International Journal of Advanced and Applied Sciences, 13(7) 2026, Pages: 222-230

Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: https://www.science-gate.com/IJAAS.html

Modeling student retention in higher education: A time-to-event survival analysis of key determinants

Joseph B. Campit *

Science and Mathematics Department, College of Teacher Education, Pangasinan State University, Pangasinan, Philippines

ARTICLE INFO

Article history:
Received 9 April 2026
Received in revised form
22 July 2026
Accepted 30 July 2026

Keywords:
Student retention
Survival analysis
Higher education
Student dropout
Persistence

ABSTRACT

Student retention is a critical indicator of both student success and institutional effectiveness in higher education. This study examines student retention in Philippine higher education institutions (HEIs) using a time-to-event modeling approach through survival analysis, providing a longitudinal perspective on student persistence. Institutional records from 961 undergraduate students at Pangasinan State University, Bayambang Campus, covering School Years 2019–2020 to 2022–2023, were analyzed. The dataset included semester-level enrollment and dropout data, along with demographic and academic characteristics such as sex, course, and class schedule. Retention patterns were assessed using life table analysis, Kaplan–Meier (KM) estimation, Log-Rank tests, and discrete-time survival modeling via generalized estimating equations (GEE). The results indicate that the first semester has the highest observed dropout rate and that survival probabilities decline in subsequent semesters. Observed differences were found across sex, academic program, and class schedule: female students, students in the Bachelor of Science in Business Administration (BSBA), and day-class students showed longer persistence. Discrete-time modeling confirmed these patterns, indicating higher odds of dropout among male students, students in certain programs, and evening-class students. By applying established time-to-event analysis methods to semester-level institutional data, this study provides context-specific evidence on observed retention patterns and the timing of student dropout. These findings offer practical insights for administrators and policymakers to design data-driven retention strategies and targeted support programs, particularly in developing-country higher education contexts.

1. Introduction

Student retention is a critical measure of both student success and institutional effectiveness in higher education. High dropout rates can negatively impact institutional efficiency, student outcomes, and resource allocation, particularly in developing-country contexts where interventions may be constrained (Patacsil, 2020). Understanding when and which students are most likely to leave a program allows institutions to implement targeted strategies to improve persistence and completion rates. Research has shown that student retention is influenced by demographic and academic factors, including sex, academic program, and class schedule (Astin, 1984; Bean, 1980; York et al., 2015). Differences across these characteristics may highlight student groups that are at greater risk of early attrition and help institutions identify periods of higher vulnerability, such as the first semester, which has consistently been identified as a critical stage in student persistence.

Time-to-event modeling approaches, including Kaplan-Meier estimation, life-table analysis, and discrete-time survival analysis using generalized estimating equations (GEE), provide a rigorous framework for studying retention. Recent studies have further demonstrated the value of longitudinal and predictive approaches for identifying students at risk of dropout, including survival-based analyses and machine-learning models that incorporate academic and enrollment characteristics (Vaarma and Li, 2024; González-Morales et al., 2025). These methods account for both the timing of dropout events and censored cases, such as students who remain enrolled or graduate, allowing for more

* Corresponding Author.
Email Address: josephcampit@psu.edu.ph
https://doi.org/10.21833/ijaas.2026.07.021
Corresponding author's ORCID profile:
https://orcid.org/0009-0007-3538-3637

accurate assessment of semester-level retention patterns. Discrete-time modeling is particularly suitable for institutional datasets with semester-based observations, thereby ensuring alignment between the methodology and the data structure.

Although multiple factors may influence student retention, this study focuses on sex, course, and class schedule, which are routinely recorded in institutional records. Applying established time-to-event methods to these predictors enables a context-specific analysis of observed retention trends and the identification of periods where dropout is most likely to occur.

The present study contributes to the literature by providing empirical evidence on observed retention patterns using semester-level institutional data in a Philippine higher education setting. The findings aim to inform administrators and policymakers in designing data-driven strategies and targeted support programs, particularly for students at higher risk of early attrition.

2. Review of related literature

Student retention is a key indicator of both student success and institutional effectiveness. Traditional perspectives highlight that persistence is influenced by academic integration, student engagement, and external conditions (Bean, 1980; Astin, 1984). While these frameworks remain foundational, earlier studies often treated dropout as a fixed outcome rather than a time-dependent event.

Recent research has increasingly applied survival analysis to examine student retention, capturing both the occurrence and timing of dropout and accounting for censored cases, such as students who remain enrolled or graduate. Kaplan-Meier and life table methods have been used to identify critical periods of attrition (Patacsil, 2020; Martínez-Carrascal et al., 2023; Shafiq et al., 2022; Cardona et al., 2023). These studies demonstrate the utility of time-to-event approaches in highlighting when students are most vulnerable to leaving and in identifying at-risk groups.

Demographic and academic factors continue to be examined as predictors of retention and dropout, with recent studies using large institutional datasets to identify characteristics associated with student attrition (González-Morales et al., 2025; Yan et al., 2025). Sex has been analyzed because male and female students may exhibit different persistence patterns (Astin, 1984; Bean, 1980). An academic program or course can influence retention due to differences in curriculum, workload, and program support (York et al., 2015; Pascarella and Terenzini, 2005). The class schedule is also relevant, with evidence that evening students may have lower persistence due to external responsibilities, such as work or family obligations (Simpson, 2012).

This study focuses specifically on sex, course, and class schedule because these variables are available in standardized institutional records. By applying time-to-event analysis to these predictors, the research provides a context-specific examination of dropout timing and observed differences in retention patterns across student groups.

3. Methodology

3.1. Research design

This study employed a quantitative research design using both descriptive and inferential approaches to examine student retention in undergraduate programs. Because student dropout was observed by academic semester rather than exact calendar time, a discrete-time survival analysis was conducted. The dataset was restructured into a person-period format, where each student contributed one record for each semester in which they remained at risk of dropout.

The dependent variable indicated whether a student dropped out during a given semester. Semester indicators were included to account for variation in baseline dropout risk across academic periods. A binary logistic model within a generalized estimating equations (GEE) framework was used to estimate semester-specific odds of dropout, accounting for repeated observations within students. This approach enables the analysis of attrition timing while accounting for censored observations during the observation period (Kleinbaum and Klein, 2012) and supports the identification of at-risk groups over time.

3.2. Study context and population

The study was conducted at Pangasinan State University Bayambang Campus, specifically within the College of Arts and Sciences. The college offers four non-board undergraduate programs: Bachelor of Arts in English (ABE), Bachelor of Public Administration (BPA), Bachelor of Science in Business Administration (BSBA), and Bachelor of Science in Information Technology (BSIT), each with a prescribed duration of 4 years (8 semesters).

The study followed a cohort of students enrolled in the School Year 2019 - 2020 and tracked their academic progress through School Year 2022 - 2023. A total of 961 students were included using complete enumeration. Transferees and shifters were excluded to ensure consistency in longitudinal tracking. Cohort-based approaches are widely used in retention studies to examine persistence over time (Pascarella and Terenzini, 2005).

3.3. Data source and description

The study utilized secondary data obtained from institutional records. The dataset covered eight academic semesters (four years), with no summer terms included. The data included selected demographic and academic characteristics, specifically sex, course, and class schedule, as well as students' enrollment status for each semester.

The selection of these variables was based on their availability in the institutional student portal database, which stores only routinely recorded, standardized information. Although this limits the inclusion of more complex socioeconomic and psychosocial variables, the selected predictors remain analytically relevant, as prior studies have consistently identified demographic and program-related characteristics as significant determinants of student persistence (Bean, 1980; Astin, 1984; York et al., 2015). The use of institutional data also ensures consistency and reliability in longitudinal tracking.

The outcome variable included the event of interest (student dropout) and the time-to-event, measured in semesters. Students who graduated or remained enrolled at the end of the observation period were treated as censored cases, under the assumption of non-informative censoring, meaning that censored cases were not systematically different in unmeasured ways that could bias the dropout estimates.

3.4. Data gathering procedure

Institutional approval was obtained before data collection. Relevant student records were accessed through the Registrar's Office and extracted from the university's student portal database via the Information and Communication Technology Management Office. The dataset was subsequently organized, cleaned, and validated to ensure accuracy and completeness before analysis.

3.5. Data preparation and coding

The dataset was structured to support the discrete-time survival analysis. Each student record included a time variable and an event indicator. The time variable ranged from the first to the eighth semester, while the event variable was coded as 1 for dropout and 0 for censored cases.

Predictor variables were coded to facilitate analysis. Sex was coded as 1 = male and 2 = female; course as 1 = ABE, 2 = BPA, 3 = BSBA, and 4 = BSIT; and class schedule as 1 = day and 2 = evening. This coding scheme allowed for comparing dropout odds across groups and aligning with the person-period dataset for logistic regression under GEE.

3.6. Censoring assumptions

Students who graduated or remained enrolled at the end of the observation period were treated as censored because they did not experience the event of interest (dropout). Graduation was considered a terminal non-dropout outcome, while students who remained enrolled had no recorded dropout event within the eight-semester observation period. Treating these cases as censored is appropriate because the exact time to dropout was either unobserved or no longer applicable. The analysis assumes non-informative censoring, meaning that censored cases were not systematically different in their unobserved dropout timing after accounting for the available predictors. Since censoring primarily resulted from graduation or continued enrollment, as recorded in institutional records, it is unlikely to bias the observed retention patterns substantially.

3.7. Statistical treatment of data

Statistical analyses were conducted in alignment with the study objectives. Descriptive statistics, including frequency counts and percentages, were used to summarize students' demographic and academic profiles.

Life table analysis was employed to examine the pattern of student retention across semesters, providing estimates of the number at risk, number of dropouts, and survival probabilities over time.

Kaplan-Meier estimation was used to generate survival curves representing the probability of students remaining enrolled.

The Log-Rank (Mantel-Cox) test was used to assess differences in survival time across groups defined by sex, course, and class schedule.

For the primary predictive analysis, a discrete-time survival model was fitted using binary logistic regression within a GEE framework, with semester as a time-varying dummy variable. This accounted for repeated observations within students and estimated semester-specific odds of dropout. Results were reported as odds ratios (OR = Exp(B)), with OR > 1 indicating higher odds of dropout and OR < 1 indicating lower odds of dropout. The modeled response category was dropout. Odds ratios were computed by exponentiating the estimated logistic regression coefficients, OR = Exp(B).

4. Results and discussion

4.1. Profile of students

Table 1 presents the demographic and academic profiles of the students included in the study. The profile is described in terms of sex, course, and class schedule, which serve as the key grouping variables in the subsequent survival analyses.

As shown in Table 1, the majority of the students are female (61.2%), while male students account for 38.8% of the sample. This indicates that female students comprise a larger proportion of the population included in the analysis.

In terms of academic program, students are distributed across four undergraduate courses. The largest proportion of students is enrolled in the BSBA (29.2%), followed by BPA (26.6%), BSIT (25.5%), and ABE (18.6%). This distribution suggests a relatively balanced representation across programs, with a slightly higher concentration in business-related courses.

Regarding the class schedule, the majority of students are enrolled in the day class (63.3%), while

36.7% are enrolled in the evening class. This indicates that most students attend regular daytime classes, while a smaller proportion are enrolled in evening schedules, which may be associated with different academic and personal circumstances.

Table 1: Profile of undergraduate students by sex, course, and class schedule (N = 961)

Profile variables	Category	Frequency	Percent
Sex	Male	373	38.8
	Female	588	61.2
	Total	961	100.0
Course	ABE	179	18.6
	BPA	256	26.6
	BSBA	281	29.2
	BSIT	245	25.5
	Total	961	100.0
Class schedule	Day class	608	63.3
	Evening class	353	36.7
	Total	961	100.0

The profile shows that the sample includes a diverse distribution of students across sex, academic programs, and class schedules. These variables serve as key factors in examining differences in student retention and survival patterns in the subsequent analyses, as demographic and academic characteristics are commonly associated with student persistence in higher education (Pascarella and Terenzini, 2005)

4.2. Pattern of student retention

Table 2 and Fig. 1 present the pattern of student retention across eight academic semesters using life table analysis and Kaplan-Meier survival estimation. Table 2 provides a numerical summary of student progression, including the number entering each interval, the number of dropouts, and the survival probabilities.

In contrast, Fig. 1 illustrates the survival curve, which shows the probability of students remaining enrolled over time.

Table 2: Life table showing the pattern of student retention

Interval start time (semester)	Number entering interval (currently enrolled) (n)	Number of terminal events (dropouts) (d)	Number graduating during interval (c)	Proportion terminating (h)	Proportion surviving (p)	Cumulative proportion surviving at the end of interval (s)
0	961	0	0	0.00	1.00	1.00
1	961	87	0	0.09	0.91	0.91
2	874	9	0	0.01	0.99	0.90
3	865	26	0	0.03	0.97	0.87
4	839	15	0	0.02	0.98	0.86
5	824	14	0	0.02	0.98	0.84
6	810	16	0	0.02	0.98	0.83
7	794	35	0	0.04	0.96	0.79
8	759	39	720	0.10	0.90	0.71

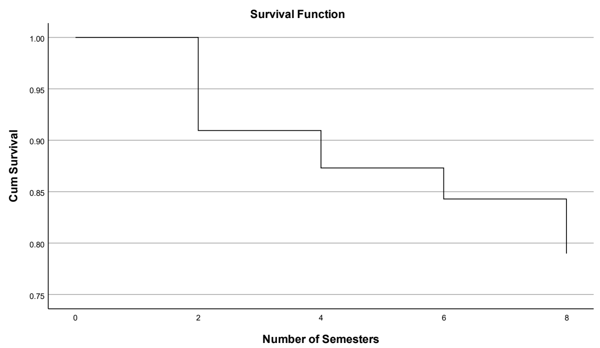

Fig. 1: Kaplan-Meier survival curves of student retention

As shown in Table 2 and reflected in the steep initial drop in Fig. 1, the highest dropout occurred during the first semester, with 87 students leaving the program, corresponding to a termination proportion of 0.09. In Semesters 2 to 6, the number of dropouts was relatively low, ranging from 9 to 16 students, with proportions terminating between 0.01 and 0.02. The cumulative survival probability decreased gradually from 0.91 after the first semester to 0.83 by the end of Semester 6, indicating greater stability in retention among students who persisted beyond the early stage.

A moderate increase in dropout was observed in Semester 7 ($d = 35$; $h = 0.04$). In the final semester, 39 students dropped out while 720 students graduated, resulting in a cumulative survival probability of 0.71. The higher proportion terminating in the final semester ($h = 0.10$), as shown in Table 2 and Fig. 1, reflects the reduced number of students at risk due to graduation.

Table 2 and Fig. 1 together provide a clear depiction of retention dynamics across the undergraduate program, emphasizing the first semester as a critical period for student attrition and showing relative stability in retention during the middle semesters, with a slight increase in dropout toward the end. These patterns are consistent with prior research on early attrition in higher education (York et al., 2015).

4.3. Comparison in survival time of students when grouped according to their selected profile variables

Table 3 and Fig. 2 present the Kaplan-Meier estimates of survival time and the corresponding survival curves comparing student retention between male and female students. Table 3 summarizes the mean survival time and statistical test results, while Fig. 2 provides a graphical representation of survival probabilities across semesters. Together, these analyses offer a comprehensive view of retention patterns by sex.

The Kaplan-Meier estimates, as shown in Table 3, indicate that female students had a higher mean survival time ($M = 7.156$ semesters, 95% CI: 6.99 7.33) than male students ($M = 6.751$ semesters, 95%

CI: 6.51 - 6.99). The Log-Rank test indicated a statistically significant difference in survival distributions between the groups (χ^2 = 26.36, df = 1, p < .001). The overall mean survival time for all students was 6.999 semesters (95% CI: 6.86 - 7.14), which falls between the estimates for male and female students.

Table 3: Kaplan–Meier estimates of mean survival time by sex and log-rank test of differences

Sex	Mean survival time (M)	95% confidence interval (CI)	Overall comparison log-rank (Mantel-Cox)
Male	6.751	6.51 – 6.99	χ^2 = 26.36,
Female	7.156	6.99 – 7.33	df = 1,
Overall	6.999	6.86 – 7.14	p < 0.001

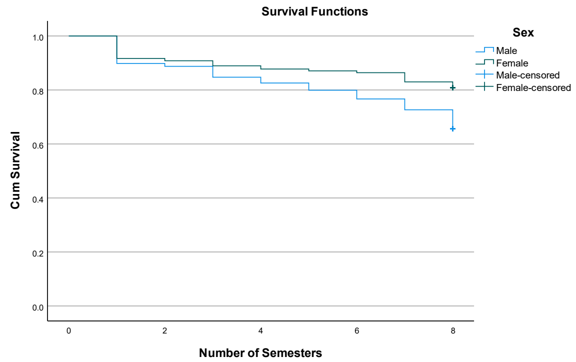

Fig. 2: Kaplan-Meier survival curves of student retention by sex

Fig. 2 shows that the survival curve for female students consistently remains above that of male students across all semesters. Male students show a steeper initial decline in survival probability, particularly during the early semesters.

These observations align with previous research reporting gender-related differences in student persistence (Astin, 1984; Bean, 1980), while other studies have found minimal differences, indicating variability in retention patterns across institutions (Simpson, 2012).

Table 4 and Fig. 3 present Kaplan-Meier estimates of student retention time and the corresponding survival curves across different undergraduate programs. Table 4 presents the mean survival time and statistical comparisons, while Fig. 3 illustrates survival probabilities over time, allowing a visual comparison of retention patterns across courses.

Table 4: Kaplan-Meier estimates of mean survival time by course and log-rank test of differences

Course	Mean survival time (M)	95% confidence interval (CI)	Overall comparison log-rank (Mantel-Cox)
ABE	7.01	6.67 – 7.35	
BPA	6.86	6.59 – 7.14	χ^2 = 45.49,
BSBA	7.36	7.13 – 7.53	df = 3,
BSIT	6.72	6.42 – 7.03	p < 0.001
Overall	6.99	6.86 – 7.14	

As shown in Table 4, students in the BSBA program had the highest mean survival time (M = 7.36 semesters, 95% CI: 7.13 7.53), while BSIT students had the lowest mean survival time (M = 6.72 semesters, 95% CI: 6.42 - 7.03). Students in the ABE and BPA programs had intermediate survival times (M = 7.01 and 6.86 semesters, respectively).

The Log-Rank test indicated a statistically significant difference in survival distributions across programs (χ^2 = 45.49, df = 3, p < .001).

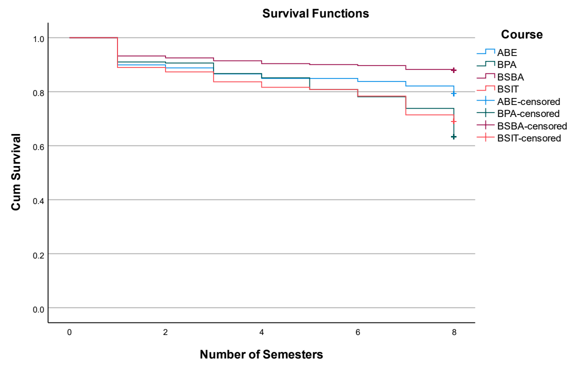

Fig. 3: Kaplan-Meier survival curves of student retention by course

Fig. 3 shows that the survival curve for BSBA students consistently remains above the other programs across all semesters, while BSIT students have the lowest survival probabilities. ABE and BPA curves fall between these extremes, illustrating moderate retention levels.

To identify which specific programs differ in retention, pairwise comparisons using the Log-Rank test were conducted with a Bonferroni-adjusted significance level (α= 0.0083). The results are presented in Table 5.

Table 5: Pairwise comparisons of student retention by course using the log-rank test with Bonferroni adjustment

Pairwise comparisons (course1 vs course2)	Log-rank (Mantel-Cox) χ^2	p-value
ABE vs. BPA	10.850	0.001
ABE vs. BSBA	6.033	0.014
ABE vs. BSIT	5.234	0.022
BPA vs. BSBA	41.273	0.000
BPA vs. BSIT	1.054	0.305
BSBA vs. BSIT	27.222	0.000

Bonferroni-adjusted significance level (α = 0.0083)

The results indicate statistically significant differences in survival distributions between ABE and BPA (χ^2 = 10.850, p = .001), BPA and BSBA (χ^2 = 41.273, p < .001), and BSBA and BSIT (χ^2 = 27.222, p < .001). No significant differences were observed between ABE and BSBA (p = .014), ABE and BSIT (p = .022), or BPA and BSIT (p = .305) using the Bonferroni-adjusted significance level. Notably, BSBA exhibits a distinct retention profile, differing from both BPA and BSIT. These observed differences are consistent with prior research suggesting that program-specific factors, including variations in curriculum or institutional context, may be associated with differences in student persistence (York et al., 2015; Pascarella and Terenzini, 2005; Bean, 1980).

Table 6 and Fig. 4 present the Kaplan-Meier estimates of survival time and the corresponding survival curves comparing student retention between day and evening class schedules. Table 6 presents the mean survival time and statistical test results, while Fig. 4 illustrates survival probabilities across semesters, enabling a visual comparison of retention patterns between the two groups.

Table 6: Kaplan-Meier estimates of mean survival time by time schedule and log-rank test of differences

Class schedule	Mean survival time (M)	95% confidence interval (CI)	Overall comparison log-rank (Mantel-Cox)
Day class	7.09	6.91 – 7.26	χ^2 = 7.73, df = 1, p = 0.005
Evening class	6.85	6.61 – 7.09	
Overall	6.99	6.86 – 7.14	

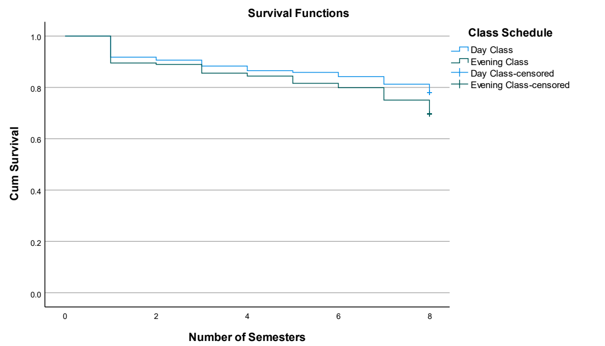

Fig. 4: Kaplan-Meier survival curves of student retention by class schedule

Table 6 and Fig. 4 present the Kaplan-Meier estimates of mean survival time and survival curves for students grouped by class schedule. Table 6 provides the mean survival time with 95% confidence intervals and the results of the Log-Rank (Mantel-Cox) test, while Fig. 4 illustrates the survival probabilities across semesters.

As shown in Table 6, students enrolled in the day class had a slightly higher mean survival time (M = 7.09 semesters, 95% CI: 6.91 - 7.26) than those in the evening class (M = 6.85 semesters, 95% CI: 6.61 - 7.09). The Log-Rank test indicated a statistically significant difference in survival distributions between the two groups (χ^2 = 7.73, df = 1, p = .005).

Fig. 4 shows that the survival curve for day class students consistently remains above that of evening class students across all semesters. Both groups start at the same survival probability, but the curve for evening class students declines more steeply in the early semesters. In contrast, the curve for day class students decreases more gradually. The evening class subgroup is small, and the observed retention trends may be more variable. These observed differences align with prior research indicating that retention may vary by class schedule (Simpson, 2012; Bean, 1980; Astin, 1984).

4.4. Discrete-time survival model

The discrete-time model examined the odds of dropout in a given semester, conditional on the student having remained enrolled up to that semester. Semester indicators were included to account for variation in baseline dropout risk across academic periods. Table 7 presents the tests of model effects for the discrete-time survival model predicting student dropout.

Table 7: Tests of model effects for discrete-time survival model predicting student dropout

Predictor	Wald Chi-square	df	p-value
Semester	107.428	7	0.000
Sex	17.045	1	0.000
Course	29.713	3	0.000
Class schedule	5.596	1	0.018

As shown in Table 7, semester was a significant predictor of dropout, Wald $\chi^2(7)$ = 107.428, p = .001. This indicates that the odds of dropout varied significantly across academic semesters, supporting the need to account for time when modeling student retention. Sex was also statistically significant, Wald $\chi^2(1)$ = 17.045, p = .001, indicating that dropout odds differed between male and female students. The course was likewise significant, Wald $\chi^2(3)$ = 29.713, p = .001, indicating that dropout odds varied across academic programs. Class schedule was also a significant predictor, Wald $\chi^2(1)$ = 5.596, p = .018, suggesting that dropout odds differed between day and evening class students.

The results indicate that semester, sex, course, and class schedule significantly contributed to the discrete-time survival model predicting student dropout. Table 8 presents the parameter estimates and odds ratios for the discrete-time survival model predicting student dropout. The model estimated the semester-specific odds of dropout while controlling for sex, course, and class schedule.

Table 8: Parameter estimates and odds ratios for discrete-time survival model predicting student dropout

Predictors	B	SE	Wald Chi-square	p-value	Odds ratio	95% CI for OR
Semester 1	0.565	0.2023	7.789	0.005	1.76	1.18–2.61
Semester 2	-1.711	0.3767	20.639	0.000	0.18	0.09–0.38
Semester 3	-0.614	0.2601	5.573	0.018	0.54	0.33–0.90
Semester 4	-1.139	0.3110	13.419	0.000	0.32	0.17–0.59
Semester 5	-1.189	0.3148	14.270	0.000	0.3	0.16–0.56
Semester 6	-1.021	0.3018	11.431	0.001	0.36	0.20–0.65
Semester 7	-0.175	0.2404	0.530	0.467	0.84	0.52–1.34
Male	0.570	0.1382	17.045	0.000	1.77	1.35–2.32
ABE	-0.190	0.2180	0.756	0.385	0.83	0.54–1.27
BPA	0.177	0.1698	1.086	0.297	1.19	0.86–1.67
BSBA	-0.948	0.2225	18.145	0.000	0.39	0.25–0.60
Day Class	-0.357	0.1510	5.596	0.018	0.7	0.52–0.94

Reference categories were Semester 8 for semester, female for sex, BSIT for course, and evening class for class schedule; The outcome variable was dropout; OR > 1 indicates higher odds of dropout, whereas OR < 1 indicates lower odds of dropout; Odds ratios were calculated as OR = Exp(B)

As shown in Table 8, dropout odds differed significantly across several semesters when compared with Semester 8. Students had significantly higher odds of dropout in Semester 1 than in Semester 8, OR = 1.76, p = .005. In contrast, the odds of dropout were significantly lower in Semester 2, OR = 0.18, p = .000; Semester 3, OR = 0.54, p = .018; Semester 4, OR = 0.32, p = .000; Semester 5, OR = 0.30, p = .000; and Semester 6, OR = 0.36, p = .001. Semester 7 did not significantly differ from Semester 8, OR = 0.84, p = .467. These results are consistent with the retention literature, which emphasizes the importance of early intervention (York et al., 2015).

For sex, male students had significantly higher odds of dropout than female students, OR = 1.77, p < .001. This means that, controlling for semester, course, and class schedule, male students were more likely to drop out in a given semester than female students.

For course, BSBA students had significantly lower odds of dropout than BSIT students, OR = 0.39, p = .000. This indicates stronger retention among BSBA students compared with the reference group. However, ABE, OR = 0.83, p = .385, and BPA, OR = 1.19, p = .297, did not significantly differ from BSIT.

For the class schedule, day class students had significantly lower odds of dropout than evening class students, OR = 0.70, p = .018.

The findings indicate that semester, sex, course, and class schedule were important factors in explaining student dropout, with higher dropout odds observed among first-semester students, male students, and evening class students. The findings provide actionable insights for institutional retention strategies, emphasizing targeted support for first-semester students, male students, students in high-risk programs, and those enrolled in evening classes.

These findings are consistent with prior research showing that demographic characteristics, such as sex, and program affiliations are associated with differences in student persistence (Astin, 1984; Bean, 1980; York et al., 2015; Pascarella and Terenzini, 2005). Similarly, previous studies indicate that variations in class scheduling and enrollment type can be linked to differences in retention outcomes (Simpson, 2012).

4.5. Sensitivity analysis

To assess the robustness of the main findings, a sensitivity analysis was conducted by redefining dropout as leaving before the fourth semester. An early dropout variable was created, with students who dropped out in Semesters 1 to 3 coded as 1, and students who remained enrolled or graduated at Semester 4 or later treated as censored cases. Table 9 presents the distribution of early dropout status.

As shown in Table 9, 122 students were classified as early dropouts, representing 12.7% of the total sample, while 839 students, or 87.3%, were classified as censored under this alternative definition. This result is consistent with the main life table analysis, which showed that the largest number of dropouts occurred during the first semester and that early attrition was concentrated in the initial academic periods.

Table 9: Distribution of early dropout status

Early dropout status	Frequency	Percent
Censored	839	87.3
Early dropout	122	12.7
Total	961	100.0

Table 10 presents Kaplan-Meier estimates of the mean survival time for early dropouts. Early dropout was defined as dropout during Semesters 1–3. Students who did not experience early dropout, including those who remained enrolled, graduated, or dropped out in Semester 4 or later, were treated as censored cases.

Table 10: Kaplan–Meier estimate of mean survival time under the early-dropout definition

Statistic	Estimate	SE	95% CI
Mean survival time	7.175	0.070	7.037 – 7.313

Kaplan-Meier estimates were recalculated using the early dropout definition. As shown in Table 10, the mean survival time was 7.175 semesters, with a 95% confidence interval of 7.037 to 7.313.

Fig. 5 presents the Kaplan-Meier Survival Curve for Early Dropout.

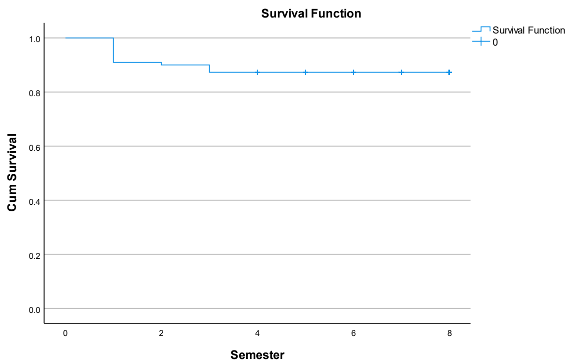

Fig. 5: Kaplan-Meier survival curve for early dropout

The Kaplan-Meier curve in Fig. 5 shows that most students persisted beyond the early semesters, while a smaller proportion dropped out before Semester 4. The sensitivity analysis supports the robustness of the main findings by showing that early dropout remains concentrated in the first three semesters when dropout is redefined as leaving before the fourth semester. This is consistent with the main analysis, which identified the first semester as the period with the highest dropout occurrence and showed lower dropout levels in the succeeding semesters.

5. Conclusions

This study examined student retention using a discrete-time survival modeling approach, providing a longitudinal perspective on the odds of dropout across eight academic semesters. The findings indicate that student retention is significantly influenced by demographic and academic characteristics, particularly sex, course, and class schedule, confirming that persistence patterns vary across student groups.

The analysis revealed that the first semester has the highest dropout odds, highlighting the importance of early monitoring and intervention. Differences in dropout odds across academic programs suggest that program-specific factors influence persistence. Additionally, students enrolled in evening classes had higher odds of dropping out than day students, suggesting an association between class schedule and student retention.

By applying discrete-time survival analysis, the study provides a robust framework for capturing both the timing and occurrence of dropout, offering insights into student persistence dynamics. This research contributes to the literature by providing a context-specific, evidence-based analysis of student retention in Philippine higher education. It emphasizes the importance of targeted, time-sensitive interventions to support student persistence.

6. Implications for policy and practice

The findings of this study provide actionable guidance for higher education institutions seeking to improve student retention.

First, identifying the first semester as the period with the highest odds of dropout underscores the need for early, proactive support programs, including structured orientation, academic advising, and mentoring systems, to enhance student adjustment and engagement.

Second, differences in dropout odds across academic programs suggest that program-level evaluation is essential. Institutions should assess curriculum design, instructional strategies, and learning environments to ensure that program-specific factors support persistence.

Third, the higher dropout odds among male students underscore the importance of gender-sensitive support strategies, including monitoring, engagement initiatives, and targeted academic support to address groups at elevated risk.

Fourth, the observed differences by class schedule indicate that evening students may require tailored support mechanisms, such as flexible learning options, accessible academic advising, and guidance for balancing external responsibilities with study demands.

Finally, this study demonstrates the value of data-driven decision-making. By leveraging discrete-time survival analysis of institutional data, universities can identify critical periods and at-risk student groups, allowing for the development of targeted, evidence-based retention strategies that enhance student persistence and academic success.

List of abbreviations

ABE	Bachelor of Arts in English
BPA	Bachelor of Public Administration
BSBA	Bachelor of Science in Business Administration
BSIT	Bachelor of Science in Information Technology
CI	Confidence interval
df	Degrees of freedom
Exp(B)	Exponentiated coefficient (odds ratio in logistic regression)
GEE	Generalized estimating equations
HEIs	Higher education institutions
KM	Kaplan–Meier
M	Mean
N	Sample size
OR	Odds ratio
p	p-value
SE	Standard error
χ^2	Chi-square statistic

Compliance with ethical standards

Ethical considerations

Ethical approval was obtained from the University Research Ethics Board of Pangasinan State University under Protocol No. 2024-0012a-CAMPIT-MODELING, dated March 18, 2024. Permission to access the institutional data was obtained from the appropriate authorities. All data were treated confidentially, and no personal identifiers were included in the dataset. The data were used solely for academic and research purposes.

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Astin AW (1984). Student involvement: A developmental theory for higher education. Journal of College Student Personnel, 25(4): 297-308.

Bean JP (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. Research in Higher Education, 12: 155-187. https://doi.org/10.1007/BF00976194

Cardona T, Cudney EA, Hoerl R, and Snyder J (2023). Data mining and machine learning retention models in higher education. Journal of College Student Retention: Research, Theory & Practice, 25(1): 51-75. https://doi.org/10.1177/1521025120964920

González-Morales MO, López-Aguilar D, Álvarez-Pérez PR, and Toledo-Delgado PA (2025). Dropping out of higher education: Analysis of variables that characterise students who interrupt their studies. Acta Psychologica, 252: 104669. https://doi.org/10.1016/j.actpsy.2024.104669 **PMid:39689520**

Kleinbaum DG and Klein M (2012). Survival analysis: A self-learning text. 3rd Edition, Springer, New York, USA. https://doi.org/10.1007/978-1-4419-6646-9

Martínez-Carrascal JA, Hlosta M, and Sancho-Vinuesa T (2023). 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): 1-21.
https://doi.org/10.19173/irrodl.v24i1.6589

Pascarella ET and Terenzini PT (2005). How college affects students: A third decade of research. 2nd Edition, Jossey-Bass, San Francisco, USA.

Patacsil FF (2020). Survival analysis approach for early prediction of student dropout using enrollment student data and ensemble models. Universal Journal of Educational Research, 8(9): 4036-4047.
https://doi.org/10.13189/ujer.2020.080929

Shafiq DA, Marjani M, Habeeb RAA, and Asirvatham D (2022). Student retention using educational data mining and predictive analytics: A systematic literature review. IEEE Access, 10: 72480-72503.
https://doi.org/10.1109/ACCESS.2022.3188767

Simpson O (2012). Supporting students for success in online and distance education. 3rd Edition, Routledge, New York, USA.
https://doi.org/10.4324/9780203095737

Vaarma M and Li H (2024). Predicting student dropouts with machine learning: An empirical study in Finnish higher education. Technology in Society, 76: 102474.
https://doi.org/10.1016/j.techsoc.2024.102474

Yan X, Fan J, Levine RA, Jacobs E, and Stronach J (2025). Analysis of risk factors for student retention. Journal of College Student Retention: Research, Theory & Practice.
https://doi.org/10.1177/15210251251378413

York TT, Gibson C, and Rankin S (2015). Defining and measuring academic success. Practical Assessment, Research & Evaluation, 20(5): 1-20.