

Research Article

Joint Modeling of Exam Results and Attrition Status of Students at Hawassa College of Education, Ethiopia

Mideksa Tola,¹ Denekew Bitew Belay ,² Melkamu A. Zeru,² and Senait Cherie Adegeh²

¹Department of Mathematics, Hawassa College of Education, Ethiopia

²Department of Statistics, College of Science, Bahir Dar University, Ethiopia

Correspondence should be addressed to Denekew Bitew Belay; denekew.t.h@gmail.com

Received 16 October 2021; Accepted 4 March 2022; Published 19 March 2022

Academic Editor: Yuqing Geng

Copyright © 2022 Mideksa Tola et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Student attrition is a challenge for higher education institutions across the world. The purpose of this study was to examine the application of joint model using students' commutative grade point average and attrition status. A total of 258 college student samples were used in this study. A mixed effect model for students' grade point average and a Cox hazard model for students' attrition status were modeled independently, and both submodels were modeled jointly by linking random effects through a shared parameter model. This study focused on gender issues, academic background, peer support, and residence. From the finding of this study, 26.4% were attrition, and 73.6% were retained. The estimated trend of commutative grade point average was found to be negatively associated with attrition status. The major factors that encounter students' attritions were academic background and institutional factors.

1. Introduction

Education enables individuals and society to make all-rounded participation in the development process by acquiring knowledge, ability, skills, and attitudes [1]. Education is a development of physical, mental, moral (spiritual), and social values of individuals for a life of dedicated service [2]. A significant and imposing challenge in education is to provide equitable instruction and assessment to all students, like students' performance is considered as objectively as possible for the field and in relation to their peers [3, 4].

Student attrition is one of the greatest areas of interest in higher education, and it is usually defined as the number of noncompleting students (i.e., students who have not yet finished their study program) who are enrolled in a specific university, college, school, discipline, or program in a given year, but not enrolled in that same program the following year [5]. In Ethiopia, student attrition is considered to be of significant importance within the education sector due to its critical role in meeting national goals and institutional objectives. Because Ethiopia's various national policies which determine the development of human capital as a

key strategic tool for meeting the country's aspiration of becoming a middle-income country by 2025 [6], student attrition has upsetting and costly impacts the country's economy, and it has drawn a considerable attention globally and nationally in colleges and universities, due to negative consequences on individual students' lives, their families, and a country's economies [2, 7].

Some have argued that student completion rates are a fundamental measurement of student success, and it is often depend upon the prerequisite skills, knowledge, and commitment of students. The author goes on to mention that low levels of basic skills, an inadequate knowledge base, and low self-confidence were contributing to the failure and attrition of students in undergraduate programs. Confidence, self-efficacy, and self-esteem also seem paramount to new student success [8].

Numerous studies have sought to identify models and sets of variables to explain what causes students force to leave the tertiary education system. For most students, deciding to leave tertiary education is not the result of one factor. Rather, it is the result of a combination of complex, interconnected factors that develop over time. Many studies

have identified previous academic performance, mismatches between student expectations and experiences, student disorientation or socialization, and other factors being key predictors of attrition. There are many variables that are likely to affect the academic success of students who enrolled to university through enabling programs. The study conducted by Alter and Haydon assessed that age, gender, educational and family circumstances, ability, self-confidence, achievement goals, and approaches to self-regulation of academic behavior were some of the main variables that affect students' performance severely [9].

The time constraints of college terms and the amount of academic rigor required in college courses can lead to student stress and dissatisfaction [10]. If students had an accurate estimate of time remaining to a person's retention, they could concentrate on those students whose studies are going to be prolonged and help them to remove the obstacles from timely retention. In a situation, when both separate and joint outcomes are observed in one subject, separate modeling does not take into account the dependence between the two types of responses. So, this approach has enabled researchers to carry out questions of factors influencing the longitudinal measure of students' result, the risk account for student retention and to test the association between cumulative grade point average (CGPA) repeated measurements and time-to-event data.

2. Data and Methods

The study has been carried out at Sidama Regional State, Hawassa College of Teacher Education, Ethiopia. The sample size used for this study was determined based on [11] formula designed for the appropriate size determination. The samples of 285 students were included in this study. Data for both the longitudinal and time-to-event outcomes were collected by using questionnaire, in which the questionnaire was prepared by the researcher based on literature which states about students' attrition and factors affecting attrition of the students. In the study, a batch of 2009 E.C regular natural science students was included, and others were excluded and studied repeatedly for the first five semesters. The systematic random sampling was employed, and samples were selected using their ID.

The analysis of this study consists of both exploratory and inferential analysis. In exploratory data analysis, we had used mean profile plot for longitudinal data and Kaplan-Meier survival plot graph for time-to-event data. In the inferential analysis, linear mixed effect model for the longitudinal data Cox proportional-hazard model for the time to-event data and joint modeling for the two data altogether were used. As a first step of analysis, the data was explored in different ways in order to get details that helps to make decisions in the subsequent steps of the analysis. The mean and the correlation structures were also explored through graphical techniques. In parallel to defining the fixed effect model, a random effect model was chosen to define a covariance model. After deciding the fixed effects, the study selected a set of random effects to be included in the model.

In the longitudinal data analysis, the variable CGPA was used as outcome measure, and covariates were sex, age, distance of residence, high school result, family size, level of peer support, entrance result, marital status, and house head education, while in the survival model, time to attrition was considered as the response variable which can be affected by sex, entrance result, peer support, marital status, interest of field study, income, study time, and place of residence of students.

In longitudinal data analysis, mean profile plots, correlation structure, and variance structure plots were obtained in order to gain some insight of the data [12]. The individual profile plots and the variance structure were used to gain insight of the variability in the data and to determine whether random effects (random intercepts and slopes) were to be considered in the analysis [13].

For longitudinal data, two sources of variations are considered. Modeling with in subject variation helps us to study changes over time, while modeling between subject variations helps us to understand differences between subjects [14].

A mixed model is one that contains both fixed and random effects part. For the continuous case, the linear mixed effect model (LLM) provides a general and flexible modeling framework where subject-specific random effects assumed to follow a normal distribution are included to account in the correlation [15].

Let β denote a $P \times 1$ vector of unknown population coefficients for the fixed effects and X_i be known $n_i \times P$ design matrix values of the fixed predictors linking β to set of longitudinal measurements Y_i .

Let b_i denote a $K \times 1$ vector of unobservable individual random effects and Z_i be a known $n_i \times 1$ design matrix values of the random factors linking b_i to Y_i , and ε_i is $n_i \times 1$ vector of unknown random errors.

Then, the general LMM of the longitudinal data is given by:

$$\begin{cases} Y_i = X_i\beta + Z_i b_i + \varepsilon_i \\ b_i \sim N(0, D) \\ \varepsilon_i \sim N(0, \sigma^2 I), b_i \text{ and } \varepsilon_i \text{ are independent} \end{cases}, \quad (1)$$

where ε_i distributed as $N(0, \sigma^2 I)$ is a vector of residual components, combining measurement error and serial correlation.

Survival analysis is an area of statistics that studies the time until a prespecified event of interest occurs.

Let T be a nonnegative random variable representing the time to attrition status of the student. The more optimistic survival function $S(t)$ at time t , $S(t) = P(T > t)$ is defined to be the probability that a randomly selected individual will survive beyond time t . We regard T for the i^{th} respondent and C as the corresponding censoring time. Let $\delta_i = I(T_i \leq C_i)$, where $I(\cdot)$ is an indicator function and takes the value

$$\delta_i = \begin{cases} 1, & \text{if } T_i \leq C_i \\ 0, & \text{if } T_i > C_i \end{cases}. \quad (2)$$

The Cox proportional hazard (CPH) model is the most widely used semiparametric survival regression model in which for a set of covariates X_i for the i^{th} subject, and β is the $\rho \times 1$ parameter of coefficients; the hazard at time t is expressed as follows:

$$\lambda_i(t) = \lambda_0(t) \exp(\beta^T X_i), \quad (3)$$

where $\lambda_i(t)$ represents the hazard of attrition for a subject i at time t . $\lambda_0(t)$ is a baseline hazard function that describes the risk for individuals with $X_i = 0$.

A novel use of joint model, which gains increasing interest in recent years, refers to the statistical analysis of the resulting data while account of any association between the repeated measurement and time-to-event outcomes [16]. Joint modeling of longitudinal and survival data can be formed where the association between the two endpoints is due to shared random effects that means random effects account for both the association between the longitudinal and time-to-event outcomes and the correlation between the repeated measurements in the longitudinal process. This type of joint model is also called a shared parameter model as both processes shared these random effects [17].

The longitudinal and survival components of the joint model are typically linked through the trajectory function. Specifically, the shared random-effect models at time t can be written as follows:

$$\lambda_i(t; m_i(t), \psi_i) = \lambda_0(t) \exp(\psi^t \gamma + \alpha m_i(t)), t > 0, \quad (4)$$

where $m_i(t)$ represents the history of unobserved longitudinal response up to time t , Ψ_i represents the vector of baseline covariates with corresponding parameter estimates γ , and α measures the effect of the longitudinal outcome to the risk of an event; the risk of an event at time t depends on the true value of the longitudinal endpoint at that time.

Parameters in the three models were estimated mainly through the use of maximum likelihood (ML) estimation which is a very general approach to statistical estimation which widely used to handle many difficult estimation problems. Models are compared with based on the value of Akaike's information criterion (AIC), the Bayesian information criterion (BIC), and the likelihood ratio test (LRT) methods for nested model assessment criterion for model selection [16].

3. Result

In this study, 1290 observations were considered to collect the CGPA from 285 students which was evaluated at fixed time points and measurements. All the CGPA were taken at two semesters of first, second, third, fourth, and fifth which had equal time intervals of 5 consecutive semester CGPA measurements.

Among the total study subjects included in this study, 183 (70.9%) were males. In terms of marital status, about 204 (79.1%) were single, and the remaining 54 (20.9%) were married. When we look at their place of residence, 168

(65.1%) were from rural areas, and 90 (34.9%) were from urban areas. According to the field of study, the student enrolled to the higher education institution, about 34 (13.2%) were in mathematics department, 50 (19.4%) were in chemistry department, 49 (19.0%) were in physics department, 49 (19.0%) were in biology department, 16 (6.2%) were in integrated science, and 60 (23.3%) were in MNS. All students did not enroll with their interest to field study. About 150 (58.1%) students were enrolled to study program without their interest, and 108 (41.9%) students were placed with their interest of study program. From the same result, with regard to the education level of house head, about 114 (44.2%) were illiterate, 86 (33.3%) household head had primary education, about 29 (11.2%) had secondary education status, and 29 (11.2%) were certificate and above. The mean age students enrolled to the Hawassa Teacher's Education College was 19.91 years with standard deviation of 1.99 (19.91 \pm 1.99).

The longitudinal response variable, CGPA, was measured from semester one to semester five consecutively. As it can be seen from Table 1, common measurements are used for all respondents at these five semesters; in semester I, there were 258 (100%) students; in semester II, there were 250 (96.9%) students; in semester III, there were about 208 (80.62%) students; in semester IV, there were about 195 (75.58%) students, and there were 190 (73.6%) students in semester V. From this result, we clearly observe that there was a sharply increase in degree of attrition over five consecutive semesters. There are so many reasons for attrition of students from schooling such as academic dismissal, readmission, withdraw from the program due to health problem or financial problem, and dropouts from their batch. Among the students who included in this study, about 190 (73.6%) were retained up to the time duration of the study. The average mean score of CGPA for five semesters was about 2.28 out of 4.00 point with a standard deviation of 0.57 per individual. This is relatively low score as compared to 4.00 point which is the highest score in Ethiopian grading system.

The survival response variable was the length of time from enrollment semester until the semester of attrition or retained. Among the students involved in this study, about 73.6% were retained until 5th semester, and 26.4% attrite were due to academic delay or any other individual cases (Table 1).

As we observe from Figure 1, the average progress of CGPA for male students was higher than the female students over the five semesters (Figure 1(b)), and the survival probability of male students was greater than that of the female students (Figure 1(a)). These two explanatory analyses revealed that male students were more survived from the program with better CGPA than female students which implies that there were high attrition rates among female students as compared to male students. Good models that best describe the observed average trends and also reflect the observed correlation structures were sought for the data sets. To identify the appropriate covariance structure, we had test three different commonly used covariance structures: compound symmetry (CS), unstructured (UN), and first order autoregressive (AR (1)) could be considered.

TABLE 1: Demographic profile of respondents (N = 258).

Characteristics	Category	Semester admission and student number				
		Semester I (M, F)	Semester II (M, F)	Semester III (M, F)	Semester IV (M, F)	Semester V (M, F)
Sex	Male	M = 183	M = 178	M = 148	M = 139	M = 136
	Female	F = 75	F = 72	F = 60	F = 54	F = 54
Marital status	Single	(164, 40)	(161, 3)	(140, 2)	(133, 22)	(131, 20)
	Married	(19, 35)	(17, 35)	(8, 35)	(6, 34)	(5, 34)
Residence	Urban	(60, 30)	(57, 30)	(49, 23)	(47, 21)	(46, 20)
	Rural	(123, 45)	(121, 4)	(99, 37)	(92, 35)	(90, 34)
Department	Mathematics	(25, 9)	(25, 7)	(21, 6)	(19, 6)	(19, 6)
	Chemistry	(38, 12)	(38, 10)	(33, 7)	(31, 7)	(31, 7)
Department	Physics	(40, 9)	(39, 8)	(32, 7)	(30, 7)	(30, 7)
	Biology	(40, 20)	(25, 24)	(21, 19)	(19, 18)	(19, 16)
Department	Integrate S.	(11, 5)	(11, 5)	(9, 4)	(8, 4)	(8, 4)
	MNS	(40, 20)	(40, 18)	(32, 17)	(30, 16)	(29, 14)
In. field choice	Interest	(130, 20)	(127, 1)	(108, 1)	(101, 12)	(101, 12)
	Not interested	(53, 55)	(51, 54)	(40, 47)	(38, 44)	(38, 44)
House head edu	Illiterate	(87, 27)	(85, 26)	(72, 20)	(67, 19)	(66, 18)
	Primary	(66, 20)	(65, 19)	(54, 16)	(50, 16)	(48, 16)
House head edu	Secondary	(11, 18)	(10, 18)	(7, 17)	(7, 15)	(7, 14)
	Certificat.ab	(19, 10)	(18, 9)	(15, 7)	(15, 6)	(14, 6)
Peer support	High	(28, 16)	(28, 15)	(23, 13)	(21, 13)	(20, 13)
	Low	(58, 29)	(58, 27)	(48, 23)	(44, 22)	(44, 20)
Distance of residence	Medium	(97, 30)	(95, 27)	(77, 24)	(74, 21)	(72, 21)
	Far	(97, 35)	(95, 33)	(78, 28)	(72, 27)	(70, 27)
Status	Moderate	(58, 25)	(55, 25)	(45, 22)	(42, 21)	(42, 19)
	Near	(28, 15)	(28, 14)	(25, 10)	(25, 8)	(24, 8)
Status	Retained	190	(178, 7)	2508	(139, 56)	195
	Not retained	68	(5, 3)	(30, 12)	(9, 4)	13
Mean, median, CGPA	Mean	Mean = 2.28, median = 2.35	Mean = 2.10, median = 2.11	Mean = 1.90, median = 2.15	Mean = 1.85, median = 2.14	Mean = 1.91, median = 2.23
	Sd	Var = 0.32, Sd = 0.57	Var = 0.53, Sd = 0.73	Var = 1.07, Sd = 0.57	Var = 1.30, Sd = 1.14	Var = 1.20, Sd = 1.10
(δ^2, δ) , Semr, CGPA	Family size	mean = 8.2, St.dv = 2.05	mean = 512.28, St.dv = 348.8	mean = 62.42, St.dv = 10.24	mean = 3 : 30, St.dv = 1 : 54	Age: mean = 19.91, St.dv = 1.99
	High school res	mean = 71.68, St.dv = 11.32	Income: mean = 512.28, St.dv = 348.8	Entrance score: mean = 62.42, St.dv = 10.24	Study time per day: mean = 3 : 30, St.dv = 1 : 54	Age: mean = 19.91, St.dv = 1.99

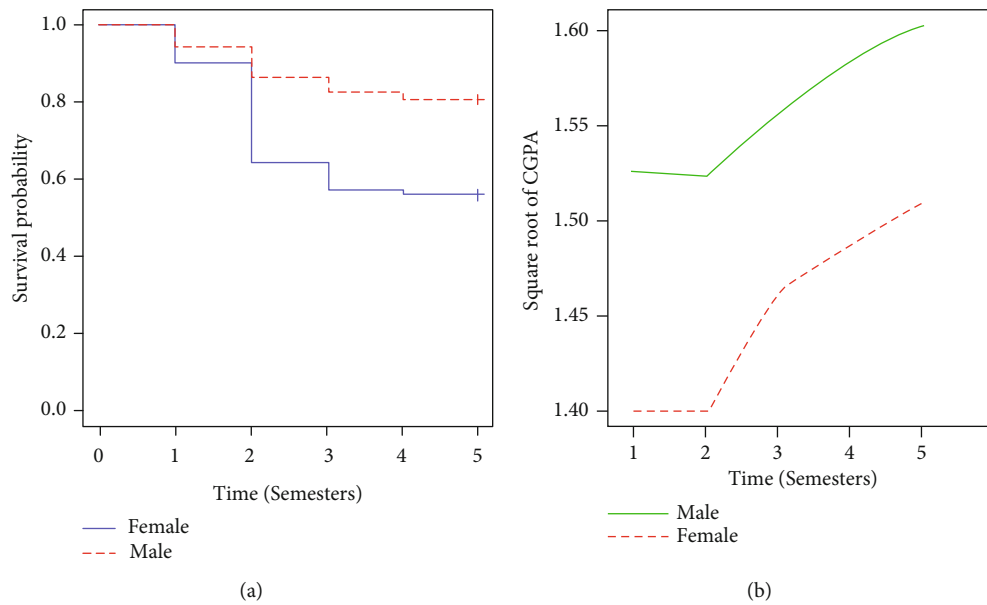


FIGURE 1: Explanatory data analysis. (a) Kaplan-Meier survival plots. (b) Mean profile plot.

Based on the smallest value AIC and BIC, AR (1) covariance structure was used in model building with intercept and linear age slope.

Study results from Table 2 show that the fitted linear mixed effect model for preliminary final model containing significant main effects and possible interaction effect reveal that sex, high school result, family size, peer support, college entrance result, and house head education level of students were statistically significant (p value < 0.05) while the distance from the college to the student residence does not affect the student CGPA.

To explore the survival process, we assessed each factor through Cox regression model and found that the variable sex, college entrance result, peer support, interest of field of study, and marital status were statistically significant under separate model of survival analysis.

From the result displayed in Table 3, it can be seen that sex, entrance result, level of support, interest of field study, and marital status are statistically significant at 5% level of significance.

The estimates of the parameter in the separate and joint models are quite similar to each other but not identical. "Association" is in fact parameter (α) in equation (4) that measures the effect of $m_i(t)$, where $m_i(t)$ represents the history of unobserved longitudinal variable CGPA. The estimated trend of CGPA was provided that CGPA is negatively associated with attrition. This shows that the effect of longitudinal measure of CGPA on the attrition status of the student. The quantification of the effect of this repeatedly measures of CGPA is fundamental to understand the trend and take appropriate intervention mechanism.

In the same result for a one unit increment of entrance result, there were 9.751 unit increase in the average change in the CGPA of students. Therefore, this revealed that higher entrance result reduces the attrition of students from their

study because of better entrance result mostly scored by better performing student who will have higher probability of retention to the study program from enrollment up to graduation (Table 4).

Result from Table 5 shows that the highest variability of residuals was from the random intercepts in both the separate linear mixed effect and survival models. It also shows that the variance of the random intercepts was higher than that of the random slopes. The residual variability was smaller in joint analysis (15.3876) compared to the relative linear mixed effect analysis (18.943). Why the variability in joint becomes lower was probably due to the reason of the standard errors were adjusted for the correlation between the responses in the joint model analysis.

3.1. Implication. This study tried to assess the effect of longitudinal measure of the commutative grade point average (CGPA) of students in each semester on the attrition status of the student. The quantification on the effect of this repeatedly measure CGPA is fundamental to understand the trend and take appropriate intervention mechanism. Understanding the effect of each semester's result for student is fundamental for timely intervention in particular, and it gives important baseline information for policy makes in general. Furthermore, identifying particular risk factors is important for student's specific intervention. This study can be the baseline for other further related study in the area of education.

4. Discussion

This study was focused on the student attrition based on the data obtained from Hawassa College Teacher's Education, students who enrolled in 2009 E.C and attended the first five semesters. In the longitudinal analysis, we used the square

TABLE 2: Parameter estimates for the marginal linear-mixed effect model.

Variable	Estimate	Std.Err	95% CI	<i>p</i> value
Intercept	0.934	0.288	(0.368, 0.499)	0.009*
Sex (ref. = male)				
Female	-0.224	0.036	(-0.294, -0.154)	≤0.001*
Age	0.007	0.009	(-0.009, 0.024)	0.392
Distance (ref. = far)				
Near	-0.083	0.043	(-0.167, 0.003)	0.051
Moderate	-0.054	0.035	(-0.123, 0.015)	0.126
High school result	0.005	0.002	(0.001, 0.009)	0.002*
Family size	-0.024	0.007	(-0.038, -0.009)	0.004*
Support (ref. = high)				
Low	-0.063	0.062	(-0.184, 0.058)	0.305
Medium	-0.157	0.038	(-0.232, -0.082)	≤0.001*
Entrance result	0.017	0.002	(0.014, 0.022)	0.001*
Residence (ref. = rural)				
Urban	0.043	0.041	(-0.038, 0.123)	0.094
House head edu (ref. = certificate and above)				
Illiterate	0.023	0.059	(-0.092, 0.139)	≤0.001*
Primary	0.099	0.052	(-0.004, 0.201)	0.004*
Secondary	0.019	0.061	(-0.101, 0.139)	0.754

* indicates the significance of covariates at 5% level of significance.

TABLE 3: Parameter estimates, HR, and 95% CI under the survival modeling analysis.

Variables	Estimate	HR 95% CI	<i>p</i> value
Sex (ref. = female)			
Male	-0.929	0.395 (0.245, 0.636)	0.0050*
Entrance result	-0.179	0.836 (0.791, 0.883)	0.0010*
Support (ref. = high)			
Low	2.919	18.534 (4.513, 76.117)	0.3984
Medium	0.649	1.914 (0.424, 8.635)	0.0252*
Study time per day	-0.444	0.643 (0.518, 0.796)	0.4422
Place of residence			
Urban (ref. = rural)	0.425	1.412 (.383, 1.234)	0.0853
Income	-0.003	0.997 (0.996, 0.998)	0.3567
Interest of study			
Interested (ref. = no interest)	-2.558	0.077 (0.037, 0.162)	0.0040*
Marital status			
Single (ref. = married)	-1.592	0.204 (0.126, 0.329)	0.0042*

* indicates the significance of covariates at 5% level of significance.

root transformation of CGPA measurements to satisfy one of the basic assumption in data analysis supported by [18]. The trend of CGPA was attended, and the average stable level was noted till the end of the study period. CGPA was found evolving differently between female and male respondents based on the result from the two models, (i.e., separate and joint models). The progress level was higher for males as

compared to the females. This result also confirms and parallel to the study result obtained by [19].

From this study, the linear mixed effect model result revealed that sex of students, peer support, high school result of students, entrance result, family size, and house head education level were the significant predictors for the CGPA result of students in Hawassa College Teacher's Education.

TABLE 4: Parameter estimates for the joint modeling analysis.

Survival model				Longitudinal model			
Parameters	Estimate	SD	<i>p</i> value	Parameters	Estimate	SD	<i>p</i> value
Sex (ref = female)				Intercepts	14.363	0.078	≤0.001
Male	0.045	0.221	0.001	Sex (ref. = female)			
Support (ref. = high)				Male	7.541	0.010	0.0014
Low	0.805	0.385	0.0036	Entrance result	9.751	0.001	0.0016
Medium	0.546	0.567	0.0019	Support (ref. = high)			
Entrance result	0.150	0.018	≤0.001	Low	0.037	0.017	0.9704
Interest study				Medium	-3.915	0.011	≤0.001
Interest (ref. = no interest)	-0.840	0.012	0.0011	High school result	2.412	0.001	0.0314
Marital status (ref. = married)				Family size	3.085	0.002	0.0020
Single	-0.640	0.015	0.0221				
Association	-0.009	0.032	0.0056				

TABLE 5: Covariance parameter estimates under separate and joint analyses.

Separate analysis		Joint analysis	
Parameter	Estimate	Parameter	Estimate
Var(b_{0i})	8.271	Var(b_{0i})	11.058
Var(b_{1i})	0.035	Var(b_{1i})	0.006
Var(ϵ_i)	18.942	Var(ϵ_i)	15.387

This finding was agreed with the study result of [20]. The baseline CGPA was shown to be significantly determining the student's retention progression rate. This implies that the higher value of CGPA results is one means of a better retain of students in the study program which is in line with the study findings of [21].

From the plots of survival analysis indicated in Figure 1, male students' retention had slightly higher survival rate than female student's retention. Thus, the survival times are found to be significantly different among male and female students. We assessed each factor through univariate Cox regression model and found the variables sex, entrance result, peer support, interest of study, and marital status were statistically significant; which implies the variables had either positive or negative impact on the retention of students while the other variables income, study time, and place of student were not found to be significant. This finding is consistent with the study findings of [22].

Furthermore, this study also tried to assess the effect of longitudinal measure of the CGPA of students in each semester on the attrition status of the student. The quantification on the effect of this repeatedly measure CGPA is fundamental to understand the trend and take appropriate intervention mechanism.

Moreover, the estimates of the parameters for the separate and joint models are quite similar to each other but not identical. The estimate of the association parameter due to the slope of CGPA is negative (-0.00907), indicating that CGPA result is negatively associated with the risk of attrition of students from semester retention. This indicates

that an increasing trend in the CGPA result of students undergoing class handling significantly reduces the risk of attrition of students. This finding was also in agreement with the studies of [22, 23] that shows the significance of the shared parameter which links the two processes and the reduction in the standard error of the parameter estimates when compared to independent model estimates. This suggests the need for a joint analysis of this data compared to the use of separate models. The estimated association parameter (α) in the joint model is -0.00907 and statistically significant at 5% level of significant. This indicates that there is strong evidence of association between the effects of the longitudinal outcome to the risk of attrition, implying initial higher values of the CGPA result associated with a better retention of students in favor with the study result of [24]. The residual variability was smaller in joint analysis (15.3876) as compared to the relative linear mixed effect analysis (18.943) that was adjusted for the correlation between the responses.

5. Conclusion

Several studies were conducted on students' CGPA and attrition status separately. However, there were only few studies conducted using joint model of the two together about students' CGPA and attrition status in education. The mixed effect model was confirmed to be adequate for the prediction of CGPA of students based on the available variable of health determinants. The pattern of mean change in CGPA revealed a linear distribution that decreased over time in the five semesters.

The predictor variables included in this study sex, high school result, family size, peer support, residence distance, entrance exam, and house head educational level were found statistically significant. In the same expression, the covariates sex, entrance exam, peer support, marital status, residence, and interest of field study were found to be significantly associated with time to event. Among the subjects of the study, 73.6% were retained and 26.4% attrite due to academic delays or other individual cases. This study also confirmed strong association between the effects of the

longitudinal outcome to the risk of attrition. The results of both the separate and joint modeling are consistent. However, the use of a joint modeling compared to independent models adjusted for correlation between the responses indicates that more adequate and efficient inferences can be made using joint model estimates. This means that joint modeling can benefit the analysis of the longitudinal measure of CGPA and survival time-to-event outcomes. This study tried to assess the effect of longitudinal measure of the commutative grade point average (CGPA) of students in each semester on the attrition status of the student.

5.1. Limitations of the Study. A limitation exists in the ability to use the findings of this study to make generalizations about use of joint modeling to study of students' CGPA and attrition status in college education. One of the limitations of this study was it was conducted in only one college in a particular location in the Hawassa, Ethiopia. The other limitation of this study was it includes only a batch of 2009 E.C enrolled students. This study was also restricted only on regular students but not included other programs. Lastly, this study takes sample from only natural science students who were more homogenous but not from other colleges, which may not be reliable generalization to be made about attrition in college education in general.

Abbreviations

AIC: Akaike information criterion
 BIC: Bayesian information criterion
 CGPA: Cumulative grade point average
 HCTE: Hawassa College of Teacher Education
 HR: Hazard ratio
 LMM: Linear mixed model
 CPH: Cox proportional hazard.

Data Availability

The data, which is available with the the first author, will be made available upon reasonable request.

Conflicts of Interest

No potential conflict of interest was reported by the authors.

Authors' Contributions

MT and DBB designed, drafted, analyzed, and interpreted the results. MAZ and SCA participated in designing the methodology and data analysis, critically read the manuscript, and gave constructive comments for the development of the manuscript. All authors have contributed in manuscript preparation.

Acknowledgments

The authors would like to acknowledge all natural science students in Hawassa College of Teacher Education for their participation in this study.

References

- [1] C. D. J. N. J. Lamichhane, "Understanding the Education Philosophy and Its Implications," *NCC Journal*, vol. 3, no. 1, pp. 24–29, 2018.
- [2] F. Eshetu, A. Guye, G. Kelemework, and S. Abebe, "Determinants of students vulnerability to attrition in higher education: evidence from Arba Minch University, Ethiopia," *Educational Research and Reviews*, vol. 13, no. 15, pp. 570–581, 2018.
- [3] D. D. J. T. A. B. T. Tarasi, "An analysis of students' makeup-exam results and performance in undergraduate biology courses," *Courses*, vol. 82, no. 5, pp. 308–314, 2020.
- [4] Y. Geng and Y. Yan, "Higher education and science popularization: Can they achieve coordinated growth?," *PLoS one*, vol. 16, no. 9, p. e0256612, 2021.
- [5] R. Morgan, S. Bryan, K. Meldrum, B. Mathiesen, and A. Hill, *Using an Adaptive e-Learning Curriculum to Enhance Digital Literacy: Challenges and Opportunities*, National Association Of Biology Teachers, Reston Va, 2016.
- [6] MOE, *Ministry of education report in Ethiopia*, 2015.
- [7] M. Hussain and P. S. A. Khader, "Analyzing student attrition and retention in engineering colleges of South India," *International Journal of Data Engineering (IJDE) Singaporean Journal of Scientific Research*, vol. 6, no. 4, pp. 213–218, 2014.
- [8] C. T. Bull, C. Manceau, J. Lydon, H. Kong, B. A. Vinatzer, and M. Fischer-le Saux, "*Pseudomonas cannabina* pv. *cannabina* pv. *nov.* and *Pseudomonas cannabina* pv. *alisalensis* (Cintas Koike and Bull, 2000) comb. nov., are members of the emended species *Pseudomonas cannabina* (ex Šutič & Dowson 1959) Gardan, Shafik, Belouin, Brosch, Grimont & Grimont 1999," *Systematic and applied microbiology*, vol. 33, no. 3, pp. 105–115, 2010.
- [9] P. Alter, T. J. T. E. Haydon, and S. Education, "Characteristics of effective classroom rules: a review of the literature," *Teacher Education and Special Education*, vol. 40, no. 2, pp. 114–127, 2017.
- [10] E. F. Villella and M. J. N. J. Hu, "A factor analysis of variables affecting the retention decision of nontraditional college students," *NASPA Journal*, vol. 28, no. 4, pp. 334–341, 1991.
- [11] R. N. Cochran and F. H. J. A. C. Horne, "Statistically weighted principal component analysis of rapid scanning wavelength kinetics experiments," *Analytical Chemistry*, vol. 49, no. 6, pp. 846–853, 1977.
- [12] G. Verbeke, G. Molenberghs, and G. Verbeke, "Linear Mixed Models for Longitudinal Data," in *Linear Mixed Models in Practice*, pp. 63–153, Springer, 1997.
- [13] D. MeleseChekole, G. Gebrerufael, and Z. Asfaw, *Statistical joint modeling on longitudinal body weight and CD4 cell progression with survival time-to-death predictors on HIV/AIDS patients in Mekelle General Hospital, Ethiopia*, 2020.
- [14] A. H. Dessiso and A. T. Goshu, "Bayesian joint modelling of longitudinal and survival data of HIV/AIDS patients: a case study at bale robe general hospital, Ethiopia," *American Journal of Theoretical and Applied Statistics*, vol. 6, no. 4, pp. 182–190, 2017.
- [15] S. R. Lipsitz, J. Ibrahim, and G. Molenberghs, "Using a Box-Cox transformation in the analysis of longitudinal data with incomplete responses," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 49, no. 3, pp. 287–296, 2000.
- [16] G. L. Hickey, P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona, "Joint modelling of time-to-event and multivariate

- longitudinal outcomes: recent developments and issues,” *BMC medical research methodology*, vol. 16, no. 1, pp. 1–15, 2016.
- [17] D. Rizopoulos, “JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data,” *Journal of Statistical Software*, vol. 35, no. 9, pp. 1–33, 2010.
- [18] S. Geremew, D. Tesfaw, and T. Getiye, “Application of longitudinal measured CD4+ count on HIV-positive patients following active antiretroviral therapy: a case of Debre Berhan Referral Hospital,” *El Hospital*, vol. 3, no. 2, p. 34, 2018.
- [19] G. Grover, P. Swain, V. Deo, and M. Varshney, “A joint modeling approach to assess the impact of CD4 cell count on the risk of loss to follow up in HIV/AIDS patients on antiretroviral therapy,” *International Journal of Statistics and Applications*, vol. 5, no. 3, pp. 99–108, 2015.
- [20] P. J. S. O. Bawa, “Retention in online courses: exploring issues and solutions—a literature review,” *Sage Open*, vol. 6, no. 1, p. 215824401562177, 2016.
- [21] V. J. L. E. Tinto, “Building community,” *Liberal education*, vol. 79, no. 8, pp. 16–21, 1993.
- [22] V. Yazisiz, M. Göçer, F. Erbasan et al., “Survival analysis of patients with Sjögren’s syndrome in Turkey: a tertiary hospital-based study,” *Clinical rheumatology*, vol. 39, no. 1, pp. 233–241, 2020.
- [23] A. Lawrence Gould, M. E. Boye, M. J. Crowther et al., “Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group,” *Statistics in medicine*, vol. 34, no. 14, pp. 2181–2195, 2015.
- [24] A. Abossie and M. Seid, “Assessment of the prevalence of intestinal parasitosis and associated risk factors among primary school children in Chencha town, Southern Ethiopia,” *BMC public health*, vol. 14, no. 1, pp. 1–8, 2014.