

# Application of the Joint Frailty Copula Model for Analyzing Time to Relapse and Time to Death of Women with Cervical Cancer

Firomsa Shewa Gari , Tashome Fenta Biru, Selamawit Endale Gurmu

Department of Statistics, Assosa University, Assosa, Ethiopia

Correspondence: Firomsa Shewa Gari, Tel +251912114186, Email firomsashewa.stat@yahoo.com

**Background:** Worldwide, there were 12.7 million new cervical cancer cases, of which 5.6 million took place in industrialized nations and 7.1 million in underdeveloped nations. In eastern, western, middle, and southern Africa, it is the main cancer-related cause of death in female patients. In Ethiopia, cancer was responsible for roughly 5.8% of all fatalities. This study makes use of sophisticated statistical models that take into account population heterogeneity in terms of frailty and dependence between two endpoints in terms of copulas.

**Methods:** Based on hospital registry data, this retrospective study intends to examine the time to relapse and time to death of cervical cancer. This study analyzes 907 cervical cancer-positive women from various parts of Ethiopia. The copula model was used to link time to relapse and time to death of women with cervical cancer. Shared frailty model was used to incorporate unexplained heterogeneity for women with cervical cancer patients.

**Results:** Of the 907 cervical cancer patients, 275 (30.32%) experienced a relapse, 353 (38.92%) died, and 554 (61.08%) were censored. Age, smoking status, family planning, HIV status, family history, abortion, and stage are the most reliable predictors of both time to relapse and time to death of cervical cancer patients. The estimate of the copula parameter ( $\theta = 1.476$ , 95% CI: 1.082, 1.870) shows moderate amount of dependence between time to relapse and time to death (Kendall's rank correlation ( $\tau$ ) = 0.425). The estimate of the variability (heterogeneity) parameter in the population of clusters (region) is  $\eta = 0.495$ , 95% CI: 0.101, 0.889.

**Conclusion:** Age, smoking status, family planning, HIV status, family history, abortion, and more advanced stage significantly increase the risk of relapse and death of female cervical patients. There was a significant association between the time to relapse and the time to die for women with cervical cancer. There was a significant heterogeneity effect in the Tikur Anbessa Specialized Hospital.

**Keywords:** Clayton copula, dependence, frailty model, heterogeneity, relapse

## Introduction

Worldwide, an estimated 604,127 women were diagnosed with cervical cancer in 2020.<sup>1</sup> More than 270,000 women die from it every year, and less developed nations account for more than 85% of these fatalities.<sup>2</sup> Cervical cancer

time worldwide; the majority of them have not received a diagnosis and lack access to medical care that could save or extend their lives.<sup>5</sup>

In eastern, western, middle, and southern Africa, it is the main cancer-related cause of death in female patients.<sup>6</sup> Poor pathology services, a lack of facilities for diagnosis and treatment, and a lack of healthcare infrastructure all contribute to the difficulty of treating cervical cancer in Africa. More than 443,000 women a year are predicted to die from cervical cancer by the year 2030, the majority of them in sub-Saharan Africa.<sup>7</sup> In Ethiopia, cervical cancer was responsible for roughly 5.8% of all fatalities. An estimated 60,960 cases of cancer are diagnosed there each year, and over 44,000 people pass away from the disease there each year.<sup>8,9</sup> According to data collected by the radiotherapy department at Tikur Anbessa Specialized Hospital (TASH), cervical cancer is the second most common form of female cancer among patients who visit the oncology department.<sup>10</sup> There are 20.90 million women in Ethiopia who are 15 years of age or older and at risk of acquiring cervical cancer.<sup>11</sup> Further, one in four women had their first sexual encounter before turning 15.<sup>12</sup> This is thought to raise the risk of contracting human papilloma virus (HPV), the virus that causes cervical cancer. Additionally, investigations have revealed that Ethiopia had a cervical cancer screening rate of less than 10%.<sup>13</sup>

Two survival outcomes are frequently recorded for each patient in medical investigations. For instance, patient's medical records include information about the date of relapses and deaths of cancer patients.<sup>14</sup> For patients with cervical cancer, there is a direct relationship between time to relapse and time to death. Relapse and death are clearly associated because patients frequently pass away right after relapse progression.<sup>15</sup> This implies that after going through recurring progression, the likelihood of dying can significantly increase. In addition, prediction power of time to death improves by taking such dependence structures into account.<sup>16</sup> These suggest the requirement for a suitable statistical model for the dependence between the time to relapse and time to death of cervix cancer patients. In spite of the numerous advantages of the modern statistical models, their applications to cancer data are rarely seen in the medical literature.

When considering two survival outcomes, censoring increases the difficulty of analyzing survival data. Survival outcomes may be censored if their follow-up is stopped before endpoints are seen. The occurrence of relapse, for instance, can be censored by an early death. When the time to tumor progression (such as a cancer relapse) is censored by an informative terminal event, the problem of dependent censoring arises (eg, death).<sup>17</sup> Dependent censoring is the term used when an endpoint of interest is blocked by a mechanism connected to the endpoint.<sup>18</sup> However, due to the dependence between relapse and death, classical statistical methods (eg, Cox regression) on time to relapse may be biased. Analyzing survival statistics for patients who are gathered from various regions (cluster) presents even greater challenges.

The classic Cox regression, which can only handle a single event, a single study, independent censoring, and a limited number of prognostic factors, is insufficient for the analysis of survival data with two events. This study makes use of novel statistical models that take into account population heterogeneity in terms of frailty and dependence between two endpoints in terms of copulas.

## Methods

### Study Area

The study area was Tikur Anbessa Specialized Hospital. It started an organized oncology service in the 1998 Ethiopian calendar. The cancer unit at the TASH provides chemotherapy, radiation therapy, hormone therapy and other supportive and palliative care. It is the main center for cancer registry, early

positive women from various parts of Ethiopia. The time to relapse and time to death of cervical cancer patients commencing from the day of the patients' hospital registration are the response variables in this study. The time is assessed in months. Patients who lost follow-up but are still alive are censored. Age, smoking status, HIV status, family plan, abortion, family history, age at marriage, age at first birth, treatments received, and stage are explanatory variables.

## Inclusion and Exclusion Criteria

All cervical cancer patients registered with full information including in the registration log book or in the patient's identification card were considered to be eligible for the study. Patients with insufficient information about one of the vital variables either in the registration book or in the card were not eligible. As a result, 73 patients were ruled out of the study due to not meeting the inclusion criteria.

## Statistical Methods

Multivariate survival analysis is a branch of survival analysis that deals with more than one event time per subject.<sup>18</sup> In analysis of such multivariate survival data, the key element is an appropriate account for dependence between event times. In survival analysis, the term survival time refers to the time elapsed from an origin to the occurrence of an event.

## Copulas for Bivariate Event Times

The copula model was used to join two event times (time to relapse and time to death of women cervical cancer) by specifying their dependence between failure times.<sup>19,20</sup>

The bivariate survival functions of two end points are given by:

$$P(X > x, Y > y | \mathbf{Z}) = C_{\theta}\{S_X(x | \mathbf{Z}_1), S_Y(y | \mathbf{Z}_2)\}$$

where a function  $C_{\theta}$  is called bivariate survival copula or simply survival copula<sup>21</sup> and a parameter  $\theta$  describes the degree of dependence between two end points. The most popular bivariate copulas are listed below:

The Clayton copula:

$$C_{\theta}(x, y) = (x^{-\theta} + y^{-\theta} - 1)^{-1/\theta}, \theta > 0$$

Kendall's tau ( $\tau$ ) is used to assess the degree of dependence between time to relapse and time to death of women cervical cancer patients and given as:  $\theta/(\theta + 2)$  for Clayton.<sup>22</sup>

## Shared Frailty Model

Shared frailty model was used to incorporate unexplained heterogeneity in the risks of experiencing an event for women cervical cancer patients. In this study, the patients are collected from different regions of Ethiopia. In shared frailty models, it is assumed that each region has its own unobserved factor (called the frailty term) influencing the risks of

## Joint Frailty-Copula Model

The joint frailty-copula model is defined as

$$\begin{cases} r_{ij}(t|u_i) = u_i r_0(t) \exp(\beta_1' \mathbf{Z}_{1,ij}) \\ \lambda_{ij}(t|u_i) = u_i^\alpha \lambda_0(t) \exp(\beta_2' \mathbf{Z}_{2,ij}) \\ P(X_{ij} > x, D_{ij} > y | u_i) = C_\theta[S_{X_{ij}}(x|u_i), S_{D_{ij}}(y|u_i)] \end{cases}$$

## Penalized Likelihood with Splines

Penalized likelihood approach used in this study because the idea of Cox partial likelihood does not carry out in a simple manner, since the integration over frailty induces a complicated form of this likelihood.<sup>23</sup> Thus, the penalized likelihood approach has the advantage that while making no parametric assumption on the hazard or intensity functions, it yields smooth estimates of these functions. The spline method aims to obtain smooth estimate for  $\lambda_0(\cdot)$  and  $r_0(\cdot)$  as a weighted sum of cubic polynomial functions, called basis functions.<sup>23</sup> In this study, the goodness of fit of the models is provided by an approximate likelihood cross-validation criterion (LCV).<sup>24</sup>

- R software version 4.0.5 with joint. Cox package was used for data analysis.

## Result

This study includes 907 women patients who received cervical cancer treatment at TASH from September 2011 to September 2015. Of this, 275 (30.32%) experienced a relapse, 353 (38.92%) died, and 554 (61.08%) were censored (Table 1). The overall survival time ranged from 7 months to 48 months, with a median overall survival time of 34 months. The median relapse time was 29 months, ranging from 24 months to 48 months.

In this study, the explanatory variables include age, smoking status, HIV status, family plan, abortion, family history, age at marriage, age at first birth, treatment received, and stage. Analyses with a single variable and many variables are

**Table 1** Descriptive Summary of Cervical Cancer Data (2011–2015)

| Variable Names | Category | The Number of Observed Events (Event Rates %) |            |             |
|----------------|----------|-----------------------------------------------|------------|-------------|
|                |          | Relapse                                       | Death      | Censored    |
| Age            | ≤ 40     | 33 (3.64%)                                    | 40 (4.41%) | 96 (10.58%) |
|                |          |                                               |            |             |

Table I (Continued).

| Variable Names     | Category          | The Number of Observed Events (Event Rates %) |              |              |
|--------------------|-------------------|-----------------------------------------------|--------------|--------------|
|                    |                   | Relapse                                       | Death        | Censored     |
| Age at marriage    | ≤15               | 164 (18.08%)                                  | 233 (25.69%) | 311 (34.29%) |
|                    | 16–20             | 96 (10.58%)                                   | 107 (11.80%) | 213 (23.48%) |
|                    | >20               | 15 (1.65%)                                    | 13 (1.43%)   | 30 (3.31%)   |
| Age at first birth | ≤15               | 113 (12.46%)                                  | 169 (18.63%) | 184 (20.29%) |
|                    | 16–20             | 129 (14.22%)                                  | 144 (15.88%) | 285 (31.53%) |
|                    | >20               | 33 (3.64%)                                    | 40 (4.41%)   | 85 (9.37%)   |
| Treatment taken    | Surgery           | 63 (6.95%)                                    | 85 (9.37%)   | 110 (12.13%) |
|                    | Chemotherapy      | 60 (6.62%)                                    | 69 (7.61%)   | 84 (9.26%)   |
|                    | Radiation         | 81 (8.93%)                                    | 115 (12.68%) | 206 (22.71%) |
|                    | Combination of ≥2 | 71 (7.83%)                                    | 84 (9.26%)   | 154 (16.98%) |
| Stage              | I                 | 17 (1.87%)                                    | 34 (3.75%)   | 61 (6.73%)   |
|                    | II                | 68 (7.50%)                                    | 95 (10.47%)  | 147 (16.21%) |
|                    | III               | 144 (15.88%)                                  | 172 (18.96%) | 290 (31.97%) |
|                    | IV                | 46 (5.07%)                                    | 52 (5.73%)   | 56 (6.17%)   |
| Region             | Addis Abeba       | 71 (7.83%)                                    | 114 (12.57%) | 120 (13.23%) |
|                    | Oromia            | 56 (6.17%)                                    | 65 (7.17%)   | 142 (15.66%) |
|                    | Amh               |                                               |              |              |

**Table 2** The Gamma-Clayton Archimedean Copula Multivariable Analysis

| Variables      | Categories | Relapse  |       |                     | Death    |       |                     |
|----------------|------------|----------|-------|---------------------|----------|-------|---------------------|
|                |            | Estimate | SE    | RR[95% CI]          | Estimate | SE    | RR[95% CI]          |
| Age            | ≤ 40       | Ref      |       |                     |          |       |                     |
|                | 41–60      | 0.383    | 0.188 | 1.466[1.014, 2.121] | 0.287    | 0.172 | 1.333[0.951, 1.868] |
|                | ≥61        | 0.037    | 0.237 | 1.038[0.652, 1.653] | 0.905    | 0.188 | 2.473[1.711, 3.572] |
| Smoking status | Non-smoker | Ref      |       |                     |          |       |                     |
|                | Smoker     | 0.436    | 0.151 | 1.546[1.15, 2.08]   | 0.812    | 0.127 | 2.252[1.756, 2.888] |
| Family plan    | No         | Ref      |       |                     |          |       |                     |
|                | Yes        | 0.263    | 0.126 | 1.301[1.016, 1.667] | 0.267    | 0.112 | 1.306[1.048, 1.627] |
| HIV status     | No         | Ref      |       |                     |          |       |                     |
|                | Yes        | 0.443    | 0.121 | 1.557[1.227, 1.975] | 0.406    | 0.109 | 1.501[1.213, 1.857] |
| Family history | No         | Ref      |       |                     |          |       |                     |
|                | Yes        | 0.397    | 0.150 | 1.488[1.109, 1.995] | 0.628    | 0.131 | 1.875[1.45, 2.424]  |
| Abortion       | No         | Ref      |       |                     |          |       |                     |
|                | Yes        | 0.358    | 0.136 | 1.432[1.097, 1.     |          |       |                     |

increases the risk of relapse of cervical cancer. The relative risk of women using family plan on time-to-death is 1.306 times more than women did not use family plan [RR = 1.306, 95% CI: 1.048, 1.627].

The relative risk of cervical cancer relapse for women with HIV positive is 1.557 times more than women with HIV negative (RR = 1.557, 95% CI: 1.227, 1.975). This indicated that HIV status was significantly increases the risk of relapse of cervical cancer. The relative risk of women with HIV positive on time-to-death is 1.501 times more than as compared to of the women with HIV negative (RR = 1.501, 95% CI: 1.213, 1.857) which implies that HIV status was significantly increases the risk of deaths of cervical cancer patients.

Regarding family history of women cervical cancer, the relative risk of cervical cancer relapse for having family history is 1.488 times more than women with no family history (RR = 1.488, 95% CI: 1.109, 1.995), which implies that women having family history have a significant effect on the increasing relapse of cervical cancer. The relative risk of having family history on time-to-death is 1.875 times more than as compared to women with no family history (RR = 1.875, 95% CI: 1.45, 2.424).

For women with cervical cancer, having an abortion increases the chance of relapse by 1.432 times compared to not having one (RR = 1.432, 95% CI: 1.097, 1.868), which implies that women having abortion have a significant effect on the increasing relapse of cervical cancer. The relative risk of having abortion on time-to-death is 1.647 times more than as compared to women with no abortion (RR = 1.647, 95% CI: 1.302, 2.084).

Looking for stages of women with cervical cancer, the relative risk of stage III and IV on time-to- relapse were 1.778 with [95% CI: 1.075, 2.940] and 2.955 with [95% CI: 1.694, 5.157], respectively, and this indicates that women with advanced stages have a greater risk for relapse of cervical cancer. Thus, the relative risk of women with relapse of cervical cancer increased by 1.778 and 2.955 for stage III and IV, respectively, as compared to stage I. Also, the relative risk of cervical cancer for stage IV is 2.349 times more than as compared to women with cervical cancer of stage I (RR = 2.349, 95% CI: 1.523, 3.624) and this indicates that women with advanced stages have a greater risk for death of cervical cancer.

Finally, the relative risk of age at marriage who are grouped in 16–20 years on time-to-death is 0.771 times less than as compared to that of the patients age less than or equals to 15 [RR = 0.771, 95% CI: 0.613, 0.970]. Also, the relative risk of age at first birth who are grouped in 16–20 years on time-to-death is 0.677 times less than as compared to that of the patients age less than or equals to 15 [RR = 0.677, 95% CI: 0.542, 0.846].

The correlation between the time of relapse and the time to death from cervix cancer was shown using the scatter plot of the joint survival distribution. The scatter plot (Figure 1) indicates the

## Discussion

This study used the state-of-the-art statistical Joint Frailty Copula model on a dataset on cervical cancer obtained from Tikur Anbessa Specialized Hospital. According to this study, the patients' time to relapse and time to death were associated. This demonstrates that the risk of death can increase dramatically after experiencing reoccurring progression, which could be explained by the fact that the same patient was engaged in both events. This solidifies the idea that the time to bi-variate events are related.<sup>25–28</sup> The study also shows that there was a clustering (frailty) effect, which may have been influenced by the heterogeneity in the patients' countries of origin, given that those from the same region share similar risk factors for cervical cancer.<sup>29,30</sup>

The study found that age significantly increased both the death and relapse rates for women with cervical cancer. This is because the likelihood of causing the body genetic harm (mutations) rises as we live longer. In addition, our bodies' capacity to repair genetic damage declines with age. This aligns with past studies.<sup>31,32</sup> Abortion had a strong positive correlation with both the time to relapse and the time to death for women with cervical cancer. This proves that cervical cancer patients who have previously had an abortion have a higher risk of dying and experiencing a relapse than those who have not. These findings are consistent with a study that was conducted.<sup>33</sup>

Another important factor that significantly affected how long it took for women with cervical cancer to die and experience a relapse of their disease was the usage of family plans. Using a family plan may increase your risk of dying from cervical cancer and its relapse, according to this information. These investigation's findings are consistent with earlier ones.<sup>34,35</sup> Smoking status is a significant risk factor that significantly increases the relapse and death of cervical cancer in women. The most recent study confirms earlier results.<sup>36,37</sup> The results of this study also showed that a woman's family history has a major impact on her chances of dying from cervical cancer and having it come back. The study that served as the basis for this one was.<sup>38</sup>

Additionally, consistent with research, the HIV status at diagnosis was significantly linked to an increased risk of cervical cancer death and relapse.<sup>39</sup> The stages of the disease have a major impact on the time to relapse and the time to death for women with cervical cancer. This shows that women with advanced cervical cancer are more likely to die and experience relapse. Additionally, this result supports the investigations.<sup>40</sup> The study's concluding finding was that women's death from cervical cancer is significantly influenced by their marital and childbearing ages. These findings are consistent with a study that was conducted.<sup>41,42</sup>

## Conclusion

The most significant risk variables for relapse and death in women cervical patients were age, smoking status, family planning, HIV status, family history, abortion, and more advanced stage. While the risk of dying from cervical cancer in women is greatly influenced by age at marriage and age at first birth. The significant association between the time to relapse and the time to die for women with cervical cancer suggests that the copula model must be used to account for such a dependence. In the Tikur Anbessa Specialized Hospital, there was a significant heterogeneity effect, which reveals that one needs to employ the suitable clustered time to event frailty models to account for this clustering effect.

## Abbreviations

HPV, Human papilloma virus; LCV, Likelihood Cross-validation Criterion; MPL, Maximum Penalized Log-likelihood; SNNP, South Nation Nationality and People of Ethiopia; TASH, Tikur Anbessa Specialized Hospital; WHO, World Health Organization.

medical records of patients. Informed consent was not taken from all respondents enrolled in the study since data were collected from the patient's follow-up chart. Confidentiality during all phases of research activities was kept and data was held on a secured password-protected system. All the procedures are based on the principles of Helsinki declaration.

## Acknowledgment

The authors gratefully acknowledge Tikur Anbessa Specialized Hospital for the provision of the data.

## Author Contributions

F.S. Conceived the idea. All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Disclosure

The authors declare that there were no conflicts of interest in this study.

## References

1. Singh D, Vignat J, Lorenzoni V, et al. Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO global cervical cancer elimination initiative. *Lancet Glob Health*. 2023;11(2):e197–e206. doi:10.1016/S2214-109X(22)00501-0
2. Tugizova D, Amonova M, Farmonova D. Prognostic factors of cervical cancer in pregnant women. *Sci Innov*. 2022;1(D7):292–296.
3. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249. PMID: 33538338. doi:10.3322/caac.21660
4. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359–E386. PMID: 25220842. doi:10.1002/ijc.29210
5. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2011;61(2):69–90. Erratum in: *CA*

23. Nielsen G, Gill R, Andersen P, Sorensen T. A counting process approach to maximum likelihood estimation in frailty models. *Scand J Stat.* 1992;19:25–43.
24. Ramsay J. Monotone regression spline in action. *Stat Sci.* 1988;3:425–461.
25. Gari FS, Gelcho GN. Bivariate survival copula analysis of glaucoma patients during blindness: glaucoma cases at Alert Hospital in Addis Ababa City of Ethiopia. *J Res Health Sci.* 2022;22(2):e00547. PMID: 36511259; PMCID: PMC9818039. doi:10.34172/jrhs.2022.82
26. Shewa F, Endale S, Nugussu G, Abdisa J, Zerihun K, Banbeta A. Time to kidneys failure modeling in the patients at adama hospital medical college: application of copula model. *J Res Health Sci.* 2022;22(2):e00549. PMID: 36511261; PMCID: PMC9818037. doi:10.34172/jrhs.2022.84
27. Joe H. *Multivariate Models and Multivariate Dependence Concepts*. CRC Press; 1997.
28. Sun T, Ding Y. CopulaCenR: copula based regression models for bivariate censored data in R. *R J.* 2020;12(1):266. doi:10.32614/RJ-2020-025
29. Gorfine M, Hsu L, Parmigiani G. Frailty models for familial risk with application to breast cancer. *J Am Stat Assoc.* 2013;108(504):1205–1215. PMID: 24678132; PMCID: PMC3963469. doi:10.1080/01621459.2013.818001
30. Mazroui Y, Mathoulin-Pélissier S, Macgrogan G, Brouste V, Rondeau V. Multivariate frailty models for two types of recurrent events with a dependent terminal event: application to breast cancer data. *Biom J.* 2013;55(6):866–884. PMID: 23929494. doi:10.1002/bimj.201200196
31. Ioka A, Tsukuma H, Ajiki W, Oshima A. Influence of age on cervical cancer survival in Japan. *Jpn J Clin Oncol.* 2005;35(8):464–469. doi:10.1093/jjco/hyi125
32. Rositch AF, Nowak RG, Gravitt PE. Increased age and race-specific incidence of cervical cancer after correction for hysterectomy prevalence in the United States from 2000 to 2009. *Cancer.* 2014;120(13):2032–2038. doi:10.1002/cncr.28548
33. Miller B, Morris M, Rutledge F, et al. Aborted exenterative procedures in recurrent cervical cancer. *Gynecol Oncol.* 1993;50(1):94–99. doi:10.1006/gyno.1993.1170
34. Mwaka AD, Orach CG, Were EM, Lyratzopoulos G, Wabinga H, Roland M. Awareness of cervical cancer risk factors and symptoms: cross-sectional community survey in post-conflict northern Uganda