# Survival Analysis in Education: A Systematic Literature Review, Bibliometric Insights, and Methodological Comparison

V WOS vse “Survival analysis” AND “Education” in glede n ate v abstract / keyword Koliko posamenih.

Priloga 10 najbolj citiranih SLR

## Abstract

## Keywords

Survival Analysis, Education Research, Dropout Prediction, Student Retention, Time-to-Event Analysis, Systematic Literature Review, Bibliometric Analysis, Methodological Comparison

## Introduction

Educational attainment plays a crucial role in personal development, economic mobility, and societal progress. However, dropout remains a significant issue across various educational levels, affecting both individual students and educational institutions. Understanding the factors influencing student retention and dropout rates is essential for developing effective policies and interventions. One powerful methodological approach to studying time-dependent educational outcomes is survival analysis, a statistical framework designed to model time-to-event data. This study aims to explore the application of survival analysis in education by conducting a systematic literature review, performing a bibliometric analysis, and comparing methodological approaches within the field.

### Dropout in education

Dropout rates in education vary across different regions, demographics, and educational levels, influenced by factors such as personal and economic issues, academic performance and satisfaction, social and institutional support, psychological determinants and demographic variables (Addison & Williams, 2023; Pusztai et al., 2022; Wild & Schulze Heuling, 2020). High dropout rates can lead to long-term negative consequences, including individual reduced impact like employment opportunities (Kiprianos & Mpourgos, 2025) and societal impacts (Ressa & Andrews, 2022). Policymakers and educators have long sought ways to mitigate dropout rates through targeted interventions. However, accurately predicting and understanding dropout behaviour requires robust analytical methods that account for time-dependent risk factors and censoring, making survival analysis an ideal tool for such investigations (Wild & Schulze Heuling, 2020).

### Survival analysis methods

Survival analysis encompasses a suite of statistical techniques designed to model the time until the occurrence of a specific event, such as student dropout or graduation. In the educational context, these methods are particularly valuable due to their ability to account for censoring—cases where the event of interest has not yet occurred—and to incorporate time-varying covariates that influence risk over time (Wild & Schulze Heuling, 2020).

The most commonly used techniques include, but are not limited to:

* Kaplan-Meier Estimator (Lee, 2023): A non-parametric method used to estimate survival probabilities over time without assuming an underlying distribution.
* Cox Proportional Hazards Model (Lee, 2023): A semi-parametric model that assesses the impact of covariates on the hazard function, widely used for its flexibility and interpretability.
* Discrete-Time Survival Models (Schmid & Berger, 2021): Suitable for data collected in regular intervals (e.g., academic years), allowing for logistic regression frameworks.
* Competing Risks Models (Schmid & Berger, 2021): Applied when individuals are exposed to multiple potential events, enabling the analysis of cause-specific outcomes.

These methods enable researchers to capture the dynamic and multifactorial nature of educational trajectories, offering nuanced insights into factors that influence student persistence and attrition over time.

### Existing literature

Despite the long-standing application of survival analysis in fields like medicine and engineering, its systematic use in educational research remains relatively limited. One of the most influential and methodologically rigorous contributions is the systematic literature review by (Willett & Singer, 1991) titled "From Whether to When: New Methods for Studying Student Dropout and Teacher Attrition." This foundational work introduced educational researchers to the core principles of survival analysis, emphasizing its relevance for modelling time-to-event educational outcomes. The authors showcased a range of survival methods—such as Kaplan-Meier estimators and discrete-time hazard models—and applied them to teacher attrition and student dropout, arguing for the necessity of analysing when events occur, not just if they occur. Their review highlighted the challenges of censoring in educational datasets and offered a conceptual and methodological framework that continues to inform survival-based educational studies today​.

However, as survival analysis techniques have advanced—particularly through the integration of machine learning, expanded covariate modelling, and broader educational contexts—there is a need to revisit this body of literature through the lens of newer empirical contributions, even if they do not meet the standard of being review type.Tu bi dodal še kaj, ampak SLR/Meta je edini na naših rezultatih, bi vseeno dodali kaj pravi npr. 5 najbolj citiranih?.

### Research Questions

This study seeks to address the following research questions:

1. What are the predominant survival analysis methods used in education research?
2. How has survival analysis been applied to model student retention and dropout rates?
3. What are the key trends and contributions in the literature, as identified through bibliometric analysis?
4. How do different survival analysis methods compare in terms of predictive accuracy and applicability to educational data?

## Methods

### Literature search

To assess the state of research on survival analysis in education, a systematic search was conducted using the Web of Science (WoS) database. The last search was performed in April 2025 using the following query:

("survival analysis" OR "event history analysis" OR "hazard model" OR "time-to-event analysis" OR "Kaplan-Meier" OR "Cox regression" OR "discrete-time survival" OR "competing risks") AND ("education" OR "educational attainment" OR "student retention" OR "dropout" OR "graduation" OR "academic success" OR "school completion" OR "college completion" OR "student success").

This search yielded NNN results, which were subsequently screened based on predefined inclusion and exclusion criteria as summarized in Table X.

|  |  |
| --- | --- |
| Inclusion Criteria | Exclusion Criteria |
| Written in English | Book chapters, conference papers, or non-peer-reviewed works |
| Published between 2019 and 2024 | Expressions of concern or retracted publications |
| Classified as peer-reviewed journal articles |  |
| Utilizing qualitative or quantitative methodologies relevant to survival analysis in education |  |
| In subject area |  |

Search resulted

The years prior to XXX were omitted because survival analysis in education was still an emerging topic during those years, making older studies less relevant to current trends. Additionally, 2025 was excluded because the available data for that year was still limited and emerging, with only XXX articles meeting the criteria.

To ensure the comprehensiveness and robustness of our bibliometric analysis, we extended our search beyond WoS and conducted an additional query in the Taylor & Francis database. The search string was structured as follows:

[[All: "survival analysis"] OR [All: "event history analysis"] OR [All: "hazard model"] OR [All: "time-to-event analysis"] OR [All: "kaplan-meier"] OR [All: "cox regression"] OR [All: "discrete-time survival"] OR [All: "competing risks"]] AND [[All: "education"] OR [All: "educational attainment"] OR [All: "student retention"] OR [All: "dropout"] OR [All: "graduation"] OR [All: "academic success"] OR [All: "school completion"] OR [All: "college completion"] OR [All: "student success"]] AND [All Subjects: Education]

This query yielded 367 results, reflecting a slightly broader pool of publications compared to WoS. To assess the overlap and uniqueness of these results, we employed a fuzzy matching algorithm to compare the retrieved articles against the WoS dataset. This technique allows for minor variations (e.g., one-character differences, formatting inconsistencies, or slight modifications in metadata) while still identifying potentially identical studies across databases. The results of the fuzzy matching analysis revealed that: 44 articles (12%) were a 100% match between the two databases. 48 additional articles (13%) exhibited a 90% or higher similarity score, indicating potential duplication with slight variations in metadata. The remaining 275 articles (75%) did not meet the matching threshold, suggesting a relatively low overlap between Taylor & Francis and WoS datasets. These findings highlight that while Taylor & Francis includes a number of relevant articles, its overlap with WoS is limited, emphasizing the importance of multi-database searches in conducting a comprehensive systematic review. The relatively low alignment percentage also suggests a degree of publication exclusivity between these databases, which may be influenced by indexing policies and journal coverage differences.

The selected studies were further filtered based on their methodological approach, as detailed in the Data Analysis section.

### Data analysis

The final selection of NNN articles was analysed using the Python programming language to construct and visualize the bibliometric network. While Python offers existing libraries such as pyBibX for bibliometric and scientometric analysis, we opted to develop a customized solution tailored to our specific research needs.

## Results

The final dataset consisted of 297 documents, with no duplicates, covering a timespan from 1991 to 2025. The majority of the records were peer-reviewed journal articles (256), followed by articles in press (15), conference proceedings (22), book chapters (2), and review papers (2). The dataset spanned 50 countries and 302 institutions, with research published in 110 different sources. The dominant language was English (290 documents), with a few studies published in German (2), Spanish (3), Portuguese (1), and Dutch (1).

In terms of bibliometric completeness, most entries contained abstracts (98.99%), author affiliations (99.66%), and DOIs (93.60%), while 78.45% included author keywords and 88.55% had Keywords Plus. Collaboration was notable, with 768 authors contributing to 297 documents, leading to an average collaboration index of 2.94. Of these, 62 were single authored while 235 were multi-authored.

Regarding impact, the dataset accumulated 7,015 total citations, translating to an average of 23.62 citations per document. The highest recorded h-index was 7, indicating a moderate level of influence within the field. Additionally, citation distribution revealed an average of 9.13 citations per author, 23.23 per institution, and 57.3 per source.

To assess whether our search strategy effectively captured the key themes of survival analysis in education, we conducted a word cloud analysis of the extracted dataset. The word cloud (Figure X) visualizes the most frequently occurring terms in the titles, abstracts, and keywords of the selected studies. Larger words indicate higher frequency, reflecting dominant topics within the dataset. The results confirm that our search query successfully encompassed the core concepts of survival analysis and education. The most prominent terms include "student," "study," "school," "time," "survival," "analysis," "teacher," "dropout," "retention," and "academic"—all of which align with our research focus. The presence of terms related to statistical methods (e.g., "model," "risk," "data," "event," "rate") further validates that the dataset is methodologically relevant. Additionally, words like "higher education," "university," and "college" indicate that the dataset captures various levels of the education system, ensuring broad applicability.

A close-up of words

AI-generated content may be incorrect.

Figure : Word cloud analysis.

To analyze the geographical distribution of research on survival analysis in education, we mapped the number of documents produced by each country, considering the affiliations of all contributing authors (Figure X). Countries shaded in darker colours indicate higher research output, while lighter colours represent lower contributions. This analysis highlights the uneven distribution of research output, suggesting that survival analysis methods in education are predominantly explored in North America and select European countries, while developing regions remain underrepresented.

A map of the world

AI-generated content may be incorrect.

Figure : Production for each Country (count is made considering each doc author).

The collaboration network presented in Figure X illustrates the global research partnerships in survival analysis applied to education. Each node represents a country, while the lines (edges) between them indicate international co-authorships. The thickness of the edges reflects the strength of collaboration, with thicker lines representing more frequent research partnerships.

From the visualization, we observe that:

European countries exhibit dense collaboration networks, particularly among the United Kingdom, Germany, France, Portugal, and the Netherlands. North America (United States and Canada) plays a central role, collaborating extensively with European and Australian researchers. Australia emerges as another key player, engaging in significant partnerships with both Asian and European countries. South America and Africa show some engagement, but their collaboration networks are relatively sparse, suggesting limited international research connections. India and Pakistan stand out as key contributors from Asia, actively collaborating with other regions.

The strong interconnectedness of researchers from Western countries suggests a well-established academic exchange, while developing regions show fewer international collaborations. These findings highlight the potential for expanding global partnerships, particularly by fostering research collaborations with underrepresented regions to diversify methodological approaches and broaden the scope of survival analysis in education.

A map of the world

AI-generated content may be incorrect.

Figure : Collaboration Analysis Between Countries.

## Discussion

## Limitations and future directions

## Data availability

## References

Addison, L., & Williams, D. (2023). Predicting student retention in higher education institutions (HEIs). *Higher Education, Skills and Work-Based Learning*, *13*(5), 865–885. https://doi.org/10.1108/HESWBL-12-2022-0257

Kiprianos, P., & Mpourgos, I. (2025). School failure and life trajectories. *Journal of Adult and Continuing Education*, 14779714251314681. https://doi.org/10.1177/14779714251314681

Lee, S. W. (2023). Kaplan-Meier and Cox proportional hazards regression in survival analysis: Statistical standard and guideline of Life Cycle Committee. *Life Cycle*, *3*, e8. https://doi.org/10.54724/lc.2023.e8

Pusztai, G., Fényes, H., & Kovács, K. (2022). Factors Influencing the Chance of Dropout or Being at Risk of Dropout in Higher Education. *Education Sciences*, *12*(11), 804. https://doi.org/10.3390/educsci12110804

Ressa, T., & Andrews, A. (2022). HIGH SCHOOL DROPOUT DILEMMA IN AMERICA AND THE IMPORTANCE OF REFORMATION OF EDUCATION SYSTEMS TO EMPOWER ALL STUDENTS. *International Journal of Modern Education Studies*, *6*(2). https://doi.org/10.51383/ijonmes.2022.234

Schmid, M., & Berger, M. (2021). Competing risks analysis for discrete time‐to‐event data. *WIREs Computational Statistics*, *13*(5), e1529. https://doi.org/10.1002/wics.1529

Wild, S., & Schulze Heuling, L. (2020). Student dropout and retention: An event history analysis among students in cooperative higher education. *International Journal of Educational Research*, *104*, 101687. https://doi.org/10.1016/j.ijer.2020.101687

Willett, J. B., & Singer, J. D. (1991). From Whether to When: New Methods for Studying Student Dropout and Teacher Attrition. *Review of Educational Research*, *61*(4), 407–450. https://doi.org/10.3102/00346543061004407

## Supplementary files