Thailand Statistician
July 2024; 22(3): 509-532
http://statassoc.or.th
Contributed paper

Frailty-Based Competing Risks Model for the Analysis of Events in Transition to Adulthood

Jayanta Deb* **and Tapan Kumar Chakrabarty**
Department of Statistics, North-Eastern Hill University, Shillong, Meghalaya, India
*Corresponding author; e-mail: jdeb888@gmail.com

Received: 25 February 2021
Revised: 7 April 2022
Accepted: 13 April 2022

Abstract

The analysis of clustered time-to-event data is carried out using random effects models, popularly known as frailty models in the literature of event history analysis. The present work demonstrates an application of competing risks frailty model for analysing adulthood transitions that are clustered into geographical regions. Observations from the same cluster are assumed to be correlated because these usually share certain unobserved characteristics. Ignoring such correlations may lead to incorrect standard errors of the estimates of parameters of interest. Besides making the comparisons between usual competing risks model and competing risks model with frailty for analysing geographically clustered time-to-event data, important demographic and socio-economic factors that may affect the duration of transition to adulthood events namely: transition from leaving study to work and/or marriage of Indian youths are also reported in this paper. The data from the study ”The Youth in India: Situation and Needs 2006-2007” which was implemented by the International Institute for Population Sciences, Mumbai and the Population Council, New Delhi (IIPS and PC 2010), is used. The results of the analysis highlight the significant transition differentials among Indian youths by gender, place of residence, religion, caste, wealth quintile, among others. We found that after leaving study men join the work much earlier than women, and prefer to postpone their marriage. But women have higher likelihood of entering into marriage early compared to men. Rural residents have significantly higher likelihood of joining work and lower likelihood of entering into marriage compared to their urban counterparts at their early age. Wealth quintile has been observed to have a mild or no significant effect on the hazards of adulthood transitions.

Keywords: Frailty, heterogeneity, competing risks, gamma distribution, log-likelihood.

1. Introduction

Demographers in their scientific pursuits seek to understand population dynamics through the study of major vital events such as birth, death, migration, family formation, and marriage dissolution etc. All of these vital processes contribute to changes in population characteristics for societies, communities, geographical regions or nations. In recent years, a renewed interest among the researchers in demography is noticed which are particularly infused by the policy managers through the developmental political discourse.

The life course research, is one of such areas founded on a holistic theoretical framework that seeks to study various schemes of life course trajectories, their links, inter-connectedness and the causality aspects using event history analysis (EHA). EHA helps in developmental policy framing through the understanding of the effects of several background variables and contextual factors at different levels of aggregation on individual behaviour (Blossfeld, 1995).

EHA deals with the methods for the analysis of length of time from a specific time of origin until the occurrence of an event of interest or a specified endpoint. The dependent variable is the duration until event occurrence. In life course studies, one may be interested in analysing the time to leaving full time education or time to entry into a full time employment after leaving study etc. The body of methods are also known as survival analysis, hazard modelling or duration analysis. The longitudinal records of exposure period from individuals under observation focusing on the events occurring for the individuals are usually collected retrospectively from panel or cohort studies of partnerships, birth, employment and housing histories. Important applications, where such type of data arise, include life course studies of humans in demography, life insurance mathematics, epidemiology, and sociology. The interest is in modeling individual event histories, which in some disciplines is termed as micro-data approach, as opposed to the aggregate-data approach. The basic data are the times to occurrence of the events and the types of events that occur. Today, the standard approach to the analysis of such data is to use multistate models; a basic example is finite-state Markov processes in continuous time.

Typically, the dependent variable, the time or duration, has some special features in EHA. First, the durations are always positive and their distribution is positively skewed (long tail to the right). Second is the censoring due to which the exact duration of time remains unobserved or incomplete. Usually, there are individuals who have not yet experienced the event when we observe them, but may do so at an unknown time in the future.

With multi-state transition studies in EHA, multiple causes for failure of a subject is observed in many observational studies (Noordzij, 2013; Dignam, 2012; Austin, 2016). Individuals are exposed to the risk of experiencing events other than that of interest which alter the probability of experiencing the event of interest (Bernoulli, 1760). For instance, in analysing the pathways to adulthood, the probability of entering into work/labour force after having left full time study, which is of primary interest, may be altered by the occurrence of the event of early marriage in an Indian society. In a study of transition to marriage after cohabitation for the individuals in USA, Shaw (2011), reported that returning back to the state of being single, could be a potential competing event, and one might choose to determine what influences the transition of individuals from cohabitation to marriage or being single, possibly using a competing risks model. In a "risk for arrest" study by Cusick (2012), competing risks models are used to distinguish among arrests for drug-related, nonviolent, or violent crimes. An attempt was made by Lentine (2008) to identify the factors of cerebrovascular events for a group of patients post kidney transplantations. However, the estimates were found to be inconsistent (Varadhan, 2010), because individuals who die of non-cerebrovascular causes were also treated as being at risk for stroke even after their death. Ignoring such causes of failure, that may be potential modifier of the chance of occurrence of the event of principal interest, can lead to biased estimates and misleading interpretations of the analyses. The models developed to analyze the time to failure of an event of interest with multiple other causes of secondary interest are known as the Competing Risks models.

In studies with competing risks, there are several types of predictors or explanatory variables whose effects on the time to events, we wish to assess or control. Among those predictors some are observable and others are not. An examination of covariates that influence the distribution of the duration or time variable to one of the competing states from an initial state is of great importance for understanding the insights of the transition trajectory patterns across societies. Moreover, individuals in the same cluster share certain observable and unobservable characteristics and as a result duration data from the same cluster tend to be correlated. Because of such a correlation, two main problems generally occur: one of them, is the modeling of dependence in a clustered situation, and secondly, lack of fit of the usual competing risks models. The inclusion of frailties, in competing risks model can be a potential aid to both the problems. Frailty-based competing risks models can help in understanding the dependence in sequential transitions more efficiently, and can be useful in explaining some strange phenomena in the effect of covariates in competing risks models.

Moreover, event history data collected in many surveys, often provide highly complex longitudinal record of events such as leaving study, transition to work, entry into marital union and parenthood etc., with common features including multiple states and multiple types of events (competing risks)(Steele, 2004). These types of longitudinal data are usually clustered according to geographical regions due to their sampling design and thus are correlated, as these share some common attributes and characteristics. Ignoring these dependencies among the observations obtained from a clustered sampling scheme, can lead to incorrect inferences if, an usual competing risks model is used without considering a frailty-based competing risks model. Competing risks model can be used to identify only the observable important factors for duration of entry into work or marriage whereas, the random effect models for time to event data in a competing risk, which are known as Competing risks frailty models can be used to estimate the effects of both the observed and unobserved factors affecting the duration of entry into work or marriage after leaving full time study.

Over the last one and a half decade, researchers have proposed several modifications and extensions on competing risk modeling that are based on frailty. The competing risks model of the Prentice (1978) has been extended by Gorfine (2011), where a new class of frailty-based competing risks models for clustered failure times data is suggested. In 2012, Gauss-Seidel and BFGS methods are used by Tang (2012) as an iterative algorithm for stable and efficient calculation of log-normal frailties in the presence of competing risks with missing cause of failure. A hierarchical likelihood method based on sub-distribution hazard model for clustered competing risks data is developed by Ha (2016). A hierarchical likelihood approach for fitting the cause-specific proportional hazards model with a shared frailty in the presence of missing cause of failure for competing risks is proposed by Lee (2017). Furthermore, the identifiability of dependent competing risks models induced by bivariate frailty is introduced by Lee (2017).

The present article proposes a new frailty-based competing risks model to analyse the effect of various socio-demographic determinants on the transition to Work/marriage after leaving study in six Indian states. The sections are organised as follows. Subsection 2.1 gives an introduction to the dataset, Subsection 2.2 represents the motivation of this study, Subsection 3 describes the methodology part and Section 4 provides the pretest for visualising the existence of heterogeneity in the data. Section 5, describes the analytical results of the methodology applied to the Youth study dataset considered in this study. Also, some general concluding remarks are given in Section 6.

2. Background of the Present Study

The purpose of this section is to portray the backdrop of the present study. We shall first introduce the dataset used herein and later we brief the motivation behind taking up the study.

2.1. Dataset used

The unit level data from the study “The Youth in India: Situation and Needs 2006-2007” (IIPS and PC, 2010) has been used in this article. Being implemented by the International Institute for Population Sciences, Mumbai and the Population Council, New Delhi (IIPS and PC, 2010), it is the first-ever sub nationally representative study that provides data on young people’s transition to various adulthood events. Survey has been conducted in a total of six states of India namely: Andhra Pradesh, Bihar, Jharkhand, Maharashtra, Rajasthan and Tamil Nadu and these six states were purposively selected to represent the geographic variations and socio-cultural and demographic plurality within the country and together they represent two-fifths of the country’s population. It provides a wealth of evidence on married and unmarried young women and men (aged 15-24 and 15-29 respectively) from both rural and urban settings of each state. The surveys were undertaken in a phased manner and took place between January 2006 and April 2008. In all, 58,728 young people were contacted, of which a total of 50,848 married and unmarried young women and men were successfully interviewed. Using the information on time to entry into work and time to entry into marital union after living full time study from Youth Study survey, the present article has made an attempt to identify potential factors and contextual correlates accounting for such differential transitions.

2.2. Evidence of adulthood transitions

Over the past few decades the transitions to adulthood have undergone major changes in the events that occur in terms of their timing, sequence and quantum (Studer, 2018). A two stage model is considered to view the decision to live together and the decision whether to live as a cohabiting or a married couple (Yamaguchi, 1991). Alternatively, Berrington (2000) considered direct entry into marriage and entry into cohabitation to be two distinct process or competing risks, where the occurrence of the event 'entry into marriage' removes the individual from the possibility of experiencing the other event 'entry into cohabitation'. Such an approach is investigated and used by several researchers (Thornton, 1995; Liefbroer, 1992; Manting, 1994; Ermisch, 1996; Hoem, 1986). The experiment which is done, in order to explore the influence of education on cohabitation and marriage shows that, school enrolment decreases the rate of union formation and has greater effects on marriage than on cohabitation (Thornton, 1995). The competing risks model also assumes that the competing events are independent of each other. There are unobserved factors which are relevant to both the events and attempts are made to control for correlation due to shared unobserved factors by using a 'shared unmeasured risk factor model' of entry into marriage and cohabitation in USA (Hill, 1993). A multinomial logit model is used to estimate simultaneously the effect of covariate on the risk of marriage or cohabitation (Berrington, 2000).

The major determinants attending the transition from adolescence to adulthood include participation in the workforce and marriage. Transition into these adult roles is assumed to take place at certain ages and to follow a normative sequence involving leaving education, entry into work force, entry into marital union etc..

Work force participation is one of the proximate key features of youth study survey. Data suggest that over two-third of young men and one-half of young women are sometime engaged in work force. Almost all of married young men and almost two-thirds of unmarried young men are in work force compared with three-fifths and two-fifths of married and unmarried young women, respectively. Moreover, 74% of rural and 62% of urban young men are in work force, indicating that there may be rural-urban differential in transition patterns for young men. This differential is far more sharper for young women as 58% of rural young women are in work participation as against only 30% of them who work are from urban areas. Dataset reveals that 61% of unmarried and 97% of married young men respectively and 37% and 43% of unmarried and married young men respectively, are engaged in work at some point in the last 12 months period preceding the interview.

Youth study data also indicates that although most youth preferred to marry after age 18, as many as 19% of young women aged 20-24 were married before age 15, 49% before age 18 and 67% before age 20. In contrast, 7% of young men aged 20-24 are married before age 18, 16% before age 20 and 70% (approx) before age 30. But the two events namely: entry into workforce and entry into marriage are competing as the people showing significant percentages for both the events after leaving education. Though several researchers consider these events separately or use separate analysis for each events yet, though from above descriptive measures it is clear that respondents are at risk of experiencing one of these two events at first after schooling, preventing the other event from occurring at the same time. Hence, competing risks modeling is an appropriate approach to study the proximate determinants of the duration of these competing events.

Figure 1 depicts the standard competing risks multi-state model. Initially, every individual is in the initial state of "leaving study" at the time origin. The individual stays in the state until occurrence of any first event. Usually, there is one event of interest i.e. first entry into workforce, modelled by transitions into state 1, and first entry into marital union is assumed into the competing event state 2. Also, we observe that all the respondents (100%) were in the initial state of "leaving study". 58% of these individuals have experienced the event "entry into workforce", 25% of them had experienced the event "entry into marital union", and rest of them were in the origin state without experiencing any of these two events.

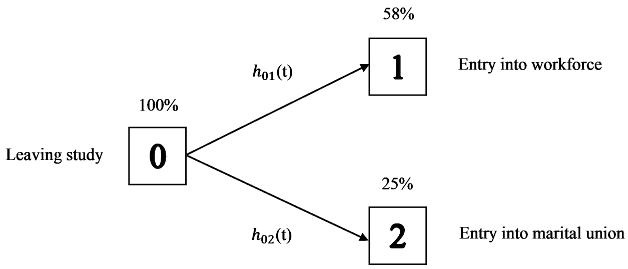

Figure 1 Competing risks multi-state model with cause specific hazards h_{0k}(t), k=1,2.

3. Methodology

The generalization of the Cox proportional hazards model (Cox, 1972) is the best to assess the observable covariate effects on the hazard function and widely applied model that allows for the random effect by multiplicatively adjusting the baseline hazard function for single cause of failure. But in case of two or multiple failures we need to consider competing risks models. In the presence of competing risks, i.e. when these two types of mutually exclusive events are occurring, a joint distribution for the time to these two different types of events can be estimated after making strong unverifiable assumption which are given as

1. Each cause leading to a particular type of event proceeds independently of every other one, at least until an event occurs.
2. The event occurs when the first of all the competing event cause reaches a destination state.
3. Each of the k event modes has a known life distribution model, where k=1 if failure occurs due to entry into job and k=2 if failure occurs due to entry into marriage.

3.1. Model specification

Let us consider there are K causes and $k = 1, 2, ..., K$ denote the k^{th} failure type, and $i = 1, 2, ..., l_v$ be the individual in a particular cluster or state v, $v = 1, 2, ..., V$. Let T_{iv} be a non-negative random variable representing the iv^{th} individual's time to failure after leaving study due to any of these two causes namely: entry into job or entry into marriage, C_{iv} be the censoring time if neither of these two events occur, and Z_{iv} be a vector of covariates. The observed time for individual i in state v is $X_{iv} = min(T_{iv}, C_{iv})$. Let e_{iv} define the event type corresponding to T_{iv} such that $e_{iv} \in 1, 2, ..., K$. The event status can subsequently be calculated as $\epsilon_{iv} = e_{iv} I(T_{iv} \geqslant C_{iv})$. The censoring status for individual i in state v can be obtained for each of the k event types such that $\delta_{ikv} = I(\epsilon_{iv} = k)$. Therefore the observed data are $(X_{iv}, \delta_{ikv}, \epsilon_{iv}, Z_{iv})$ for the iv^{th} individual for cause k. Then the proportional hazards model for the iv^{th} individual with failure type k conditional on both the covariates and frailty can be defined as,

$$h_{ikv}(t|Z_{iv}, W_{kv}) = h_{k0}(t) w_{kv} exp(Z_{iv}\beta_k) \tag{1}$$

where, $h_{k0}(t)$ is the baseline cause specific hazard for cause k at time t. Z_{iv} is a vector of covariates and β_k is the vector of regression coefficients for cause k and W_{kv} is the cause-specific shared frailty. If there is no cluster or unobserved heterogeneity present in the data then Eqn. (1) is reduced to,

$$h_k(t|Z) = h_{k0}(t) exp(Z\beta_k). \tag{2}$$

3.2. Estimation of competing risks without frailty

To deal with the data for competing risks analysis, two useful features need to be considered one, the cause specific hazard function and the other, cumulative incidence function.

Now in general, if we consider there are K causes and $k = 1, 2, ..., K$ represents the k^{th} cause ($\delta = k$ represents the cause indicator), then the cause specific hazard function $h_k(t)$ at time t, which is the instantaneous rate of failure due to cause k, conditional on survival until time t or later is defined as,

$$h_k(t) = \lim_{\Delta t \to 0} \frac{P(t \leqslant T_k \leqslant t + \Delta t \,;\, \delta = k | T_k \geqslant t)}{\Delta t}$$

Also, the cumulative incidence function denoted by $F_k(t)$, is the probability of failure due to cause k prior to time t is defined as,

$$F_k(t) = p(T \leqslant t, \delta = k), k = 1, 2, ..., K.$$

In the literature, $F_k(t)$ is also referred to as the subdistribution function because it is not a true probability distribution. It follows from these definition that,

$$F_k(t) = \int_0^t S\,(u)\, h_k(u) du = \int_0^t S\,(u)\, dH_k(u) du, k = 1, 2, ..., K$$

where, $H_k(t) = \int_0^t h_k(u)du$ is referred to as the cause-specific cumulative hazard function and $S(t) = p(T > t) = e^{-\sum_{k=1}^{K} H_k(t)}$, is the overall survival function, which is the probability of surviving beyond time t.

Let, $t_1 < t_2 < ... < t_j < ... < t_n$ be the distinct failure times from any cause. Similar to the cumulative hazard function in standard survival analysis the cumulative hazard function $H_k(t)$ for cause k can be estimated by the Nelson-Aalen estimator,

$$\widehat{H}_k(t) = \sum_{t_j \leqslant t} \frac{m_{kj}}{n_j},$$

where, m_{kj} is the number of failures from cause k and n_j is the number of subjects at risk at time t_j. The overall survival function $S(t)$ can be estimated from the Kaplan-Meier estimator $\widehat{S}(t)$. After plugging these two estimator into the equation for $F_k(t)$, the cumulative incidence function for cause k (Marubini, 2004) can be estimated as,

$$\widehat{F}_k(t) = \sum_{t_j \leqslant t} \widehat{S}_t\,(t_j - 1) \frac{d_{kj}}{n_j}$$

where, $\widehat{F}_k(t)$ is a step function that changes only at failure times t_j when d_{kj} is not zero. Also, this function is related to all of the cause-specific hazard function through $S(t)$.

To obtain the pointwise confidence intervals for the cumulative incidence function, the variance of $F_k(t)$ which was derived by Marubini (2004); Hosmer (2008) is given as,

$$\begin{aligned} Var(\widehat{F}_k(t)) = \sum_{t_j \leq t} &((\widehat{F}_k(t) - \widehat{F}_k(t_j))^2 \frac{d_j}{n_j(n_j - d_j)} + (\widehat{S}(t_{j-1}))^2 \frac{n_j - d_{kj}}{{n_j}^3} \\ &- 2(\widehat{F}_k(t) - \widehat{F}_k(t_j))(\widehat{S}(t_{j-1})) \frac{d_{kj}}{n_j}) \end{aligned}$$

where, $d_j = \sum_{k=1}^{K} d_{kj}$.

A better way to obtain confidence intervals is based on the log[-log] transformation of the estimated cumulative incidence function. Using the delta method, the standard error of $log(-log(\widehat{F}_k(t)))$ is,

$$SE(log(-log(\widehat{F}_k(t)))) = \frac{SE(\widehat{F}_k(t))}{\widehat{F}_k(t)|log\widehat{F}_k(t)|}.$$

The pointwise confidence interval of $log(-log(\widehat{F}_k(t)))$ is,

$$L = log(-log(\widehat{F}_k(t))) - Z_{\alpha/2}(SE(log(-log(\widehat{F}_k(t)))))$$

and Z_α is the $100(1-\alpha)$ percentile of the standard normal distribution. Following Kleinbaum (2005) and Putter (2007), the CIC is constructed by first estimating the hazard at ordered failure times t_j for the event type (k) of interest (Mills, 2011). This hazard estimate is simply the number of events that occur at t_j divided by the number at risk at t_j. We can write this as,

$$\widehat{h}_k(t_j) = \frac{m_{kj}}{n_j}$$

where m_{kj} is the number of events for risk k at time t_j and n_j is the number of subjects at that time. Thus, at any particular time, $\frac{m_{kj}}{n_j}$ is the estimated proportion of subjects failing from risk k. In order to be at risk for failure, the subject must have survived the previous time when a failure could have occurred. The second step entails calculating the probability of surviving the previous time t_{j-1}, which is denoted by $\widehat{S}(t_{j-1})$. $\widehat{S}(t)$ is the overall survival curve and here we calculate the overall survival rather than the cause specific survival $S_k(t)$ because the subject must have survived all other competing events.

In the third step we compute the estimated incidence of failing from event type k at time t_j, which is the probability of having survived the previous time period multiplied by the new hazard calculated in the first step. This is written as:

$$\widehat{I}_k(t_j) = \widehat{S}(t_{j-1}) \times \widehat{h}_k(t_j).$$

Finally, the CIC at time t_j is calculated as:

$$CIC(t_j) = \sum_{j'=1}^{j} \widehat{I}_k\ (t_{j'})$$

which is the cumulative sum up to time t_j of these incidence values over all event-type c failure times and is equivalent to:

$$\sum_{j'=1}^{j} \widehat{S}(t_{j-1})\, \widehat{h}_k\ (t_{j'}).$$

The overall hazard is the sum of the individual hazards for all the risk types (Kalbfleisch, 2011).

$$h(t) = h_{k1}(t) + h_{k2}(t) + ... + h_{ci}(t).$$

3.3. Competing risks with frailty

From the previous section the CIF for cause k prior to time t is

$$F_k(t) = p(T \leqslant t, \delta = k), k = 1, 2, ..., K.$$

Then for i^{th} individual in a particular cluster v, the CIF for cause k prior to time t is,

$$F_{ikv}(t) = p(T \leqslant t, \delta_{ikv} = k), i = 1, 2, ..., l_v, k = 1, 2, ..., K, v = 1, 2, ..., V.$$

Adopting a frailty modelling strategy, one might assume that,

$$p(T \leqslant t, \delta_{ikv} = k | w_{kv}) = \{B_{(kv)}(t)\}^{w_{kv}}, i = 1, 2, ..., l_v, k = 1, 2, \ldots, K, v = 1, 2, ..., V$$

where $B_{(kv)}$ are some base cumulative incidence functions and w_{kv} is a random effect or frailty term. Then,

$$F_{ikv}(t) = p(T \leqslant t, \delta_{ikv} = k) = \int \{B_{(kv)}(t)\}^{w} dF_{w_{kv}}(w) = p_{kv}[-log\{B_{(kv)}(t)\}]$$

where $F_{w_{kv}}$ is the cumulative distribution function of w_{kv} and $p_{kv}(u)$ is the Laplace transformation of w_{kv}.

3.4. Non parametric estimation

Let us consider the cause specific density as,

$$f_{ikv}(t|Z_{iv}) = h_{ikv}(t|Z_{iv})S_{iv}(t|Z_{iv}); k = 1, 2, ..., K \quad (3)$$

where, $S_{iv}(t|Z_{iv}) = e^{-\int_0^t \sum_{k=1}^{K} h_{ikv}(u|Z_{iv})du}$ Hence, the likelihood function can be written in terms of cause-specific hazards. For the cause specific survival function, we have,

$$S_{ikv}(t|Z_{iv}) = 1 - \int_0^t f_{ikv}(u|Z_{iv})du \quad (4)$$

Conditional on the clusters frailty and the observed covariates, the survival times within cluster v are assumed to be independent. At any given time $t_v = (t_{v1}, t_{v2}, ..., t_{vl_v}), w_v(t_v) = \{w_v(t_{v1})', ..., w_v(t_{vl_v})'\}'$; $v = 1, 2, ..., V$ are assumed to be independent with density denoted by $f\{w_v(t_v)|\theta(t_v)\}$ where $\theta(t)$ is a vector of unknown parameters. Now, let us consider the observed data of cluster $v = 1, 2, ..., V$ by $(T_v, Z_v, k_v, \delta_v)$, where $T_v = (T_{v1}, ..., T_{vl_v})$, $Z_v = (Z_{v1}^{'}, ..., Z_{vl_v}^{'})^{'}$, $k_v = (k_v, ..., k_{vl_v})$ and $\delta_v = \delta_{v1}, ..., \delta_{vl_v}$. Assume the censoring and failure times are independent and non-informative for frailty. Also assume that frailty is independent of the covariates. So we can get the likelihood function as Gorfine (2011),

$$L = \prod_{v=1}^{V} \prod_{t_v \geq 0} \int \prod_{i=1}^{l_v} [h_{k_{iv}0}(t_{iv}) e^{\beta_{k_{iv}}^{'} Z_{iv} + w_{vk_{iv}}(t_{iv})}]^{dN_{iv}(t_{iv})}$$
$$e^{[-\sum_{k=1}^{K} e^{\{\beta_k^{'} Z_{iv} + w_{kv}(t_{iv})\}h_{k0}(t_{iv})dt_{iv}}(\{1-dN_{iv}(t_{iv})\}Y_{iv}(t_{iv}))]}$$
$$f(w_v(t_v)|\theta(t_v))dw_v(t_v) \quad (5)$$

where,$N_{iv}(t) = \delta_{iv} I(T_{iv} \leqslant t)$ and $Y_{iv}(t) = I(T_{iv} \geqslant t)$ are failure counting process and the at-risk processes, respectively. Alternatively, let the counting process be $\{N_{iv}^k(t), t \geqslant 0\}$ be defined at time t by,

$$N_{iv}^k(t) = N_{iv}(t) I(k_{iv} = k),\ k = 1, 2, ..., K.$$

Then the above likelihood function can also be written as,

$$L = \prod_{v=1}^{V} \prod_{t_v \geq 0} \int \prod_{i=1}^{l_v} \prod_{k=1}^{K} [h_{k0}(t_{iv}) e^{\{\beta_k^{'} Z_{iv} + w_{kv}(t_{iv})\}}]^{dN_{iv}^k(t_{iv})}$$
$$e^{[-e^{\{\beta_k^{'} Z_{iv} + w_{kv}(t_{iv})\}h_{k0}(t_{iv})dt_{iv}}(\{1-dN_{iv}^{(k)}(t_{iv})\}Y_{iv}(t_{iv}))]} f(w_v(t_v)|\theta(t_v)). \quad (6)$$

Now, the non-parametric maximum likelihood estimates can be obtained by the following EM algorithm. For $k = 1, 2, ..., K$ define,

$$S_k^{(0)}(\beta_k, t) = \sum_{v=1}^{V} \sum_{i=1}^{l_v} Y_{iv}(t) e^{\beta_k^{'} Z_{iv} + w_{kv}(t)} \quad (7)$$

and

$$S_k^{(0)}(\beta_k, t) = \sum_{v=1}^{V} \sum_{i=1}^{l_v} Y_{iv}(t) Z_{iv} e^{\{\beta_k^{'} Z_{iv} + w_{kv}(t)\}}. \quad (8)$$

In M-step, we solve the complete data score equation conditional on the observed data and the current parameter estimates. Hence, for the estimation of $\beta = (\beta_1^{'}, ..., \beta_k^{'})$, we solve,

$$\sum_{v=1}^{V} \sum_{i=1}^{l_v} \int_0^{\infty} [Z_{iv} - \frac{\hat{S}_k^{(1)}(\beta_k, t)}{\hat{S}_k^{(0)}(\beta_k, t)}] dN_{iv}^{(k)}(t) = 0,\ k = 1, ..., K \quad (9)$$

where $\hat{S}_k^{(l)}(\beta_k, t); l = 0, 1$ are defined analogously to $S_k^{(l)}(\beta_k, t)$ while replacing the unknown $e^{w_{kv}(t)}$ by its conditional expectation given the observed data and the current parameters value $\hat{\mu}_{kv}(t)$. Let us consider for each $k = 1, 2, ..., K$, $w_{kv}(t) = w_{kv}^{(1)}$ if observed time is equal to or smaller than a pre-specified constant $t_0^{(k)}$, and $w_{kv}(t) = w_{kv}^{(2)}$ otherwise. Suppose $w = (w_1^{(1)}, w_1^{(2)}, ..., w_k^{(1)}, w_k^{(2)})'$ with density function $f(w|\theta)$ and unknown vector of parameter θ. Then for each cluster v, w_v is being determined by the observed data of the family and its distribution, $f(w_v|\theta v)$ and recognised as the marginal distribution of $f(w|\theta)$, since, w_v and θ_v are the sub-vectors of w and θ. Thus, for each $v = 1, 2, ..., V, i = 1, 2, ..., l_v$, and $k = 1, 2, ..., K$, let $\tilde{w}_{ikv} = w_k^{(1)} I(T_{iv} \leqslant t_0^{(k)}) + w_k^{(2)} I(T_{iv} \leqslant t_0^{(k)})$ and $\delta_{iv}^{(k)} = \delta_{iv} I(k_{iv} = k)$ and,

$$\mathcal{V}_v = e^{\sum_{i=1}^{l_v} \sum_{k=1}^{K} [\tilde{w}_{ikv} \delta_{iv}^{(k)} - e^{\{\beta_k' Z_{iv} + \tilde{w}_{ikv}\}} \Lambda_{k0}(T_{iv})]}.$$

Then the conditional expectation of $e^{w_{kv}(t)}$ becomes,

$$\frac{\int e^{w_k^{(h)}} \mathcal{V}_v f(w_k^{(h)}, w_v|\theta^2) dw_k^{(h)} dw_v}{\int \mathcal{V}_v f(w_v|\theta_v) dw_v} \tag{10}$$

with $h = 1$ if $t \leqslant t_0^{(k)}$ and $h = 2$ otherwise. Here, θ^2 is a sub-vector of θ and is determined by the joint distribution of $(w_k^{(h)}, w_k)$

$$\prod_{v=1}^{V} \prod_{t_v \geq 0} \prod_{i=1}^{l_v} \prod_{k=1}^{K} [h_{k0}(t_{iv}) e^{\{\beta_k' Z_{iv} + log_{\hat{\mu}_{kv}(t_{iv})}\}}] e^{\{-e^{\{\beta_k' Z_{iv}\}} h_{k0}(t_{iv}) dt_{iv} (1 - dN_{iv}^k(t_{iv})) Y_{iv}(t_{iv})\}}. \tag{11}$$

For the estimation of cumulative hazard function,

$$\Lambda_{k0}(t) = \int_0^t h_{k0}(u) du; \; k = 1, 2, ..., K \tag{12}$$

we define the estimator of the k^{th} cumulative baseline hazard by a step function with jumps at the observed failure times of type k and is given by,

$$\hat{\Lambda}_{k0}(t) = \int_0^t \frac{\sum_{v=1}^{V} \sum_{i=1}^{l_v} dN_{iv}^k(s)}{\hat{S}_k^{(0)}(\beta_k, s)}; \; k = 1, 2, ..., K. \tag{13}$$

To summarize,

(i) Given the values of θ and $\Lambda_{k0}; \; k = 1, 2, ..., K$ estimate $\beta_k; \; k = 1, 2, ..., K$ by solving Eqn. (9).

(ii) Given the values of θ and $\beta_k; \; k = 1, 2, ..., K$ estimate $\Lambda_{k0}; \; k = 1, 2, ..., K$ using Eqn. (13).

(iii) Given the values of β_k and $\Lambda_{k0}; \; k = 1, 2, ..., K$ estimate θ by maximizing $\sum_{v=1}^{V} \hat{E}[\log f\{w_v(.)|\theta(.)\}]$.

(iv) Repeat these three steps until convergence obtained with respect to all the parameters' estimates.

Alternatively,

(i) Estimate $e^{w_{kv}}$ by μ_{kv} which are defined as,

$$\frac{\int e^{w_k} exp[\sum_{i=1}^{l_v} \{w_k \delta_{iv}^{(k)} - exp(\beta_k' Z_{iv} + w_k) \Lambda_{k0}(T_{iv})\}] f(w_k|\theta) dw_k}{\int exp[\sum_{i=1}^{l_v} \{w_k \delta_{iv}^{(k)} - exp(\beta_k' Z_{iv} + w_k) \Lambda_{k0}(T_{iv})\}] f(w_k|\theta) dw_k}$$

with replacing the unknown parameters with current estimated values.

(ii) Estimate β_j and the parameters involved with $h_{k0}(.)$ by maximizing,

$$\prod_{v=1}^{V}\prod_{i=1}^{l_v}[h_{k0}(T_{iv})exp\{\beta_k^{'}Z_{iv}+log\hat{\mu}_{kv}\}]^{\delta_{iv}^{k}}exp[-exp\{\beta_k^{'}Z_{iv}+log\hat{\mu}_{kv}\}\Lambda_{k0}(T_{iv})].$$

(iii) Estimate $\theta=\sigma_k^2$ by,

$$\frac{\int w_k^2 exp[\sum_{i=1}^{l_v}\{w_k\delta_{iv}^{(k)}-exp(\beta_k^{'}Z_{iv}+w_k)\Lambda_{k0}(T_{iv})\}]f(w_k|\theta)dw_k}{\int exp[\sum_{i=1}^{l_v}\{w_k\delta_{iv}^{(k)}-exp(\beta_k^{'}Z_{iv}+w_k)\Lambda_{k0}(T_{iv})\}]f(w_k|\theta)dw_k} \quad (14)$$

while replacing the unknown parameters with their current estimated values. Repeat steps (i)-(iii) till convergence.

4. Exploratory Analysis of Adulthood Transitions

In this section, we carry out several exploratory and graphical analyses to examine the patterns of transitions and their typological differences across time and geographical regions.

4.1. Age specific distribution of youths by their adulthood states

While the period of transitions to adulthood and family formation is marked by discontinuation of schooling and entry into workforce or marriage for young people, some combine schooling and work, some combine schooling and marriage and others are neither in school nor in work, nor in marital union. The information, which was collected through the calendar component of Youth Study (IIPS and PC, 2010) provides us an opportunity to explore the various age specific adulthood states (viz., not studying-not working-unmarried (NS_NW_UM), not studying-not working-married (NS_NW_M), not studying-working-married (NS_W_M), not studying-working-unmarried (NS_W_UM) studying-not working-married (S_NW_M), studying-working-married (S_W_M), studying-working-unmarried (S_W_UM) and studying-not working-unmarried (S_NW_UM)) occupancy patterns in young people's lives through density plots by gender, residence and geographical regions (Ram, 2010) and these are presented in Figure 2.

An examination of these 28 Panels of Figure 2 shows that the patterns of age specific state occupancy varied widely by state, sex and place of residences. First, a significant unusual decline in the proportion of youths attending school by age 15 is observed across all groups. While 79% of urban young men, 76% of urban young women, 69% of rural young men and 59% of rural young women were in study (a very small part of them were also working) at age 12, the percentages who remained in school at age 15 fell to 55%, 50% , 43%, and 31% respectively for these categories.

Second, very few of the youths i.e, 1% or fewer urban young men, 0.7% or fewer rural young men, 1% or fewer urban young women and 0.4% or fewer rural young women reported having combined study and work together at any age, whereas, 0.5% or fewer urban young men, 0.6% or fewer rural young men, 0.9% or fewer urban young women and 0.5% or fewer rural young women reported having been continued their study after marriage at any age. Third, exit from school was accompanied by a steady rise in work participation over the ages among young men whereas, a steady rise in marital union was seen over the ages among young women. Moreover, while more young rural women than men were working at early ages (12-13), a reverse pattern was evident after age 13, and the gender gap widened with age thereafter. Fourth, at the age of 12, 6% and 1% of the urban young women were in workforce and marital union respectively whereas 18% and 2% of the rural young women were in work force and marital union respectively. After the age of 15, urban women, and after the age of 17, the rural women showed greater tendency to enter into marital life than work force. After 16 years of age, men from both areas showed a slow increasing trend in marital relations. The percentage of unmarried urban young women who were neither studying nor working, was highest at the age of 17. Finally, 38% of urban young women were married at the age of 18 but it reached

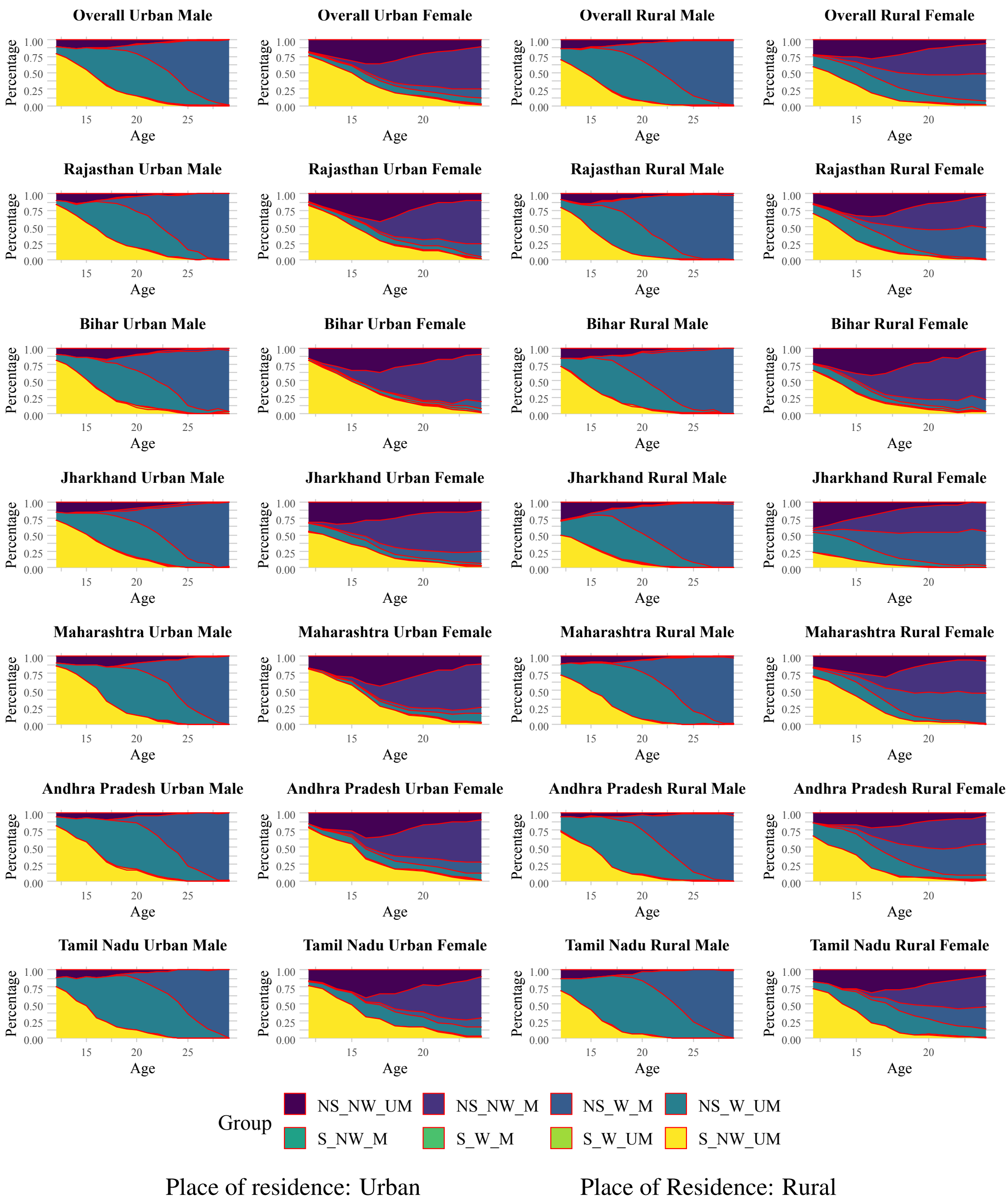

Figure 2 Respondent-type wise density plots for various states

to 79% at the age of 24 whereas, 51% of rural young women were unmarried at the age of 18 but it reached 88% at the age of 24. At the age of 21, 84% of the urban young men and 91% of the rural young men were working but when the age was about to reach 30, the percentages reached upto 100% and 98% for urban and rural young men.

Significant differences in age specific adulthood progression among the states by gender and place of rural-urban residence are clearly visible from Figure 2. 5% (lowest) of urban men in Maharashtra and 14% (highest) of urban men in the state of Tamil Nadu were in work force at age 12 whereas, 12% (lowest) of rural men in Rajasthan and 23% (highest) of rural men in Jharkhand were in workforce at age 12. Also, 79% (lowest) of urban men in Jharkhand and 89% (highest) of urban men in the state of Tamil Nadu were in work force by age 21 whereas, 87% (lowest)of rural men in Bihar and 94% (highest) of rural men in Tamil Nadu were in workforce by age 21. 17%(lowest) of urban men in Maharashtra as well as in Tamil Nadu and 35%(highest)of urban men in Bihar were in marital union by the age of 21, whereas, 22%(lowest) of rural young men in Tamil Nadu and 58% of the total rural young men in Rajasthan were in marital union by the age of 21. Moreover, in case of females, 2% (lowest) of urban women in Maharashtra and 13% (highest) of urban women in the state of Jharkhand were in work force at age 12 whereas, 11% (lowest) of rural women in Bihar and in Tamil Nadu and 34% (highest)of rural women in Jharkhand were in workforce at age 12. Percentage reached to 17%(lowest) for urban young women in Bihar, 27%(highest) urban young women in Andhra Pradesh and 18%(lowest)of rural young women in Bihar, 55%(highest)of rural women in Jharkhand were in work force by the age of 24. By the age of 18, 26%(lowest) of urban young women in Tamil Nadu, 54%(highest) of urban young women in Jharkhand and 30%(lowest) of rural young women in Tamil Nadu and 75% of rural young women in Jharkhand were married. Percentages reached to 71%(lowest) for urban young women in Maharashtra, 86%(highest) for urban young women in Rajasthan and 78%(lowest) for rural young women in Tamil Nadu and 96% for rural young women in Rajasthan as well as Jharkhand by the age of 24.

4.2. Age-wise Theil's Index for comparing state level heterogeneity

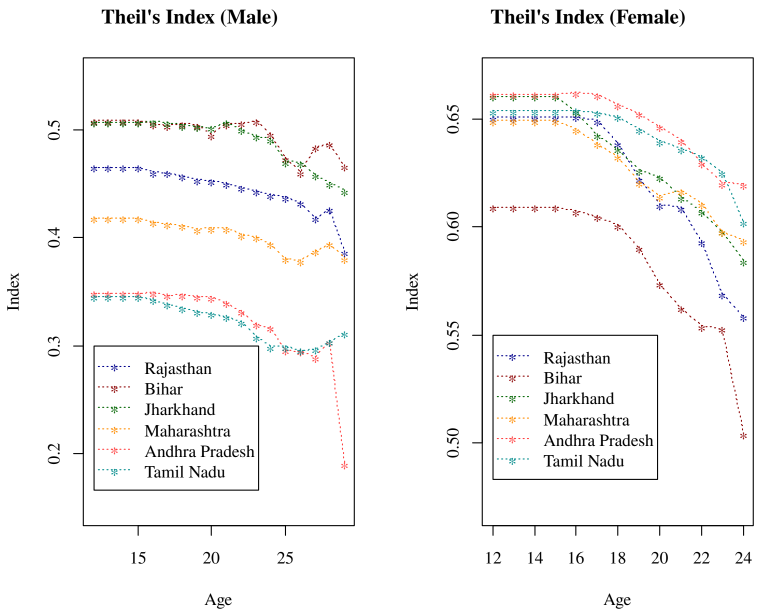

Figure 3 Respondents Age and Statewise Theil's inequality index

To figure out the degree of unobserved heterogeneity quantitatively, we have used Theil's Index (Theil, 1972) as suggested in Billari (2001) and is written as $E = \sum_{s=1}^{S} p_s log(p_s)$ where S is the number of the type of transition and p_s is the relative frequency of the transition S. One might simply examine the heterogeneity of five transitions namely; leaving study to work at same age (LSW), entry into work atleast after one year of leaving study (LS $\rightarrow$ W), leaving study to marital union at same age (LSM), entry into marriage atleast after one year of leaving study (LS $\rightarrow$ M) and left study but neither entered into work nor entered into marital union (LS), at each age and compare the distribution of their values for different states. Both the panels in Figure 3 depict that age specific values of the Theil's index are well separeted for the states and justify the use of methodology in heterogeneity analyses. Moreover, we have considered Theil's index and calculated the values for male and female separately at various ages for the six states and are presented in the following Table 1 and Table 2 respectively.

Table 1 Theil's Index for interpreting heterogeneity for male

Age	State					
	Rajasthan	Bihar	Jharkhand	Maharashtra	Andhra Pradesh	TamilNadu
Age 12	0.46513	0.50838	0.50678	0.41807	0.34863	0.34549
Age 13	0.46513	0.50838	0.50678	0.41807	0.34863	0.34549
Age 14	0.46513	0.50838	0.50678	0.41807	0.34863	0.34549
Age 15	0.46513	0.50838	0.50678	0.41807	0.34863	0.34549
Age 16	0.46038	0.50548	0.50709	0.41432	0.34939	0.34281
Age 17	0.46023	0.50367	0.50609	0.41287	0.34677	0.33848
Age 18	0.45683	0.50546	0.50408	0.41156	0.34729	0.33546
Age 19	0.45338	0.50410	0.50265	0.40772	0.34546	0.33167
Age 20	0.45239	0.49556	0.50206	0.40881	0.34502	0.32959
Age 21	0.45004	0.50515	0.50678	0.40901	0.33961	0.32682
Age 22	0.44640	0.50601	0.49959	0.40279	0.33178	0.32241
Age 23	0.44307	0.50850	0.49396	0.40053	0.31986	0.30858
Age 24	0.43960	0.49608	0.49137	0.39434	0.31679	0.29928
Age 25	0.43771	0.47326	0.47031	0.38026	0.29574	0.29924
Age 26	0.43215	0.46053	0.46961	0.37921	0.29496	0.29615
Age 27	0.41844	0.48401	0.45803	0.38756	0.28949	0.29675
Age 28	0.42648	0.48715	0.45058	0.39430	0.30429	0.30389
Age 29	0.38695	0.46580	0.44376	0.38106	0.19028	0.31161

Table 2 Theil's Index for interpreting heterogeneity for female

Age	State					
	Rajasthan	Bihar	Jharkhand	Maharashtra	Andhra Pradesh	TamilNadu
Age 12	0.65066	0.60907	0.66039	0.64922	0.66144	0.65372
Age 13	0.65066	0.60907	0.66039	0.64922	0.66144	0.65372
Age 14	0.65066	0.60907	0.66039	0.64922	0.66144	0.65372
Age 15	0.65066	0.60907	0.66039	0.64922	0.66144	0.65372
Age 16	0.65093	0.60728	0.65346	0.64525	0.66216	0.65394
Age 17	0.64918	0.60466	0.64237	0.63848	0.66117	0.65284
Age 18	0.63861	0.60079	0.63586	0.63250	0.65665	0.65099
Age 19	0.62219	0.59019	0.62626	0.62074	0.65227	0.64510
Age 20	0.61038	0.57379	0.62295	0.61428	0.64646	0.63975
Age 21	0.60897	0.56230	0.61374	0.61673	0.64019	0.63615
Age 22	0.59327	0.55413	0.60714	0.61094	0.62941	0.63250
Age 23	0.56908	0.55281	0.59778	0.59779	0.62023	0.62533
Age 24	0.55849	0.50449	0.58452	0.59386	0.61974	0.60212

Figure 3 shows the differences among the states in terms of values of the index. Among all states, Bihar and Jharkhand have shown the highest level of diversity among these five transitions, whereas the states of Andhra Pradesh and Tamil Nadu have the lowest level of diversity. Similarly, Andhra Pradesh and Bihar have the maximum and minimum degree of diversity for the women respectively in respect of adulthood transitions.

4.3. Kaplan-Meier failure estimates of entry into workforce and entry into marriage

Kaplan-Meier (Kaplan, 1958) failure estimates were used to visualize the failure patterns of both the events i.e. entry into work and entry into marital union after leaving study. Figure 4 depicts that patterns of entry into work as well as marriage are not only different by place of rural-urban residence, but also considerably by gender and geographical regions. The timing at experiencing entry into work and marriage generated different patterns specific to different groups of population (Mejia, 2012).

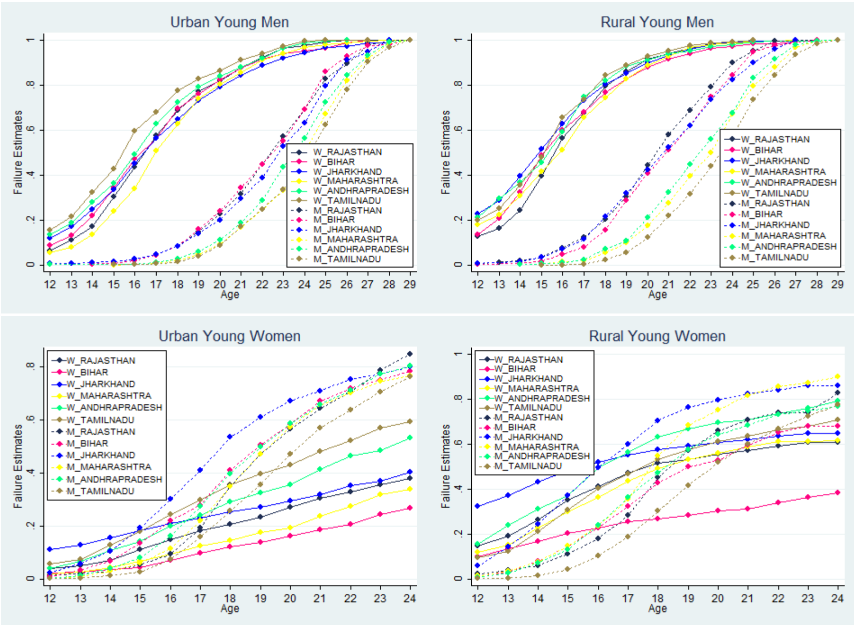

Figure 4 Kaplan-Meier failure estimates for six states by sex and place of residence-wise

We notice that failure curves of entering into work and marriage are well separated both for urban and rural young men indicating that men join work much early in their life before entering into marital union across all states. Contrary to that, the failure curves for rural and urban young women are not well separated and cross each other across the states. After the age of 15, almost all the failure curves of entering into marriage surpass those of the entering into work, depicting how young women from these states enter early into marital life after leaving their schools.

Among all the states, young men of Tamil Nadu have been witnessed to enter into workforce earlier than those of all other states across all ages and also postpone their entry into marital unions. Young men from the state of Jharkhand is observed to enter into marriage much earlier followed by Bihar and Rajasthan as compared to other states.

In case of females, Jharkahnd's urban and rural young women showed early entry into marital union than that of other states, whereas, Tamil Nadu's rural and urban young women delayed their entry into marriage than that of other states. Jharkhand women are showing remarkable early entry into workforce than the women in other states, whereas, Bihar women are delaying entry into workforce than the women in other states. After 15 years of age urban women in Andhra Pradesh and Tamil Nadu entered into work faster than that of any other states. In the rural areas, after 16 years of age, women in Andhra Pradesh, and after 19 years of age, women in Tamil Nadu showed faster trend in entry into work in comparison to others. Summarily, There was a significant delay in entry into work for women rather than men but, in case of marital union, results are reversed. If we compare women only, then we can see that women in rural areas are entering into work earlier than urban women.

4.4. Proportion of young men and women having followed different trajectories

Since the Kaplan-Meier failure estimates produced cumulative proportions of transitions at a given age, so the estimates provide patterns that did not consider individual trajectories between the transition from leaving education to entry into work or marriage. Table 3 displays the different trajectories achieved by age 12-29 from leaving education to entry into work as well as entry into marriage by states, respondent type and place of residence. The first trajectory includes respondents that left education and subsequently entered the work ($E \rightarrow W$); the second trajectory is that in which leaving study and entry into workforce occurred during the same year of age (EW simultaneously); the third trajectory includes respondents that left education and subsequently entered the marital union ($E \rightarrow M$); the fourth trajectory is that in which leaving study and entry into marriage occurred during the same year of age (EM simultaneously). Finally, the last sequence correspond to those respondents who did experience neither of these two social transitions and were in initial state leaving study(LS).

Two genders, two areas and six states were considered for this analysis along with five possible outcomes. This means that there are upto 120 different results to look at. Therefore the main patterns that come out on the analysis are summarized as follows.

Table 3 shows important differences between urban and rural young men as well as urban and rural young women in the experience of social trajectories. The probability of entry into work within the same year is significantly observed among rural young men than the urban men. If we compare the females only, we see that, rural young women showed a greater tendency to enter into work rather than that of urban women although women from both areas showed greater tendency to go to marital union after leaving education. Where a little percentage of them were in home. In case of rural females, Tamil Nadu's and Andhra's women show the higher percentages in work force, Rajasthan's women show the higher percentages in work and at home both, Bihar's women show the higher percentages in marital union and at home and lastly, Jharkhand's and Maharashtra's women show their tendency to go to work and marital life at the same time. But in the case of urban women, only two trends have been observed, one of them being in the house, the other is marital union. Finally, men have shown tendency to enter into workforce within a year of leaving the education or within a few years.

Table 3 Proportions of young men and women having followed different social trajectories by gender and place of residence

State	Respondent type	Male		Female	
	Place of residence	Urban	Rural	Urban	Rural
Rajasthan	LS→W	26%	26%	11%	12%
	LSW(Simul)	63%	59%	12%	29%
	LS→M	3%	5%	34%	19%
	LSM(Simul)	5%	6%	9%	7%
	Initial State(LS)	3%	4%	35%	32%
	N	699	906	980	1138
	Total	100%	100%	100%	100%
Bihar	LS→W	32%	30%	7%	8%
	LSW(Simul)	51%	55%	7%	15%
	LS→M	6%	6%	35%	22%
	LSM(Simul)	3%	4%	14%	9%
	Initial State(LS)	9%	6%	38%	46%
	N	481	435	865	732
	Total	100%	100%	100%	100%
Jharkhand	LS→W	29%	32%	9%	11%
	LSW(Simul)	52%	55%	14%	33%
	LS→M	6%	5%	34%	31%
	LSM(Simul)	3%	4%	13%	11%
	Initial State(LS)	10%	5%	30%	14%
	N	976	812	1293	2090
	Total	100%	100%	100%	100%
Maharashtra	LS→W	35%	23%	10%	10%
	LSW(Simul)	55%	69%	11%	27%
	LS→M	3%	2%	40%	32%
	LSM(Simul)	1%	2%	9%	12%
	Initial State(LS)	7%	4%	30%	20%
	N	942	775	1268	1401
	Total	100%	100%	100%	100%
Andhra Pradesh	LS→W	24%	15%	15%	16%
	LSW(Simul)	67%	79%	19%	39%
	LS→M	1%	1%	34%	21%
	LSM(Simul)	1%	1%	9%	7%
	Initial State(LS)	7%	5%	24%	17%
	N	955	865	1153	1438
	Total	100%	100%	100%	100%
Tamil Nadu	LS→W	29%	25%	24%	23%
	LSW(Simul)	66%	70%	22%	32%
	LS→M	1%	1%	28%	23%
	LSM(Simul)	1%	0%	5%	5%
	Initial State(LS)	4%	5%	21%	18%
	N	1043	1151	1344	2068
	Total	100%	100%	100%	100%

5. Results and Discussion

Using the time to event data from the youth study, we defined the duration by the age difference of leaving study and entry into workforce or marriage for this analysis. A total number of 25810 respondents were selected for the analysis, of which 10040 were men and 15770 were women. 21378 of them experienced both the events, of which 14854 experienced the event entry into workforce and rest of them experienced entry into marital union. In this analysis a small part of the observations were found as censored i.e. respondents who neither experienced entry into work nor entry into marital union after leaving their study. In addition to the commonly used competing risks model for analysing time to event data, competing risks with frailty model is also considered for examining the effects of different demographic and socio-economic factors on the duration simultaneously for both the causes. Competing risks with frailty models is also considered for adjusting the heterogeneity in the transitions in various geographical regions i.e. states and the gamma distribution is used as the frailty distribution for the states. In both the models same set of covariates are used which are, place of residence, religion, caste, work status of father, work status of mother, father's education, mother's education, result of last exam, type of school last attended and total number of siblings. Moreover, wealth quintle is also considered as an important covariate because the economic condition of a family can significantly affect the timing of transitions of the individuals. For analytical purpose, the most versatile statistical software packages like R, SPSS and STATA have been used.

Table 4 Parameter estimates using CIF for transition from leaving study to work or marriage without frailty for male

Covariate	category	exp(coef)	se(coef)	p-value	lower.95	upper.95
Place of residence	Rural(Work)	1.04763	0.02340	0.047*	1.00070	1.09680
	Rural(Marriage)	1.22855	0.09649	0.033*	1.01690	1.48430
Religion	Muslim(Work)	0.97559	0.03469	0.476	0.91150	1.04420
	Muslim(Marriage)	0.55755	0.16473	0.000*	0.40370	0.77000
	Others(Work)	1.00175	0.04803	0.971	0.91170	1.10060
	Others(Marriage)	0.91664	0.19899	0.662	0.62060	1.35390
Caste	OBC(Work)	1.04431	0.02648	0.102	0.99150	1.09990
	OBC(Marriage)	0.91444	0.10611	0.399	0.74270	1.12580
	GEN(Work)	1.03501	0.03486	0.324	0.96670	1.10820
	GEN(Marriage)	0.92239	0.14270	0.571	0.69730	1.22000
	Others(Work)	1.04600	0.08578	0.600	0.88410	1.23750
	Others(Marriage)	0.93047	0.34889	0.836	0.46960	1.84360
Father's education	Literate(Work)	0.96612	0.02382	0.148	0.92210	1.01230
	Literate(Marriage)	1.18734	0.09712	0.077	0.98150	1.43630
Mother's education	Literate(Work)	1.01355	0.02740	0.623	0.96050	1.06950
	Literate(Marriage)	0.58452	0.12512	0.000*	0.45740	0.74700
Work status of father	Yes(Work)	0.96896	0.02278	0.166	0.92660	1.01320
	Yes(Marriage)	1.02065	0.09420	0.828	0.84860	1.22760
Work status of mother	Yes(Work)	1.11451	0.02363	0.000*	1.06410	1.16730
	Yes(Marriage)	0.63216	0.10914	0.000*	0.51040	0.78290
Result of last exam	Pass(Work)	0.97912	0.02502	0.399	0.93230	1.02830
	Pass(Marriage)	1.25876	0.11421	0.044*	1.00630	1.57460
Type of school last attended	Govt.(Work)	1.03607	0.02967	0.232	0.97750	1.09810
	Govt.(Marriage)	0.84163	0.11763	0.143	0.66830	1.05990
	Others(Work)	0.86520	0.04780	0.002*	0.78780	0.95020
	Others(Marriage)	0.56064	0.21024	0.006*	0.37130	0.84650
Wealth quintile	3^{rd},4^{th}& 5^{th}(Work)	1.06196	0.02642	0.023*	1.00840	1.11840
	3^{rd},4^{th}& 5^{th}(Marriage)	1.18330	0.10823	0.120	0.95710	1.46290
Total no of siblings	Num(Work)	0.98428	0.00621	0.011*	0.97240	0.99630
	Num(Marriage)	1.00989	0.02540	0.699	0.96080	1.06140
Log-likelihood	-81795.56					
AIC	163651.1					

Table 5 Parameter estimates using CIF for transition from leaving study to work or marriage with frailty for male

Covariate	category	exp(coef)	se(coef)	p-value	lower.95	upper.95
Place of residence	Rural(Work)	1.03410	0.02358	0.160	0.98740	1.08300
	Rural(Marriage)	1.21030	0.09645	0.048*	1.00180	1.46210
Religion	Muslim(Work)	0.97620	0.03476	0.490	0.91190	1.04500
	Muslim(Marriage)	0.55790	0.16484	0.000*	0.40390	0.77070
	Others(Work)	1.03180	0.04875	0.520	0.93780	1.13530
	Others(Marriage)	0.95420	0.19894	0.810	0.64610	1.40920
Caste	OBC(Work)	1.04540	0.02667	0.096	0.99210	1.10150
	OBC(Marriage)	0.91530	0.10595	0.400	0.74370	1.12660
	GEN(Work)	1.04340	0.03601	0.240	0.97230	1.11980
	GEN(Marriage)	0.93620	0.14291	0.640	0.70750	1.23890
	Others(Work)	1.06440	0.08870	0.480	0.89460	1.26660
	Others(Marriage)	0.93910	0.34982	0.860	0.47310	1.86410
Father's education	Literate(Work)	0.98430	0.02416	0.510	0.93880	1.03210
	Literate(Marriage)	1.21130	0.09722	0.049*	1.00110	1.46560
Mother's education	Literate(Work)	1.00930	0.02796	0.740	0.95550	1.06620
	Literate(Marriage)	0.58030	0.12529	0.000*	0.45400	0.74190
Work status of father	Yes(Work)	0.96700	0.02284	0.140	0.92460	1.01120
	Yes(Marriage)	1.01930	0.09422	0.840	0.84740	1.22600
Work status of mother	Yes(Work)	1.09080	0.02409	0.000*	1.04050	1.14350
	Yes(Marriage)	0.61610	0.10925	0.000*	0.49730	0.76320
Result of last exam	Pass(Work)	0.97910	0.02521	0.400	0.93190	1.02870
	Pass(Marriage)	1.26250	0.11388	0.041*	1.01000	1.57820
Type of school last attended	Govt.(Work)	1.02970	0.02988	0.330	0.97120	1.09180
	Govt.(Marriage)	0.84060	0.11767	0.140	0.66740	1.05860
	Others(Work)	0.89720	0.04891	0.027*	0.81520	0.98750
	Others(Marriage)	0.58390	0.20981	0.010*	0.38700	0.88090
Wealth quintile	3^{rd},4^{th}& 5^{th}(Work)	1.01370	0.02757	0.620	0.96040	1.07000
	3^{rd},4^{th}& 5^{th}(Marriage)	1.13040	0.10837	0.260	0.91410	1.39790
Total no of siblings	Num(Work)	0.99310	0.00652	0.280	0.98050	1.00580
	Num(Marriage)	1.01860	0.02546	0.470	0.96900	1.07070
Frailty				0.000		
Variance	0.03632793					
Log-likelihood	-81774.01					
AIC	163608					

Table 4 (male) and Table 6 (female) show the estimated hazard ratios with standard errors and confidence intervals corresponding to the parameters and the corresponding p-values for the Competing risk model without frailty. We discuss first the effects in non-frailty Competing risks model.

Examining the total number of siblings in the model, we see that, with a hazard ratio of 0.98428 for male and 0.96020 for female, each sibling decreases the hazard of entry into workforce significantly. Wealth quintile found to be an important covariate for the transition to work as the analysis shows that the 3^{rd}, 4^{th} and 5^{th} quintile has significantly higher likelihood of experiencing entry to workforce compared to 1^{st} and 2^{nd} quintile (for male), but has a significantly lower effect on entry into work compared to 1^{st} and 2^{nd} quintile for female youths.

The results of the analysis show that place of residence is another important factor for both the adulthood transitions, where respondents from rural areas have 1.04763 and 1.43750 times higher likelihood of entry into workforce for male and female respectively, and 1.22855 times higher likelihood of entry into marriage for men and a lower hazard of experiencing marital union (insignificant), compared to that of urban regions of these six states.

Religion shows significant effect on entry into workforce (for females) and marriage (for both males and females) where Muslims have a significantly lower likelihood of entry into workforce (for females) as well as entry into marital union (for both males and females) than that of Hindus. Respondents in other categories are associated with an increased hazard of entry into workforce and

Table 6 Parameter estimates using CIF for transition from leaving study to work or marriage without frailty for female

Covariate	category	exp(coef)	se(coef)	p-value	lower.95	upper.95
Place of residence	Rural(Work)	1.43750	0.03023	0.000*	1.35480	1.52520
	Rural(Marriage)	0.98590	0.02806	0.614	0.93320	1.04170
Religion	Muslim(Work)	0.80220	0.04974	0.000*	0.72770	0.88430
	Muslim(Marriage)	0.71520	0.04049	0.000*	0.66060	0.77420
	Others(Work)	1.30970	0.04409	0.000*	1.20130	1.42790
	Others(Marriage)	0.75070	0.05972	0.000*	0.66780	0.84390
Caste	OBC(Work)	0.93830	0.03161	0.044*	0.88190	0.99830
	OBC(Marriage)	1.13360	0.03456	0.000*	1.05940	1.21310
	GEN(Work)	0.68950	0.04675	0.000*	0.62920	0.75570
	GEN(Marriage)	1.25590	0.04197	0.000*	1.15670	1.36360
	Others(Work)	0.60960	0.12615	0.000*	0.47610	0.78060
	Others(Marriage)	1.48450	0.09479	0.000*	1.23280	1.78760
Father's education	Literate(Work)	0.89900	0.02888	0.000*	0.84950	0.95130
	Literate(Marriage)	0.98470	0.02886	0.592	0.93050	1.04200
Mother's education	Literate(Work)	0.96140	0.03422	0.250	0.89900	1.02810
	Literate(Marriage)	0.92540	0.03155	0.014*	0.86990	0.98440
Work status of father	Yes(Work)	0.97680	0.03116	0.451	0.91890	1.03830
	Yes(Marriage)	1.05260	0.03068	0.095	0.99110	1.11780
Work status of mother	Yes(Work)	1.92670	0.02847	0.000*	1.82220	2.03730
	Yes(Marriage)	0.88550	0.02798	0.000*	0.83820	0.93540
Result of last exam	Pass(Work)	1.03080	0.03447	0.379	0.96340	1.10280
	Pass(Marriage)	1.07690	0.03505	0.034*	1.00540	1.15350
Type of school last attended	Govt.(Work)	1.03900	0.04197	0.362	0.95700	1.12810
	Govt.(Marriage)	0.85880	0.03664	0.000*	0.79930	0.92270
	Others(Work)	1.15460	0.05672	0.011*	1.03310	1.29040
	Others(Marriage)	0.94520	0.05491	0.305	0.84880	1.05260
Wealth quintile	3^{rd},4^{th}& 5^{th}(Work)	0.86810	0.03059	0.000*	0.81760	0.92180
	3^{rd},4^{th}& 5^{th}(Marriage)	1.04350	0.03291	0.196	0.97830	1.11300
Total no of siblings	Num(Work)	0.96020	0.00779	0.000*	0.94570	0.97500
	Num(Marriage)	1.02230	0.00727	0.002*	1.00790	1.03700
Log-likelihood	-105857.3					
AIC	211774.6					

a decreased hazard of marriage after leaving education. Compared to SC/ST categories, OBC, GEN and Other categories show a higher likelihood of entry into workforce for males and a significantly higher likelihood of marital union in case of females.

Female youths having literate parents, have significantly lower likelihood of entry into work as well as marriage than their counterparts having illiterate parents. Male Respondents, whose mothers are literate, have higher likelihood of entry into work. For all respondents, whose fathers do work have a significantly lower likelihood of entry into work and a significantly higher likelihood of marriage than those whose fathers don't work. Among those whose mothers are working, the tendency to enter work before marriage is observed. Women respondents, those who passed their last examinations, have higher likelihood of entering into work or marriage compared to those who failed. In case of men, those who have failed in their last exam they are showing a higher likelihood of entering into workforce after leaving education.

In addition to the general competing risks model, competing risks model with frailty is used to examine the effects of all the covariates discussed above in the presence of heterogeneous clusters/states (cluster/frailty term) on the hazards of transitions. Estimated hazard ratios with confidence interval and p-values for competing risks with frailty models are also reported in Table 5 and Table 7 for males and females respectively. Results show that the frailty terms are statistically significant and the competing risks model with frailty are found to be the superior as indicaded by their AIC values for analysing the adulthood transitions in Indian scenario.

Table 7 Parameter estimates using CIF for transition from leaving study to work or marriage with frailty for female

Covariate	category	exp(coef)	se(coef)	p-value	lower.95	upper.95
Place of residence	Rural(Work)	1.43820	0.03036	0.000*	1.35510	1.52640
	Rural(Marriage)	0.99000	0.02826	0.720	0.93670	1.04640
Religion	Muslim(Work)	0.80480	0.04980	0.000*	0.72990	0.88730
	Muslim(Marriage)	0.71260	0.04053	0.000*	0.65820	0.77160
	Others(Work)	1.26690	0.04451	0.000*	1.16110	1.38240
	Others(Marriage)	0.73010	0.05996	0.000*	0.64910	0.82110
Caste	OBC(Work)	0.94810	0.03171	0.093	0.89100	1.00890
	OBC(Marriage)	1.15120	0.03467	0.000*	1.07550	1.23210
	GEN(Work)	0.67740	0.04795	0.000*	0.61660	0.74420
	GEN(Marriage)	1.24020	0.04296	0.000*	1.14010	1.34920
	Others(Work)	0.60390	0.12834	0.000*	0.46960	0.77660
	Others(Marriage)	1.47610	0.09754	0.000*	1.21920	1.78710
Father's education	Literate(Work)	0.91170	0.02905	0.002*	0.86130	0.96520
	Literate(Marriage)	1.00210	0.02905	0.940	0.94660	1.06080
Mother's education	Literate(Work)	0.97690	0.03456	0.500	0.91290	1.04540
	Literate(Marriage)	0.93890	0.03191	0.048*	0.88200	0.99950
Work status of father	Yes(Work)	0.97530	0.03122	0.420	0.91740	1.03680
	Yes(Marriage)	1.05130	0.03073	0.100	0.98980	1.11650
Work status of mother	Yes(Work)	1.91190	0.02879	0.000*	1.80700	2.02290
	Yes(Marriage)	0.87940	0.02837	0.000*	0.83190	0.92970
Result of last exam	Pass(Work)	1.03150	0.03465	0.370	0.96380	1.10400
	Pass(Marriage)	1.08660	0.03524	0.018*	1.01410	1.16430
Type of school last attended	Govt.(Work)	1.04390	0.04211	0.310	0.96120	1.13370
	Govt.(Marriage)	0.86600	0.03676	0.000*	0.80580	0.93070
	Others(Work)	1.11200	0.05903	0.072	0.99050	1.24840
	Others(Marriage)	0.91310	0.05698	0.110	0.81660	1.02100
Wealth quintile	3^{rd},4^{th}& 5^{th}(Work)	0.85960	0.03128	0.000*	0.80850	0.91390
	3^{rd},4^{th}& 5^{th}(Marriage)	1.03220	0.03354	0.350	0.96650	1.10230
Total no of siblings	Num(Work)	0.96270	0.00805	0.000*	0.94760	0.97800
	Num(Marriage)	1.02540	0.00752	0.001*	1.01040	1.04060
Frailty				0.000		
Variance	0.5443056					
Log-likelihood	-105822.2					
AIC	211704.4					

Except for the covariate Father's education, the sign of the estimates are found to be the same for both the models, but the size of the estimates along with confidence interval and p-values are found to be different for some covariates considered in the models. After adjusting the heterogeneity due to states, the analysis of the competing risks with shared frailty shows that there is a significant effect of this covariate on the likelihood of marriage among the male respondents, which was otherwise insignificant in the previous model without frailty. This may be due to the strong correlation between education of father and the duration of entry into marriage. It is also found that, women whose fathers are literate have (90% for without frailty and 91% for with frailty) lower likelihood of entry into workforce compared to those with illiterate fathers. After incorporating frailty, it is observed that, male respondents with literate parents have significantly higher likelihood of experiencing marital union as compared to their counterparts with illiterate parents. Moreover, in case of male, the effect of wealth quintile and total number of siblings on the duration of entry into workforce and the effect of fathers education on the hazards of transition to marriage are found to be insignificant. The effect of caste(OBC) and the effect of the type school (other than private and government) on the duration of transition to work are found to be insignificant.

Since the size of the variance parameter is large for shared frailty (in case of females) it indicates that clusters are heterogeneous with respect to the risk of entry into work/marriage. In our analysis, substantial estimates of variance parameter of sizes 0.03633 (male) and 0.54431 (female) are obtained

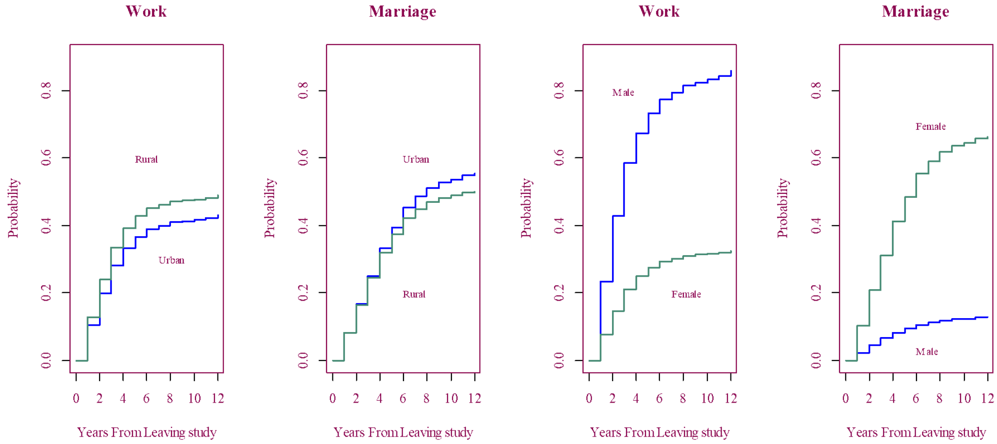

Figure 5 Cumulative Incidence Function for work and marriage covariate: place of residence and respondent type

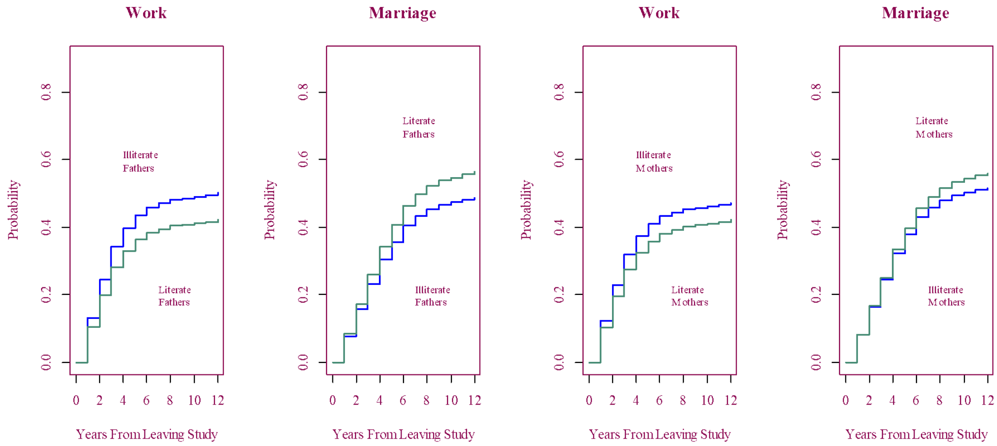

Figure 6 Cumulative Incidence Function for work and marriage covariate: education of father and mother

for the frailty term 'state'. Furthermore, if we compare the AIC for the two models, we see that it is 163651.1 and 163608 (with frailty) and 211774.6 and 211704.4 (with frailty) for male and female respectively, which means that the frailty term makes a significant contribution.

Finally, it is also possible to produce and plot the cumulative incidence function for various covariates which are shown in Figure 5, Figure 6, Figure 7 and Figure 8. The left panel of all the plots show the clear impact of various covariates on hazards of transition to workforce and the right panels show the impact of various covariates on hazards of transition to marriage. X-axis represents the time (in years) from leaving study and Y-axis represents the probability. It can be seen that, the impact of each covariate on workforce and marriage is different. As the time moves on, it is observed that hazards of transitions to work are more for respondents from rural areas and for male youths, whereas the opposite picture is seen for the hazards of transitions to marriage. The roles of literacy of parents are observed to act in a reverse way for the transitions to work and marriage. Illiteracy of

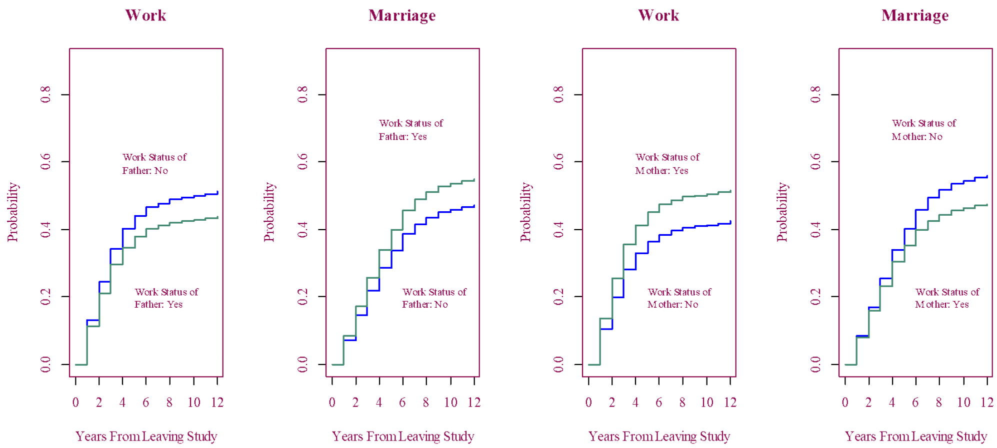

Figure 7 Cumulative Incidence Function for work and marriage covariate: work status of father and mother

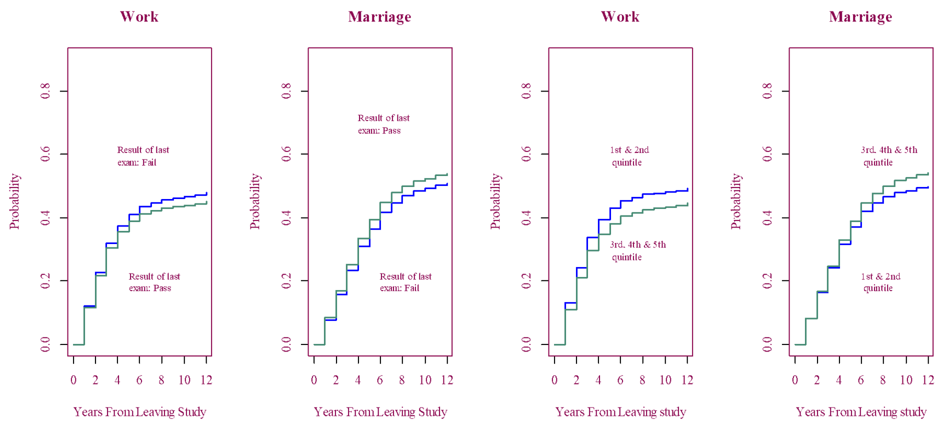

Figure 8 Cumulative Incidence Function for work and marriage covariate: result of last exam and wealth quintile

parents increases the hazards of transition to work but reduces the hazards of transition to marriage. Respondents whose fathers are non-working have increased probability of joining work and reduced probability of marriage; whereas, the respondents with working-mothers have increased probability of joining work and reduced probability of marriage. Also, failure in the examination increases the chances of entering into work and decreases the probability of marriage compared to the individuals who passed out their examinations. Last but most importantly, wealth quintle plays a significant role in explaining the hazards of transitions to the adulthood events. Respondents from lower quintiles have higher likelihood of joining work early compared to their counter parts in the upper quintiles. But the reverse is the effects of the wealth quintile in case of transition to marriage. Respondents from the lower quintiles prefer to postpone their entry into marriage as compared to their counter parts from upper quintiles.

6. Conclusions

Timing of transitions to work and marriage are considered to be important indicators for describing the overall labour force pattern and growth of population level of a country. Competing risks model is used to examine the effects of various covariates on the duration of competing events, but geographical regions can produce unobserved heterogeneity in the data which can't be captured through an usual competing risks model. In this analysis, states are considered as clusters and it is assumed that time of transition for competing events for male and female residing in the same clusters are correlated because they share the same environment. Some of the covariates, such as gender, place of residence, parental education and work, religion and caste are found to have significant effects on the hazards of transition to work/marriage for both the respondent types. Results with frailty model in case of males show that the effect of wealth quintile on the duration of both entry into workforce and entry into marriage are insignificant which was not being observed by the usual competing risks model. However, the wealth quintile is found to be a significant covariate for increasing the hazards of transition to work for female respondents.

Acknowledgements

We are grateful to Government of India, as the government has served with NNF for our research work. We are also thankful to International Institute for Population Sciences, Mumbai, as they provided the valuable dataset for this research work. Lastly we are thankful to this journal as it provides us a good place for publishing our work. Also, we would like to thank the referees for their comments and suggestions on the manuscript.

References

Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. Circulation. 2016; 133(6): 601−609.

Bernoulli D. Réflexions sur les avantages de linoculation; 1760. p. 173−190.

Berrington A, Diamond I. Marriage or cohabitation: A competing risks analysis of firstpartnership formation among the 1958 british birth cohort. J R Stat Soc Ser A Stat Soc. 2000; 163(2): 127−151.

Billari FC. The analysis of early life courses: Complex descriptions of the transition to adulthood. J Popul Res. 2001; 18(2): 119−142.

Blossfeld HP, Rohwer G. West germany. The New Role of Women: Family Formation in Modern Society. Westview Press. Boulder. Colo; 1995.

Cox DR. Regression models and life-tables. J R Stat Soc Series B Stat Methodol. 1972; 34(2): 187−220.

Cusick GR. Havlicek JR, Courtney ME. Risk for arrest: The role of social bonds in protecting foster youth making the transition to adulthood. Am J Orthopsychiatry. 2012; 82(1):19.

Dignam JJ, Zhang Q, Kocherginsky M. The use and interpretation of competing risks regression models. Clin Cancer Res. 2012; 18(8): 2301−2308.

Ermisch J, Francesconi M. Partnership formation and dissolution in great britain. Technical report. Institute for Social and Economic Research; 1996.

Gorfine M, Hsu L. Frailty-based competing risks model for multivariate survival data. Biometrics. 2011; 67(2): 415−426.

Ha ID, Christian NJ, Jeong JH, Park J, Lee Y. Analysis of clustered competing risks data using subdistribution hazard models with multivariate frailties. Stat Methods Med Res. 2016; 25(6):2488−2505.

Hill DH, Axinn WG, Thornton A. Competing hazards with shared unmeasured risk factors. Sociol Methodol. 1993; 23: 245−277.

Hoem JM. The impact of education on modern family-union initiation. Eur J Popul/Revue Europeenne de D emographie. 1986; 2(2):113−133.

Hosmer D, Lemeshow S, May S. Applied survival analysis regression model of time to event data. hoboken; 2008.

IIPS, PC. Youth in india: Situation and needs 2006−2007; 2012.

Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data, volume 360. John Wiley & Sons; 2011.

Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc J AM STAT ASSOC. 1958; 53(282): 457−481.

Kleinbaum DG, Klein M. Competing risks survival analysis. Survival Analysis: A self-learning text; 2005. p. 391−461.

Lee M, Ha ID, Lee Y. Frailty modeling for clustered competing risks data with missing cause of failure. Stat Methods Med Res. 2017; 26(1):356−373.

Lentine KL, Rey LAR, Kolli S, Bacchi G, Schnitzler MA, Abbott KC, Xiao H, Brennan DC. Variations in the risk for cerebrovascular events after kidney transplant compared with experience on the waiting list and after graft failure. Clin J Am Soc Nephrol. 2008; 3(4):1090−1101.

Liefbroer A, de JGJ. Social background and life course influences on patterns of union formation among young adults in the netherlands. Dynamics of Cohort and Generations Research. Amsterdam: Thesis Publishers; 1992. p. 481−514.

Manting D. Dynamics in marriage and cohabitation: an inter-temporal life course analysis of first union formation and dissolution; 1994.

Marubini E. Valsecchi MG. Analysing survival data from clinical trials and observational studies. John Wiley & Sons. 2004; 15.

Mejia PG. A life course perspective on social and family formation transitions to adulthood of young men and women in Mexico. PhD thesis. The London School of Economics and Political Science (LSE); 2012.

Mills M. Introducing survival and event history analysis. Sage; 2011.

Noordzij M, Leffondre K, van SKJ, Zoccali C, Dekker FW, Jager KJ. When do we need competing risks methods for survival analysis in nephrology? Nephrol Dial Transplant. 2013; 28(11): 2670−2677.

Prentice RL, Kalbfleisch JD, Peterson JAV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. Biometrics. 1978; 34(4): 541−554.

Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat. Med. 2007; 26(11): 2389−2430.

Ram U, Mohanty S, Singh A. Youth in india: Situation and needs; 2010.

Shaw TS. Transitions from cohabitation: A competing risk analysis. Rev Mark Integr. 2011; 3(2): 121−159.

Steele F, Goldstein H, Browne W. A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. Stat Model. 2004; 4(2): 145−159.

Studer M, Liefbroer AC, Mooyaart JE. Understanding trends in family formation trajectories: An application of competing trajectories analysis (cta). Adv Life Course Res. 2018; 36: 1−12.

Tang X. Two-level lognormal frailty model and competing risks model with missing cause of failure; 2012.

Theil H. Statistical decomposition analysis; with applications in the social and administrative sciences. Technical report; 1972.

Thornton A, Axinn WG, Teachman JD. The influence of school enrollment and accumulation on cohabitation and marriage in early adulthood. Am Sociol Rev. 1995; 60(5): 762−774.

Varadhan R, Weiss CO, Segal JB, Wu AW, Scharfstein D, Boyd C. Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications. Med Care. 2010; 48(6 Suppl): S96−S105.

Yamaguchi K. Event history analysis. Sage. 1991 Vol28.