Wiley
Journal of Probability and Statistics
Volume 2025, Article ID 6808479, 11 pages
https://doi.org/10.1155/jpas/6808479

WILEY

*Research Article*

# Bivariate Copula Model on Fitting Correlated Time-to-Event Outcomes: Age at First Sex and Age at First Marriage Among Youth in Tanzania

**Jacqueline Materu**[1,2] **Eveline T. Konje**[2] **Ties Boerma**[3] **Mark Urassa**[1]
**Milly Marston**[4] **and Jim Todd**[1,2]

[1]*Department of Sexual and Reproductive Health, National Institute for Medical Research, Mwanza Centre, Mwanza, Tanzania*
[2]*Department of Biostatistics, Epidemiology and Behavioural Sciences, School of Public Health, Catholic University of Health, and Allied Sciences, Mwanza, Tanzania*
[3]*Department of Community Health Science, University of Manitoba, Winnipeg, Canada*
[4]*Department of Population Health, London School of Hygiene and Tropical Medicine, London, UK*

Correspondence should be addressed to Jacqueline Materu; jacquelinemateru2@gmail.com

Received 12 November 2024; Accepted 22 May 2025

Academic Editor: Muhammad Ahsan



Traditionally, age at first sex (AFS) and age at first marriage (AFM) have been analysed independently. While useful for summarising risk factors for each outcome individually, these approaches offer limited insight into the interdependence between these events. This study used an Archimedean copula model for bivariate right-censored data to jointly model AFS and AFM reported by 9726 young people aged 15–24 years in Kisesa, Tanzania. The dependence structure was identified, the degree of association between these events and their associated factors assessed, and the trends of predicted medians examined. Various Archimedean copulas (Ali–Mikhail–Haq, Clayton, Frank, Gumbel, Copula2, and Joe) were evaluated. Copula function selection was based on the Akaike information criterion (AIC), Bayesian information criterion (BIC), and log-likelihood values, with the Frank copula and a log-logistic marginal distribution performing best. The Frank copula's dependency parameter ($\theta$) was highly significant, with an estimated $\theta$ of 39.49, translating to a Kendall's $\tau$ of 0.903 in the unadjusted model, which included only sex as a covariate, indicating a strong positive correlation between AFS and AFM. Similar results were observed in the adjusted model (Kendall's $\tau = 0.89$), which incorporated additional variables such as education and residence area. Trends show a better estimation of AFS and AFM for both females and males over the period 1994–2016 when analysed jointly rather than separately. The strong positive correlation suggests these events are highly correlated; hence, using joint models captures interdependence and provides more accurate estimates. This approach can inform targeted interventions to improve youth health outcomes.

**Keywords:** age at first marriage; age at first sex; Archimedean copula models; bivariate right censored; dependence

## 1. Introduction

Early sexual debut and marriage have been associated with diverse health risks, particularly for young individuals, exposing them to the vulnerabilities of acquiring HIV, other sexually transmitted infections (STIs), unintended pregnancies, childbirth complications, and fertility-related issues [1–3]. A better understanding of age at first sex (AFS) and age at first marriage (AFM) is essential for improving sexual and reproductive health (SRH) among young people, especially in low- and middle-income countries (LMICs) [4–8]. Reported AFS and AFM have shown that trends and variations in these ages can significantly impact health outcomes and the effectiveness of interventions designed to address these risks [9, 10].

The analysis of the association between risk factors and the age at the onset of sexual activity and marriage has traditionally been examined independently for each outcome, neglecting the potential interdependence of these two events [4–8]. However, historically, many societies esteemed virginity, and sexual initiation commonly occurred within the context of marriage, whether at a younger or older age [11]. In contemporary times, however, the onset of sexual intercourse has shifted to earlier ages, while the age at marriage has increased. Consequently, adolescent sexual initiation is likely to occur earlier and before marriage [11].

Traditionally, AFS and AFM have been analysed as independent observations. Some studies highlight the importance of choosing the right method for analysing these two events [12, 13], and a recent study conducted in Tanzania emphasizes the significance of using the proper choice of survival analysis methods to find factors associated with AFS [14]. However, none of these studies indicated the possibility of dependency between these two events. Recognising and accounting for this interdependence through joint modelling can enhance the statistical analysis, provide better model estimates, and offer valuable insights into the interactions between early sexual debut and marriage within the socio-cultural context. These methodological improvements can lead to a more comprehensive understanding of the risk factors associated with AFS and AFM jointly.

There are few joint models for the analysis of two survival functions. The copula survival model is a popular approach for modelling correlated bivariate time-to-event data. Bivariate time-to-event endpoints are frequently used in clinical trials studying bilateral diseases (e.g., eye diseases) or complex diseases (e.g., cancer and psychiatric disorders) [15]. These endpoints can also arise in other fields, including sexual behavioural studies, where two events coming from the same subject are likely to be correlated. Unlike other approaches such as the marginal model [16] or frailty model [17], the copula approach models the joint survival distribution by directly connecting the two marginal distributions through a copula function [15]. It offers the distinct advantage of modelling the marginal distributions and the dependence parameter(s) separately, thereby allowing for greater flexibility in marginal models and easier interpretation of covariate effects [15]. Additionally, the challenge of censoring can be naturally handled through the marginal distributions within the copula function. This model also provides a measurement of association between the two failure time variables, quantified by Kendall's $\tau$, which helps assess the magnitude of the relationship between the two time-to-event outcomes [15].

This study aims to identify the dependence structure between AFS and AFM among young individuals, by measuring the degree or magnitude of association between these events. It also observed the risk factors associated with these events jointly using a copula survival model. The study utilized data collected at an average interval of three year from the Magu Demographic Health Surveillance System (HDSS), conducted between 1994 and 2016 in the Mwanza region of Tanzania, focusing on the young population aged 15–24 years.

## 2. Methods

*2.1. Data Source.* Cleaned data from eight surveys nested within the Magu HDSS between 1994 and 2016 were used. Magu HDSS is an open community cohort of over 54,000 residents in 2022, located 20 km east of Mwanza city. Detailed descriptions of the sampling strategies and survey methods used in Magu HDSS are available in other sources [18]. In brief, the surveys included all adults aged 15 and above who were listed in the respective HDSS follow-up rounds, with data collected using a structured questionnaire. The analysis focused on the young population aged 15–24 years, who provided information on whether they had experienced the events (sex or marriage) and at what age they first experienced these events. The analysis was based on the dataset provided in the supporting information (Dataset.xlsx) (available here).

*2.2. Measures.* The Magu HDSS survey included two questions about sexual intercourse to ascertain the end event (AFS or censoring): "Have you ever had sex? (i.e., ever had sexual intercourse)" and "How old were you when you first had sexual intercourse?" Similarly, to determine the end event for marriage experience (AFM or censoring), the survey asked: "Have you ever married? (i.e., ever been married or in a cohabiting union)" and "How old were you when you first married or lived with a sexual partner?"

The dependent variables in this analysis for each individual were AFS and AFM, which represent the age in years from birth to the respective event. The end events were defined as AFS and AFM, where AFS refers to the age at first sexual intercourse among those who have ever had sex (retrospective reports of AFS), and AFM refers to the age at first marriage or cohabitation among those who have ever been married or cohabited (retrospective reports of AFM). Individuals who reported not being sexually active, or never being married were right-censored at their age at the time of the interview. Data on various independent variables were also collected, including the respondent's sex, level of formal education, area of residence, and HIV status. Descriptive statistics for these variables, as well as for the event variables (ever had sex, ever married), are summarised in Table 1.

*2.3. Statistical Analysis.* The demographic characteristics of the youth in the survey were summarised using the mean and standard deviation for continuous variables, and frequencies and percentages for categorical variables. A copula model for bivariate right-censored data was used to jointly model the two events—AFS and AFM. To fit this copula-based regression model, the marginal distribution was chosen based on the characteristics of the two random variables. Log-logistic and Weibull distribution functions were used in the analysis, as they are suitable for modelling positive continuous random variables (i.e., AFS and AFM) with flexibility in their shape, as noted in a previous study with similar data [14]. The copula family was selected from the most popular copula families for bivariate censored data

Table 1: Descriptive statistics of responses and variables of participants.

| Independent variables | Overall $N = 9726$ n (%) | Female $n = 5751$ (59.1) n (%) | Male $n = 3975$ (40.9) n (%) |
|---|---|---|---|
| *Residence area* | | | |
| Rural | 5733 (58.9) | 3439 (59.8) | 2294 (57.7) |
| Semiurban | 3989 (41.0) | 2309 (40.2) | 1680 (42.3) |
| *Level of formal education* | | | |
| No education | 1104 (11.4) | 747 (13.0) | 357 (9.0) |
| Primary education | 6365 (65.4) | 3867 (67.2) | 2498 (62.8) |
| Secondary or higher education | 2248 (23.1) | 1130 (19.7) | 1118 (28.1) |
| *HIV status* | | | |
| Negative | 7982 (82.1) | 4666 (81.1) | 3316 (83.4) |
| Positive | 207 (2.1) | 161 (2.8) | 46 (1.2) |
| **Event variables** | | | |
| *Ever had sex* | | | |
| No | 4048 (41.6) | 2059 (35.8) | 1989 (50.0) |
| Yes | 5678 (58.4) | 3692 (64.2) | 1986 (50.0) |
| *Ever married* | | | |
| No | 6846 (70.4) | 3185 (55.4) | 3661 (92.1) |
| Yes | 2880 (29.6) | 2566 (44.6) | 314 (7.9) |
| **Response variables** | | **Mean ± SD** | |
| Age at first sex | 16.4 ± 1.8 | 16.3 ± 1.7 | 16.4 ± 1.9 |
| Age at first marriage | 17.5 ± 2.3 | 17.3 ± 2.1 | 17.8 ± 2.5 |

(i.e., the Archimedean copula family), which has an explicit form of:

$$C_n = \phi_n\left\{\phi_n^{-1}(u) + \phi_n^{-1}(v)\right\}, \tag{1}$$

where $u$ and $v$ are two uniformly distributed margins.

$\phi_n$ is the generator function, which is a continuous, strictly decreasing and convex function; $\phi_n^{-1}$ is the inverse of $\phi_n$.

*2.4. Selection of the Copula Function and Model Fitting.* A variety of bivariate distribution functions meet the criteria of a copula. Among the most commonly used parametric bivariate copulas are the Clayton, Frank, and Gumbel and others [19]. The Archimedean copulas examined in this study are detailed in Table 2.

All of these are defined by only one parameter that captures the dependence structure, except for Copula2, which has two dependence parameters ($\alpha$ and $\theta$) that are explicitly connected to Kendall's tau ($\tau$) [15]. Where:

$\theta$ (theta): Is association parameter that controls the strength and direction of dependence between the two variables.

$\alpha$ (Alpha): A secondary dependence parameter in Copula2, which adjusts the shape of the dependence structure alongside $\theta$.

$\tau$ (Kendall's tau): A rank correlation coefficient that quantifies the degree of concordance between two time-to-event outcomes.

$u$ and $v$: The marginal cumulative distribution functions (CDFs) of the two time-to-event variables, transformed to the unit interval [0, 1].

Since copulas are invariant under monotone transformations, the concordance measure Kendall's tau ($\tau$) was used to observe the degree of dependence [15, 20]. The association between AFS and AFM was evaluated using Kendall's ($\tau$), which calculates the difference between the probability of concordance and the probability of discordance for the two realisations of time-to-event outcomes. The impact of copula selection on association shape and correlation values varies: certain formulations produce symmetric associations (e.g., Frank), others do not. When association structure is unknown, copula choice can be guided by data, using criteria like Akaike information criterion (AIC), Bayesian information criteria (BIC), or observed marginal correlations [21]. The Gumbel and Clayton copulas are frequently used to account for upper and lower dependence between two margins respectively, while the Frank copula is more appropriate for dealing with symmetric dependence [15, 19, 20]. For this analysis, all copulas were considered, but the final results were presented with the Frank copula, as it was observed to capture dependence better compared to the others.

The Frank copula $C_\theta(u, v)$ for two variables $u$ and $v$, where $u, v \in [0, 1]$, is given by:

$$C_\theta(u, v) = -\frac{1}{\theta}\ln\left[1 + \frac{\left(e^{-\theta u} - 1\right)\left(e^{-\theta v} - 1\right)}{e^{-\theta} - 1}\right], \tag{2}$$

Table 2: Archimedean copula family.

| Copula name | Bivariate copula $C(u,v;\theta)$ | Domain of $\theta$ | Possible correlation $\tau$ |
|---|---|---|---|
| Ali–Mikhail–Haq | $((uv)/(1-\theta(1-u)(1-v)))$ | $\theta\in[-1,1]$ | $\tau\in[-0.18,0.33]$ |
| Clayton | $[\max\{u^{-\theta}+v^{-\theta}-1;0\}]^{(1/\theta)}$ | $\theta\in[-1,\infty)\backslash\{0\}$ | $\tau\in[-1,1]\backslash 0$ |
| Frank | $-(1/\theta)\log[1+((e^{-\theta u}-1)(e^{-\theta v}-1)/(e^{-\theta}-1))]$ | $\theta\in\mathbb{R}\backslash\{0\}$ | $\tau\in[-1,1]\backslash 0$ |
| Gumbel | $\exp[-((-\log(u))+(-\log(v)^{\theta})^{(1/\theta)})]$ | $\theta\in[-1,\infty)$ | $\tau\in[0,1]$ |
| Copula2 | $(1+((u^{-\theta}-1)/\theta)^{\alpha}+((v^{-\theta}-1)/\theta)^{\alpha})^{(-1/\alpha)}$ | $\alpha\in(0,1],\theta>0$ | $\tau\in[0,1]$ |
| Joe | $1-[(1-u)^{\theta}+(1-v)^{\theta}-(1-u)^{\theta}(1-v)^{\theta}]^{(1/\theta)}$ | $\theta\in[-1,\infty)$ | $\tau\in[0,1]$ |

where $\theta$ is a parameter controlling the dependence strength: $\theta>0$ indicates positive dependence, $\theta<0$ indicates negative dependence and $\theta=0$ indicates independence.

$u$ and $v$ are uniform marginals transformed to the unit interval, representing the CDF values of the variables.

$C_\theta(u,v)$ denotes the Frank copula evaluated at $u$ and $v$, giving the joint probability that the two variables are less than or equal to $u$ and $v$ respectively under the copula model. The Frank copula Kendall's tau ($\tau$) values range from −1 to 1 and assume a zero value when the two time-to-event outcomes are independent (Table 2) [15, 21].

The copula model for bivariate censored AFS and AFM can be presented as follows:

Let $(T_{1,}T_2)$ be the true bivariate event times (i.e., AFS and AFM), with marginal survival functions $S_j(t_j)=P(T_j>t_j), j=1,2$ and joint survival function $S(t_1,t_2)=P(T_1>t_1,T_2>t_2)$. When $(T_{1,}T_2)$ are subject to right censoring for subject $i=1,\ldots,n$, we observe:

$$D_i=\left\{\left(Y_{ij},\Delta_{ij},Z_{ij}\right):\ Y_{ij}=\min\left(T_{ij},C_{ij}\right),\ \Delta_{ij}=I\left(T_{ij}\le C_{ij}\right), j=1,2\right\}, \tag{3}$$

where $C_{ij}$ is the censoring time of $T_{ij}$, $\Delta_{ij}$ is the censoring indicator, and $Z_{ij}$ is the covariate vector.

The R package *CopulaCenR*, using the function *rc_par_copula*, was used to fit the copula model for bivariate censored data under R version 4.4.1. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm was used to find the parameter values that maximise the likelihood function. This method is effective for complex models with multiple local maxima, as it uses gradient information to navigate the likelihood surface efficiently. It is also known for its fast convergence compared to other optimisation methods [22]. The selection of the copula and marginal distribution was based on AIC, BIC, and log-likelihood values. The model with the lowest AIC and BIC values and the highest log-likelihood was chosen. Detailed selection steps are presented in Table 3.

*2.5. Trends Estimation of AFS and AFM.* In estimating the trends in median AFS and AFM among young people over time, three methods were applied based on reports from individuals aged 15–24 years at different time points (i.e., for each survey round), separately for males and females. First, the traditional Kaplan–Meier method was used to estimate the median AFS and AFM, based on survival defined as

$$\hat{\mathrm{S}}(\mathrm{t})=\prod_{\mathrm{t_i}\le \mathrm{t}}\left(1-\frac{\mathrm{d_i}}{\mathrm{n_i}}\right), \tag{4}$$

where $\hat{\mathrm{S}}(\mathrm{t})$ represents the estimated survival probability at time $t$, $d_i$ is the number of events (e.g., first sex or first marriage) occurring at time $t_i$, and $n_i$ is the number of individuals at risk just before $t_i$. Second, accelerated failure time (AFT) models with a log-logistic error distribution were used to separately model AFS and AFM [14]. The AFT model describes a linear relationship between the logarithm of the survival time and the covariates, defined as

$$\log[(T)]=\beta_0+x'\beta+\varepsilon, \tag{5}$$

where $T$ represents the survival time (AFS or AFM), $x$ is the vector of independent variables (e.g., sex), $\beta_0$ is the intercept, $\beta$ is the vector of coefficients, and $\varepsilon$ is a random error term assumed to follow some parametric distribution.

Finally, the Frank copula model was applied to jointly model AFS and AFM, capturing the dependence between these two time-to-event outcomes.

## 3. Results

Table 1 shows the demographic characteristics of the study participants overall and by sex. The available numbers for certain variables are slightly lower than the total number of participants due to missing responses. Overall, more than half of the participants were female (59.1%), had a primary education level (65.4%), resided in rural areas (58.9%), had ever had sex (58.4%) and had married (29.6%). A small percentage had a positive HIV status (2.1%), with a higher prevalence among females (2.8%) compared to males (1.2%). Among the 5678 who had sex, the mean AFS was 16.4 years with a standard deviation of 1.8. Among the 2880 who had married, the mean AFM was 17.5 years with a standard deviation of 2.3.

Table 3 presents the results of fitting different copula models (Clayton, Gumbel, Frank, Ali–Mikhail–Haq, and Copula2) to the data, with sex as a variable in the model. For each copula, the AIC, BIC, and log-likelihood values are provided for Weibull and log-logistic distributions. Lower AIC

Table 3: Fitted maximum likelihood values, AIC values, and BIC values under various copula specifications with sex as variable in the model; Joe copula are omitted as inversion of the hessian was numerically unstable.

| | | Clayton | Gumbel | Frank | Ali–Mikhail–Haq (AMH) | Copula2 |
|---|---|---|---|---|---|---|
| Weibull | AIC | 34,085.3 | 17,313.1 | 5164.2 | 23,241.4 | 51,127.0 |
| | BIC | 34,114.1 | 17,341.8 | 5195.7 | 23,270.1 | 51,162.9 |
| | Log-likelihood | −17038.7 | −8652.5 | −2578.1 | −11616.7 | −25558.5 |
| Log-logistic | AIC | 32,170.4 | 16,669.4 | 2357.1 | 23,110.9 | 23,345.2 |
| | BIC | 32,199.1 | 16,698.2 | 2385.8 | 23,139.7 | 23,381.2 |
| | Log-likelihood | −16081.2 | −8330.7 | −1174.5 | −11551.5 | −11667.6 |

and BIC values indicate a better-fitting model, and higher log-likelihood values suggest a better explanation of the observed data. For both Weibull and log-logistic distributions, the Frank copula shows the best fit, evidenced by the lowest AIC and BIC values and the highest log-likelihood values (AIC = 5164.2, BIC = 5195.7, and log-likelihood = −2578.1 for Weibull, and AIC = 2357.1, BIC = 2385.8, and log-likelihood = −1174.5 for log-logistic, respectively). To fit a copula-model, it is also necessary to select a distribution model for the margins. When comparing the Weibull and log-logistic marginal distributions within the Frank copula, the log-logistic distribution, which has lower AIC and BIC values and a higher log-likelihood value, was the better fit for this data.

Table 4 presents result from a Frank copula model with log-logistic distribution, jointly modelling AFS and AFM outcomes with sex as the only independent variable (crude model). The parameters $\lambda$ (24.17) and $k$ (−3.15) describe the scale and shape of the log-logistic distributions for AFS and AFM, respectively, with the negative shape parameter indicating that the likelihood of these events decreases with age. Females are observed to take a shorter time to experience both AFS and AFM compared to males (−2.38). The high value of theta (39.46) indicates a strong positive dependency between AFS and AFM, suggesting that changes in AFS are associated with corresponding changes in AFM. Kendall's tau (0.91) further confirms this strong positive association, indicating that individuals who experience earlier AFS are likely to experience earlier AFM, and vice versa (Table 4 and Figure 1(a)).

Table 5 expands on these results by fitting both the crude (univariate) and adjusted models. In the crude analysis, being female was associated with an earlier AFS and AFM (parameter = −2.38) compared to males, but the effect became much smaller (parameter = −0.029) after adjusting for residence area, education, and HIV status (adjusted model). Although the effect was minimal, the result remained statistically significant at the borderline level ($p = 0.045$), likely due to the precision of the estimate and the large sample size. Semiurban residence was associated with a later age at these events both before (0.07) and after adjustment (0.37). When compared to individuals with no formal education, those with primary education experienced earlier AFS and AFM in the crude analysis but later in the adjusted model, while those with secondary or higher education experienced these events significantly later in both analyses. HIV-positive individuals had an earlier AFS and AFM in both crude and adjusted analyses compared to HIV-negative individuals, although the difference was insignificant in the adjusted analysis. The strong positive correlation between AFS and AFM remained high after adjustments (0.89), indicating that individuals who experience an earlier AFS are likely to experience an earlier AFM as well (Table 5 and Figure 1(b)).

Table 4: Fitted parameters (with 95% CI) for the Frank copula model with log-logistic margins, jointly modelling age at first sex and age at first marriage outcomes with sex as a variable.

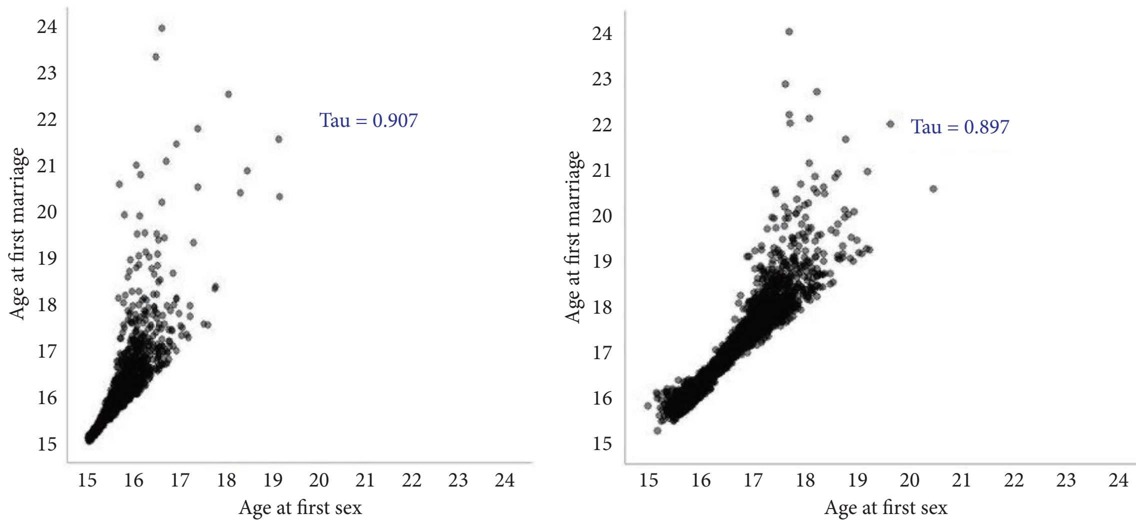
| Parameter | Fitted values (95% CI) |
|---|---|
| *Sex* | |
| Male | 1 |
| Female | −2.3773 (−2.3774; −2.3772)*** |
| Scale ($\lambda$) | 24.1670 (24.1659; 24.1681)*** |
| Shape ($k$) | −3.1467 (−3.1478; −3.1455)*** |
| Theta ($\theta$) | 39.4853 (39.4842; 39.4864)*** |
| Kendall's tau ($\tau$) | 0.9069 (0.8969; 0.9090)*** |

***$p$ value < 0.001.

Separate analyses for females and males showed similar patterns (Tables 6 and 7). For females, semiurban residence and secondary or higher education were associated with a later AFS and AFM in both crude and adjusted analyses. HIV-positive status was associated with an earlier AFS and AFM in the crude analysis but was insignificant in the adjusted model. For males, semiurban residence and higher education delayed AFS and AFM in both analyses, while HIV-positive status resulted in earlier AFS and AFM but was insignificant in both the crude and adjusted models. The strong positive correlation between AFS and AFM remained high for both sexes (Figures 2(a) and 2(b)).

When comparing the estimates of median AFS and AFM across different methods, distinct patterns emerge. For AFS, females show a gradual increase over the years with the copula model, although the copula model produces slightly lower values than the AFT and Kaplan–Meier (KM) methods, with less fluctuation over time. Males also show a similar upward trend with the copula model, while the KM method shows a flat trend from 2010 to 2015–16, and the AFT method shows a slight decrease over the same period. In some years, the estimates are closely aligned across all methods, such as in 2003-04 for females and 2006–07 for males (Figures 3(a) and 3(b)).

For AFM, a similar trend is observed for females, with both the AFT and copula models showing a gradual increase and closely aligned values from 2003–04 to 2015–16, with the




(a) (b)

Figure 1: Scatter plot of bivariate age at first sex and age at first marriage generated from Frank copula model. (a) A model with sex only as a variable. (b) A model with sex adjusted for education level, residence area, and HIV status.

Table 5: Crude and adjusted fitted parameters for Frank copula model with log-logistic margins, jointly modelling age at first sex and age at first marriage outcomes.

| **Parameter** | **Crude$^{\$}$ fitted values (95% CI)** | **Adjusted$^{\dagger}$ fitted values (95% CI)** |
|---|---|---|
| *Sex* | | |
| Male | 1 | 1 |
| Female | −2.3773 (−2.3772; −2.3774)$^{***}$ | −0.0289 (−0.0305; −0.0272)$^{*}$ |
| *Residence area* | | |
| Rural | 1 | 1 |
| Semiurban | 0.0699 (0.0686; 0.0713)$^{***}$ | 0.3711 (0.36979; 0.3724)$^{***}$ |
| *Level of formal education* | | |
| No education | 1 | 1 |
| Primary education | −1.6010 (−1.6017; −1.6003)$^{***}$ | 0.1805 (0.1804; 0.1806)$^{***}$ |
| Secondary or higher education | 0.1143 (0.0626; 0.1659)$^{***}$ | 0.4551 (0.4549; 0.4552)$^{***}$ |
| *HIV status* | | |
| Negative | 1 | 1 |
| Positive | −0.7187 (−0.7199; −0.7177)$^{*}$ | −0.2185 (−0.2257; −0.2114) |
| Kendall's tau ($\tau$) | | 0.8971 (0.8852; 0.8991)$^{***}$ |

$^{*}p$ value = 0.045.
$^{***}p$ value < 0.001.
$^{\$}$Univariate model.
$^{\dagger}$Adjusting for sex, residence area, education, and HIV status.

copula model yielding slightly lower estimates. In contrast, the Kaplan–Meier method shows a consistently flat trend from 1999–2000 to 2015–16. For males, both the AFT and copula models show a gradual increase, with the AFT model showing higher values than the other methods. The copula model exhibits less fluctuation, while the Kaplan–Meier method shows lower values and a flat trend from 2003–04 to 2015–16 (Figures 4(a) and 4(b)).

Figure 5 presents the cumulative incidence curves for AFS and AFM by birth cohorts for both males and females. For females, the more recent cohort (1990–99) experiences these events later on average at earlier ages, with AFS occurring between the ages of 16 and 18, and AFM between the ages of 17.5 and 20, compared to earlier cohorts (1970–79 and 1980–89), which show minimal differences. A similar pattern is observed for males, with the 1990–99 cohort experiencing later events, up to age 21 for AFS and across almost the entire age range for AFM, while earlier cohorts (1970–79 and 1980–89) overlap across the age range. Overall, the curves for females shift to the left, indicating earlier events, whereas the male curves shift to the right, indicating later events. The overall shift in the cumulative incidence curves supports the copula model's finding of strong interdependence between AFS and AFM while also visually showing cohort-specific trends in timing. This means that as AFS shifts to later ages in recent cohorts, with

Table 6: Crude and adjusted fitted parameters for Frank copula model with log-logistic margins, jointly modelling age at first sex and age at first marriage outcomes for females.

| Parameter | Crude[$] fitted values (95% CI) | Adjusted[†] fitted values (95% CI) |
|---|---|---|
| *Residence area* | | |
| Rural | 1 | 1 |
| Semiurban | 0.1675 (0.1661; 0.1687)*** | 0.2271 (0.2160; 0.2389)*** |
| *Level of formal education* | | |
| No education | 1 | 1 |
| Primary education | −1.1770 (−1.1833; −1.1716)*** | 0.3120 (0.2868; 0.3371)*** |
| Secondary or higher education | 0.8753 (0.8494; 0.9013)*** | 0.8734 (0.8661; 0.8808)*** |
| *HIV status* | | |
| Negative | 1 | 1 |
| Positive | −0.6896 (−1.7235; −0.8056)* | −0.1093 (−1.1129; −0.1734) |
| Kendall's tau ($\tau$) | | 0.8909 (0.8846; 0.8973)*** |

*$p$ value = 0.038.
***$p$ value < 0.001.
[$]Univariate model.
[†]Adjusting for residence area, education, and HIV status.

Table 7: Crude and adjusted fitted parameters for Frank copula model with log-logistic margins, jointly modelling age at first sex and age at first marriage outcomes for males.

| Parameter | Crude[$] fitted values (95% CI) | Adjusted[†] fitted values (95% CI) |
|---|---|---|
| *Residence area* | | |
| Rural | 1 | 1 |
| Semiurban | 0.5166 (0.5155; 0.5177)*** | 0.2560 (0.2584; 0.2707)*** |
| *Level of formal education* | | |
| No education | 1 | 1 |
| Primary education | −2.5640 (−2.5641; −2.5637)*** | −1.5506 (−1.5568; −1.5445)*** |
| Secondary or higher education | 1.2923 (1.2639; 1.3202)*** | 1.8087 (1.7924; 1.8248)*** |
| *HIV status* | | |
| Negative | 1 | 1 |
| Positive | −1.5125 (−1.5135; −1.5116) | −1.2624 (−1.2653; −1.2595) |
| Kendall's tau ($\tau$) | | 0.8977 (0.8916; 0.9040)*** |

***$p$ value < 0.001.
[$]Univariate model.
[†]Adjusting for residence area, education, and HIV status.

AFM following a similar trend. The leftward shift in female curves suggests that they generally experience these events earlier than males, whose curves shift rightward, indicating a later occurrence of both AFS and AFM.

## 4. Discussion

AFS, or sexual debut, is a significant event that influences subsequent life events and future SRH [23]. However, statisticians have often analysed AFS as an independent event [14], neglecting the additional information derived from related events such as AFM. This study identified the dependence structure between AFS and AFM, incorporating this interdependence into the statistical model. The results yield a deeper understanding of the timing and influences surrounding these events in adolescents and young adults, highlighting the need for better models that relate AFS to other life events in order to explore statistical associations more effectively.

A copula model for bivariate right-censored data was applied in a joint analysis of the factors associated with both AFS and AFM, taking into account the dependency between the two events. Without prior knowledge of the association structure, it is difficult to know which specific copula model to use. Six different Archimedean copulas were evaluated, with the Frank copula observed to perform better than the others. The Frank copula is particularly effective when there is symmetric lower and upper tail dependence [19]. These results suggest that dependence was detected in both tails, with Kendall's ($\tau$) correlation of 0.903, indicating a strong positive correlation between the AFS and AFM events. This positive correlation was observed in both unadjusted and adjusted models. The marginal log-logistic distribution was found to perform better than the Weibull distribution for this analysis which accords with the study for AFS survival analysis approach [14].

Using joint modelling in this context offers several advantages over marginal single outcome analysis: it is more

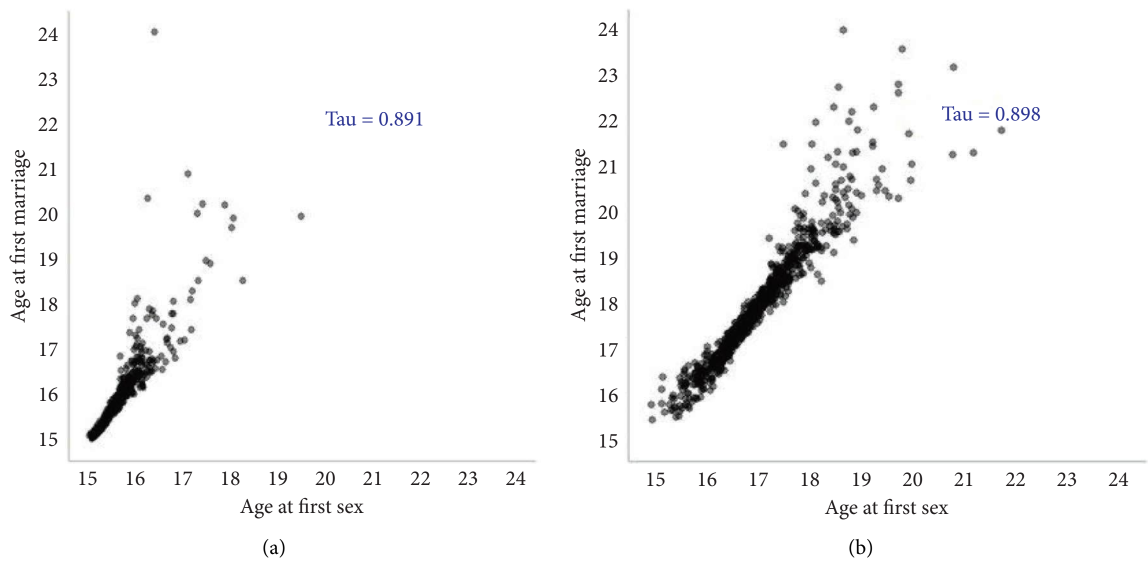



Figure 2: Scatter plot of bivariate age at first sex and age at first marriage generated from adjusted model. (a) Females ($N$ = 4827). (b) Males ($N$ = 3362).

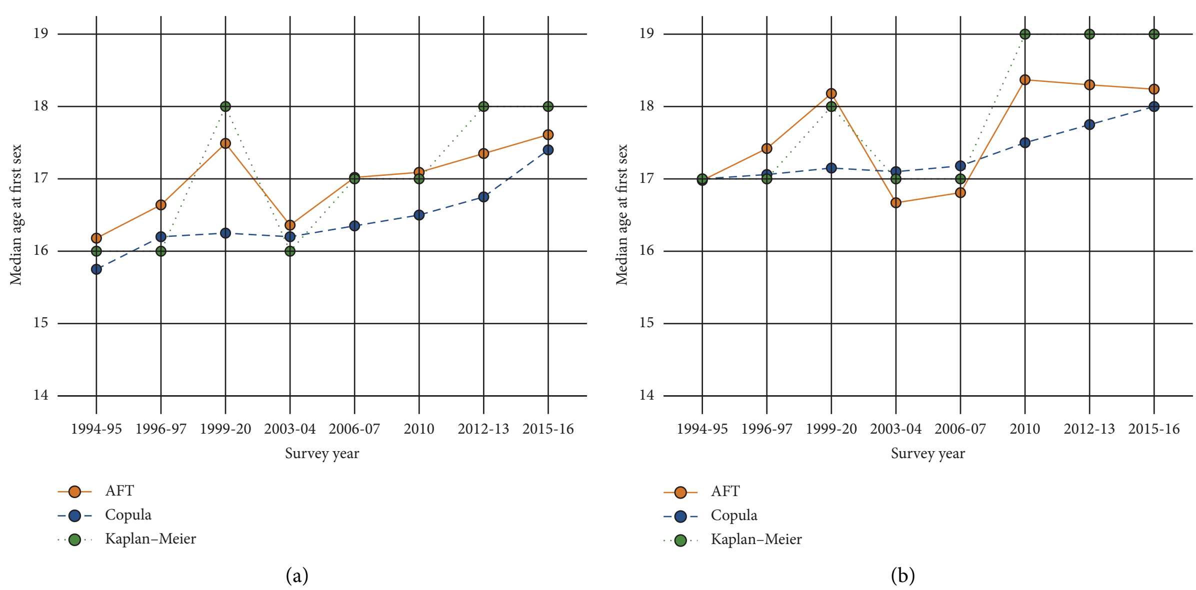



Figure 3: Trends in median age at first sex estimated using AFT, copula (joint) and Kaplan–Meier methods. (a) Females. (b) Males.

time-efficient, improves model fit, increases statistical power, and handles bivariate censoring efficiently [15, 24]. Simultaneously modelling AFS and AFM provides deeper insights into how sexual debut influences marriage timing and vice versa, avoiding biased conclusions that might arise from independent analyses. There have been very few joint models for two time-to-event outcomes, and none we could find for life events such as AFS and AFM. Direct comparisons of copula results with other studies for SRH data are lacking, although copula models have been used. Other studies have explored joint modelling of multiple outcomes, including survival and binary outcomes, using different approaches to model completed schooling year, marriage, and age at first birth [25, 26]. This study provides new insights into the interdependence of AFS and AFM using copula models. More research is needed to determine the best joint model for fitting these two survival functions, and in particular whether common marginal models are appropriate. Future methodological research could employ alternative methods, such as B-splines or Bayesian approaches, to model these bivariate outcomes with the addition of other socio-demographic and risk factors.

Significant associations were also observed between socio-demographic factors and the timing of AFS and AFM, both overall and when stratified by sex. Overall, females tended to experience both events earlier than males in both

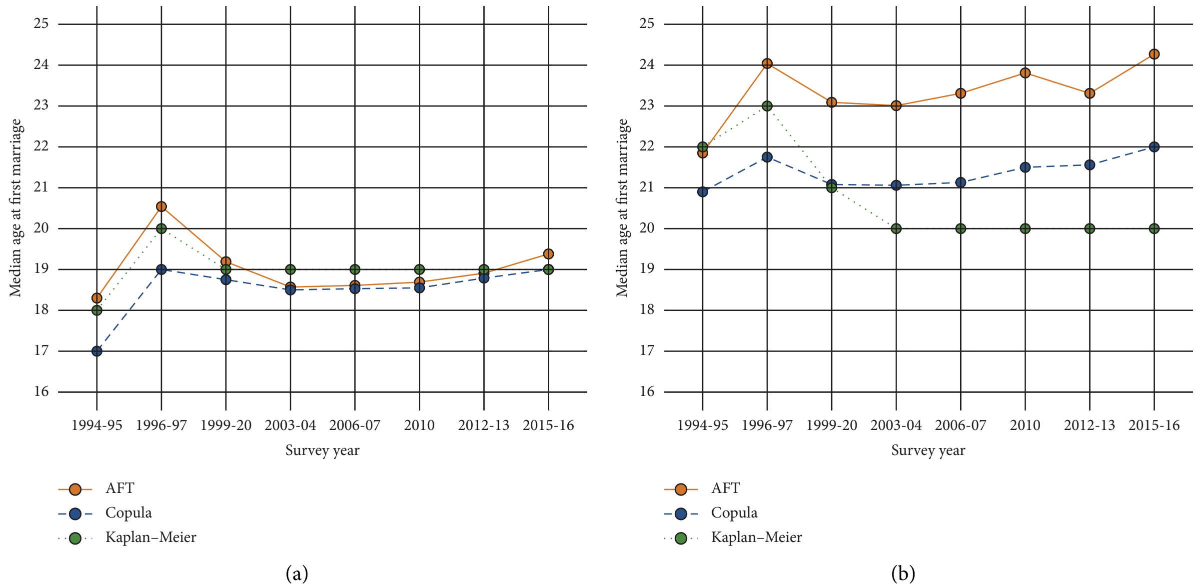



Figure 4: Trends in median age at first marriage estimated using AFT, copula (joint) and Kaplan–Meier methods. (a) Females. (b) Males.

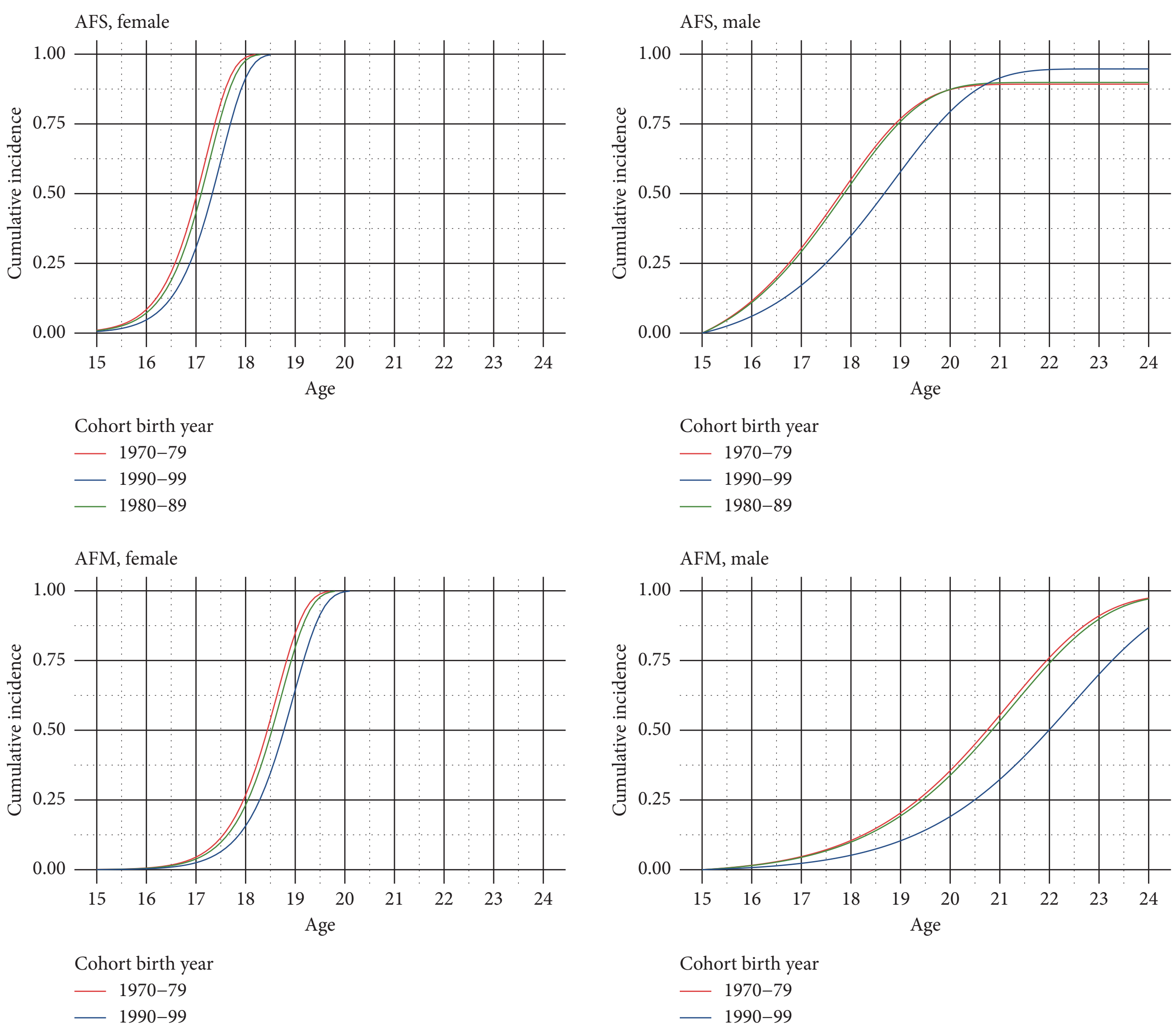



Figure 5: Cumulative incidence probability for age at first sex (AFS) and marriage (AFM) by sex and birth cohort.

crude and adjusted analyses. Those residing in semiurban areas tended to take longer to experience these events in both analyses, with the delay being slightly more pronounced in the adjusted model. Education level also played a crucial role: individuals with primary education experienced AFS and AFM earlier in the crude analysis. Conversely, in the adjusted model, those with primary and secondary or higher education experienced these events significantly later than those with no education. These findings underscore the complexity of socio-demographic influences on the timing of life events and highlight the importance of considering multiple factors simultaneously. Males and females have different pressures regarding first sex and marriage. An analysis of AFS and AFM for each sex separately showed that the magnitude of the effect for each factor varied by sex. Further analysis, incorporating more explanatory factors, is needed to better understand the significance of these influences on AFS and AFM [27].

Trends in AFS and AFM from 1994 to 2016 were conducted separately for cohorts of females and males using three different approaches. The descriptive method using Kaplan–Meier estimation showed flat trends in some years for both events but had the greatest range of values across the different cohorts. The AFT model followed the Kaplan–Meier estimation slightly smaller changes from one cohort to the next. The copula model showed a smoother, more gradual increase in both AFS and AFM over time in some years, reflecting a trend towards older ages at these life events in the later cohorts. All three models provide evidence of an increase in the AFS for both males and females over 22 years, but little evidence of a change in AFM in either males or females.

## 5. Conclusion

A strong positive correlation was observed between AFS and AFM events, which can be used to model these life events. This correlation provides vital information for designing targeted interventions for adolescents and young adults, as early sexual debut often correlates with a higher likelihood of early marriage, which is linked to adverse health outcomes such as increased maternal mortality, reduced educational attainment, and limited socioeconomic opportunities [1–3]. Policymakers and researchers can develop tailored SRH programmes that consider the broader context of young people's lives, ultimately leading to more effective strategies for delaying both sexual debut and early marriage, thereby improving health outcomes for youth. The application of a joint model is able to borrow strength from both outcomes for more accurate and efficient parameter estimates. Although this study is the first to apply a copula bivariate right-censored approach in the context of AFS and AFM, it paves the way for further methodological exploration. Joint models for these bivariate outcomes with the addition of other socio-demographic and risk factors would provide a deeper understanding of the timing and influences on these events, particularly for adolescents and young adults. These results would be crucial for effective policymaking and programme design in SRH research.

## Data Availability Statement

The data that supports the findings of this study are available in the supporting information of this article.

## Ethics Statement

This study, approved by the Research Ethics and Review Committee of the Catholic University of Health and Allied Sciences (CUHAS) and Bugando Medical Centre (CREC/585/2022), utilised secondary data from the Magu HDSS. Ethical clearance for the Magu HDSS had previously been obtained from the Lake Zone Institutional Review Board (LZIRB), the Medical Research Coordinating Committee of Tanzania, and the London School of Hygiene and Tropical Medicine.

## Consent

Participants provided informed consent, with written confirmation obtained prior to survey interviews. The consent process included an oral explanation, allowing participants to agree to the use of their data for research and publication under conditions of confidentiality. For those aged 15–17, parental or guardian consent was required, along with the minor's assent. Eligible participants were invited to a private setting to complete the questionnaire in Swahili, with confidentiality safeguarded through unique anonymised identification numbers.

## Disclosure

The funding body had no role in the study design, data analysis and interpretation, or the writing of the paper.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Funding

The training to conduct this study was supported by the Fogarty International Center (D43 TW011826).

## Acknowledgements

Gratitude is extended to the Magu Health and Demographic Surveillance System (MaguHDSS) study team for their dedicated efforts in data collection and processing. Special thanks to the data managers for their support in data preparation, and sincere appreciation to the survey participants for their invaluable contributions.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section. *(Supporting Information)*

Dataset (.xlsx): The dataset used to produce the results for this study.

## References


[1] P. Agaba, L. K. Atuhaire, and G. Rutaremwa, eds., *Determinants of Age at First Marriage Among Women in Western Uganda* (European population conference, 2010).

[2] E. O. Amoo, "Trends and Determinants of Female Age at First Marriage in Sub-saharan Africa (1990-2014): What Has Changed?" *African Population Studies* 31, no. 1 (2017): https://doi.org/10.11564/31-1-1024.

[3] S. Clark, "Early Marriage and HIV Risks in sub-Saharan Africa," *Studies in Family Planning* 35, no. 3 (2004): 149–160, https://doi.org/10.1111/j.1728-4465.2004.00019.x.

[4] Z. Mtepa, "Factors that Influence Early Sexual Debut Among Female Adolescents," in *Mtwara District, Tanzania* (Sokoine university of agriculture, 2021).

[5] M. Ghebremichael, U. Larsen, and E. Paintsil, "Association of Age at First Sex with HIV-1, HSV-2 and Other Sexual Transmitted Infections Among Women in Northern Tanzania," *Sexually Transmitted Diseases* 36, no. 9 (2009): 570–576, https://doi.org/10.1097/olq.0b013e3181a866b8.

[6] H. S. Whitworth, K. J. Baisley, S. Nnko, et al., "Associations between Age of Menarche, Early Sexual Debut and High-risk Sexual Behaviour Among Urban Tanzanian Schoolgirls: A Cross-sectional Study," *Tropical Medicine and International Health* 28, no. 3 (2023): 237–246, https://doi.org/10.1111/tmi.13858.

[7] M. G. Gobena and Y. M. Alemu, "Analyzing Factors Associated with Time to Age at First Marriage Among Women in Ethiopia: Log Logistic-Gamma Shared Frailty Model," *BMC Women's Health* 22, no. 1 (2022): 191, https://doi.org/10.1186/s12905-022-01775-1.

[8] B. M. Magnusson, J. A. Nield, and K. L. Lapane, "Age at First Intercourse and Subsequent Sexual Partnering Among Adult Women in the United States, a Cross-Sectional Study," *BMC Public Health* 15 (2015): 98–99, https://doi.org/10.1186/s12889-015-1458-2.

[9] B. Zaba, R. Isingo, A. Wringe, M. Marston, E. Slaymaker, and M. Urassa, "Influence of Timing of Sexual Debut and First Marriage on Sexual Behaviour in Later Life: Findings from Four Survey Rounds in the Kisesa Cohort in Northern Tanzania," *Sexually Transmitted Infections* 85, no. Suppl 1 (2009): i20–i26, https://doi.org/10.1136/sti.2008.033704.

[10] L. A. Lara and C. H. Abdo, "Age at Time of Initial Sexual Intercourse and Health of Adolescent Girls," *Journal of Pediatric and Adolescent Gynecology* 29, no. 5 (2016): 417–423, https://doi.org/10.1016/j.jpag.2015.11.012.

[11] B. S. Mensch, M. J. Grant, and A. K. Blanc, "The Changing Context of Sexual Initiation in Sub-saharan Africa," *Population and Development Review* 32, no. 4 (2006): 699–727, https://doi.org/10.1111/j.1728-4457.2006.00147.x.

[12] B. Zaba, T. Boerma, E. Pisani, and N. Baptiste, "Estimation of Levels and Trends in Age at First Sex from Surveys Using Survival Analysis," *MEASURE Evaluation Project, Carolina Population Center Working Paper* (2002).

[13] B. Zaba, E. Pisani, E. Slaymaker, and J. T. Boerma, "Age at First Sex: Understanding Recent Trends in African Demographic Surveys," *Sexually Transmitted Infections* 80, no. suppl_2 (2004): ii28–ii35, https://doi.org/10.1136/sti.2004.012674.

[14] J. Materu, E. T. Konje, M. Urassa, M. Marston, T. Boerma, and J. Todd, "Comparison of Survival Analysis Approaches to Modelling Age at First Sex Among Youth in Kisesa Tanzania," *PLoS One* 18, no. 9 (2023): e0289942, https://doi.org/10.1371/journal.pone.0289942.

[15] T. Sun and Y. Ding, "CopulaCenR: Copula Based Regression Models for Bivariate Censored Data in R. R J" (2020).

[16] L.-J. Wei, D. Y. Lin, and L. Weissfeld, "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions," *Journal of the American Statistical Association* 84, no. 408 (1989): 1065–1073, https://doi.org/10.1080/01621459.1989.10478873.

[17] D. Oakes, "A Model for Association in Bivariate Survival Data," *Journal of the Royal Statistical Society: Series B: Statistical Methodology* 44, no. 3 (1982): 414–422, https://doi.org/10.1111/j.2517-6161.1982.tb01222.x.

[18] M. Urassa, M. Marston, C. Mangya, et al., "Cohort Profile Update: Magu Health and Demographic Surveillance System, Tanzania," *International Journal of Epidemiology* 53, no. 3 (2024): dyae058, https://doi.org/10.1093/ije/dyae058.

[19] R. B. Nelsen, *An Introduction to Copulas* (Springer, 2006).

[20] M. Ayuso, L. Bermúdez, and M. Santolino, "Copula-based Regression Modeling of Bivariate Severity of Temporary Disability and Permanent Motor Injuries," *Accident Analysis & Prevention* 89 (2016): 142–150, https://doi.org/10.1016/j.aap.2016.01.008.

[21] A. Gasparini and K. Humphreys, "A Natural History and Copula-Based Joint Model for Regional and Distant Breast Cancer Metastasis," *Statistical Methods in Medical Research* 31, no. 12 (2022): 2415–2430, https://doi.org/10.1177/09622802221122410.

[22] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer, 1999).

[23] K. Wellings, M. Collumbien, E. Slaymaker, et al., "Sexual Behaviour in Context: a Global Perspective," *Lancet (London, England)* 368, no. 9548 (2006): 1706–1728, https://doi.org/10.1016/s0140-6736(06)69479-8.

[24] D. Rizopoulos, *Joint Models for Longitudinal and Time-To-Event Data: With Applications in R* (CRC Press, 2012).

[25] P. Glick, C. Handy, and D. E. Sahn, "Schooling, Marriage, and Age at First Birth in Madagascar," *Population Studies* 69, no. 2 (2015): 219–236, https://doi.org/10.1080/00324728.2015.1053513.

[26] P. Baizán, P. Baizán, and T. Martín-García, "Endogeneity and Joint Determinants of Educational Enrolment and First Birth Timing in France and West Germany [Joint Determinants of Educational Enrolment and First Birth Timing in France and West Germany]," *Genus* (2006): 89–117.

[27] D. MacKinnon, *Introduction to Statistical Mediation Analysis* (Routledge, 2012).