STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, Month 2025, pp 0–17.
Published online in International Academic Press (www.IAPress.org)

Survival time in higher education program after a dropout using modified survival function: A retrospective study to predict average graduation time and factors leading to early dropout

Intisar Ahmad Siddiqui[1,2,*], Siti Meriam Zahari[1], Nor Azura Md Ghani[1], Muhanad S. Alhareky[3], Jehan, A. Alhumaid[3], Maram A. Alghamdi[4], Abdur Rasheed[5]

[1] *School of Mathematical Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*
[2] *Department of Dental Education, College of Dentistry, Imam Abdulrahman Bin Faisal University, Saudi Arabia*
[3] *Department of Preventive Dental Sciences, College of Dentistry, Imam Abdulrahman Bin Faisal University, Saudi Arabia*
[4] *Department of Substitutive Dental Sciences, College of Dentistry, Imam Abdulrahman Bin Faisal University, Saudi Arabia*
[5] *School of Public Health, Dow University of Health Science, Karachi, Pakistan*

Abstract The case recurrence survival model in terms of case retention after at least one dropout is still difficult to investigate, and some concrete framework is required to derive the survival model in order to achieve more precise results and reduce the magnitude of bias that may occur during dropout, reappearance, and retention either until the last or dropout again. The data for this retrospective, longitudinal study was retrieved based on undergraduate program students spanning eight years. All relevant information was entered into the data worksheet of SPSS-29.0 (IBM product, USA). Syntax programming of survival algorithm was developed using the statistical programming software R version-4.2.1, and survival parameters were generated. The survival probability of the existing model compared to the modified function showed minimal differences. The survival rate was 94% in the first year of study, with a gradual decline of 1% - 3% annually, reaching 91.6% by the end of the fifth year. The average survival time for the existing survival function was 4.666 ± 3.25 years, whereas the modified function exhibited a higher average of 5.584 ± 3.61 years. Similarly, the mean graduation time was slightly higher for the modified function (6.10 ± 0.302 years) compared to the existing model (6 ± 0.0 years). Model performance, assessed using the R^2 value revealed the modified survival model than existing model was more accurate and preferable (i.e. 0.903 vs. 0.808). It was concluded that dropout cases, which were censored in the existing survival model, played a significant role in estimating students' survival time and the program's graduation time. Hence, the modified function can be preferred when the first event to time doesn't represent the final outcome.

Keywords Survival analysis, student dropout, higher education, retention, recurrence

AMS 2010 subject classifications 62M10, 93A30

DOI: 10.19139/soic-2310-5070-2499

1. Introduction

Survival analysis, commonly referred to as time-to-event analysis [1], examines the occurrence of an event over a specific time-period "T". It assesses the likelihood of an event occurring at any time point "t" within this period. Such events may include death, the end of functional life, defects, collapse or outbreak under the certain conditions or exposures [2, 3]. The exposure may be any diagnosis, intervention, process or trail which is followed up for a

*Correspondence to: Intisar Ahmad Siddiqui (Email: iasiddiq@iau.edu.sa). Department of Dental Education, College of Dentistry, Imam Abdulrahman Bin Faisal University, Saudi Arabia (13441).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

certain length of time "T". The time clock starts from the exposure and progress till occurrence of event is called survival time.

1.1. A brief review of existing methods

An intriguing application of survival analysis is observed in academia, particularly in higher education programs, where it tracks students' progress from enrollment to a program completion. However, it is not guaranteed that all students who enroll will persist until the end, let alone graduate on time. Student attrition presents significant challenges for both individuals and institutions [4]. Multiple factors may contribute to a student's dropout from a program, such as program selection, shifting priorities regarding program scope, lack of interest, social or financial constraints [5, 6]. Over the years, various strategies have been implemented to mitigate student attrition and enhance retention, benefiting both students and the institutions. These measures include promoting equity in education, providing financial assistance through institutional programs for underprivileged students, offering counseling and advisory services for social challenges, organizing extra training sessions to improve learning skills, allowing fair opportunities to retake courses or components, encouraging participation in sports and extracurricular activities to manage stress, and facilitating individual consultation with faculty during school and outside school hours [7, 8].

An exploratory analysis of higher education revealed that 64.1% of students remained censored after the first year, while 34.4% dropped out. The highest dropout rates were observed in information technology (51.6%) and food technology (47.7%). In contrast, engineering programs had the highest retention rates, with hazard ratios of 0.803, 0.458, and 0.565, respectively. Low secondary school knowledge and lack of motivation to study engineering were identified as significant factors to attrition in engineering programs [9].

Several factors influencing higher education dropouts, retention, and graduation were identified in a recent Brazilian study. The study highlighted that higher education dropout cannot be attributed solely to the students. Instead, operational and strategic factors also play significant roles. From a strategic perspective, government policies related to access to higher education and institutional efforts to attract and retain undergraduate students are critical. From an operational standpoint, department heads emphasized the need for educators to revise their teaching and assessments methods. The study further demonstrated that variables such as age, marital status, race, and type of high school (public or private) had no bearing on the length of time students took to graduate from high school or the percentage of students dropped out [10].

Survival analysis is a valuable tool for analyzing time-to-event data. However, longitudinal data often include variations, such as permanently lost cases, time-specific attrition, retention and resumption. In these scenarios, survival functions face challenges in accurately accounting for attrition and estimating the actual event at a given time point "t". If the time-varying nature of outcomes is overlooked, it can introduce a "time-based bias" into the survival analysis [11].

Many studies employing survival analysis to predict time-to-event models primarily address missing data for covariates rather than outcome variable such as attrition. However, developing analytical models that account for retention following at least one dropout remains a significant challenge. A robust methodology is required to construct survival models that yield more accurate results while maintaining bias associated with dropout, re-enrollment, and retention until completion or subsequent drop out.

1.2. Problem statement

In this context, survival analysis serves as a valuable tool for decision makers and policymakers in estimating survival time, probabilities of hazardous events, standard costs and pricing, as well as customer preferences and satisfaction. However, there remains significant potential for the evolution and deeper understanding of its fundamental concepts, applications, and presentations in research publications. Enhancing the accessibility and ease of use of statistical software could promote the broader adoption of underutilized survival analysis techniques, such as the "accelerated failure time model". The topic requires special attention, both in research and application, as numerous model extensions could prove beneficial in the future. Accordingly, this research aims to derive a survival function for sample retention and resumption over a specified time-period, defined as an "event" at a certain "time" point. It seeks to develop a framework for sample attrition at time point t_i (where $t_i \neq t_i$), which resumes at another time point t_j (where $t_j \neq t_j$) as a non-event and reintegrate into the study sample.

1.3. Scope/Outcomes & benefits of the presented research

i. This research may offer a survival tool particularly useful in educational and healthcare follow-up studies to estimate survival rates for participants' dropouts and retention.
ii. The research approach could be valuable for long-term higher education programs, particularly self-financed full-time bachelor programs, as well as those offered through virtual platforms, and postgraduate programs that anticipate a larger risk of student dropout and resumption at a certain time point.
iii. The findings may provide future researchers with enhanced statistical models to utilize early censored data after recurrence and predict key covariates contributing to the risk of dropouts and be helpful for resumption.
iv. The model may assist institutions in developing targeted interventions, such as academic counseling or financial aid, before dropout risk increases, as well as assessing the impact of these interventions on long-term student recurrence in the program.

2. Conceptual framework of proposed Survival Function for Dropout, Recurrence & Retention

2.1. Existing survival and hazard function

Three possible events of survival function for real-life data of higher education program students during the stipulated time period may be denoted as follows:

- Retention: Attended the program in continuation till the end.
- Dropout: Discontinuation of the program for a period of one or more semesters (or > 6 months) with a zero-grade point average (GPA).
- Retention after one dropout: Resumption of program after one dropout and retained in continuation till the end.

According to Kalbfleisch (2002), the Kaplan-Meier survival model [12] can be expressed as follows:
If Y is the outcome, it can be expressed as:

$$Y = \text{Surv}(t, e) \tag{1}$$

where $t(0 \leq t \leq n)$ denotes time and $e(0, 1)$ represent the event.
The probability of survival is given by the survival function, $S(t)$:

$$S(t) = P(T > t) \tag{2}$$

This represents the probability of survival (retention until the end of a stipulated time period), where $t = 0, 1, 2, \ldots, T$.
The Kaplan-Meier survival estimate,

$$S(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i} \tag{3}$$

or equivalently,

$$S(t) = \prod_{t_i < t} 1 - \frac{d_i}{n_i} \tag{4}$$

The probability of dropout or a censored event is represented as:

$$HAZ = P(T < t + \Delta | T > t) \tag{5}$$

This is the probability of dropout or censoring at time T, conditional on retention.
According to Cox (1972), the Cox regression hazard ratio (HR) is defined [13] as follows:

$$HR = \frac{(HAZ_{x=1})}{(HAZ_{x=0})} \tag{6}$$

where $x = 1$ indicates exposure and $x = 0$, indicates non-exposure.
Additionally, the probability of survival can be expressed as:

$$P(Surv) = 1 - HAZ \tag{7}$$

The cumulative distribution function (CDF) for survival time is T, representing the probability that survival time is less than or equal to time t, is given as:

$$F(t) = P(T \leq t), \quad \text{where } T = 0, 1, 2, \ldots, t. \tag{8}$$

2.2. Parametric model (Weibull distribution)

The parametric approach for the probability of survival may be Weibull distribution. The basic Weibull distribution follows shape parameter $k \in (0, \infty)$, is a continuous distribution on $(0, \infty)$, with probability density function:

$$f(t; k, \lambda) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} \exp\left(-\frac{t}{\lambda}\right)^{k}, \quad t \geq 0 \tag{9}$$

The probability density function g is given by

$$F(t, k, \lambda) = 1 - \exp\left(-\frac{t}{\lambda}\right)^{k}, \quad t \geq 0 \tag{10}$$

Where:
k is the shape parameter (determine the behavior of the dropout rate)
λ is the scale parameter (determines the spread of the distribution)
t is the time to event.
Mean $= E(t) = \lambda \prod \left(1 + \frac{1}{k}\right)$
Variance $=$ Var $(t) = \lambda^2 \prod \left(1 + \frac{1}{k}\right) - \left[\lambda \prod \left(1 + \frac{1}{k}\right)\right]^2$

2.3. Proposed modified survival model (Algorithm)

In the proposed model, the time-to-event (t) during the specified time period T, encompassing retention until the end, dropout and resumption, follows a probability density function p(t). The survival function $S(t)$ can be characterized as:

$$\begin{aligned} S(t) &= P(T > t) \\ &= \int_{t}^{\infty} P(T)dT \quad \text{for } t > 0 \text{ (continuous data).} \end{aligned}$$

The survival function $S(t)$ may capture the probability that the event e does not occur until time t. Hence,

$$S(t) = \prod_{t_i < t} \frac{n_i + c_i - d_i}{n_i} \tag{11}$$

for the censored retained survival function, where d_i and c_i denote dropouts and censored observations, respectively, for all $i = 0, 1, 2, \ldots, n$. Alternatively, it can be expressed as:

$$S(t) = \prod_{t_i < t} 1 + \frac{c_i}{n_i} - \frac{d_i}{n_i} \tag{12}$$

The hazard function is defined as:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t)}{\Delta t} \tag{13}$$

where Δt represents the time interval for the censored event. This measures the event rate at time t, conditioned on survival until t:

$$\lambda(t) = \frac{P(t)}{S(t)} \tag{14}$$

The cumulative survival function is:

$$S(t) = e^{-\int_0^t \lambda(\tau)dr} \tag{15}$$

2.4. Survival Time Estimation

i. Mean survival time: $E(T) = \int_0^\infty S(t)\,dt$,

ii. Variance survival time: $\mathrm{Var}(T) = E(T^2) - \big[E(T)\big]^2$,

where $E(T^2) = 2\int_0^\infty t\,S(t)\,dt$ and $\big[E(T)\big]^2 = \Big[\int_0^\infty S(t)\,dt\Big]^2$ and the standard deviation, SD, is given by $\sqrt{\mathrm{Var}(T)}$.
Hence, 95% confidence interval for mean survival time:

$$E(T) \pm Z_{0.05/2} * S.E(T), \tag{16}$$

where $S.E(t) = S.D(t)/\sqrt{n}$.

2.5. Testing of performance model (Curve fitting)

To explore the best curve estimation, linear regression and exponential curves were fitted:
Linear regression model:

$$y = f(x, \beta) + \epsilon \tag{17}$$

where, y is the dependent variable (time-to-event t), x is the independent variable (dropout), β is the slope, and ϵ is error term.
Exponential function:

$$S(t) = e^{-\lambda t}, \tag{18}$$

where t is time until the event occurs.
The coefficient of determination, R^2 was calculated as follows:

$$R^2 = \frac{SS\ residual}{SS\ total} \tag{19}$$

where, $SS\ residual = \sum(y_i - y_i')^2$, where y_i' is the estimated time.
$SS\ total = \sum(y_i - E(y_i))^2$, $E(y_i)$ is the mean of y "time"'.

2.6. Assumptions of existing survival model

i. The survival times of different subjects are assumed to be independent of each other.
ii. The events are independent.
iii. The model assumes the correct functional form of the relationship between the covariates and the hazard function.
iv. Censoring is non-informative, i.e. the reason for censoring is not related to the likelihood of the event occurring.

2.7. Assumptions of modified survival model

i. The survival times of different subjects are assumed to be independent of each other.
ii. The number of samples is not fixed for a specified time, but it may change due to recurrent events over the time.
iii. The event is non-mutually exclusive, and occurs at a single time point only once, while a terminal event was considered if the subject dropout and didn't continue till to the end of study.
iv. The probability of occurrence of an event may change from time to time due to recurrence of the event.
v. Censoring is non-informative, i.e. the reason for censoring is not related to the likelihood of the event occurring.
vi. The model assumes the correct functional form of the relationship between the covariates and the hazard function.

3. Application of Modified Survival Function

3.1. Materials and Methods

This retrospective, longitudinal study was conducted at the College of Dentistry, Imam Abdulrahman Bin Faisal University, Saudi Arabia, from December 15, 2023 to April 14, 2024. Ethical approval for this study was obtained from the Institutional Review Board (IRB) of Imam Abdulrahman Bin Faisal University (IRB No. 2023-02-237), to ensure compliance with research ethics and appropriate data retrieval protocols.

Data from undergraduate program enrollment cohorts over the last eight years (2015-2016 to 2022-2023) were retrieved from the College of Dentistry of Imam Abdulrahman Bin Faisal University. Enrollment data were obtained from the Deanship of Student Registration, while information on potential causes of dropout was sourced from the registrar's office. Students were considered eligible for the study if they enrolled in the undergraduate dental surgery program at Imam Abdulrahman Bin Faisal University in Saudi Arabia, regardless of gender or age, and had at least 50% attendance in the first semester after the program began. Students who were expelled from the program due to violations or disciplinary proceedings, deported or imprisoned for criminal activity, or had special needs or disabilities were not eligible for the study.
Demographic characteristics included: gender, parents alive, parental education level, socio-economic status, and marital status. Students' dropout, retention, and survival outcome related data included:

- Student's outcome: Pass or Fail
- Graduation time duration (years)
- Student's performance grade point average
- Causes of dropout:
- Constantly low performance, lack of interest in the opted program, health issues, financial concerns, migration and social upset (e.g. accident, death of loved ones, violence, life threats etc.).

3.2. Sample Size and Sampling Technique

Assuming a student's attrition rate in higher education in Malaysia of 17.5% reported by Shilbayeh et al. [14], the representative sample size for a specified higher education program was calculated using the formula:

$$n = \frac{Z^2_{(1-\propto/2)}P(1-P)}{d^2}$$

where $Z^2_{(1-\propto/2)} = (1.96)^2$, $P = 0.175$, and $d = 0.05$.
Given a 5% margin of error (α), 80% power ($1-\beta$), a 95% confidence interval and an anticipated proportion $p = 0.175$, the required representative sample size was calculated to be 222 students.
Non-probability, consecutive sampling technique was used to select the consecutive students' information in the cohorts from the student's registration database.

3.3. Data Collection Procedure

Following the approval of the research protocol by the ethical board, real-life data for the examination of the survival model was retrieved. The data encompassed information from higher education programs spanning 10 consecutive years, adhering to the sample selection criteria.

Inclusion and exclusion criteria were strictly followed to control bias. The student's release for permanent dropout (i.e., disciplinary actions, violations, etc.) and reasons for temporary dropout were available to document, in order to rule out students as per exclusion criteria. Relevant data included students' enrollment, attrition, retention, demographic characteristics including gender, marital status (married or single), both parents alive, parents' education level in terms of postgraduate, bachelor, and intermediate-secondary school qualifications, socio-economic status based on family income per month (i.e., high (>SAR 15000), middle (SAR5000-15000), and low class (<SAR5000)), reasons for dropout, and graduation durations. All data were systematically entered into a data worksheet for analysis.

3.4. Statistical Analysis

Syntax programming was developed in R-Gui version 4.2.1 to implement the proposed algorithm, and the test data were processed accordingly. The Shapiro-Wilk test for normality was applied to determine the distribution of the data. A non-parametric model was used for non-Gaussian distributions, while a semi-parametric model was applied for Gaussian distributions. Categorical variables including gender, both parents alive, parents' education level, socio-economic status, marital status of the student, permanent and temporary dropout, and reason of dropouts were presented as frequencies and percentage. The survival time and graduation time were presented in terms of mean $\pm$ SD. Statistical significance was set at $p \leq 0.05$.

3.5. Measures of Model Performance

The model's performance was evaluated using the Piccolo method, applying an autoregressive model to cluster time-series data. Curve estimation for the best fit was determined using the least squares method. The accuracy of predicted graduation times was assessed by comparing the existing Kaplan-Meier (KM) survival model with the modified survival function. The coefficient of determination (R^2) was used to evaluate and compare the curve estimation results.

Sensitivity analysis was performed by applying the receiver operating characteristic curve (ROC) approach to compare the area under the curve (AUC) of conventional versus modified models. A model was considered sensitive if the AUC value was higher. The coordinate points were compared for both models to show the highest sensitivity, specificity, PPV, and NPV values at a certain survival time value.

4. Results

4.1. Demographic Characteristics

Real-life data from a higher education program were utilized to test the modified survival function. The study included 324 students enrolled since 2016 in the six-year (maximum) study program of Bachelor of Dental Surgery (BDS) after completing the first preparatory year. These students represented six enrolment cohorts, comprising three male cohorts and three female cohorts, expected to complete the program in 2021, 2022, and 2023, respectively. The number of females students, 167 (51.5%) was slightly higher than the number of male students 157 (49.5%). The proportion of married students was 17.9%, consistent with socio-cultural norms about early weddings. However, 82.1% of students were single and relied on their parents. A total of 34 (10.5%) had only one parent alive, either a father (5.6%) or a mother (4.9%). A bachelor's degree was the most common type of parental education, with 78.4% of fathers and 77.2% of mothers having earned one. The vast majority of participants (81.8%) were from the middle to upper socioeconomic class, while 14.8% decided not to disclose their family income, and only 11 (3.4%) were from the lower socioeconomic class.

The final outcome revealed that 27 students (8.3%) permanently dropped out, 28 students (8.7%) temporarily dropped out, and 269 students (83.0%) completed the six years program, irrespective of their final examination status (passed or delayed). The majority of temporary dropouts (4.3%) among a total of 324 students were due to severe COVID-19 infections, followed by domestic problems (4.0%) and roadside accidents (0.3%). Transferring to another program was the leading cause of permanent dropouts (3.4%), followed by consistent poor performance (3.4%).

4.2. Student's Retention Life-Tables Using Existing and Modified Survival Functions

In the existing survival model, the outcome of students was denoted as d_i (event: permanently dropout case) and, $d_i^{\wedge}$ (temporary dropout but active student). The probability of survival was 94% during the first year of study, with a gradual annual decrease of 1% to 3%, and 91.6% survival by the end of the fifth year. The parametric approach by using the Weibull distribution showed a very nominal difference of survival probability distribution such that almost similar survival till the fourth year, and then a drop of 1% and 3% till fifth and sixth year respectively.

Table 1. Survival probability distribution for existing non-parametric KM survival function and parametric Weibull distribution approaches.

# of years	**# of student entering (n_i)**	**# of permanently Dropout (d_i)**	**# of temporary Dropout ($d_i^{\wedge}$)**	$(1-d_i/n_i)$	**KM Survival function $f(S)=\prod(1-d_i/n_i)$**	**Weibull function $f(t;k,\lambda)$**
1	324	0	0	1.0000	1.0000	1.0000
2	324	18	2	0.9444	0.9444	0.9999
3	304	4	10	0.9868	0.9320	0.9989
4	290	5	3	0.9828	0.9159	0.9946
5	282	0	9	1.0000	0.9159	0.9827
6	273	0	269	1.0000	0.9159	0.9589

d_i: Event, i.e. permanently dropout case; $d_i^{\wedge}$: Temporarily dropout but active student

The modified survival function expanded the categorization to include c_i, representing students who resumed the program after a temporary dropout. Notably, 28 students who temporarily dropped out and were censored in the existing model resumed the program after deferring one or more semesters and continued with subsequent cohorts. The survival probabilities calculated through the modified model showed minimal differences compared to the existing model (Table 2).

Table 2. Modified Survival function for dropout and resumption.

# of years	**# of student at risk (n_i)**	**# of permanently dropout (d_i)**	**# of temporary dropout ($d_i^{\wedge}$)**	**# of Resume (c_i)**	$(1-d_i/n_i)$	**Survival function $f(S)=\prod(1-d_i/n_i)$**
1	324	0	0	0	1.00000	1.0000
2	324	18	2	1	0.94444	0.9444
3	305	4	10	7	0.98689	0.9321
4	298	5	3	6	0.98322	0.9164
5	296	0	9	6	1.00000	0.9164
6	293	0	4	6	1.00000	0.9164
7	295	0	269	2	1.00000	0.9164

d_i: Event, i.e. permanently dropout case; $d_i^{\wedge}$: Temporarily dropout but active student; c_i: resumed after a dropout

4.3. Comparison of Student's Survival Time Using Existing and Modified Survival Functions

When comparing survival times, the area under the survival curve revealed a mean ± SD survival time of 4.666 ± 3.25 years (95% confidence interval: 4.31–5.02) using the existing model. However, the modified survival model demonstrated a significantly higher mean survival time of 5.584 ± 3.61 years (95% confidence interval: 5.19–5.98), as illustrated in Figures 1 and 2. This supports the hypothesis that incorporating resumed students into the analysis results in longer survival times, highlighting the importance of accounting for temporary dropouts when measuring student retention. However, a non-comparable mean ± SD survival time of 10.02 ± 2.81 years was found by applying Weibull distribution parameters (shape "$k = 4$" and scale "$\lambda = 11.04$").

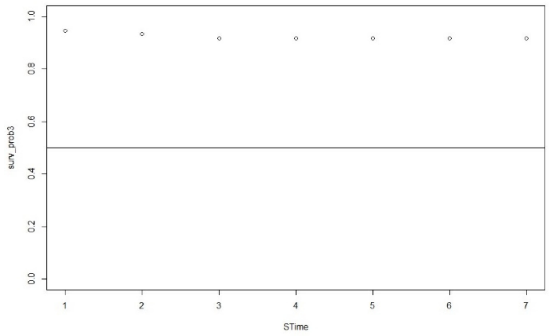

Figure 1. Survival curve of existing KM survival function. Mean ± S.D Survival Time: 4.666 ± 3.252. x-axis: Survival time, and y-axis: probability of survival

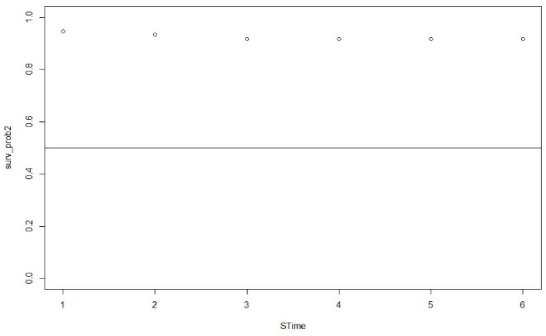

Figure 2. Survival curve of modified survival function. Mean ± S.D Survival Time: 5.584 ± 3.61. x-axis: Survival time, and y-axis: probability of survival

4.4. Comparison of graduation time using existing and modified survival functions

Graduation time comparisons also underscored the limitations of the existing model. The mean graduation time for 269 graduates using the existing model was 6 ± 0.0 years. In contrast, the modified model accounted for 297 graduates, including those who initially dropped out and resumed their studies, yielding a mean graduation time of 6.10 ± 0.302 years (95% confidence interval: 6.07–6.13). These findings demonstrate that measuring graduation time solely from the initial event may not provide an accurate representation of program completion dynamics.

4.5. Comparison of performance of two survival models

The performance of the two survival models was evaluated using the R^2 value. The modified survival model outperformed the existing Kaplan-Meier model (i.e. 0.903 vs. 0.808), indicating greater accuracy and reliability in predicting student retention and survival outcomes (Table 3). This further validates the modified model as a preferable approach for studying enrollment program retention.

Table 3. Performance of model comparison

Models	**R**	**Coefficient of Determination (R^2)**	**Adjusted R square**	**Std. Error of the Estimate**
Existing survival model	0.808	0.685	0.684	0.156
Modified survival model	0.950	0.903*	0.902	0.086

*The R^2 value of the modified survival model is greater than the value of existing survival model, revealing the modified survival model as the preferable and accurate model.

4.6. Factors affecting student's retention and cause early dropouts

Finally, factors influencing student retention and early dropout were analyzed. Due to confidentiality constraints, a limited number of factors including gender, parental education, socio-economic status, marital status of the student and reason for dropout were considered as covariates in the Cox regression analysis. Among these, the reason for dropout emerged as a significant factor affecting student survival in higher education programs (Table 4). This highlights the critical role of understanding dropout reasons in designing interventions to improve retention rates.

Table 4. Cox regression analysis for predictive factors affecting the studentś survival in a higher education program

Models	Covariates	B	SE	Wald	df	Sig.	Exp(B)
Existing survival model	Gender	-0.277	0.497	0.311	1	0.958	0.970
	Both parents alive	0.019	0.418	0.002	1	0.964	1.019
	Father's education	-1.027	0.599	2.938	1	0.087	0.358
	Mother's education	0.460	0.580	0.628	1	0.428	1.584
	Socio-economic status	0.084	0.284	0.087	1	0.767	1.088
	Marital status	-0.632	0.842	0.563	1	0.453	0.532
	Reason of dropout	-1.228	0.300	16.736	1	0.000*	0.293
Modified survival model	Gender	-0.001	0.596	0.000	1	0.999	0.999
	Both parents alive	-0.122	0.426	0.082	1	0.774	0.885
	Father's education	-0.979	0.592	2.737	1	0.098	0.376
	Mother's education	0.434	0.556	0.610	1	0.435	1.543
	Socio-economic status	0.023	0.277	0.007	1	0.934	1.023
	Marital status	-0.844	0.862	0.957	1	0.328	0.430
	Reason of dropout	-1.311	0.294	19.949	1	0.000*	0.270

* The reason of dropout was the significant covariate in both existing as well as modified survival models.

4.7. Comparison of model sensitivity

Figure 3 shows that the modified model surpassed the existing survival model in terms of efficiency, as indicated by the area under the normal curve AUC (95% C.I.) values of 1.00 (1.0-1.0) and 0.990 (0.982-0.999), respectively. Table 5 compares coordinate points for the highest sensitivity and specificity values. At a cutoff of over 3.5 years, the existing survival model had the highest sensitivity (96.3%), specificity (100%), PPV (100%), and NPV (99.7%). In contrast, the adjusted survival model predicted $>$ 5-year survival with 100% sensitivity, specificity, PPV, and NPV.

Table 5. Comparison of coordinate points for a high sensitivity, specificity, positive predictive value, and negative predictive value of survival models

Survival time (existing model)					Survival time (modified model)				
Cut-off points	Sen (%)	Spe (%)	PPV (%)	NPV (%)	Cut-off points	Sen (%)	Spe (%)	PPV (%)	NPV (%)
>2.5	99.3	67.7	21.2	99.9	>2.5	100.0	67.7	21.4	100.0
>3.5	96.3	100.0	100.0	99.7	>3.5	100.0	81.5	32.9	100.0
>5	92.7	100.0	100.0	99.3	>5	100.0	100.0	100.0	100.0

Sen: Sensitivity, Spe: Specificity, PPV: Positive predictive value, NPV: Negative predictive value

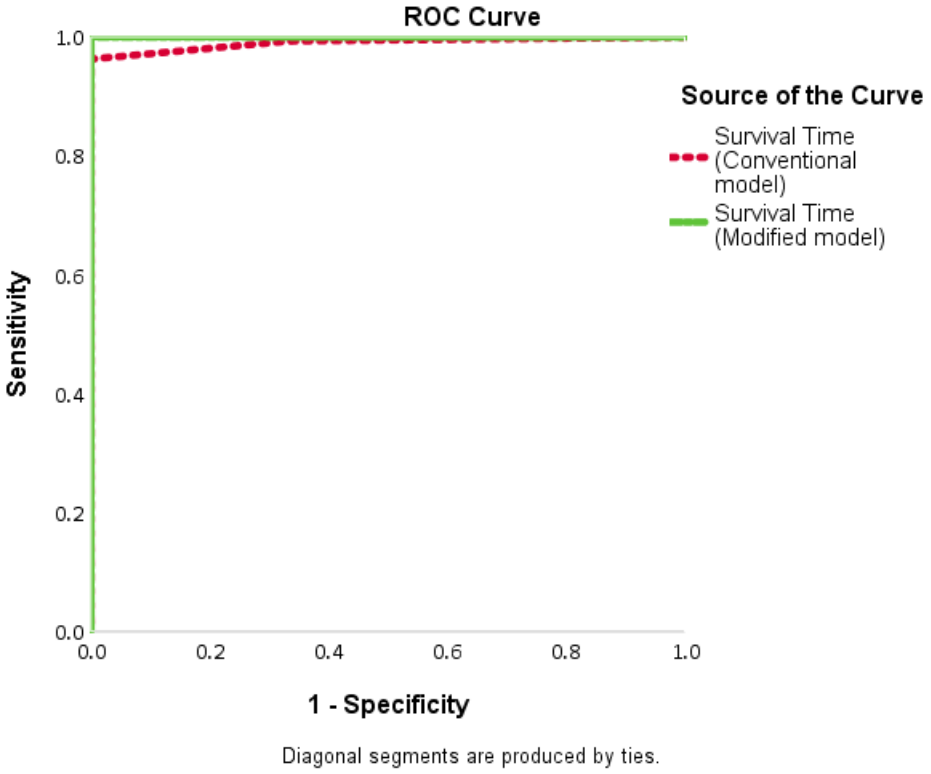

Figure 3. Sensitivity analysis of the existing versus modified survival models: Area under the normal curve AUC (95% C.I.) for the conventional survival model 0.990 (0.982-0.999), and of modified model was 1.00 (1.0-1.0). It revealed a high sensitivity of modified survival model over conventional model

5. Discussion

5.1. Novelty of model approach

Numerous researchers worldwide have applied Kaplan-Meier and Markov chain Monte Carlo survival models to predict student dropout trends and potential risk factors over longitudinal periods [15]–[18]. Most analytical models, however, focus solely on the initial occurrence of events, neglecting subsequent events. To address this limitation, numerous statistical models have been proposed to assess following multiple events [19]–[22]. Our study adopts a novel approach by comparing predictive survival models using both an existing survival model and a modified model in which the survival time of permanent as well as temporarily dropped-out students was also taken into account while calculating the survival time and time-to-event analysis. This comprehensive framework allows for a more comprehensive time-to-event analysis.

5.2. Comparison of outcomes with previous models

Comparing the two models revealed a substantial higher mean survival time with the modified survival function than with the existing function. While there was only a nominal difference in the survival curves and probabilities of survival, the modified function better accounts for the complexities of student dropout and retention, supporting the hypothesis that it is more accurate. Worldwide, the majority of university programs, particularly bachelor programs, require a student to attend a set number of semesters without fail and make completion essential. Typically, policies define a maximum term of temporary absence, such as one to two years [23, 24]. There may also be restrictions on the time of the leave, such as requiring students to complete at least one semester before applying [25]. As a result,

it was challenging to collaborate with university programs that had a flexible deferment or temporary dropout policy and use it to test the algorithm. Furthermore, the institutional regulations of confidentiality and publication of students' personal information and performance statistics are deemed particularly sensitive, and they prohibit research studies [26, 27], unless the university board requests policy adjustments or financial reasons. Therefore, only one program was utilized to analyze the algorithm.

In epidemiological research, recurrent event states are often conceptualized as a special case of multistate models. Aronim (2015) emphasized that as individuals transition across states over time, intermediate states may emerge, which can be analyzed using techniques such as counting occurrences, examining inter-event relationships, and understanding biological processes. Multistate models, such as total or gap times, are typically employed in cases with limited recurrent events and varying probabilities of recurrence. This framework aligns with the approach of the modified survival function in the present study, where censored observations and resumed, states are explicitly addressed to improve predictive accuracy [28].

Liu (2012) proposed a modification to the Kaplan-Meier survival estimator to reduce bias from censored data. This adjustment accounted for the average size of the risk set in intervals with uniformly distributed censored observations. By refining the survival function estimator, Liu demonstrated the potential for improving the Kaplan-Meier model's accuracy in predictive modeling. Similarly, the present study highlights the importance of incorporating censored and resumed states in the survival model for better predictive performance [29].

Dehon (2024) investigated student outcomes categorized as dropout, graduation, and right-censored data, focusing primarily on the most frequently studied outcomes: graduation and dropout. Unlike the present study, students classified as "stop-out" cases (temporary breaks from study) were excluded to avoid bias. This exclusion overlooks a significant group whose inclusion, as demonstrated by the modified survival model, can provide a more accurate depiction of survival curves and outcomes [30].

5.3. Factors Associated with Student Dropout

In the present study, Cox regression analysis revealed that the reason for dropout was a significant factor influencing student survival in higher education programs. Previous studies, such as Ali (2023), found that female students, students with lower admission exam scores, and those from English-medium schools had higher dropout probabilities, with p-values of 0.200, < 0.001, 0.04, and 0.02, respectively [5]. These findings align with the current study, emphasizing the need for targeted interventions to address dropout risks. Further evidence from Vallejos (2017) highlighted the importance of program selection as a risk-reducing factor for delayed graduations and voluntary withdrawals. Conversely, the risk of dropout increased with longer gaps between high school graduation and university entrance. This aligns with the present study's findings that emphasize the importance of understanding dropout dynamics to enhance retention strategies [31].

5.4. Model Performance and Efficiency

The modified survival model demonstrated superior performance, with an R^2 value greater than that of the existing Kaplan-Meier model. This finding underscores the modified model's accuracy and reliability in predicting student survival and retention patterns. The model performance in the present study is the highest as compared to previously reported study by Ameri S, 2016 [15], who reported the highest AUC of time-dependent Cox-model was 0.84. The study on handling recurrent events [28] compared five different models but did not report model performance. Thus, the present study on recurrent events in academia has this unique feature.

5.5. Practical Implication

The modified survival model can also be used to forecast survival time and early dropout rates with enhanced statistical models to utilize early censored data after recurrence and predict key covariates contributing to the risk of dropouts and be helpful for resumption. Although the evaluated program in the study was a health-track free-of-cost program with financial aid and loan programs, there was very little chance of financial difficulty for students who dropped out. However, the modified survival model can also be a first choice for private institutions to reconstruct the program structure, resources, and financial aid for student retention and program sustainability.

Developing national guidelines for analyzing student survival and dropout risks in higher education programs is essential, particularly in professional tracks where significant public resources and governmental commitments are invested. The approach can be adapted for other full-time postgraduates and virtual educational systems containing potential risks of dropouts, providing a robust framework for longitudinal student outcomes research.

The dropout event and recurrence are common practices in many fields, like participant dropout in clinical trials, nutritional programs, training, cultural programs, etc. There is a massive shortage of literature regarding the recurrent event survival models. The modified survival function would be applied to patient survival data, considering factors such as quality of treatment, risks of long-term care, appointment adherence, and patient retention. This approach would also benefit business models by informing policy development and decision-support analyses, especially in scenarios involving high dropout rates and inconsistent customer behavior.

5.6. Limitations of Study

The study focused on data from a single higher education program, specifically a health-track professional program with restricted student intake per cohort. Very limited work was previously done with regard to recurrent event survival time in the field of academia, particularly in higher education. Therefore, it was challenging to compare study outcomes with similar models.

Another limitation, combined with the fact that it was a free, government-supported institution with financial aid and a limited number of seats, may have contributed to the comparatively low dropout rate and stronger student commitment compared to other programs. Additionally, confidentiality constraints prevented the inclusion of potentially influential covariates, such as assessment grades, program preference in application, and cultural diversity, in the Cox regression analysis.

5.7. Recommendations

Hence, it is advised that future research be planned for diversified programs in various geographical locations particularly in low-income countries and low-ranked university programs that are striving to thrive and compete with world-class university programs on a national and international scale. Institutional collaboration should be established to ensure the accessibility of accurate and comprehensive student data. A synthetic approach can also be employed to generate confidential information like student track records, high school performance grades, overall psychic state, financial support, etc.

Future studies should test the modified survival function using large-scale datasets and long-term follow-up period to measure the terminal event of delayed students across multidisciplinary professional programs, incorporating multi-factor dropout risks.

6. Conclusion

The results of this study support the hypothesis that the survival time calculated through the modified survival function is higher than that derived from the existing survival function. The modified model outperformed the existing survival model in terms of efficiency, sensitivity, and specificity. It also promotes the use of a modified survival model for full-time higher education programs that are spread out over a longer period of time and involve the risk of a dropout event at some point, followed by a return to study.

Furthermore, dropout cases, which are censored in the existing survival model, play a significant role in estimating student survival time and the mean graduation time of graduates. Hence, the modified function can be preferred when the first event to time doesn't represent the final outcome.

REFERENCES

1. Tolles, J. and Lewis, R.J., *Time-to-event analysis*, Jama, 315(10), pp. 1046–1047, 2016.
2. Tee, K.F., Pesinis, K. and Coolen-Maturi, T., *Competing risks survival analysis of ruptured gas pipelines: A nonparametric predictive approach*, International Journal of Pressure Vessels and Piping, 175, p. 103919, 2019.

3. Brevik, K., Schoville, S. D., Mota-Sanchez, D., & Chen, Y. H., *Pesticide durability and the evolution of resistance: A novel application of survival analysis*, Pest management science, 74(8), 1953–1963, 2018.
4. Ameri, S., *Survival analysis approach for early prediction of student dropout*, 2015.
5. Ali, D.A. and Hussein, A.M., *Analysis of cox proportional hazard model for dropout students in university: case study from SIMAD university*, Journal of Applied Research in Higher Education, 16(3), pp. 820–830, 2024.
6. Gutierrez-Pachas, D.A., Garcia-Zanabria, G., Cuadros-Vargas, E., Camara-Chavez, G. and Gomez-Nieto, E., *Supporting decision-making process on higher education dropout by analyzing academic, socioeconomic, and equity factors through machine learning and survival analysis methods in the Latin American context*, Education Sciences, 13(2), p. 154, 2023.
7. Choi, H.J. and Kim, B.U., *Factors affecting adult student dropout rates in the Korean cyber-university degree programs*, The Journal of Continuing Higher Education, 66(1), pp. 1–12, 2018.
8. Simon, F., Małgorzata, K. and Beatriz, P.O.N.T., *Education and training policy no more failures ten steps to equity in education: Ten steps to equity in education*, oecd Publishing, 2007.
9. Paura, L. and Arhipova, I., *Cause analysis of students' dropout rate in higher education study program*, Procedia-social and behavioral sciences, 109, pp. 1282–1286, 2014.
10. Costa, F.J.D., Bispo, M.D.S. and Pereira, R.D.C.D.F., *Dropout and retention of undergraduate students in management: a study at a Brazilian Federal University*, RAUSP Management Journal, 53(1), pp. 74–85, 2018.
11. Austin, P.C. and Platt, R.W., *Survivor treatment bias, treatment selection bias, and propensity scores in observational research*, Journal of clinical epidemiology, 63(2), pp. 136–138, 2010.
12. Kalbfleisch, J.D. and Prentice, R.L., *The statistical analysis of failure time data*, John Wiley & Sons, 2002.
13. Cox, D.R., *Regression models and life-tables*, Journal of the Royal Statistical Society: Series B (Methodological), 34(2), pp. 187–202, 1972.
14. Shilbayeh, S. and Abonamah, A., *Predicting student enrollments and attrition patterns in higher educational institutions using machine learning*, Int. Arab J. Inf. Technol., 18(4), pp. 562–567, 2021.
15. Clemens, E.V., Lalonde, T., Klopfenstein, K. and Sheesley, A., *Early warning indicators of dropping out of school for teens who experienced foster care*, Child Welfare, 97(5), pp. 65–88, 2020.
16. Seidel, E. and Kutieleh, S., *Using predictive analytics to target and improve first year student attrition*, Australian Journal of Education, 61(2), pp. 200–218, 2017.
17. Martínez-Carrascal, J.A., Hlosta, M. and Sancho-Vinuesa, T., *Using survival analysis to identify populations of learners at risk of withdrawal: Conceptualization and impact of demographics*, International Review of Research in Open and Distributed Learning, 24(1), pp. 1–21, 2023.
18. Cobre, J., Tortorelli, F.A.C. and de Oliveira, S.C., *Modelling two types of heterogeneity in the analysis of student success*, Journal of Applied Statistics, 46(14), pp. 2527–2539, 2019.
19. Chatterjee, M. and Sen Roy, S., *A copula-based approach for estimating the survival functions of two alternating recurrent events*, Journal of Statistical Computation and Simulation, 88(16), pp. 3098–3115, 2018.
20. Su, C.L., Platt, R.W. and Plante, J.F., *Causal inference for recurrent event data using pseudo-observations*, Biostatistics, 23(1), pp. 189–206, 2022.
21. Kelly, P.J. and Lim, L.L.Y., *Survival analysis for recurrent event data: an application to childhood infectious diseases*, Statistics in medicine, 19(1), pp. 13–33, 2000.
22. Wang, M.C. and Chang, S.H., *Nonparametric estimation of a recurrent survival function*, Journal of the American Statistical Association, 94(445), pp. 146–153, 1999.
23. King Fahad University of Petroleum and Minerals (KFUPM) unified regulations manual: https://cgis.kfupm.edu.sa/academics/unified-regulations
24. Dublin City University guidelines https://www.dcu.ie/sites/default/files/registry_access/2023-09/GuidelinesonDeferral_2024.pdf
25. University of Technology Mara (UiTM) regulations, https://www.dcu.ie/sites/default/files/registry_access/2023-09/GuidelinesonDeferral_2024.pdf
26. Family Educational Rights and Privacy Act (FERPA) at 34 CFR Part 99, http://www.ed.gov/policy/gen/guid/fpco/ferpa/index.html
27. "Research in Schools and with Education Records" and the University of Kentucky, https://registrar.uky.edu/ferpa
28. Amorim, L.D. and Cai, J., *Modelling recurrent events: a tutorial for analysis in epidemiology*, International journal of epidemiology, 44(1), pp. 324–333, 2015.
29. Liu, X., *Survival analysis: models and applications*, John Wiley & Sons, 2012.
30. Dehon, C. and Lebouteiller, L., *A comparison between two systems of university education: years of study versus credit accumulation*, Education Economics, 33(2), pp. 258–276, 2025.
31. Vallejos, C.A. and Steel, M.F., *Bayesian survival modelling of university outcomes*, Journal of the Royal Statistical Society Series A: Statistics in Society, 180(2), pp. 613–631, 2017.

Appendix 1

R software programming layout for existing survival function:

I. R software programming layout for existing KM survival function: Idata $<-$ FinalDataConv
Idata $<-$ as.data.frame(Idata)
attach(Idata)

```
Idata$Time

# calculate permanently dropout cases
P1=sum(data$Sem.1 ==' 2', na.rm = TRUE) + sum(data$Sem.2=='2', na.rm=TRUE)
P2=sum(data$Sem.3 ==' 22', na.rm = TRUE) + sum(data$Sem.4=='22', na.rm=TRUE)
P3=sum(data$Sem.5 ==' 222', na.rm = TRUE) + sum(data$Sem.6 == '222', na.rm=TRUE)
P4=sum(data$Sem.7 ==' 2222', na.rm = TRUE) + sum(data$Sem.8 == '2222', na.rm=TRUE)
P5=sum(data$Sem.9 ==' 22222', na.rm = TRUE) + sum(data$Sem.10=='22222', na.rm=TRUE)
P6=sum(data$Sem.11 ==' 222222222', na.rm = TRUE) + sum(data$Sem.12 == '22222', na.rm=TRUE)

# calculate Temporary dropout cases
T1=sum(data$Sem.1 ==' 1', na.rm = TRUE) + sum(data$Sem.2=='1',na.rm=TRUE)
T2=sum(data$Sem.3 ==' 11', na.rm = TRUE) + sum(data$Sem.4=='11',na.rm=TRUE)
T3=sum(data$Sem.5 ==' 111', na.rm = TRUE) + sum(data$Sem.6 == '111',na.rm=TRUE)
T4=sum(data$Sem.7 ==' 1111', na.rm = TRUE) + sum(data$Sem.8 == '1111',na.rm=TRUE)
T5=sum(data$Sem.9 ==' 11111', na.rm = TRUE) + sum(data$Sem.10 == '11111',na.rm=TRUE)
Idatasum=Idata[-c(13:15)]
sum=rowSums(Idatasum,na.rm=TRUE)
T6=sum(sum=='0',na.rm=TRUE)
D=c(P1,P2,P3,P4,P5,P6)
D_cap=c(T1,T2,T3,T4,T5,T6)
n_total <- nrow(Idata)
ni[j] <- n_total
surv_prob2 <- numeric(length = 6)
surv_prob2[1] <-(-(1-D[1]/ni[1]))
for (j in 2:6) {
ni[j] = ni[j-1]-D[j-1]-D_cap[j-1]
surv_prob2[j] <-surv_prob2[j - 1] * (1 - (D[j] / ni[j]))
}
surv_prob2
STime=c(1:6)
result3 <- data.frame(STime=STime, D=D, D_cap=D_cap, surv_prob2 = surv_prob2)
print(result3)
plot(STime, surv_prob2, ylim=c(0,1), abline(0,5,0))

# Initialize area
area <- 0
# for mean
# Calculate area under the curve
for (i in 2:length(STime)) {
# Calculate the width of the rectangle
width <- STime[i] - STime[i-1]

# Calculate the height of the rectangle
height <- (surv_prob2[i] + surv_prob2[i-1]) / 2

# Add the area of the rectangle to the total area
area <- area + width * height
}
```

# Print the result
cat("Total area under the survival curve (Mean Survival time):", area)

for $t^2.S(t)$
area2 $<-$ 0
for (i in 2:length(STime)) {
# Calculate the width of the rectangle
width2 $<-$ STime[i] - STime[i-1]

Calculate the height of the rectangle
height2 $<-$ ($STime[i]^2$*surv_prob2[i] + $STime[i-1]^2$*surv_prob2[i-1]) / 2

Add the area of the rectangle to the total area
area2 $<-$ area2 + width2 * height2
}
# Print the result
cat("Total area under the survival curve:", area2)
# For variance and SD
variance $<-$ area2 - $area^2$
variance
SD $<-$ sqrt(variance)
SD

II.R software programming layout for parametric (Weibull) survival function: surv_obj $<-$ Surv(time = ldata$Time, event = ldata$Status)
weibull_model $<-$ survreg(surv_obj 1, dist = "weibull")
shape = 4
scale = 11.05
print(paste("Shape:", shape))
print(paste("Scale:", scale))
time_points $<-$ seq(0, max(ldata$Time), by = 1)
survival_probs $<-$ pweibull(time_points, shape = shape, scale = scale, lower.tail = FALSE)
print(data.frame(Time = time_points, Survival Probability = surv_probs))

III. R software programming layout for modified survival function: Idata $<-$ FinalDataRes
Idata $<-$ as.data.frame(Idata)
attach(Idata)
Idata$Time

# calculate permanently dropout cases
P1=sum(data$Sem.1 ==' 2', na.rm = TRUE) + sum(data$Sem.2=='2',na.rm=TRUE)
P2=sum(data$Sem.3 ==' 22', na.rm = TRUE) + sum(data$Sem.4=='22',na.rm=TRUE)
P3=sum(data$Sem.5 ==' 222', na.rm = TRUE) + sum(data$Sem.6 == '222',na.rm=TRUE)
P4=sum(data$Sem.7 ==' 2222', na.rm = TRUE) + sum(data$Sem.8 == '2222',na.rm=TRUE)
P5=sum(data$Sem.9 ==' 22222', na.rm = TRUE) + sum(data$Sem.10=='22222',na.rm=TRUE)
P6=sum(data$Sem.11 ==' 222222222', na.rm = TRUE) + sum(data$Sem.12 == '222222222',na.rm=TRUE)
P7=sum(Idata$Sem.13 ==' 2222222', na.rm = TRUE)

```
# calculate Temporary dropout cases
T1=sum(data$Sem.1 ==' 1', na.rm = TRUE) + sum(data$Sem.2=='1',na.rm=TRUE)
T2=sum(data$Sem.3 ==' 11', na.rm = TRUE) + sum(data$Sem.4=='11',na.rm=TRUE)
T3=sum(data$Sem.5 ==' 111', na.rm = TRUE) + sum(data$Sem.6 == '111',na.rm=TRUE)
T4=sum(data$Sem.7 ==' 1111', na.rm = TRUE) + sum(data$Sem.8 == '1111',na.rm=TRUE)
T5=sum(data$Sem.9 ==' 11111', na.rm = TRUE) + sum(data$Sem.10 == '11111',na.rm=TRUE)
T6=sum(Idata$Sem.11 ==' 111111', na.rm = TRUE) + sum(Idata$Sem.12== '111111',na.rm=TRUE)

Idatasum <- Idata[-c(14:16)]
sum <- rowSums(Idatasum, na.rm = TRUE)
T7 <- sum(sum == '0', na.rm = TRUE)

# calculate resume cases
R1=sum(Idata$Sem.1 ==' 9', na.rm = TRUE) + sum(Idata$Sem.2=='9',na.rm=TRUE)
R2=sum(Idata$Sem.3 ==' 99', na.rm = TRUE) + sum(Idata$Sem.4=='99',na.rm=TRUE)
R3=sum(Idata$Sem.5 ==' 999', na.rm = TRUE) + sum(Idata$Sem.6=='999',na.rm=TRUE)
R4=sum(Idata$Sem.7 ==' 9999', na.rm = TRUE) + sum(Idata$Sem.8=='9999',na.rm=TRUE)
R5=sum(Idata$Sem.9 ==' 99999', na.rm = TRUE) + sum(Idata$Sem.10=='99999',na.rm=TRUE)
R6=sum(Idata$Sem.11 ==' 999999', na.rm = TRUE) + sum(Idata$Sem.12=='999999',na.rm=TRUE)
R7=sum(Idata$Sem.13 ==' 9999999', na.rm = TRUE)

D=c(P1,P2,P3,P4,P5,P6,P7)
D_cap=c(T1,T2,T3,T4,T5,T6,T7)
R=c(R1,R2,R3,R4,R5,R6,R7)

n_total <- nrow(Idata)
ni[1] <- n_total
surv_prob3 <- numeric(length = 7)
surv_prob3[1] <- (1 - (D[1] / ni[1]))

for (j in 2:7) {
ni[j] <- ni[j-1] - D[j-1] - D_cap[j-1] + R[j-1]
surv_prob3[j] <- surv_prob3[j-1] * (1 - (D[j] / ni[j]))
}

surv_prob3
STime <- c(1:7)
resultR <- data.frame(STime = STime, D = D, D_cap = D_cap, R = R, surv_prob3 = surv_prob3)
print(resultR)

plot(STime, surv_prob3, ylim = c(0,1), abline(a = 0.5, b = 0))
surv_prob3
STime
# Initialize area
area <- 0
# for mean
# Calculate area under the curve
for (i in 2:length(STime)) {
width <- STime[i] - STime[i-1]
```

Calculate the height of the rectangle
height $<-$ (surv_prob3[i] + surv_prob3[i-1]) / 2
Add the area of the rectangle to the total area
area $<-$ area + width * height
}
# Print the result
cat("Total area under the survival curve (Mean Survival Time):", area)
surv_prob3

for t^2.S(t)
area2 $<-$ 0
for (i in 2:length(STime)) {
width2 $<-$ STime[i] - STime[i-1]

Calculate the height of the rectangle
height2 $<-$ ($STime[i]^2$*surv_prob3[i] + $STime[i-1]^2$*surv_prob3[i-1]) 2

Add the area of the rectangle to the total area
area2 $<-$ area2 + width2 * height2
}
# Print the result
cat("Total area under the survival curve:", area2)
# For variance and SD
variance $<-$ area2 - $area^2$
variance
SD $<-$ sqrt(variance)
SD