 1.

2. Input data requirements

This section describes the data required for the calculation of age profiles on the transitions between various living arrangements and concerning fertility. Considering micro level data originated from a retrospective survey, in the starting data file we need the individual complete biography concerning family and fertility behavior. More specifically, we assume to have information on the date of birth, the date of the interview and the date of each event experienced within the range considered. All dates are expected to be available on a monthly time scale, i.e. to be expressed in calendar month (format MM: 1 to 12) and year (format YYYY).

Moreover, for each specific transition, we need a *status* variable at the interview that is, for a generic transition *TR*:

- 0 if the individual has never experienced *TR* at the time of the interview
- *1* if the individual experienced *TR* before the interview
- *9* if the case is not applicable, i.e. the individual has never been at risk to experience *TR*.

As an example, we consider the transition *TR*="first child→second child". The status variable is 1 if the second child was born before the interview; is 0 if the individual has only one child; is 9 if the individual is still childless at the interview.

2.1 Marital status

The state space for marital status is composed by four states:

- never married
- married
- divorced
- widowed

These categories imply that we shall not consider explicitly the order of marriage. That is, the behavior of persons in their second marriage cannot be distinguished from persons in their first marriage. The qualitative shape of the transition matrix is as follows:

From \ to	Never married	Married	Divorced	Widowed
Never married	O	TR1		
Married		O	TR2	TR3
Divorced		TR4	O	
Widowed		TR5		O

"TRX"   transition between categories
"empty space" impossible event
"O"   non-event

All these transitions are experienced when one of these events (marriage, divorce, death of spouse) occurs. In the starting data file we need dates for these events (if any) and 5 status variables at the interview:

- date of first marriage
- date of second marriage
- date of divorce
- date of death of spouse

- status for TR1 (first marriage)
- status for TR2 (divorce)
- status for TR3 (death of spouse)
- status for TR4 (second marriage after a divorce)
- status for TR5 (second marriage after death of spouse)

2.2 Living arrangement

The choice of possible states for the living arrangements strongly depends on the characteristics of the country examined. As a general rule (see de Beer et al., 2006), we could consider the following states: living with parents, living without a partner, living with a partner, living with other persons, living in an institution. However, some of these states and/or transitions are relatively rare in a specific country. This is, for example, the case of Italy, where the distinction between living alone (without a partner), with other persons, and in institution cannot be adequately studied. Therefore, we prefer to aggregate some states in order to have an adequate number of cases. In detail, we consider: living with parents, living with a partner, living alone or with other people. The qualitative shape of the transition matrix is:

From \ to	at parental home	with a partner	alone or with other persons
at parental home	O	TR6	TR7
with a partner	TR8	O	TR9
alone or with other persons	TR10	TR11	O

"TRX"   transition between categories
"empty space" impossible event
"O"   non-event

We need all dates of the events that cause these transitions and a status variable for each transition:

- date of exit from parental home
- date of (first, second, third) return into parental home
- date of the beginning of (first, second, third) union
- date of the end of (first, second, third) union
- status for TR6 (exit from parental home with a partner)
- status for TR7 (exit from parental home alone or with other persons)
- status for TR8 (return into parental home after a union)
- status for TR9 ("alone or with other persons" after a union")
- status for TR10 (return into parental home after "alone or with other persons")
- status for TR11 (union after "alone or with other persons")

Here we need to underline that these transitions could be experienced twice or more by the same individual (for example, one can experience T6, then TR9, then TR10 and T6 again). If we have a very detailed biography that gives us all the needed information, we can consider more than one record for each individual in the data set if he/she experienced a certain transition more than once. However, it results very difficult to rely on the complete set of information mentioned above. As we will see in the case of Italy, we are obliged to totally exclude some transitions and limit the others only at the first experience (for example, we can consider only the first exit from parental home).

2.3 Fertility (number of children ever born)

Women and men are distinguished by the number of children ever born. The possible states are then childless, 1 child, 2 children, 3 children, 4 or more children. The qualitative shape of the transition matrix is:

From \ to	childless	1 child	2 children	3 children	4+ children
Childless	O	TR12			
1 child		O	TR13		
2 children			O	TR14	
3 children				O	TR15
4+ children					O

"TRX"   transition between categories
"empty space" impossible event
"O"   non-event

Transition such as 0 → 2, 1→3, etc. caused by multiple births are not taken into account. A childless woman who has a twin birth simply experiences the transition 0→1 and 1→2 at the same date.

The i-th status variable accounts for the i-th child even born at the time of the interview (is 1 if the woman gave birth to at least i children, 0 if the woman gave birth to only $i-1$ children, 9 if the woman has less than $i-1$ children). Evidently, for TR12 the status variable can only take value 0 or 1.

Then, we need:

-   date at first birth
-   date at second birth
-   date at third birth
-   date at fourth birth
-   status for TR12 (first child)
-   status for TR13 (second child)
-   status for TR14 (third child)
-   status for TR15 (fourth child)

2.4 Children in the household

Differently from the previous section, now we consider the presence of children in the parental home, thus not simply the births. Given that in many applications, a dichotomy is sufficient (see D2), we consider two states: with and without children. The qualitative shape of the transition matrix is:

From \ to	Without children	With children
Without children	O	TR16
With children	TR17	O

"TRX"   transition between categories
"empty space" impossible event
"O"   non-event

The events that cause transition TR16 are the birth of the first child or the entry into a relationship with a partner who has one or more children. Transition TR17 is experienced when the last child exits from parental home (or dies) or following a divorce or separation causing children living at ex-partner's.

The status variables are dichotomous variables assuming value 0 when no event that causing transitions is experienced before the interview, and 1 otherwise.

Briefly, we need:

- date of entry into the status "with children"
- date of exit from the status "without children"
- status for TR16 (without children→ with children)
- status for TR17 (with children → without children)

2.5 Structure of the initial data set

The initial dataset should contain all the dates and variables specified in the previous sections. Then, the structure of records should be as follow:

ID	Weight	m.birth	y.birth	m.inter	y.inter	m.event1	y.event1	m.event2	y.event2	….
Identification number	weight	Month of birth	Year of birth	Month of interview	Year of interview	Month of event 1	Year of event 1	Month of event 2	Year of event 2	….

….	TR1	TR2	…	TR17	Sex	Edu
….	Status for transition TR1	Status for transition TR2	…	Status for transition TR17	Sex	Level of educational attainment

3. Data preparation

Given the initial dataset, the first step towards the estimation of age profiles is the calculation of exact ages in which each event occurred. Then, we must define a *window of observation* for our data. Our final aim is the estimation of the transition rates that will be used as input in the models developed within MicMac for population forecasting. Nevertheless, we easily understand that in order to forecast, or "project", individual behaviors in the future, we need to start from recently observed behaviors and, in case, manipulate the outcomes of such observations. Since we are working with retrospective data, we have information on the whole biography of individuals interviewed at all ages, including events (and transitions) experienced many years ago. This is why we limit our observation to the more recent past, excluding all the events that happened before. The choice of this window of observation is evidently subjective. The third step is the computation of a new status variable, that we can call *censor* and that tells us if the individual must be considered in the analysis, if she/he has experienced the specific transition in the fixed window of observation or if it is left censored.

Through ages and status variable, we have all we need for the effective calculation of relative risks. Depending on the type of observational plan, transition rates can be distinguished in *period rates*, based on period-age observations (calendar year in which an event occurs and the age at the time of event), *cohort rates* based on cohort-age observations (cohort to which a person belongs and the age in calendar years) and *period-cohort rates* based on period-cohort observation (calendar year in which the event occurs and the cohort to which the person belongs). Since we are using retrospective data, we use cohort rates for which the observation period extends over two calendar years and the age is in completed years. The last step is the determination of a smoother function for the observed age profile. We will often make direct reference to R syntax in order to represent computations.

3.1 Transformation of dates into ages

If we have, for each individual, the date of birth in the format month (MM: 1 to 12) – calendar year (YYYY), we can calculate the exact age at the interview with a precision of two digits. First of all, we have to transform a date from the format month-year in a decimal expression. Considering that an event that occurs in the month *m* can be pointed at the middle of this month, the R syntax for the computation of exact age at birth is:

```
birth.dec<- y.birth -1900 + round((m.birth-0.5)/12,2)
```

where round is the rounding function and 2 is the number of digits that we want to consider; *y.birth* and *m.birth* are respectively the year and the month of birth. Following the same procedure for the date of the interview (*m.inter*, *y.inter*), we can calculate the exact age at the interview as

```
ageint<-inter.dec-birth.dec
```

For example, if the interviews were conducted on November 2003 (decimal expression=103.88) and the individual was born on May 1955 (decimal expression=55.38), the exact age at the interview is 48.50 years. These formulas can be applied to the calculation of all the specified events: given *y.event* and *m.event* respectively the year and the month of a specified event, the expression

```
ageev<-y.event-1900+round((m.event-0.5)/12,2)-birth.dec
```

gives the exact age at the considered event.

3.2 The window of observation

All the information on the events in the living arrangement and fertility fields, are collected through surveys based on interviews to respondents aged at least 15 or 18 years at the interview. This is consistent with the dynamics of such behaviors in contemporary Europe (see, e.g., Billari et al., 2005). From now on, we consider all the individuals aged at least 18 years at the interview.

We have already stressed the necessity to focus our attention only on the more recent past. A plausible period could be the last five years before the interview. Then, we can consider all the events experienced within the age interval [x-5, x], where x is the exact age at the interview. However, as we have already underlined, the type of data used requires transition rates calculated as cohort rates (based on cohort-age observations) and then it seems reasonable to consider the beginning of the window of observation at the (x-5)-*th* birthday. The R syntax is

```
agesw<-trunc(age.inter)-5
```

For example, if an individual were interviewed on November 2003 at the exact age 40.35, the window of observation starts at the 35th birthday and ends at 40.35 years. Fig. 1 shows the windows of observation on the Lexis diagram.

Fig. 2. The window of observation in the Lexis diagram. Individuals are interviewed in a precise point in time during year t .
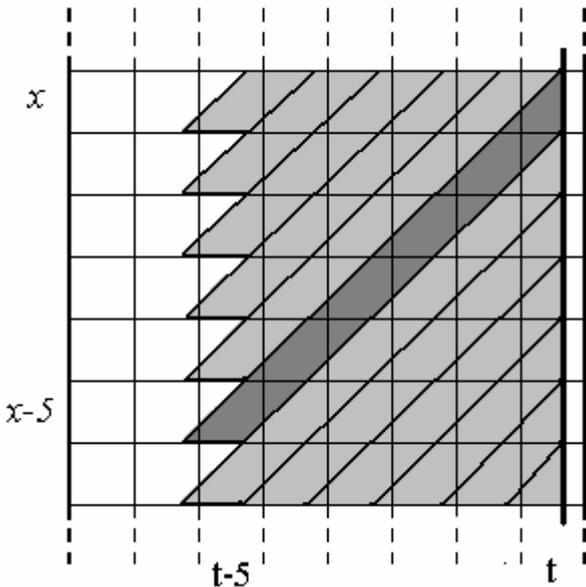
3.3 Episode limits and censoring

From this point onward, the procedure is transition-specific and it has to be repeated for each transition. Once we know the window of observation, we can define the limits of the episode that must be included in the window. The episode is the time interval spent in the initial state (the individual is at risk to experience the transition) and it is delimited by a starting age and a final age.

The starting age could be the age at the beginning of the window (*agesw*) of observation or a higher age, in case the individual enters the initial states after *agesw*. For example, in the transition TR1 (never married → married), the starting age is always the beginning of the window of observation, whereas for TR2 (married → divorced) the starting age follow *agesw* (age at the beginning of the window), if the age at marriage is higher than *agesw* (see Section 4 for more details).
The age at the end of the episode is alternately:

- the age at the interview if the case is censored at the interview (*ageint*);
- the age at the considered event, if the individual experienced the transition (*ageev*);
- the age at any other event that provokes the exit from the observation (*agecens*).

In other words, given that age is NA (not available) for events that are not experienced, the age at the end of the episode is the minimum among the three specified ages

```
agefin<-pmin(ageint, ageev, agecens, na.rm=TRUE)
```

where the command *pmin* determines the minimum values among vectors and the option *na.rm=TRUE* removes the NA values, i.e. not available ages are not taken into account.

The next step is the computation of a new status variable called *censor*, which indicates if the individual has to be included in the analysis, if it is censored or not and, in case, the kind of censoring (interview or other events). More in detail, for a generic transition TR*X*, *censor* assumes the following values:

0 if the case is censored at the interview
1 if the case experienced the event in the window of observation
2 if the case experienced another event in the window of observation that causes the exit of observation. Since we consider retrospective data, we do not have deaths and migrations. However, a competitive event could cause a censoring. For example, in the transition TR1, death of spouse causes the exit from observation.
9 if the case cannot be considered in the analysis. This situation emerges when the transition or an event that causes the exit of observation are experienced before *agesw*. This means that in the period considered, the case cannot contribute to the time of exposure.

Considering TR1 and calling *agemarr* the exact age at marriage (if any) and *marr* the TR1 status, the R syntax for defining of starting age (*agestart*), final age (*agefin*) and censor variable (*cens*) is:

```
agestart<-agesw
agefin<-pmin(agemarr,ageint, na.rm=TRUE)
cens<-ifelse(marr==1, ifelse (agemarr<agestart,9,1),
+ ifelse(marr==0,0,9))
```

3.4 Transition-specific data matrix

In order to estimate transition rates, the number of events and time of exposure have to be measured. For every transition, we consider single years of age from 0 to 100+. Considering a generic transition TR*X* from a state *A* to a state *B*, for the *j-th* individual we have a window of

observation included between x_j^{IN-WIN} and $x_j^{FIN-WIN}$, and an episode that starts at the age x_j^{in} and ends in x_j^{fin}.

In general, we have that

$$\underbrace{x_j^{IN-WIN} \leq \overbrace{x_j^{in} \leq x_j^{fin}}^{episode} \leq x_j^{FIN-WIN}}_{\textit{Window of observation}}$$

The *transition rate r_x* at age x is the ratio between the number of events E_x experienced at age x and the amount of time spent in the initial states (time of exposure PY_x) by individuals at the same age. Therefore, the next step is the computation of E_x *and* PY_x for every age x. Considering N as the number of individuals, we have:

$$E_x = \sum_{j=1}^{N} E_{j,x}$$

$$PY_x = \sum_{j=1}^{N} PY_{j,x}$$

where, for the *j*-th individual:

$$E_{j,x} = \begin{cases} 1 & \text{if } x_j^{in} < x < x_j^{fin} \text{ and the transition is experimented at the exact age x} \\ 0 & \text{otherwise} \end{cases}$$

$$PY_{j,x} = \begin{cases} 1 & \text{if } x_j^{in} < x < x_j^{fin} \text{ and neither transition nor exit from observation} \\ & \qquad \text{are experienced at the age x} \\ \delta_{j,x} & \text{if } x_j^{in} < x < x_j^{fin} \text{ and transition or exit from observation is experienced} \\ & \qquad \text{at the age x} \\ 0 & \text{if } x < x_j^{in} \text{ or } x > x_j^{fin} \end{cases}$$

where $\delta_{j,x}$ is the fraction of year spent in the initial state at the exact age in which the individual experiences the transition or the exit from observation. For example, let us suppose that in the transition (married → divorced) the *j*-th individual has an episode that start at the 40th birthday and ends at the exact age 42.31 with a divorce, for the age 41 his contributions for the event is 0 and his contribution for the time of exposure is 1; for the age 42 he contributes 1 event and 0.31 years to the time of exposure.

We can also include individual post-stratification weights w_j in the computation. The formulas become:

$$E_x = \sum_{j=1}^{N} E_{j,x} \cdot w_j$$

$$PY_x = \sum_{j=1}^{N} PY_{j,x} \cdot w_j$$

It is also possible to take into account some covariates by computing events and time of exposure separately for any combination of two or more categorical variables. In order to do so, we need to select sub-samples (defined for each combination of levels) to which apply the previous calculations.

We will consider two *timed-fixed* covariates, in the sense that their values remain fixed for the whole window of observation, sex and level of education, coded as follow:

sex categories

- 1. Men
- 2. Women

level of educational attainment categories

- 1. Primary education and less (ISCED 01)
- 2. Lower secondary education (ISCED 2)
- 3. Upper secondary education (ISCED 3-4)
- 4. Tertiary (ISCED 5A-6, 5B)

In table 1, we show a segment of a transition-specific data matrix calculated taking into account sex and education. For the age x we have 8 rows, one for each combination of levels for the covariates sex and education. We have a column with the number of cases (CASES) relating to a specific row (combination of age x and the specific level of covariates); unweighted (EVENT and EXPOS) and weighted (EVENTW and EXPOSW) events and time of exposure.

Table 1. Transition-specific data matrix

ID	AGE	CASES	EVENT	EVENTW	EXPOS	EXPOSW	SEX	EDU
[110,]	27	46	3	2.7286948	41.83	45.99903	1	1
[111,]	27	669	41	44.4088040	596.41	622.97497	1	2
[112,]	27	783	32	34.2193542	693.11	715.09001	1	3
[113,]	27	275	6	4.9138153	252.44	269.64722	1	4
[114,]	27	21	3	3.6084344	19.27	16.23176	2	1
[115,]	27	324	36	37.1002991	280.89	262.49605	2	2
[116,]	27	607	47	46.0331352	528.09	516.12050	2	3
[117,]	27	347	31	35.4466726	306.03	312.34037	2	4
[118,]	28	45	6	7.0276129	37.83	43.06423	1	1
[119,]	28	658	46	52.2252426	579.47	602.05693	1	2
[120,]	28	717	53	58.0949265	615.20	653.72342	1	3
[121,]	28	272	7	5.3103960	250.33	265.40648	1	4
[122,]	28	23	2	1.1703283	19.18	16.29679	2	1
[123,]	28	300	34	33.5822526	258.96	245.48345	2	2
[124,]	28	528	53	54.3586481	444.91	447.92291	2	3
[125,]	28	320	17	23.9958731	285.22	291.48234	2	4

4. GAM models and transition rates

If we consider the transition rate for a specific event as the dependent variable, we should model it as a function of age and a set of covariates. However, age profiles for a specific transition should never be considered as a linear function. Smoothing or graduating rates, or more specifically the age profile of rates, has been a traditional issue in various disciplines, including demography and actuarial science. Traditional approaches based on polynomials have been criticized in the literature since a long time, with authors proposing the use of spline functions as a solution (see, e.g., McNeil et al., 1977); recent developments include Smith et al., (2004) and, on age-specific fertility rates, Schmertmann (2003).

For our purpose a suitable solution are the so-called A*dditive Models* (Hastie & Tibshirani, 1990; Chambers & Hastie, 1992; Hastie *et al*, 2001) that are a generalization of linear model where the dependent variable Y can be modeled as a sum of non-linear (smoother) functions.

The model structure is

$$g(\mu) = \beta_0 + f(age) + \sum_k \beta_k X_k \tag{1}$$

where

$$\mu = E(Y) \quad \text{and} \quad g(.) \text{ is the link function}$$

and Y is the response variable (distributed as some exponential distributions); X_k is a generic covariate and β_k the corresponding parameter; β_0 is the intercept and *f(age)* is the smooth function of age.

Since transition rates at age x for a specific event is given by the ratio between number of events (*Events*) and the time of exposure (*Exp.time*), considering natural logarithm as link function, for each i-th row of data matrix[2] we can write:

$$\ln\left(\frac{Events_i}{Exp.time_i}\right) = \beta_0 + f(age_i) + \sum_k \beta_k X_{ki} + \varepsilon_i$$

where ε_i is a random error term. Then,

$$\ln(Events_i) = \ln(Exp.time_i) + \beta_0 + f(age_i) + \sum_k \beta_k X_{ki} + \varepsilon_i$$

or, considering the expected value

$$\ln(E[Events]) = offset[\ln(Exp.time)] + \beta_0 + f(age) + \sum_k \beta_k X_{ki} \tag{2}$$

where

---

[2] We remember that each row of the data matrix is given by a specific combination of age x and the levels of categorical covariates.

Events ~ Poisson

It is important that the term ln(*Exp.time*) has no coefficient to be estimated.

The smooth function *f* is a *piecewise cubic spline*, a curve made up of sections of cubic polynomial joined together so that they are continuous in value, as well as first and second derivatives. The points at which the sections join are known as the *knots* of the spline, that are placed at quantiles of the distribution of unique *x* values. The number of knots defines the *degree of smoothness* (i.e. number of knots + 2).

In order to avoid the choice of the is parameter, that is essentially arbitrary, the degree of smoothness of the *f* is estimated by Generalized Cross Validation[3] (Wood, 2006). The *mgcv* package contains a GAM implementation in which the degree of smoothness of model terms is estimated as part of fitting (see Wood, 2006). The R syntax for the model estimation without covariates is

```
require (mgcv)
m<-gam(event~offset(log(expos))+ s(age),family=poisson)
```

Calling $\hat{y}$ the fitted values of this model, the transition rate will be estimated as

$$\hat{r} = \frac{\hat{y}}{Exp.time} \qquad (4)$$

We estimate relative risks for each transition separately for men and women. The covariates in the models are:

1. level of education at the interview (considered constant throughout the window of observation)
2. be married (Yes/No) (*time-varying*)
3. having children in the household (Yes/No)

A time-varying variable is obtained by splitting the episode at the point where the event occurs (see Blossfeld and Rohwer, 2002). Each sub-episode will be characterized by a unique value of the variable. For example, let's consider the transition to the first child and an episode that starts at 22 and ends at 25.83 years of age. If the individual married at 23.24 years, then the episode is splitted into the two sub-episodes (22, 23.24) and (23.24, 25.83). The covariate "be married" has value "No" in the first sub-episode and "Yes" in the second.

The variable "having children in the household" requires a particular procedure: we must account for the entry of the first child in the household and for the exit of the last child. Therefore, if needed, we will split our subinterval twice.

We must underline three crucial aspects emerging from this approach:

a. In our dataset *Events* are calculated starting from individual weighted information. As a consequence, number of events and time of exposure are not integers. Since the *Poisson* distribution is defined only for integers, we need to round the number of weighted events. Empirical analyses (here not shown) suggest that this approximation appears acceptable.

---

[3] The way to control smoothness by altering the basis dimension, is to keep the it fixed at a size a little larger than it is believed could reasonably be necessary, but to control the model's smoothness by adding a "wiggliness" penalty to the least squares fitting objective (penalized regression spline). (Wood, 2006).

b. The effect of covariates should be considered as differences from the grand mean, i.e. from the mean risk for the whole sample[4]. Therefore, we use the "deviation coding" system that permits to compare the mean of the dependent variable for a given level to the overall mean of the dependent variable. If we consider, for example, the categorical covariate *Education* with 4 levels (primary school, lower secondary school, upper secondary school, tertiary school), the deviation coding is accomplished by assigning value "1" to level 1 for the first comparison (because level 1 is the level to be compared to all others), to level 2 for the second comparison (because level 2 is to be compared to all others), and to level 3 for the third comparison (because level 3 is to be compared to all others). The value "-1" is assigned to level 4 for all three comparisons (because it is the level that is never compared to the other levels). The value "0" is assigned to all other levels. In R we do not need to create the contrast matrix manually because the contr.sum function creates it for us. The syntax is[5]:

```
contrasts(edu) = contr.sum(4)
```

This command creates three dummies with values specified in table 2.

Table 2. Deviation coding for level of education

Level of education	Dummy 1 (Primary vs. mean)	Dummy 2 Low. sec. vs mean	Dummy 3 Upp. sec vs mean
Primary	1	0	0
Lower secondary	0	1	0
Upper secondary	0	0	1
Tertiary	-1	-1	-1

Fig. 3 Multiplicative effects of covariates estimated with additive model.

---
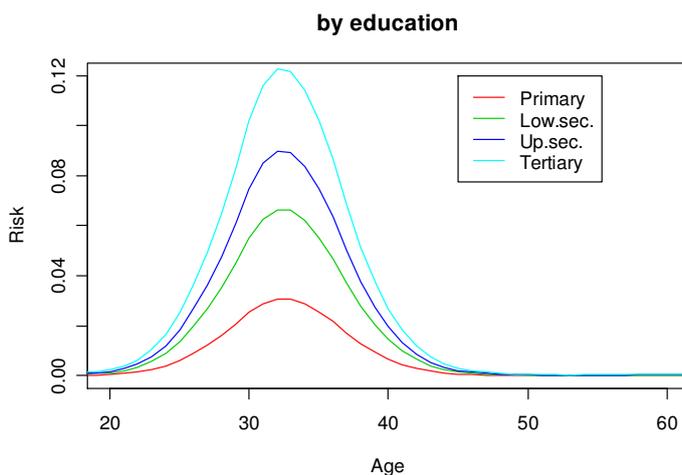
[4] We need this feature because in the development of the MicMac project we will use the effect of covariates estimated with micro data to the baseline age profile based on macro data.

[5] The simple coding, where each level is compared to the reference level is achieved with the R function contr.treatment.

The contrast estimate gives the proportional effect to be applied to the baseline risk. Given that the expected values of the dummies specified in such a way are always zero[6], we can obtain the baseline transition rate as:

$$baseline_i = e^{\beta_0 + f(age_i)}$$

c. The use of GAM models allows to include covariates in the equation and to evaluate their proportional effect on the smoothed age. Therefore, the estimated coefficients express multiplicative changes to be applied at the baseline age profile in order to evaluate the estimated risk for each year of age.

The simplest way to do so is to consider a vertical shift throughout the whole range of age. For example, in fig. 3 it is shown the multiplicative effect of the level of education on an unspecified transition.

However, very often the effect of a covariate shows a combination of vertical and horizontal shifts. In order to take into account this feature, a solution could be the estimation of the vertical shift for different specific sub-interval of age. In our analysis we split the range of age into 3 sub-intervals at two specific knots. The knots are fixed automatically at the 33rd and at the 67th percentiles (i.e. at the ages x_1 and x_2 at which, respectively, the 33% and the 67% of all the events are experienced before these ages).

The R syntax for the calculation of knots is:

```
knot<-c(0,0)
prop<-cumsum(events/sum(events))
knot[1]<- round(length(prop[prop<0.33])/8)
knot[2]<- round(length(prop[prop<=0.67])/8)
# define int: categorical variable indicating the sub-interval of age (1,2, or 3)
int<-as.factor(ifelse(age<knot[1],1,ifelse(age>=knot[1] & age<knot[2],2,3)))
```

The resulting model takes into account the baseline transition rate and the interaction between covariates and the age sub-intervals. The covariate deviation coding is adapted to this new feature: we evaluate the effect of each level of categorical variable within each subinterval. For example, R syntax for the definition of dummy 1 (primary level vs. mean) is

```
edu11<-ifelse(edu==1 & int==1,1,ifelse(edu==4 & int==1,-1,0))
```

The R syntax for the complete model (with level of education and one time-varying covariate that we call "timed") and the computation of baseline transition rates is

```
m<-gam(event~offset(log(expos))+ s(age)+ edu11+edu21+edu31+
edu12+edu22+edu32+edu13+edu23+edu33+timed1+timed2+timed3,family=poisson)
baseline<-exp(predict.gam(m,type="terms",terms="s(age)")+m$coeff[1])
```

In fig. 2 we can see an example of the effects of education on transition TR1 (never married-married) obtained by dividing the age range into three subintervals. The plotted baseline rates are reported analytically in table 3.

---

[6] More precisely, the expected values are zero if the number of cases is (approximately) the same for each levels. In our analysis this condition is satisfied given the structure of our data-matrix (equal number of rows for each combination of levels of covariates).

Fig. 4 Proportional effects of education on the transition TR1.

TR1 (never married->married)

Relative risks (deviation coding:differences from Baseline)			
Age interval	15-27	27-30	30-70
primary	2.189	1.196	0.878
lower sec.	1.533	1.08	0.81
upp. sec.	0.727	1.031	1.142
tertiary	0.41	0.751	1.232
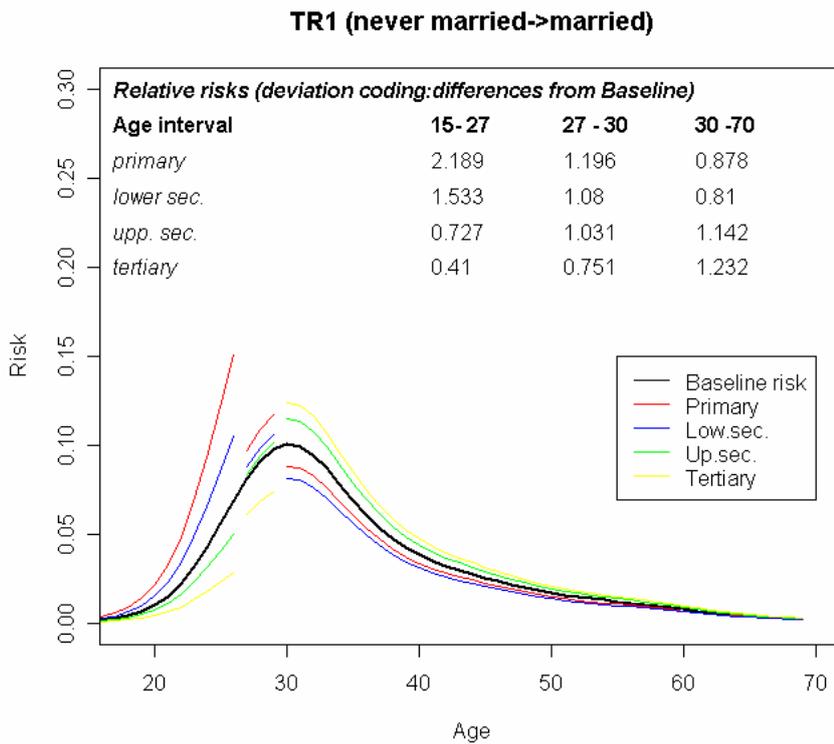
Table 3. Baseline transition rates for the transition TR1.

	Baseline transition rates			Baseline transition rates	
Age	Women	Men	*Age*	Women	Men
16	0.002	0.0008	*43*	0.023	0.045
17	0.0037	0.0012	*44*	0.0207	0.0421
18	0.0067	0.002	*45*	0.0183	0.0397
19	0.0117	0.0033	*46*	0.0158	0.0377
20	0.0195	0.0054	*47*	0.0134	0.0361
21	0.0307	0.0087	*48*	0.0112	0.0349
22	0.0451	0.0137	*49*	0.0092	0.0337
23	0.0617	0.021	*50*	0.0077	0.0327
24	0.0789	0.031	*51*	0.0066	0.0317
25	0.095	0.0439	*52*	0.0058	0.0305
26	0.1088	0.0591	*53*	0.0053	0.0292
27	0.1195	0.0753	*54*	0.005	0.0276
28	0.1267	0.0908	*55*	0.0049	0.0259
29	0.1299	0.1034	*56*	0.0049	0.0241
30	0.1289	0.1117	*57*	0.0049	0.0221
31	0.1234	0.1151	*58*	0.0049	0.0201
32	0.1139	0.114	*59*	0.0048	0.0182
33	0.1015	0.1092	*60*	0.0045	0.0164
34	0.0875	0.1022	*61*	0.0041	0.0147
35	0.0736	0.094	*62*	0.0036	0.0132
36	0.0611	0.0855	*63*	0.0031	0.0118
37	0.0507	0.0775	*64*	0.0025	0.0106
38	0.0424	0.0701	*65*	0.0021	0.0095
39	0.0362	0.0635	*66*	0.0017	0.0085
40	0.0316	0.0577	*67*	0.0014	0.0077
41	0.0281	0.0528	*68*	0.0011	0.0069
42	0.0254	0.0486	*69*	0.0009	0.0062

Finally, we can test the statistical significance of an additional covariate in the model by dropping it and noting the change in the deviance. The fitted models are compared using an analysis of deviance table. The tests are usually approximated, unless the models are un-penalized (Wood, 2006). In R this is accomplished by using the `anova.gam` function included in the MGCV package. We will only consider covariates that significantly increase the fit of the model when we add them into the equation.

5. Age profiles of family and fertility events in Italy

4.1 Data adaptations

Data for the Italian case come from the multipurpose survey called "Famiglia e soggetti sociali (FSS)". Carried out at the end of 2003, these data contain wide retrospective information on life course trajectories and the transition to adulthood, including data on the history of marital unions, cohabitations (followed by a marriage or not) and marital disruption, for a large sample of the resident population. The longitudinal nature of the survey makes it possible to update the collected information and to follow the same individual over time.

However, in this dataset we find limitations in the data available that requires some adaptation.

a. We do not have the date of birth of the respondent, but only the age (in completed years) at the interview (i.e. on 15 November 2003): an individual with the age x (say, 48) has an exact age that is in the interval (48.00, 48.99) and the date of birth is included between 16 November 1954 and 15 November 1955. This means that, even if we have dates in month and year for every event, we cannot define the exact age at which these events were experienced.

Since we do not have further information, we can avoid this indetermination adding a random month of birth extracted from a discrete uniform distribution $U(1,12)$. The R syntax is

```
m.birth<- trunc(runif(id,1,13))
```

b. For some events, only the calendar year has been asked. In particular, we do not have information on the month for the following events:

- exit from parental home
- divorce
- death of spouse
- exit from parental home (or death) of the last child

If the individual experienced the event, we can consider a random month for each event extracted from a discrete uniform distribution $U(1,12)$. For example, for the divorce, we have:

```
m.div<-ifelse(div=1,trunc(runif(id,1,13))),NA)
```

where *div* is the status variable for the transition married→divorced.

A remark is needed: in this case we add a second approximation to the one already introduced in point a. Since the available information concerns the respondent's age in completed years at the interview and the calendar year in which the divorce occurred, the exact age at this event is included

in a two-years range. For example, an individual aged 48 at the interview who divorced during 1995, could have experienced this event at an exact age included in the interval (39.12, 41.11), i.e. at 39, 40 or 41. For the month of exit from parental home we can use the additional information given by the month of marriage: if the individual experienced both exit from home and first marriage and if the two events occurred in the same calendar year, we can assume that the two events occurred simultaneously. Therefore, the month of exit from home corresponds to the month of first marriage. The R syntax is

```
m.exit<- ifelse(exit==1,ifelse(marr==1 & y.exit==y.marr,
+m.marr,trunc(runif(id,1,13)))),NA)
```

In the following tables we present detailed schemes containing the needed information for transition rate calculations adapted to the ISTAT 2003 data. It is not possible to analyze all the transitions proposed in section 2 because of lack of data or very few events (rare transition). Transitions are divided into four main sections: marital status, living arrangements, number of children ever born and children in the household.

a. Marital status

Transition	Event	Events that implies the exit from observation	Cases in the analysis	Starting age of exposure	Final age of exposure
TR1 (never-married → married)	First marriage	None	Who did not experience marriage before *agesw*	*agesw* (age at the beginning of the window of observation)	Age at marriage or Age at the interview
TR2 (married →divorced)	Divorce	Death of spouse	Ever married who did not divorce before *agesw*	*agesw* or age at marriage (if it occurs after *agesw*)	Age at divorce, or age at death of spouse or age at the interview
TR3 (married→ widowed)	Death of spouse	Divorce	Ever married with alive spouse at *agesw*	*agesw* or age at marriage (if it occurs after *agesw*)	Age at death of spouse, or age at divorce or age at the interview
TR4 (divorced→married)	Second marriage	None	Ever divorced without a second marriage before *agesw*	*agesw* or age at divorce (if it occurs after *agesw*)	Age at second marriage or age at interview
TR5 (widowed→married) (Rare transition. The window of observation is the whole life course)	Second marriage	None	Who experienced the death of spouses without a second marriage	age at death of spouse	Age at second marriage or age at interview

In Italy, the status of divorced is achievable only after a period of legal separation. Then, the event "legal separation" could lead towards at least two additional transitions, namely "marriage →

18

separation" and "separation→divorce". However, in the ISTAT 2003 the date of separation is frequently not specified: a consistent percentage of people who divorced did not give the date of separation but only the date of divorce. Then, it is not possible to take into account this additional event.

b. Living arrangements

Transition	Event	Events that implies the exit from observation	Cases in the analysis	Starting age of exposure	Final age of exposure
TR6 (At parental home → with a partner)	Exit from parental home	Exit from home (without a partner)	Who has not experienced exit from home before *agesw*	*agesw*	Age at exit with a partner or at exit without a partner or age at the interview
TR7 (At parental home → alone or without a partner)	Exit from parental home	Exit from home (with a partner)	Who has not experienced exit from home before *agesw*	*agesw*	Age at exit with a partner or at exit without a partner or age at the interview
TR8 (With a partner→ at parental home)	Return parental home	No information for this transition			
TR9 (with a partner→ alone or with other persons)	Exit from union	No information for this transition			
TR10 (alone or with other persons→ at parental home)	Return parental home	No information for this transition			
TR11 (alone or with other people→ with a partner)	First union	We have information on the event that causes transition (entry into a union) but not on the event that causes exit from observation (return parental home)			

c. Number of children ever born

Transition	Event	Events that implies the exit from observation	Cases in the analysis	Starting age of exposure	Final age of exposure
TR12 (0 child → 1 child)	First child	None	Who has no children at *agesw*	*agesw*	Age at first child or age at the interview
TR13 (1 child → 2 children)	Second child	None	Who has no more than one child at *agesw*	*Agesw* or *agech1*+nine months (if the first birth occurs within the window of observation)	Age at second child or age at the interview
TR14 (2 children → 3 children)	Third child	None	Who has no more than two children at *agesw*	*Agesw* or *agech2*+nine months (if the second birth occurs within the window of observation)	Age at third child or age at the interview
TR15 (3 children → 4 children)	Fourth child	None	Who has no more than three children at *agesw*	*agesw* or *agech3* + nine months (if the third birth occurs within the window of observation)	Age at fourth child or age at the interview

d. Children in the household

Transition	Event	Events that implies the exit from observation	Cases in the analysis	Starting age of exposure	Final age of exposure
TR16 (without child(ren) → with child(ren)	First birth or entry into a relationship with a partner who has one or more children	None	Who is without children at *agesw*	*agesw*	Age at first birth or entry into a relationship with a partner who has one or more children or age at the interview
TR17 (with child(ren) → without child(ren)	Exit from parental home or dead	None	Who is with children at *agesw*	*agesw* or age at first birth or entry into a relationship with a partner who has one or more children (if these events occur after *agesw*)	Age of exit or death of the last child (including the case after a divorce or separation) .

Transition-specific age profiles are estimated separately for men and women. Moreover, we considered the effect of educational level and the following time-varying covariates:

- married (Yes/No), included in TR12, TR13, TR14, and TR15
- with children in the household (Yes/No), included in TR1,TR2, TR4 and TR5.

In table 4, we have p-values associated with the Null Hypothesis that a specific covariate does not increase the fit of the model. In the columns marked with (1) the significance Chi-square test related to the comparison between the model without covariate and the model with education is shown; in columns marked with (2) and (3) the model with education is compared with the model adding, respectively, "married or not" and "with/without children in the household". Results obtained for women suggest to include education everywhere excepting TR4, TR5 and TR15, where the difference from the base model (without covariates) is not significant (at 95% level). For men, education could be excluded in TR2, TR3, TR4, TR5 and TR15. Time-varying covariate "married" could be excluded in TR14 and TR15 for both sex whereas "with children" should be included in TR2, TR4 and TR5 for women and TR1 and TR2 for men.

Table 4. Analysis of Deviance. Significance Chi-square test (p-value) comparing model without covariate and model with education (Column 1); model with education and model with "married or not" (column 2); model with education and model with "with/without children in the household" (column 3) by sex and transition.

	Women			Men		
	education	married	with children	education	married	with children
	(1)	(2)	(3)	(1)	(2)	(3)
TR1 never married -> married	0.000	-	0.066	0.000	-	0.000
TR2 married -> divorced	0.002	-	0.009	0.590	-	0.001
TR3 married -> widowed	0.000	-	-	0.124	-	-
TR4 divorced -> married	0.193	-	0.005	0.072	-	0.193
TR5 widowed -> married	0.084	-	0.003	0.051	-	0.428
TR6 parental home -> with a partner	0.000	-	-	0.000	-	-
TR7 parental home -> alone or with other persons	0.006	-	-	0.001	-	-
TR12 childless -> 1 child	0.000	0.000	-	0.000	0.000	-
TR13 1 -> 2 children	0.001	0.000	-	0.002	0.000	-
TR14 2->3 children	0.006	0.390	-	0.000	0.238	-
TR15 3->4 children	0.439	0.709	-	0.234	0.076	-
TR16 without children -> with children	0.000	-	-	0.000	-	-
TR17 with children -> without children	0.000	-	-	0.001	-	-

4.2 Age profiles

TR1 (never married->married) - Women

Level of education. Relative risks

Age interval	15 - 24	25 - 28	29 - 50
primary	2.237	1.657	1.046
lower sec.	2.108	1.273	0.779
upp. sec.	0.712	0.782	1.163
tertiary	0.298	0.607	1.055

Baseline risk
Primary
Low.sec.
Up.sec.
Tertiary

TR1 (never married->married) - Women

Children in the household. Relative risks

Age interval	15 - 24	25 - 28	29 - 50
Without children	0.785	1.012	0.876
With children	1.273	0.989	1.142

Baseline risk
Without children
With children
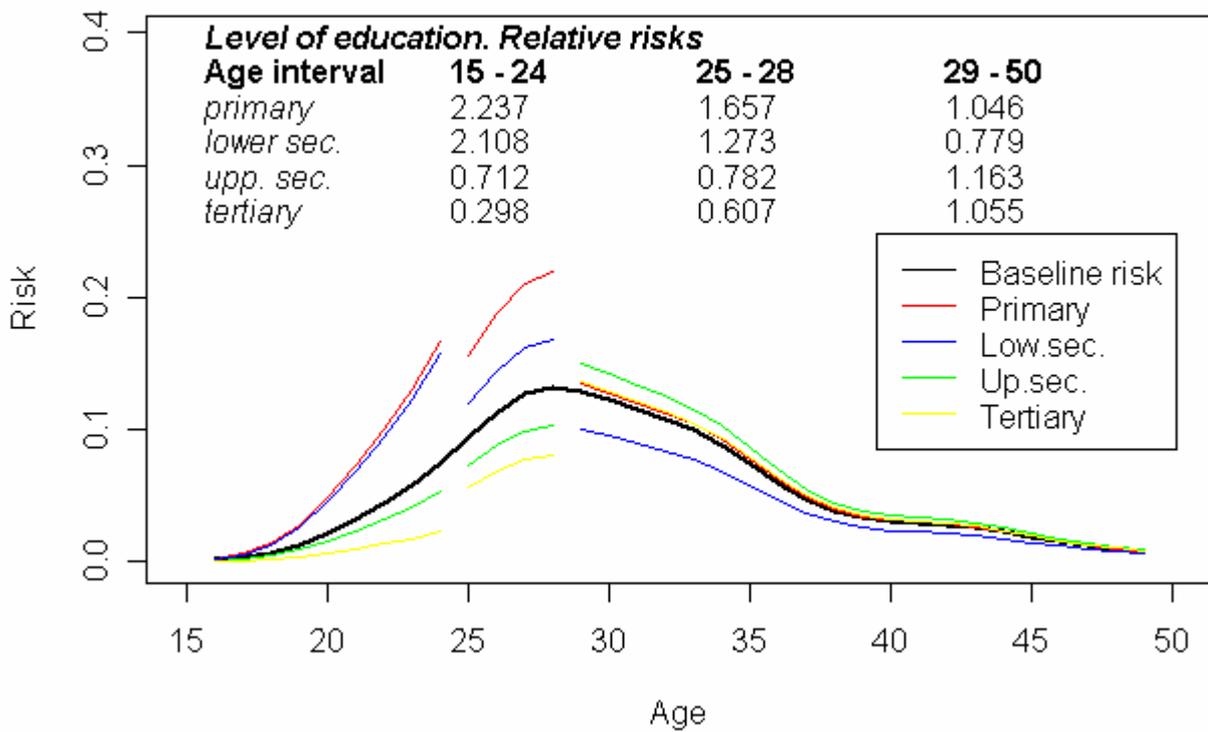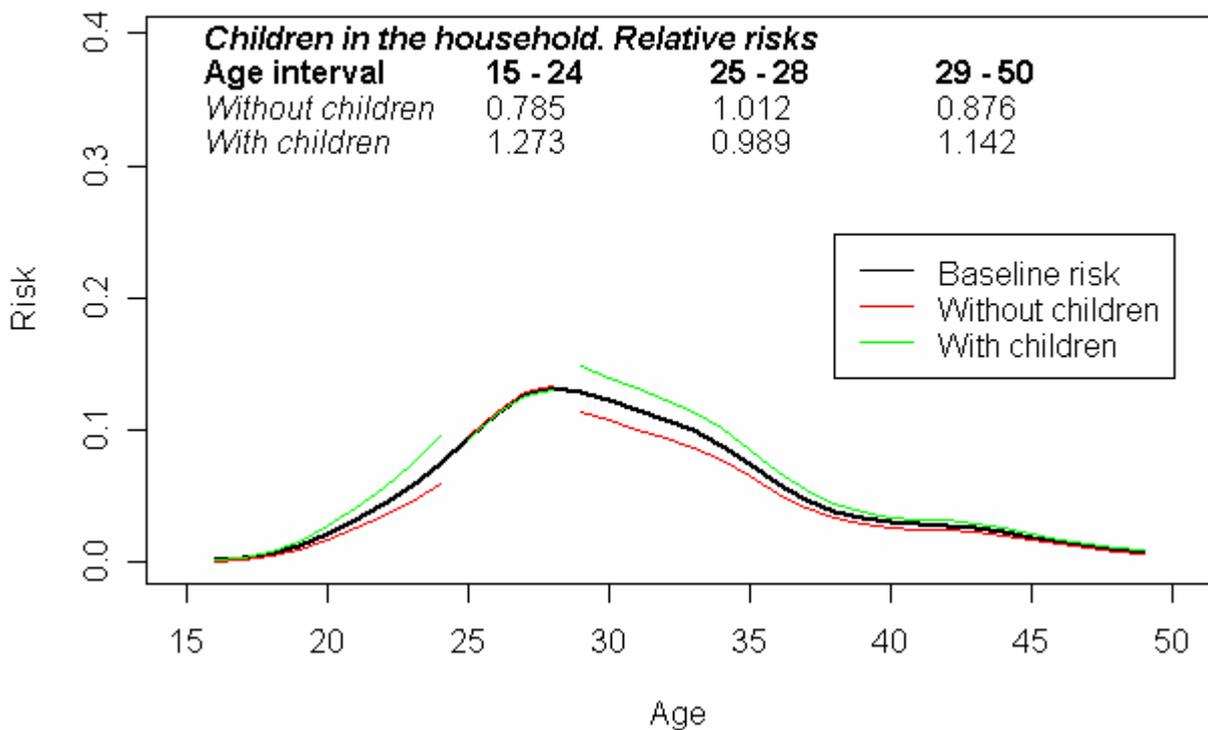
Italy. ISTAT 2003. TR1 Women. Events observed in the window of observation: 1306.

TR1 (never married->married) - Men

Level of education. Relative risks			
Age interval	15 - 27	28 - 31	32 - 50
primary	2.694	1.277	0.624
lower sec.	1.648	0.983	0.791
upp. sec.	0.799	1.122	1.177
tertiary	0.282	0.71	1.721

Legend:
- Baseline risk
- Primary
- Low.sec.
- Up.sec.
- Tertiary

TR1 (never married->married) - Men

Children in the household. Relative risks			
Age interval	15 - 27	28 - 31	32 - 50
Without children	0.573	0.939	0.739
With children	1.744	1.065	1.353

Legend:
- Baseline risk
- Without children
- With children
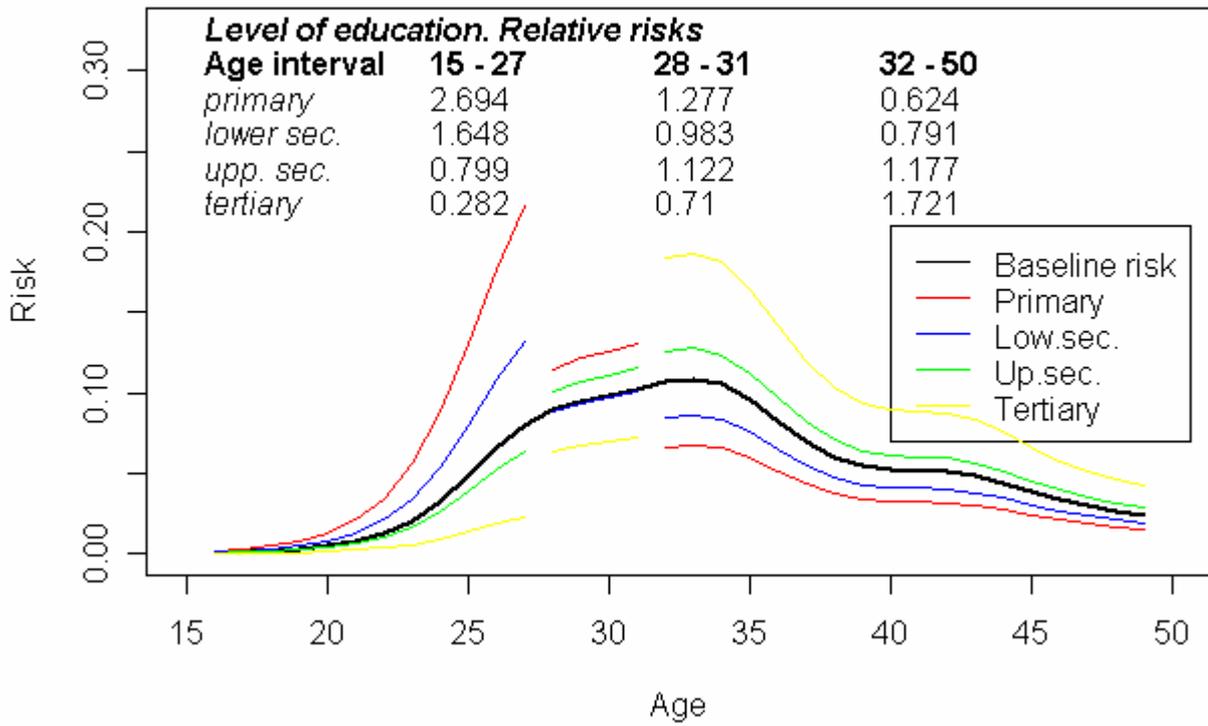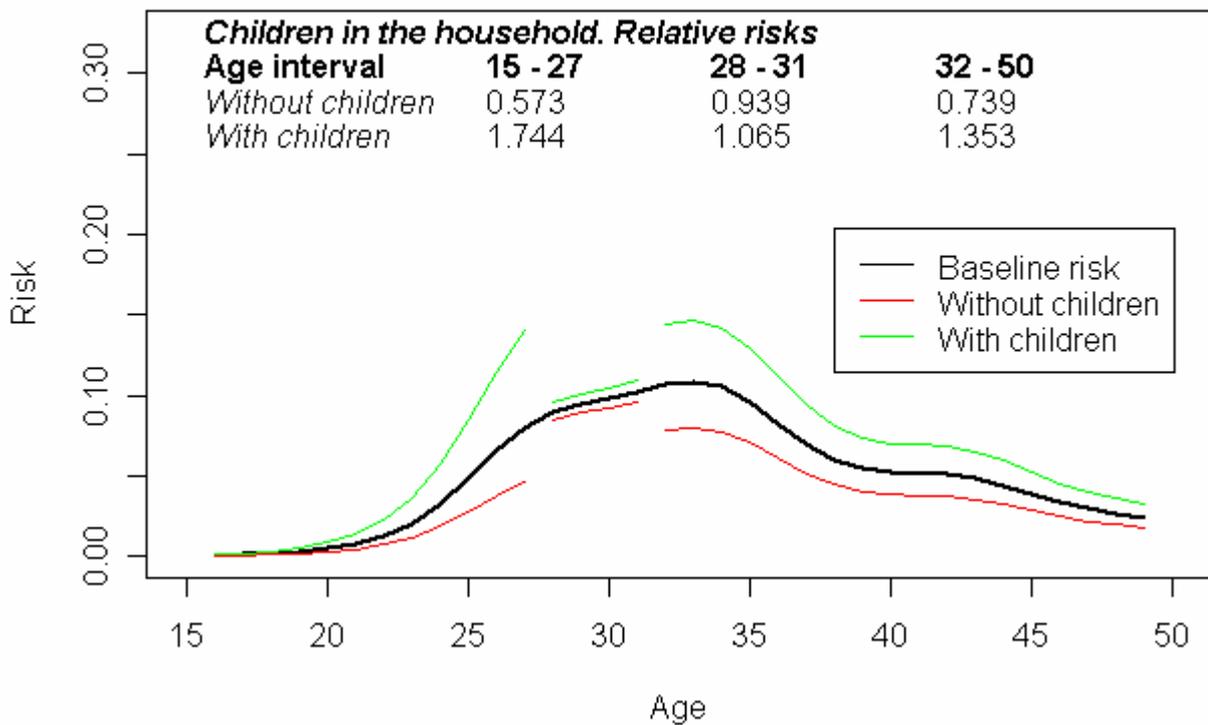
Italy. ISTAT 2003. TR 1 Men. Events observed in the window of observation: 1279.

TR2 (married -> divorced) - Women

Level of education. Relative risks

Age interval	15 - 40	41 - 54	55 - 90
primary	1.259	0.765	1.702
lower sec.	1.419	0.928	0.295
upp. sec.	0.812	0.936	1.774
tertiary	0.689	1.505	1.122

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR2 (married -> divorced) - Women

Children in the household. Relative risks

Age interval	15 - 40	41 - 54	55 - 90
Without children	1.588	1.079	0.64
With children	0.63	0.927	1.562

Legend: Baseline risk, Without children, With children
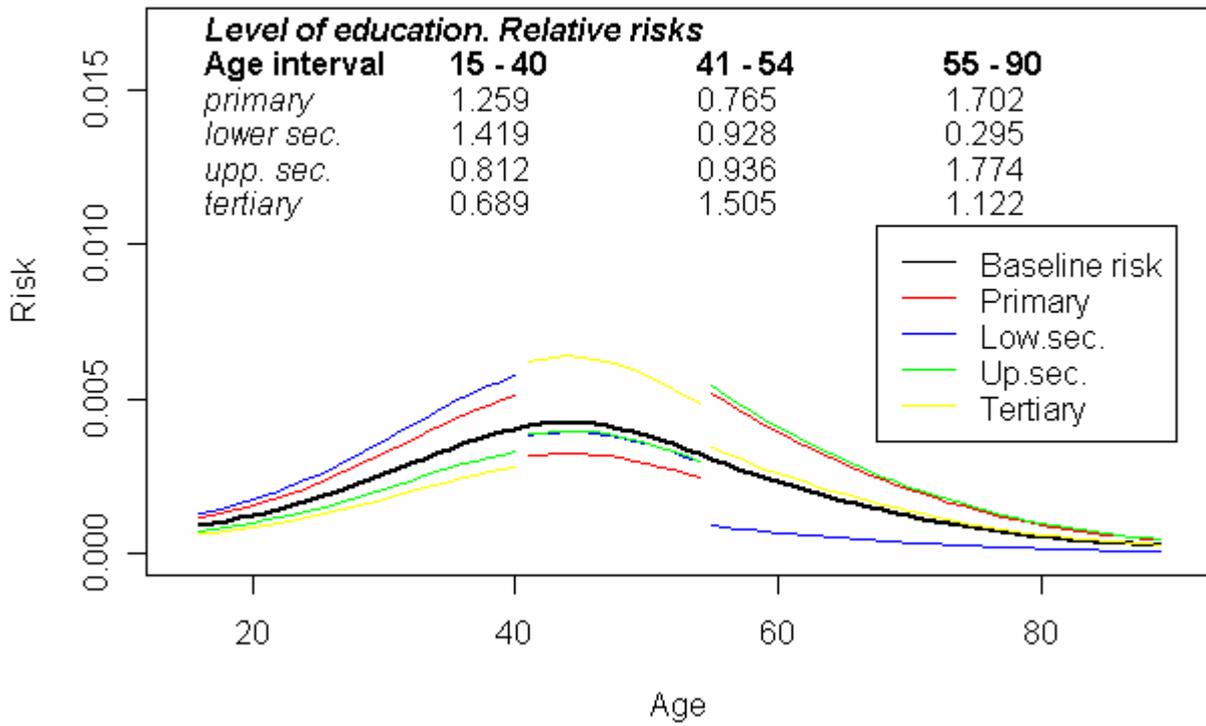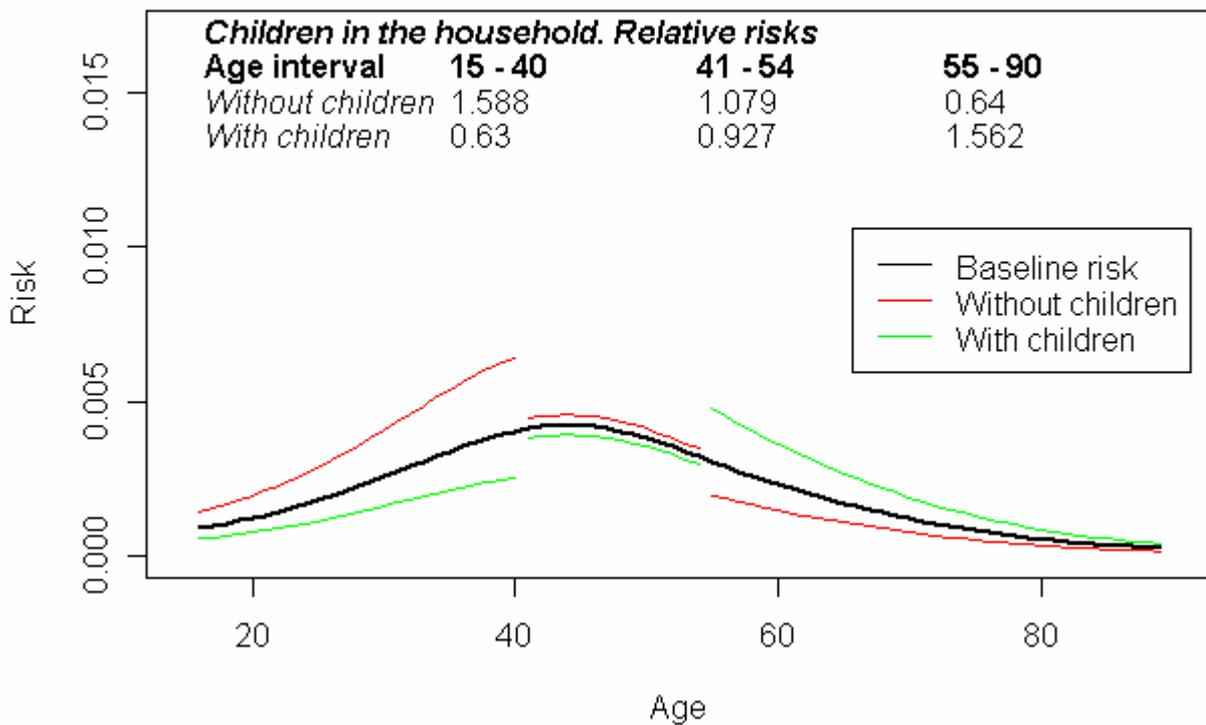
Italy. ISTAT 2003.TR2 Women. Events observed in the window of observation: 232.

25

TR2 (married -> divorced) - Men

Level of education. Relative risks			
Age interval	15 - 42	43 - 59	60 - 90
primary	1.405	1.22	1.18
lower sec.	0.995	0.848	0.661
upp. sec.	0.983	0.735	0.973
tertiary	0.727	1.316	1.318

TR2 (married -> divorced) - Men

Children in the household. Relative risks			
Age interval	15 - 42	43 - 59	60 - 90
Without children	1.467	1.176	0.443
With children	0.682	0.85	2.258
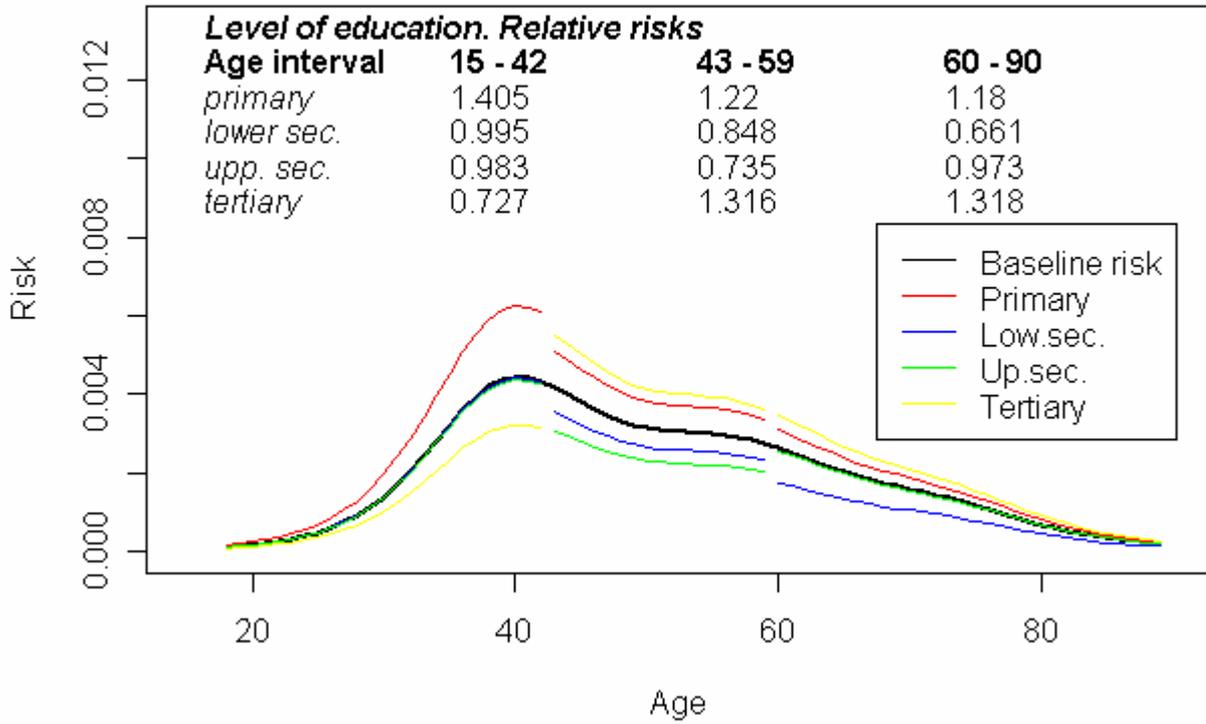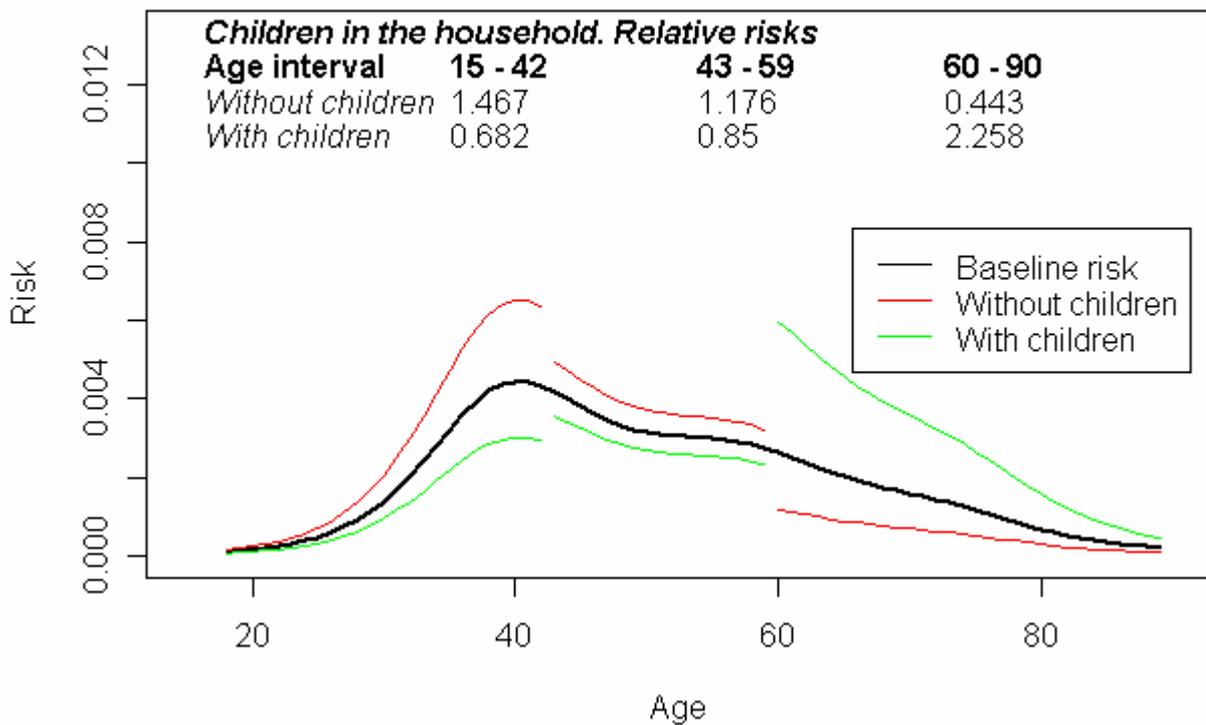
Italy. ISTAT 2003. TR2 Men. Events observed in the window of observation: 204.

TR3 (married -> widowed) - Women

Level of education. Relative risks			
Age interval	**30 - 60**	**61 - 69**	**70 - 95**
primary	1.4	1.291	1.362
lower sec.	1.118	0.831	0.995
upp. sec.	0.958	0.928	0.589
tertiary	0.667	1.005	1.253

Legend:
— Baseline risk
— Primary
— Low.sec.
— Up.sec.
— Tertiary

TR3 (married -> widowed) - Men

Level of education. Relative risks			
Age interval	**30 - 60**	**61 - 71**	**72 - 95**
primary	2.069	1.123	1.281
lower sec.	0.893	0.836	0.769
upp. sec.	0.639	1.36	1.331
tertiary	0.847	0.783	0.762

Legend:
— Baseline risk
— Primary
— Low.sec.
— Up.sec.
— Tertiary

Italy. ISTAT 2003. TR3. Events observed in the window of observation: Women 1359 ; Men 380.
NOTE: Window of observation extended to the last 10 years before the interview.

TR4 (divorced -> second marriage) - Women

Level of education. Relative risks			
Age interval	20 - 32	33 - 41	42 - 65
primary	0.401	1.252	0.838
lower sec.	1.004	0.694	0.958
upp. sec.	1.118	0.721	1.544
tertiary	2.221	1.595	0.806

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR4 (divorced -> second marriage) - Women

Children in the household. Relative risks			
Age interval	20 - 32	33 - 41	42 - 65
Without children	1.027	0.99	0.425
With children	0.974	1.01	2.352

Legend: Baseline risk, Without children, With children
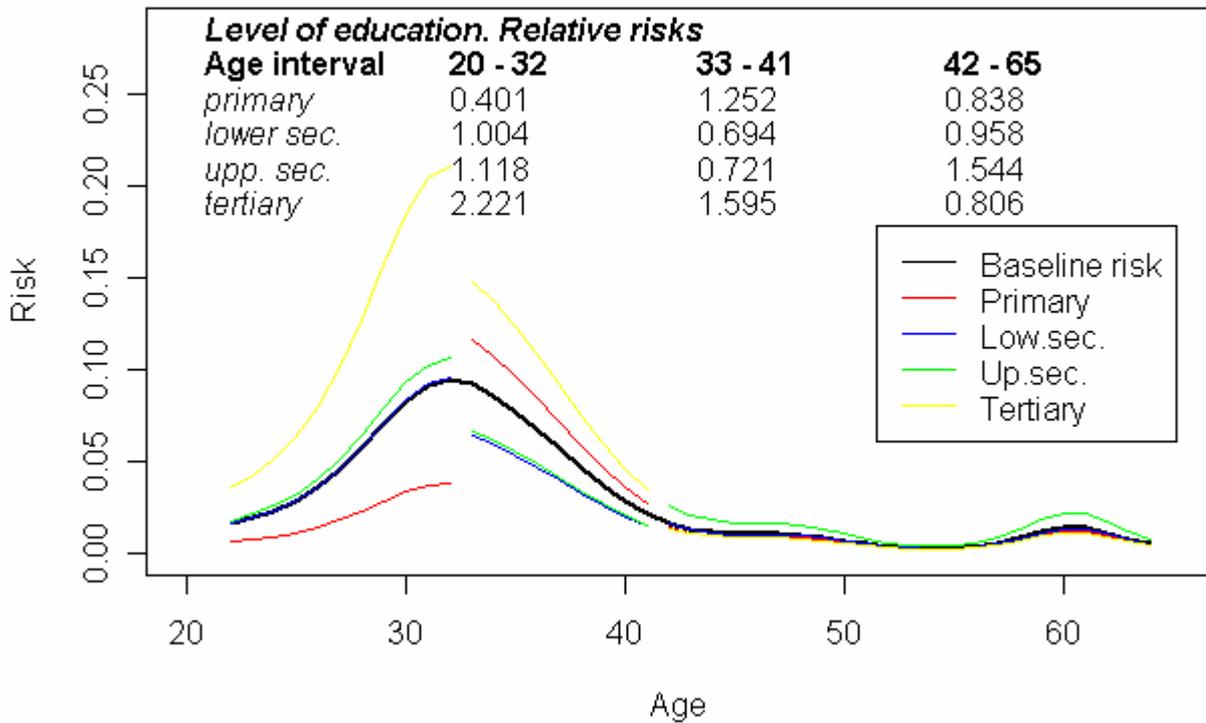
Italy. ISTAT 2003. TR4 Women. Events observed in the window of observation: 146

TR4 (divorced -> second marriage) - Men

Level of education. Relative risks			
Age interval	20 - 34	35 - 45	46 - 65
primary	0.295	0.75	1.094
lower sec.	1.21	0.614	1.118
upp. sec.	1.596	1.066	0.664
tertiary	1.754	2.037	1.231

Legend:
- Baseline risk
- Primary
- Low.sec.
- Up.sec.
- Tertiary

TR4 (divorced -> second marriage) - Men

Children in the household. Relative risks			
Age interval	20 - 34	35 - 45	46 - 65
Without children	0.98	0.704	0.684
With children	1.02	1.42	1.461

Legend:
- Baseline risk
- Without children
- With children
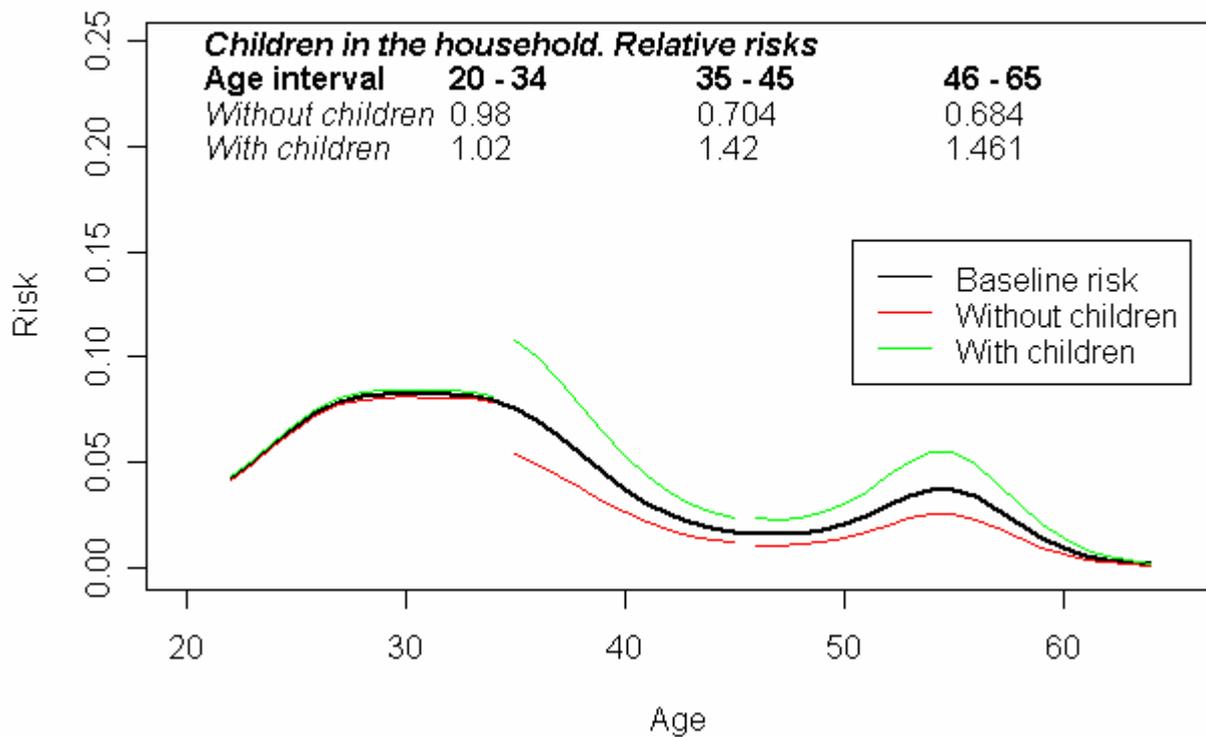
Italy. ISTAT 2003. TR4 Men. Events observed in the window of observation: 122
NOTE: Window of observation extended to the last 20 years before the interview.

29

TR5 (widowed -> second marriage) - Women

Level of education. Relative risks

Age interval	15 - 56	57 - 64	65 - 70
primary	1.306	0.913	4.156
lower sec.	0.444	0.997	1.187
upp. sec.	1.744	0.63	2.043
tertiary	0.99	1.744	0.099

Legend:
- Baseline risk
- Primary
- Low.sec.
- Up.sec.
- Tertiary

Risk (y-axis), Age (x-axis)

TR5 (widowed -> second marriage) - Women

Children in the household. Relative risks

Age interval	15 - 56	57 - 64	65 - 70
Without children	0.775	0.242	0.678
With children	1.291	4.135	1.474

Legend:
- Baseline risk
- Without children
- With children

Risk (y-axis), Age (x-axis)

Italy. ISTAT 2003. TR5 Women. Events observed in the window of observation: 138
NOTE: Window of observation extended to the last 20 years before the interview.

TR5 (widowed -> second marriage) - Men

Level of education. Relative risks

Age interval	15 - 46	47 - 58	59 - 70
primary	0.366	1.992	1.094
lower sec.	0.621	2.235	1.84
upp. sec.	2.558	0.34	0.105
tertiary	1.719	0.661	4.739

Baseline risk
Primary
Low.sec.
Up.sec.
Tertiary

TR5 (widowed -> second marriage) - Men

Children in the household. Relative risks

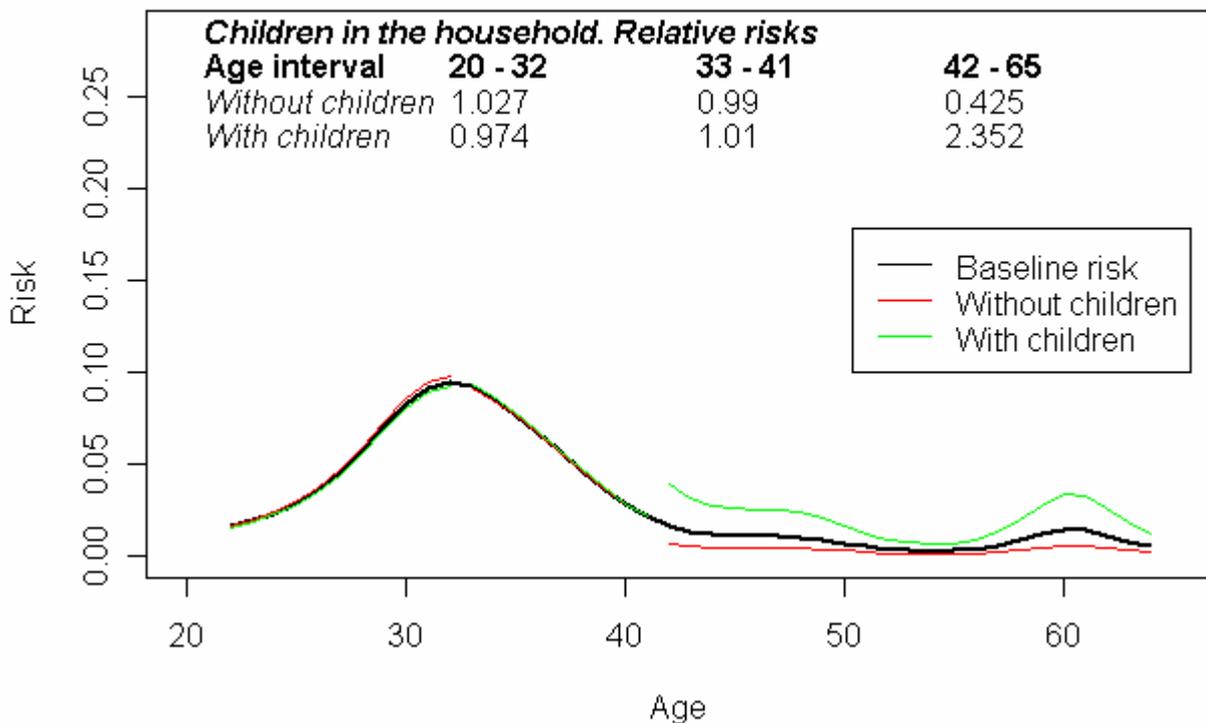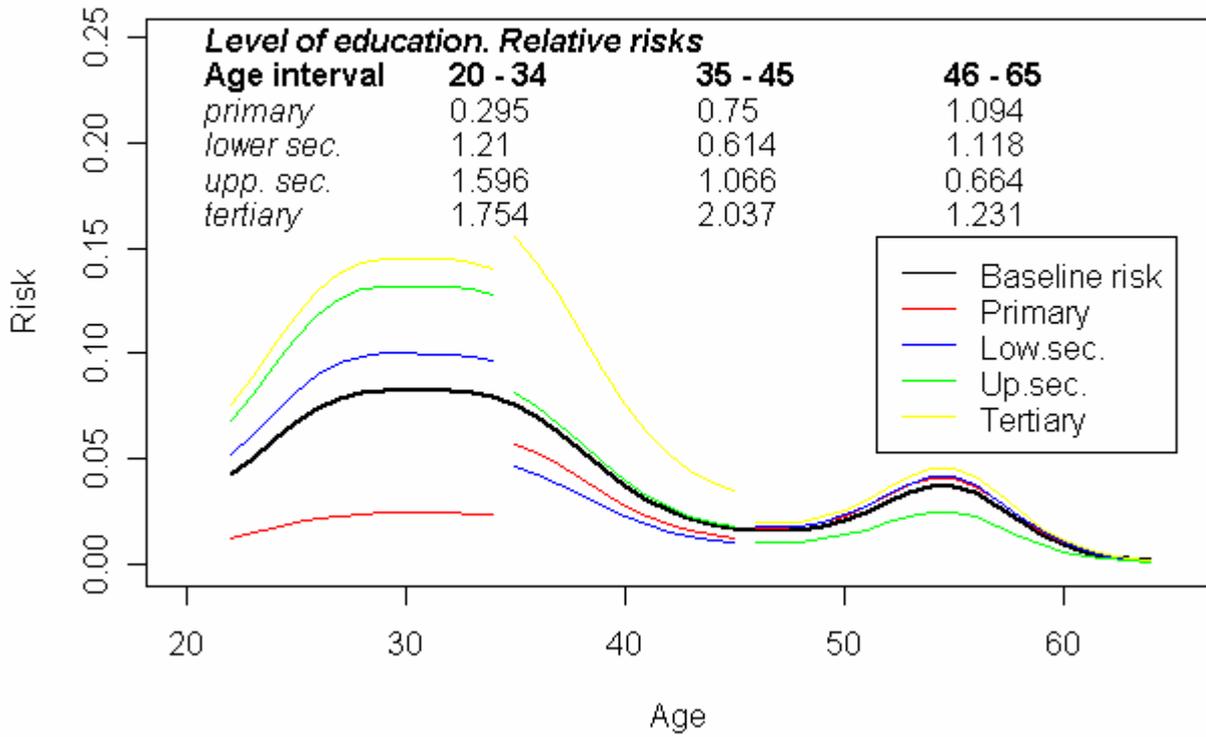| Age interval	15 - 46	47 - 58	59 - 70
Without children	1.004	0.719	0.361
With children	0.996	1.39	2.772

Baseline risk
Without children
With children
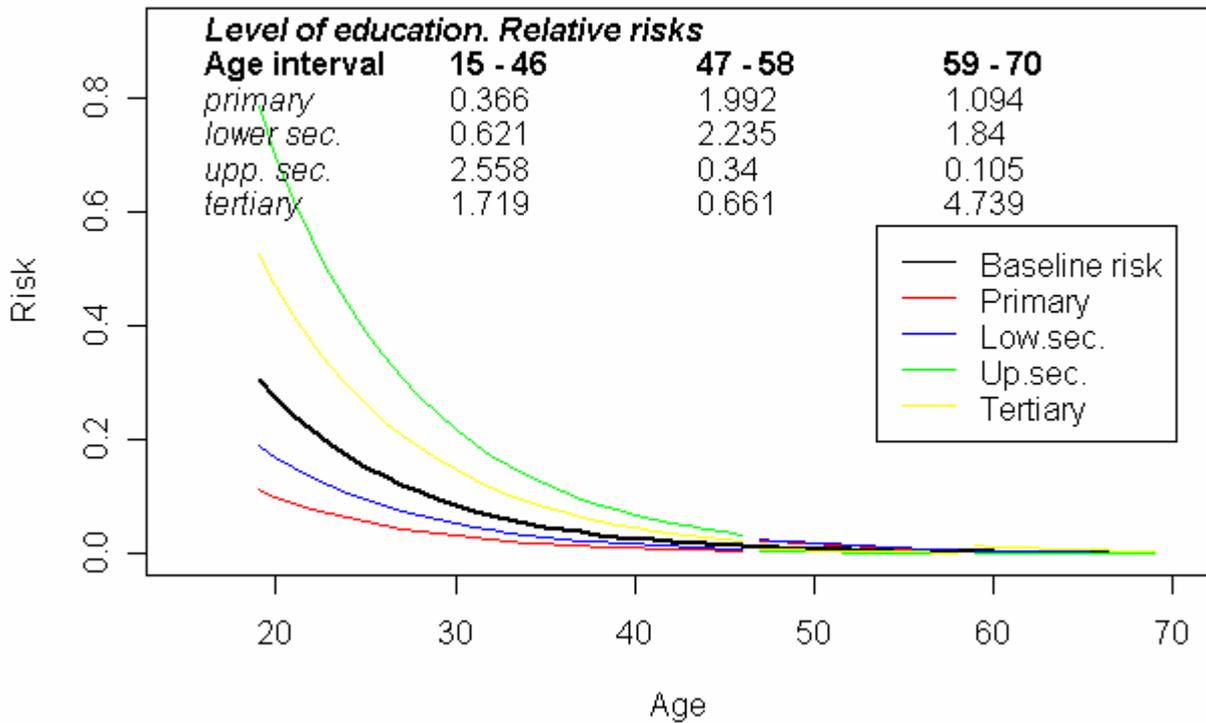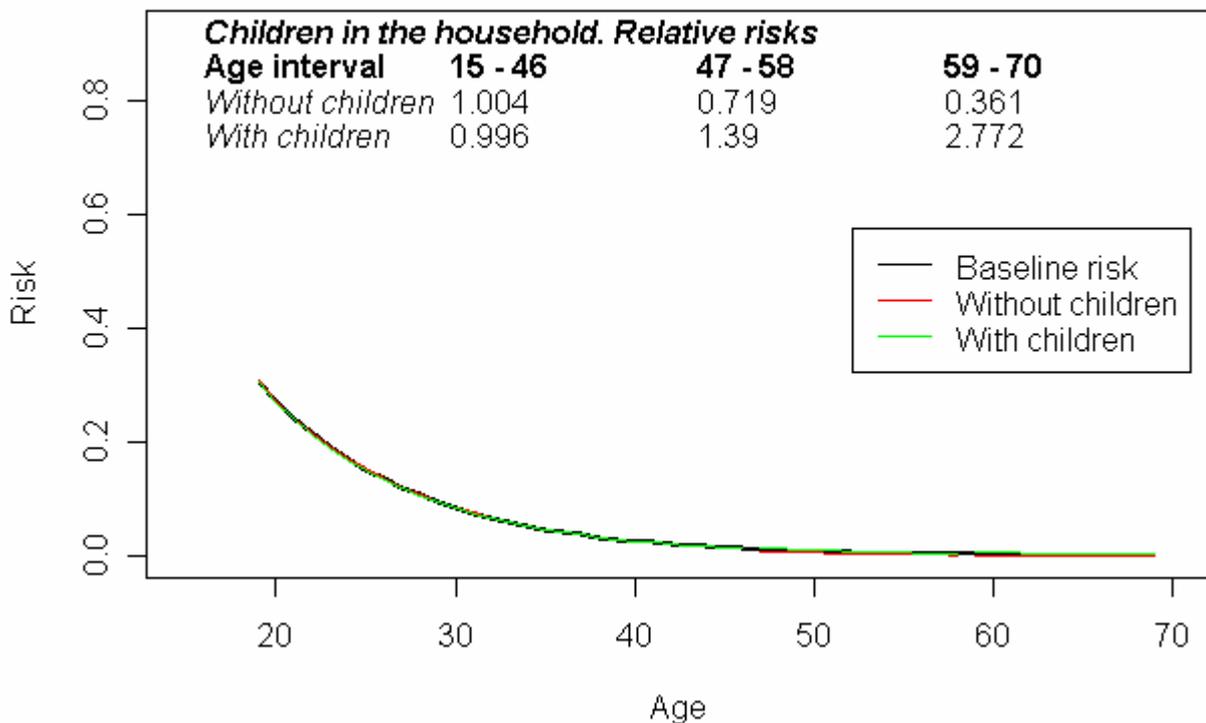
Italy. ISTAT 2003. TR5 Men. Events observed in the window of observation: 43
NOTE: Window of observation extended to the last 20 years before the interview.

TR6 (in parental home -> with a partner) - Women

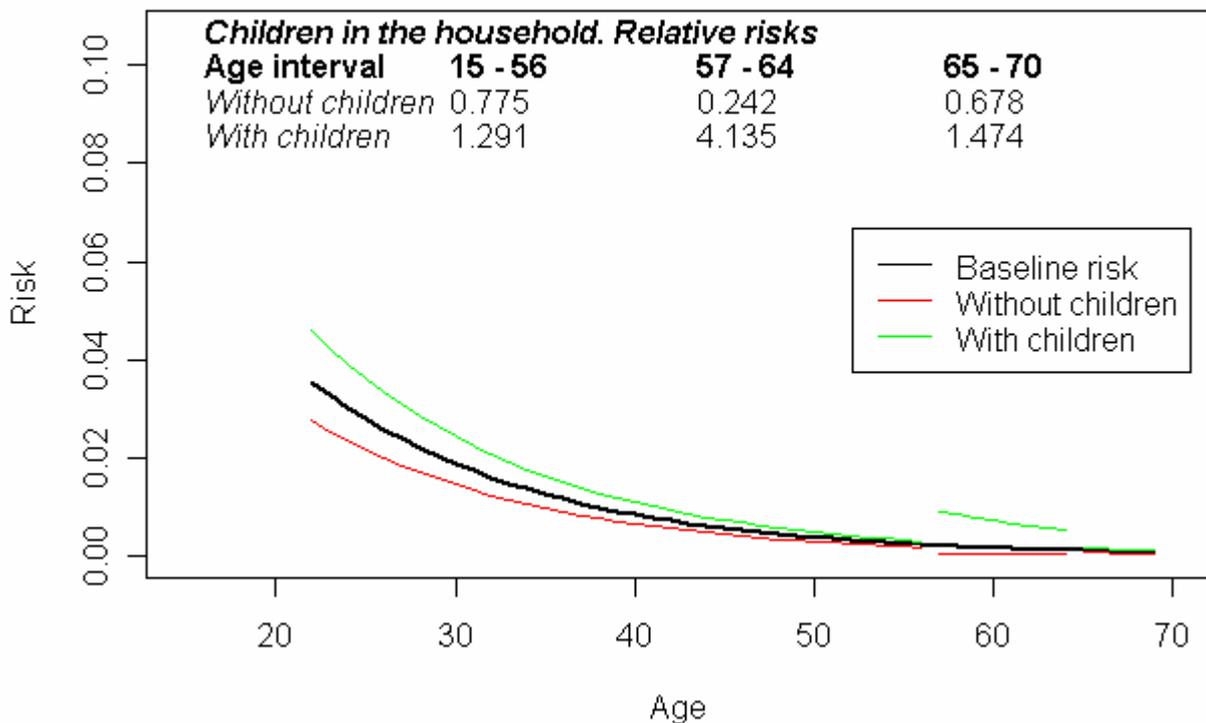| Level of education. Relative risks			
Age interval	15 - 23	24 - 27	28 - 60
primary	2.309	1.457	0.914
lower sec.	2.297	1.654	0.81
upp. sec.	0.573	0.817	1.246
tertiary	0.329	0.508	1.084

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR6 (in parental home -> with a partner) - Men

Level of education. Relative risks			
Age interval	15 - 26	27 - 30	31 - 60
primary	3.943	1.17	0.767
lower sec.	1.753	1.247	0.782
upp. sec.	0.489	1.13	1.208
tertiary	0.296	0.607	1.381

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary
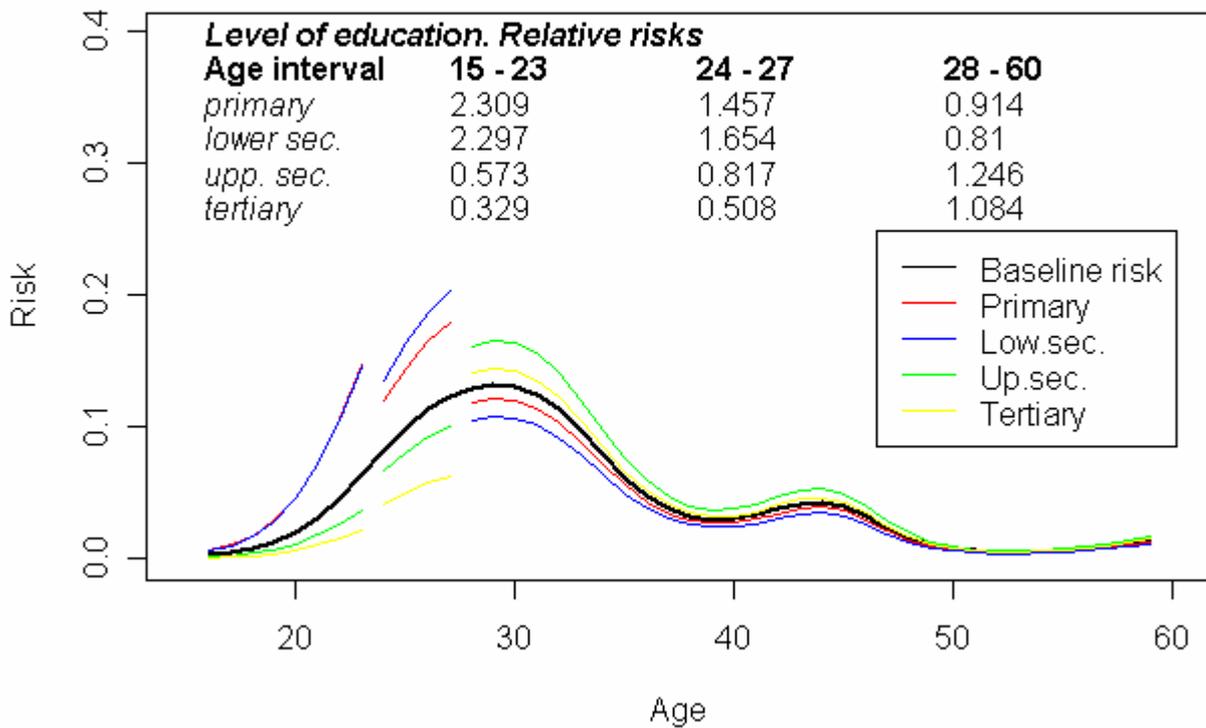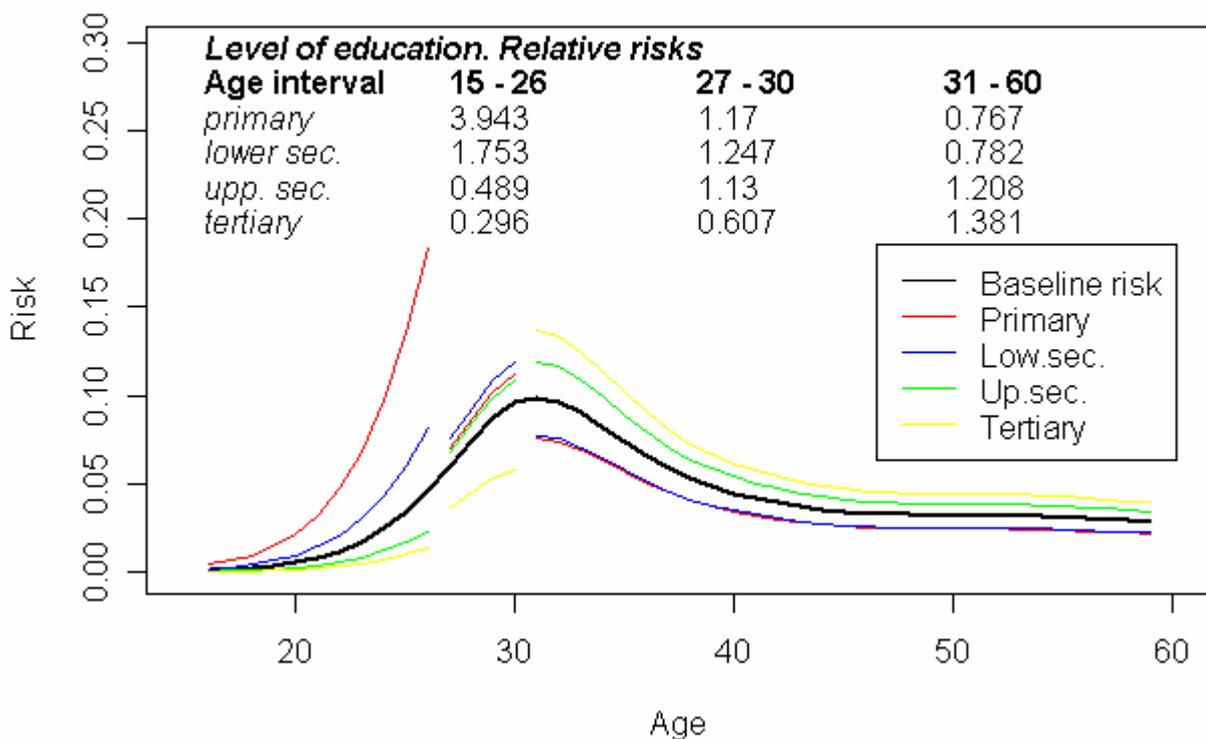
Italy. ISTAT 2003. TR6. Events observed in the window of observation: Women1131; Men 972.

TR7 (in parental home -> alone or with other persons) - Women

Level of education. Relative risks			
Age interval	10 - 19	20 - 25	26 - 50
primary	0.45	1.149	0.711
lower sec.	0.936	0.784	0.727
upp. sec.	1.557	0.925	1.109
tertiary	1.526	1.2	1.745

Baseline risk — Primary — Low.sec. — Up.sec. — Tertiary

TR7 (in parental home -> alone or with other persons) - Men

Level of education. Relative risks			
Age interval	10 - 21	22 - 27	28 - 50
primary	1.025	0.542	0.759
lower sec.	1.192	1.143	0.712
upp. sec.	1.759	1.125	1.342
tertiary	0.465	1.435	1.379

Baseline risk — Primary — Low.sec. — Up.sec. — Tertiary
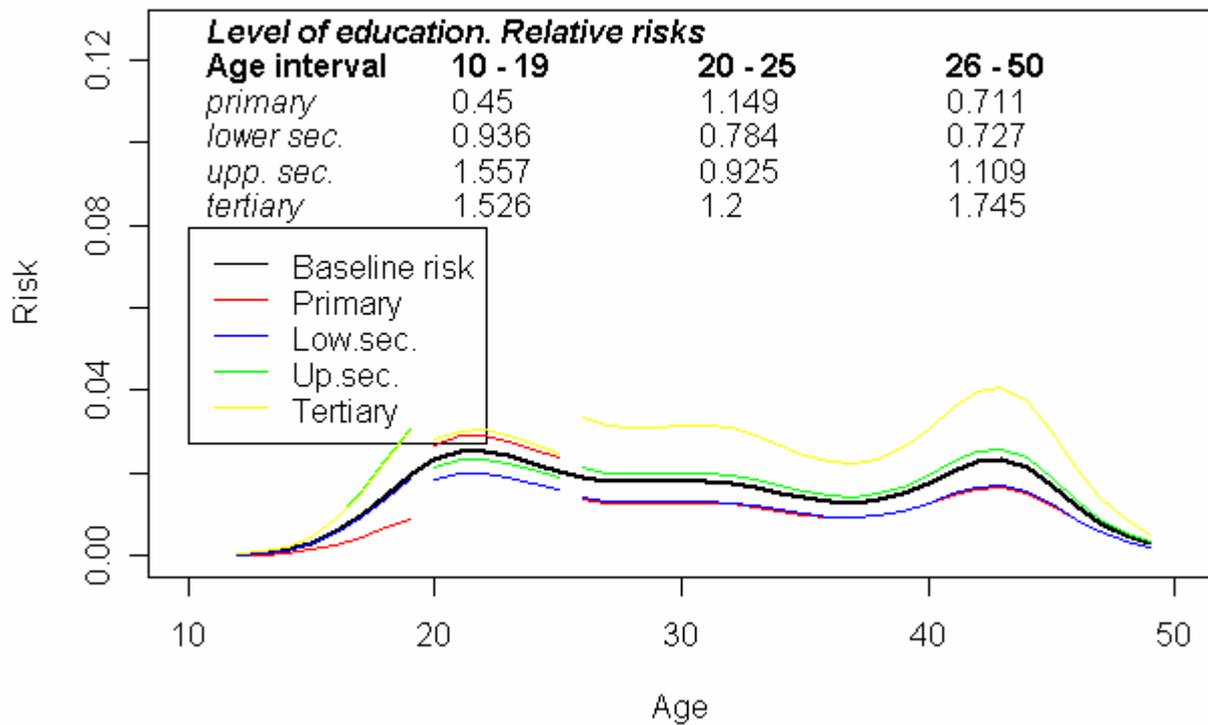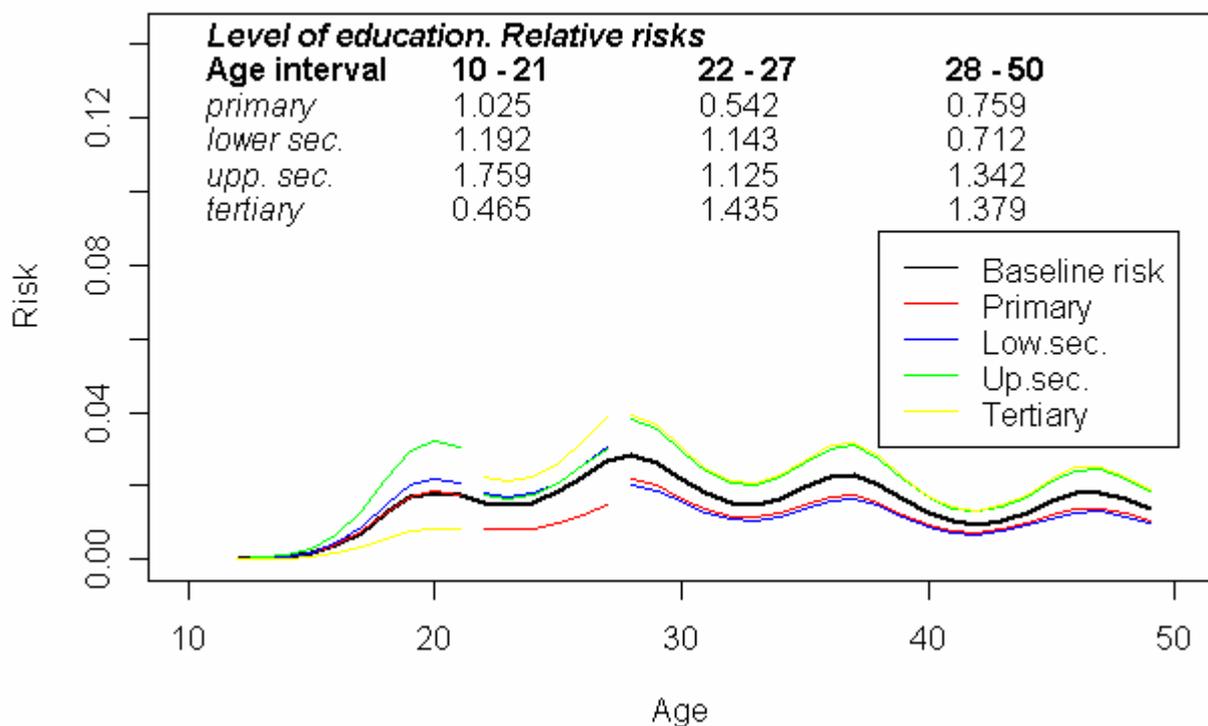
Italy. ISTAT 2003. TR7. Events observed in the window of observation: Women 391; Men 514.

33

TR12 (first birth) - Women

Level of education. Relative risks			
Age interval	**15 - 26**	**27 - 30**	**31 - 50**
primary	3.378	0.782	0.793
lower sec.	1.64	1.404	1.045
upp. sec.	0.952	1.233	1.008
tertiary	0.19	0.738	1.199

Legend:
- Baseline risk
- Primary
- Low.sec.
- Up.sec.
- Tertiary

TR12 (first birth) - Women

Married or not. Relative risks			
Age interval	**15 - 26**	**27 - 30**	**31 - 50**
Not married	0.186	0.221	0.317
Married	5.38	4.522	3.158

Legend:
- Baseline risk
- Not married
- Married
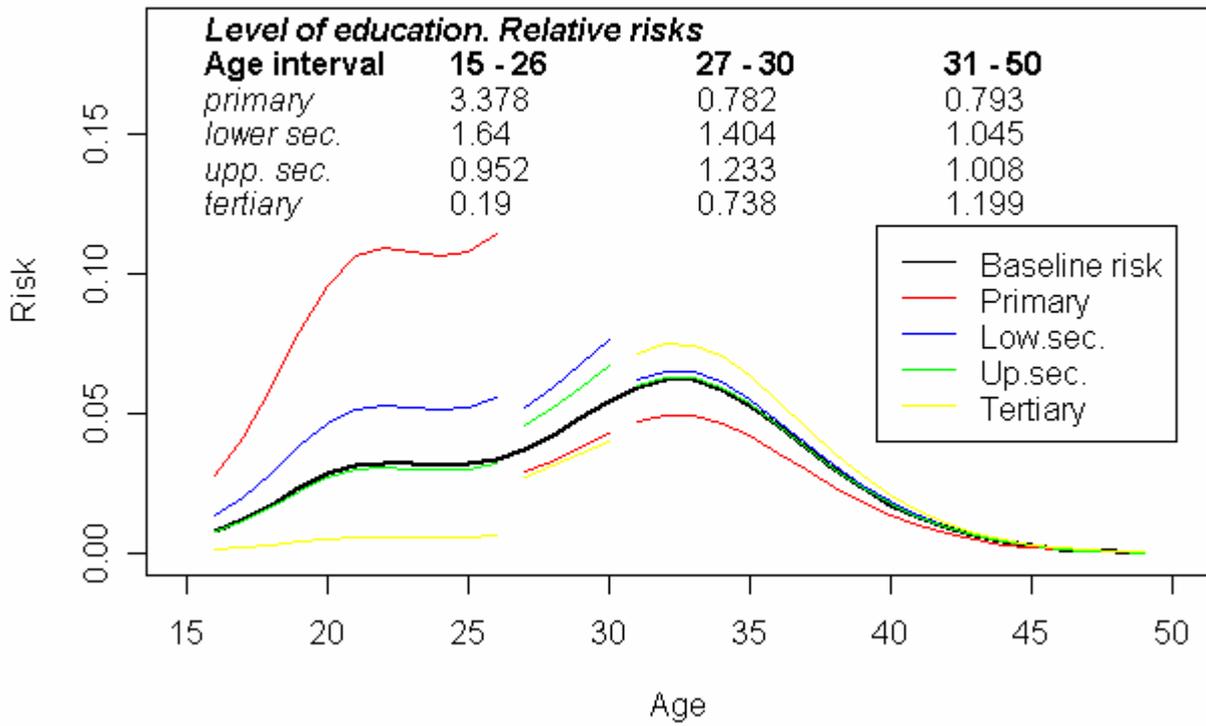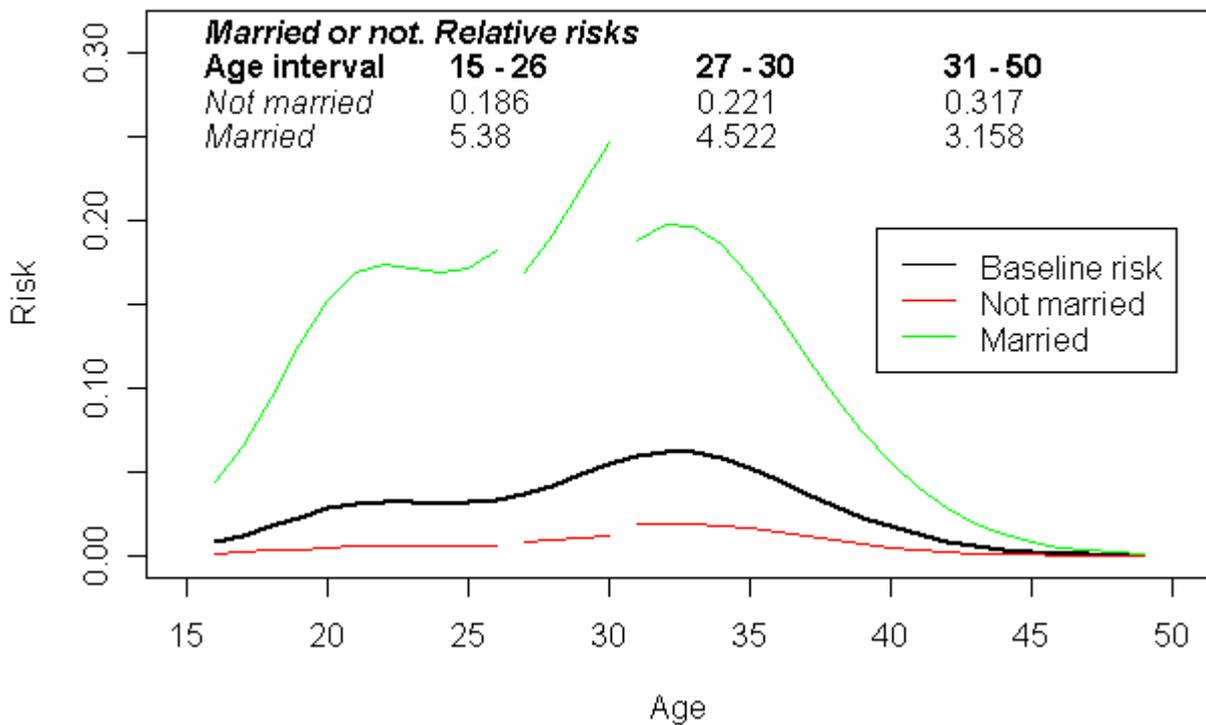
Italy. ISTAT 2003. TR12 Women. Events observed in the window of observation: 1322.

TR12 (first birth) - Men

Level of education. Relative risks			
Age interval	15 - 29	30 - 33	34 - 50
primary	2.204	1.271	0.647
lower sec.	1.245	1.005	0.879
upp. sec.	0.692	0.96	1.187
tertiary	0.526	0.815	1.481

Legend:
- Baseline risk
- Primary
- Low.sec.
- Up.sec.
- Tertiary

TR12 (first birth) - Men

Married or not. Relative risks			
Age interval	15 - 29	30 - 33	34 - 50
Not married	0.151	0.193	0.225
Married	6.635	5.174	4.443

Legend:
- Baseline risk
- Not married
- Married
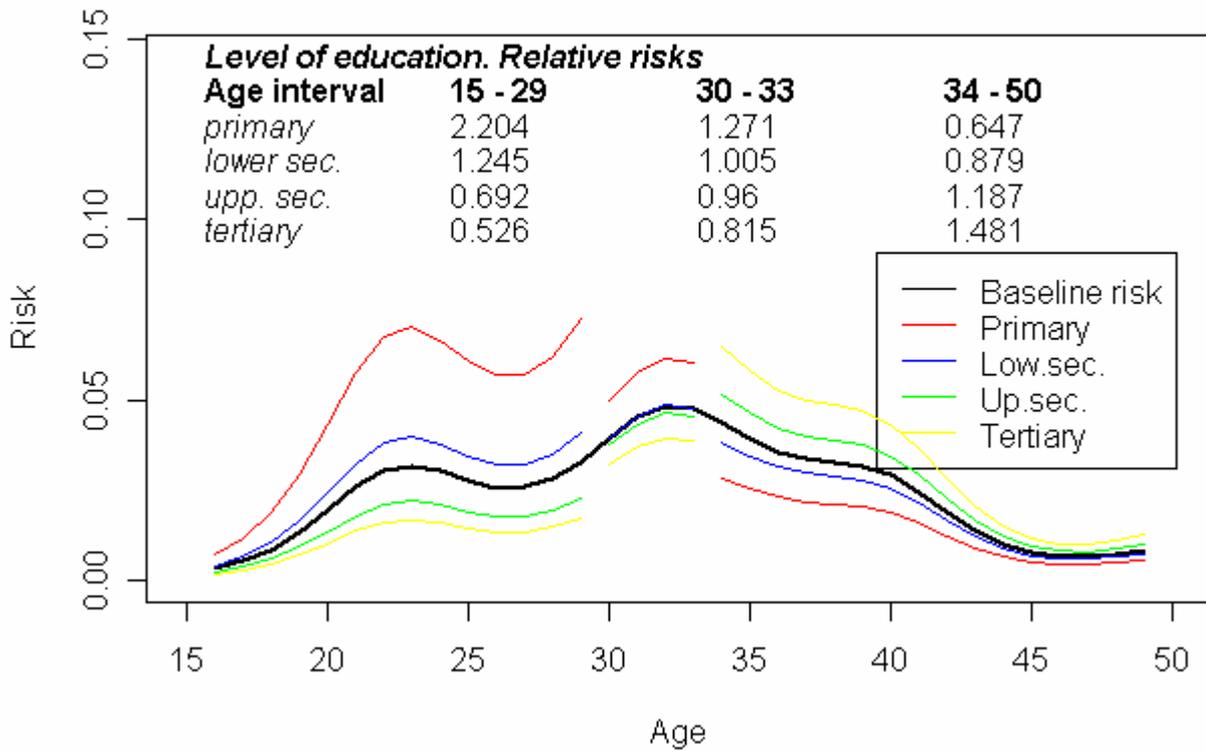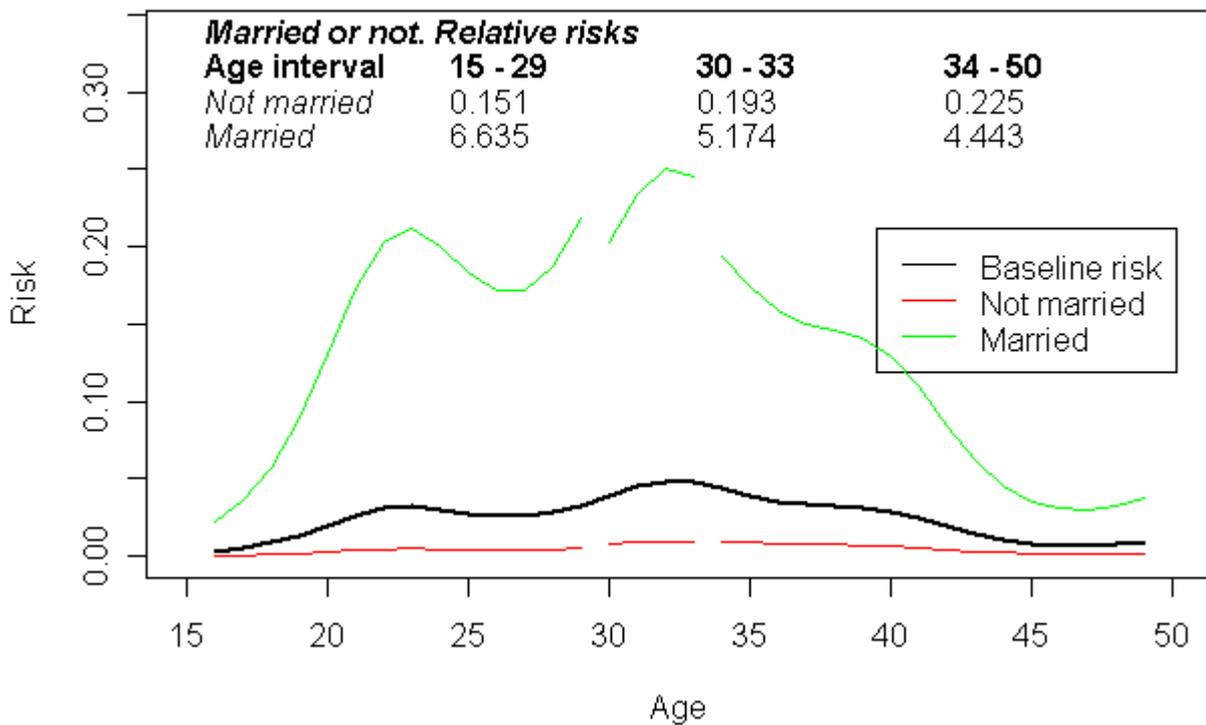
Italy. ISTAT 2003. TR12 Men. Events observed in the window of observation: 1257.

TR13 (second birth) - Women

Level of education. Relative risks			
Age interval	**15 - 28**	**29 - 32**	**33 - 50**
primary	1.133	1.098	0.666
lower sec.	0.901	1.037	0.872
upp. sec.	0.909	0.853	1.033
tertiary	1.077	1.03	1.666

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR13 (second birth) - Women

Married or not. Relative risks			
Age interval	**15 - 28**	**29 - 32**	**33 - 50**
Not married	0.669	0.492	0.519
Married	1.495	2.032	1.927

Legend: Baseline risk, Not married, Married
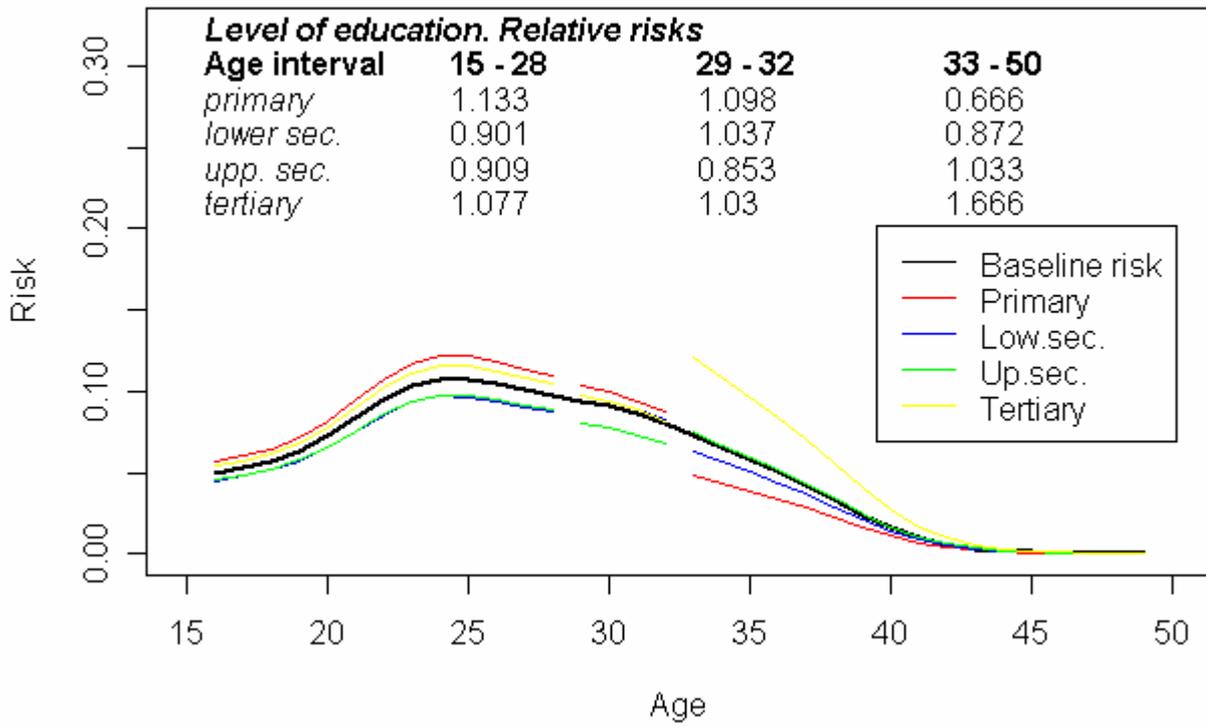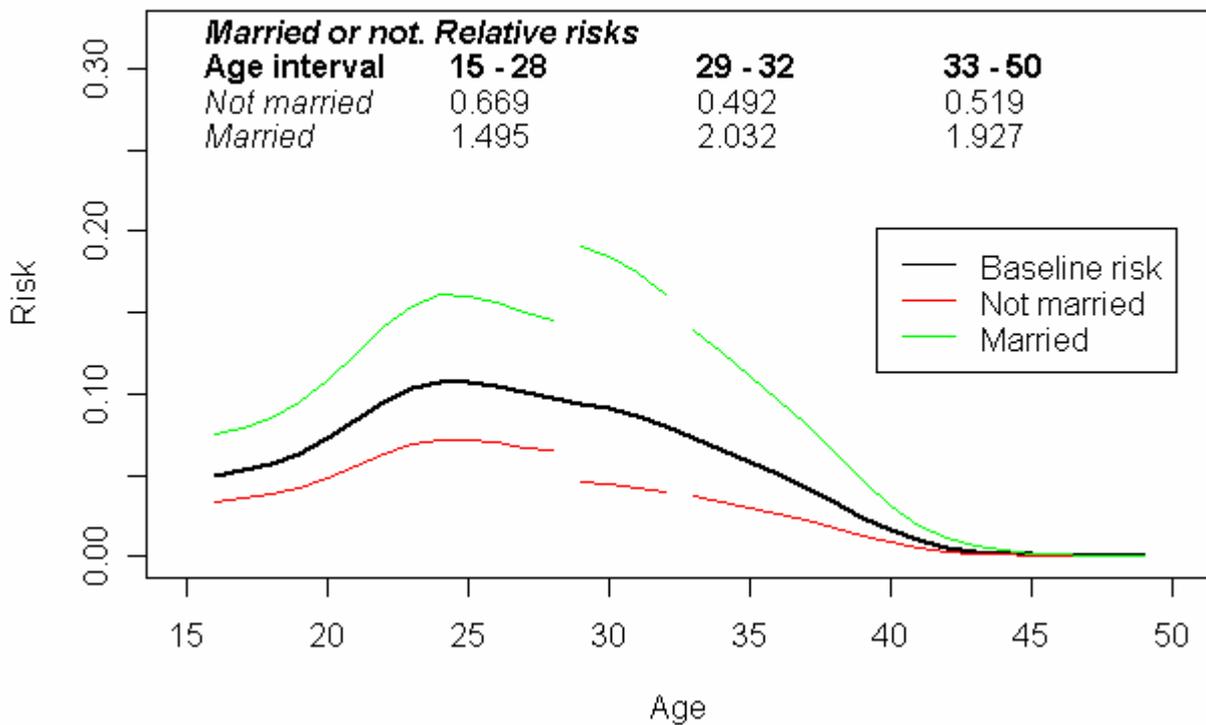
Italy. ISTAT 2003. TR13 Women. Events observed in the window of observation: 1013.

TR13 (second birth) - Men

Level of education. Relative risks

Age interval	15 - 31	32 - 35	36 - 50
primary	0.844	1.226	0.619
lower sec.	0.932	0.91	0.892
upp. sec.	0.89	0.967	1.089
tertiary	1.43	0.927	1.664

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR13 (second birth) - Men

Married or not. Relative risks

Age interval	15 - 31	32 - 35	36 - 50
Not married	0.631	0.653	0.583
Married	1.586	1.531	1.715

Legend: Baseline risk, Not married, Married
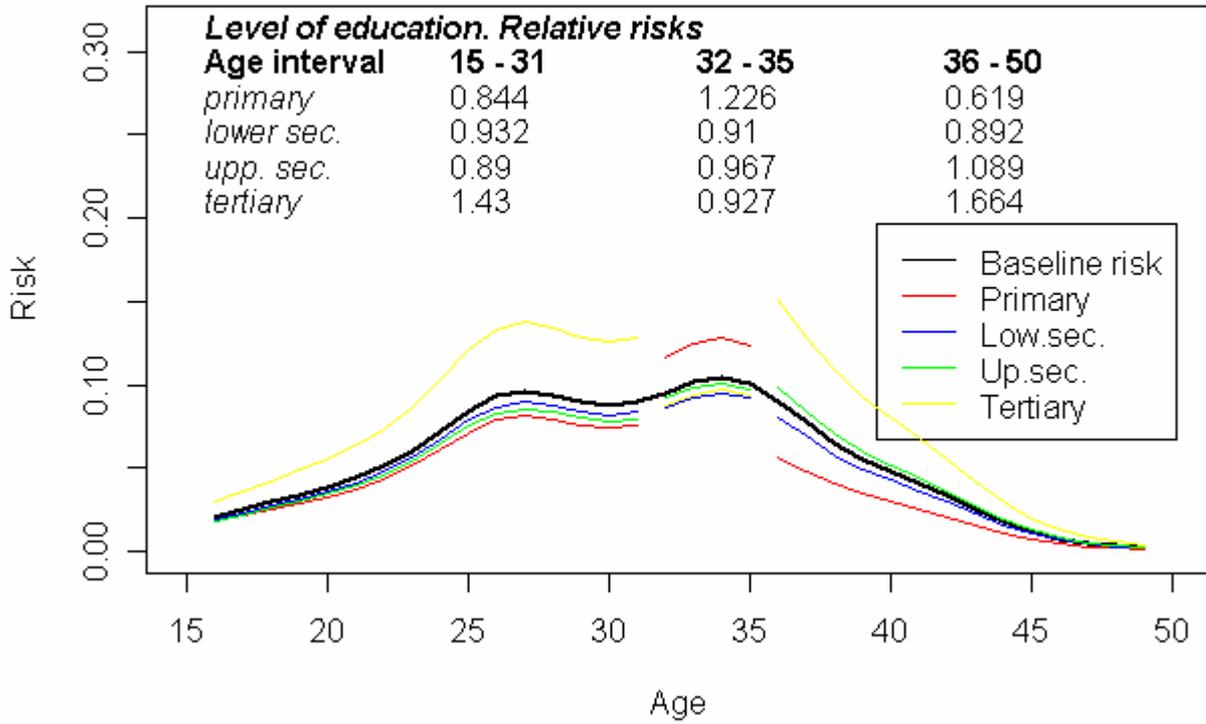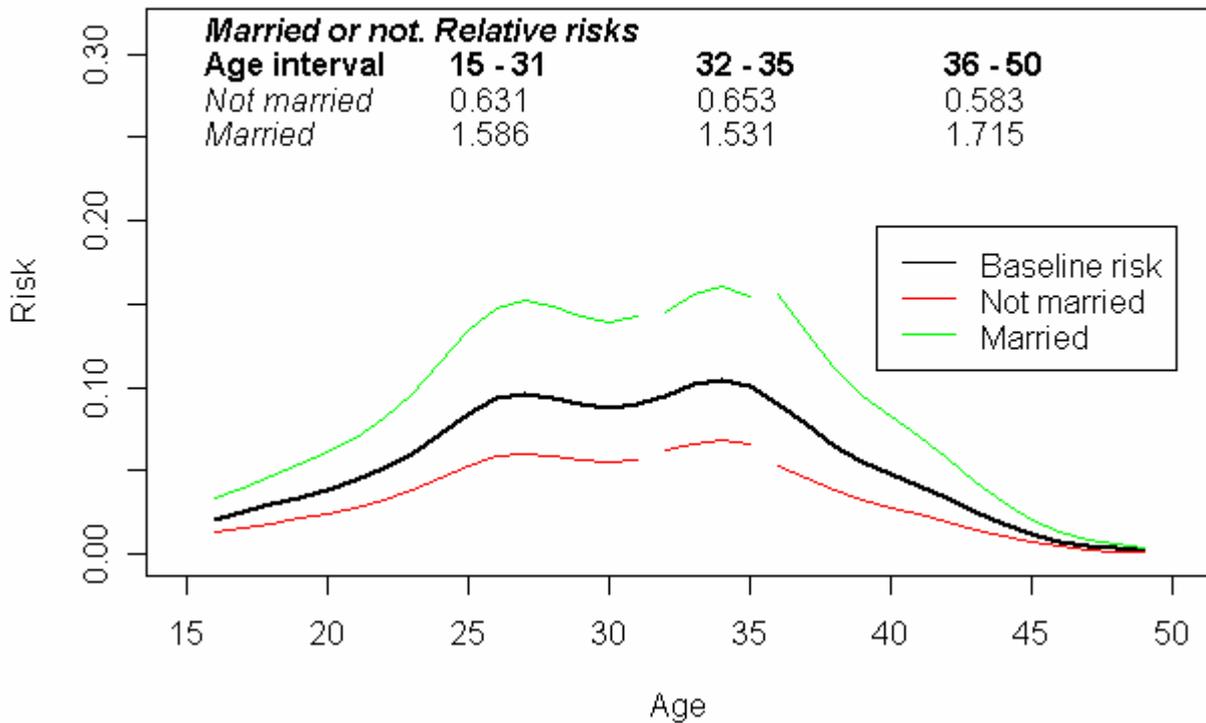
Italy. ISTAT 2003.TR13 Men. Events observed in the window of observation: 981.

TR14 (third birth) - Women

Level of education. Relative risks

Age interval	15 - 31	32 - 35	36 - 50
primary	1.078	1.052	0.632
lower sec.	0.698	0.768	0.882
upp. sec.	0.627	0.737	0.922
tertiary	2.12	1.678	1.945

Legend:
— Baseline risk
— Primary
— Low.sec.
— Up.sec.
— Tertiary

TR14 (third birth) - Women

Married or not. Relative risks

Age interval	15 - 31	32 - 35	36 - 50
Not married	1.471	1.322	1.162
Married	0.68	0.757	0.861

Legend:
— Baseline risk
— Not married
— Married
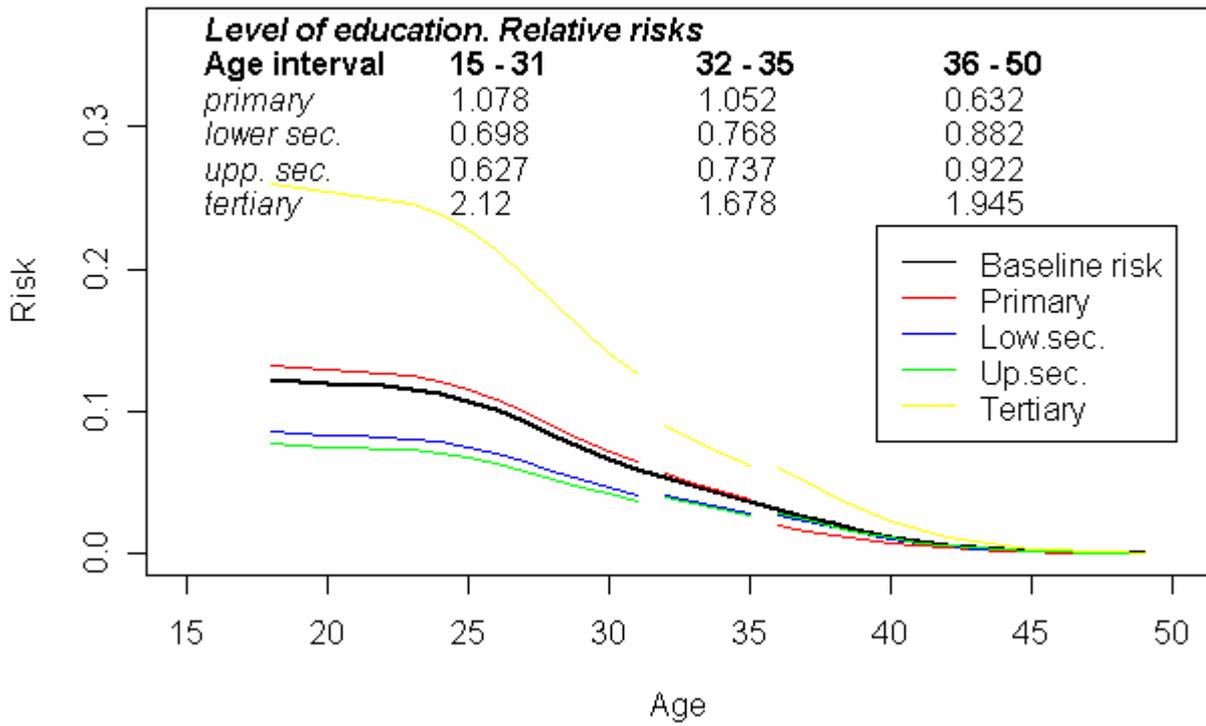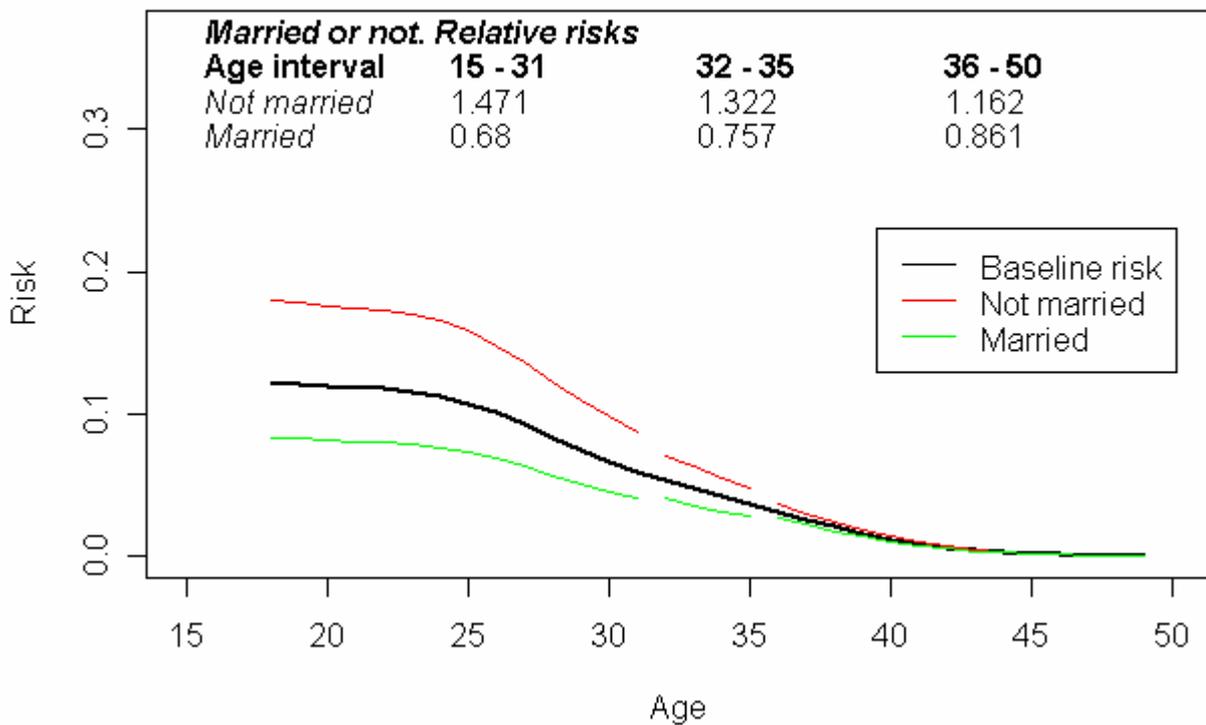
Italy. ISTAT 2003. TR14 Women. Events observed in the window of observation: Women 270

38

TR14 (third birth) - Men

Level of education. Relative risks

Age interval	15 - 35	36 - 39	40 - 50
primary	0.702	0.858	1.043
lower sec.	1.025	0.631	0.922
upp. sec.	0.524	0.824	0.596
tertiary	2.653	2.243	1.745

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR14 (third birth) - Men

Married or not. Relative risks

Age interval	15 - 35	36 - 39	40 - 50
Not married	1.463	1.071	1.177
Married	0.684	0.933	0.85

Legend: Baseline risk, Not married, Married
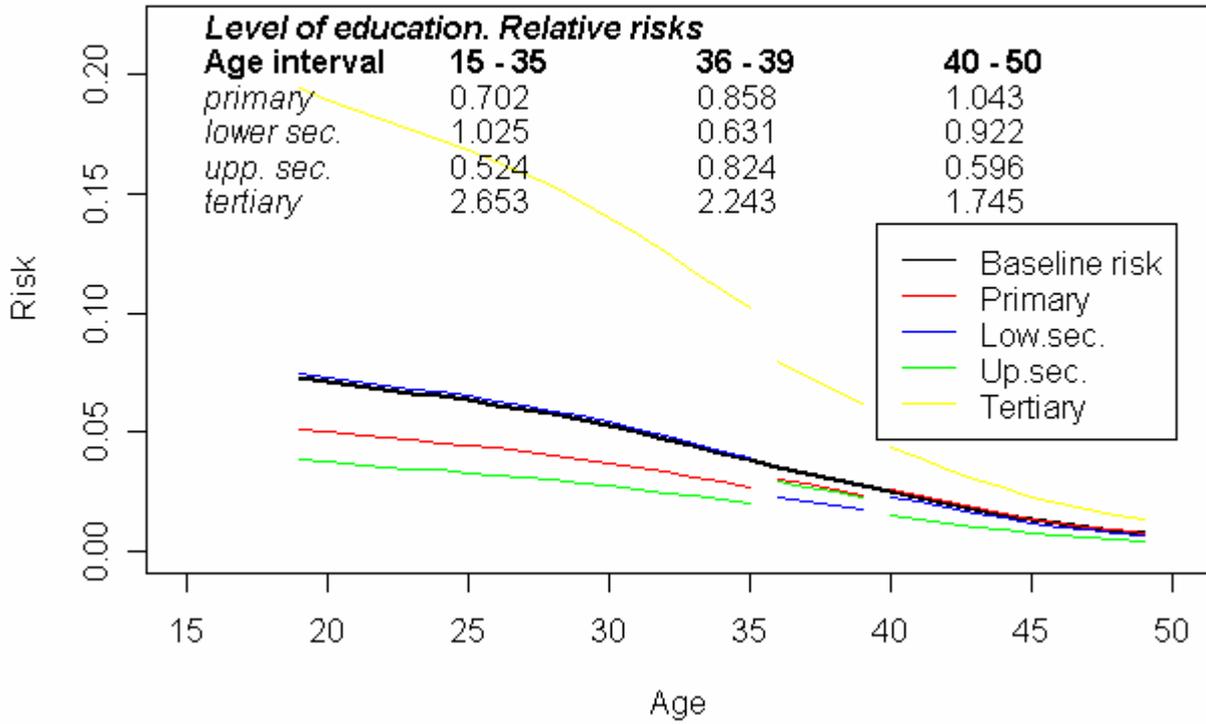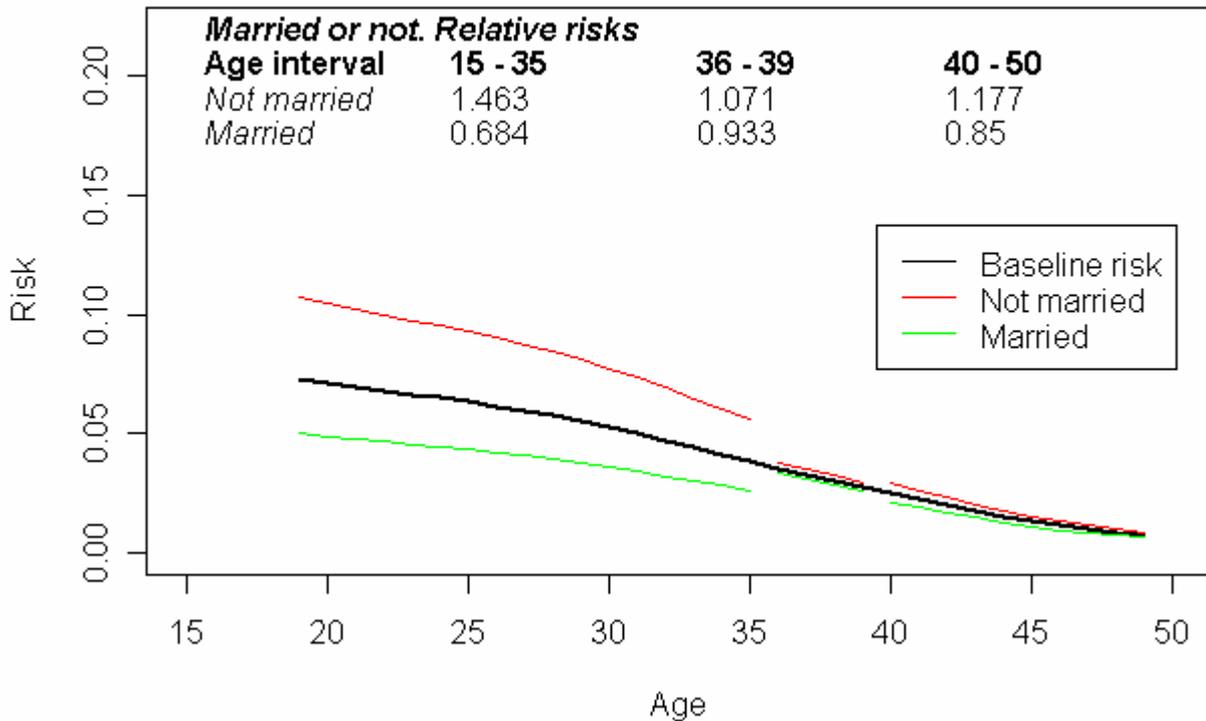
Italy. ISTAT 2003. TR14 Men. Events observed in the window of observation: 248.

39

TR15(fourth birth) - Women

Level of education. Relative risks			
Age interval	20 - 30	31 - 34	35 - 50
primary	0.823	1.447	1.508
lower sec.	0.873	0.59	0.737
upp. sec.	0.632	1.216	0.82
tertiary	2.203	0.963	1.097

Legend:
- Baseline risk
- Primary
- Low.sec.
- Up.sec.
- Tertiary

TR15(fourth birth) - Women

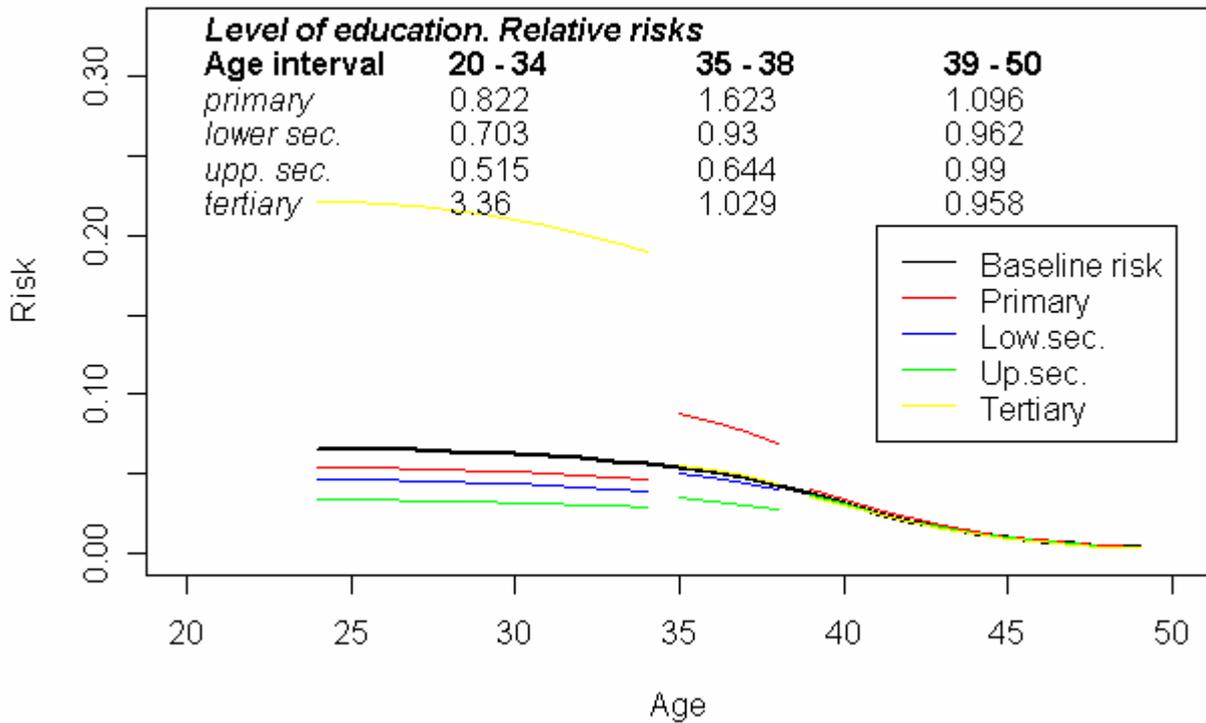| Married or not. Relative risks			
Age interval	20 - 30	31 - 34	35 - 50
Not married	0.536	0.681	0.726
Married	1.864	1.468	1.378

Legend:
- Baseline risk
- Not married
- Married
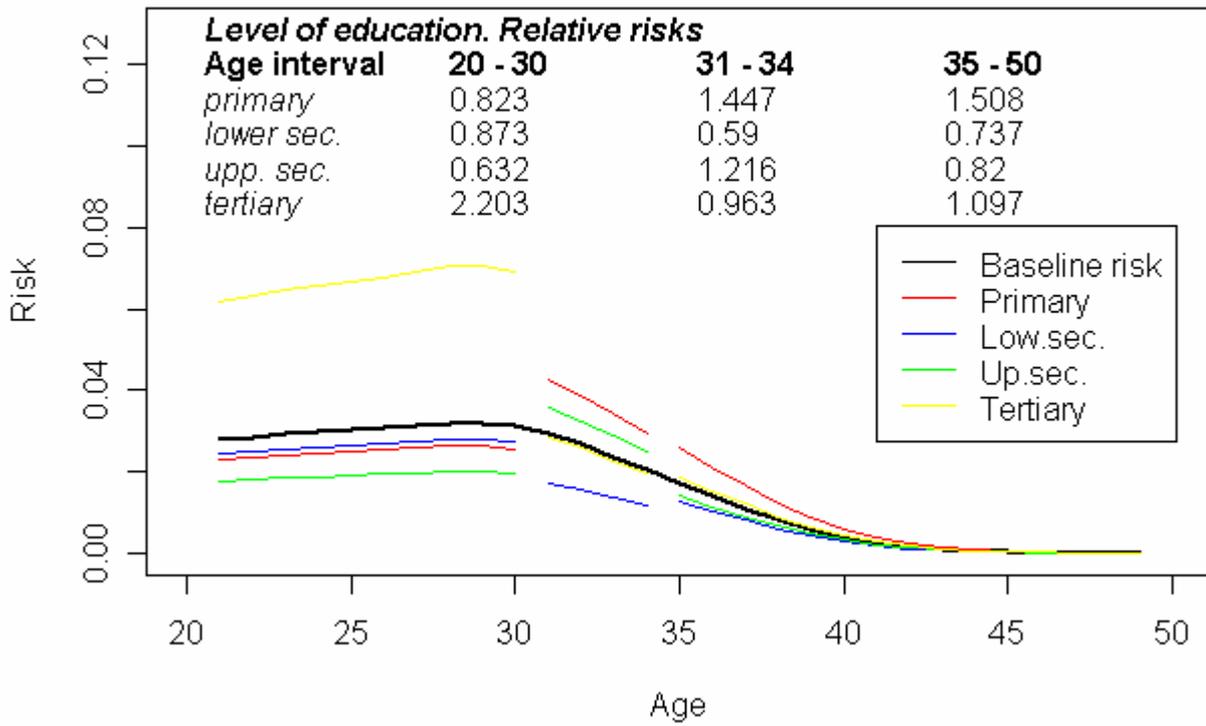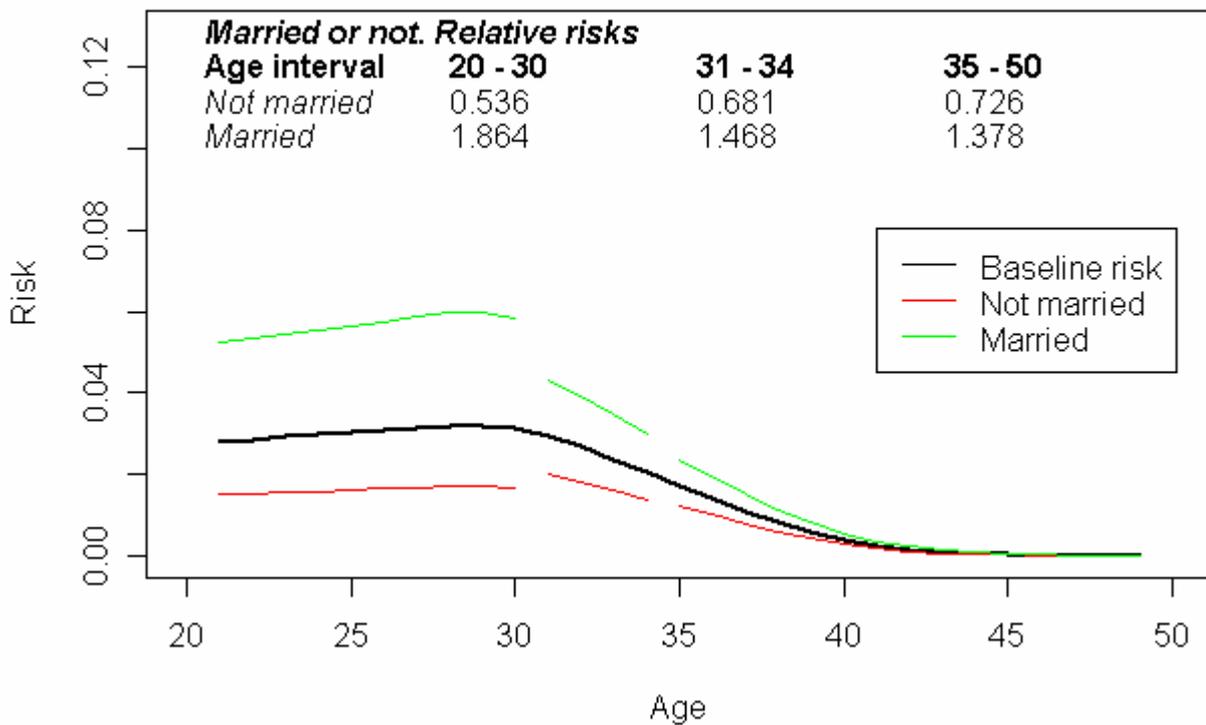
Italy. ISTAT 2003. TR15 Women. Events observed in the window of observation: 211.
NOTE: Window of observation extended to the last 10 years before the interview.

TR15(fourth birth) - Men

Level of education. Relative risks			
Age interval	20 - 34	35 - 38	39 - 50
primary	0.822	1.623	1.096
lower sec.	0.703	0.93	0.962
upp. sec.	0.515	0.644	0.99
tertiary	3.36	1.029	0.958

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR15(fourth birth) - Men

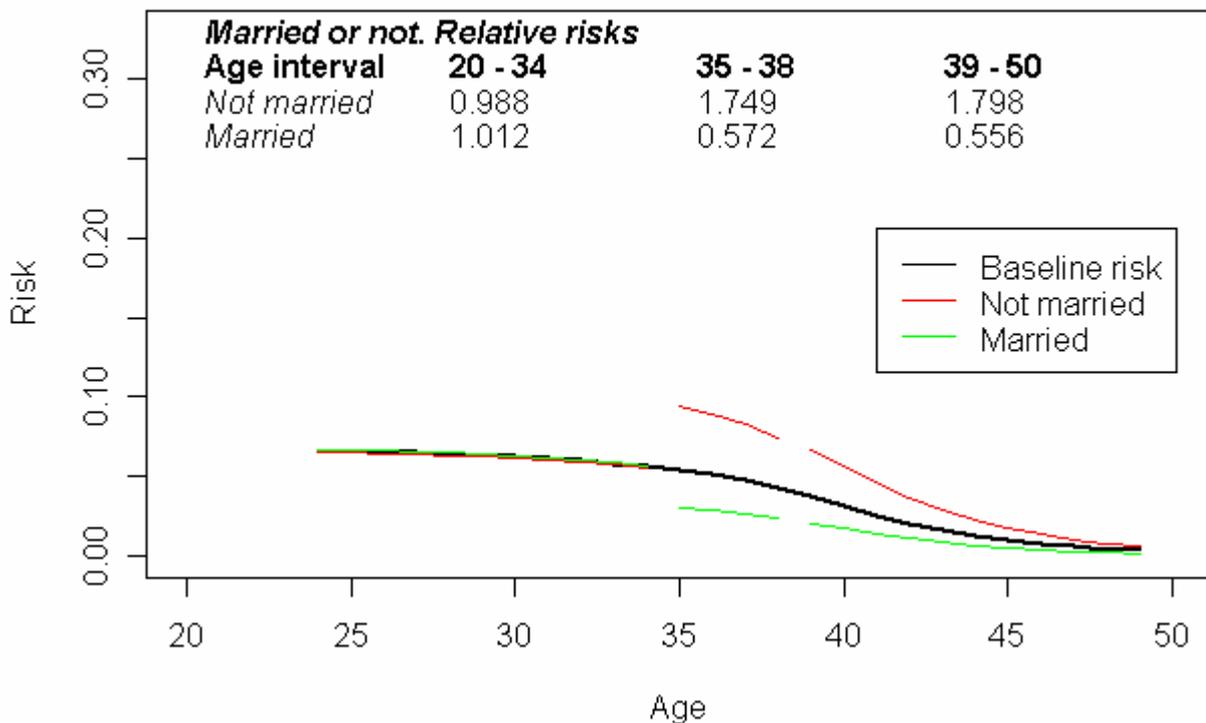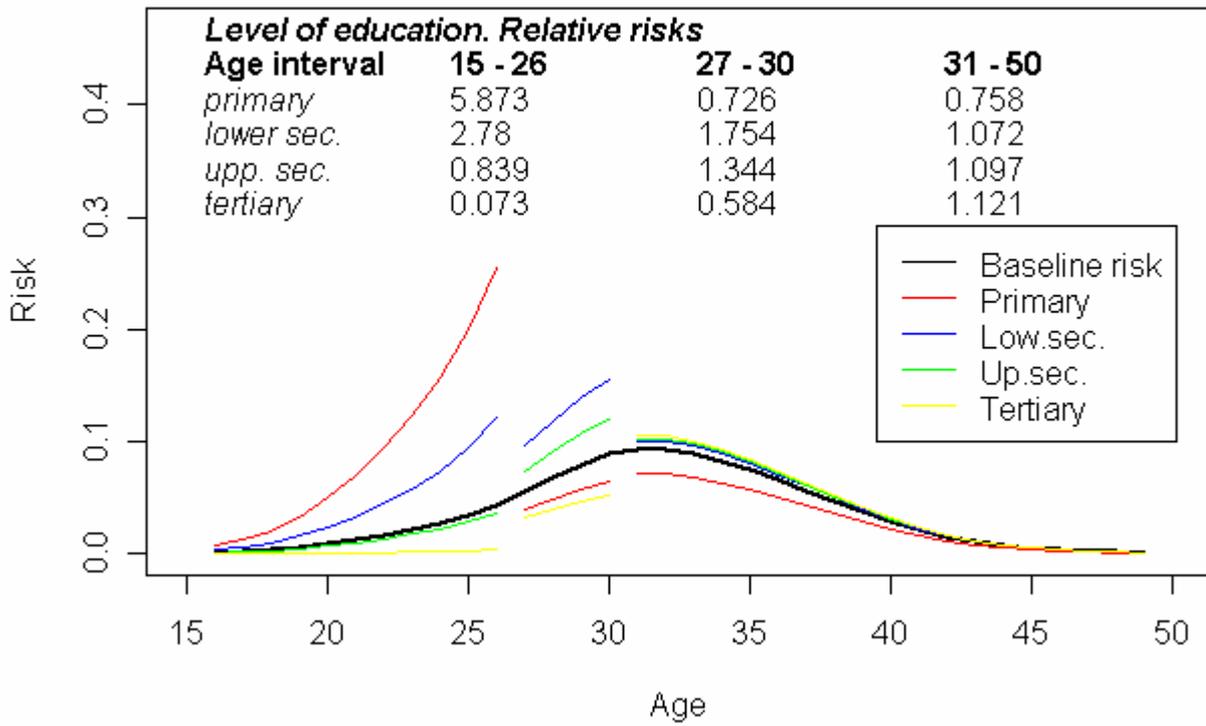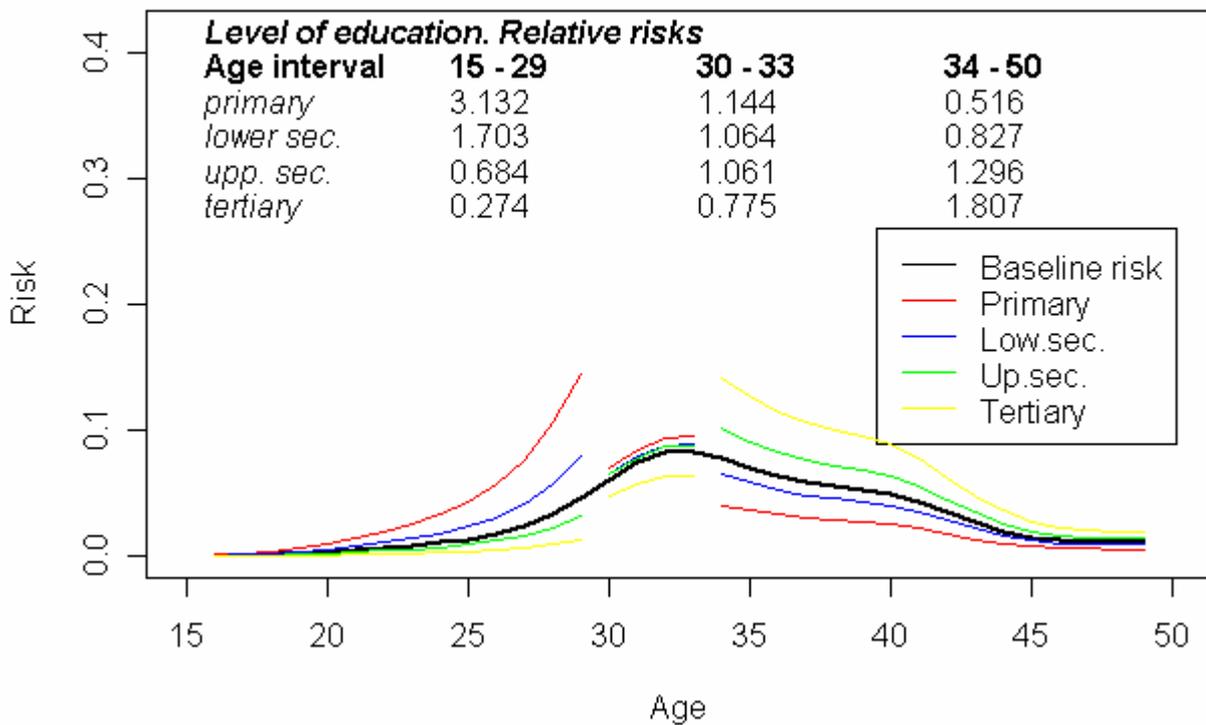| Married or not. Relative risks			
Age interval	20 - 34	35 - 38	39 - 50
Not married	0.988	1.749	1.798
Married	1.012	0.572	0.556

Legend: Baseline risk, Not married, Married

Italy. ISTAT 2003. TR15 Men. Events observed in the window of observation: 203
NOTE: Window of observation extended to the last 20 years before the interview.
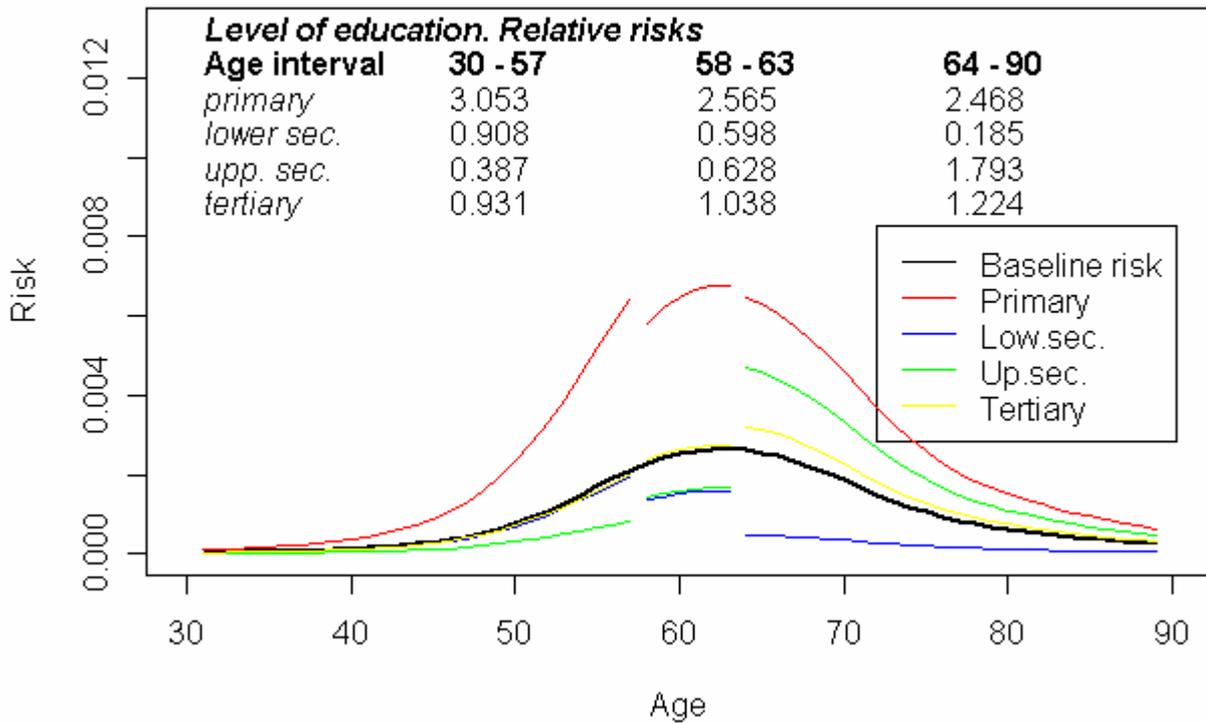
TR16(without children -> with children) - Women

Level of education. Relative risks			
Age interval	15 - 26	27 - 30	31 - 50
primary	5.873	0.726	0.758
lower sec.	2.78	1.754	1.072
upp. sec.	0.839	1.344	1.097
tertiary	0.073	0.584	1.121

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR16(without children -> with children) - Men

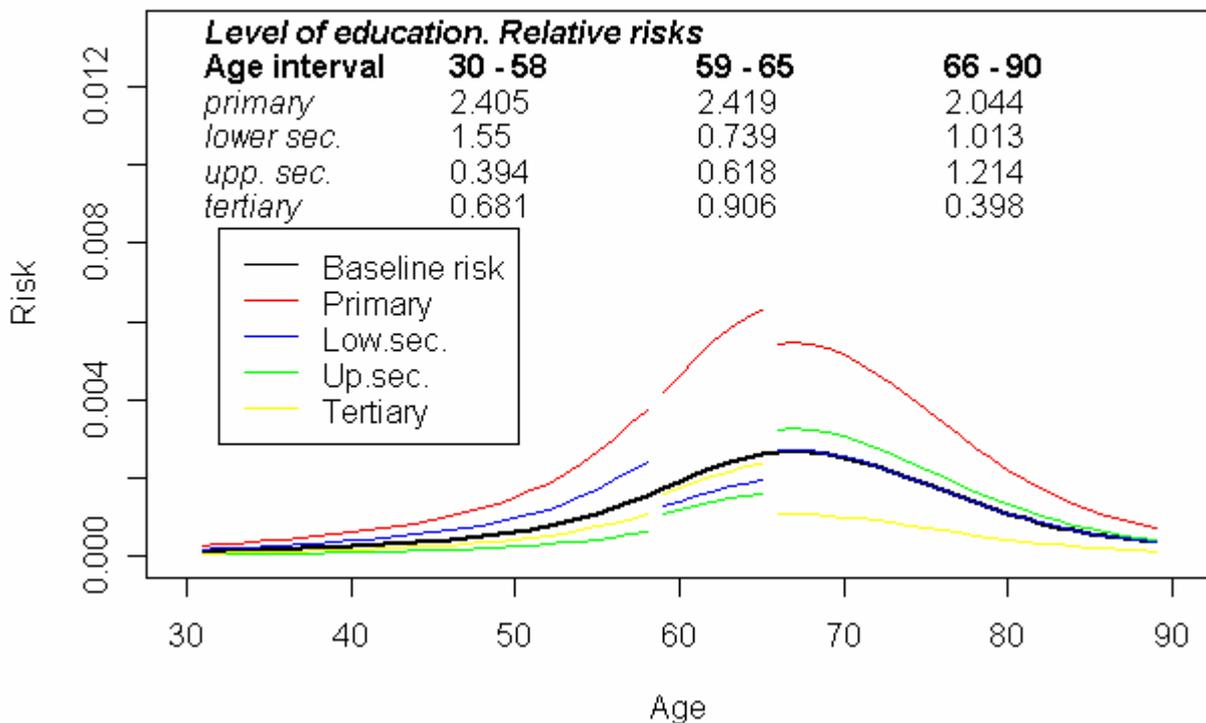| Level of education. Relative risks			
Age interval	15 - 29	30 - 33	34 - 50
primary	3.132	1.144	0.516
lower sec.	1.703	1.064	0.827
upp. sec.	0.684	1.061	1.296
tertiary	0.274	0.775	1.807

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

Italy. ISTAT 2003. TR16. Events observed in the window of observation: Women 1317; Men 1242.

TR17(with children -> without children) - Women

Level of education. Relative risks			
Age interval	**30 - 57**	**58 - 63**	**64 - 90**
primary	3.053	2.565	2.468
lower sec.	0.908	0.598	0.185
upp. sec.	0.387	0.628	1.793
tertiary	0.931	1.038	1.224

TR17(with children -> without children) - Men

Level of education. Relative risks			
Age interval	**30 - 58**	**59 - 65**	**66 - 90**
primary	2.405	2.419	2.044
lower sec.	1.55	0.739	1.013
upp. sec.	0.394	0.618	1.214
tertiary	0.681	0.906	0.398

Italy. ISTAT 2003. TR17. Events observed in the window of observation: Women 267; Men 186
NOTE: Window of observation extended to the last 20 years before the interview.

43

5. Age profiles of family and fertility events in the Netherlands

5.1 Data adaptations

We used data from Netherlands "Fertility and Family Survey" (FFS-NL). The 8145 interviews were carried out between February and June 2003. Since 1974, Statistics Netherlands organizes the Netherlands Fertility and Family Survey (FFS-NL) collecting longitudinal information on leaving home, cohabitation, marriage, and childbearing. Differently from the Italian data, in this case the exact date of birth is available, as well as the information on the date of each event, in the format month-year. Nonetheless, for some individuals it could happen that the information on the month is missing, since sometimes only a specification for the year of occurrence is provided. In this case, we input a random month extracted from a discrete uniform distribution $U(1,12)$.

For the Netherlands, data limitations and adaptation are similar to that needed in the Italian case. For example, we do not have all the required information for the analysis of all the transition proposed in Section 2. For the Dutch case we do not show detailed schemes for every transition since we still refer to the schemes proposed in Section 4 except for the following:

- children's dates of births are available only for women
- in the FSS-NL children's dates of exit (or death) are not available. Thus, we cannot study transitions related to the presence of children in the household. Besides, we have no information for evaluating the time-varying variable "with children in the household (Yes/No)".

Briefly, for the Netherlands we can analyze the following transitions:

TR1 (never married → married);
TR2 (married → divorced)
TR3 (married → widowed)
TR4 (divorced → married)
TR5 (widowed → married)
TR6 (parental home → with a partner)
TR7 (parental home → alone or with other persons)
TR12 (childless → 1 child) (only women)
TR13 (2nd birth) (only women)
TR14 (3rd birth) (only women)
TR15 (4th birth) (only women)

Transition-specific age profiles are estimated separately for men and women and the effect of educational level is considered. As time-varying covariates, we can consider "married (Yes/No)" and "having had first birth (Yes/No)" as an approximation of the missing covariate "with children in the household (Yes/No)".
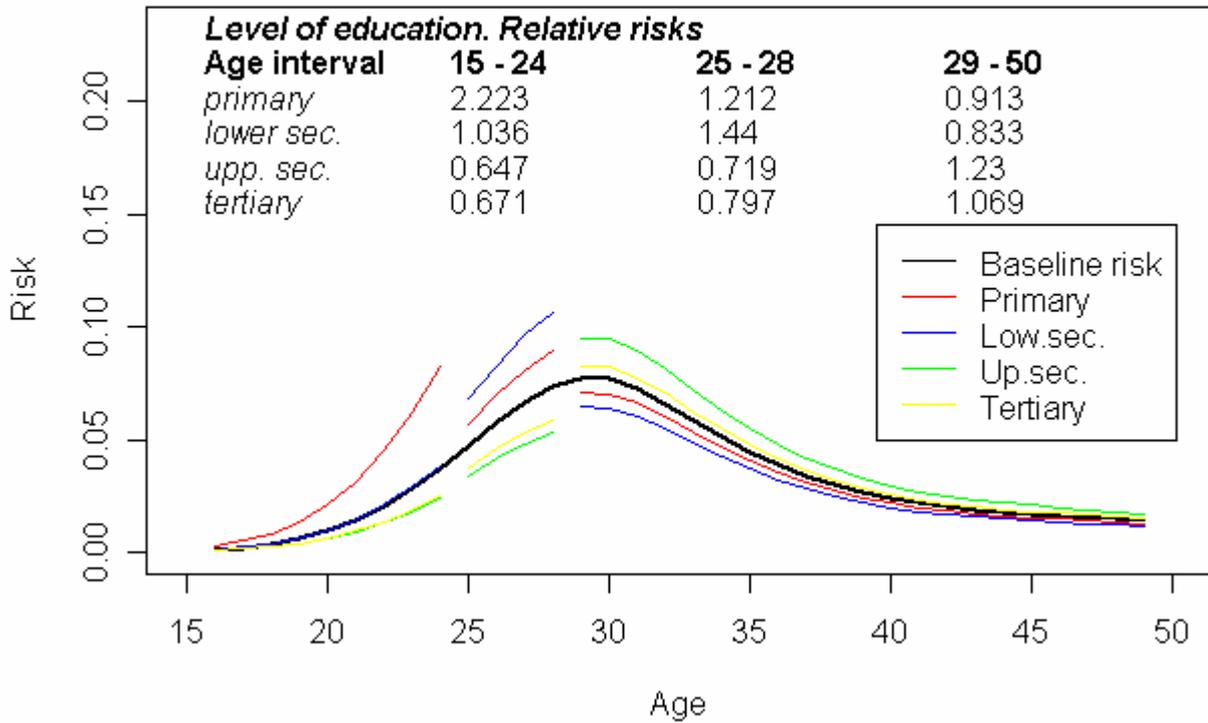
p-values associated with the Null Hypothesis that a specific covariate does not increase the fit of the model are shown in table 5. The meaning of this table is the same as table 4. Results suggest that, among women, education could be left out of the models in TR2,TR3, TR4; "with children" could be left out in all the three transitions that included this covariate whereas "married" could be excluded in TR14 and TR15. Among men, the introduction of education has a significant effect only for TR2 and TR7. We cannot consider time varying covariates for men since information on their child births is not available in the initial dataset.

Table 5. Analysis of Deviance. Significance Chi-square test (p-value) comparing model without covariate and model with education (Column 1); model with education and model with "married or not" (column 2); model with education and model with "with/without children in the household" (column 3) by sex and transition.
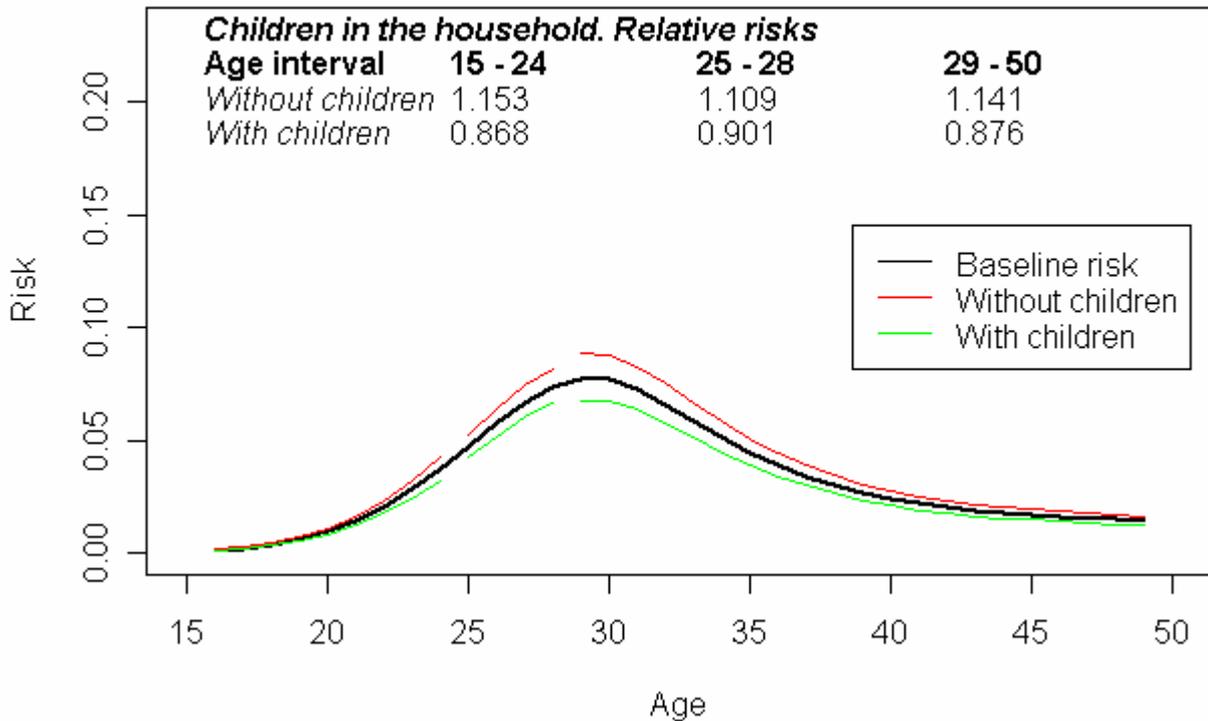
	Women			Men		
	education	married	with children	education	married	with children
	(1)	(2)	(3)	(1)	(2)	(3)
TR1 never married -> married	0.006	-	0.687	0.213	-	-
TR2 married -> divorced	0.67	-	0.36	0.046	-	-
TR3 married -> widowed	0.822	-	-	-	-	-
TR4 divorced -> married	0.269	-	0.413	0.241	-	-
TR5 widowed -> married	-	-	-	-	-	-
TR6 parental home -> with a partner	0.024	-	-	0.843	-	-
TR7 parental home -> alone or with other persons	0.000	-	-	0.000	-	-
TR12 childless -> 1 child	0.000	0.000	-	-	-	-
TR13 1 -> 2 children	0.000	0.000	-	-	-	-
TR14 2->3 children	0.012	0.124	-	-	-	-
TR15 3->4 children	0.011	0.446	-	-	-	-

TR1 (never married->married) - Women

Level of education. Relative risks

Age interval	15 - 24	25 - 28	29 - 50
primary	2.223	1.212	0.913
lower sec.	1.036	1.44	0.833
upp. sec.	0.647	0.719	1.23
tertiary	0.671	0.797	1.069

Legend:
- Baseline risk
- Primary
- Low.sec.
- Up.sec.
- Tertiary

TR1 (never married->married) - Women

Children in the household. Relative risks

Age interval	15 - 24	25 - 28	29 - 50
Without children	1.153	1.109	1.141
With children	0.868	0.901	0.876

Legend:
- Baseline risk
- Without children
- With children

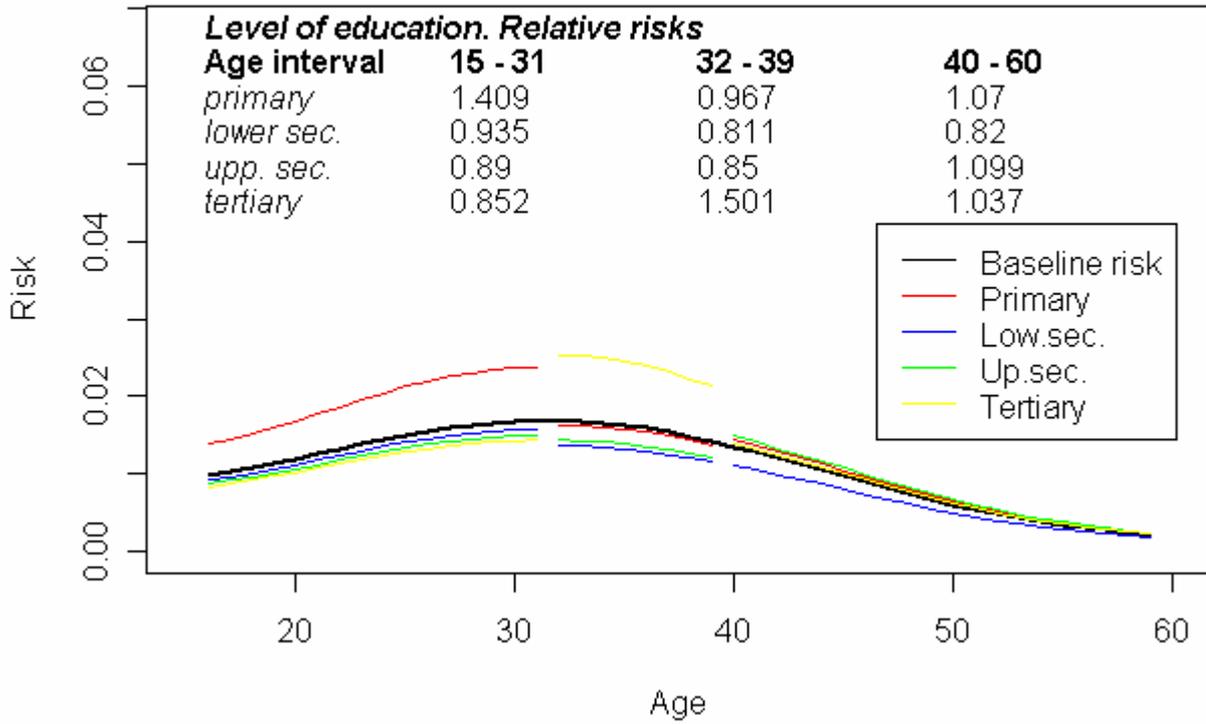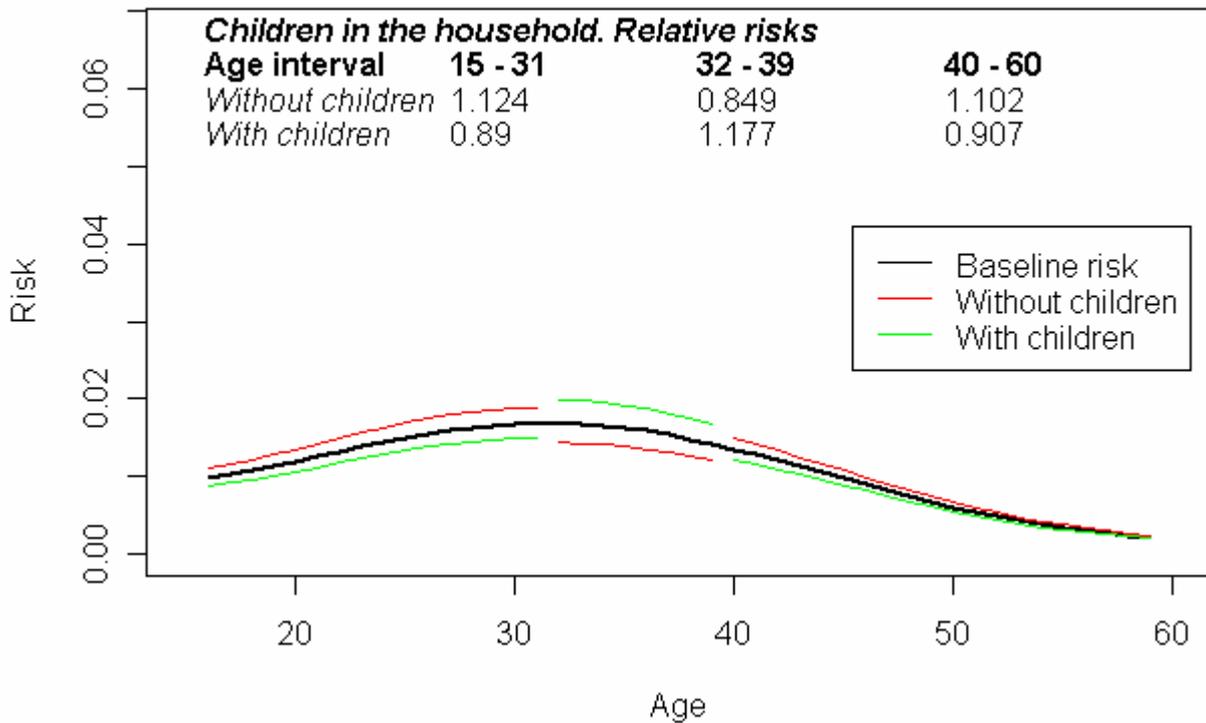*The Netherlands. FFS 2003. TR1 Women. Events observed in the window of observation: 295*

46
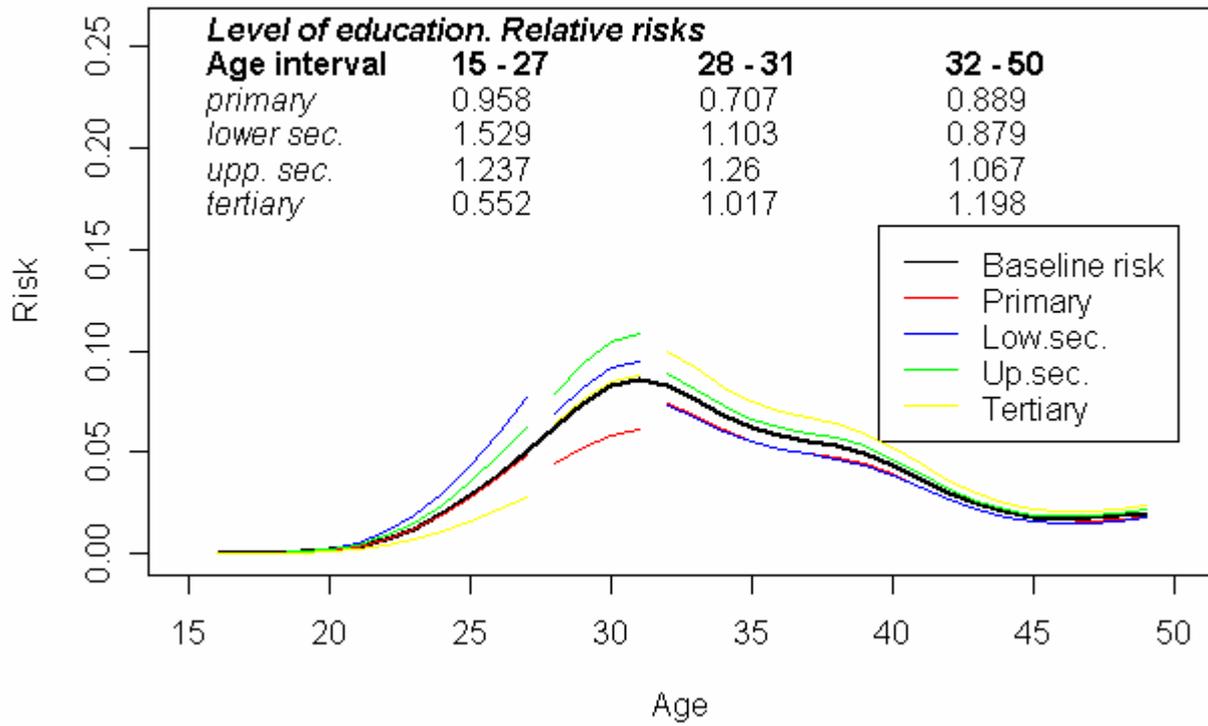
The Netherlands. FFS 2003. TR1 Men. Events observed in the window of observation: 312

TR2 (married -> divorced) - Women

Level of education. Relative risks

Age interval	15 - 31	32 - 39	40 - 60
primary	1.409	0.967	1.07
lower sec.	0.935	0.811	0.82
upp. sec.	0.89	0.85	1.099
tertiary	0.852	1.501	1.037

Legend:
- Baseline risk
- Primary
- Low.sec.
- Up.sec.
- Tertiary

TR2 (married -> divorced) - Women

Children in the household. Relative risks

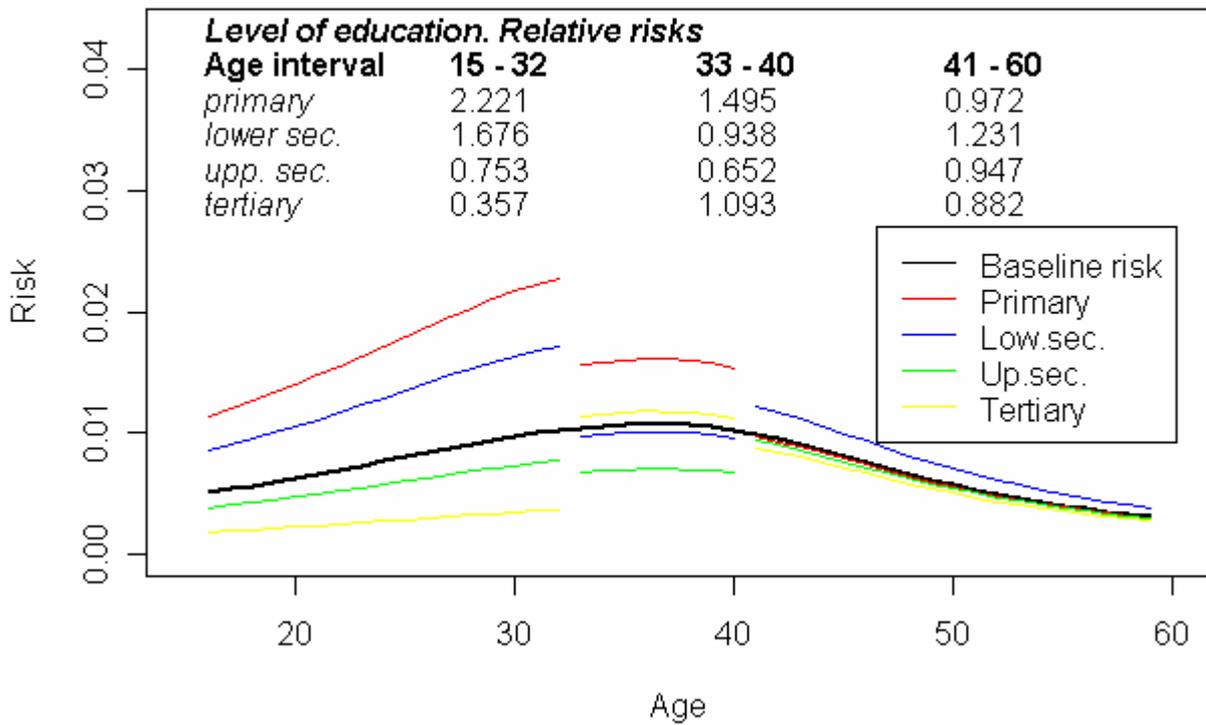| Age interval	15 - 31	32 - 39	40 - 60
Without children	1.124	0.849	1.102
With children	0.89	1.177	0.907

Legend:
- Baseline risk
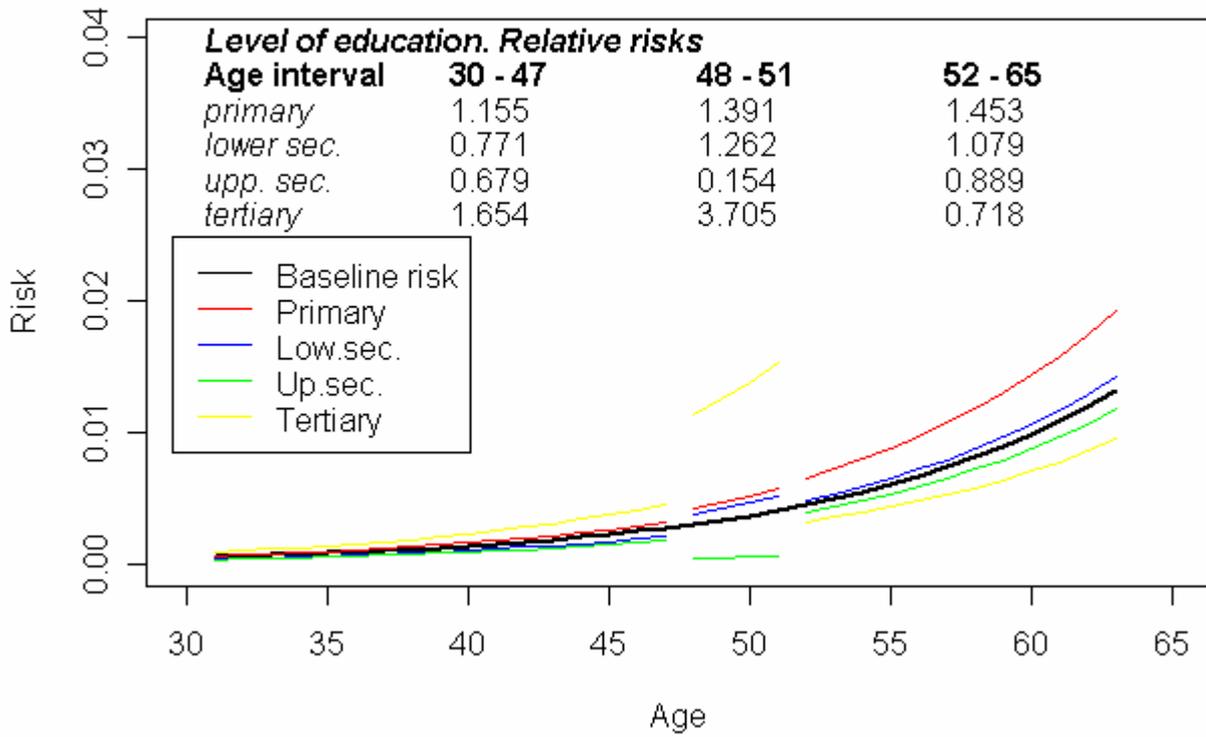- Without children
- With children

The Netherlands. FFS 2003. TR2 Women. Events observed in the window of observation: 239
NOTE: Window of observation extended to the last 10 years before the interview.

TR2 (married -> divorced) - Men

Level of education. Relative risks

Age interval	15 - 32	33 - 40	41 - 60
primary	2.221	1.495	0.972
lower sec.	1.676	0.938	1.231
upp. sec.	0.753	0.652	0.947
tertiary	0.357	1.093	0.882

Legend:
- Baseline risk
- Primary
- Low.sec.
- Up.sec.
- Tertiary

The Netherlands. FFS 2003. TR2 Men. Events observed in the window of observation: 191
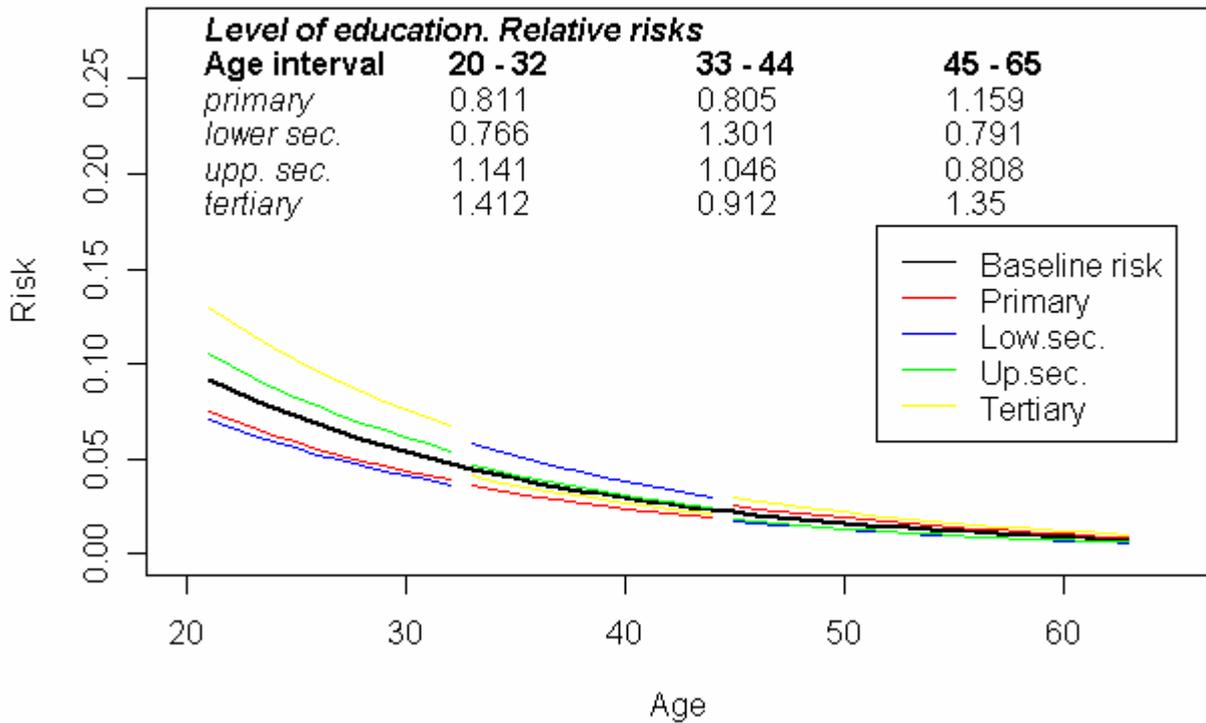*NOTE: Window of observation extended to the last 10 years before the interview.*

49

TR3 (married -> widowed) - Women

Level of education. Relative risks			
Age interval	30 - 47	48 - 51	52 - 65
primary	1.155	1.391	1.453
lower sec.	0.771	1.262	1.079
upp. sec.	0.679	0.154	0.889
tertiary	1.654	3.705	0.718

Legend:
— Baseline risk
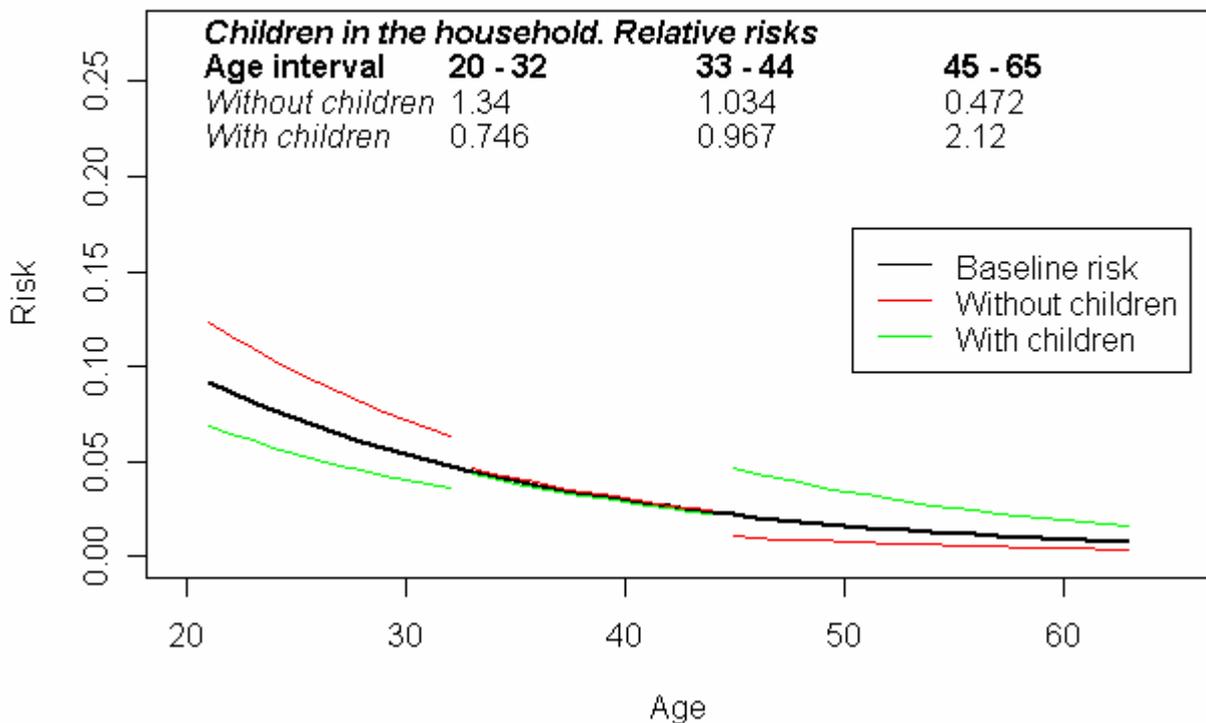— Primary
— Low.sec.
— Up.sec.
— Tertiary

The Netherlands. FFS 2003. TR3. Events observed in the window of observation: Women 55; Men 19 (too few events: baseline not shown)
NOTE: Window of observation extended to the last 10 years before the interview.

50

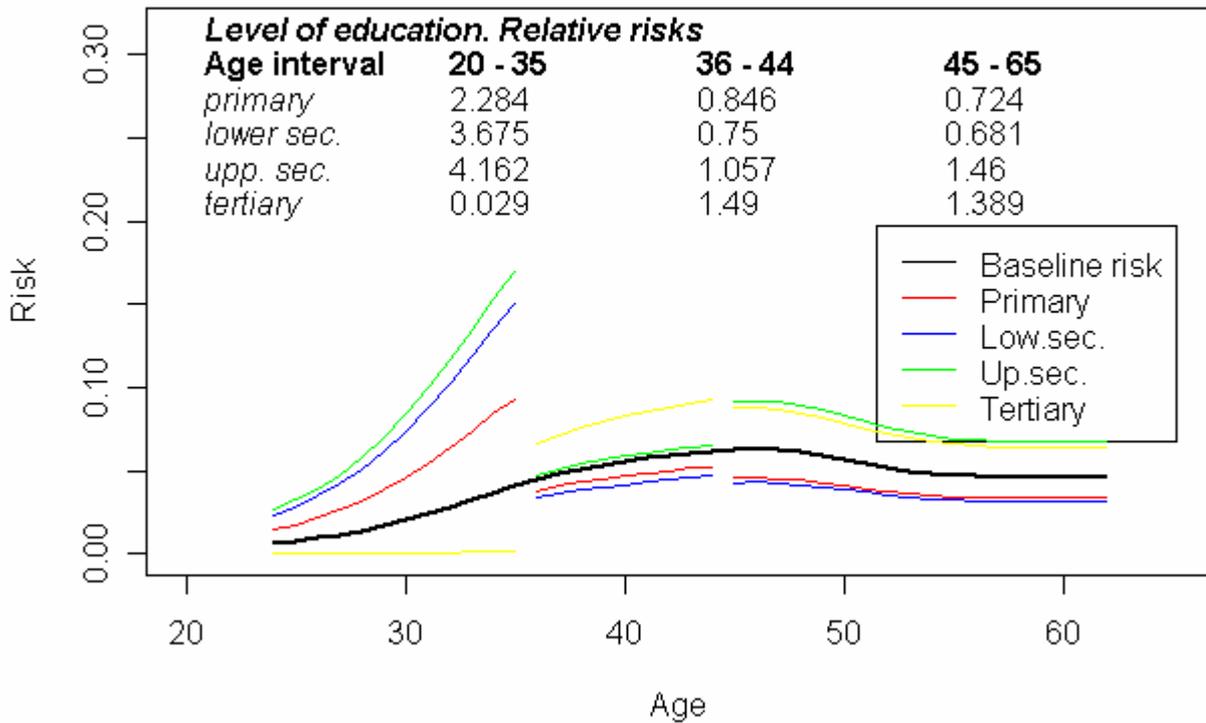## TR4 (divorced -> second marriage) - Women

Level of education. Relative risks			
Age interval	20 - 32	33 - 44	45 - 65
primary	0.811	0.805	1.159
lower sec.	0.766	1.301	0.791
upp. sec.	1.141	1.046	0.808
tertiary	1.412	0.912	1.35

Legend:
- Baseline risk
- Primary
- Low.sec.
- Up.sec.
- Tertiary

TR4 (divorced -> second marriage) - Women

Children in the household. Relative risks			
Age interval	20 - 32	33 - 44	45 - 65
Without children	1.34	1.034	0.472
With children	0.746	0.967	2.12

Legend:
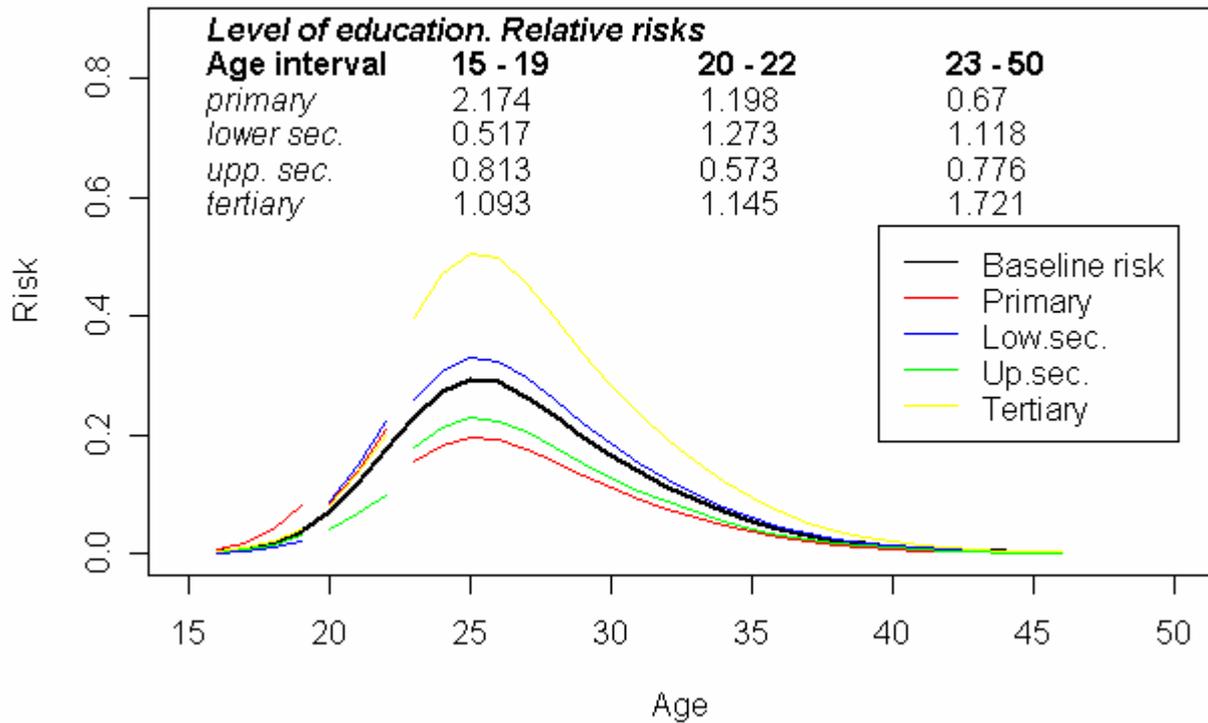- Baseline risk
- Without children
- With children

The Netherlands. FFS 2003. TR4 Women. Events observed in the window of observ.: 97
NOTE: Window of observation extended to the last 10 years before the interview.
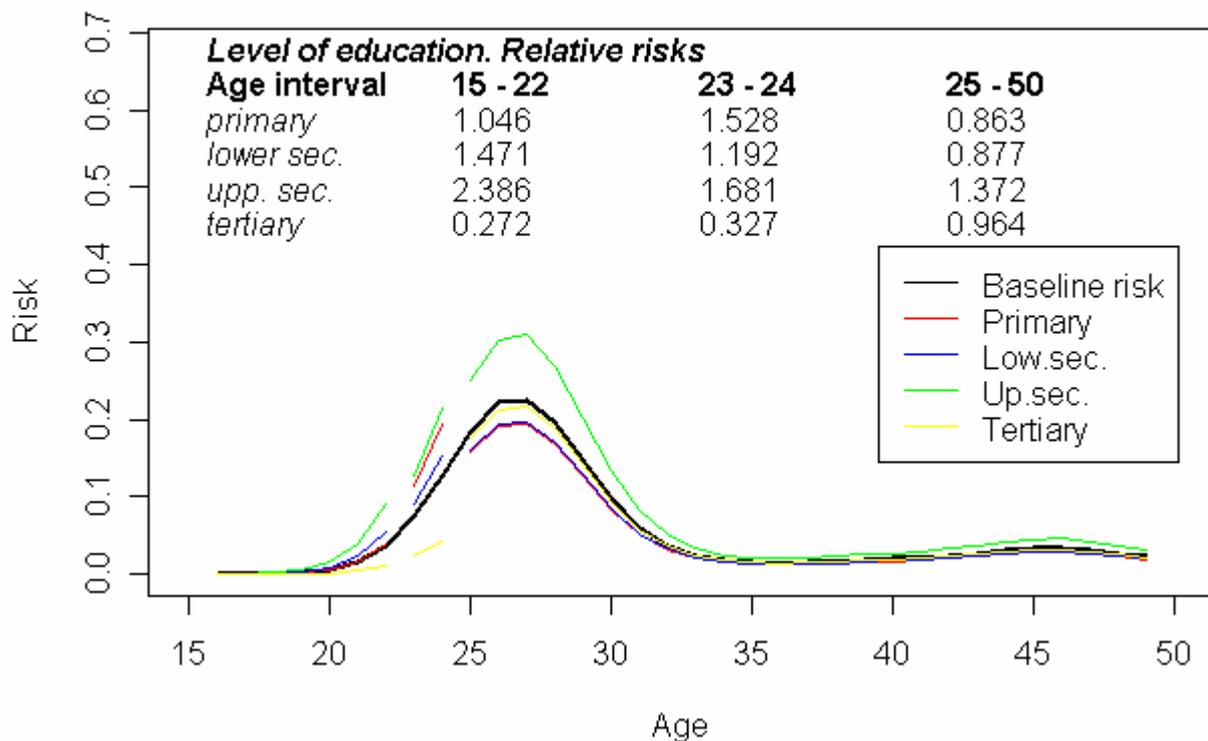
51

TR4 (divorced -> second marriage) - Men

Level of education. Relative risks			
Age interval	20 - 35	36 - 44	45 - 65
primary	2.284	0.846	0.724
lower sec.	3.675	0.75	0.681
upp. sec.	4.162	1.057	1.46
tertiary	0.029	1.49	1.389

The Netherlands. FFS 2003. TR4 Women. Events observed in the window of observ.: 105
NOTE: Window of observation extended to the last 10 years before the interview.

TR5 (second marriage after death of spouse)
The Netherlands. FFS 2003. TR5. Events observed in the window of observ.: Women 15; Men 11
NOTE: too few cases (baseline not shown)
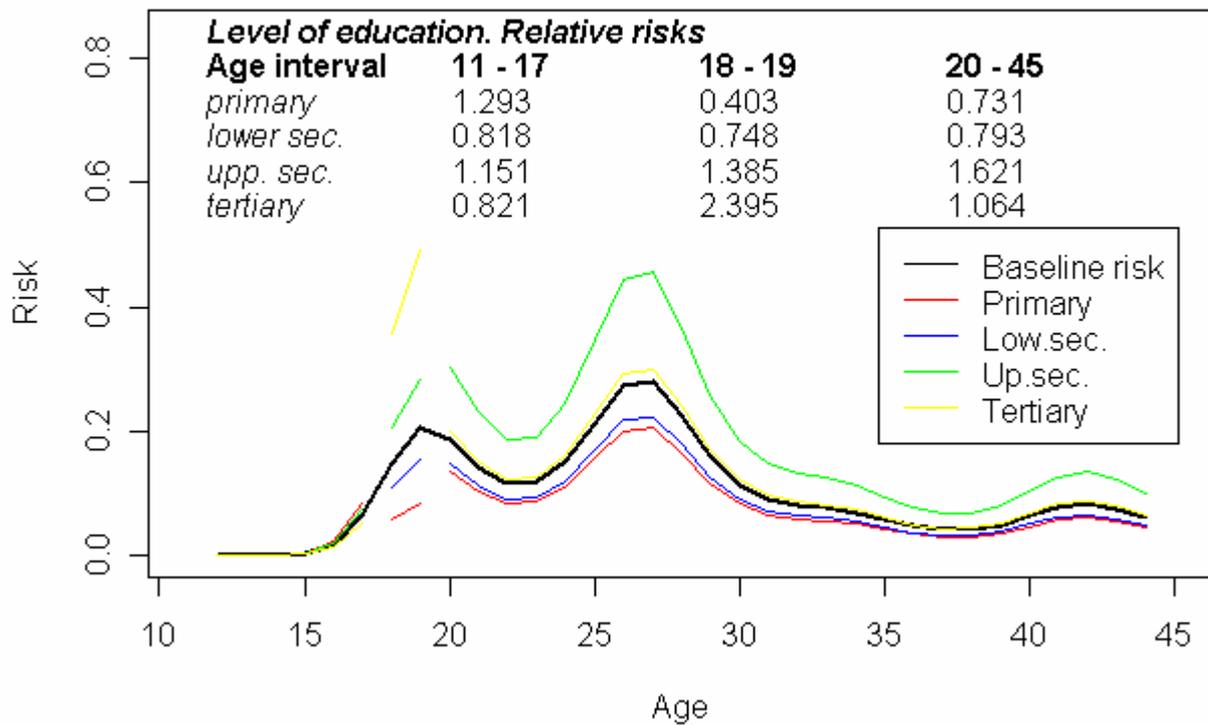
TR6 (in parental home -> with a partner) - Women

Level of education. Relative risks			
Age interval	15 - 19	20 - 22	23 - 50
primary	2.174	1.198	0.67
lower sec.	0.517	1.273	1.118
upp. sec.	0.813	0.573	0.776
tertiary	1.093	1.145	1.721

TR6 (in parental home -> with a partner) - Men

Level of education. Relative risks			
Age interval	15 - 22	23 - 24	25 - 50
primary	1.046	1.528	0.863
lower sec.	1.471	1.192	0.877
upp. sec.	2.386	1.681	1.372
tertiary	0.272	0.327	0.964

The Netherlands. FFS 2003. TR6. Events observed in the window of observ: Women 167; Men 157.

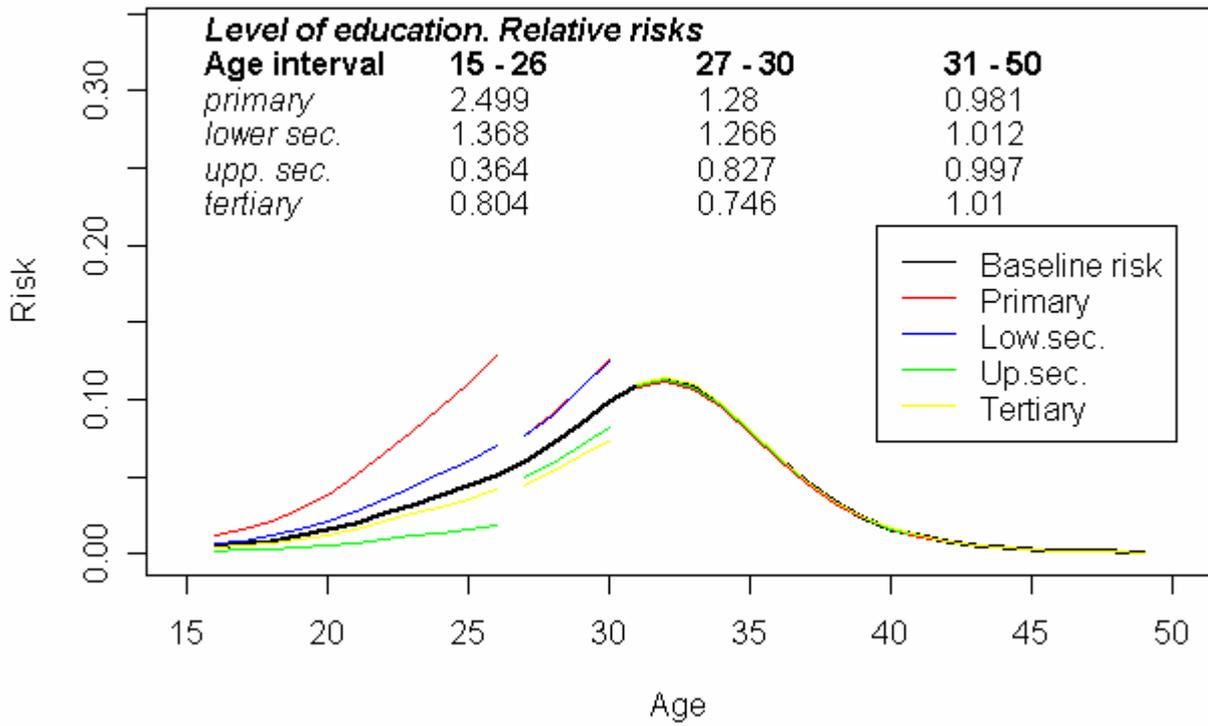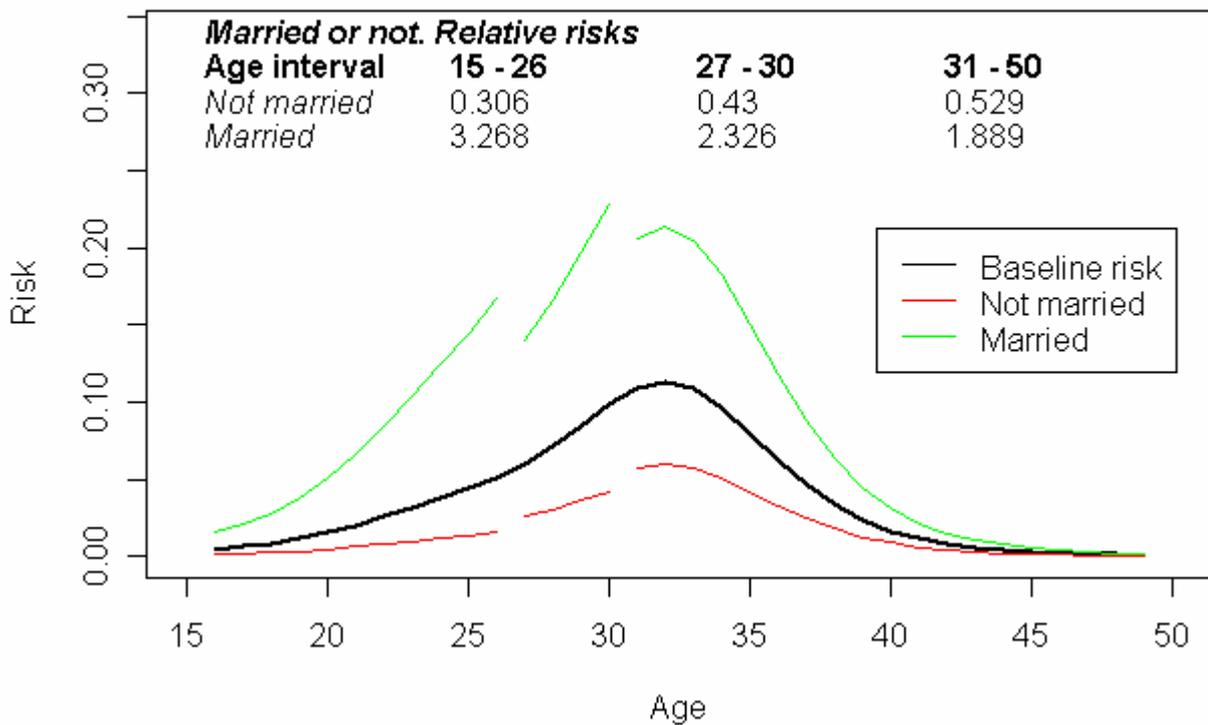## TR7 (in parental home -> alone or with other persons) - Women

Level of education. Relative risks			
Age interval	11 - 17	18 - 19	20 - 45
primary	1.293	0.403	0.731
lower sec.	0.818	0.748	0.793
upp. sec.	1.151	1.385	1.621
tertiary	0.821	2.395	1.064

TR7 (in parental home -> alone or with other persons) - Men

Level of education. Relative risks			
Age interval	11 - 18	19 - 22	23 - 45
primary	0.722	0.625	0.784
lower sec.	0.68	0.719	0.675
upp. sec.	1.173	1.113	0.913
tertiary	1.736	1.999	2.067
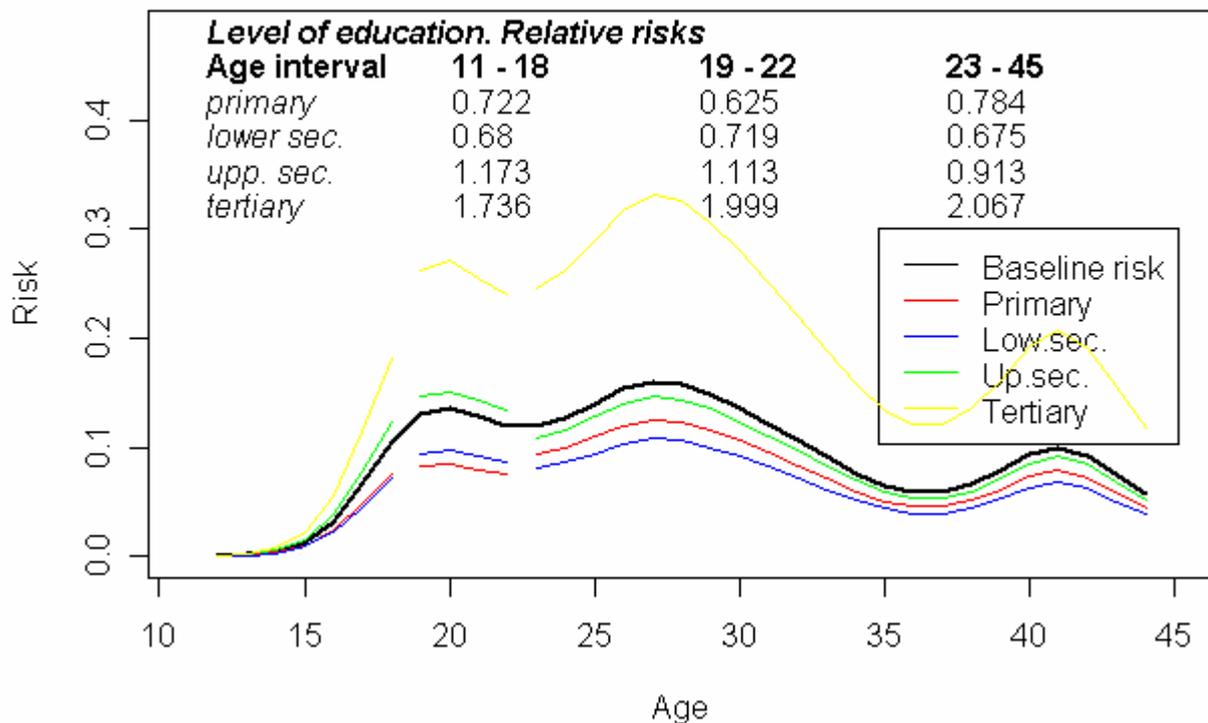
The Netherlands. FFS 2003. TR7. Events observed in the window of observ.: Women 510; Men 559
NOTE: Window of observation extended to the last 10 years before the interview.
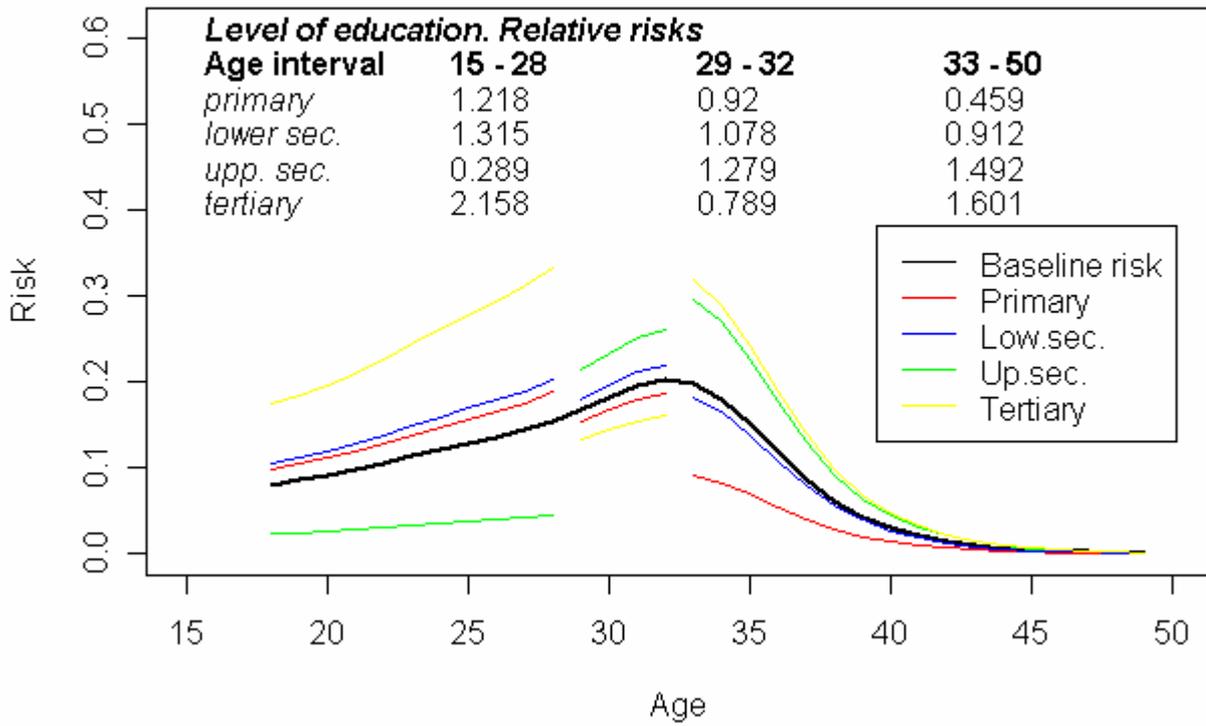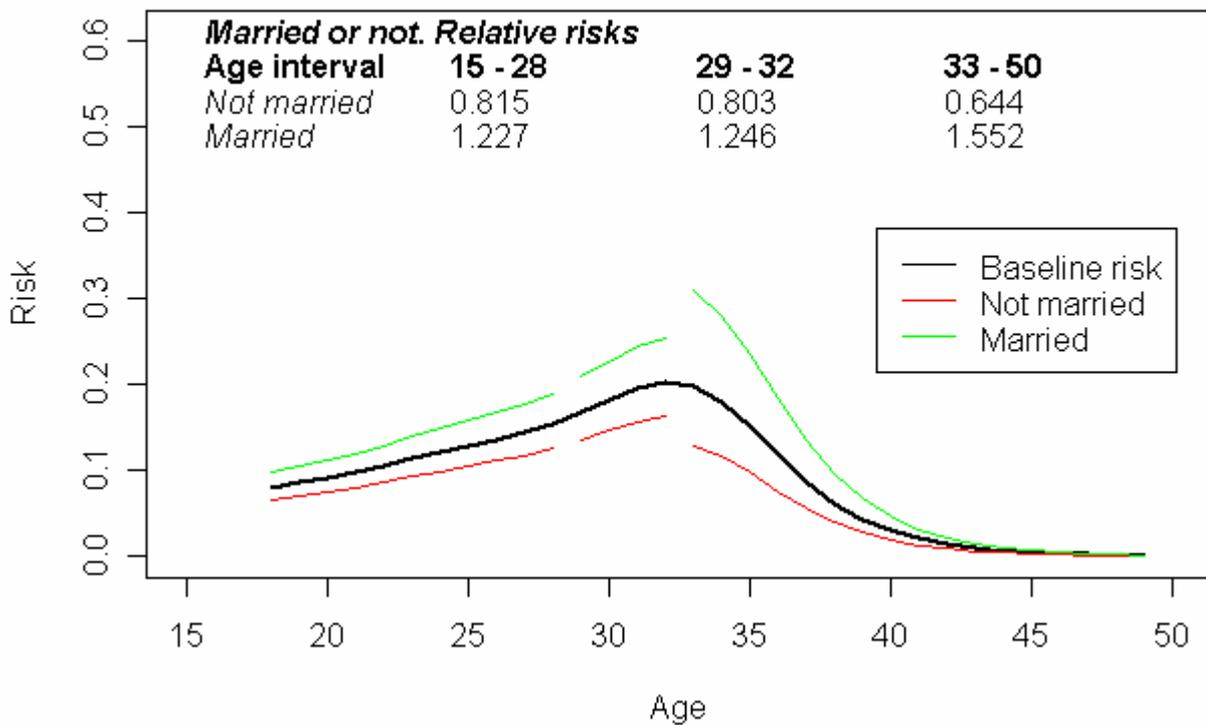
TR12 (first birth) - Women

Level of education. Relative risks

Age interval	15 - 26	27 - 30	31 - 50
primary	2.499	1.28	0.981
lower sec.	1.368	1.266	1.012
upp. sec.	0.364	0.827	0.997
tertiary	0.804	0.746	1.01

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR12 (first birth) - Women

Married or not. Relative risks

Age interval	15 - 26	27 - 30	31 - 50
Not married	0.306	0.43	0.529
Married	3.268	2.326	1.889

Legend: Baseline risk, Not married, Married

The Netherlands. FFS 2003. TR12 Women. Events observed in the window of observ.: Women 373.
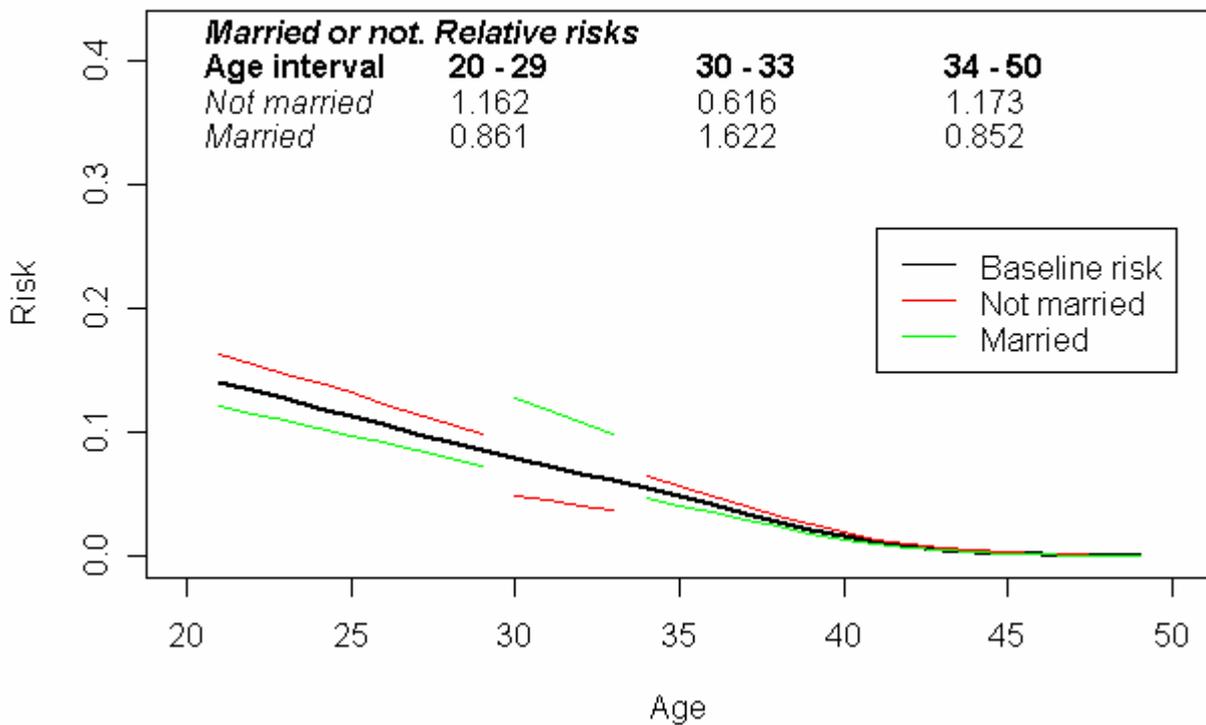
TR13 (second birth) - Women

Level of education. Relative risks

Age interval	15 - 28	29 - 32	33 - 50
primary	1.218	0.92	0.459
lower sec.	1.315	1.078	0.912
upp. sec.	0.289	1.279	1.492
tertiary	2.158	0.789	1.601

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR13 (second birth) - Women

Married or not. Relative risks

Age interval	15 - 28	29 - 32	33 - 50
Not married	0.815	0.803	0.644
Married	1.227	1.246	1.552

Legend: Baseline risk, Not married, Married

The Netherlands. FFS 2003. TR13 Women. Events observed in the window of observ.: Women 285.

TR14 (third birth) - Women

Level of education. Relative risks

Age interval	20 - 29	30 - 33	34 - 50
primary	1.028	0.661	0.505
lower sec.	0.735	0.494	0.602
upp. sec.	0.632	1.098	1.407
tertiary	2.093	2.786	2.335

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR14 (third birth) - Women

Married or not. Relative risks

Age interval	20 - 29	30 - 33	34 - 50
Not married	1.162	0.616	1.173
Married	0.861	1.622	0.852

Legend: Baseline risk, Not married, Married
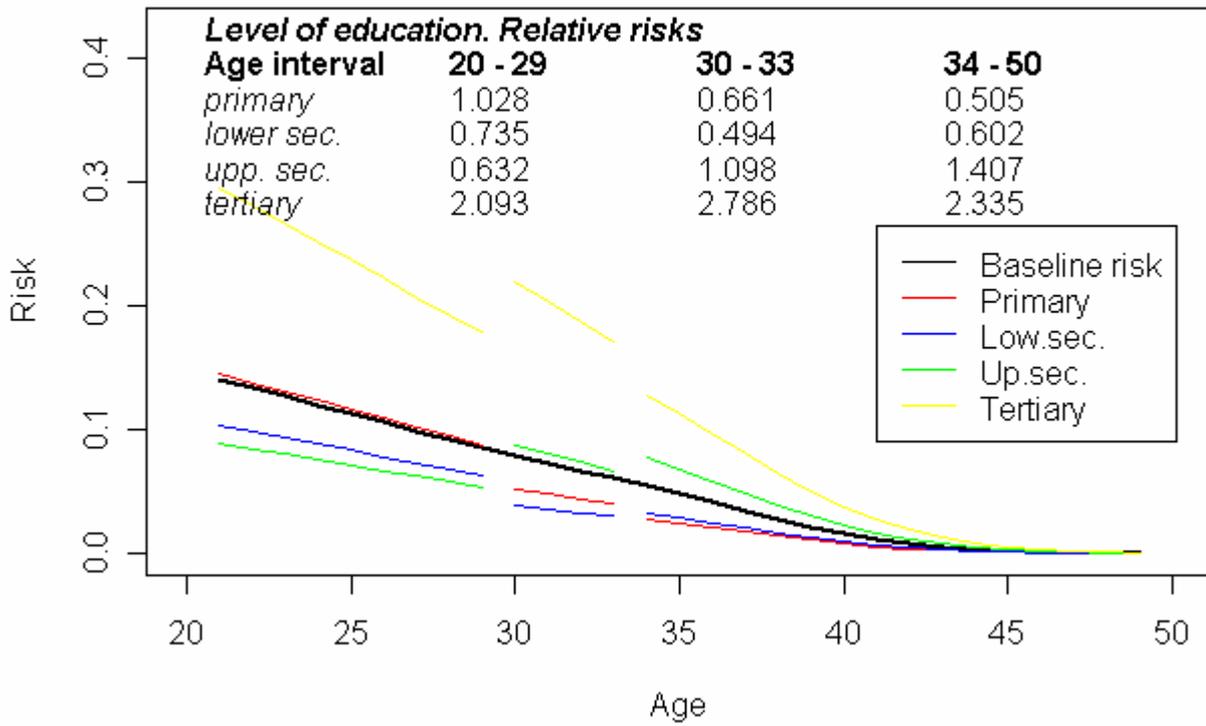
The Netherlands. FFS 2003. TR14 Women. Events observed in the window of observ.: Women 96.

TR15(fourth birth) - Women

Level of education. Relative risks

Age interval	20 - 30	31 - 33	34 - 50
primary	0.949	0.498	0.394
lower sec.	0.813	1.247	0.659
upp. sec.	2.395	0.631	1.208
tertiary	0.541	2.548	3.191

Legend: Baseline risk, Primary, Low.sec., Up.sec., Tertiary

TR15(fourth birth) - Women

Married or not. Relative risks

Age interval	20 - 30	31 - 33	34 - 50
Not married	1.107	1.703	1.225
Married	0.904	0.587	0.817

Legend: Baseline risk, Not married, Married
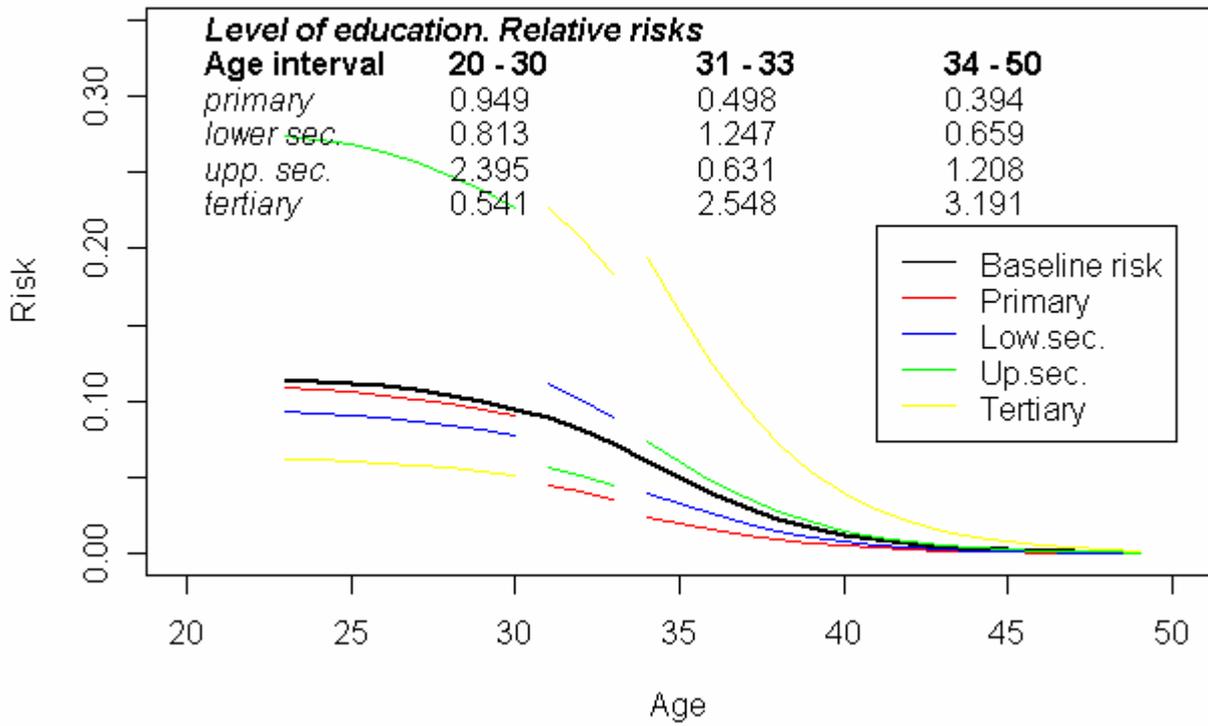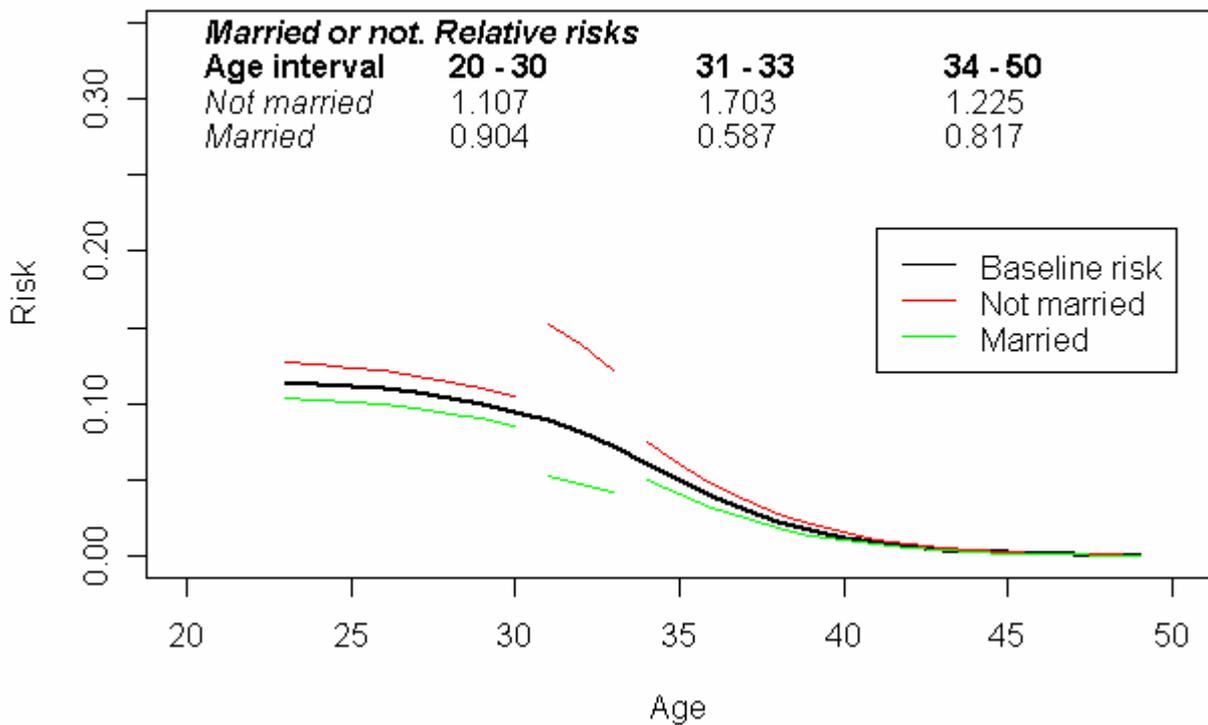
The Netherlands. FFS 2003. TR15 Women. Events observed in the window of observ.: Women 76.
NOTE: Window of observation extended to the last 15 years before the interview.

Bibliographic references

Billari F.C., Philipov D., Toulemon L., 2005, *Report on fertility, family and household data, and current and future trends in fertility and household structure*, Deliverable D20, MicMac Project "Bridging the micro-macro gap in population forecasting".

Blossfeld Hans-Peter, Rohwer Götz. 2002. *Techniques of Event History Modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chambers J.M. and Hastie T.J. (eds.), 1992, *Statistical models in S*, New York: Chapman and Hall.

de Beer J., van der Gaag N., Willekens F., 2006, *Report on Input Data Requirements of MAC*, Deliverable D2, MicMac Project "Bridging the micro-macro gap in population forecasting".

Hastie T.J and Tibshirani R.J., 1990, *Generalized additive models*, London: Chapman and Hall.

Hastie T.J., Tibshirani R.J. and Friedman J., 2001, *The elements of statistical learning. Data mining, inference and prediction*, New York: Springer-Verlag.

McNeil D.R., Trussell T.J., Turner J.C., 1977, "Spline interpolation of demographic data", *Demography*, 14 (2): 245-252.

Schmertmann C., 2003, "A system of model fertility schedules with graphically intuitive parameters", *Demographic Research*, 9 (5): 81-110.

Smith L., Hyndman R.J., Wood S.N., 2004, "Spline interpolation for demographic variables: the monotonicity problem", *Journal of Population Research*, 21(1): 95-98.

van der Gaag N., de Beer J., Ekamper P., Willekens F., 2006, *Using MicMac to project living arrangements: an illustration of biographic projections*, paper presented at the European Population Conference, Liverpool.

Venables W.N. and Ripley B.D., 1997, *Modern applied statistics with S-plus*, New York: Springer.

Wachter K.W., Blackwell D., Hammel E.A., 1998, *Testing the Validity of Kinship Microsimulation: An Update,* University of California, Berkeley, CA.

Willekens F., 2005, "Biographic forecasting: bridging the micro-macro gap in population forecasting". *New Zealand Population Review* 31 (1): 77-124.

Wood S.N, 2006, *Generalize Additive Models. An introduction with R*, Chapman and Hall/CRC

Complete R syntax with comments

1. PREPARING DATA

```
rm(list=ls())  # all objects are removed from memory

#read initial data set
d<-read.table(file="C:\\Documents        and        Settings\\Impicciatore\\Documenti\\RICERCA\\Mic
Mac\\ITALY.dat",
  header=TRUE, sep="\t",na.strings="NA",dec=".")


#(Completely) missing months as casual numbers from uniform distribution
d$m.exit<- ifelse(!is.na(d$y.exit),
  ifelse(d$marr==1 & d$y.exit==d$y.marr,d$m.marr,
    trunc(runif(d$id,1,13)))),NA)
d$m.div<-ifelse(d$div==1,trunc(runif(d$id,1,13)),NA)
d$m.ved<-ifelse(d$ved==1,trunc(runif(d$id,1,13)),NA)
d$m.without<-ifelse(d$without==1,trunc(runif(d$id,1,13)),NA)


#Decimal dates
d$inter <- 103.88  #At the interview (November 2003)
d$birdate<-  d$y.birth -1900 + round((d$m.birth-0.5)/12,2) # At birth
d$exitdate<-   d$y.exit -1900 + round((d$m.exit-0.5)/12,2) # At the exit from parental home
d$mardate<-  d$y.marr -1900 + round((d$m.marr-0.5)/12,2)   #At the first marriage
d$mar2date<-  d$y.marr2 -1900 + round((d$m.marr2-0.5)/12,2)#At the second marriage
d$divdate<- d$y.div -1900 + round((d$m.div-0.5)/12,2)      #at divorce
d$veddate<- d$y.ved -1900 + round((d$m.ved-0.5)/12,2)      #at the death of spouse
d$ch1date<-  d$y.ch1 -1900 + round((d$m.ch1-0.5)/12,2)     # at the first child
d$ch2date<-  d$y.ch2 -1900 + round((d$m.ch2-0.5)/12,2)     # at the second child
d$ch3date<-  d$y.ch3 -1900 + round((d$m.ch3-0.5)/12,2)     # at the third child
d$ch4date<-  d$y.ch4 -1900 + round((d$m.ch4-0.5)/12,2)     # at the fourth child
d$withdate<-d$y.withch -1900 + round((d$m.withch-0.5)/12,2)# enter 'with children'
d$withoutdate<-d$y.without -1900 + round((d$m.without-0.5)/12,2)# enter 'without children'

# ages at the various events (exact ages)
d$ageint<-d$inter-d$birdate        # at the interview
d$agexit<-d$exitdate-d$birdate     # at the exit from parental home
d$agemarr<-d$mardate-d$birdate     # at  first marriage
d$agemarr2<-d$mar2date-d$birdate   # at second marriage
d$agediv<-d$divdate-d$birdate      # at the divorce
d$ageved<-d$veddate-d$birdate      # at the death of spouse
d$agech1<-d$ch1date-d$birdate      # at the first child
d$agech2<-d$ch2date-d$birdate      # at the second child
d$agech3<-d$ch3date-d$birdate      # at the third child
d$agech4<-d$ch4date-d$birdate      # at the fourth child
d$agewith<-d$withdate-d$birdate     # at the entry 'with children'
d$agewithout<-d$withoutdate-d$birdate# at the entry 'without children'
d$agesw<-trunc(d$ageint)-5         #Age at (x-5)-th birthday (starting age)

# Control for ages at various events
# Cases with a very low ages are eliminated
d$marr<-ifelse(d$marr!=1,d$marr,ifelse(d$agemarr<14,9,1))
d$ch1<-ifelse(d$ch1!=1,d$ch1,ifelse(d$agech1<14,9,1))
d$ch2<-ifelse(d$ch2!=1,d$ch2,ifelse(d$agech2<14,9,1))
d$ch3<-ifelse(d$ch3!=1,d$ch3,ifelse(d$agech3<14,9,1))
d$ch4<-ifelse(d$ch4!=1,d$ch4,ifelse(d$agech4<14,9,1))
d$agexit<-ifelse(d$agexit<0,0,d$agexit)
d$agewith<-ifelse(d$agewith<14,14,d$agewith)
d$agewithout<-ifelse(d$agewithout<14,14,d$agewithout)




#_____
# FROM HERE: WE MUST SPECIFY PARAMETERS FOR EACH TRANSITION
#_____


# EPISODE DEFINITION
#agestart, agefin and cens are determined separately for each transition
#ATTENTION: AGESTART, AGEFIN AND CENS MUST BE REINIZIALED EVERY TRANSITION

# Cens is defined as
#0 the case is censored at the INTERVIEW
#1 the case experienced the event in the window of observation
#2 the case experienced a censoring event in the window of observation
#9 the case is not included in the analysis

#_____
# TR1  -  MARRIAGE
transition<-"TR1 (never married->married)"
minage<-15
```

```
maxage<-50
d$agestart<-d$agesw
d$agefin<-pmin(d$agemarr,d$ageint,na.rm=TRUE) # NA value are not considered
# control: if agestart or agefin >100 then 100
d$agestart<-ifelse(d$agestart>100,100,d$agestart)
d$agefin<-ifelse(d$agefin>100,100,d$agefin)
d$cens<-ifelse(d$marr==1,
      ifelse (d$agemarr<d$agestart,9,1),
      ifelse(d$marr==0,0,9))


# Matrix Selection (d$cens not equal 9)
dtemp<-subset(d,d$cens<9,
   select=c(id,agestart,agefin,cens,weight,withch,agewith,without,
            agewithout,sex,edu,))


#splitting (with children in the household (Y/N))
i<-1
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$withch[i]==1 & (dtemp$agewith[i]>dtemp$agestart[i]
                    & dtemp$agewith[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,withch,agewith,
        # without,agewithout,sex,edu,
        dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agewith[i],dtemp$agefin[i],
                           dtemp$cens[i],dtemp$weight[i],dtemp$withch[i],
                           dtemp$agewith[i],dtemp$without[i],dtemp$agewithout[i],
                           dtemp$sex[i],dtemp$edu[i]))
        # old row
        dtemp$agefin[i]<-dtemp$agewith[i]
        dtemp$cens[i]<-0
        k<-k+1
    }
}

# Time-varying covariate  (1 Without children   2 With children)
dtemp$timed<-ifelse(dtemp$withch==1 & dtemp$agewith<=dtemp$agestart,2,1)

#Splitting for without children
i<-1
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$timed[i]==2 & dtemp$without[i]==1 &
        (dtemp$agewithout[i]>dtemp$agestart[i]
        & dtemp$agewithout[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,withch,agewith,
        # without,agewithout,sex,edu,
        dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agewithout[i],dtemp$agefin[i],
                           0,dtemp$weight[i],dtemp$withch[i],
                           dtemp$agewith[i],dtemp$without[i],dtemp$agewithout[i],
                           dtemp$sex[i],dtemp$edu[i],1))
        # old row
        dtemp$agefin[i]<-dtemp$agewithout[i]
        dtemp$cens[i]<-1
        k<-k+1
    }
}

dx<-subset(dtemp, select=c(id,agestart,agefin,cens,weight,sex,edu,timed))
remove(dtemp)




#_____
# TR2 -  DIVORCE
transition<-"TR2 (married -> divorced)"
minage<-15
maxage<-90
d$agestart<-trunc(pmax(d$agesw,d$agemarr,na.rm=TRUE))
d$agefin<-pmin(d$agediv,d$ageint,d$ageved, na.rm=TRUE)
d$cens<-ifelse(d$div==1,
        ifelse (d$agediv<d$agestart,9,1),
        ifelse(d$div==0,ifelse(is.na(d$ageved),0,
            ifelse(d$ageved<d$agestart,9,2)),9))


# Matrix Selection (d$cens not equal 9)
dtemp<-subset(d,d$cens<9,
   select=c(id,agestart,agefin,cens,weight,withch,agewith,without,
            agewithout,sex,edu,))


#splitting (with children in the household (Y/N))
i<-1
```

```
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$withch[i]==1 & (dtemp$agewith[i]>dtemp$agestart[i]
                    & dtemp$agewith[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,withch,agewith,
        # without,agewithout,sex,edu,
        dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agewith[i],dtemp$agefin[i],
                            dtemp$cens[i],dtemp$weight[i],dtemp$withch[i],
                            dtemp$agewith[i],dtemp$without[i],dtemp$agewithout[i],
                            dtemp$sex[i],dtemp$edu[i]))
        # old row
        dtemp$agefin[i]<-dtemp$agewith[i]
        dtemp$cens[i]<-0
        k<-k+1
    }
}

# Time-varying covariate  (1 Without children   2 with children)
dtemp$timed<-ifelse(dtemp$withch==1 & dtemp$agewith<=dtemp$agestart,2,1)

#Splitting for without children
i<-1
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$timed[i]==2 & dtemp$without[i]==1 &
        (dtemp$agewithout[i]>dtemp$agestart[i]
        & dtemp$agewithout[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,withch,agewith,
        # without,agewithout,sex,edu,
        dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agewithout[i],dtemp$agefin[i],
                            0,dtemp$weight[i],dtemp$withch[i],
                            dtemp$agewith[i],dtemp$without[i],dtemp$agewithout[i],
                            dtemp$sex[i],dtemp$edu[i],1))
        # old row
        dtemp$agefin[i]<-dtemp$agewithout[i]
        dtemp$cens[i]<-1
        k<-k+1
    }
}

dx<-subset(dtemp, select=c(id,agestart,agefin,cens,weight,sex,edu,timed))
remove(dtemp)



#_____
# TR3  -   DEATH OF SPOUSE
transition<-"TR3 (married -> widowed)"
minage<-30
maxage<-100
d$agestart<-trunc(pmax(d$agesw,d$agemarr,na.rm=TRUE))
d$agefin<-pmin(d$agediv,d$ageint,d$ageved, na.rm=TRUE)
d$cens<-ifelse(d$ved==1,
        ifelse (d$ageved<d$agestart,9,1),
        ifelse(d$ved==0,ifelse(is.na(d$agediv),0,
            ifelse(d$agediv<d$agestart,9,2)),9))

dx<-subset(d,d$cens<9,select=c(id,agestart,agefin,cens,weight,sex,edu,))



#_____
# TR4 -   SECOND MARRIAGE AFTER DIVORCE
transition<-"TR4 (divorced -> second marriage)"
minage<-20
maxage<-70
d$agestart<-trunc(pmax(d$agesw,d$agediv,na.rm=TRUE))
d$agefin<-pmin(d$agemarr2,d$ageint, na.rm=TRUE)
d$cens<-ifelse(d$mar2div==1,
    ifelse (d$agemarr2<d$agestart,9,1),
    ifelse(d$mar2div==0,0,9))

# Matrix Selection (d$cens not equal 9)
dtemp<-subset(d,d$cens<9,
  select=c(id,agestart,agefin,cens,weight,withch,agewith,without,
            agewithout,sex,edu,))

#splitting (with children in the household (Y/N))
i<-1
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$withch[i]==1 & (dtemp$agewith[i]>dtemp$agestart[i]
                    & dtemp$agewith[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,withch,agewith,
```

```
            # without,agewithout,sex,edu,
            dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agewith[i],dtemp$agefin[i],
                              dtemp$cens[i],dtemp$weight[i],dtemp$withch[i],
                              dtemp$agewith[i],dtemp$without[i],dtemp$agewithout[i],
                              dtemp$sex[i],dtemp$edu[i]))
            # old row
            dtemp$agefin[i]<-dtemp$agewith[i]
            dtemp$cens[i]<-0
            k<-k+1
    }
}

# Time-varying covariate   (1 Without children    2 With children)
dtemp$timed<-ifelse(dtemp$withch==1 & dtemp$agewith<=dtemp$agestart,2,1)

#Splitting for without children
i<-1
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$timed[i]==2 & dtemp$without[i]==1 &
        (dtemp$agewithout[i]>dtemp$agestart[i]
        & dtemp$agewithout[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,withch,agewith,
        # without,agewithout,sex,edu,
        dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agewithout[i],dtemp$agefin[i],
                              0,dtemp$weight[i],dtemp$withch[i],
                              dtemp$agewith[i],dtemp$without[i],dtemp$agewithout[i],
                              dtemp$sex[i],dtemp$edu[i],1))
        # old row
        dtemp$agefin[i]<-dtemp$agewithout[i]
        dtemp$cens[i]<-1
        k<-k+1
    }
}

dx<-subset(dtemp, select=c(id,agestart,agefin,cens,weight,sex,edu,timed))
remove(dtemp)


#_____
# TR5 -   SECOND MARRIAGE AFTER DEATH OF SPOUSE
transition<-"TR5 (widowed -> second marriage)"
minage<-15
maxage<-70
# Very few cases (<50): we consider 10 years before interview
d$agestart<-trunc(pmax(0,d$ageved,na.rm=TRUE))
d$agefin<-pmin(d$agemarr2,d$ageint, na.rm=TRUE)
# control: if agestart or agefin >100 then 100
d$agestart<-ifelse(d$agestart>100,100,d$agestart)
d$agefin<-ifelse(d$agefin>100,100,d$agefin)
d$cens<-ifelse(d$mar2ved==1,
     ifelse (d$agemarr2<d$agestart,9,1),
     ifelse(d$mar2ved==0,0,9))


# Matrix Selection (d$cens not equal 9)
dtemp<-subset(d,d$cens<9,
  select=c(id,agestart,agefin,cens,weight,withch,agewith,without,
           agewithout,sex,edu,))


#splitting (with children in the household (Y/N))
i<-1
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$withch[i]==1 & (dtemp$agewith[i]>dtemp$agestart[i]
                 & dtemp$agewith[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,withch,agewith,
        # without,agewithout,sex,edu,
        dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agewith[i],dtemp$agefin[i],
                              dtemp$cens[i],dtemp$weight[i],dtemp$withch[i],
                              dtemp$agewith[i],dtemp$without[i],dtemp$agewithout[i],
                              dtemp$sex[i],dtemp$edu[i]))
        # old row
        dtemp$agefin[i]<-dtemp$agewith[i]
        dtemp$cens[i]<-0
        k<-k+1
    }
}

# Time-varying covariate   (1 Without children    2 With children)
dtemp$timed<-ifelse(dtemp$withch==1 & dtemp$agewith<=dtemp$agestart,2,1)

#Splitting for without children
i<-1
l<-dim(dtemp)[1]
```

```
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$timed[i]==2 & dtemp$without[i]==1 &
            (dtemp$agewithout[i]>dtemp$agestart[i]
            & dtemp$agewithout[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,withch,agewith,
        # without,agewithout,sex,edu,
        dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agewithout[i],dtemp$agefin[i],
                            0,dtemp$weight[i],dtemp$withch[i],
                            dtemp$agewith[i],dtemp$without[i],dtemp$agewithout[i],
                            dtemp$sex[i],dtemp$edu[i],1))
        # old row
        dtemp$agefin[i]<-dtemp$agewithout[i]
        dtemp$cens[i]<-1
        k<-k+1
    }
}

dx<-subset(dtemp, select=c(id,agestart,agefin,cens,weight,sex,edu,timed))
remove(dtemp)


#_____
# TR6  -   EXIT FROM PARENTAL HOME (FOR UNION)
transition<-"TR6 (in parental home -> with a partner)"
minage<-15
maxage<-60
d$agestart<-d$agesw
d$agefin<-pmin(d$agexit,d$ageint, na.rm=TRUE) # NA value are not considered
# control: if agestart or agefin >100 then 100
d$agestart<-ifelse(d$agestart>100,100,d$agestart)
d$agefin<-ifelse(d$agefin>100,100,d$agefin)
d$cens<-ifelse(d$exitp==1,
        ifelse (d$agexit<d$agestart,9,1),
        ifelse(d$exitp==0,ifelse(is.na(d$agexit),0,
            ifelse(d$agexit<d$agestart,9,2)),
          9))

dx<-subset(d,d$cens<9,select=c(id,agestart,agefin,cens,weight,sex,edu,))

#_____
# TR7  -   EXIT FROM PARENTAL HOME (FOR OTHER REASONS)
transition<-"TR7 (in parental home -> alone or with other persons)"
minage<-11
maxage<-50
d$agestart<-d$agesw
d$agefin<-pmin(d$agexit,d$ageint, na.rm=TRUE) # NA value are not considered
# control: if agestart or agefin >100 then 100
d$agestart<-ifelse(d$agestart>100,100,d$agestart)
d$agefin<-ifelse(d$agefin>100,100,d$agefin)
d$cens<-ifelse(d$exitwp==1,
        ifelse (d$agexit<d$agestart,9,1),
        ifelse(d$exitwp==0,ifelse(is.na(d$agexit),0,
            ifelse(d$agexit<d$agestart,9,2)),9))

dx<-subset(d,d$cens<9,select=c(id,agestart,agefin,cens,weight,sex,edu,))

#_____
# TR12  -   FIRST CHILD

transition<-"TR12 (first birth)"
minage<-15
maxage<-50
d$agestart<-d$agesw
d$agefin<-pmin(d$agech1,d$ageint, na.rm=TRUE) # NA value are not considered
# control: if agestart or agefin >100 then 100
d$agestart<-ifelse(d$agestart>100,100,d$agestart)
d$agefin<-ifelse(d$agefin>100,100,d$agefin)
d$cens<-ifelse(d$ch1==1,
    ifelse (d$agech1<d$agestart,9,1),
    ifelse(d$ch1==0,0,9))

# Matrix Selection (d$cens not equal 9)
dtemp<-subset(d,d$cens<9,
  select=c(id,agestart,agefin,cens,weight,marr,agemarr,sex,edu,))

#splitting (married (Y/N))
i<-1
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$marr[i]==1 & (dtemp$agemarr[i]>dtemp$agestart[i]
                    & dtemp$agemarr[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,marr,agemarr,sex,edu,
        dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agemarr[i],dtemp$agefin[i],
                            dtemp$cens[i],dtemp$weight[i],dtemp$marr[i],
                            dtemp$agemarr[i],dtemp$sex[i],dtemp$edu[i]))
        # old row
```

```
            dtemp$agefin[i]<-dtemp$agemarr[i]
            dtemp$cens[i]<-0
            k<-k+1
    }
}
# Time-varying covariate  (1 not married  2 married)
dtemp$timed<-ifelse(dtemp$marr==1 & dtemp$agemarr<=dtemp$agestart,2,1)

dx<-subset(dtemp, select=c(id,agestart,agefin,cens,weight,sex,edu,timed))
remove(dtemp)


#_____
# TR13  -  SECOND CHILD

transition<-"TR13 (second birth)"
minage<-15
maxage<-60
d$agestart<-trunc(pmax(d$agesw,d$agech1+0.75,na.rm=TRUE))
d$agefin<-pmin(d$agech2,d$ageint, na.rm=TRUE) # NA value are not considered
# control: if agestart or agefin >100 then 100
d$agestart<-ifelse(d$agestart>100,100,d$agestart)
d$agefin<-ifelse(d$agefin>100,100,d$agefin)
d$cens<-ifelse(d$ch2==1,
      ifelse (d$agech2<d$agestart,9,1),
      ifelse(d$ch2==0,0,9))

# Matrix Selection (d$cens not equal 9)
dtemp<-subset(d,d$cens<9,
  select=c(id,agestart,agefin,cens,weight,marr,agemarr,sex,edu,))

#splitting (married (Y/N))
i<-1
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$marr[i]==1 & (dtemp$agemarr[i]>dtemp$agestart[i]
                    & dtemp$agemarr[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,marr,agemarr,sex,edu,
        dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agemarr[i],dtemp$agefin[i],
                            dtemp$cens[i],dtemp$weight[i],dtemp$marr[i],
                            dtemp$agemarr[i],dtemp$sex[i],dtemp$edu[i]))
        # old row
        dtemp$agefin[i]<-dtemp$agemarr[i]
        dtemp$cens[i]<-0
        k<-k+1
    }
}
# Time-varying covariate  (1 not married  2 married)
dtemp$timed<-ifelse(dtemp$marr==1 & dtemp$agemarr<=dtemp$agestart,2,1)

dx<-subset(dtemp, select=c(id,agestart,agefin,cens,weight,sex,edu,timed))
remove(dtemp)


#_____
# TR 14  -  THIRD CHILD

transition<-"TR14 (third birth)"
minage<-15
maxage<-60
d$agestart<-trunc(pmax(d$agesw,d$agech2+0.75,na.rm=TRUE))
d$agefin<-pmin(d$agech3,d$ageint, na.rm=TRUE) # NA value are not considered
# control: if agestart or agefin >100 then 100
d$agestart<-ifelse(d$agestart>100,100,d$agestart)
d$agefin<-ifelse(d$agefin>100,100,d$agefin)
d$cens<-ifelse(d$ch3==1,
      ifelse (d$agech3<d$agestart,9,1),
      ifelse(d$ch3==0,0,9))

# Matrix Selection (d$cens not equal 9)
dtemp<-subset(d,d$cens<9,
  select=c(id,agestart,agefin,cens,weight,marr,agemarr,sex,edu,))

#splitting (married (Y/N))
i<-1
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$marr[i]==1 & (dtemp$agemarr[i]>dtemp$agestart[i]
                    & dtemp$agemarr[i]<dtemp$agefin[i])) {
        # new row: id,agestart,agefin,cens,weight,marr,agemarr,sex,edu,
        dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agemarr[i],dtemp$agefin[i],
                            dtemp$cens[i],dtemp$weight[i],dtemp$marr[i],
                            dtemp$agemarr[i],dtemp$sex[i],dtemp$edu[i]))
        # old row
        dtemp$agefin[i]<-dtemp$agemarr[i]
        dtemp$cens[i]<-0
        k<-k+1
```

```
    }
}
# Time-varying covariate  (1 not married  2 married)
dtemp$timed<-ifelse(dtemp$marr==1 & dtemp$agemarr<=dtemp$agestart,2,1)

dx<-subset(dtemp, select=c(id,agestart,agefin,cens,weight,sex,edu,timed))
remove(dtemp)

#_____
# TR15  -  FOURTH CHILD

transition<-"TR15(fourth birth)"
minage<-20
maxage<-60
d$agestart<-trunc(pmax(d$agesw,d$agech3+0.75,na.rm=TRUE))
d$agefin<-pmin(d$agech4,d$ageint, na.rm=TRUE) # NA value are not considered
# control: if agestart or agefin >100 then 100
d$agestart<-ifelse(d$agestart>100,100,d$agestart)
d$agefin<-ifelse(d$agefin>100,100,d$agefin)
d$cens<-ifelse(d$ch4==1,
     ifelse (d$agech4<d$agestart,9,1),
     ifelse(d$ch4==0,0,9))

# Matrix Selection (d$cens not equal 9)
dtemp<-subset(d,d$cens<9,
  select=c(id,agestart,agefin,cens,weight,marr,agemarr,sex,edu,))

#splitting (married (Y/N))
i<-1
l<-dim(dtemp)[1]
k<-dim(dtemp)[1]+1
for (i in 1:l) {
    if (dtemp$marr[i]==1 & (dtemp$agemarr[i]>dtemp$agestart[i]
                 & dtemp$agemarr[i]<dtemp$agefin[i])) {
       # new row: id,agestart,agefin,cens,weight,marr,agemarr,sex,edu,
       dtemp<-rbind(dtemp, c(dtemp$id[i],dtemp$agemarr[i],dtemp$agefin[i],
                          dtemp$cens[i],dtemp$weight[i],dtemp$marr[i],
                          dtemp$agemarr[i],dtemp$sex[i],dtemp$edu[i]))
       # old row
       dtemp$agefin[i]<-dtemp$agemarr[i]
       dtemp$cens[i]<-0
       k<-k+1
    }
}
# Time-varying covariate  (1 not married  2 married)
dtemp$timed<-ifelse(dtemp$marr==1 & dtemp$agemarr<=dtemp$agestart,2,1)

dx<-subset(dtemp, select=c(id,agestart,agefin,cens,weight,sex,edu,timed))
remove(dtemp)

#_____
# TR16  -  WITHOUT CHILDREN --> WITH CHILDREN

transition<-"TR16(without children -> with children)"
minage<-15
maxage<-50
d$agestart<-d$agesw
d$agefin<-pmin(d$agewith,d$ageint, na.rm=TRUE) # NA value are not considered
# control: if agestart or agefin >100 then 100
d$agestart<-ifelse(d$agestart>100,100,d$agestart)
d$agefin<-ifelse(d$agefin>100,100,d$agefin)
d$cens<-ifelse(d$withch==1,
     ifelse (d$agewith<d$agestart,9,1),
     ifelse(d$withch==0,0,9))

dx<-subset(d,d$cens<9,select=c(id,agestart,agefin,cens,weight,sex,edu,))

#_____
# TR17  -  WITH CHILDREN --> WITHOUT CHILDREN

transition<-"TR17(with children -> without children)"
minage<-30
maxage<-90
d$agestart<-trunc(pmax(d$agesw,d$agewith,na.rm=TRUE))
d$agefin<-pmin(d$agewithout,d$ageint, na.rm=TRUE) # NA value are not considered
# control: if agestart or agefin >100 then 100
d$agestart<-ifelse(d$agestart>100,100,d$agestart)
d$agefin<-ifelse(d$agefin>100,100,d$agefin)
d$cens<-ifelse(d$without==1,
     ifelse (d$agewithout<d$agestart,9,1),
     ifelse (d$without==0,0,9))

dx<-subset(d,d$cens<9,select=c(id,agestart,agefin,cens,weight,sex,edu,))
```

2. ITERATIONS (Data matrix with three covariates: sex, education and one time-varying)

```
#SETTING PARAMETERS _____

#Number of age classes
CA<-100

# Number of covariates (fixed)
M<-3
#Number of categories for each covariates
mod1<-2
mod2<-4
mod3<-2

#INITIALIZE CONTROL VARIABLE _____
E<-mod1*mod2*mod3    # Number of combinations of covariate categories


# Initialize matrix of event, exposure, rates and covariates
event<-matrix(rep(0,CA*E),ncol=CA)
eventw<-matrix(rep(0,CA*E),ncol=CA)
exposure<-matrix(rep(0,CA*E),ncol=CA)
exposurew<-matrix(rep(0,CA*E),ncol=CA)
cases<-matrix(rep(0,CA*E),ncol=CA)
covar<-matrix(rep(0,M*E),ncol=M)

N<-dim(dx)[1]   #Number of individuals in the sample


#_____
#ITERATIONS (3 covariates)

# In each loop we have:
#   j number of individual (from 1 to N)
#   i age class (from 1 to CA)
#   z number of covariate (from 1 to M)

k<-0
for (i1 in 1:mod1)  {
  for (i2 in 1:mod2)  {
    for (i3 in 1:mod3)  {
      dsel<-subset(dx, dx$sex==i1 & dx$edu==i2 & dx$timed==i3)
      k<-k+1
      covar[k,1]<-i1
      covar[k,2]<-i2
      covar[k,3]<-i3
      if (!is.na(dsel$agestart[1])) {
          for (j in 1:length(dsel$id)) {
                        # for the j-th individual (in the subset selected)
              for(i in trunc(dsel$agestart[j]):trunc(dsel$agefin[j])) {
                        # for the i-th age class (of the j-th individual)
                  exposure[k,i]<-exposure[k,i]+1 # time of exposure increases
                  exposurew[k,i]<-exposurew[k,i]+ dsel$weight[j] # weighted time of exposure
                  cases[k,i]<-cases[k,i]+1    # number of (not weighted) cases (person-years)

                  if (dsel$cens[j]==1 & trunc(dsel$agefin[j])==i) {  # if event occurs
                      event[k,i]<-event[k,i]+1   # number of events (not weighted) increases
                      eventw[k,i]<-eventw[k,i]+dsel$weight[j]    #  number  of  events  (weighted)
increases
                      exposure[k,i]<-exposure[k,i]-(1-(dsel$agefin[j]-trunc(dsel$agefin[j])))
                                        # correction of time of exposure
                      exposurew[k,i]<-exposurew[k,i]-((1-(dsel$agefin[j]-trunc(dsel$agefin[j])))*
dsel$weight[j])
                  }
                  if (dsel$cens[j]==2 & trunc(dsel$agefin[j])==i) { # if censor occurs
                      exposure[k,i]<-exposure[k,i]-(1-(dsel$agefin[j]-trunc(dsel$agefin[j])))
                      exposurew[k,i]<-exposurew[k,i]-((1-(dsel$agefin[j]-trunc(dsel$agefin[j])))*
dsel$weight[j])
                  }
                  if (dsel$cens[j]==0 & trunc(dsel$agefin[j])==i) { # if interview occurs
                      exposure[k,i]<-exposure[k,i]-0.5
                      exposurew[k,i]<-exposurew[k,i]-(0.5 * dsel$weight[j])
                  }
              }
          }
      }
    }
  }
}


# Recomposing data matrix

matr<-matrix(rep(0,CA*E*(M+6)),ncol=M+6)
for (i in 1:CA) {
```

```
    for (k in 1:E) {
        matr[E*(i-1)+k,1]<-i
        matr[E*(i-1)+k,2]<-cases[k,i]
        matr[E*(i-1)+k,3]<-event[k,i]
        matr[E*(i-1)+k,4]<-eventw[k,i]
        matr[E*(i-1)+k,5]<-exposure[k,i]
        matr[E*(i-1)+k,6]<-exposurew[k,i]
        matr[E*(i-1)+k,7]<-covar[k,1]
        matr[E*(i-1)+k,8]<-covar[k,2]
        matr[E*(i-1)+k,9]<-covar[k,3]
    }
}
remove(dx)
require(mgcv)
```

_____

3. MODEL ESTIMATION AND PLOT (FOR WOMEN)

```
# Data selection
# select range of age & drop not-missing exposures rows
matrx<-subset(matr, matr[,1]>minage & matr[,1]<maxage
  & round(matr[,6])>0 & matr[,7]==2)
age<-matrx[,1]
event<-round(matrx[,4])
expos<-round(matrx[,6])
sex<-as.factor(matrx[,7])
edu<-as.factor(matrx[,8])
timed<-as.factor(matrx[,9])


# defining knots and age intervals
knot<-c(0,0)
prop<-cumsum(matr[,4][matr[,7]==2]/sum(matr[,4][matr[,7]==2]))
knot[1]<- round(length(prop[prop<0.33])/8)
knot[2]<- round(length(prop[prop<=0.67])/8)
int<-as.factor(ifelse(age<knot[1],1,ifelse(age>=knot[1] & age<knot[2],2,3)))

# EDUCATION (COV1) (deviation coding : comparison with baseline risk)
#               first interval        second interval           third interval
# level/dummy   d1    d2    d3         d4 d5 d6              d7   d8    d9
# 1              1    0     0           1    0    0          1    0     0
# 2              0    1     0           0    1    0          0    1     0
# 3              0    0     1           0    0    1          0    0     1
# 4             -1   -1    -1          -1 -1 -1          -1 -1  -1

# Time-dependent variables
edu11<-ifelse(edu==1 & int==1,1,ifelse(edu==4 & int==1,-1,0))
edu21<-ifelse(edu==2 & int==1,1,ifelse(edu==4 & int==1,-1,0))
edu31<-ifelse(edu==3 & int==1,1,ifelse(edu==4 & int==1,-1,0))
edu12<-ifelse(edu==1 & int==2,1,ifelse(edu==4 & int==2,-1,0))
edu22<-ifelse(edu==2 & int==2,1,ifelse(edu==4 & int==2,-1,0))
edu32<-ifelse(edu==3 & int==2,1,ifelse(edu==4 & int==2,-1,0))
edu13<-ifelse(edu==1 & int==3,1,ifelse(edu==4 & int==3,-1,0))
edu23<-ifelse(edu==2 & int==3,1,ifelse(edu==4 & int==3,-1,0))
edu33<-ifelse(edu==3 & int==3,1,ifelse(edu==4 & int==3,-1,0))

# TIME DEPENDENT VARIABLE (deviation coding : comparison with baseline risk)
#               first interval        second interval            third interval
# level/dummy       d1                    d2                    d3
# 1                  1                     1                     1
# 2                 -1                    -1                    -1

# Time dependent covariate 1 (deviation coding)
timed1<-ifelse(timed==1 & int==1,1,ifelse(timed==2 & int==1,-1,0))
timed2<-ifelse(timed==1 & int==2,1,ifelse(timed==2 & int==2,-1,0))
timed3<-ifelse(timed==1 & int==3,1,ifelse(timed==2 & int==3,-1,0))

m0<-gam(event~offset(log(expos))+ s(age),family=poisson)
baseline0<-exp(predict.gam(m0,type="terms",terms="s(age)")+m0$coeff[1])

m1<-gam(event~offset(log(expos))+ s(age)+ edu11+edu21+edu31+
edu12+edu22+edu32+edu13+edu23+edu33 ,family=poisson)
baseline1<-exp(predict.gam(m1,type="terms",terms="s(age)")+m1$coeff[1])

m<-gam(event~offset(log(expos))+ s(age)+ edu11+edu21+edu31+
edu12+edu22+edu32+edu13+edu23+edu33+timed1+timed2+timed3,family=poisson)
baseline<-exp(predict.gam(m,type="terms",terms="s(age)")+m$coeff[1])


# Plot graph (education) _____
windows(6.5,4.5)
hgraph<-max(baseline)*3
stepy<-hgraph/15
stepx<-(maxage-minage)/4
```

```
plot(c(minage,maxage),c(0,hgraph),                    main=paste(transition,"-                Women"),
    xlab="Age",ylab="Risk",type="n")
legend (maxage-stepx*1.2,hgraph/1.6,c("Baseline risk","Primary","Low.sec.","Up.sec.","Tertiary"),
        lty=1,col=c("black","red","blue","green","yellow"))
# plot first subinterval
lines(age,baseline,lwd=2)
lines(age[int==1],baseline[int==1]*exp(m$coeff[2]),col="red")
lines(age[int==1],baseline[int==1]*exp(m$coeff[3]),col="blue")
lines(age[int==1],baseline[int==1]*exp(m$coeff[4]),col="green")
#lines(age[int==1],baseline[int==1]/(exp(m$coeff[2]+m$coeff[3]+m$coeff[4])),col="yellow")
# plot second subinterval
lines(age[int==2],baseline[int==2]*exp(m$coeff[5]),col="red")
lines(age[int==2],baseline[int==2]*exp(m$coeff[6]),col="blue")
lines(age[int==2],baseline[int==2]*exp(m$coeff[7]),col="green")
lines(age[int==2],baseline[int==2]/(exp(m$coeff[5]+m$coeff[6]+m$coeff[7])),col="yellow")
# plot third subinterval
lines(age[int==3],baseline[int==3]*exp(m$coeff[8]),col="red")
lines(age[int==3],baseline[int==3]*exp(m$coeff[9]),col="blue")
lines(age[int==3],baseline[int==3]*exp(m$coeff[10]),col="green")
lines(age[int==3],baseline[int==3]/(exp(m$coeff[8]+m$coeff[9]+m$coeff[10])),col="yellow")


# write legend
text(minage,hgraph,"Level of education. Relative risks",pos=4,font=4)
text(minage,hgraph-stepy,"Age interval",pos=4,font=2)
text(minage,hgraph-2*stepy,"primary",pos=4,font=3)
text(minage,hgraph-3*stepy,"lower sec.",pos=4,font=3)
text(minage,hgraph-4*stepy,"upp. sec.",pos=4,font=3)
text(minage,hgraph-5*stepy,"tertiary",pos=4,font=3)

text(minage+stepx,hgraph-stepy,paste(minage,"-",knot[1]-1),pos=4,font=2)
text(minage+stepx,hgraph-2*stepy,round(exp(m$coeff[2]),3),pos=4)
text(minage+stepx,hgraph-3*stepy,round(exp(m$coeff[3]),3),pos=4)
#text(minage+stepx,hgraph-4*stepy,round(exp(m$coeff[4]),3),pos=4)
#text(minage+stepx,hgraph-5*stepy,round(1/exp(m$coeff[2]+m$coeff[3]+m$coeff[4]),3),pos=4)

text(minage+2*stepx,hgraph-stepy,paste(knot[1],"-",knot[2]-1),pos=4,font=2)
text(minage+2*stepx,hgraph-2*stepy,round(exp(m$coeff[5]),3),pos=4)
text(minage+2*stepx,hgraph-3*stepy,round(exp(m$coeff[6]),3),pos=4)
text(minage+2*stepx,hgraph-4*stepy,round(exp(m$coeff[7]),3),pos=4)
text(minage+2*stepx,hgraph-5*stepy,round(1/exp(m$coeff[5]+m$coeff[6]+m$coeff[7]),3),pos=4)

text(minage+3*stepx,hgraph-stepy,paste(knot[2],"-",maxage),pos=4,font=2)
text(minage+3*stepx,hgraph-2*stepy,round(exp(m$coeff[8]),3),pos=4)
text(minage+3*stepx,hgraph-3*stepy,round(exp(m$coeff[9]),3),pos=4)
text(minage+3*stepx,hgraph-4*stepy,round(exp(m$coeff[10]),3),pos=4)
text(minage+3*stepx,hgraph-5*stepy,round(1/exp(m$coeff[8]+m$coeff[9]+m$coeff[10]),3),pos=4)


# Plot graph (married ) _____
windows(6.5,4.5)
plot(c(minage,maxage),c(0,hgraph), main=paste(transition,"- Women"),
    xlab="Age",ylab="Risk",type="n")
legend (maxage-stepx*1.2,hgraph/1.6,c("Baseline risk","Not married","Married"),
        lty=1,col=c("black","red","green"))
# plot first subinterval
lines(age,baseline,lwd=2)
lines(age[int==1],baseline[int==1]*exp(m$coeff[11]),col="red")
lines(age[int==1],baseline[int==1]/(exp(m$coeff[11])),col="green")
# plot second subinterval
lines(age[int==2],baseline[int==2]*exp(m$coeff[12]),col="red")
lines(age[int==2],baseline[int==2]/(exp(m$coeff[12])),col="green")
# plot third subinterval
lines(age[int==3],baseline[int==3]*exp(m$coeff[13]),col="red")
lines(age[int==3],baseline[int==3]/(exp(m$coeff[13])),col="green")


# write legend
text(minage,hgraph,"Married or not. Relative risks",pos=4,font=4)
text(minage,hgraph-stepy,"Age interval",pos=4,font=2)
text(minage,hgraph-2*stepy,"Not married",pos=4,font=3)
text(minage,hgraph-3*stepy,"Married",pos=4,font=3)

text(minage+stepx,hgraph-stepy,paste(minage,"-",knot[1]-1),pos=4,font=2)
text(minage+stepx,hgraph-2*stepy,round(exp(m$coeff[11]),3),pos=4)
text(minage+stepx,hgraph-3*stepy,round(1/exp(m$coeff[11]),3),pos=4)

text(minage+2*stepx,hgraph-stepy,paste(knot[1],"-",knot[2]-1),pos=4,font=2)
text(minage+2*stepx,hgraph-2*stepy,round(exp(m$coeff[12]),3),pos=4)
text(minage+2*stepx,hgraph-3*stepy,round(1/exp(m$coeff[12]),3),pos=4)

text(minage+3*stepx,hgraph-stepy,paste(knot[2],"-",maxage),pos=4,font=2)
text(minage+3*stepx,hgraph-2*stepy,round(exp(m$coeff[13]),3),pos=4)
text(minage+3*stepx,hgraph-3*stepy,round(1/exp(m$coeff[13]),3),pos=4)


# Plot graph (with children) _____
```

```
windows(6.5,4.5)
plot(c(minage,maxage),c(0,hgraph), main=paste(transition,"- Women"),
        xlab="Age",ylab="Risk",type="n")
legend (maxage-stepx*1.3,hgraph/1.6,c("Baseline risk","Without children","With children"),
        lty=1,col=c("black","red","green"))
# plot first subinterval
lines(age,baseline,lwd=2)
lines(age[int==1],baseline[int==1]*exp(m$coeff[11]),col="red")
lines(age[int==1],baseline[int==1]/(exp(m$coeff[11])),col="green")
# plot second subinterval
lines(age[int==2],baseline[int==2]*exp(m$coeff[12]),col="red")
lines(age[int==2],baseline[int==2]/(exp(m$coeff[12])),col="green")
# plot third subinterval
lines(age[int==3],baseline[int==3]*exp(m$coeff[13]),col="red")
lines(age[int==3],baseline[int==3]/(exp(m$coeff[13])),col="green")


# write legend
text(minage,hgraph,"Children in the household. Relative risks",pos=4,font=4)
text(minage,hgraph-stepy,"Age interval",pos=4,font=2)
text(minage,hgraph-2*stepy,"Without children",pos=4,font=3)
text(minage,hgraph-3*stepy,"With children",pos=4,font=3)

text(minage+stepx,hgraph-stepy,paste(minage,"-",knot[1]-1),pos=4,font=2)
text(minage+stepx,hgraph-2*stepy,round(exp(m$coeff[11]),3),pos=4)
text(minage+stepx,hgraph-3*stepy,round(1/exp(m$coeff[11]),3),pos=4)

text(minage+2*stepx,hgraph-stepy,paste(knot[1],"-",knot[2]-1),pos=4,font=2)
text(minage+2*stepx,hgraph-2*stepy,round(exp(m$coeff[12]),3),pos=4)
text(minage+2*stepx,hgraph-3*stepy,round(1/exp(m$coeff[12]),3),pos=4)

text(minage+3*stepx,hgraph-stepy,paste(knot[2],"-",maxage),pos=4,font=2)
text(minage+3*stepx,hgraph-2*stepy,round(exp(m$coeff[13]),3),pos=4)
text(minage+3*stepx,hgraph-3*stepy,round(1/exp(m$coeff[13]),3),pos=4)



anova.gam(m0,m1,test="Chisq")
anova.gam(m1,m,test="Chisq")
sum(event)

#tab_f<-cbind(age[edu==1 & timed==1],round(baseline[edu==1 & timed==1],4))
```