Microeconometric Analysis of the Relationships Between Early Alert Systems and Student Retention

Scott Harrison

BEcon, BEcon (Hons)

A thesis submitted for the degree of Doctor of Philosophy

of the University of New England

UNE Business School

University of New England

Armidale, New South Wales, Australia

November 2015

Certification

I certify that the substance of this thesis has not already been submitted for any degree and is not currently being submitted for any other degree or qualification.

I certify that any help received in preparing this thesis and all sources used have been acknowledged in this thesis.

Signed:

[redacted]

Scott Harrison

23/11/2015

Abstract

The main objective of this study is to evaluate the relationship between Early Alert Systems (EAS) and student retention. Specifically, the study aims to: (i) examine the effects of demographic, institutional and learning environment variables on student retention, (ii) examine the effects of EAS on student retention, and (iii) assess the financial implications of the interaction between EAS and student retention. Selected microeconometric models were estimated using data for 16,124 undergraduate students extracted from a case study university. The data was captured over three years between 2011 and the beginning of 2014.

Key findings of this study show that demographic, institution, student performance and workload variables all exhibit statistically significant relationships with retention measures at the case study institution. Furthermore, the EAS had a positive effect on increasing students' length of enrolment. Females are more likely to discontinue, but are also more likely to complete their course. Aboriginal and Torres Strait Islander (ATSI) students are more likely to be retained than non-ATSI students. Institutional factors such as the type of course, the school a student enrols in, or mode of enrolment all affect student's retention rate. Variables capturing student performance and workload further affect retention. Periods of inactivity during students' enrolment was one of the strongest factors affecting measures of student retention. The study also finds that demographic, institution, learning environment and EAS variables are subject to significant temporal effects. Using weekly observations, temporal effects were captured up to 156 weeks (3 years) of student enrolment, yielding a total of 1,119,170 observations. Using survival modelling, the study provides an unprecedented degree of accuracy in estimating the relationship between explanatory variables and the hazard of discontinuing over time.

Finally, the financial implications of the EAS was evaluated using treatment effects modelling. On average, students identified by the EAS for targeted support remained enrolled for an extra 14 weeks than students not identified by the EAS. The additional revenue in tuition fees caused by EAS identification is estimated to be $4,004 per student. It is concluded that early alert systems have significant financial benefits, initiating support services that positively impact on student outcomes.

Acknowledgements

I would like to acknowledge the tireless efforts of my supervision team, Associate Professor Renato (Rene) Villano, Adjunct Professor Grace Lynch and Dr George Chen in assisting me throughout this study. I have been fortunate to have a team of supervisors from such diverse research backgrounds, who were all so like-minded.

Initially my undergraduate lecturer in 2007, Rene Villano continually encouraged and supported me to develop a career in research. As my honours supervisor, he guided me to attaining First Class Honours, which permitted me to enter the Ph.D. program. As a colleague, Rene fostered my understanding and interest in the learning process, with opportunities to be a tutor, lecturer and unit coordinator. For five years I have been mentored in the finer aspects of teaching; the supportive and inclusive approach that Rene demonstrates, I have tried to emulate. During this time and with Rene's encouragement, choosing to study a Ph.D. in learning analytics in many ways was serendipitous. The fact that I made it to completion of this study is a testament to Rene's time, energy, patience, encouragement and support. For the major role Rene has had in my career and personal development, I am forever grateful.

My sincere gratitude also extends to Grace Lynch. As an expert in the field, Grace introduced me to learning analytics, furthered my understanding of the field, introduced me to the learning analytics community, and provided me with the opportunity to follow my passion for both learning and economics. The time and support Grace has shown me throughout this process has been nothing short of amazing.

I am thankful for the time and thoughtful commentary from George Chen. I have been honoured to be both George's student and colleague during this study. The encouragement I received was definitely needed at times, and George's sense of humour helped lighten this journey.

I would like to acknowledge the University of New England, in particular Milly Taylor and the staff of the Corporate Intelligence Unit (CIU). Your support of this project from the outset has been paramount to the success of this research. I would like to acknowledge the support of the UNE Business School. The financial, academic and social support I received throughout my time with the school has been invaluable. I would like to thank in particular Pauline Fleming for initial introductions and help in finding this topic, and Professor Euan Fleming for the supervision through my honours and early days of commencement of the Ph.D. Without the support of the school, this study may not have happened.

I also want to acknowledge the support from the Society for Learning Analytics Research (SoLAR). Being accepted to attend the doctoral consortium at the annual Learning Analytics and Knowledge conference in 2015 (LAK15), was one of the highlights of my time as a student. The feedback I received from the many experts in the field assisted me to develop this study further, and inspired me to completion.

To my editors, Peter Hammond, Robin Hammond and Dr Stephen Norris. The prompt, detailed feedback and discussion helped immensely. I enjoyed the many conversations we had not only on the finer points of the thesis, but the broader societal implications of learning analytics and discussions on the future of education.

I also need to acknowledge the vast social network of friends that I had support me. To Georgina and Nigel Watkins for inspiring me to study. Thanks goes to Liam, Rob, Peter, Ben and Kylie, who have assisted greatly in my extra-curricular education in Japanese arts. To Wendy, who brought many unique experiences to my world. To Emily, thanks for being the kind and caring friend I needed on so many days. To my colleague Dr Cornelia Cohrs, thanks for all the help and good conversations. To Ed and Mitch, thanks for the many debriefs that needed to occur along the way. To the floor of W39: Kanar, Derek, Bright, Valantine, Kodjo, Wasanthi, Ethan, Omphile and Peter, the informal conversations, cultural lessons and on-going humour has made the study so much easier. The collegial atmosphere fostered between students was only further enriched by the ‘Emperor of W39’. To Enactus UNE, I am proud to have been part of community projects that make such a huge difference to people’s lives. We will always have Mexico! Dan, Colette, Adrienne, Nicole, Tonia, Jo and Rachel for just always being there.

Finally, to my family. To Clinton and Nina Harrison, thank you for some great memories. Clint, your initial support when I started university helped me make it through. Living close to Jocelyn and Josh Reynolds, and my beautiful nieces Matilda, Claudia and nephew Tristan, has provided me with a much needed outlet of magic, fun and laughter. Finally, to Joy and John Harrison, who have suffered what no parents should go through, a child still living at home in their 30s. To all of you, this would not have been possible without your love and support.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ASB	All-Students-Base: a model that uses all students as the sample space with only a single instance of EAS triggers occurring in a given week.
ASM	All-Students-Multiple: a model that uses all students as the sample space with multiple instances of EAS triggers occurring in a given week.
ATE	Average Treatment Effects
ATET	Average Treatment Effects on the Treated
ATN	Australian Technology Network
ATSI	Aboriginal and Torres Strait Islander
AWE	Automated Wellness Engine: The Early Alert System developed at the University of New England
CASB	Controlled All-Students-Base: a model that uses all students as the sample space with only a single instance of EAS triggers occurring in a given week
CASM	Controlled All-Students-Multiple: a model that uses all students as the sample space with multiple instances of EAS triggers occurring in a given week
CISB	Controlled All-Students-Base: a model that uses only EAS identified students as the sample space with only a single instance of EAS triggers occurring in a given week
CISM	Controlled All-Students-Multiple: a model that uses only EAS identified students as the sample space with multiple instances of EAS triggers occurring in a given week
CIU	Corporate Intelligence Unit (for University of New England)
EAS	Early Alert System
FYE	First Year Experience
G8	Group of Eight
GPA	Grade Point Average
HELP	Higher Education Loan Program
IIA	Independence of Irrelevant Alternatives
IRU	Innovative Research Universities
ISB	Identified-Students-Base: a model that uses only EAS identified students as the sample space with only a single instance of EAS triggers occurring in a given week
ISM	Identified-Students-Multiple: a model that uses only EAS identified students as the sample space with multiple instances of EAS triggers occurring in a given week
LR	Likelihood Ratio
MSE	Mean Squared Error
NCES	National Centre for Education Statistics
OAAI	Open Academic Analytics Initiative

OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PH	Proportional Hazards
POM	Predicted Outcome Means
PSM	Propensity Score Matching
QUT	Queensland University of Technology
RMIT	Royal Melbourne Institute of Technology
RRR	Relative Risk Ratio
RUN	Regional Universities Network
SET	Special Extension of Time
UNE	University of New England, Australia

Chapter 1 - Introduction

1.1 Preface

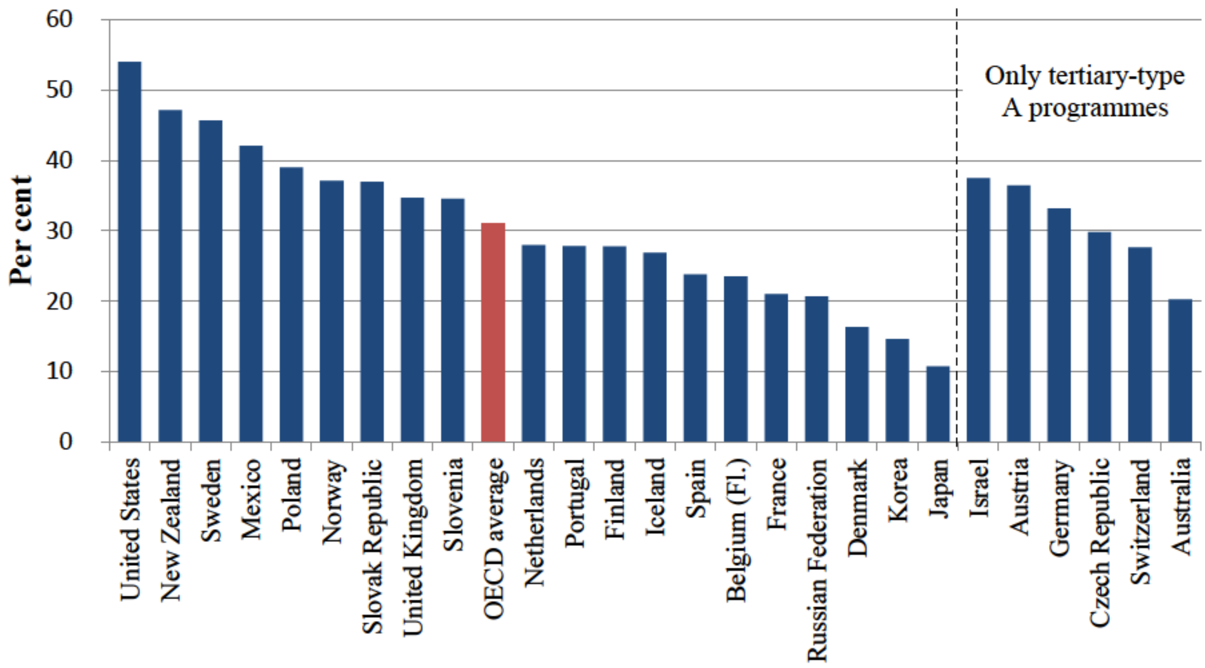
Not every student who starts an undergraduate degree completes the course to graduation. It is naive to think, however, that universities can stop students from discontinuing their studies. Circumstances change, affecting people's decisions and commitment to university study. With information technology now an important aspect of universities, Early Alert Systems (EAS) are becoming a major foundation of support programs within universities. This study explores the links between EAS and student retention.

There are many different institutional settings that affect how retention is defined. Generally, a student is considered retained if they remain enrolled, enabling them to continue their education. This chapter introduces the current trends in student retention, both globally and within Australia to frame the magnitude of retention as an issue. A discussion of the research problem is presented, followed by the main research questions and objectives of the study. Finally, this chapter concludes with a thesis outline and discussion on the significance of the research presented.

1.2 Recent retention trends

Internationally, a key publication on education in general is the 'education at a glance' report, published by the Organisation for Economic Co-operation and Development (OECD). This annual report on education uses data provided by member nations to allow comparisons of education indicators. Past reports used indicators of educational attainment, how many students complete tertiary education, how much is spent per student and public costs and benefits of attaining tertiary education (OECD, 2013, pp. 5-11).

In producing the annual report, many published indicators are included across years. The report also provides occasional indicators such as the proportion of students who enter tertiary education without graduating. This indicator last appeared in the 2010 report, using data from 26 member nations to provide a comparative analysis. The indicator indirectly measures student retention by capturing student attrition rates within each nation. The 2010 report uses 2008 data to compare attrition rates, presented in Figure 1.1 below.



Note: Some students who have not graduated may be enrolled, or may have finished their education at a different institution than the one they started at, like in the United States

1. Includes students entering single courses who may never intend to study all courses needed for a degree

Figure 1.1 - Proportion of students who enter tertiary education without graduating, 2008

(Adapted from OECD, 2010, p. 72)

Inferences made from direct comparison between countries are limited. This is due to differences in how tertiary systems are structured, regulated and funded in each nation. One valid inference is that relative to other OECD nations, Australia is performing below the OECD average on students not reaching graduation, with Denmark, Korea and Japan all experiencing lower proportions of student who enter tertiary education without graduating.

Within Australia, the recently formed Department of Education and Training is responsible for universities at a federal level. Before 2014, universities were part of the Department of Industry. Australian universities report operating information to the department, including information on student retention. Published information is available through the department's website or through the MyUniversity portal. For Australian reporting purposes, the Department of Industry defines the student retention rate as:

> Retention rate for year (x) is the number of students who commenced an undergraduate course in year (x) and continue in year (x+1) as a proportion of students who commenced an undergraduate course in year (x) and did not complete the course in year (x) (Department of Industry, 2012, Table 4.9).

Student retention rates in Australia over the last decade have improved slightly as indicated in figure 1.2, rising from around 81% in 2002 to 83% in 2011.

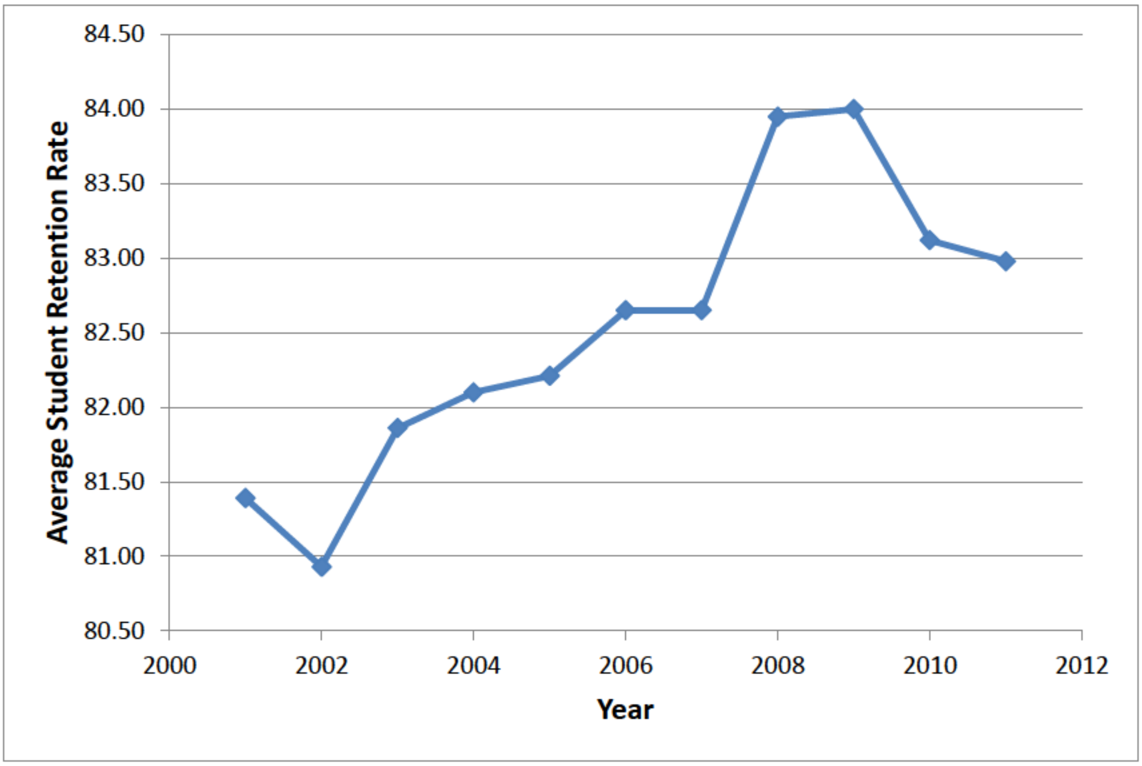


Figure 1.2 - Average student retention rate in Australia over time

(Department of Industry, 2012, Table 4.9)

From the data provided by the Department of Industry, analysis of retention for peer groups provides a deeper understanding of retention rate differences within Australian universities. Four major peer groups exist within the Australian tertiary education sector, being the Group of Eight (G8), the Regional Universities Network (RUN), Australian Technology Network (ATN) and Innovative Research Universities (IRU) (Australian Education Network, 2014a). Figure 1.3 shows the different average retention rates for the four major university groupings as well as the average of all remaining institutions. The Department of Industry 2012 figures show a disparity in average retention rates in Australia between university groups.

The Australian higher education system serviced over 1.3 million students in 2013 (Department of Education, 2014a) at a cost of 29 billion dollars in the 2012/2013 financial year (Australian Bureau of Statistics, 2014). Among all students studying in Australia, 70.4 per cent of students studied in a full-time capacity and 25 per cent of these students were international students. Higher education is a significant sector within the economy with 41.28

per cent of the workforce aged between 25 and 65 years having attained a tertiary education qualification (OECD, 2014).

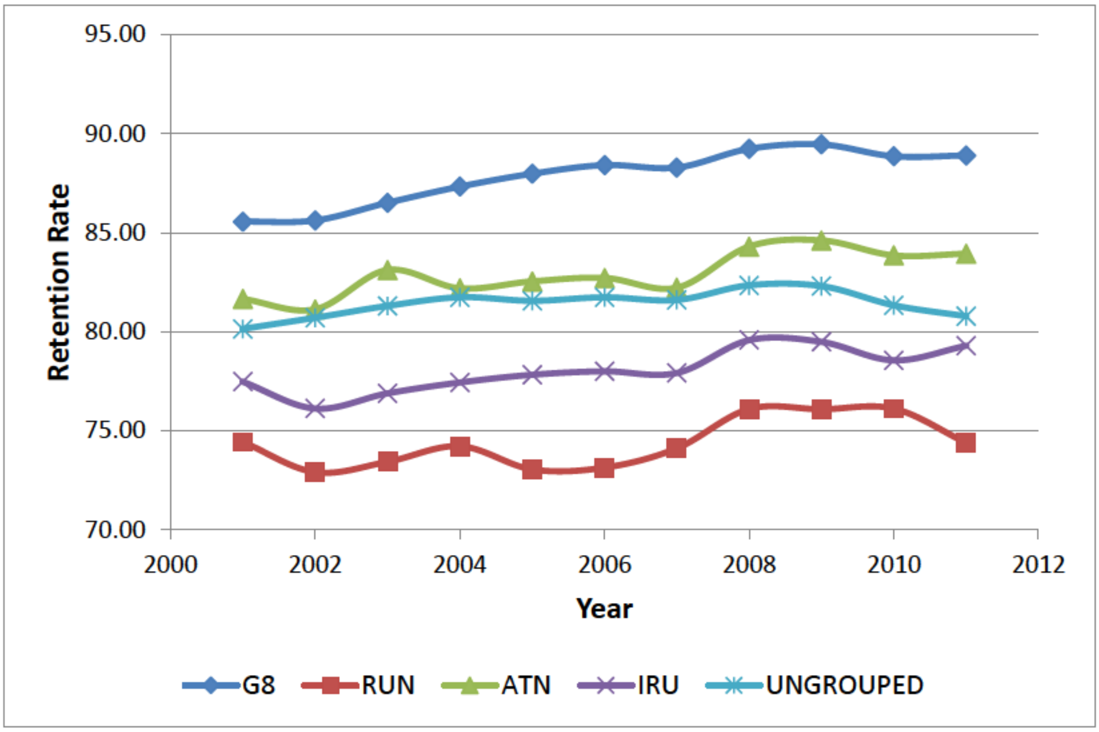


Figure 1.3 – Average student retention rates by university grouping

The review of Australian higher education (Bradley, Noonan, Nugent, and Scales, 2008) outlined ways in which the sector can be improved through to 2020. One recommendation of the review was that "to support the achievement of the target to increase the proportion of students from disadvantaged groups to 20 per cent by 2020, 4 per cent of all funds for teaching will be directed to outreach and retention initiatives" (Bradley et al., 2008, p. xiv). The focus on funds being directed to outreach and retention is a major motivation for this study. As more funding is moved into outreach and retention, it is paramount that initiatives and programs are informed and evaluated by an evidence based approach.

1.3 The research problems

The learning environment is a complex space of interactions between students, teachers, technology, institutions and the broader environment. When discussing student retention problems, it is easy to mistake that the overall problem is students discontinuing their studies. There will always be some level of attrition which cannot be controlled. The research problem is actually understanding whether universities are maximising student retention in their respective university setting.

This is an important distinction for it places focus on the universities' responsibility to support students during their study. When a student fails to be retained by an institution in their studies, there are three key areas impacted by this decision. The first and foremost is the student, for which the effects of discontinuing can be many fold. This may be the effect of failing to attain qualifications which can affect employment prospects or their mental health is possibly affected through negative feelings associated with leaving their studies. The institution itself is affected by the student leaving. This is not only financially in terms of their bottom line and tuition fees paid but the institutions' reputations are affected. For society as a whole, there is a social cost associated with educating someone over being employed and when that student fails to attain the qualification, the overall net social benefits from a more educated population fail to be passed on.

The second key research problem is finding the strategies that are used to address student retention related issues. There are many approaches by which the student retention rate can be improved. These can come from varying levels in terms of government, institution, student support and peer support initiatives. There is a wide breadth of strategies that have been developed historically to address student retention. However this poses a problem in that governments, institutions, student support and peers all have limited resources, so how does one select the best strategy to address student retention with the limited resources available?

The rise of information technology has enabled universities to collect massive data sets on students and their interactions with the learning environment. This has given rise to the field of research termed learning analytics. One aspect of learning analytics has been to use data to identify students in need of targeted student support. Targeted support systems require significant university resources in development, deployment and ongoing operations. Where there is an interaction between the institutions providing support to the student, the effect of the support is unquantified. A key problem for universities generally is justifying the implementation of these complex programs and determining what (if any) effect this has had on student retention. So the problem becomes, do targeted student retention programs, driven by learning analytics, work? Furthermore, do they represent fair value for money given the resource expenditure required to develop, implement, and maintain sophisticated information technology?

1.4 Research questions

In view of the research problems outlined above, the main research question for this study

1: What are the links between EAS and student retention?

Specifically, using a microeconometric approach, this study is aimed to answer the following research questions:

2: What variables affect student retention rate of undergraduate students?

3: What is the effect of the EAS on student retention rates?

4: What is the relationship between the variables affecting student engagement and variable affecting student retention?

5: What are the financial implications of improving student retention rates?

1.5 Objectives of the study

In light of the above research questions, this study has three main objectives:

1. Investigate, analyse and understand the demographic, institution, learning environment and temporal variables affecting undergraduate student retention rates.
2. Analyse in a quantitative framework the relationship of EAS on the undergraduate student retention rate.
3. Quantify the potential financial benefit of an EAS.

1.6 Case study institution

To answer the research questions and achieve the objectives of this study, a data set was obtained from the University of New England (UNE), Australia. The University of New England (UNE) is one of thirty-nine universities in Australia (Australian Education Network, 2014b), and is part of the Rural Universities Newtwork discussed in section 1.2. Established in Armidale, New South Wales in 1938 as a campus of the University of Sydney, UNE was Australia's first regional university campus. UNE transitioned from a regional campus to an autonomous university in its own right in 1954. During this period, UNE pioneered teaching to external students by correspondence (University of New England, 2014a), which continues to be a fundamental aspect of the university. In 2013, UNE had 22,389 students enrolled, of which 78.9 per cent were off-campus students. UNE has a significantly larger off-campus student cohort when compared to national averages (Department of Education, 2014b), making UNE a unique institution in the higher education sector. Additionally, UNE, through distance education, has been able to create a strong focus on supporting students from disadvantaged groups. More detail on UNE is provided in the discussion of the data set in Chapter 4.

The data set contains student-level data covering demographic, institutional, student performance and workload variables. Additionally, the data set contains detailed weekly information on the identification process used by an EAS implemented at the institution from 2011. The EAS was designed to identify students at risk of disengaging from their studies. As such, this study analyses the EAS and how it may also extend to effects on student retention.

1.7 Approval process

This study uses detailed student level data. Given the privacy and security concerns surrounding student data, extensive ethical and legal processes were undertaken. The ethics approval process for data extraction required all information to be de-identified and could not be re-identifiable. Additionally, all data was required to be stored securely throughout the study, with the data to be returned to the institution upon completion. Ethics approval was granted to extract data from the 1st of July, 2013 under approval number HE13-152. Additionally, a confidentiality deed was signed between the institution and researcher on the 18th of February, 2014. This was to ensure confidential information was appropriately handled within the legal framework of the case study institution.

1.8 Thesis outline

This thesis is comprised of nine chapters in total. Chapters one to four provide the contextual and background information, with chapters five to eight provide empirical results which address the research questions outlined above. Below are brief descriptions of each chapter.

Chapter 1 establishes the topic of the thesis. The chapter introduces key concepts associated with student retention, research problems and research questions. It then outlines how the thesis will address the research questions in presenting the thesis outline.

Chapter 2 provides a review of current literature pertinent to this study. The study starts with a discussion of student retention theory and empirical analysis. It then reviews literature on EAS, covering: how EAS have developed over time, which institutions are currently using EAS; issues associated with early alert systems including privacy and ethics; and how the effects of EAS have been quantified to date.

Chapter 3 provides a review of previous empirical studies on student retention, providing the foundation literature to develop models used in this study. The statistical methods used in this study are also presented in full, including multinomial logistic regression analysis, multiple regression analysis, survival analysis and treatment effects modelling. To conclude, model selection, calibration and specifications used in the empirical chapters are presented.

Chapter 4 discusses the data set acquisition process, variables captured within the data set, how data cleaning was conducted and relevant descriptive statistics on the data set. This establishes the underlying trends that are present within the data set used in the empirical chapters.

Chapter 5 presents empirical results of the likelihood of discontinuation, enrolment or completion based on demographic, institutional, student performance and workload variables. This model identifies both the factors that can affect student retention and the effects on student retention associated with the Early Alert System. This commonly used empirical method of analysis within the student retention literature highlights advantages and disadvantages of this method. From this, temporal models are developed and expanded on in chapters 6 to 8.

Chapter 6 links increased retention to increased time enrolled. Introducing time as the dependent variable, the model captures effects associated with students' enrolment with the length of enrolment. This model identifies both the variables that can affect student retention and the effects on student retention associated with the EAS.

Chapter 7 provides a comprehensive analysis of student retention under temporal effects. The chapter explores multiple model specifications to capture variables affecting student retention rates using a survival analysis approach. The Cox proportional hazard model allows for temporal effects and unbalanced data to be analysed in a non-parametric form. The model makes few statistical assumptions and can handle the complex nature of student retention data. The effect of the EAS is also tested in various model specifications, to determine both short-run and long-run effects of the system. The chapter concludes by establishing a causal relationship between the EAS and increased length of enrolment. The results of this chapter form an important benchmark for future student retention analysis using complex temporal data.

Chapter 8 estimates the effect of the EAS on the key financial metric, student tuition fees. Using treatment effects modelling of student retention, the dependent variable is student tuition fees paid and the treatment effect variable is the EAS. The statistical model provides a causal inference on the effect of the EAS in terms of the financial benefit. This chapter presents a new standard to estimate the financial implications of student retention programs.

Chapter 9 concludes with detailed discussion on the results of the thesis, contextualising the findings and indicating areas of future research on both EAS and student retention. Some key conclusions of the chapter include: the importance of appropriate methods required to analyse temporal data; the significance of demographic, institutional, student performance and workload variables effecting student retention; and financial implications of EAS and the financial effect of improving retention.

1.9 Contributions of the study

This study provides significant contributions to multiple areas of study in student retention and EAS analysis. A significant contribution comes from the analysis of the relationship between the EAS and student retention in a quantitative framework. The study demonstrates the complexity of modelling retention with detailed temporal data. In Chapter 7, complex temporal interactions are required to adequately capture how factors fluctuate in significance, over time. Importantly, the modelling provides solid statistical evidence as to the program's efficacy. It identifies areas where the system can be improved and a significant contribution to the literature can come from exposing weakness in EAS and their underlying assumptions.

While treatment effects models have been extensively used in other fields, this is the first study to use this method in evaluating possible causal effects of an EAS on student retention. Through using detailed student data, the study sets the standard for program evaluation with respect to such systems. A dual benefit is then gained in demonstrating the need for institutions to provide detailed data for future analyses to be performed.

A significant contribution from the study is combining detailed student data with financial information to value the effect of the EAS. Estimating the causal financial implications on student tuition fees provides an important metric for institutions to evaluate the effect of the system and its worth to the institution. For institutions and administrators,

> the hard truth is that there are many programs, retention or otherwise, that make claims upon institutional resources. In such circumstances, retention programs have to provide empirical evidence that resources committed to them are an investment that yields long-term benefits to the institution (Tinto, 2006, p. 10).

The research aims to provide the empirical evidence required to measure the financial implications to the case study institution. The research establishes a best practice approach for analysing quantitatively the causal effects of Early Alert Systems. This makes a significant contribution to advancing the literature and the learning analytics community in identifying appropriate approaches to valuing programs that effect learning.

The final area where the study makes a significant contribution, is demonstrating the need for large institutional data sets in learning analytics. The study uses data extracted from a data warehouse, capturing eight underlying information systems. The data is further enhanced with daily student observations attained by the EAS. The analysis conducted in this study shows a level of detail that is only possible when large data sets are made available. Removing restrictions caused by information silos, while maintaining ethical standards, provides major benefits for institutions by capturing the true complexity of the learning environment.

Chapter 2 - Review of Literature

2.1 Introduction

The relevant literature for this study comes from two distinct fields: student retention and learning analytics. The area of student retention research is a mature and well established body of knowledge, focusing on the problems of retention and attrition from both theoretical and practical viewpoints. In contrast, learning analytics is a comparatively new field, and universities globally are exploring new ways of leveraging the growing accessibility and quantity of organisational and digital data. Learning Analytics is not simply about counting hits or mapping discussions, it is about intelligent and thoughtful integration and interpretation of data in the context of human activity. The increased adoption of technologies to mediate activities in the education space enables access to large sets of data sources about student learning. The analysis of such digital traces can provide deep insights into numerous aspects related to teaching quality and student learning experience. The area of learning analytics has recently emerged to study the storage, analysis, visualisation and actions derived from data mining procedure applied to these data sources. Learning analytics is a bricolage of disciplines combining expertise from areas such as business intelligence, web analytics, academic analytics and educational data mining to support decision-making and planning in education at multiple institutional levels. The field is still in its infancy and the understanding of what constitutes learning analytics remains somewhat in flux.

Each of these areas of research frame the study as both a temporal analysis of student retention, and an evaluation of an EAS using student level data. A review on the background on student retention theory is provided with the associated empirical studies which tested the validity of the theories developed. This provides the foundations for model development in assessing student retention. The chapter then focuses on more recent developments in the field of learning analytics, covering recent approaches to studying retention. The chapter concludes with a discussion on recent early alert initiatives, the motivation for implementing EAS and how this study contributes to ongoing research in learning analytics.

2.2 Student retention theory

Student retention has been an issue of interest since the 1920s, with development of literature in the field a keen research interest, especially since the 1960s. This means “the literature on student retention is voluminous and arguably capricious” (Simons, 2011, p. 13). This occurred due to the wide variety of academic disciplines contributing to research in student retention, including psychology, education and social science. Arguably, the most relevant student retention literature was framed by strong discussions on theoretical models of retention. Theoretical approaches to understanding student retention have been a major component of student retention literature of the 1970s and 1980s. While these models themselves are now dated due to ongoing changes to the higher education sector, the theories create a solid foundation for understanding retention as a concept. Additionally, researchers and institutions frequently referenced four major works to discuss student retention. These major works come from Tinto (1975), Bean (1980), Astin, (1984) and Bean and Metzner (1985). These studies proposed a variety of models, definitions and frameworks used to describe and understand the nature of student retention at universities.

The first major student retention theory can be traced back to Tinto's (1975) seminal paper entitled "Dropout from higher education: A theoretical synthesis of recent research". Tinto began by making critical points that previous research had failed to disaggregate different paths students can take to dropping out from university. It was important to differentiate between students who leave an institution due to poor academic outcomes versus students who leave due to personal circumstances beyond their control. In response, Tinto developed the integration model of student attrition, represented in Figure 2.1.

Tinto used Durkheim's theory of suicide (1961) in combination with benefit-cost analysis concepts from economics to develop an institutional model of student retention. Like suicide, which Durkheim relates to a lack of social integration, Tinto postulated that the student dropout behaviour occurred due to a lack of academic and social integration.

Tinto's integration model described the longitudinal nature of university progression. Starting with background variables of family, individual attributes and pre-college schooling, background variables are related to goal and institutional commitment developed before entering university. Upon entering university, the level of commitment to goals and the institution will affect components of the academic system such as grade performance or peer-group interactions. How a student performs in the academic and social setting of the university are related to the level of academic and social integration experienced by the student.

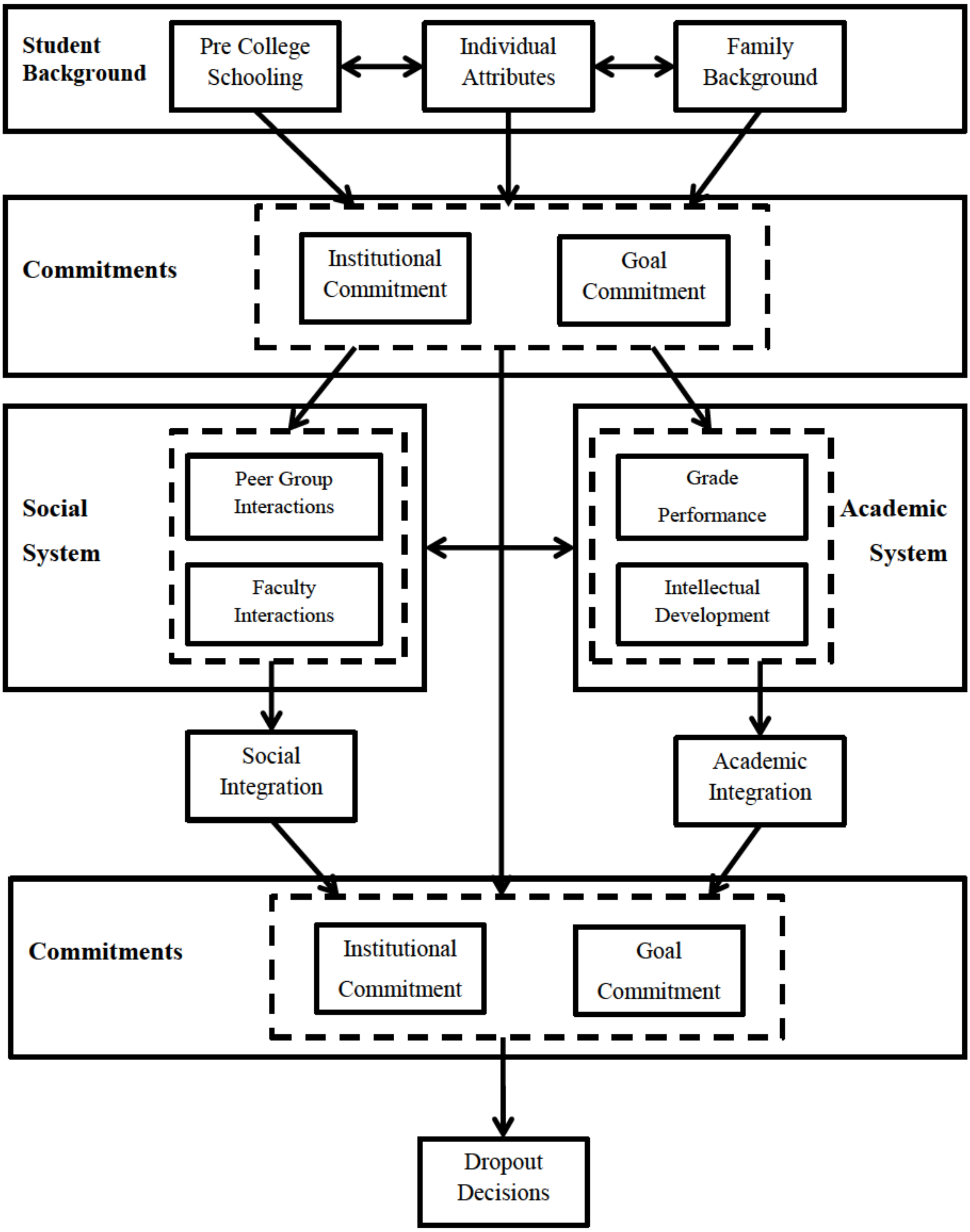


Figure 2.1 – Adaptation of Tinto's integration model of student integration

(Adapted from Tinto, 1975, p. 95)

Over time, students' commitment to the goal of completing the award and being committed to the institution need to be re-evaluated. This in turn affects the decision of a student to drop out from university. As Tinto writes:

> This theoretical model of dropout, diagrammed in [Figure 2.1], argues that the process of dropout from college can be viewed as a longitudinal process of interactions between the individual and the academic and social systems of the college during which a person's experiences in those systems (as measured by his normative and structural integration) continually modify his goal and institutional commitments in ways which lead to persistence and/or to varying forms of dropout (Tinto, 1975, p. 94).

When a student evaluates the decisions to remain at university or seek alternative activities, Tinto employs benefit-cost analysis. Students will continue to remain at university when the perceived benefits to cost ratio is greater than that for other alternatives such as paid employment, incorporating the influence of external events in the decision to drop out or not. Tinto concluded by discussing that the nature of goal and institutional commitment are key elements in distinguishing dropout behaviour. It is conjectured that those who exhibit low goal commitment, for example, commitment to the goal of graduating, coupled with low commitment to the institution, will be more likely to withdraw voluntarily not through low grades but due to a lack of benefit within the social systems.

Tinto's integration theory has served as a foundation in the research due to the model indicating the importance of social involvement in the student retention process. Tinto (1975, pp. 89-90) made a distinction between different forms of dropout behaviour, disaggregating the term dropout to account for the difference between voluntary versus compulsory withdrawal and temporary versus permanent withdrawal. This changed the tone of discussion for future research to look at how social aspects of university affected the student retention process.

Following Tinto's work was the model proposed by Bean in 1979 at the annual meeting of the American Educational Research Association and later published in 1980, as seen in Figure 2.2.

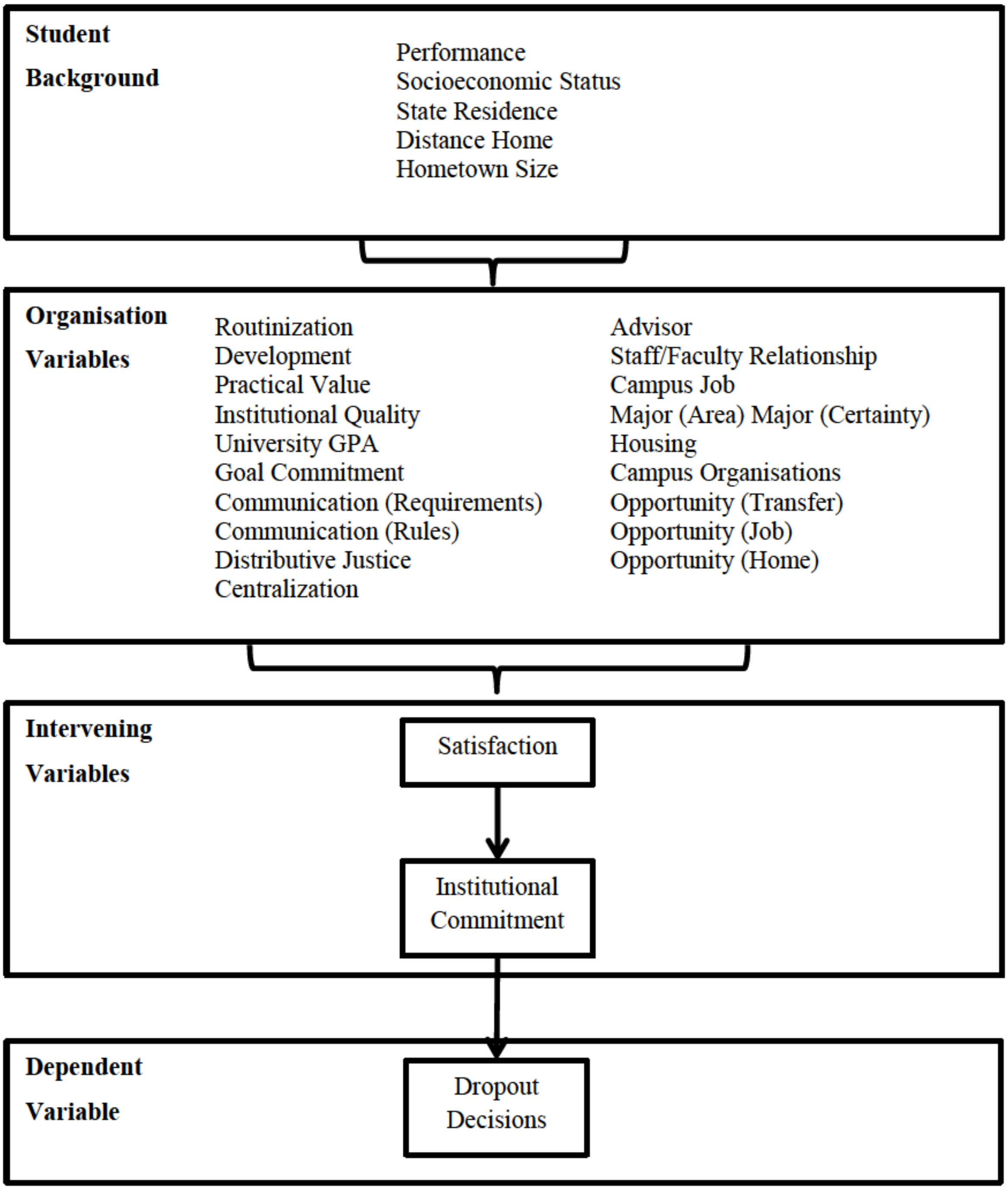


Figure 2.2 – Adapted Bean's student attrition model of student attrition

(Adapted from Bean, 1980a, p. 160)

Unlike Tinto (1975), whose model was developed from Dunkheim's suicide theory, Bean's "causal model of student attrition is similar to Price's (1977) model of employee turnover" (Bean, 1980a, p. 161). A criticism of models based on suicide theory was that "the link between dropping out of school and suicide is suggested as a theoretical basis for these models, but there is insufficient evidence for this premise" (Bean, 1979, p. 156). Another key distinction to previous models by Tinto (1975) and other authors was the model proposed by Bean could be tested using path analysis enabling casual analysis to validate the relationships between variables.

The variables used to develop the model presented in Figure 2.2 are categorised into background variables, organisational determinants, intervening variables and the dependent variable. Like Tinto's integration model, this captured the longitudinal nature of university study. The background variables express student attributes and past performance prior to entering university. The student attrition model showed background variables affect components of the organisational variables once they enter university. For example, a student's past performance and socio-economic status will positively affect their grade point average grade point average (GPA) at university. The organisational variables impact on students' level of satisfaction with university and the course they are undertaking. This was the key aspect of the model taken from the employee turnover model. The students' level of satisfaction was positively correlated to their level of institutional commitment and this ultimately influenced a student's decision to drop out of university.

The third model of importance was Astin's (1984) interaction theory of student involvement developed from a behavioural sciences background. The model focuses on student developmental theory and involvement for higher education; it provides an important perspective which aided research in student retention theory. Astin (1984, p.523) noted that "the theory of student involvement has its roots in longitudinal study of college dropouts". Astin defines the term 'involvement' as "the quantity and quality of the physical and psychological energy that students invest in the college experience" (Astin, 1984, p. 528).

Astin reviewed past theories of pedagogy used to explain student retention, which are subject-matter theory, resource theory, and individualised (eclectic) theory. In reviewing these theories, Astin contended there are flaws in their application in treating students like a "black box" process. Relating this to the student retention theory, Astin asserted that "the theory of involvement … provides a conceptual substitute for the black box that is implicit in the three traditional pedagogical theories" (Astin, 1984, p. 521).

Variables used in Astin’s theory include place of residence, academic involvement, honours programs, student-faculty interaction, athletic involvement and involvement in student governance. Astin hypothesised that students who live on-campus are able to “achieve in such extracurricular areas as leadership and athletics and to express satisfaction with their undergraduate experience, particularly in the areas of student friendship, faculty-student relations, institutional reputation, and social life” (Astin, 1984 p. 525). In the modern university setting, this has implications for differences between on-campus and off-campus online students.

Another variable from Astin’s theory was academic involvement, which is similar to the variable used by both Tinto and Bean of academic integration. Overall, a positive relationship should exist between academic involvement and student retention as suggested by previous models. However Astin also asserted “intense academic involvement tends to retard those changes in personality and behaviour that normally result from college attendance” (Astin, 1984, p. 525). With respect to honours programs, intellectual self-esteem and interpersonal self-esteem derived from undertaking these programs had a significant impact on the students’ ability to persist. However, a potential negative relationship exists with honours affecting a student’s ability to persist, where they feel isolated from their peers (Astin, 1984 p. 525).

Astin concluded that “all college personnel – councillors and student personnel workers as well as faculty and administrators – can assess their own activities in terms of their success in encouraging students to become more involved in the college experience”(Astin, 1984, p. 529). This is in line with conclusions of both Tinto (1975) and Bean (1984) that institutional variables and involvement are critical in student retention outcomes.

The models by Tinto (1975), Bean (1980) and Astin (1984) referred to “traditional” student cohorts: internal undergraduate students under the age of 25, which represented the majority of students enrolled in university prior to the 1990s. Bean and Metzner (1985) recognised the need to account for students who had “non-traditional” study environs. A non-traditional student was defined as:

> older than 24, or does not live in a campus resident (e.g., is a commuter), or is a part-time student, or some combination of these three factors; is not greatly influenced by the social environment of the institution; and is chiefly concerned with the institutions academic offerings (especially courses, certification, and degrees) (Bean and Metzner, 1985, p. 489).

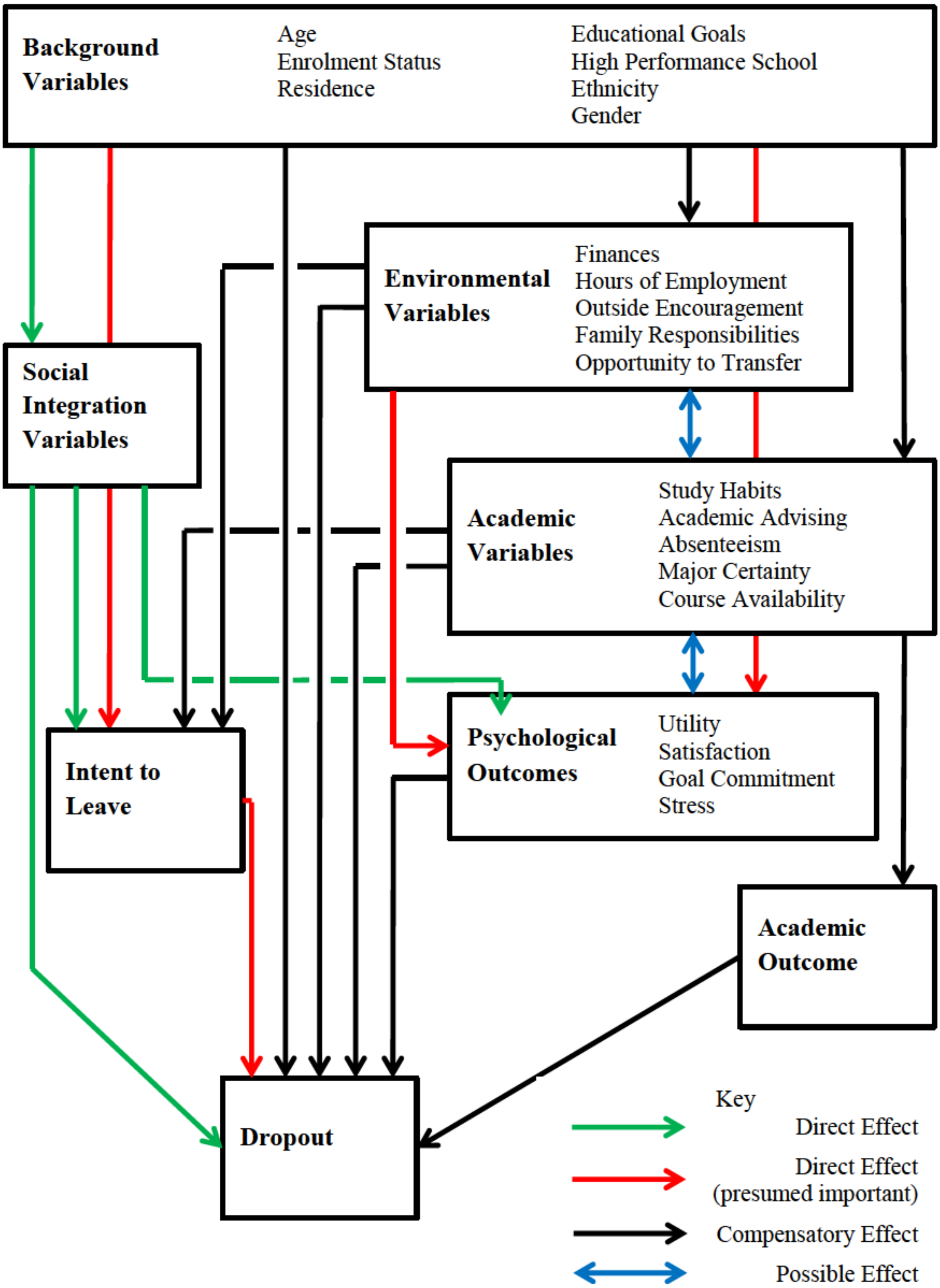


Figure 2.3 - Bean and Metzner's conceptual model of non-traditional student attrition

(Adapted from Bean and Metzner, 1985, p. 491)

The non-traditional conceptual model was developed with the knowledge that the social integration aspects of on-campus study used in past models (Tinto, 1975), (Bean, 1980), (Pascarella and Terenzini, 1980) was not relevant to "non-traditional" students. The model developed by Bean and Metzner (1985) is presented in Figure 2.3.

The model contained four main categories of variables; background/defining variables, academic variables, environmental variables and psychological outcomes. The model incorporated several features from Tinto's (1975) model and Bean's (1980) model of traditional student attrition. It is important to note while social integration appeared in the model, "social integration variables should have only minimal effects on retention, partly due to the way non-traditional students were defined and partly because social variables from the outside environment are expected to be of greater importance than college social integration variables" (Bean and Metzner, 1985, p. 530).

The inclusion of psychological outcomes gathered key terms from previous models including goal commitment, which was a fundamental aspect of Tinto's integration model. Bean and Metzners (1985) model included stress as a determining variable, which added to the scope in which psychological elements effected a student's decision to drop out. The term utility was transferred over from Bean's (1980) model where it was referred to the practical value the student perceives attaining a degree will give them.

The four theories formed an important basis for discussion on student retention. Summarising the contribution of the theories, the models incorporated several common themes, none more important than the role of institutional and social involvement in student outcomes. This changed the discourse on student retention to acknowledge that universities had the capacity to influence and affect student outcomes. The theories laid the foundations for institutions to take proactive steps to determine how to improve student retention. These theories were the precursors to the current body of literature developed in learning analytics, in particular with respects to Early Alert Systems that aim to affect retention. The models provide important arguments on what factors are likely to affect student retention. As such, the theories should guide some of the selection of variables that should be included in an Early Alert System, including the demographic, institutional, student performance and workload measures. Additional variables should also be sourced from within the learning environment to ensure that the EAS can identify students in real time based on their learning.

2.3 Empirical analysis of retention

The foundation studies in student retention theory have produced numerous articles comparing and contrasting theories, empirically testing for statistical rigour to assist in developing a deeper understanding of student retention. Cabrera, Nora, Castaneda, and Hengstler (1992) tested Tinto's integration model and Bean's student attrition model for convergence using a three stage methodology. This included: 1) assessment of the measurement properties for each model, 2) assessment of the predictive validity of each model independently and 3) a test for convergence (Cabrera et al., 1992, p. 146). In comparing the two models, it was noted that:

> both models regard persistence as the result of a complex set of interactions over time. The two models also argue that precollege characteristics affect how well students subsequently adjust to their institution. Further, the two models argue that persistence is affected by the successful match between the student and the institution. A close examination of the two theories, for instance, reveals that what the Student Integration Model refers to as Institutional Commitment, the Student Attrition Model identifies as Institutional Fit (Cabrera et al., 1992, p. 145).

A key conclusion was that Tinto's student integration model provided a more robust theory, however Bean's student attrition model accounted for more variance in student intent to persist and persistence. It was found that "the results of this study suggest that both the Student Integration Model and the Student Attrition Model add relevant knowledge to the understanding of the college persistence process, but that a model integrating the leading factors in each theory may contribute to explain this process better" (Cabrera et al., 1992, p. 160). As such, a unified theory was suggested as the future direction of student retention theory and literature which leads to Milem and Berger's (1997) study.

Milem and Berger (1997) analysed the overlap that existed between Astin's theory of involvement and Tinto's integration model. A key finding of Milem and Berger was that while Tinto and Astin asserted that student involvement in extracurricular activities had a positive relationship to a student's level of engagement and involvement, "going 'overboard' with involvement in this area has a detrimental effect on students" (Milem and Berger, 1997, p. 398). This is similar to Astin's (1984, p. 525) comment on excessive academic activity; diminishing marginal returns may be present in the positive impact of involvement and student retention. Comparing the levels of social integration to academic integration as determinants of student persistence, Milem and Berger (1997) showed that social integration had greater influence on persistence. Both Milem and Berger (1997) and Cabrera, Nora, Castaneda, and Hengstler (1992) show that there is significant overlap in the foundation theories of Tinto (1975), Bean (1980) and Astin (1984), complementing each other. However “in the world of action, what matters are not our theories per se, but how they help institutions address pressing practical issues of persistence. Unfortunately, current theories of student leaving are not well-suited to that task” (Tinto, 2006, p. 6). This conclusion is echoed by other empirical studies in that “despite its popularity, Tinto’s theory has only modest empirical support” (Kuh, Kinzie, Buckley, Bridges, and Hayek, 2007, p. 14). Many empirical models have tested elements from both Tinto’s student integration model, Bean’s student attrition model and Astin’s student interaction theory. Studies investigating the determinants of student retention have used a variety of methods to estimate the nature of the relationship between determinants, but no one theory has been supported more than the other competing theories possibly due to the diverse nature of institutions and the students attending the institution.

Empirical studies using probit or logit multiple regression models included Lin, Yu and Chen (2012), Jones-White, Radcliffe, Huesman Jr., and Kellogg (2010), Singell and Waddell (2010), while DesJardins, Ahlburg, and McCall (2002), Ishitani and Desjardins (2003), Ishitani (2003) utilise survival analysis to capture temporal effects in student retention. Stratton, O’Toole, and Wetzel (2007) used a two stage sequential decision model that tested “the impact of personal, household, academic, institutional, and economic factors have on long term dropout behaviour separately for those students who enrolled full-time and for those who enrolled part-time” (Stratton et al., 2007, p. 47). With a wide variety of methods testing factors covered in the theoretical literature, it is difficult to conclude that any one approach is superior, that any of the theories proposed are robustly valid. The diverse range of studies and sheer volume of student retention analyses make it difficult to conduct a comprehensive review of empirical methods. Nevertheless, what can be concluded is that both the theoretical and early empirical studies laid the foundation for the learning analytics field to evolve. The increased adoption of technologies to mediate activities in the education space enables access to large sets of data sources about student learning. The analysis of such digital traces can provide deep insights into numerous aspects related to teaching quality and student learning experience. The area of learning analytics has recently emerged to study the storage, analysis, visualisation and actions derived from data mining procedure applied to these data sources. Given the significant social, technological and systematic changes that have occurred in higher education since the retention theories were developed, these constructs are becoming less relevant in understanding student retention in the modern age and highlights that there is scope for new theories of student retention that can be developed to help understand student retention in the information age.

2.4 Learning analytics and retention

The 21st century opened with advances in information technology enabling significant increases in both the quantity and complexity of data available for analysis. The broad term ‘analytics’ captures “the discovery and communication of meaningful patterns of data” (Jayaprakash, Moody, Lauría, Regan, and Baron, 2014, p. 8). Patterns that existed within data enabled managers in a wide variety of fields to make informed decisions and enhance outcomes. Analytics in various forms has proved to be a powerful and now necessary tool for many industries. Generally, “analytics marries large data sets, statistical techniques and predictive modelling” (Campbell, DeBlois, and Oblinger, 2007, p. 42), with the aim “to assist scientists, researchers, and academics to make sense of the connective structures that underpin their field of knowledge” (Siemens, 2013, p. 1381). The higher education sector is one area where analytic techniques are becoming common place. Universities are integrating information systems to capture data to understand the dynamic learning environment.

Learning analytics is defined as “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (Society of Learning Analytics Research, 2012). Under the learning analytics banner, there are multiple problem domains which exist which have interested researchers. This includes areas such as intervention strategies, social learning analytics, curriculum analysis, text analytics and relevant to this study, student retention. The learning analytics field is maturing to include all aspects of the learning process and how it can be enhanced. The breadth of research topics the definition encompasses “includes techniques such as predictive modelling, building learner profiles, personalized and adaptive learning, optimizing learner success, early interventions, social network analysis, concept analysis, and sentiment analysis” (Siemens, 2012, p. 5).

From an empirical perspective, learning analytics can be viewed as a process that shows multidimensional aspects of learning. Figure 2.4 over page shows the learning process as a part of both the institution and the enrolled course. The simplified model relates the areas of study that fall under the learning analytics, presented in the green boxed areas. This is not exhaustive of all topics covered in learning analytics research, but rather relates some of the relevant areas which are researched under the learning analytics banner. A key aspect of the diagram is the institution is defined as both an academic and social space. This incorporates an important concept from student retention theory, where social integration is just as important as academic integration

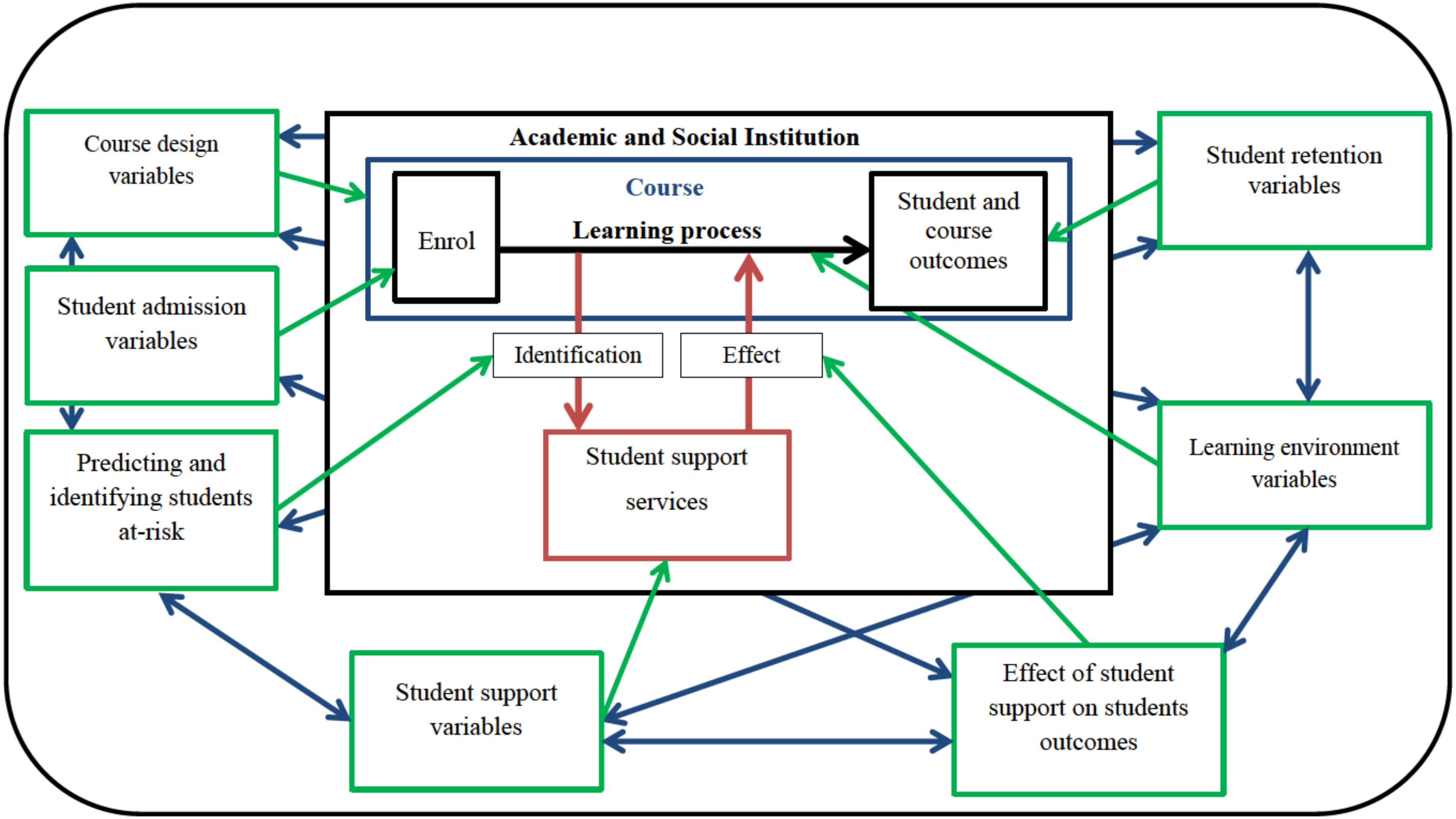


Figure 2.4 – Learning analytics and student retention

(Source: Author's own conceptualisation)

The student support services exist within this space, showing that both the identification and effects of support can be academic and social in nature. The diagram also indicates that observations of students can produce variables in many areas. For example, the course design variables may include the type of course taught, the number of assessments required to be completed within a course, or the budget allocated to funding a course. Importantly, areas within the learning analytics field are interconnected. Course design variables will affect learning environment variables, which in turn can affect student retention and early alert variables. Finally, the model shows that support services should affect the learning process directly as a means of improving student outcomes.

The blue linkage indicates that learning analytics not only focuses on the learning process and enhancing student outcomes, but also includes academic and social factors as fundamental to this process of enhancement and the interconnected nature of learning.

2.4.1 Retention studies in learning analytics

With the dramatic increase in the availability of learning data, new and more complex analysis is being conducted to understand student retention. Méndez, Ochoa, and Chiluiza (2014) investigated the relationship between student retention and the curriculum design. This study "proposed sequence mining for the identification of enrolling paths defined by students within a curriculum that frequently lead to dropouts scenarios [and] finding critical paths to avoid dropout and to reduce the negative impact of certain courses in desertion rates is a key input for curricular design." (Méndez et al., 2014, p. 155). The results showed that a particular unit of study, for example Physics A, was the most common unit that students enrolled in prior to dropping out. The sequence approach used has allowed more detailed understanding of the relationship between course design and student retention.

Wolff, Zdrahal, Nikolov, and Pantucek (2013) used data mining techniques to predict module failure at Open University UK. The study used state vector machines and decision trees to predict drops in performance of final course outcomes. It was found that "the best predictor is based on changes in the student's own virtual learning environment activity, compared to their previous activity" (Wolff et al., 2013, p. 149). The use of forum and log event data was a major component of the research, which allowed detailed patterns of online learning behaviour to be analysed.

Predictive tools form an important section of the learning analytics research utilising detailed student level data to predict outcomes. These predictive tools can form part of an EAS. Essa and Ayad (2012) developed the student success system (S3 system), which used ensemble methods "designed to boost the predictive generalizability by blending the predictions of multiple models" (p. 159). The aim of the tool was to predict students at-risk from adverse academic outcomes. Barber and Sharkey (2012) developed a similar tool, using logistic regression to predict student outcomes. The aim of the project was to connect students and advisors, where "the advisor can help the student by pointing them to necessary resources, coaching them on time management, or even advising early withdrawal" (Barber and Sharkey, 2012, p. 259).

One of the earliest initiatives in learning analytics focusing on EAS development came from Purdue University with the development of Course Signals. Course Signals was built from the work done by Campbell (2007), with the program being launched after pilot testing in 2009. Arnold and Pistilli described Course Signals as "a student success system that allows faculties to provide meaningful feedback to students based on predictive models" (Arnold and Pistilli, 2012, p. 267). The system functions as an on-demand tool by instructors which, when activated, processes student data from several sources to evaluate students' risk of being successful. The system relays information to instructors, where "a red light indicates a high likelihood of being unsuccessful; yellow indicates a potential problem of succeeding; and a green signal demonstrates a high likelihood of succeeding in the course" (Arnold and Pistilli, 2012, p. 268). The simplicity of the system is one of the major benefits and highlights the potential of learning analytics to provide meaningful enhancement of the learning environment through improved retention.

Being one of the first EAS developed and implemented, Course Signals has garnered strong interest from the learning analytics community. Arnold and Pistilli (2012) estimated the benefits of Course Signals program improved graduation rates by 21 per cent. This study however was criticised by Caufield (2013b) for not making it clear if the number of courses were controlled for. As such, it was not possible to disaggregate the effects of "students taking more Course Signals courses because they persist, [compared to] persisting because they are taking more Signals courses" (Caufield, 2013a). The resulting discussions in this area have highlighted the need for strong quantitative analysis of the effect of EAS on retention.

The scale at which learning data is captured has allowed increased micro analysis of relationships within the learning space. EAS form the IT infrastructure by which students are identified. In a broader context, EAS only form part of the overall early alert initiatives at any given institution. Institutional differences and different administrative perspectives mean many early alert initiatives have developed at various scales, with similar objectives, using different approaches. Early alert initiatives are also being seen as a technological solution to maximising retention. The information revolution within education holds great promise, so the following section reviews some of the early alert initiatives that have been implemented to date, looking at how institutions are going about realising the promise of EAS.

2.4.2 Early alert initiatives

With increased awareness of learning analytics as a field, there is developing interest in using Early Alert Systems as part of maximising student retention. The aim of this section is to review what different initiatives consider an EAS to be, what are the motivations and objectives of implementing an EAS as part of early alert initiative, and what are the salient features of EAS developed as part of early alert initiatives. This section does not review all early alert initiatives, but does aim to cover many of those institutions which first implemented early alert systems. The early movers in this area provide a solid understanding of the different objectives and methods used to implement early alert systems.

A survey of institutions implementing EAS was done as part of a doctoral dissertation by Simons (2011), who obtained survey responses from 529 institutions in the USA. In summarising institutions implementing early alert initiatives, “programs are often designed for specific populations or found on-campuses with small populations” (Simons, 2011, p. 33). The results “established the fact that early alert program implementation is considerably new in terms of higher education retention initiatives” (Simons, 2011, p. 104). More recent research indicated that “the number of initiatives that have been able to transition from concept to implementation is still scarce” (Jayaprakash et al., 2014, p. 11).

The importance of data

A characteristic of early alert initiatives is that data is fundamental to success of the project. A policy brief prepared by Siemens, Dawson, and Lynch (2013) for the Australian government on learning analytics reviewed how initiatives can improve quality and productivity in the higher education sector. From reviewing ten case study institutions, it was concluded that several issues inhibit the development and implementation of learning analytics. These include issues around data access where data silos limit analytic projects. The review suggested a more open and collaborative approach to data usage needs to be developed to allow for detailed analysis of the learning environment. This needs to occur while also satisfying the ethical issues around data sharing and utilisation (Siemens et al., 2013, p. 22). It can be concluded that for early alert initiatives to succeed, data needs to be available which captures the breadth of the learning environment.

With the push for data to become available, important conversations have arisen around data ownership and the rights to utilise student data. Jones, Thomson, and Arnold (2014) discuss the student and institutional perspectives on ownership, concluding “students have strong grounds for claiming ownership of their identifiable data, especially when it can be or already is used to influence their academic, professional and personal lives” (Jones et al., 2014, p. 5). From the institutions’ perspective, the data generated by students in the learning environment has become central to the provision of basic services. Institutions have a need to maintain ownership of data just to function even in the most basic capacity. It was proposed that “a shared ownership model would support the institutions’ data needs, protect students’ privacy, and inform individual students about personally identifiable data use on-campus and what rights they have to it” (Jones et al., 2014, p. 6). This means that for early alert initiatives, important features are either informed consent from students to participate in the initiative, or the capacity to opt out of programs at any stage.

Data collected also needs to be a reliable and meaningful measure of the learning environment. Traditionally, student support services and academic advisors rely on information that is either outdated or unrepresentative of the student’s true need for support. This was the motivation behind the development of Student Explorer at the University of Michigan, where previously:

> Advisors relied on students’ self-reported grades that students brought to monthly meetings. According to advisors, the monthly meeting schedule did not provide frequent enough interactions between students and advisors. For example, once a student had failed an exam or assignment it was often too late to correct a student’s academic trajectory (Krumm, Waddington, Teasley, and Lonn, 2014, p. 106).

The data contained within the Learning Management System (LMS) presented a more reliable source of information which included students’ grades and login data (Krumm et al., 2014, p. 107). This underscores the importance of EAS using reliable sources of data to drive interventions. Overall, data availability, ownership and meaningfulness form the fundamental basis of all Early Alert Systems.

The types of data

As indicated, data is fundamental for early alert initiatives. The main source of data for analysis has come from LMSs which capture important characteristics of the students' learning outcomes, including grade and assessment information. LMS data provides a rich source of information from which early alert initiatives can build. However, the learning analytics community has developed more novel approaches to analysing students learning to identify students at risk.

For example, student engagement in the learning environment is of keen interest, but difficult to measure. To overcome this limitation, Baker, D'Mello, Rodrigo, and Graesser (2010) developed a protocol to observe students directly in the learning environment. While the students used intuitive tutorial tools, observers coded students' engagement with the environment. Blending qualitative and quantitative approaches, the data yielded from this approach when coupled with log files capturing students usage allowed detailed understanding of student disengagement within the learning environment.

Romero-Zaldivar, Pardo, Burgos, and Kloos (2012) used a virtual machine approach to create an artificial learning environment tool. The virtual machine was an "appliance fully configured for the course activities, [where] students have a self-contained environment, yet highly versatile (it is a fully equipped computer), to work in the course" (Romero-Zaldivar et al., 2012, p. 3). One of the major benefits to teaching using the virtual computer approach was the capacity to monitor student behaviour using the virtual machine. Capturing data in this way allows more detailed data on student behaviour than is possible using just the LMS.

Baker, Lindrum, Lindrum, and Perkowski (2015) investigated early at-risk factors of students in the online environment. The online environment Webtexts combine a "mix of original, permissioned, and open content, combining text, images, audio, video, hosted and linked artefacts, and tools for study" (Baker et al., 2015, p. 2). Analysing student interactions it was possible to "specifically identify that a student is at risk because he/she has failed to access the resources, or because he/she failed to complete the assignments on time, or because he/she has scored poorly on the assignments" (Baker et al., 2015, p. 6). The demonstrated capacity to use data generated outside of the LMS indicates the scope for early alert initiatives to incorporate data from a wide range of sources within the institution. One of the major issues for the learning analytics community is to determine from the volumes of data types available for analysis, which types of data are best when designing an early alert initiative.

Early alert systems that reached implementation

While many institutions have expressed interest in developing EAS, very few have reached the stages of implementing the systems. As mentioned previously, Course Signals at Purdue University was one of the first EAS. The system "detects early warning signs and provides interventions to students who may not be performing to the best of their abilities before they reach a critical point" ("Course Signals," 2013). The program was enabled for 500 'entry courses', which are seen as gateway units to further units in the students chosen course. For

> the students the red, yellow and green Signals message is intuitive and simple, but behind the uncomplicated interface lies a sophisticated data mining and analytics algorithm that checks more than 20 data points, focusing more on the student's effort than just their grades (Tally, 2009).

The program implements many aspects of the underlying student retention theory, where "Signals combines demographic information with online engagement" (Straumsheim, 2013) allowing the program to factor in some measures of academic integration.

The Open Academic Analytics Initiative (OAAI) initiated at Marist College was an attempt to develop an open source EAS, assessing the portability of such a system between different institutions. The first objective of the program was to "identify challenges, solutions and benefits associated with developing a completely open source early alert solution" (Jayaprakash et al., 2014, p. 7). The emphasis on open source system solutions diverged from Course Signals which used a proprietary algorithm. The EAS designed draws on data from four sources:

> a) student demographic and aptitude data; b) course grades and course related data; c) Sakai-generated data on student interaction with the learning management system; d) partial contributions to the student's final grade collected by Sakai's gradebook tool (ie., student grades on specific grading events, such as assignments and exams) (Jayaprakash et al., 2014, p. 14)

The initiative also tested two possible treatment outcomes; one being alerts sent to students, the other participating in an online academic support environment. One of the successes of the program was being able to test the portability of the EAS on data from other institutions. The results showed a "positive impact on the effectiveness of interventions on students at academic risk" (Jayaprakash et al., 2014, p. 41). A key difference to Course Signals is the scale of the system design, which included all students enrolled at all levels within the institution, not just entry courses.

In reviewing student engagement best practices in Australia and New Zealand, Nelson and Creagh (2013) covered eight case study institutions which implemented early alert initiatives in various forms. All of the programs listed one of three objectives of their respective initiatives:

1) Improved student retention
2) Reduced student attrition
3) Supporting students at-risk of disengaging.

Those indicated institutions saw early alert initiatives as an appropriate method for addressing student retention and engagement issues. A common feature of the initiatives was proactive contact with students when identified as being at-risk of disengaging, discontinuing or failing a unit. Contact was initiated via email alerts or telephone calls, which outlined a range of support options for the student. The data used to classify students at risk was similar across institutions, with information from LMSs and data warehouses being the major sources. The types of data consisted of current student enrolment information, background demographic variables and academic performance. Despite the common objectives of the programs, how these programs were implemented by institutions varied. For example, Curtin University's JumpSTART program "targets first year, first semester undergraduate students in selected first year units" (Nelson and Creagh, 2013, p. 58), while the University of New England's (UNE) Early Alert Program was implemented to capture all students regardless of year of enrolment or school. The varying scope indicates different levels of institutional acceptance of the initiatives which is a significant issue for the adoption of learning analytics in general.

Early alert initiatives can be summarised as a proactive evidence based approach to connecting students to student support in a timely manner. To be evidence-based, institutions needed to make reliable student data available for the implementation of the programs. To be proactive, the early alert initiative needed to provide actionable data to advisors and support staff. To connect students to support, students needed to be contacted directly with options to assist them in making informed decisions about their study and support options. To implement an early alert initiative does not come without issue, it "requires strong leadership and awareness to instil a coherent vision and strategy and to navigate the complexities and resistance to change" (Siemens et al., 2013, p. 29). To truly understand the benefits of implementing an EAS initiative however, an evaluation is required which is covered in section 2.4.4.

2.4.3 Early Alert System at UNE

The University of New England is used as the case study institution for this study. UNE's Early Alert Program is "the foundation of student engagement and retention activities at UNE. Early Alert uses contemporary technology to collate data and identify students who may be at risk of disengagement and attrition" (Nelson and Creagh, 2013, pp. 85-86). Custom designed software called the Automated Wellness Engine (AWE) is a type of EAS. The EAS used student level information from the data warehouse to analyse, flag and report students deemed to be at risk of disengaging. The data warehouse itself collects and stores data from eight IT platforms from around the institution to aid university management and decision making. This includes unit monitoring reports on the teaching of various units of study, course level information and university workforce data. The EAS used 34 triggers to identify these students, with each trigger assigned a positive or negative weight, summated to give a final score for a student (see Appendix C). Each day, the 200 students with the highest negative scores were sent an initial email outlining support options and if they choose to, could contact the student support team to opt-in to a tailored support program (Leece, 2013).

Nelson and Creagh (2013) outlined challenges for the Early Alert Program at UNE. This included the "acceptance of social media as a legitimate platform for student learning engagement, support for centralised approach to identification of student need, ability to create a sense of community for distributed learnings, acceptance of the role of data intelligence to drive student support activities" (Nelson and Creagh, 2013, p. 87). The challenge around the acceptance of data intelligence was an important observation that corresponds to one of the motivations for this study.

The early alert initiatives reviewed showed common objectives in trying to improve student retention. A common link between the systems was the integration of the Learning Management System with other information systems within the institution. The variables used by the systems varies between the institutions, but can be generally surmised into three main categories of variables. The first group was demographic variables, which capture the students' background factors that are independent of their enrolment within the institution. The second group was institutional variables, which capture institutional level data relevant to the students' learning. This included information on the course undertaken, tuition fees charged and the type of study the student is undertaking, being on-campus or off-campus online. The third group of variables was the learning environment variables which corresponded to data points taken from within the learning process. This included assessment information taken throughout a unit of study.

While many programs have been developed, the next phase in early alert initiatives is to evaluate the effectiveness of the systems. Evaluation of any project is an important final step to establish the usefulness of the project. In the case of early alert initiatives, it is important to have solid evaluation of efficacy if the early alert initiatives are to be adopted globally.

2.4.4 Early Alert System evaluation

There has been significant focus on the design and implimentation of early alert systems through various initiaitives. Generally, project management lifecycles indicate that after implementation, evaluation of the project should follow. This was an important component of the evaluation model developed by Tynan and Buckingham Shum (2013) on how learning analytics can drive student success. The model was implemented as part of the Open University (OU) strategic analytics investment program (Ferguson et al., 2015, p. 133) and presented in Figure 2.5 over page.

Evaluating becomes a reflective process used to analyse whether the project achieved its initial objectives. Where objectives are not met, the evaluation process helps drives the change required to improve system design. As part of the program,

> interventions are evaluated and then become the evidence base for factors that drive student success. For example, as part of the universities quality assurance process, a 'module pass rate model' is used to compare actual module pass rates with those expected based on a statistical analysis of the previous achievement of students over the preceding five years. Use of the model has given the university an improved understanding of the characteristics and behaviours of students who are more likely to struggle with their studies (Ferguson et al., 2015, p. 135)

It is important to establish what can be learnt from the process of designing and implimenting an EAS, and to measure the size of the effect it has on student retention. In the case of EAS, "it will be important to understand the 'value added' of learning analytics over what an instructor is capable of doing on his or her own" (Jayaprakash et al., 2014, p. 42).

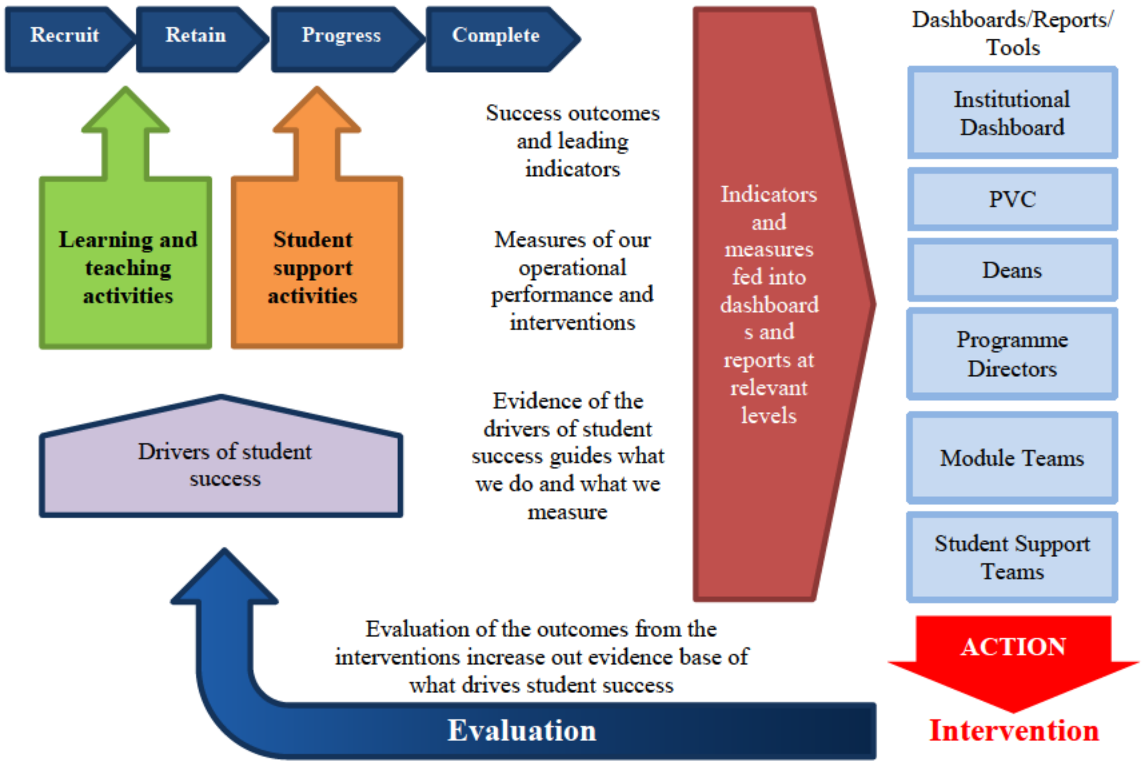


Figure 2.5 – Evaluation cycle of the Open University strategic analytic model

(Adapted from: Tynan and Buckingham Shum, 2013, slide 21)

Several of the early alert initiatives reviewed have reported improvements in retention. In the pilot study conducted at the University of Sydney, it was seen the project had "moderate success, particularly among the Business participants, whose failure rate was 50 per cent below that of the BUSS1001 cohort" (Khamis and Kiernan, 2013, p. 4). In the case of RMIT's student success program, it was reported the program increased student retention from 57.8 per cent to 64.8 per cent (an extra 40 students). In turn, it was estimated by RMIT to bring in an additional $1,135,000 of revenue after the costs of program administration are taken into account (Nelson and Creagh, 2013, p. 80). QUT reported similar success, finding "estimated retained income through the retention of an additional 227 students is $3,745,000 for every remaining year of their enrolment" (Nelson and Creagh, 2013, p. 72). These estimates show that early alert initiatives have the potential to be prudent investments for institutions implementing EAS.

As highlighted previously, claims about the success of Course Signals at Purdue have been met with scepticism. The online exchange highlighted one of the major issues in evaluating EAS. Over-simplified measures of percentage changes in retention rates do not accurately reflect the true benefit of the systems. In the case of QUT, it was reported that "in three out of five units (units 1, 2 and 4), the SSP intervention with at-risk students had a statistically significant impact upon their achievement" (Marrington, Nelson, and Clarke, 2010, p. 2). This indicates that for units 3 and 5, there was no significant improvement in student outcomes. This raises questions over the validity of conclusions drawn on the program's effectiveness in the general context, especially if institutions are seeking to implement initiatives on a large scale. Evaluating EAS is important to justifying the expenditure of university resources on such programs.

The issues highlighted here show that the literature on evaluating EAS has only just started developing. This is going to be an increasingly important area of research as organisations become more skilled with learning analytical tools and strategies. This study contributes to the field by providing a detailed evaluation of an EAS using microeconometric approaches.

2.5 Chapter summary

Student retention theory from the 1970s and 1980s has been frequently referenced and used as a basis for student retention analysis, along with the subsequent empirical models which tested the validity of the theories developed. While now outdated due to systematic and social shifts over time, the models have helped guide the selection of variables that should be used to develop EAS. This includes the background demographic information, institutional aspects of students' enrolment and measures of student performance and workload. Retention theory also assisted in forming expectations of the types of patterns that should be observed in data, now that more detailed information on the learning environment is available.

There are a range of early alert initiatives that have now been implemented at a number of institutions. These programs all use learning analytics as a means of understanding complex patterns within student data to improve student outcomes. The data used for analysis blends demographic, institutional and learning environment variables to identify students at risk of adverse academic outcomes. One caveat however is there are limited studies which currently evaluate effectiveness. Of those studies, there were no statistically rigorous approaches used to determine the effect the EAS have on student retention. This study hopes to begin addressing this gap in the research by using microeconometric approaches to estimating the effect of the EAS on both retention and revenue.

This study is framed in the cross section between retention analysis and EAS evaluation. The first research objective is to investigate, analyse and understand the demographic, institutional, student performance, workload and temporal variables affecting undergraduate student retention rates. Using detailed data from the case study university, this study will contribute to research in student retention by providing empirical analysis of the factors affecting retention in a temporal framework with an unprecedented level of detail. The second objective is to analyse in a quantitative framework the relationship of EAS on the undergraduate student retention rate. This will be the first published study to evaluate the effect of an EAS on student outcomes using rigorous statistical modelling. The final research objective is to quantify the effect of an EAS using a financial instrument. Typically, the effect of the EAS has been measured in terms of retained student tuition fees. While this has been done in a few previous studies, measuring the effect of the EAS on retention rates needs to be analysed with appropriate methods. The use of microeconometric modelling with detailed data allows such an analysis to be performed. The outline of the method of analysis is presented in the succeeding chapter, Chapter 3.

Chapter 3 - Methods of Analysis

3.1 Introduction

When it comes to quantitative analysis, the approaches used to analyse the information matters. Different methods allow discovery of different effects and patterns within the data. Choosing which method is appropriate for the objectives of the research needs justification, accompanied by the underlying assumptions and statistical limitations of the method. This chapter reviews previous methods used in student retention analysis, summarising the significant conclusions found. The review guides the choice of methods and models used for this study. The chapter concludes with hypotheses that test the research questions with models calibrated to include variable specifications for each model.

3.2 Review of previous methods

A wide variety of approaches have been utilised in analysing student retention. Some of the qualitative approaches used to understand student retention include survey and student feedback, case study, and focus group approaches. The qualitative approaches can yield important information on students thoughts, opinions and feelings on the factors that affect their decision to stay at university. However, qualitative approaches have limitations: it can be difficult to construct testable hypotheses; it can be time consuming to collect and attain all the information necessary for research; the identified effects may not be easily generalised to the wider population.

Quantitative analysis can address many of these limitations. Within quantitative analysis, there are many approaches to understanding student retention. These include: descriptive analysis, focused on the salient features of data to develop an understanding of the status quo; predictive analysis attempts to predict future student outcomes based on previous patterns and information; time-series analysis focuses on how effects change over time; causal analysis which attempts to establish the causal relationship between different variables. To implement these various approaches, a range of statistical tools have been developed. These include: Ordinary Least Squares (OLS), structural equation modelling (path analysis), probit, logit, sequential logit, multinomial logit, survival analysis (hazard modelling) and sequence analysis. The strength of statistical inferences from these models varies based on the underlying assumptions and estimation approaches used by the models. This section reviews previous studies and methods used to understand the student retention.

3.2.1 Quantitative methods

Ordinary Least Squares and Path analysis approaches

Ordinary least squares and path analysis were the methods used by Bean (1980) to test the validity of his theoretical model of student dropout behaviour. The objective of Bean's paper was to establish a testable theoretical model that can establish causality between factors affecting student retention. The main problem with "previous models [prior to Bean's study] of student attrition lies in the fact that the definition of variables used in the analysis rendered the models unsuitable for path analysis" (Bean, 1980, p. 156). Path analysis formed an important methodology for establishing the causes of student attrition.

Bean found some empirical support for his theoretical model using this methodology. Estimating separate models stratified on gender, Bean found that for female students, factors of institutional commitment, performance, campus organisation involvement, the practical value of the degree and the opportunity to transfer impacted on a student's decision to drop out. For male students, institutional commitment, university grade point average, satisfaction, development and level of study routine developed by the student all impacted on a student's decision to drop out. A limitation in this method of analysis is that intervening variables are both dependent and independent variables within the model specification. This inhibits meaningful interpretations of the intervening variables. One way around this limitation was to use probability-based models that allowed the relationships between variables to be better understood without the limitations of path analysis.

Limited dependent variable approaches

The first of two major probability based methods of analysis used in student retention literature is probit modelling. Probit modelling was recently used by Singell and Waddell (2010) in a longitudinal study trying to determine if students at-risk of dropping out can be identified early enough to be treated. Probit modelling is an extension from multiple regression analysis allowing for the treatment of a categorical binary dependent variable. This allows the estimation of the relationship between the dependent variable (dropout) with independent variables in terms of probability. Singell and Waddell (2010) draw on theoretical bases of both Tinto (1975) and Bean (1980) to develop the general probit model used in analysis. Significant variables in the model include background and demographic variables (gender, high school, SAT score, race, age), institutional variables (financial aid, first year interest group involvement) and student performance in the first year. Additionally, the model used GPA as a time variable with the results of first, second and third term GPA being used. Singell and Waddell (2010, p. 569) "find that at-risk students can be identified using accessible statistical models and information available at the time a student enrols and that observed performance in college improves the model's ability to predict retention". The conclusions of this model are significant in that the study used teaching period and annual time variables. This has important implications for the development of any time-dependent model attempting to identify students at risk in a meaningful period of time whereby intervention strategies can be used to assist the student.

Logistic (Logit) regression is an alternate method to probit modelling and is a special class of regression analysis. Comparing logit and probit models, the two approaches differ on the underlying distribution of the cumulative probability function. However, logit models have become the preferred approach of modelling student retention rates compared to probit modelling. In a systematic review of literature on logistic regression models used in higher education research, it was noted that "the trend in higher education is for researchers to recognise limitations of Ordinary Least Squares (OLS) regression and turn increasingly to logistic regression for explaining relationships between a categorical outcome variable and a mixture of continuous and categorical predictors" (Peng, So, Stage, and John, 2002, p. 260). Over the period under review, significant contributions to the literature included Hinkle, Austin, and McLaughlin (1989), Austin, Yaffee, and Hinkle (1992), Dey and Astin (1993) and Cabrera (1994).

More recent studies using logistic regression in a wide variety of student retention settings include Stratton et al. (2007), Fike and Fike (2008), and Jamelske (2009). Stratton et al. (2007) investigated the relationship between initial enrolment intensity and dropout behaviour using data from 4,655 students attained from the National Centre for Education Statistics (NCES). Using 28 explanatory variables to model retention, the significant findings were that

> parental education, the timing of enrolment, college GPA, and the local unemployment rate are significantly associated with attrition for full-time students, they do not appear to be as significant for part-time students... it is not part-time enrolment per se that is correlated with attrition; rather it is underlying differences in observable factors between the two groups that leads to the correlation with attrition (Stratton et al., 2007, pp. 478-479).

In Fike and Fike (2008) the objective was to find predictors of student retention in first year community college students. A logit model was used to determine the relationship between retention rates and three explanatory variables, including variables such as the number of hours enrolled, using a sample of 9,200 first year students. A significant finding is that "the strongest predictor for retention is passing a developmental reading course" (Fike and Fike, 2008, p. 80).

Jamelske (2009) uses a logit model to determine the relationships between university First Year Experience (FYE) programs and the dependent variables of student GPA and retention. A sample of 1997 full-time students was used in combination with 20 indicators. The logit model revealed that "there was no positive effect on retention, but the GPA for FYE students was higher than non-FYE students" (Jamelske, 2009, p. 389). The conclusion with respect to retention may be at odds initially to other literature; upon focusing the sampling frame a significant relationship between FYE programs and retention was observed.

Multinomial variable approaches

An alternative model which extends the binary logit model is the multinomial logit which allows for multiple categories of dependent variables. Stratton, O'Toole, and Wetzel (2005) constructed a multinomial logit model of college stopout and dropout behaviour. Using a sample of 4,226 students to model random utilities, three categorical dependent variables were used, continuous enrolment, short-term stopout and long-term dropout. The conclusions of this model showed that given the 40 explanatory variables, the "multinomial logit indicated that there are distinct differences between those who stop out and those who drop out. This suggests that parameter estimates from a standard logit model of attrition that fails to distinguish between stopout and dropout behaviour will be biased" (Stratton et al., 2005, p. 19). This is an important finding in developing any model dealing with student retention rates provided the categories in the dependent variable satisfy the assumptions of IIA.

Jones-White, Radcliffe, Huesman Jr., and Kellogg (2010) used a variety of probability regression based methods to investigate suitable techniques for studying student graduation across institutions of higher education. A binary logit model was constructed using graduation within six years or not as the dependent variable, and multinomial logit and probit models used varying degree levels (associate-level degree, bachelors-level degree, bachelors-level degree from the entry institution, did not graduate within six years). The binary model produced similar results to previous studies done at the same institution on student retention, with key factors of academic preparation and performance measures being significant (Jones-White et al., 2010, p. 162). Both multinomial models were run to test assumptions of independence of irrelevant alternatives and no significant difference between the two models was found. A key conclusion from the study was that "One of the strengths of the multinomial logit approach is that is produces estimates that are relatively more easily interpreted than those of the multinomial probit model" (Jones-White et al., 2010, p. 166). This ease of interpretation was a key factor as to why multinomial logit is the preferred probability regression model.

Survival analysis approaches

Survival analysis (also known as hazard or event history) modelling is similar to probability regression models used in understanding the nature of student retention when temporal effects are present. Initially, survival analysis was developed as a method for analysing the survival rates of cancer patients under different treatment conditions. The key feature of survival analysis is that it can account for varying times to a failure event such as discontinuation of studies. In allowing for varying lengths of time, survival analysis has become the preferred method for analysing unbalanced data sets where censoring of data occurs. On survival analysis and its application to analysis of student retention,

> since students may depart at any given time while they are enrolled, selecting an arbitrary point in time to specify enrolment status of students in structural equation modelling [such as path analysis] fails to examine differences in departure behaviour that may exist at various times (Ishitani 2006, p. 865).

Ishitani (2003) modelled attrition behaviour among first-generation students and developed the following conceptual model (based on previous research) to describe the process of student departure, presented in Figure 3.1. In the framework, the observation period refers to the period of time where student information is gathered. Each time period *t* is a teaching period for the *i* number of students.

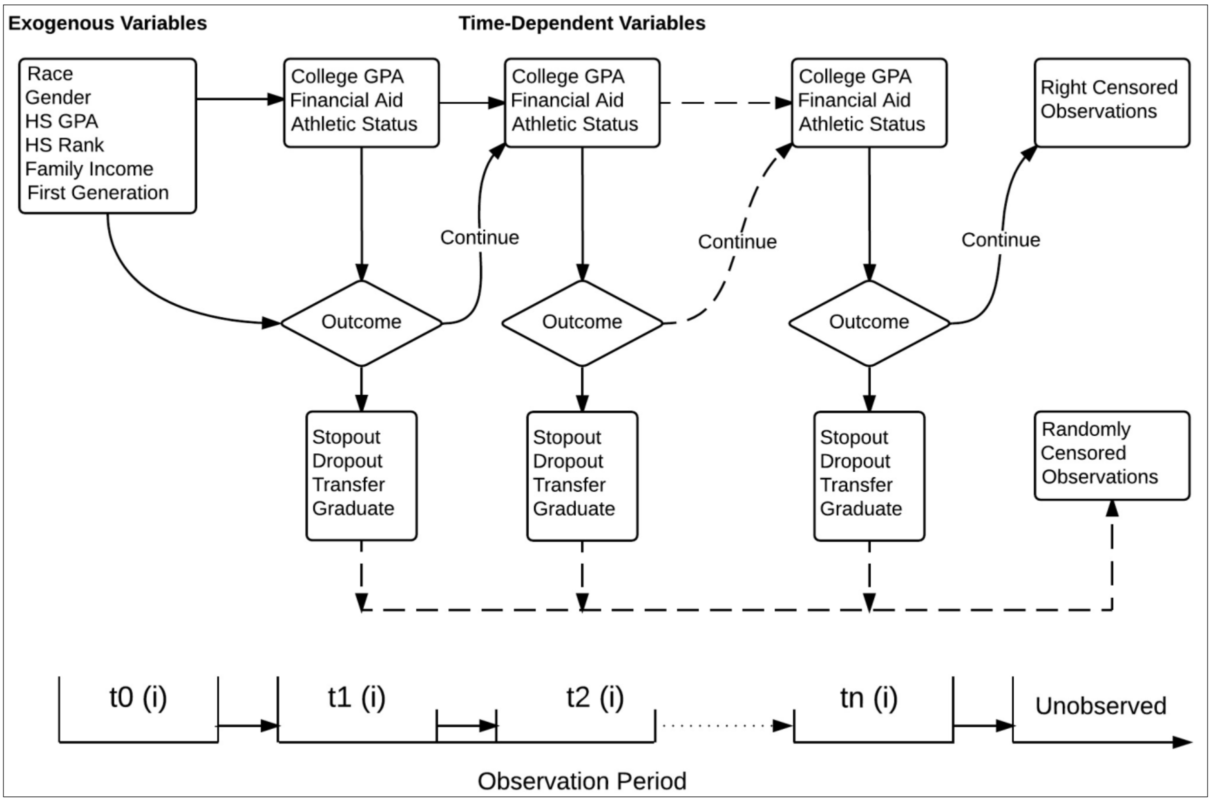


Figure 3.1 – Conceptual framework of student departure process

(Adapted from: Ishitani, 2003, p. 437)

This framework shows the variables of interest, and how there is a sequential process over time, to attaining a degree, or more importantly, students have the potential to discontinue their studies at the institution during any time period.

Ishitani (2006) continued to investigate attrition and degree completion of first generation students using the survival analysis framework in a multinomial logistic regression model. The events history model used dummy variables (1 for graduates, 0 for not graduates) to determine which students graduated in fourth, fifth and sixth year of study. Using the Kaplan-Meier method which allows graphical analysis of the hazard function, the proportion of students who were still enrolled in each year was estimated to generate Figure 3.2. This showed the differences in survival rates based on the level of parental college education.

[redacted]

Figure 3.2 – Kaplan-Meier survivor function by parent's educational attainment

(Ishitani, 2006, p. 871)

Following on from the initial survival modelling, it was concluded that:

> first generation students were exposed to higher risks of departure through college years than their counterparts were. Moreover, they were less likely to complete their degree programs in a timely manner [where] being a first generation student reduced the odds of graduating in four or five years by 51 per cent and 32 per cent [respectively] (Ishitani 2006, p. 880).

This demonstrates the importance of survival analysis in understanding temporal effects occurring in the student retention problem. Other studies using this methodology to analyse student retention related areas include DesJardins, Ahlburg, and McCall (1999), DesJardins, Ahlburg, and McCall (2002), Ishitani and DesJardins (2003) and Radcliffe, Huesman, and Kellogg (2006).

Sequence analysis approaches

A more recent method of analysis from the learning analytics field is sequence analysis. This refers to the group of models which analyse the sequence of events. This is particularly pertinent when analysing student retention, to estimate if the order of events within a student's course affect outcomes, or what is the most common event affecting retention. If the sequence of the events matter, then this can inform course design to ensure it maximises student retention. This method was recently used by Méndez et al. (2014) to analyse the computer science course, where 610 dropouts occurred out of the sample of 1,591 students. Using an algorithm originally designed by Zaki (2001), it was found that the Physics A unit was the largest stumbling block, with 61% of student failing this program out of all of students who discontinued. This has important implications for course design in determining critical moments in the students progression in a course.

A relevant statistical approach that is yet to be applied in EAS statistical method that is pertinent to student retention and the effect of EAS is treatment effects modelling. This class of causal models has been used in a few key studies, however its application to EAS is new. Schudde (2011) applied propensity score matching methods to investigate the causal effects of campus residency on college student retention. The data was pre-processed into two strata, the treatment group (on-campus students) and the control group (off-campus students). Taking this approach allowed matched estimation of outcomes using regression and logistic regression methods. The results indicated that there was a causal effect where living in college improved retention. More importantly, the study demonstrated that causal modelling is possible on student retention data.

The various methods of analysis allow different effects and strengths of inferences to be made with respect to student retention. In summary, path analysis and structural equation modelling can provide strong causal inferences, however it is harder to make intelligible interpretation of intermediate results. Logistic regression is commonly used to analyse the likelihood of binary outcomes, while multinomial logistic regression can explore situations where the outcome variable has more than two discrete states. Survival analysis analyses the time to event, estimating the Cox hazard ratio as a measure of risk associated with discontinuing. Sequence analysis can measure the sequence of events that precede discontinuation to identify common factors in students' journeys that correlate to the decision to leave. Finally, treatment effects modelling provided another causal model that allows interpretation of the average effects associated with a particular treatment option on student retention. From these methods, models are developed for this study which will allow the effects of an EAS on student retention to be captured.

3.3 Early Alert System framework and method selection

The review of the methods provides a foundation for the studies' model selection. However, selecting models needs to be done in the context of the EAS used in this study. The EAS design covered in this study comes from the University of New England, Australia. This system was briefly discussed as part of the review of literature in chapter 2. The system was originally designed for the purposes of identifying students at risk of disengaging from their studies. A more detailed description is provided to allow appropriate model section to capture EAS effects on retention.

3.3.1 Case Study EAS

The EAS and course progression framework from UNE is presented in Figure 3.3. The flowchart outlines the processes students go through from enrolment to outcomes, factoring in student support services and the EAS. Several key events can occur throughout students' enrolment. These events are denoted with X, where a student will either be identified by the EAS (X_1) or not identified by the EAS (X_2). If the student is identified by the EAS, the next critical event is being contacted by the student support team. If the student was contacted (X_{11}) then the student may opt for targeted student support (X_{111}). The default path is coloured in blue, where a student will not be identified by the EAS and will continue (Y_1) their studies. This creates the circular path that surrounds the EAS system and represents the most common path of students. The EAS identification process occurs daily, so a circuit following the blue path takes one day.

The framework captures the reality of study and the decisions faced by students. On any given day, a student can choose to not continue their studies. One outcome, completion (Y_2), is limited only to those who have satisfied the conditions to be admitted to the award of their course. The remaining outcomes have different meanings for situations students choose when stopping or dropping out of UNE. The first is inactive (Y_3) which captures students who are enrolled in the institution, but choose to not enrol in any units of study in a teaching period. This is one of two stopout scenarios, capturing the situation where students have not formally notified the university of their intention to defer studies. The second outcome is intermittent enrolment (Y_4). This is the situation where a planned stopout of studies is taken, formally applying for studies to be deferred. The two remaining outcomes (Y_5 and Y_6) relate to students stopping their studies altogether.

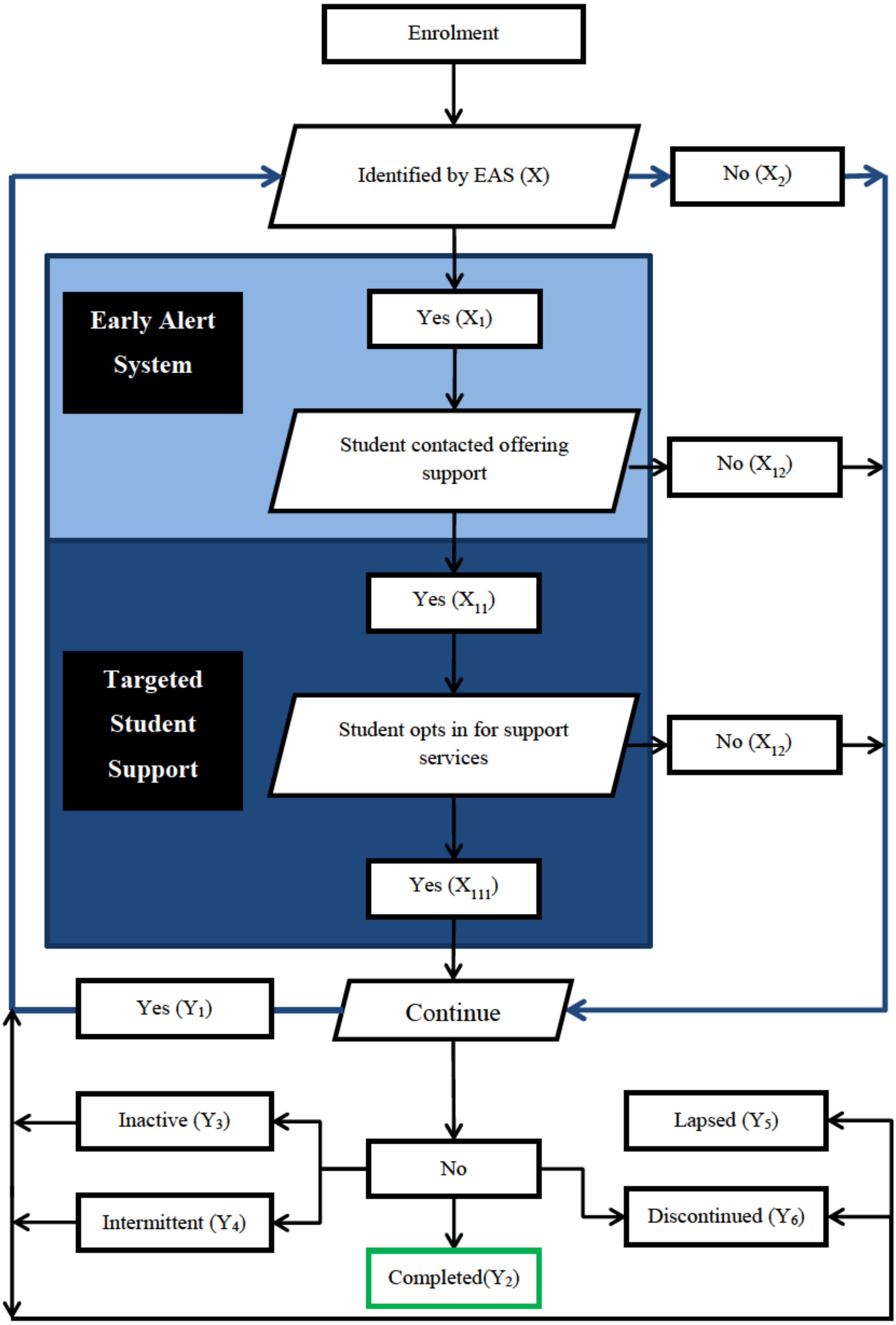


Figure 3.3 – EAS Process at UNE
(Authors own concept)

Lapsed (Y_5) refers to the situation where a student has been inactive for at least two years. As such, lapsed captures a student who has stopped their studies but has not administratively discontinued their studies. The final outcome, discontinued (Y_6) is where the proper administrative processes have been followed, and the student ceases to study their enrolled course.

The identification process of the EAS is the critical entry point to future events associated with support services. The identification process utilises 34 “triggers” reflecting data points collected on students throughout the learning process. Descriptions of the 34 triggers are supplied in Table 3.1. Some of these triggers are constant throughout enrolment. Other triggers capture information on the current or previous teaching period. The most granular triggers analyse student log data, updating daily. Each trigger carries a positive or negative weighting, which is added up each day, yielding a score which is assigned to the student. Students are then ranked on the basis of their score attained. Those students with the lowest 200 scores that day are considered to be the most at-risk of disengaging. These 200 at-risk students form the short list (X_1) which is then passed to the student support team for further action. The choice of identifying the most at-risk students is an arbitrary design parameter of the EAS. The sequence of events after identification relate to both if the student was contacted, and if so, whether the student chose to have tailored support. A key aspect of the system is the capacity for the student to opt out of support, giving students self-determination in the support process after identification.

The table of triggers in Table 3.1 show a wide variety of data elements captured by the EAS. While many of these triggers are intuitive, several are specific to tools and functions at UNE. The first is the alternative entry pathway. This is where the admission scores of the student are not taken into account, but rather the student is admitted on the recommendation of the high school principal. Students admitted this way may be at increased risk due to entering university with lower than necessary entrance scores. Another key trigger is the in-house tool called “e-Motion”. This is a set of emoticons in the student portal which appear next to each unit the student is currently enrolled in. The student can select several states to represent how they feel about their studies, including happy, neutral, I do not want to say (opt out), unhappy and very unhappy. The e-Reserve activity refers to students’ use of the online library portal as a means of accessing information relevant to studies. Finally, the teacher enabling course is a special course for students wishing to have a career in education and teaching. These courses are designed to assist students who do not currently have skills that align with the required skills sets by various education boards and institutes.

Table 3.1 – EAS Triggers

Trigger	Description
1	Student admitted through alternate entry pathway
2	Student is Aboriginal or Torres Strait Islander
3	Unit is a currently high attrition unit
4	Unit is a historically high attrition unit
5	Student is a college resident
6	Student registered "Happy" in e-Motion
7	Student registered "I do not want to say" in e-Motion
8	Student registered "Neutral" in e-Motion
9	Student registered "Unhappy" in e-Motion
10	Student registered "Very Unhappy" in e-Motion
11	Student has high e-Reserve usage inactivity (31-40 days)
12	Student has low e-Reserve usage inactivity (10-20 days)
13	Student has medium e-Reserve usage inactivity (21-30 days)
14	Student has very high e-Reserve usage inactivity (41+ days)
15	Student has been granted 1-2 assignment extension in current teaching period
16	Student has been granted more than 2 assignment extension in current teaching period
17	Student has submitted 1-2 assignments late in current teaching period
18	Student has submitted more than 2 assignments late in current teaching period
19	Student enrolment has involved > double their number of currently enrolled units in current teaching period, post start of teaching
20	Student has appeared in High Risk Category in a previous teaching period
21	Student is an international student
22	Student has no prior enrolment at UNE
23	Student is enrolled in 5 or more units in a single teaching period
24	Student was previously enrolled in a pathways enabling course
25	Student has been flagged for contact by the retention team in current teaching period
26	Student has been flagged for contact by the retention team in a previous teaching period
27	Student is carrying over Special Extension of Time (SET) exams from a previous teaching period, and is enrolled in current teaching period
28	Student has high portal usage inactivity (31-40 days)
29	Student has low portal usage inactivity (10-20 days)
30	Student has medium portal usage inactivity (21-30 days)
31	Student has very high portal usage inactivity (41+ days)
32	Student was enrolled in the Teacher Enabling course
33	Student received a fail in a unit in a prior teaching period
34	Student received a fail incomplete in a unit in a prior teaching period

(University of New England, 2014c)

3.3.2 Model selection

Having reviewed the models used for analysis in section 3.2 and discussed the EAS framework; appropriate methods of analysis need to be selected to address the research questions. Before the appropriate model can be selected, a few caveats with respect to the data used in this study need to be discussed. As shown in the framework, the EAS is comprised of a sequence of events. In an ideal study, this would enable conditional likelihood models and sequence analysis. However, detailed data from within the student support office was not captured in the data extractions for this study. This means who, how and when students were contacted, or what support was offered is unknown. This limits the models which can be used to analyse the process. Two critical data elements were collected for this study which can be used. The first is the data for the identification process. That is, for each day over a three year period, the ranks of all students and the respective triggers activated can be analysed. This enables the models to capture which students were identified by the EAS and why. The second critical data point is the outcome. It is known which students were inactive or intermittent in their enrolment and when. The data also captures which students were still enrolled, who completed and who had discontinued or lapsed. To simplify the enrolment states, it is assumed that the outcomes are independent of the student advising the institution of intent to change states. That is, a student who lapses from the institution should be considered to have discontinued. A student who is intermittently enrolled should be treated the same as a student who is inactive. This means that the final enrolment states collapse down to four possible states, enrolled, completed, discontinued or inactive.

Knowing which students are identified by the EAS and the final state of students allows a black box approach to analysing the EAS system. This eliminates sequential analysis as a possible method; however the remaining modelling options can estimate the relationship between the EAS and student retention. Four methods are identified which help address the research questions. The first modelling approach is multinomial logistic regression. This frequently used method will act as the benchmark for the study, both in model specification and interpretation of results. It is expected that the EAS should have a significant relationship to the probability of the final enrolment states. Additionally, this method has been frequently used in other retention analysis, so will act as a benchmark method of analysis allowing the variables that affect retention to be identified. Second, multiple regression analysis can be used to estimate the length of students' enrolment. It is expected that if the EAS has a positive effect on student retention, then students should be enrolled for longer. Third, through survival analysis, the temporal effects of retention can be explored. Using flexible non-parametric estimation and the removal of many limiting assumptions, it is expected this method will yield consistent results in varying model specifications. Finally, treatment effects modelling can be used to capture the causal effects of the EAS on student retention outcomes such as length of time enrolled and the amount of student tuition fees paid.

3.4 Theoretical underpinnings

With four models selected for use in the empirical analysis, the theoretical underpinnings of these four models are explored to ensure they are valid in the context of this study. The method of estimation is explored, along with the base interpretations of the models. Both dependent and independent variables used in the models are represented in italics throughout the study to aide easy identification of variables.

3.4.1 Likelihood analysis using multinomial logistic regression

The first model is the multinomial logistic regression model. Following Liao (1994, p. 48), the model estimates the probability of discrete enrolment outcomes such as *enrolled*, *completed* and *discontinued*. Mathematically, this is expressed as

$$P(y = j) = \frac{e^{\sum_{i=1}^{K} \beta_{ij} x_i}}{1+\sum_{j=1}^{J-1} e^{\sum_{i=1}^{K} \beta_{ij} x_i}} \tag{3.1}$$

where

j is the response category
$P(y = j)$ is the probability of response category j outcome
β_{ij} is the estimated coefficient for the ith variable contributing to the jth response category
x_i is a vector of K explanatory variables

This is one of the frequently used models, Liao (1994) writes "the rationale is that we should go with a statistical model that requires fewer or weaker assumptions" (Liao, 1994, p. 48). The main assumption associated with the multinomial logistic model is the assumption of independence of irrelevant alternatives (IIA). To illustrate this assumption, a single student can choose between two states, discontinuing or being enrolled. The choice probability of these two states is 50/50 or, 0.5 for each state. If, however, the means by which the student discontinues is factored in, for example whether to discontinue by informing the institution or not, then the choice probabilities, if the assumption holds, should be approximately 0.33 each. In reality however, the choice probability for being enrolled is still 0.5 and the choice probability of discontinuing and informing the university is only 0.25. This means that the independence assumption is violated. As discussed in section 3.3, there exist multiple enrolment states a student can be in. By combining lapsed and discontinued students together into a single category of response variable, this avoids logical violation of the assumption.

The estimated multinomial logistic regression model produces three key statistical measures that assist in determining the significance of the model. The first is the significance of the overall model which is measured through the likelihood ratio chi-square test statistic (LR χ^2). It is easiest to interpret the LR χ^2 test statistic using the p-value approach. This finds the probability of finding a more extreme test statistic than the calculated test statistic. Comparing this to a given level of significance indicates whether the model is statistically significant. The second statistic useful in interpreting multinomial logistic regression models is the significance of the individual coefficients. Again, using the p-value approach, each variable is compared to the standard error of estimation for the coefficient. The p-value is a calculation of the probability of finding a more extreme estimated coefficient. Comparing the p-value to a given level of significance reveals how much the variable contributes to estimating the overall probability of the outcome variable. The final measure is the estimated probabilities of a particular outcome. In this case, the model can estimate the probability of being *enrolled*, *completed* or *discontinued.*

As discussed in section 3.2, multinomial logistic regression is a frequently used statistical tool due to its capacity to handle categorical dependent variables. For student retention analysis, estimating the likelihood of the various enrolment states provides important information on the variables that affect retention.

3.4.2 Multiple regression analysis using ordinary least squares

Ordinary least squares are considered the basic statistical tool used to analyse the relationship between variables. The general form model is mathematically represented as

$$Y_j = \beta_0 + \sum_{i=1}^{k} \beta_i X_{ij} + \varepsilon_j \qquad (3.2)$$

where

Y_j is the dependent variable for observation j
$X_{1j}, X_{2j}, \ldots, X_{ij}$ is the vector of k independent variables for the jth observation
β_0 is the intercept coefficient
$\beta_1, \beta_2, \ldots, \beta_k$ are the slope coefficients parameters for the k independent variables
ε_j is the error term of the jth observation

The model has a linear relationship between the dependent variable Y and the parameters, where effects associated with the independent variables are additive. The assumptions associated with ordinary least squares are as follows:

All variables must be measured at the interval level and without error

For each set of values for the k independent variables ($X_{1j}, X_{2j}, \ldots, X_{kj}$), the mean value of the error term ε_j is zero

For each set of values for the k independent variables, the variance of ε_j is equal to σ^2(ie. the variance of the error term is constant)

For any two sets of values for the k independent variables, the covariance of error terms (ε_j, ε_h) equals zero (ie. the error terms are uncorrelated; thus no autocorrelation)

For each X_i, the covariance of the variable and error term is equal to zero (ie. each independent variable is uncorrelated with the error term)

There is no perfect collinearity – no independent variable is perfectly linearly related to one or more of the independent variables in the model

For each set of values for the k independent variables, ε_j is normally distributed

(Berry and Feldman, 1985, p. 10)

The parameters of interest are estimated using sample data in the estimated regression model. The estimated regression model is presented in equation 3.2.

$$\hat{Y}_j = b_0 + \sum_{i=1}^{k} b_i X_{ij} \qquad (3.3)$$

where

$\hat{Y}_j$ is the estimated dependent variable for observation j
$X_{1j}, X_{2j}, \ldots, X_{ij}$ is the vector of k independent variables for the j^{th} observation
b_0 is the estimated intercept coefficient
$b_1, b_2, \ldots, b_k$ are the estimated slope coefficients for the k independent variables

Several important statistical measures are frequently used to test the significance of ordinary least squares models. These are the R^2 value, the F-test statistic, and the significance of the estimated slope coefficients (b_1, $b_{2, \ldots}$, b_k), where b_i is the estimation of β_i.

The R^2 value is given by the formula

$$R^2 = \frac{\sum_{j=1}^{n}(\hat{Y}_j - \bar{Y})^2}{\sum_{j=1}^{n}(Y_j - \bar{Y})^2} \qquad (3.4)$$

(Berry and Feldman, 1985, p. 15)

where

Y_j is the observed dependent variable.
$\bar{Y}$ is the average of the dependent variable.
$\hat{Y}_j$ is the estimated dependent variable from the estimated regression model.

The R^2 value is a goodness of fit measure of the overall model. It is interpreted as the proportion of variance of the observed variable that can be explained by the regression model. R^2 can take a value between 0 and 1, where 1 indicates that the regression model perfectly explains all variation within the observed dependent variable. While it is desirable to have a statistical model with a high R^2 value close to 1, it needs to be interpreted in the context of the model with supporting information from the F-test statistic.

The F-test statistic is an estimation of the overall significance of the regression model. It tests the hypothesis that all the estimated slope coefficients are equal to zero. The F-test statistic is calculated as

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{3.5}$$

Comparing the F-test statistic to the F-critical value, it can be concluded whether the null hypothesis is rejected. A common approach to determining the significance of the overall model is to calculate the probability of finding a more extreme F-test statistic, or the p-value. If the p-value is less than a given level of acceptable error, or significance, then the null hypothesis can be rejected. The most common level of significances used are 10%, 5% and 1%. The lower the significance of the model, the more statistically significant the results are.

The last statistical measure of importance is the significance of the estimated coefficients $b_1, b_2, \ldots, b_i$. Each estimated coefficient has a standard error (se) that can be calculated as per Berry and Feldman (1985, p. 13). The test for significance for each independent variable follows a t-distribution and is calculated for the general b_i variables as follows:

$$t_{b_k} = \frac{b_i - \beta_i}{se_{b_i}} \sim t_{n-k-1} \tag{3.6}$$

The t-test statistic is compared to the t-critical value with n-k-1 degrees of freedom for a given level of significance. The p-value approach also applies, where the probability of attaining a more extreme t-test statistic is calculated. If the p-value is less than the chosen level of significance, then the slope coefficient is deemed to be statistically significant.

Overall, the OLS approach serves as a suitable statistical tool to analyse student retention. The three key statistical measures outlined will help identify the variables which significantly affect suitable measures of student retention.

3.4.3 Survival analysis using the Cox proportional hazards model

The Cox proportional hazards model is a common regression technique within survival analysis. The model estimates the hazard function $\lambda(t)$, which captures a level of risk associated with a defined failure event. In this case, the failure event is that the student lapsed or discontinued their enrolment. The main focus of the model is in assessing "the relation between the distribution of failure time and $[x]$" (David R Cox, 1972, p. 189), where x captures the array of variables that can affect the hazard function.

The general model is often represented mathematically

$$\lambda(t{:}\,x) = \lambda_0(t)e^{\beta x} \tag{3.7}$$

where

$\lambda(t{:}\,x)$ is the hazard function for time period t given x explanatory variables
β is the coefficients to the vector of explanatory variables x
$\lambda_0(t)$ is the baseline hazard function

The general form of the model can be rewritten in the forms of probability and relative risk as

$$\frac{\lambda_0(t_{(i)})e^{x' j(i)\beta}}{\sum_{j \epsilon R_i} \lambda_0(t_{(i)})e^{x' j(i)\beta}} \tag{3.8}$$

The multiplicative relationship between the baseline hazard function and the exponent capturing the effect of explanatory variables means that the function can be simplified. The baseline hazard $\lambda_0(t_{(i)})$ can be cancelled out to show that the relative risk associated with the failure time is independent of the baseline hazard function. This is presented in Equation 3.9.

$$\frac{e^{x' j(i)\beta}}{\sum_{j \epsilon R_i} e^{x' j(i)\beta}} \tag{3.9}$$

Using this form of the equation, the multiplication of 3.9 across all failure times t produces ordinary likelihood estimation. This is presented in equation 3.10.

$$L = \prod_{i=1}^{m} \frac{e^{x' j(i)\beta}}{\sum_{j \epsilon R_i} e^{x' j(i)\beta}} \tag{3.10}$$

Robust inferences can be made on the estimated model as discussed in Lin and Wei (1989).

The Cox proportional hazards model extends to dealing with time varying covariates. This means the estimated hazards associated with each explanatory variable (β) can change over time, allowing dynamic temporal modelling. The Cox proportional hazards assumption states that the baseline hazard function is independent of time. It is critical for this assumption to be tested to ensure valid inferences of the coefficients.

Censoring is another issue associated with survival analysis. In student retention analysis, this is where students discontinue their studies after the end of the period of analysis. The student would appear enrolled at the end of the data capture period, however this does not represent the final outcome for the student. In the proportional hazards model,

> it can be assumed that censorings can only occur immediately after failures. This requirement does conflict slightly with the model in which censoring times are fixed constants, but can usually be viewed as a reasonable approximation, as the information about β contributed by an exact observed censoring time c_i will generally be small (Cox and Oakes, 1984, p. 93).

For student retention analysis, this is not a significant issue if the time period captured by the data set is sufficient. It is expected that within the survival model, most students who would discontinue do so before the end of the three year data capture period. This means any error associated with censoring is minimal. Overall, the Cox proportional hazards model is a useful approach of analysing the variables that affect student retention.

3.4.4 Treatment effects modelling

Treatment effects modelling is a method that allows causal inferences to be made on observational data. It is an appropriate method to use when “the cost of performing the equivalent randomised experiment to test all treatments is prohibitive; there are ethical reasons why the treatments cannot be randomly assigned; estimates based on results of experiments would be delayed many years” (Rubin 1974, p. 688). The following scenario of a university with an EAS is used to understand the theoretical underpinnings of the treatment effects model.

A student is enrolled at a university where an EAS is actively identifying students at risk of discontinuing. Two possible outcomes can occur for this student: either they are identified by the EAS, or not identified. To measure the causal effect of EAS identification on student outcomes, the difference between the two outcomes should be measured. This however is impossible, where only one of these outcomes can be observed, while the other outcome remains unknown. This is the fundamental issue with causal inferences. The solution to the problem is to attain an overall average effect by comparing two groups of students, one group identified by the EAS, the other acting as a pseudo control where no EAS identification occurred. The model works by estimating the expected outcome of the alternate unobserved outcome. The treatment effects model estimates what would have happened if students identified by the EAS were not identified. The opposite applies, estimating the expected outcome of students not identified by the EAS, as opposed to if they were identified. Comparing the observed and expected outcomes, the average treatment effect of EAS identification can be calculated.

Three parameters of interest compose the treatment effects model. These are the Predicted Outcome Means (POM), the Average Treatment Effects (ATE) and the Average Treatment Effects on the Treated (ATET).

The POM is simply the expected value of for the treatment group. For example, if measuring the length of enrolment was measuring the effect of the EAS, the POM would be the expected length of enrolment if students were identified by the EAS.

The ATE estimates the effect of the treatment within the whole sample. It is mathematically represented as

$$ATE = E(Y_{1i} - Y_{0i}) = E(Y_{1i}) - E(Y_{0i}) \tag{3.11}$$

where

$E(Y_{1i} - Y_{0i})$ is the average treatment effect capturing the difference in outcomes Y for treatment group 1 compared to the non-treatment group 0.
$E(Y_{1i})$ is the expected outcomes for the treatment group
$E(Y_{0i})$ is the expected outcomes for the non-treatment group

The ATET estimates the treatment effect only for the treatment group. Mathematically, this is represented as

$$ATET = E[Y_{1i} - Y_{0i}|D_i = 1] = E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1) \quad (3.12)$$

where

$E[Y_{1i} - Y_{0i}|D_i = 1]$ is the conditional average treatment effect on the treated
$E(Y_{1i}|D_i = 1)$ is the expected outcome of the treatment group under the condition of the treatment
$E(Y_{0i}|D_i = 1)$ is the expected outcome of the non-treatment group under the condition of the treatment

The ATE and ATET statistical measures can be interpreted as the average effect of the treatment in the population and the treatment group respectively. As highlighted in the example, the purpose of the treatment effects model is to determine what, if any, causal effect the EAS has on student outcomes.

3.5 Model specification, calibration and hypothesis testing

The specific models used in this study can be broken down into containing several classes of explanatory variables, *demographic*, *institutional*, *student performance*, *workload* and *EAS* variables. For the purposes of the study, lapsed and discontinued students are grouped together as they represent the same outcome with administrative interaction the only difference. To capture the effects of students being inactive or intermittently enrolled, a weighted average workload variable was created. The weights are relative to the time spent in a particular state: being inactive, part-time or full-time. Students who had a significant period of inactivity have a workload value close to 0, whereas students who had maintained a fulltime workload would have a value close to 2. Part-time students have a value close to 1 on this continuum. For the temporal modelling using survival analysis, this becomes a categorical variable, representing the workload of the student in any given week.

3.5.1 Specification and calibration of likelihood approach

Using multinomial logistic regression, the approach estimates the relationship between the four classes of variables to the probability of three enrolment outcomes. These discrete states are whether the student discontinued, completed or was still enrolled at the end of the data capture period. Mathematically, the general model specification in this case is

$$P(Y = j) = f(\beta_{j0} + \sum \beta_{jp} D_p + \sum \beta_{jq} I_q + \sum \beta_{jr} L_r + \beta_j E + \varepsilon) \quad (3.13)$$

where

$P(Y = j)$ is the probability of the outcome Y equalling event j

β_{j0} is the constant term
β_{jp} is the coefficients for the vector of p demographic variables D
β_{jq} is coefficients for the vector of q institution variables I
β_{jr} is the coefficients for the vector of r student performance and workload variables L
β_{js} is the coefficient for the EAS variables E
ε is the error term of the model

For this approach, the EAS can take on two possible specifications to capture the effects of the EAS. Firstly, students can be divided into identified and not identified categories, making the EAS variable a dummy variable. The second option is to split students into different treatment categories. Since the EAS runs daily, it is possible for students to be identified multiple times, sometimes within the same week. As such, the number of times a student was identified can be broken down into five classes of treatment. The first is that the student was never identified by the EAS. The second group is comprised of students identified one to four times. The third group is comprised of students identified five to nine times. The fourth group is comprised of students identified ten to nineteen times and the final group is comprised of students identified twenty times or more.

The main hypothesis tested in this model relates to the coefficient of the EAS variable. It is expected that students identified by the EAS will have a significantly different likelihood of discontinuing than students not identified by the EAS. The null hypothesis is that there is no significant difference in the estimated EAS coefficients:

H_0: $\beta_{discontinued}$|Identified = $\beta_{discontinued}$|Not Identified

3.5.2 Specification and calibration of multiple regression approach

The multiple regression OLS model correlates the relationship between the dependent variable and the four classes of explanatory variables. The dependent variable is the number of weeks enrolled at the institution. This is calculated by taking the difference between the commencement date and either the discontinuation date, completion date or end of data capture date. The number of weeks enrolled captures both teaching and non-teaching periods, since students can make the decision to leave the institution at any time. The expected outcome of the model is that if the EAS works, students should be retained and as such, the average length of enrolment will be higher for students identified by the system. The general model specification is as follows:

$$Y = \beta_{j0} + \sum \beta_{jp} D_p + \sum \beta_{jq} I_q + \sum \beta_{jr} L_r + \beta_{js} E + \varepsilon \qquad (3.14)$$

where:

Y is the length of the students' enrolment
β_{j0} is the constant term
β_{jp} is the coefficients for the vector of p demographic variables D
β_{jq} is coefficients for the vector of q institution variables I
β_{jr} is the coefficients for the vector of r student performance and workload variables L
β_{js} is the coefficient for the EAS variables E
ε is the error term of the model

The EAS variable will again be expressed in two versions, an overall treatment where students are divided into identified and not identified categories. The second model will estimate the same relationship using varying treatment levels. The main hypothesis the multiple regression case is that there is a significant positive relationship between the EAS and the length of students' enrolment. The null hypothesis is then that the EAS has no significant relationship to the length of enrolment.

H_0: $\beta_{js} = 0$

3.5.3 Specification and calibration of survival analysis approach

This study uses an extension of the survival framework used by Ishitani (2003, p. 437) to calibrate the models using the survival analysis approach. Two general frameworks are presented which capture the scope of analysis, both in terms of the variables used and the time frames for the respective sections. The first is the teaching period framework in Figure 3.4, which shows the process of students' initial enrolment during the first teaching period. The second framework is depicted in Figure 3.5 and shows a course level model. Both frameworks show progression as a measurement of time (t). For any given student, there are *i* weeks in a teaching period, and *j* weeks in a course.

Figure 3.4 represents the sequential decision process faced by students on a weekly basis during a teaching period. The exogenous variables in the sequence are the pre-enrolment factors that affect student retention, including *demographic* variables. Once the teaching period commences, the teaching period variables are held constant which includes *workload*, *college residential status* or *student performance* for that teaching period.

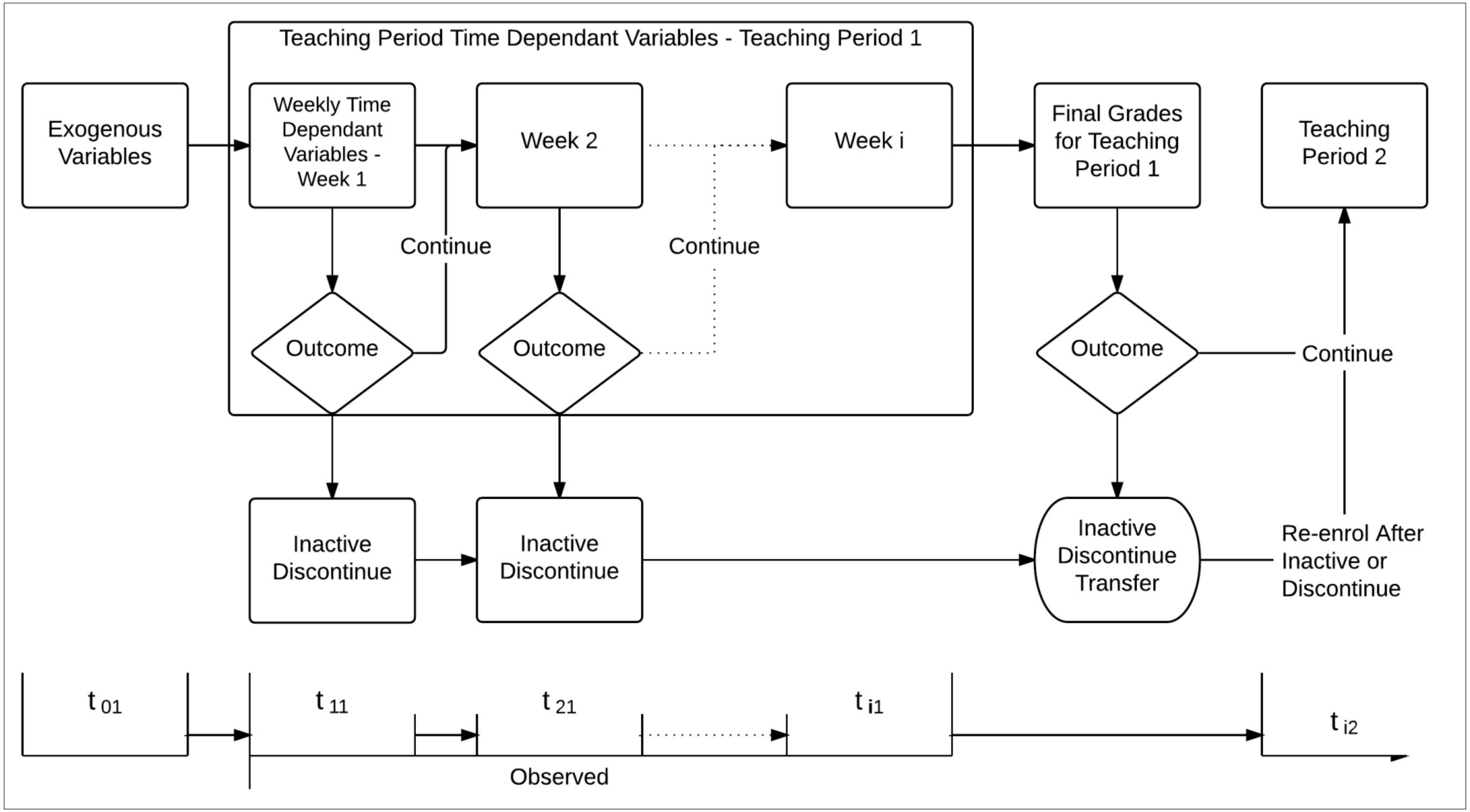


Figure 3.4 – Teaching period survival analysis framework (Authors own concept)

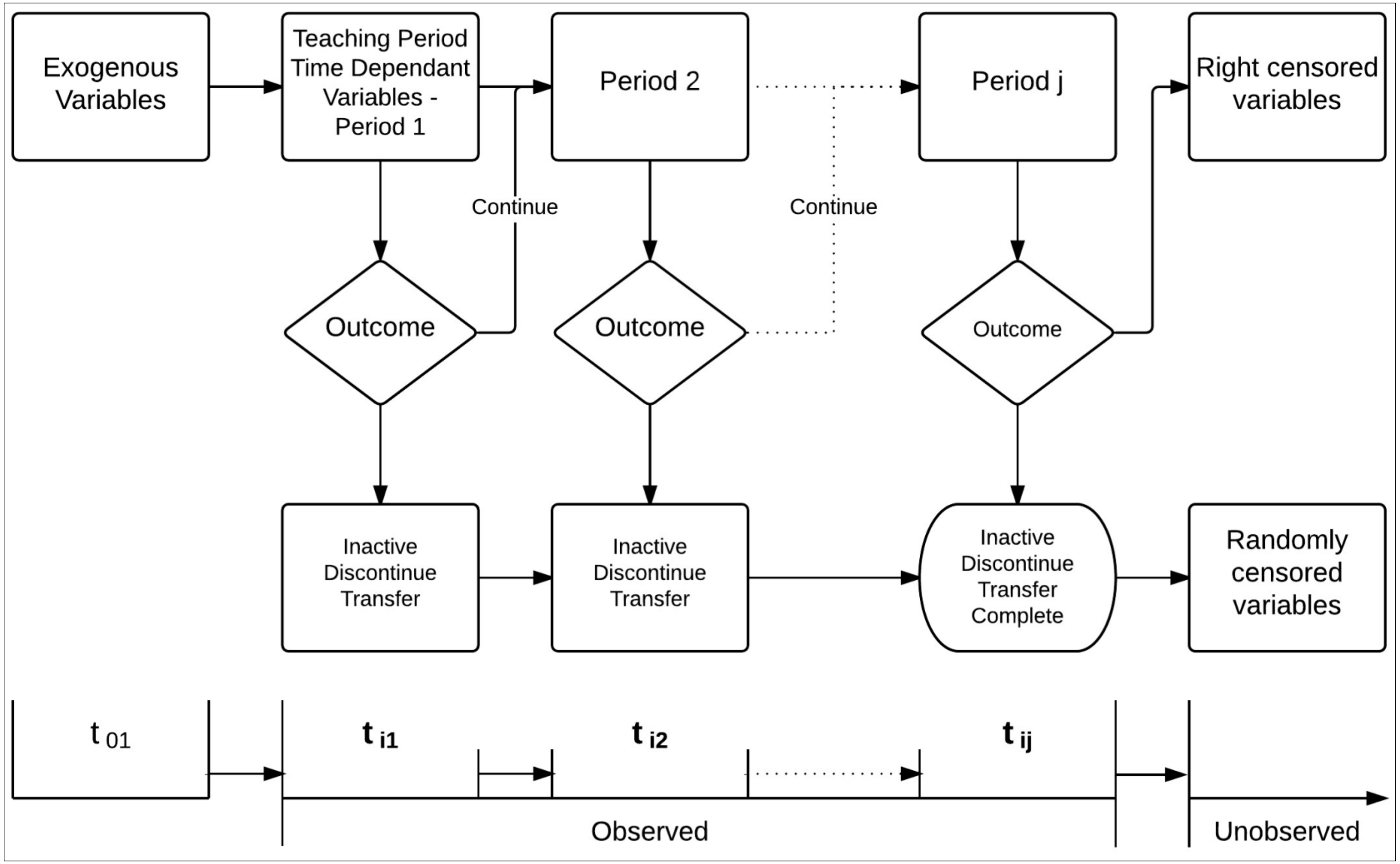


Figure 3.5 – Course period survival analysis framework
(Authors own concept)

Within the teaching period, weekly variables can capture student-institution interactions which can reveal important timely information on the students' progress within the teaching period. In reality, these interactions are occurring constantly throughout the teaching period. However, for the purposes of analysis and simplification of modelling, weekly variables are used which aggregate data that is captured daily or even hourly.

Discontinued or *inactive* outcomes can occur at any time during the teaching period. Institutions typically have fixed admissions periods, typically between teaching periods, and transferring students typically will want recognition of studies already completed. As such, transferring to another institution is assumed to only occur at the end of the teaching period. *Completion* of the qualification also only occurs at the end of the teaching period and course; it is the sum of the successful completion of the necessary units to be admitted to the award.

Using the frameworks developed, five models are estimated using the Cox proportional hazards model over weekly time periods. The first model (Equation 3.15) estimates the proportional hazards of students under the three different EAS specifications. It is referred to as the base model. Generally, the first model can be expressed as follows:

$$\lambda(t\colon z) \propto \beta_{j0} + \sum \beta_{jp} D_p + \sum \beta_{jq} I_q + \sum \beta_{jr} L_r + \beta_{js} E + \varepsilon \qquad (3.15)$$

Where

$\lambda(t\colon z)$ is the hazard function over time which is proportional to the regressors
β_{j0} is the constant term
β_{jp} is the coefficients for the vector of *p* demographic variables D
β_{jq} is coefficients for the vector of q institution variables I
β_{jr} is the coefficients for the vector of *r* student performance and workload variables L
β_{js} is the coefficient for the EAS variables E
ε is the error term of the model

Within 3.15, the EAS variable can be expressed in three different configurations. First, the short run effects of the EAS are estimated by taking the value of 1 if the student was identified in a given week and 0 if the student was not identified in a given week. Second, the enduring effect associated with being identified is estimated. The EAS variable remains 0 until the week the student is first identified, after which it remains as 1. The third configuration captures the long run overall effects of the EAS system. Students are divided into two categories for the duration of the data set, students identified and students not identified by the EAS. The three configurations are expected to reveal different features of the EAS. The short run effect should be that students identified by the EAS are at a higher risk of discontinuing. The enduring effect is unpredictable, but the long run effect is expected to reduce the risk of discontinuing.

The second model (Equation 3.16) analyses within the long run model for differences between the identified and not identified student cohorts. Logically, this is a conditional model which identifies the factors of retention under the condition of EAS identification or not identified. The model is referred to as the conditional model. The first configuration captures the effect of the explanatory variables given that the students were never identified by the EAS. The second configuration uses the same variables to capture differences of effects for students identified by the EAS. Since analysis is occurring within the EAS variable, the general form of the model is

$$(t:z|Identified) \propto \beta_{j0} + \sum \beta_{jp} D_p + \sum \beta_{jq} I_q + \sum \beta_{jr} L_r + \varepsilon \qquad (3.16)$$

where

$\lambda(t:z|Identified)$ is the hazard function over time which is proportional to the regressors under the condition the student was identified
β_{j0} is the constant term
β_{jp} is the coefficients for the vector of demographic variables D
β_{jq} is coefficients for the vector of institution variables I
β_{jr} is the coefficients for the vector of student performance and workload variables L
ε is the error term of the model

The third model (Equation 3.17) analyses the interaction effect associated with the EAS. Logically, this is the joint condition model where a given variable and EAS identification are tested together. The model is referred to as the interaction model. Two alternate model configurations are used to capture the interactions of the EAS. The first uses the short run expression of the EAS used in Equation 3.15. Interpreting the effects of the EAS in the interactions model is relating the effects in any given week. The second uses the long run specification of the EAS used in Equation 3.15. Interpreting the results captures overall interaction between variables and the EAS. It is expected that the two configurations will indicate differences between the independent variable compared to the independent variable jointly conditioned with EAS identification. Generally, the model can be expressed as

$$\lambda(t{:}\,z) \propto \beta_{j0} + \sum \beta_{jq} D_q + \sum \beta_{jr} I_r + \sum \beta_{js} L_s + \beta_{jt} E\left(D_q + I_r + L_s\right) + \varepsilon \qquad (3.17)$$

where

$\lambda(t{:}\,z)$ is the hazard function over time which is proportional to the regressors
β_{j0} is the constant term
β_{jq} is the coefficients for the vector of demographic variables D
β_{jr} is coefficients for the vector of institution variables I
β_{js} is the coefficients for the vector of student performance and workload variables L
$\beta_{jt} E(D_q + I_r + L_s)$ is the multiplicative interaction effect with the demographic, institution and student performance and workload variables
ε is the error term of the model

The fourth model (3.18) diverges from the previous model specifications, where the focus of the model is the EAS triggers. This model is called the EAS trigger model. The general form of the model is expressed as

$$\lambda(t{:}\,z) \propto \beta_{j0} + \beta_{js} ET_s + \varepsilon \qquad (3.18)$$

Where

$\lambda(t{:}\,z)$ is the hazard function over time which is proportional to the regressors
β_{j0} is the constant term
β_{js} is the coefficients for the vector of *s* EAS trigger variables ET
ε is the error term of the estimation

The 34 triggers which comprise the identification process of the EAS are used as regressors. The relationship between the hazard ratio and the triggers can be expressed in four configurations. The first captures the relationship between the triggers and the hazard function for all students. The triggers are presented as binary variables, where 1 indicates if the trigger was activated in a given week. The second configuration also relates the activated triggers for all students, but instead treats them as continuous variables. This allows the second model to test the effects associated with a student activating the same trigger several times in a given week. The third configuration limits the activated triggers only to students identified by the EAS. Like the first configuration it uses a binary variable approach to representing the triggers. By isolating the triggers to only those identified by the EAS, it is expected that the most prominent triggers linked with retention within the EAS will be identified. The final configuration builds from the previous approach, but expresses the triggers as continuous variables to further test the effects of being identified by the same trigger multiple times within a week.

The fifth model (Equation 3.19) combines the base survival (Equation 3.15) and trigger survival (Equation 3.18) models together. The EAS variable from the base model is replaced by the underlying triggers of the EAS from the trigger model. The model captures the effect of triggers in the presence of other established explanatory variables. It is expected this will control for variations associated with *demographic, institutional, student performance and workload* effects. This model is known as the controlled-trigger model. The general form of the model is expressed as

$$\lambda(t{:}\,z) \;\propto\; \beta_{j0} + \sum \beta_{jp} D_p + \sum \beta_{jq} I_q + \sum \beta_{jr} L_r \; + \beta_{js} ET_s + \; \varepsilon \tag{3.19}$$

Where

$\lambda(t{:}\,z)$ is the hazard function over time which is proportional to the regressors
β_{j0} is the constant term
β_{jp} is the coefficients for the vector of *p* demographic variables D
β_{jq} is coefficients for the vector of *q* institution variables I
β_{jr} is the coefficients for the vector of *r* student performance and workload variables L
β_{js} is the coefficient for *s* EAS trigger variables ET
ε is the error term of the model

The five survival models are expected to reveal detailed information on the operations of the EAS under various configurations. The different approaches ensure a detailed treatment of the effects associated between the EAS and any link to student retention. Importantly, these models will allow temporal effects, which will indicate which variables are significant at different times during students' enrolment.

3.5.4 Specification and calibration of survival treatment effects approach

One additional model is covered under the banner of survival analysis is a combination of two statistical methods. Survival treatment effects uses survival analysis data to estimate the ATE and ATET associated with a particular treatment. While the five survival models provide solid analysis of the relationship between the EAS and retention, the survival treatment effects model will establish causal inference. This is an important step in evaluating the effectiveness of the EAS in addressing retention.

The model will use the same *demographic*, *institutional*, *student performance*, *workload* and *EAS* variables specified in the previous models. The dependent variable will be the students' length of enrolment. It is expected that the results will support the multiple regression model specified in 3.5.2.

3.5.5 Specification and calibration of treatment effects approach

The objective of the treatment effects model is to analyse if a causal relationship exists between the EAS and institution revenue in terms of student tuition fees. To develop such a model, it is important to know the fees charged for each student. While ideally the exact fees paid by a student would be provided as part of the data set covered in chapter 4, this was not possible. In lieu of this, it is possible to estimate the fees paid by a student given the number of units study in a given school. Table 3.2 was extracted from the UNE website to show the breakdown of fees depending on course and student type.

Table 3.2 – Domestic Fee Schedule by Band

	Domestic HELP	Domestic Fee Paying
	Per 6 CP	Per 6 CP
Band 1	$755.00	$679.50
Band 2	$1,076.00	$968.40
Band 3	$1,260.00	$1,134.00
Band 4	$755.00	$679.50
Band 5	$1,260.00	$1,134.00
Band 6	$1,076.00	$968.40
Band 7	$1,076.00	$968.40

(University of New England, 2014b)

From this information, it is possible to create an estimated fee paid by students, and the model can test for a significant difference in the fees paid to the university under the effect of the early alert system. The bands refer to the fees charged for different schools within the case study institution. Six credit points (6 CP) is the equivalent of one unit of study, with different qualifications and courses requiring a set number of units to be undertaken to be admitted to the award.

International student fees are taken from the UNE 2013 international student prospectus. (University of New England, 2013). To create parity with the domestic fee schedule by band, the international student fees which are charged annually were divided by 8 (a typical full-time annual load) and correlated to the most representative band of study. The final step is to then calculate the total amount of revenue generated from each course. The revenue function can be expressed as:

$$R_{cx} = \sum_{i=1}^{I} C_{xDef} + \sum_{j=1}^{J} C_{xDom} + \sum_{k=1}^{K} C_{xInt} \tag{3.20}$$

where

R_c is the course revenue for course x
I is the total number of deferred students
J is the total number of upfront fee paying domestic students
K is the total number of international students
C_x is the respective costs faced by students to enrol in those courses

The calculations of R_{cx} combined with the retention data allow the calculations of the expected total fees paid by a student. The expected total fees paid will form the dependent variable in a treatment effects model, with the EAS variable acting as the treatment variable. The treatment effects model will also include demographic, institution and learning environment variables to control for variations observed in previous models.

3.6 Chapter summary

This chapter outlined the methods of analysis previously used which informed the development of both temporal and non-temporal models. These models make a significant contribution to the literature in the areas of student retention analysis and learning analytics. Using big data, the models will be able to provide a level of detail of the student retention problem that was previously not possible. Furthermore, these models set the framework to be expanded for more complex, real time retention modelling in the future. This will have significant economic implications for tertiary institutions enabling them to identify not only which students to target, but with sufficient information, how to target students with support tailored to their needs.

Chapter 4 - Data Set

4.1 Introduction

Data is the foundation of learning analytics and evidence-based approaches to enhancing the learning environment. The data set used for analysis represents a rich source of information on students within a single institution. To fully understand the complexity of the data set, chapter 4 looks at four key aspects of the data set. First, the background of the University of New England identifies common and unique characteristics to the institution. This section also captures the establishment and implementation of the EAS currently in use. Second, a description of the data extraction process along with the ethics and privacy aspects of using student data. Additionally, the section will address the process of balancing the data for analysis. Third, overall descriptive statistics of the data set are provided with respect to the key variables used in analysis. Finally, an overview of the data in temporal setting shows the importance of analysing student retention in a temporal setting.

4.2 Background of the University of New England

The University of New England (UNE) is one of 39 public universities in Australia (Australian Education Network, 2014b). Established in Armidale, New South Wales, in 1938 as a campus of the University of Sydney, UNE was Australia's first regional university campus. UNE transitioned from a regional campus to an autonomous university in its own right in 1954. During this period, UNE "pioneered teaching to external students by correspondence" (University of New England, 2014a), which continues to be a fundamental aspect of the university. In 2013, UNE had 22,389 students enrolled, of which 78.9 per cent were off-campus students. UNE has a significantly larger off-campus student cohort when compared to national averages (Department of Education, 2014b), making it a unique institution in the higher education sector. Additionally, UNE, through distance education, has been able to create a strong focus on supporting students from disadvantaged backgrounds.

At UNE, student retention information has been published as a part of the Department of Industry data sets last published in 2012 (Department of Industry, 2013c). Using the same definitions of retention as provided by the Department of Industry, the retention rate for UNE is shown in Figure 6 along with both the national and RUN average student retention rates.

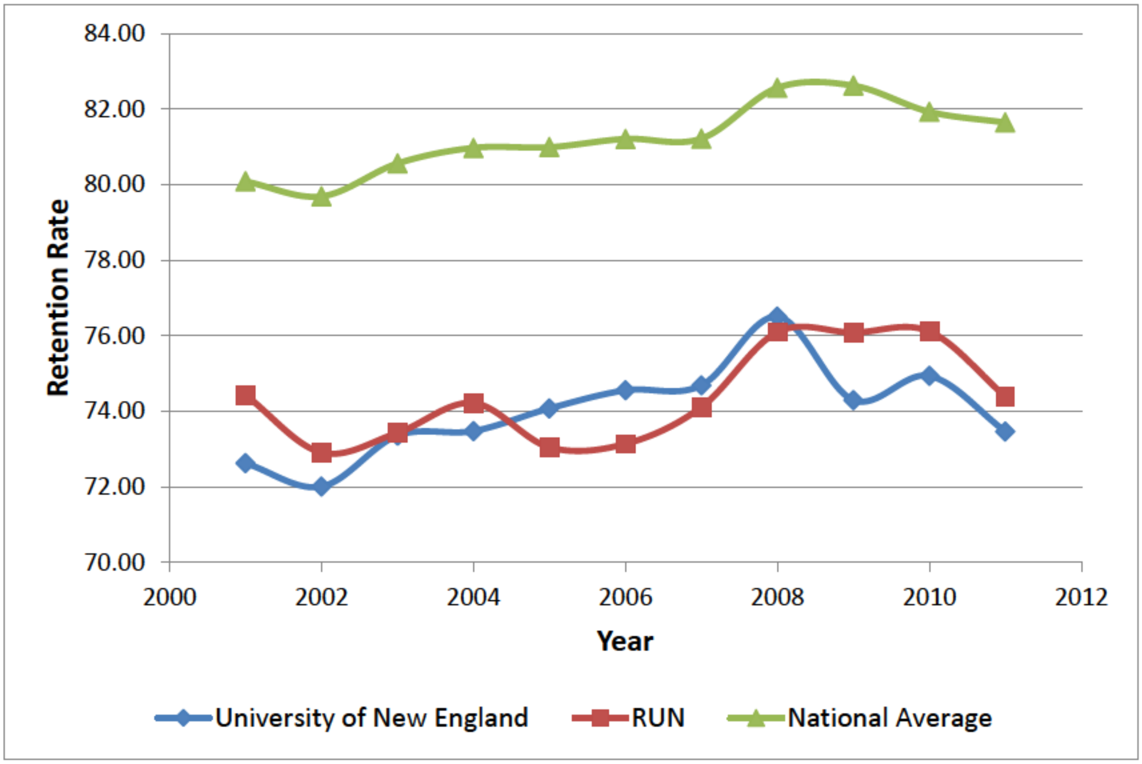


Figure 4.1 – Average student retention rates at UNE

The Department of Industry 2012 data set shows that UNE has below average student retention levels and is within ±2% of the average for RUN. The data published by the Department of Industry provides an important starting point in analysing student retention rates with UNE as the case study institution.

UNE established the Early Alert Program (EAP) to provide targeted student support, with the hope of also improving student engagement. The EAP commenced in 2010 with the objective of identifying students at risk of disengaging from their studies and providing tailored support solutions to students. To reach the objectives of the program, consultations with university staff lead to the development of the Automated Wellness Engine (AWE). The AWE is an in-house EAS, used to identify students at risk of disengagement. This enables the student support team to reach the EAP objectives of increased student engagement through the provision of targeted student support. This innovative program was nationally recognised with ALTC awards in 2011 and was regarded as the cornerstone of student engagement and retention activities at UNE (University of New England, 2012, p. 16).

4.3 The data extraction process

The data set used in this study is a rich data set containing detailed information on students enrolled at UNE between 2011 and 2014. The data set supplied for this study comes with the cooperation of UNE's Corporate Intelligence Unit (CIU). Because of the detail of the data set, there are significant ethical and privacy issues that needed addressing in the data extraction process. With support from the CIU, an ethics application made to UNE's ethics committee was approved to start data extraction from 01/07/2013. A key condition of the ethics approval is that the CIU provide a de-identified data set to prevent any breaches in student's privacy. Further to the ethics approval process, the CIU was required to obtain approval from the legal office of UNE to release the data. The unprecedented request for detailed student data required substantial time and resources, with a confidentiality deed resulting. The deed outlined the legal aspects of obtaining, storing and using the confidential data. The deed was signed by the researchers and representatives from CIU on the 18/02/2014.

Selecting the variables included in the dataset required negotiation between the researchers and CIU. The researcher provided a list of variables to CIU that was constructed using previous literature on student retention modelling and other possibly significant variables that relate to student retention that had not previously been captured. The CIU assisted in identifying variables that would be available for analysis, with limitations from the underlying data structures of UNE's data warehouse and information systems. Bounded by time, the data capture period was from the 01/01/2011 to the 14/03/2014 when the CIU was able to provide the majority of data for the project. The data set extracted consisted of multiple tables capturing different aspects of student data. The tables capture

1. Student attributes data, containing predominantly demographic information.
2. Student learning data, containing information on students' enrolment in specific units and their interaction with the learning environment.
3. Student grade data, capturing both what grade a student received for a particular unit and when the grade was received.
4. Early alert systems data, capturing the identification process used by student support to identify students in need of assistance. The raw EAS data contains daily observations.

The final data sets used collate the raw data tables supplied from CIU into two main data sets. The first data set gathers tables 1 to 3 together to provide a cross-sectional data set with no temporal effects. This data set is used by multinomial logistic regression models in chapter 5, multiple regression models in chapter 6, survival treatment effects model in chapter 7 and the treatment effects model in chapter 8. The second data set is more complex, incorporating table 4 to capture detailed temporal effects. The EAS data is aggregated into weekly observations to ensure sufficient observations within each time period. The second data set is used in survival analysis models in chapter 7.

4.3.1 Balancing the data

The balancing process amalgamates observations from multiple data sources for statistical analysis. Each record within the raw data tables is composed of a unique student identifier, course identifier and unit identifier. The data set captured 18,324 students with only one course enrolment record, 2,008 with two course enrolments, 181 with three enrolments, 25 with four enrolments and four students with five different course enrolments. This composes a total of 20,542 students enrolled in 23,003 courses. Some students had either simultaneous enrolment in multiple courses, or the students had discontinued one course and started another within the institution. For purposes of analysis, these student records are deemed to be independent. Where a student has discontinued their studies and taken up another course within the institution, the records are included to help identify the reasons why the discontinuations occurred.

Balancing the data set required matching observations across all four tables. The process eliminated students where data was missing from one or more tables. Students enrolled in cross institutional programs were removed from the data set due to unobtainable data on studies conducted at partner institutions. The final cross-sectional data set captured 13,445 students enrolled in a single course, 1206 students enrolled in two courses, 81 students enrolled in three courses and 4 students enrolled in four courses. The result was 14,736 students enrolled in 16,124 courses. Therefore the number of observations used through the analysis is 16,124. The final temporal data set uses the same 16,124 records across 156 weeks of data obtained from the EAS. Due to varying lengths of enrolment and the associated censoring of observations, the total number of observations in the temporal model is 1,119,710.

4.4 Overview of the data set

UNE is unique in the characteristics of the population it services. Descriptive statistics for the main variables captured by the data set and used in analysis are presented in Table 4.1 using 16,124 observations. The averages are weighted relative to the number of units attempted.

Table 4.1 – Continuous variable descriptive statistics

Variable	Mean	Std. Dev.	Min	Max
Age	27.675	10.474	15	89
Withdrawn	0.69	1.531	0	16
Withdrawn Early	0.944	1.61	0	16
Fail Incomplete	0.589	1.497	0	18
Fail	0.58	1.307	0	12
Pass	2.443	3.177	0	19
Credit	2.868	3.123	0	18
Distinction	2.4	3.013	0	20
High Distinction	0.979	2.167	0	28
Other	0.362	1.359	0	12
Units Enrolled	2.292	0.858	0	6.3

The descriptive statistics show a high variance in age amongst the student body. With an average age of just over 27 years, this reflects the diverse student cohort that studies at UNE. Taking a graphical analysis approach, the distribution of ages of students is shown in Figure 4.2. The distribution of students by age is heavily skewed, with the main body of students being around 19 years of age. There are some outliers with respect to age as identified by the minimum and maximum ages. The minimum age is 15 years, which is possible for an advanced youth who satisfies the conditions of entry to the university. The oldest student is 89 years of age.

Grade distributions show that the average student attains over two pass, credit and distinction grades, while high distinctions are less frequently awarded. The average number of units enrolled is just over two units per teaching period. This captures the varying workload levels students can undertake while studying. The average is just above what constitutes a part-time workload.

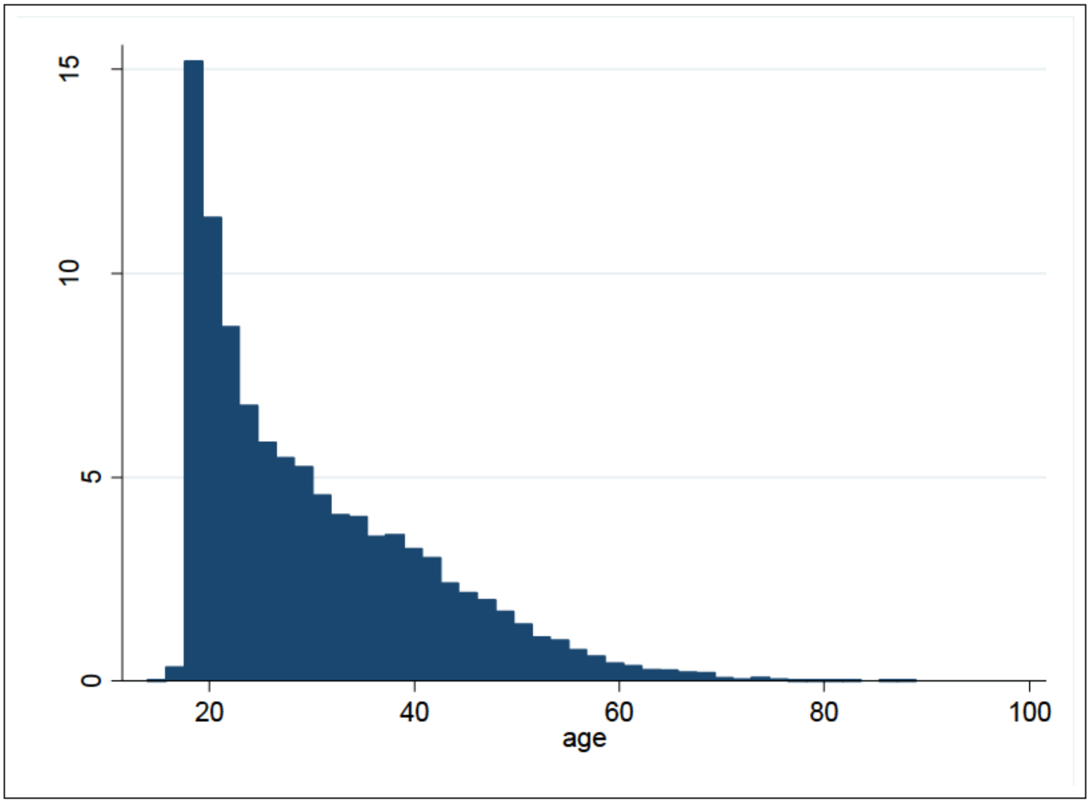


Figure 4.2 – Distribution of students by age

Aboriginal and Torres Strait Islander students (ATSI) comprise 3.2 per cent of the sample student population. This is higher than the Australian average of ATSI students in higher education (Australian Bureau of Statistics, 2013). UNE also provides focused support for ATSI students through the Oorala Centre. The Oorala Centre specialises in the cultural, social and study needs of ATSI students. This facility plays a key role in supporting ATSI students throughout their enrolment.

Breaking students down into fee categories, only 0.1 per cent of domestic students pay fees up front. The small number of students in the domestic fee paying category may impact on the capacity to meaningfully analyse students within the group. of sampled students, 4.9 per cent are international fee paying students. Compared to other Australian institutions, this is a small proportion of the international student market (Department of Education, 2014b). The remaining 95 per cent of students defer student fees using the Higher Education Loan Program (HELP).

UNE has specialised in distance education since the 1940s. This is reflected in the proportion of students who study on-campus. Within the sample data set, 38.2 per cent of students study on-campus. This means 61.8 per cent of students are off-campus, online students. This has important implications for the analysis. UNE is representative of trends being experienced globally, where higher education is being moved to the online environment.

Nearly 20 per cent of students represented in the sample data have registered undertaking some form of previous study prior to enrolment. The type of study undertaken does not matter, it could be a previous degree or qualification at university, or a certificate in skills-based areas such as hospitality. The variable should capture students who are more prepared for the challenges of university study. It is expected this should have a significant relationship to student retention.

Courses undertaken by students can be divided into five categories. These are diplomas, advanced diplomas, bachelors, bachelors by graduate entry, and bachelors with honours. The proportions of students undertaking diplomas and advanced diplomas are 0.7 per cent and 2.8 per cent respectively. These comprise a small part of the sample. Bachelor students make up the majority, with 83.6 per cent of students enrolled in a bachelor level course. Graduate entry students have already graduated from a previous university qualification and make up 5.8 per cent of the sample. Honours program students comprise 1.7 per cent of the sample. It is expected that both graduate students and honours course students will have significantly different relationships from the retention rate than other students. This is because of expected preparedness for university study.

UNE is comprised of ten schools that offer courses in a wide variety of areas, including business, arts, psychology, law, rural medicine, and environmental and rural science. To assess between school differences, the school variable is categorical, meaning that all analysis will be conducted comparatively to a base case school. The distribution of students between the schools is shown in Figure 4.3.

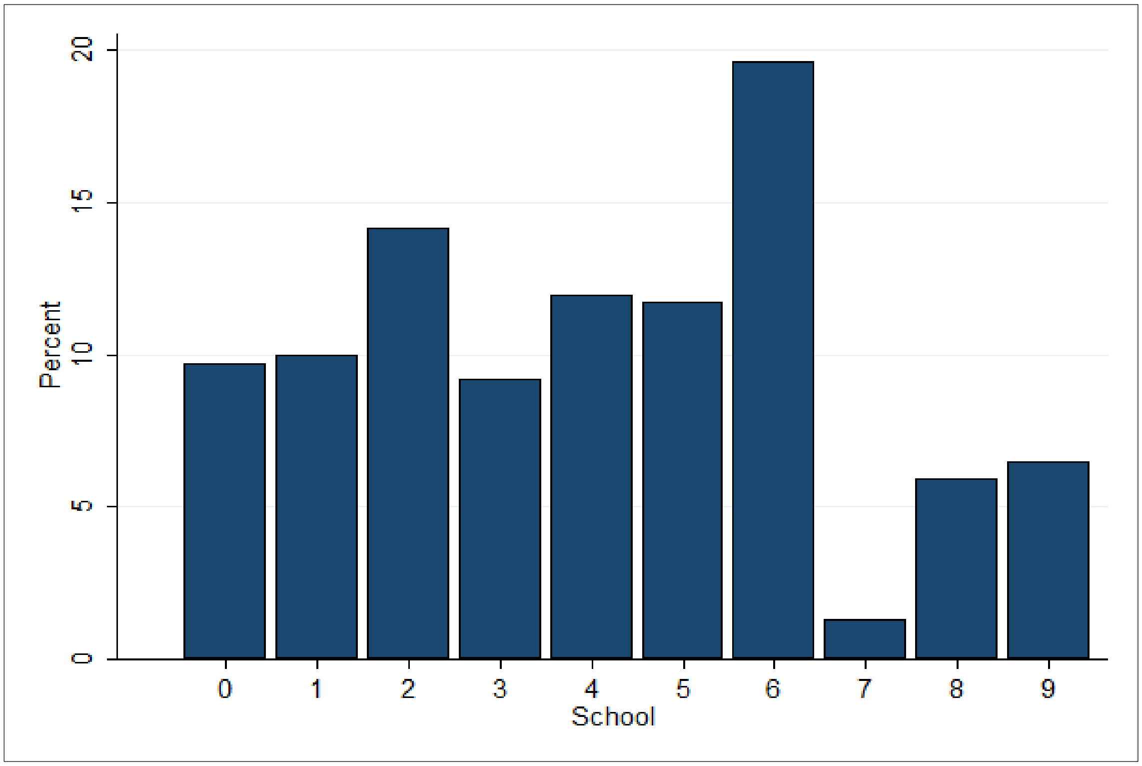


Figure 4.3 – Distribution of students by school

Figure 4.3 shows the variation between the school numbers within the data set. School 0 corresponds to the base level school, which has just under 10 per cent of students enrolled. School 6 has the largest number of student enrolments, with around 20 per cent of students. School 7 in contrast has around 2 per cent of enrolments. It is likely that retention rates and any effect of the EAS will vary between schools, making this an important variable to include in modelling.

The grade variables capture the weighted average of the grades attained. The distributions of the grades are presented in Figure 4.4 over page. The average numbers of units completed are low because of the number of students who do not obtain a particular grade. In other words, the averages will be dragged down by students who discontinue early and do not complete their studies. The estimated standard deviations around these means are large, indicating large variability within the grades attained. Visually, the distribution of the withdrawn grade outcome is presented in Figure 4.5 over page.

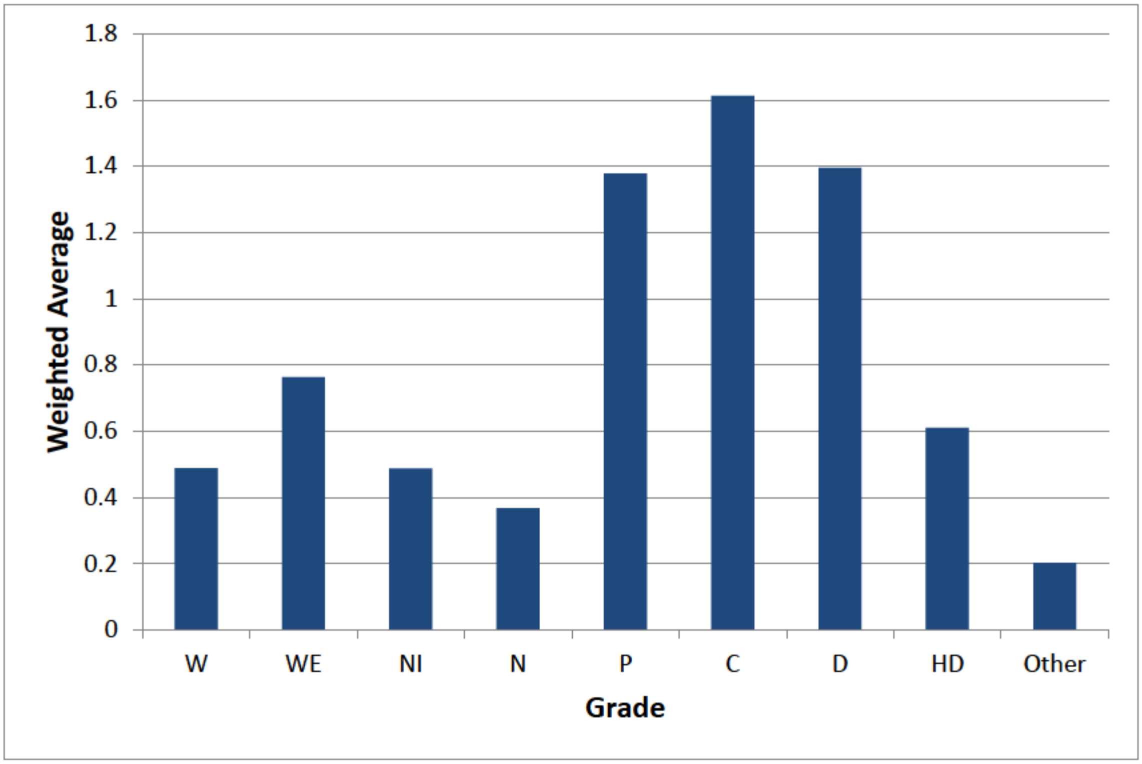


Figure 4.4 – Grade distribution weighted average

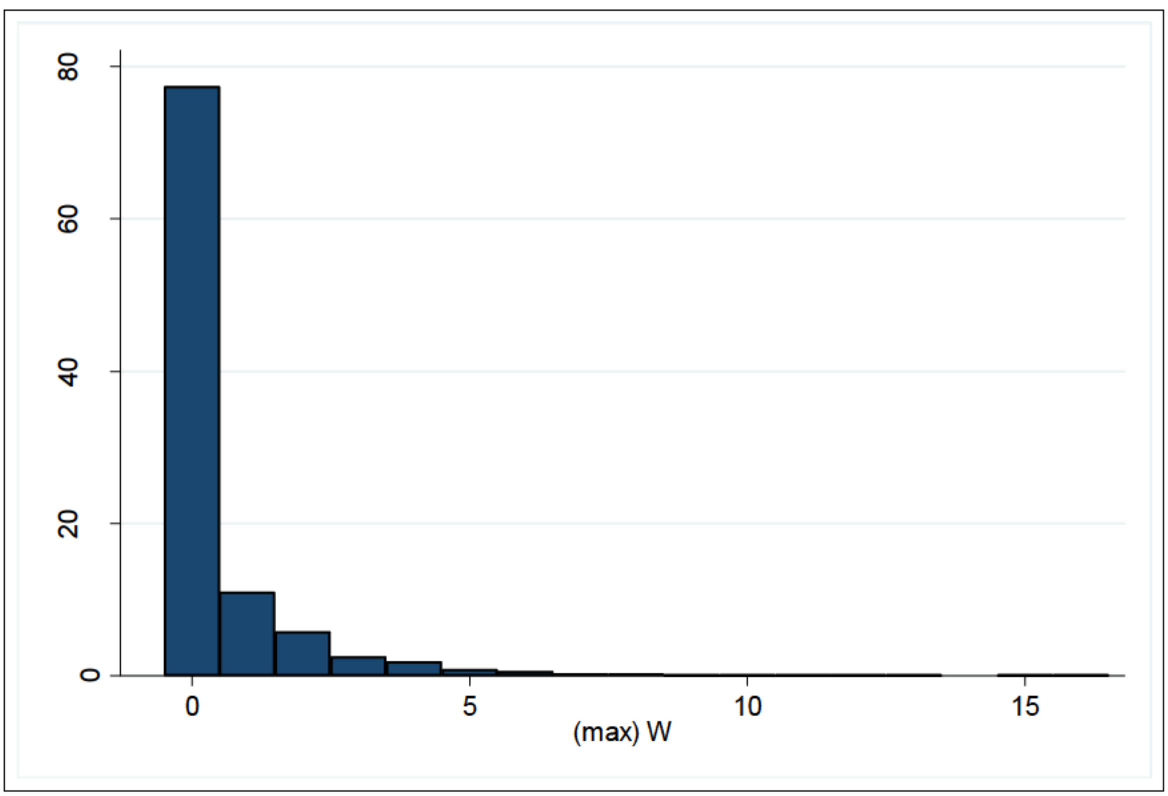


Figure 4.5 – Withdrawn grade distribution

The significant skew in the grades attained means statistical models could be sensitive to assumptions of normality.

Finally, units enrolled captures the weighted average number of units a student enrols in each teaching period. Weights are assigned relative to the length of enrolment. The distribution is slightly skewed to the right due to the peak around one unit per teaching period.

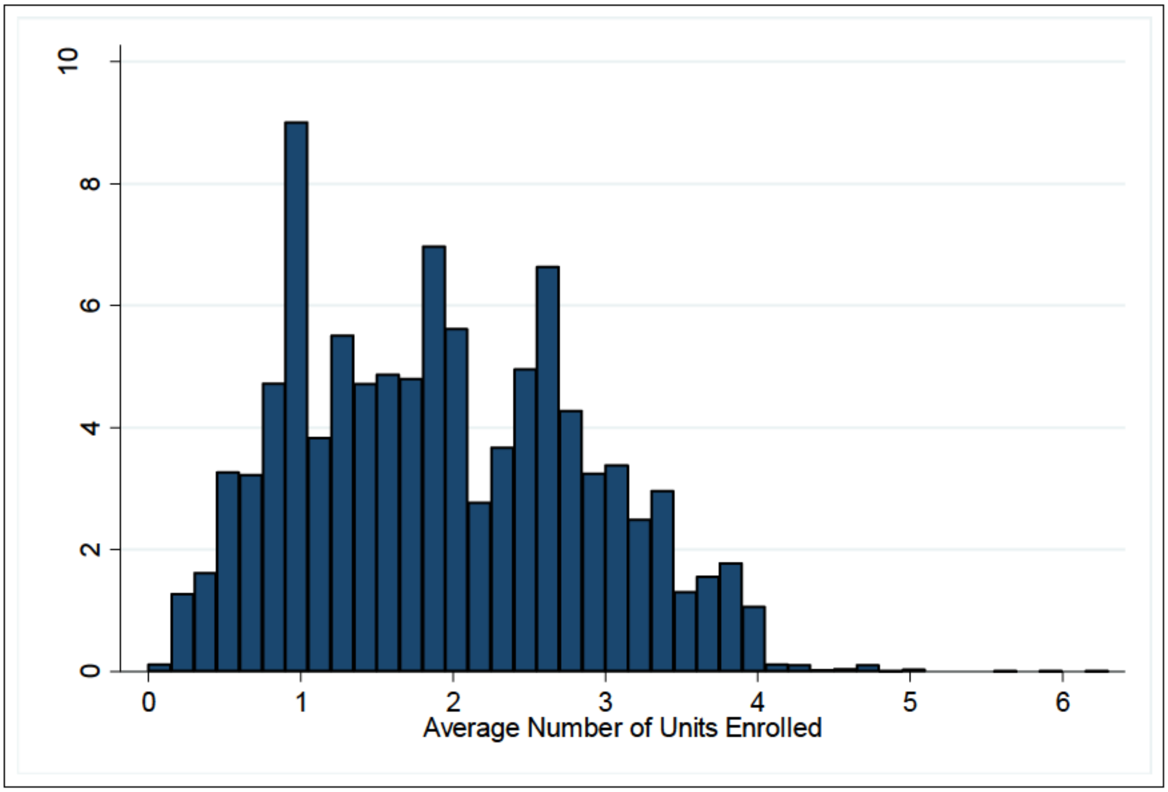


Figure 4.6 – Distribution of units enrolled

The skew is likely to occur due to the significant number of students studying part-time, taking a reduced workload of between one or two units per teaching period. Additionally, the number of units per teaching period captures students who take periods of inactivity, where zero units are undertaken.

4.5 Temporal overview of the data

As discussed in Chapter 2, the timing of events is critical to understanding the student retention problem. In section 4.5, several types of variables are discussed in detail with respect to temporal effects. These are discontinuation variables, demographic variables, institutional variables, student performance, workload variables and the early alert system variables.

4.5.1 Discontinuation variables

One way of viewing the temporal effect is to analyse the timing of when students discontinue their studies. From the sample data set, Figure 4.7 is constructed of when students have lapsed or discontinued their studies over the data capture period from 2011 to 2014.

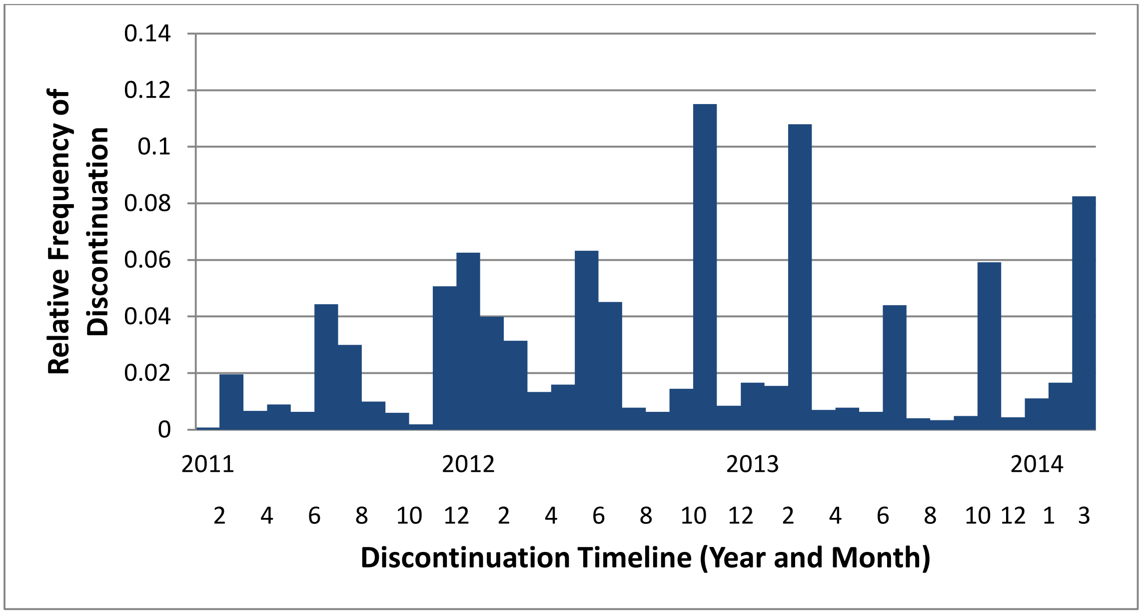


Figure 4.7 – Timing of discontinuation

The timeline shows distinct peaks and troughs occurring at regular intervals over time. The timing of the peaks and troughs correspond to the beginning and ending of teaching periods. In the context of when students decide to leave the institution, this makes sense. Students are likely to discontinue at the start of the teaching period with proactive decision-making. This is when students weigh up the future time spent studying versus time spent elsewhere.

A limitation of Figure 4.7 is that the starting times of students throughout the data set varies. Some students enter the set during intakes in 2011, while other enter during university intakes in 2013. To adjust for this, each student's starting time is left adjusted, where the first week of commencement is time period 1. Each subsequent week is coded as an additional week on this timeline. The plot of adjusted data on the discontinuation timeline is shown in Figure 4.8.

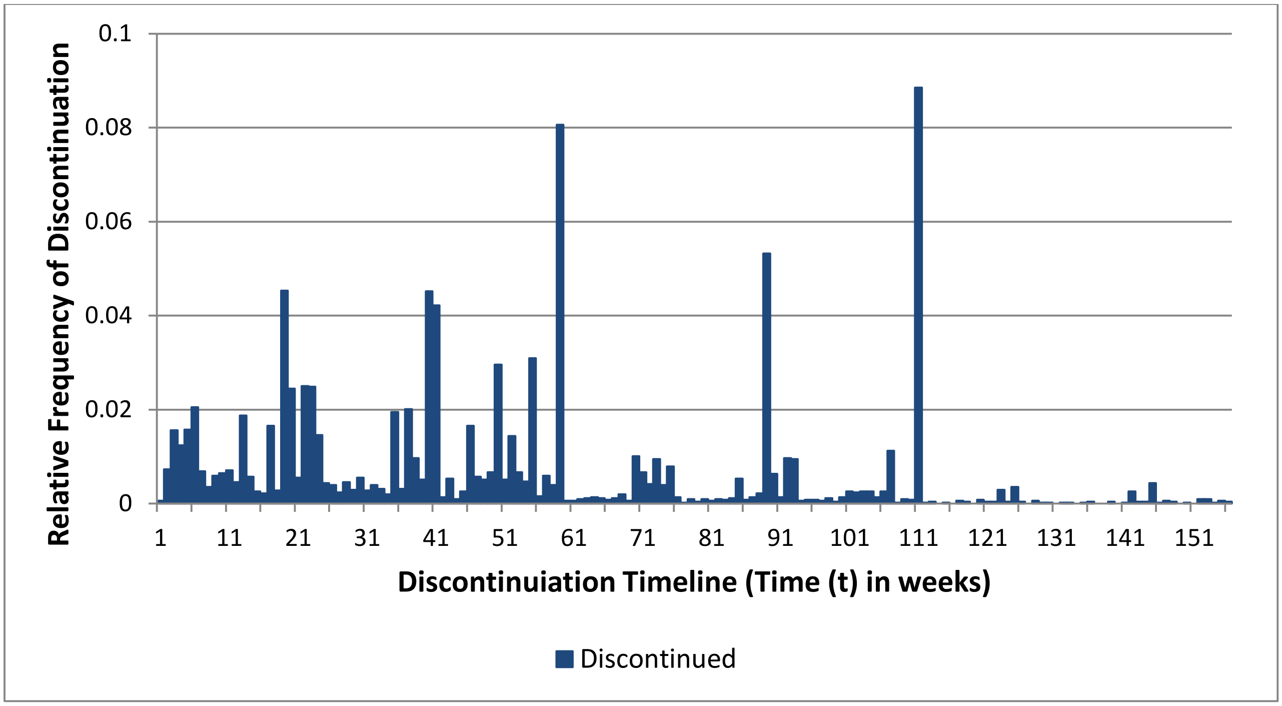


Figure 4.8 – Adjusted timing of discontinuation

The graph shows several distinct peaks at weeks 19, 41-42, 59, 89, and 111. The first teaching period from commencement of enrolment generally includes 12 weeks of teaching, two weeks of mid-trimester break and two weeks of examination. This means that the first peak corresponds to shortly after the first teaching period has concluded and would align with students results being released. It can be inferred that the other peaks also correlate to the beginning or end of teaching periods, the most likely time for students to make the decision to discontinue.

4.5.2 Demographic variables

Three demographic variables are captured in the dataset that are used for analysis. These are *gender*, *age* and *ATSI* status. Analysing *gender* over time in Figure 4.9 shows that the proportion of female to male students captured in the data set is not strictly constant. Initially, the proportion of female students is just above 66 per cent. As time progresses, there is a distinct upward trend with a maximum value 69.25 per cent occurring around week 50. The variation that occurs within this pattern indicates that there is a difference between male and female discontinuation rates, with more female students remaining enrolled longer than male students.

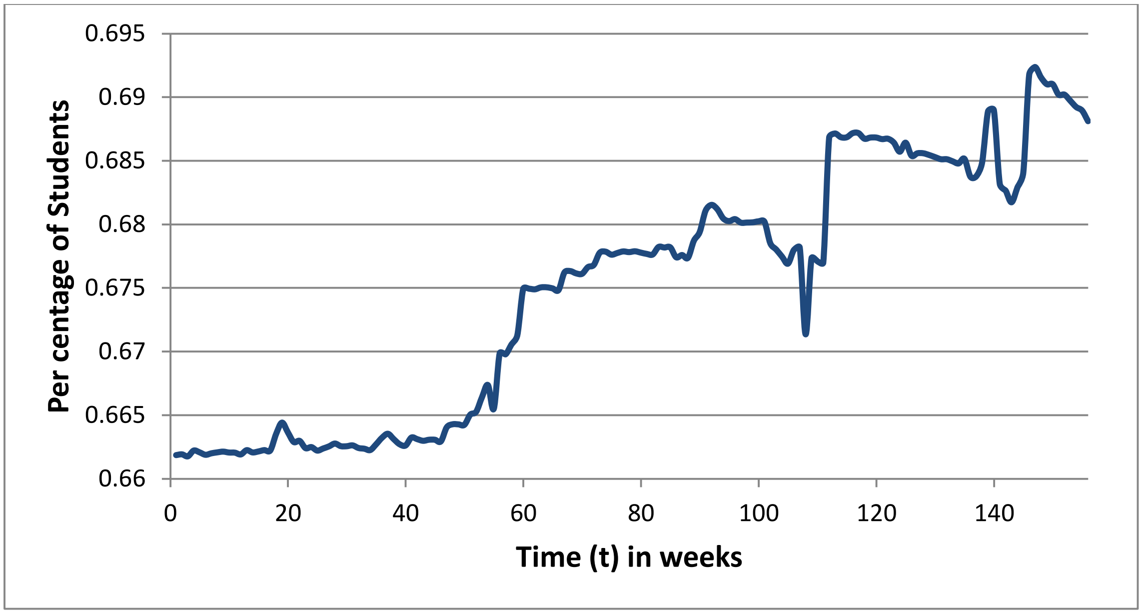


Figure 4.9 – Proportion of female students over time

It is expected that in the temporal models calibrated in chapter 3 and tested in chapter 7, *gender* could have a significant relationship to measures of student retention.

The second demographic variable is *age*. Analysing the average age of students over time indicates that younger students are enrolled for longer, causing a decreasing trend over time.

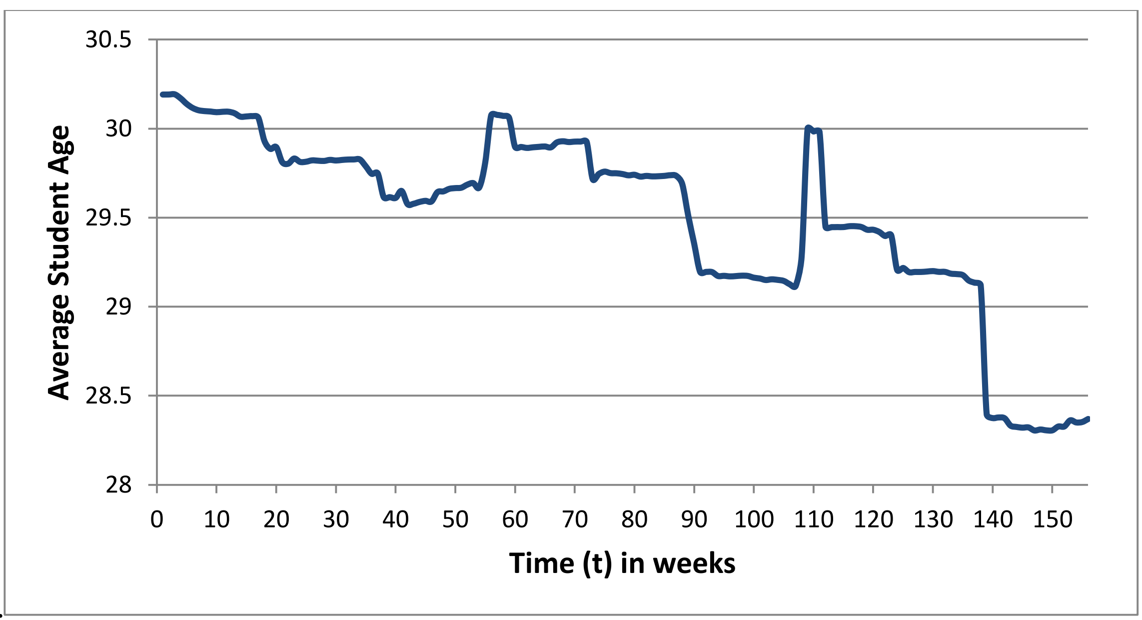


Figure 4.10 – Average student age over time

While the magnitude of the change is not large, the initial average age of 30.2 years appears different from the final average age of 28.4 years. This may indicate one of two things. Firstly, younger students are more likely to remain enrolled for longer. The second is that different courses undertaken by students have different age profiles. This means that both scenarios need to be accounted for in the models for analysis.

The third demographic variable is Aboriginal and Torres Strait Islander status (*ATSI*). Students who identify as ATSI are able to access specialist support resources to assist with their studies. Plotting the proportion of ATSI students over time reveals a relatively stable proportion.

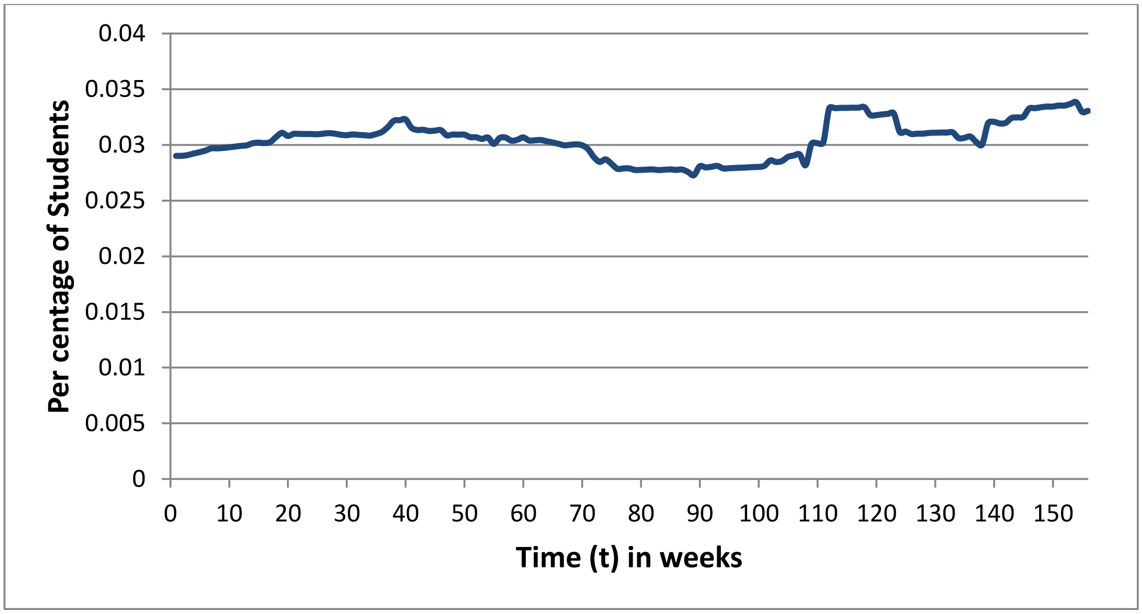


Figure 4.11 – Proportion of ATSI students over time

The initial proportion of ATSI students is around 3 per cent and by the end of 156 weeks (3 years), the proportion is slightly higher at around 3.4 per cent. The slight increase over time may indicate that the retention rate for ATSI students compared to non-ATSI students is different.

4.5.3 Institutional variables

The institutional variables capture characteristics of the relationship between the institution and the student. The key measures described include fee type, prior studies indicator, on-campus indicator and course type. The fee type variable captures the three main types of student tuition fees collected by the institution. The majority of students defer payment of tuition fees through the government-run Higher Education Loan Program (HELP). The second fee type is domestic fee paying students. The third type is international fee paying students. Figure 4.12 plots both domestic fee and international fee paying students together, from which trends in domestic HELP students can be inferred.

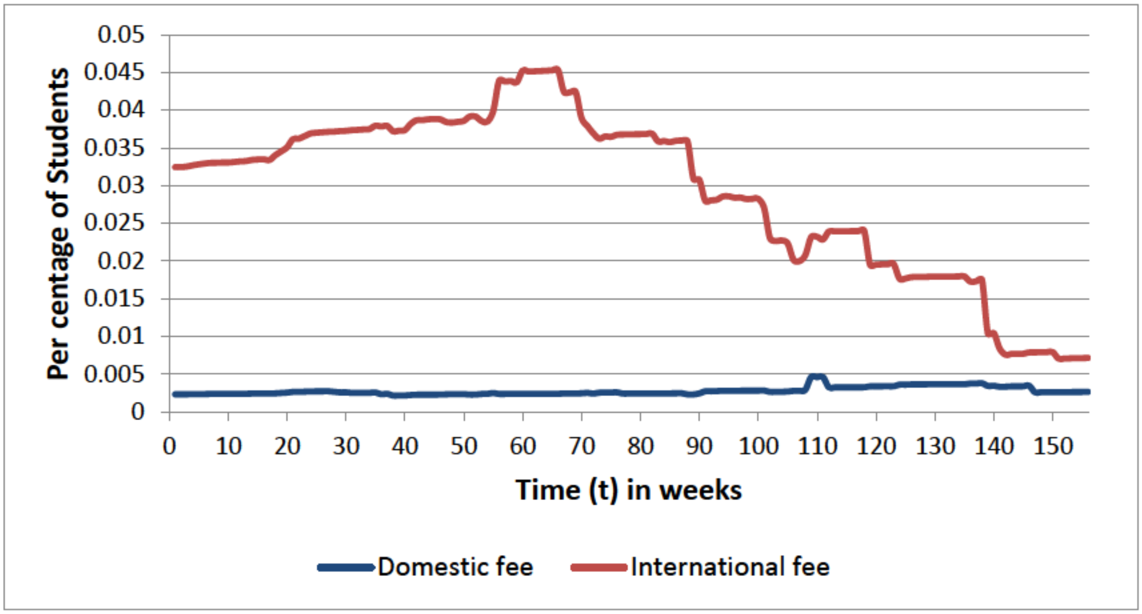


Figure 4.12 – Fee types over time

The graph shows that the proportion of domestic fee paying students remains relatively constant over time. International students, however, show an increasing proportion until around week 65, which then decreases over time. The increasing proportion indicates that international students are not discontinuing their studies at the same rate as domestic HELP students. The decreasing trend after week 65 indicates that fewer international students are captured in the data set. This trend is a combination of discontinuation and completion student outcomes. This change over time indicates that detailed temporal analysis is required to understand which of the two outcomes it is.

The next institution variable captures students who have undertaken some prior studies before enrolling at the institution. Figure 4.13 shows that at initial enrolment, around 17 per cent of students have some record of having undertaken prior studies. Relative to the remaining student cohort, the per cent increases over the first 60 weeks of enrolment. This indicates that students with prior study experience are not leaving the institution at the same rate as students without prior study experience.

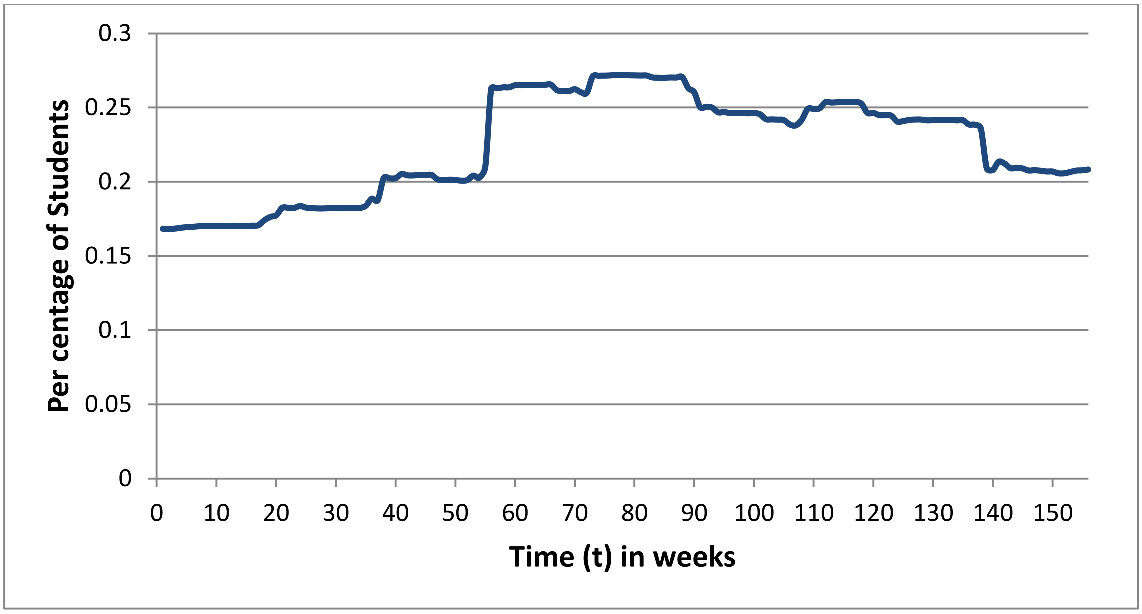


Figure 4.13 – Students with prior study experience over time

After week 60, there is a slight decrease in the per cent of students with prior study experience. The decreasing trend could either indicate a change in the rate at which students discontinue or complete.

The on-campus variable captures the per cent of students who study at the institution on-campus and access on-campus facilities. The remaining students are considered off-campus online students. Figure 4.14 over page indicates that initially, just over 25 per cent of the students studied on-campus. The per cent of on-campus students increases over time, indicating a difference in the discontinuation and completion rates compared to off-campus online students. One limitation of the variable is that a student deemed to study on-campus remains an on-campus student throughout the data set. In reality, students have the ability to change how they study between on-campus to off-campus online and vice versa. The assumption required to include this variable in the dynamic model is that the rate and timing of students changing from on-campus to off-campus is cancelled out by off-campus online students changing to on-campus study.

The last institutional variable of interest is the course of study. Most students are enrolled in a bachelor degree, which is used as the base case for analysis. The other four courses offered at the institution include diploma, advanced diploma, bachelor (graduate entry) and bachelor (honours).

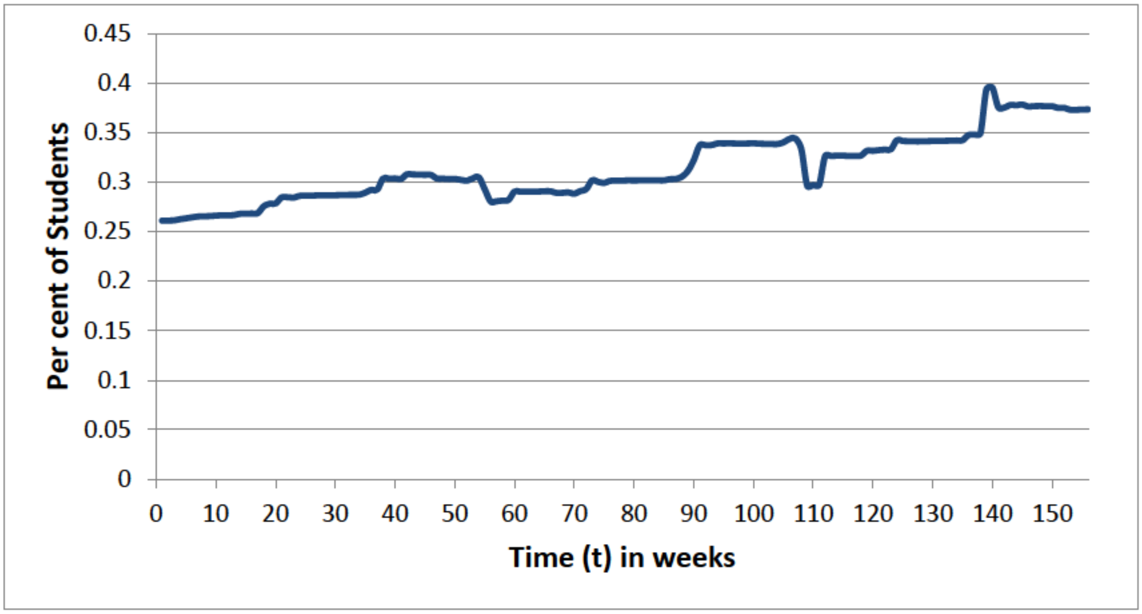


Figure 4.14 – On-campus students over time

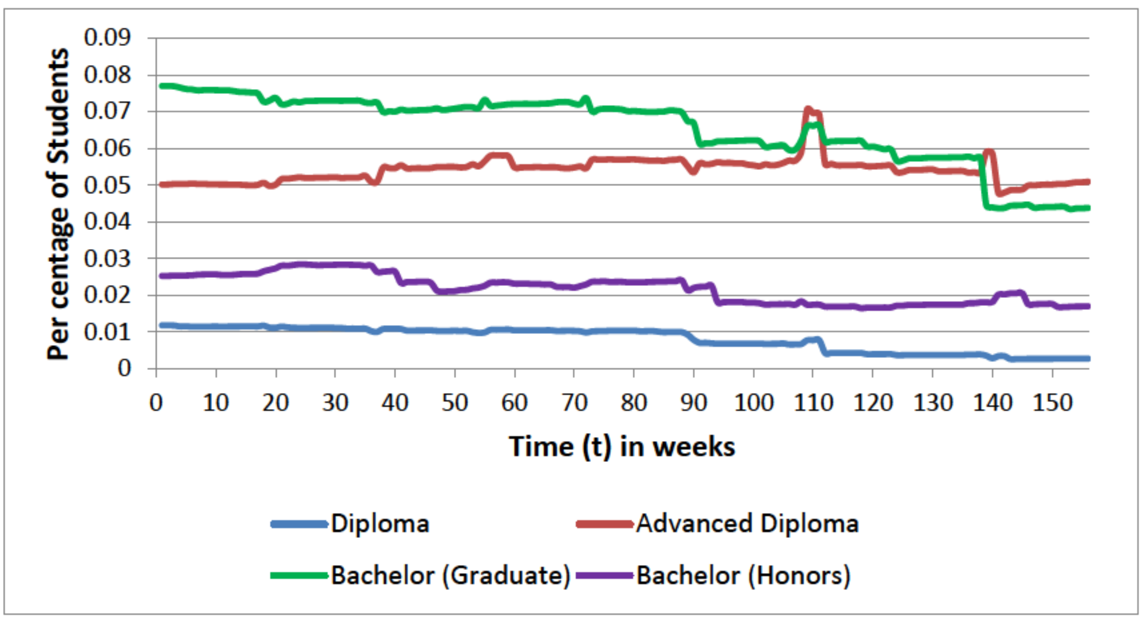


Figure 4.15 – Course type over time

Figure 4.15 shows initially, 1 per cent of students enrolled in a diploma, 5 per cent in an advanced diploma, 7.7 per cent in bachelor degree via graduate entry and 2.5 per cent in a bachelor degree with honours.

The proportion of students in diploma courses stays relatively constant until week 90, after which the proportion decreases over time. This is logical, given that diploma courses only require eight units to be completed, which in a full-time study load, takes one year of study. The proportion of advanced diploma and bachelor honours remains relatively constant over the three years of data capture. The most noticeable change occurs for bachelor students admitted by graduate entry. The proportion decreases over time, most rapidly in the past week. The final proportion of graduate entry student enrolled at the end of three years is just above 4 per cent. It is expected that some of this effect may be due to early completion of qualifications. This may also be capturing an issue with increased discontinuation rates compared to bachelor students.

4.5.4 Student performance and workload variables

Several variables capture aspects of the student performance and workload. These are the grades that students attain while studying, reflecting assessed classroom performance. The variable is the workload the students undertake, that is, how many units they attempt in any given teaching period. Grades can be broken into two main categories: negative grades which indicate students have failed to progress after enrolling in a unit of study; positive grades indicate a level of completion within a particular unit of study.

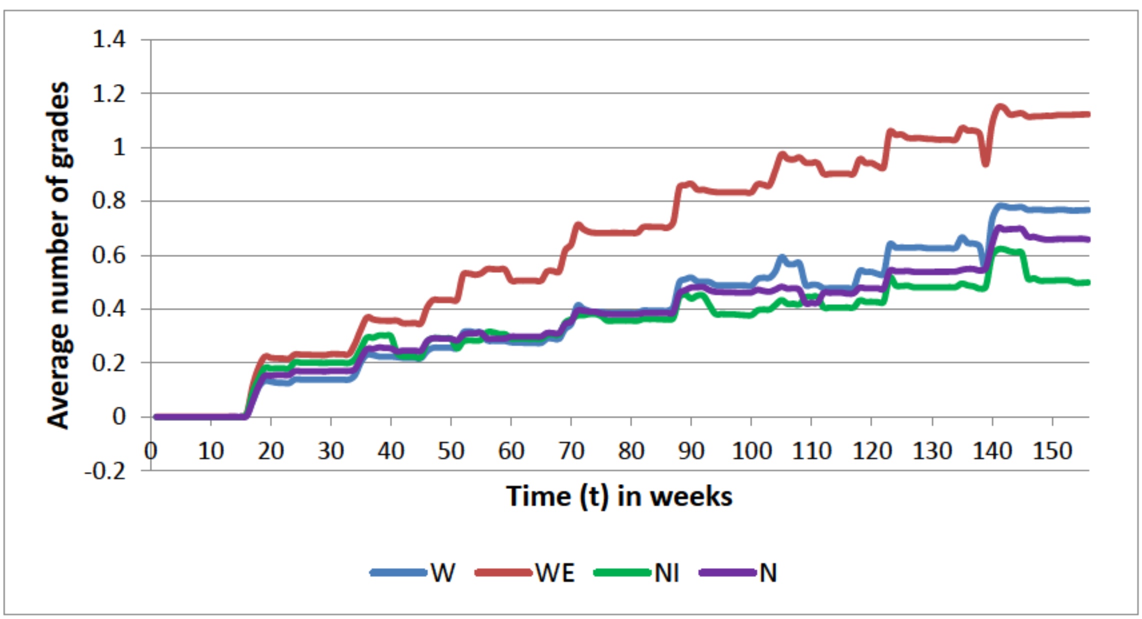


Figure 4.16 – Average negative grades over time

Figure 4.17 captures the four negative grade outcomes, withdrawn (W), withdrawn early (WE), fail incomplete (NI) and fail (N). The figure indicates positive trends over time which is logical, indicating that more of these grades are attained the longer a student is enrolled. The withdrawn early trend however diverges from the other three, indicating that students attain this grade on average more than the other negative grade outcomes. In part this is logical, since a student who withdraws from a unit faces neither academic nor financial penalty. This captures in part students' willingness to 'game the system' and try units out before they are academically and financially committed.

The average numbers of positive grades is presented in Figure 4.16, which captures pass (P), credit (C), distinction (D) and high distinction (HD) grades. The order represents the increasing difficulty of attaining these grades, with high distinctions the highest level of achievement within a unit.

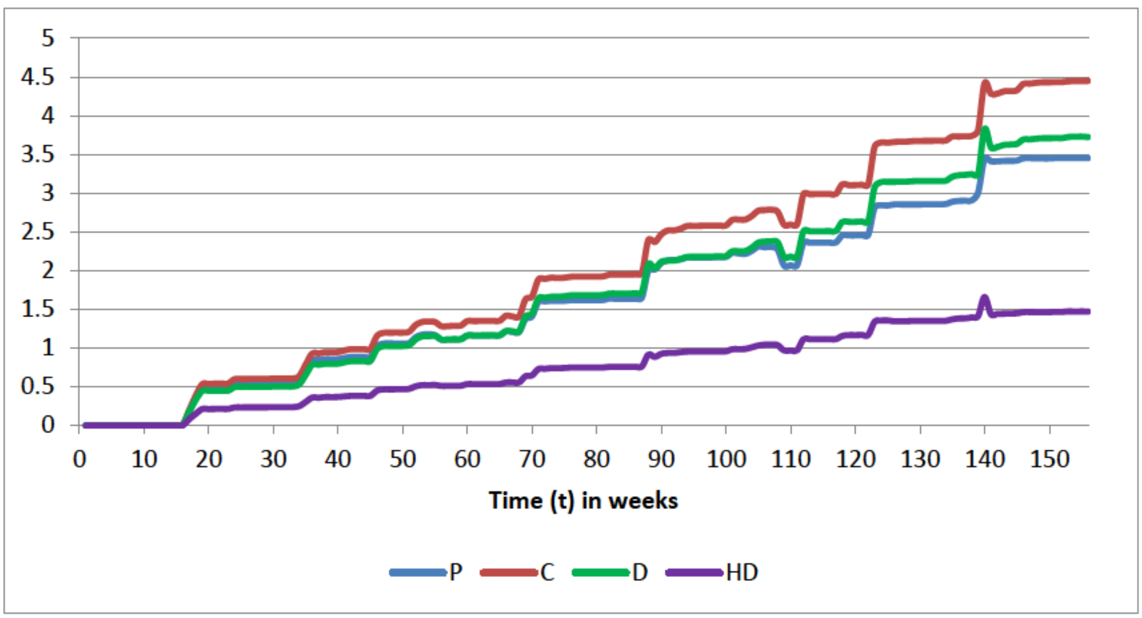


Figure 4.17 – Average positive grades over time

The chart indicates that compared to negative grades, positive grades occur more often. For example, at the end of three years, students on average had attained 4.5 credit grades. Compared to withdrawn early grades, after three years, students had on average around 1.1. The more difficult to attain high distinctions occur the least frequently, with the average student attaining 1.5 high distinctions after three years of study.

The average number of units attempted is a suitable measure of workload undertaken by students.

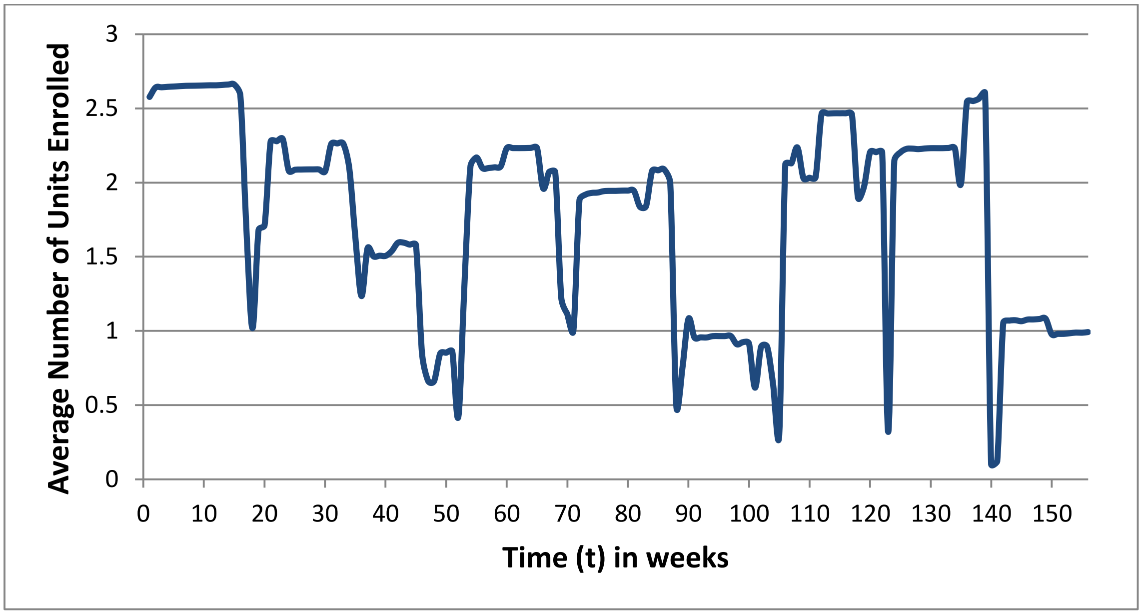


Figure 4.18 – Average number of units enrolled over time

Figure 4.18 shows a series of troughs which separate teaching periods undertaken. The first teaching period which runs until around week 18 from commencement of enrolment, has an average number of units enrolled of around 2.6 units. Troughs at weeks 18, 36, 52 71, 88, 104, 123 and 140 capture the periods in between teaching periods where students would take holidays. There are still some students enrolled during these periods, however, due to enrolling in yearlong units, or enrolling in courses that follow non-standard academic calendars. An interesting observation can be made with respect to the teaching periods that occur between weeks 37 to 51, weeks 89 to 104 and past week 140. The average number of units attempted during these teaching periods is less than the other teaching periods. This captures the third trimester which is a teaching period added to the academic calendar from 2012 onwards. During this time, many students will opt out of studying for a teaching period to work. As such, the average number of units undertaken by students is significantly less.

4.5.5 Early Alert System variables

The early alert system is a key focus of this study. The system as explained in previous chapters identifies students deemed at risk of disengaging from their studies on a daily basis. Three key measures can then be used to understand how the system functions. The first is the total number of identifications made over time. The second is the total number of students identified over time. The third and final combines the two to determine the average number of identifications per student over time.

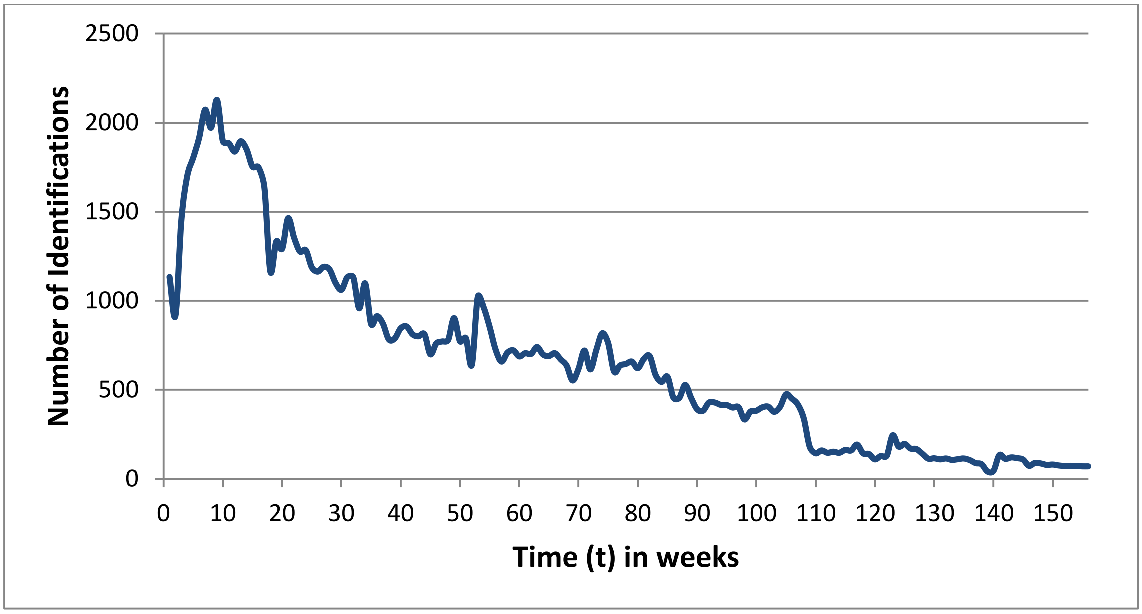


Figure 4.19 – Number of identifications over time

Figure 4.19 presents the number of EAS identifications in any given week. The chart indicates an initial low number of identifications that then increases to peak in week seven to nine. This indicates that most identification occurs during the first teaching period in which students are enrolled. It then decreases gradually over time. This is evidence that the EAS might be identifying students in need. The first teaching period forms the foundations of students' future studies, so early identification may assist students in being more aware of support available.

One limitation of this metric is that it does not account for the number of people identified. It may be possible that the same students are identified many times within a single week. To address this, Figure 4.20 plots the number of students identified by the EAS in any given week.

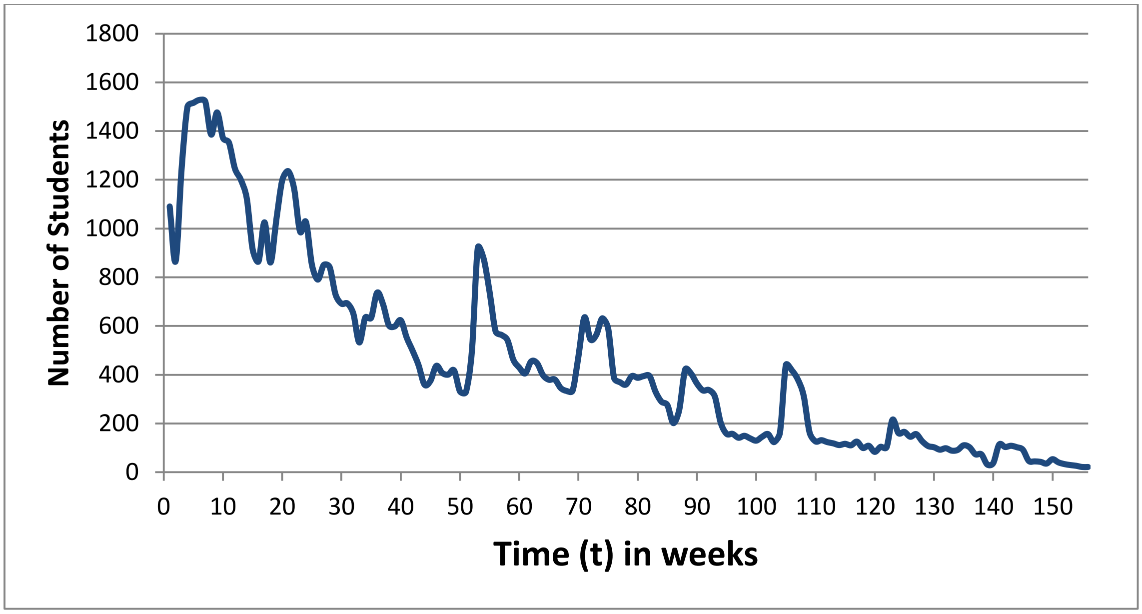


Figure 4.20 – Number of students identified by EAS over time

Figure 4.20 has a similar distribution, showing that the first teaching period has the most number of students identified by the EAS. This decreases over time with students in their third year of study being least identified. A key difference between the frequency of identification in Figure 4.19 and the number of students identified in Figure 4.20, is the spikes that occur in Figure 4.20 at weeks 21, 53, 71, 88 and 105. These peaks correspond to the start of teaching periods after the first teaching period. This is logical given the EAS evaluates a number of variables which change state at the start of each teaching period. As such, the system is likely to pick up a more diverse group of students.

Dividing the frequency of identifications by the number of students identified yields the average number of identifications per student in any given week. The results are plotted in Figure 4.21. The results show a distinct pattern over the first 108 weeks of enrolment where the average number of identifications per student increases as the teaching period progresses. At the beginning of the next teaching period the average number of identifications per student resets back to a value close to one. This indicates a key property of the EAS in terms of how it is functioning. The system identifies the same student more times as the teaching period progresses.

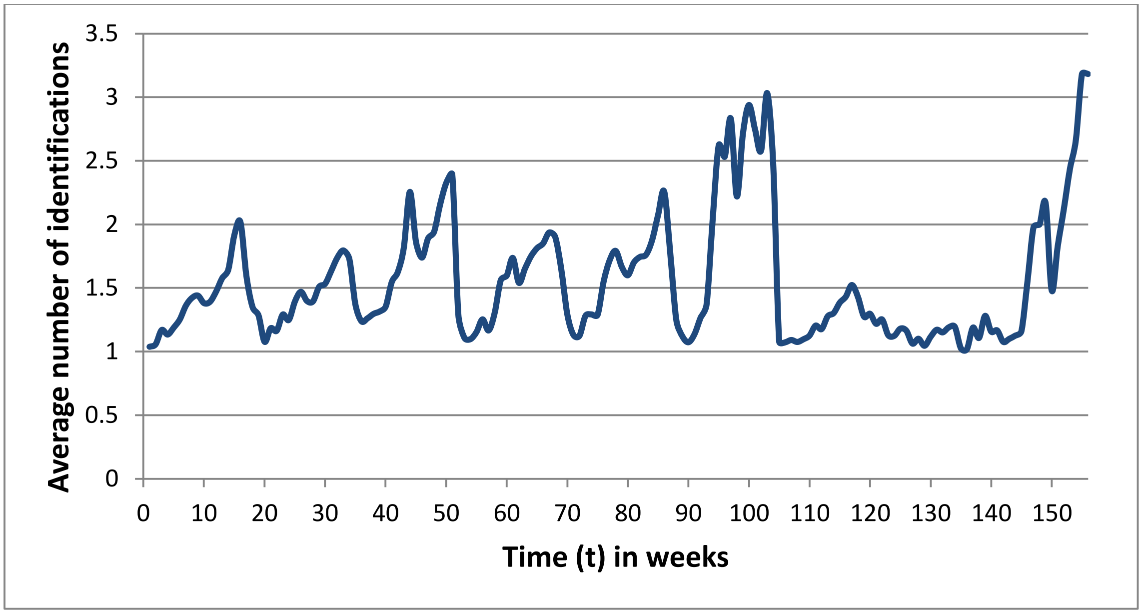


Figure 4.21 – Average number of identifications per student over time

The results may highlight a potential area of important analysis, which is the number of identifications per student. More importantly, does any EAS effect on student outcomes vary with the number of times the student is identified? This in part is covered in the model specifications covered in chapter 3, where varying severity levels of EAS identification are specified. Given the results in Figure 4.21, it can be expected to see some significant differences between the results based on the number of times students are identified.

4.6 Chapter summary

According to the descriptive statistics, the average student in this research sample is female, still enrolled after 594 days. She has not studied previously, not identified as being either an international or ATSI student. She enrols on average in two units per teaching period off-campus and defers her fees through the Australian government HELP program. On average, she will have attempted eight units with a grade distribution of one pass, two credits, two distinctions, one high distinction, one withdrawal and one withdrawal early.

Chapter 5 - A Likelihood Approach to Student Retention Analysis

5.1 Introduction

One approach to understanding student retention is the estimating the probability of events of interest. This chapter analyses the probability of three states the student can exist in. The first and primary state is that the student is enrolled in the institution. This is the base state for students and captures all students still progressing with study in their chosen course. The second state is a student who has chosen to discontinue their studies. As outlined in chapter 4, this captures students who have either formally or informally lapsed their study. This is the primary variable of interest in student retention analysis and this chapter. The final state students can attain is completion of qualifications. This is the ideal end state for students and the institution alike.

Tertiary institutions have increased their focus on early identification of students in need of support, targeting of support services and understanding student enrolment patterns (Aguilar, Lonn, and Teasley, 2014; Arnold and Pistilli, 2012; Arnold, Tanes, and King, 2010; Méndez et al., 2014; Tanes, Arnold, King, and Remnet, 2011). This has resulted in a multitude of both in-house and commercial information systems used to target student support (Arnold, 2010; Blackboard, 2014; Desire2Learn, 2013). The significance of the effect these systems have on student learning and education outcomes is an ongoing debate with some researchers finding significant results (Arnold and Pistilli, 2012), while other researchers are more sceptical about the conclusions reached (Straumsheim, 2013). This chapter contributes to research in this area in two key ways. The first is identifying the factors that affect student retention using logistic regression and the second is through estimating the likelihood of student outcomes factoring in the effect of an EAS.

5.2 Method of analysis

The multinomial logistic model estimates the likelihood of the enrolment outcomes in order to examine the factors affecting the enrolment status of students. This method has been used previously in multiple studies (Fike and Fike, 2008; Jones-White et al., 2010; Stratton et al., 2007; Waddington and Nam, 2014) to understand various aspects of EAS and student retention in tertiary institutions. Using multinomial logistic regression, the probability of the three enrolment states (discontinued, completed and enrolled) are estimated. 16,124 student observations captured from 2011 to 2014 are used to estimate the model. The model includes four main classes of variables: *demographic*, *institution*, *student performance* and *workload*. The demographic variables are *gender*, *age* at commencement of study, and Aboriginal and/or Torres Strait Islander status (*ATSI*). The institutional variables are variables which stay static during a student's enrolment at the institution, which includes *fee category*, the completion of *prior studies*, if the student was studying *on-campus*, the *course type* the students undertook and the *school* in which the course was taken. The performance variables included the *grades* of the student, including results which contribute to course progression and units that indicate a failure to progress. The level of *workload* undertaken, measured by the average number of units per teaching period, allows for varying rates of course progression. Finally, two measures for the EAS are estimated. One measure divides students into two groups, not identified and *identified* by the EAS. The second EAS measure captures the severity of risk, using the number of times a student was identified by the EAS. A student with *low severity* was identified 1 to 4 times during their study; *medium severity* captures students identified 5 to 9 times during their study; *high severity* captures students identified 10 to 19 times during their time enrolled; *very high severity* captures students identified more than 20 times by the EAS during their time enrolled. The EAS measures are estimated in separate models with complete statistical output for both models presented in Appendix C. Given the results for all variables except EAS measures are statistically similar, only one set of results are presented in the discussion of results.

The likelihood estimations are presented using Relative Risk Ratios (RRR). Interpreting the estimated coefficients is done relative to the base case of a student being enrolled. If the estimated coefficient is less than 1 for the discontinued state, this indicates the probability of the event occurring is less likely than the student being enrolled. If the value of the RRR is greater than 1, this indicates that the most likely state will be discontinued. In the multinomial model where both discontinuation and completion are estimated relative to the enrolled state, the most likely outcome is the largest RRR or the base case of enrolled if both discontinued and completed RRR are less than 1.

5.3 Overall results

The overall results presented in Table 5.1 indicate highly significant estimates. The likelihood ratio chi square value is 7431.43, which is statistically significant at the 1% level.

Table 5.1 – Multinomial logistic regression approach: overall results

Number of observations	16124
LR χ^2	7431.43
Prob > χ^2	0
Pseudo R^2	0.2747

The pseudo R^2 value is 0.2747. This has no directly interpretable meaning; however in context, it supports the conclusion of a statistically significant model.

5.4 Demographics effects

The demographic variables capture *gender*, *age* and *ATSI* status. The constant term from the estimated models are also included. The results are presented in Table 5.2.

Table 5.2 – Demographic results

	Discontinue		Complete	
Variable	RRR	Std. Error	RRR	Std. Error
Gender	1.104[b]	0.048	1.203[b]	0.109
Age	0.960[a]	0.007	1.062[a]	0.017
Age^2	1.001[a]	0.000[d]	0.999[a]	0.000[d]
ATSI	0.789[c]	0.096	0.796	0.213
Constant	1.253[c]	0.146	0.007[a]	0.002

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The model indicates that on the basis of gender, female students are 10 per cent more likely to discontinue than their male counterparts. Additionally, female students are 20 per cent more likely to complete compared to male students. Both effects are significant at a 5% level. To create parity between genders, the result indicates that female students may require more support to continue on with their studies than male students. This is contrasted by male students who may require additional support to complete their courses of study. The results may reflect underlying societal issues associated with gender inclusion in higher education.

The age variable shows a significant non-linear relationship in the two enrolment states. The probabilities for different ages are shown in Figure 5.1.

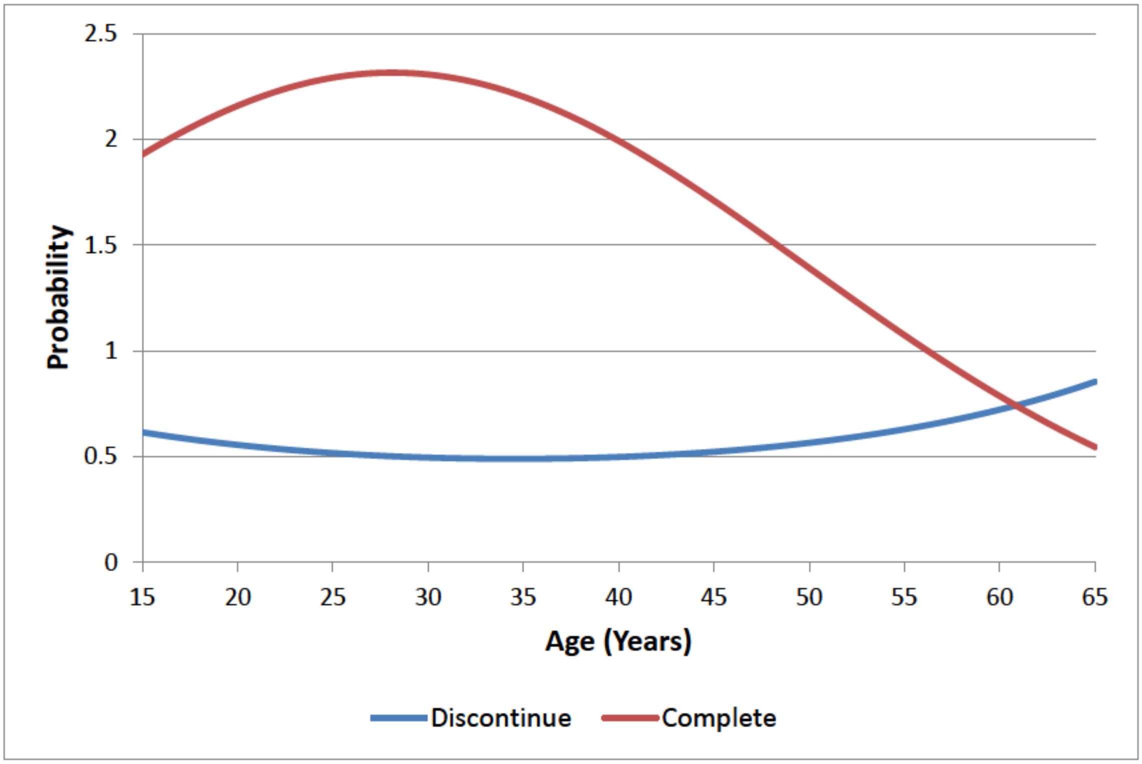


Figure 5.1 – Probability of enrolment states relative to age

The results show a small decrease in the probability of discontinuing over time, reaching a minimum point at 35 years of age. The probability of discontinuing then increases as students get older, however it never reaches the value of 1. This indicates that overall, the base case of being enrolled is more likely than discontinuing.

The results show that as age increases, the probability of completing also increases, peaking at a maximum probability around 28 years of age. The probability of completing then decreases over time and shows that mature age senior students have a relatively low probability of completing. The results are expected given the profile of the case study institution which has a relatively large mature student cohort.

Students who have identified as ATSI have a minor decrease in the probability of discontinuing compared to non-ATSI students. This effect is significant at the 10% level. This provides a promising result for inclusion of ATSI students. At the case study institution, specialist ATSI support programs are available. The result tentatively suggests that ATSI students are more likely to be retained compared to non-ATSI students.

5.5 Institutional effects

The institutional variables capture the relationship between student enrolment at the institution level, including fee type, prior study undertaken, study type (on-campus or off-campus online), course type and school enrolled in. The estimated RRR are presented in Table 5.3.

Table 5.3 – Institutional results

Variable	Discontinue		Complete	
	RRR	Std. Error	RRR	Std. Error
Domestic Fee	2.138[b]	0.805	0.497	0.410
International Fee	0.421[a]	0.072	4.161[a]	0.717
Prior Studies	3.010[a]	0.175	6.526[a]	0.589
On-campus	1.495[a]	0.101	0.531[a]	0.066
Diploma	1.254	0.213	7.813[a]	2.254
Advanced Diploma	1.491[a]	0.146	4.889[a]	0.982
Bachelors (Graduate)	0.672[a]	0.063	7.436[a]	1.396
Bachelors (Honors)	0.385[a]	0.063	33.491[a]	6.344
School 1	1.196[c]	0.114	0.146[a]	0.029
School 2	0.995	0.088	0.444[a]	0.073
School 3	1.300[a]	0.124	0.383[a]	0.070
School 4	1.250[b]	0.111	0.344[a]	0.059
School 5	1.219[b]	0.121	0.031[a]	0.009
School 6	0.966	0.083	0.341[a]	0.055
School 7	0.909	0.510	1.068	0.883
School 8	1.407[a]	0.161	0.852	0.143
School 9	0.908	0.102	0.518[a]	0.108

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Fee paying students are compared to the base case of domestic students who defer fees through the Australian governments Higher Education Loan Program (HELP). The results show that domestic students who pay fees upfront have a higher likelihood of discontinuing. Domestic fee paying students make up a small proportion of the overall student cohort. The results indicate that within that small group, the probability of discontinuing is higher than domestic HELP students. One possible explanation is that up-front fee paying students are more tentative in committing to university study. The up-front fee paying option comes with a discount on the fees and allows students to explore tertiary study at a decreased fee while deciding to commit. Further analysis through surveys within this small group of students would assist in understanding this effect.

International students are less likely to discontinue studies, which is in line with expectations. International students usually have to pay fees up-front to the institution and are charged according to the degree of study the student is entering. The upfront financial commitment coupled with cultural expectations on students completing can explain this effect. This is also confirmed by international students having a significantly higher probability of completing courses.

Students who have undertaken prior studies before entering university show contrasting estimates in Table 5.3. The results indicate students with prior studies are more likely to discontinue. This effect may be attributable to students with prior study experience having less need and less commitment to study than students without prior studies. The contrasting effect is that students with prior studies are more likely to complete their studies. This result is more in line with expectations, given that students with prior study experience will be better prepared for the tertiary environment. The result suggests that students with prior study may have varying probabilities of discontinuing and completing over time.

Comparing on-campus students to off-campus online students, on-campus students are 49.5 per cent more likely to discontinue their studies. Furthermore, on-campus students are 46.9 per cent less likely to complete their qualification. This is an unexpected result given that on-campus students have improved access to academics and support services. One explanation for this effect may be the effect of the on-campus residential colleges. One of the EAS triggers is residential college living arrangements. According to the EAS specifications, students living in residential college have an increased chance of disengaging. The model is likely to be capturing this effect as well with respect to discontinuing and completing. The EAS specification is covered in greater detail in chapter 7, where the source of this unexpected result is revealed in more detail. Given the contrary results, this effect needs to be explored in more detail to understand if on-campus students are more genuinely likely to discontinue, or if there are other factors contributing to the results found such as selection of the method of analysis.

Students at the case study institution can enrol in one of five types of courses: diploma, advanced diploma, bachelor degree, bachelor degree via graduate entry and bachelor degree with honours.

The base case used for comparative analysis is the bachelor degree, which constitutes the majority of enrolments within the university. The results show that students undertaking a diploma are significantly more likely to complete their qualification than bachelor students. This is expected given the length of a diploma is typically a year of full-time study, compared to three years for a bachelor degree. Students enrolled in an advanced diploma are 49% more likely to discontinue their studies than bachelor students. Advanced diploma students are also more likely to complete their qualification. This is expected given that the course is shorter (1.5 years at full-time study). Students admitted through graduate entry and into honour programs are significantly less likely to discontinue and more likely to complete. This is in line with expectations.

Using schools within the university is a broad way of factoring in inter-school differences and the different courses offered into the model. This controls inter-school differences to ensure model 5 produces valid estimates. The base case school was arbitrarily selected, with all other schools compared to it. The results show that students in schools 1, 3, 4, 5 and 8 have a statistically higher likelihood of discontinuing compared to the base school. Additionally, students in schools 1-6 and 9 are less likely to complete than students in the base school. The results indicate variations between the schools which is important to control for when estimating the effect of the EAS.

5.6 Student performance and workload effects

Two key measures capture learning environment. First, the grade distribution of the student captures the effect associated with attaining a specific grade. Secondly, the workload variable measures the weighted average number of units enrolled. Weights are assigned based on the time spent enrolled in units. The estimated relative risk ratios for the learning environment variables are presented in Table 5.4 over page.

The results show that all RRR are significant at the 1% level except for the estimated effect of a student failing a unit, which is significant at the 5% level. Students who withdraw, withdraw early, fail or attain an “other” grade for a unit have similar estimated outcomes. The estimated RRR are less than 1 for both discontinue and complete enrolment outcomes. This means that despite attaining one of these grades, the most likely outcome is that the student will continue to be enrolled. For the negative grades, the next likely outcome is that the student will discontinue with the least likely outcome being completing. This is an important finding, as it indicates that attaining a fail grade does not increase the probability of discontinuing.

Table 5.4 – Student workload and performance results

	Discontinue		Complete	
Variable	RRR	Std. Error	RRR	Std. Error
Withdrawn	0.739[a]	0.015	0.488[a]	0.047
Withdrawn Early	0.873[a]	0.014	0.592[a]	0.033
Fail Incomplete	1.118[a]	0.019	0.570[a]	0.068
Fail	0.941[b]	0.023	0.553[a]	0.039
Pass	0.705[a]	0.012	1.141[a]	0.021
Credit	0.722[a]	0.013	1.123[a]	0.019
Distinction	0.707[a]	0.013	1.094[a]	0.018
High Distinction	0.725[a]	0.017	1.134[a]	0.021
Other Grade	0.545[a]	0.050	0.716[a]	0.047
Workload	1.334[a]	0.036	2.736[a]	0.167

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The fail incomplete grade outcome is not in line with the other negative grades. If a student fails a unit by not completing necessary components of the unit, then the most likely outcome is that the student will discontinue. This is an important finding as this indicates that an early identifier of students at risk of discontinuing can be found within a single unit of study. If the student fails to submit compulsory work which would result in a fail incomplete grade, then the student should be contacted to be provided with support. It can be concluded that fail incomplete grades are predictive of students discontinuing.

Pass, credit, distinction and high distinction grades are all positive grades which contribute to course progression. The results from model 5 show that the RRR for completing are significantly greater than 1. This indicates the most likely outcome from attaining a positive grade is completing. The result is both logical and in line with expectations. The next most likely outcome is the base case of the students being enrolled. The least likely outcome is that a student will discontinue upon attaining a positive grade. Overall, the grade results show that to accurately estimate the likelihood of a student discontinuing, all unit results should be accounted for. From a design perspective, the EAS should estimate the true likelihood of a student discontinuing based on performance across all units, not just negative grade outcomes.

The workload variable is the weighted average of the number of units enrolled in throughout a student's enrolment. The variable is continuous and interpreted at the margin as a 1 unit increase. For example, if a student increased their workload from 2 to 3 units per teaching period, the likelihood of that student discontinuing increases by 33.4 per cent. Simultaneously, the student's likelihood of completing increases by 173 per cent. These results indicate that increasing workload has both a positive and negative effect. The most likely outcome of increasing the workload is a student completing. The next most likely outcome however is the student discontinuing. This is critical information that student support and an EAS could use to assess how changes in student workload is affecting the likelihood of either completing or discontinuing. While there are many other variables which make up the learning environment, the two tested in model 5 are incredibly useful in understanding the relationship between workload and unit performance on retention.

5.7 Early alert system effects

Two different model specifications are used to capture effects of the EAS on student outcomes. Model 5.1 uses a binary approach, where 1 indicates that a student was identified by the EAS at some stage of their enrolment, while 0 indicates that the student was never identified by the EAS. Model 5.2 uses a treatment level approach, which captures the number of times a student is identified by the EAS. The RRR for the EAS variables are presented in Table 5.5.

Table 5.5 – Early alert system results

		Discontinue		Complete	
Model	Variable	RRR	Std. Error	RRR	Std. Error
5.1	EAS Identified	1.064	0.052	0.705[a]	0.084
5.2	Low EAS severity	1.054	0.053	0.752[b]	0.090
5.2	Medium EAS severity	1.155[c]	0.087	0.441[a]	0.071
5.2	High EAS severity	1.538[a]	0.171	0.305[a]	0.059
5.2	Very high EAS severity	1.141	0.157	0.202[a]	0.047

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The results for overall EAS identification show no significant difference in the likelihood of discontinuing if a student is identified by the EAS. This inconclusive result makes it difficult to interpret any underlying effect. EAS identification does indicate that students have a significantly lower likelihood of completing. This does support the idea that the EAS is at least identifying students who may have issues with course progression. The results show that effects associated with the EAS are captured in the completion data and not the discontinuation data. This has important implications for how EAS efficacy is analysed. Discontinuation data in the logistic regression framework may not be suitable for determining changes in the likelihood of student outcomes.

Model 5.2 uses varying severity levels to capture the effect of the EAS. Low EAS severity students comprised of 34.32 per cent of the 16,124 observations sampled. Compared to the base group of students who were not identified by the EAS, there is no significant difference in the likelihood of discontinuing. There is a significant decrease in the likelihood of completing. The results do indicate that being identified at least once reduces the students' likelihood of completing. This supports the argument that completion data may be more important than discontinuation data in measuring EAS efficacy.

Medium EAS severity captures students identified by the EAS five to nine times during their enrolment. 17.55 per cent of the observations from the sample data fall into this category. Unlike previous results, medium EAS severity students are 15.5 per cent more likely to discontinue than students not identified by the AES. The increase is significant at the 10% level. This is the first instance of the EAS being associated with an increase in the probability of discontinuing. This indicates that the more times students are identified, the greater the likelihood of discontinuing, and that severity is a factor in identifying students in need of support. The likelihood of completing decreases further with medium EAS severity, indicating that students are less likely to complete than low EAS severity students. The results indicate that the EAS is correctly identifying students, however, it may take multiple instances of identification before the effect is pronounced.

High EAS severity captures students identified by the EAS 10 to 19 times and students in this category make up 11.52 per cent of the data set. The likelihood of discontinuing are 50 per cent greater for students in the high EAS severity category than students not identified by the EAS. The effect is significant at the 1% level. Additionally, the likelihood of completing decreases further from medium EAS severity. The effects captured strongly indicate that the EAS system is identifying students in need of targeted student support. For the system to capture this effect however, students need to be identified by the system multiple times.

Very high EAS severity captures the students identified 20 times or more by the EAS and should represent the students most at risk of discontinuing, however, the results show that there is no significant difference in the likelihood of discontinuing compared to students not identified by the EAS. One possible explanation for this is that for students to have been identified 20 times or more, they have needed to persevere more than the other EAS severity groups. The very high EAS severity category is capturing students who have been at-risk in the past; however, they are determined to remain at university and as such do not discontinue their studies. These students still have a significantly reduced probability of completing which results from having to re-take some units of study or receive additional support to complete which extends the students length of enrolment. This conclusion warrants further analysis which is only possible with data that extends over a longer time frame to capture students as they graduate. As such, this may be an area of important follow up analysis when more data becomes available.

5.8 Chapter summary

Overall, using the likelihood approach resulted in models statistically significant at the 1% level, capturing the likelihood of discontinuing, completing and being enrolled using *demographic*, *institutional*, *performance* and *workload* variables. For the demographic variables, gender was statistically significant which indicates that gender differences still need to be accounted for when understanding retention. Age has a non-linear relationship, where the probability of discontinuing decreases until 35 years of age before increasing again. ATSI students had a slightly lower probability of discontinuing than non-ATSI students, an effect significant at the 10% level, which may reflect the case study institution’s internal support programs for ATSI students independent of the EAS.

With respect to the institutional variables, fee category variables showed domestic students who pay fees up-front are more likely to discontinue than students who defer fees through the governments HELP program. International students are more likely to complete and least likely to discontinue. Students who have studied previously have both an increased chance of discontinuing and completing than students without prior study experience. On-campus students were more likely to discontinue and less likely to complete, an effect which needs to be explored in more detail to determine the reasons for this unexpected result. The results for both course type and schools were in line with expected results.

The learning environment variables captured the important relationship between the unit grades and workload to the student outcomes. The results showed that most of the negative grade outcomes neither increased the chances of discontinuing nor completing. The positive grade outcomes decreased the chances of discontinuing while increasing the chances of completing. The grade fail incomplete which captures students who fail to satisfy compulsory components within the unit assessments provided an important result. Unlike fail and the withdrawn grades, failing a unit incomplete was a statistically significant indicator of a student discontinuing. This indicates that compulsory in-class assessments may need monitoring as a way of early detection for discontinuing students.

Finally, the EAS results provided evidence of the system working. Estimating the overall effect of EAS identification, students identified by the EAS are significantly less likely to complete than students not identified by the EAS. The same effect was observed at increasing severity levels. The more a student was identified, the less likely the student was to complete their studies. The results were less pronounced on the discontinuation outcome. When comparing students overall, there was no significant difference in the likelihood of discontinuing between the two groups of students. At varying levels of EAS severity, there was a significant increase in the likelihood of discontinuing only when identified between 5 and 19 times. This indicates that the system may need to identify students multiple times to correctly identify students in need of targeted support with respect to discontinuation. In contrast, a student only needed to be identified once to be at-risk of not completing. This is an important finding in that measuring EAS efficacy may be better measured on changes in the probability of completing a course than on the probability of discontinuing a course. This has implications for EAS design where systems may need to be tested against both discontinuation and completion outcomes. For student support services associated with the EAS, this may also indicate that there may need to be varying approaches to supporting students. Some students may not be at-risk of discontinuing, but at-risk of not completing. While this means that the student will be retained in the system for longer, the outcomes for the student may not be ideal if support is not focused on supporting completion.

Overall, model 5 serves as a strong starting point to student retention analysis and understanding the relationship between the EAS and student retention. A limitation of this method of analysis is the inability to address for time varying aspects of student’s enrolment. For model 5, it is assumed that the effects identified remain constant throughout the students' enrolment. Pooled data was used from multiple years to estimate effects, which means that there may be some bias in the results when not accounting for temporal effects. In chapter 6, temporal effects are analysed by focusing on the students' length of enrolment as the dependent variable. This introduces the temporal correlation between improved student retention outcomes and increased length of enrolment.

Chapter 6 - An Ordinary Least Squares Approach to Analysing Factors Affecting Student Retention

6.1 Introduction

When analysing student retention, it is important to define and understand how the effect of improving retention can be captured. As explored previously, one method is to determine the probability of discontinuing studies. That is, improving retention should decrease the probability of discontinuing. However, this ignores the temporal component of improving student retention. That is, improving student retention should also translate to increased enrolment time.

In this chapter, the students' length of enrolment becomes the dependent variable. It is expected that a positive effect of the EAS will correlate to students' length of enrolment increasing. A general model is presented in Figure 6.1.

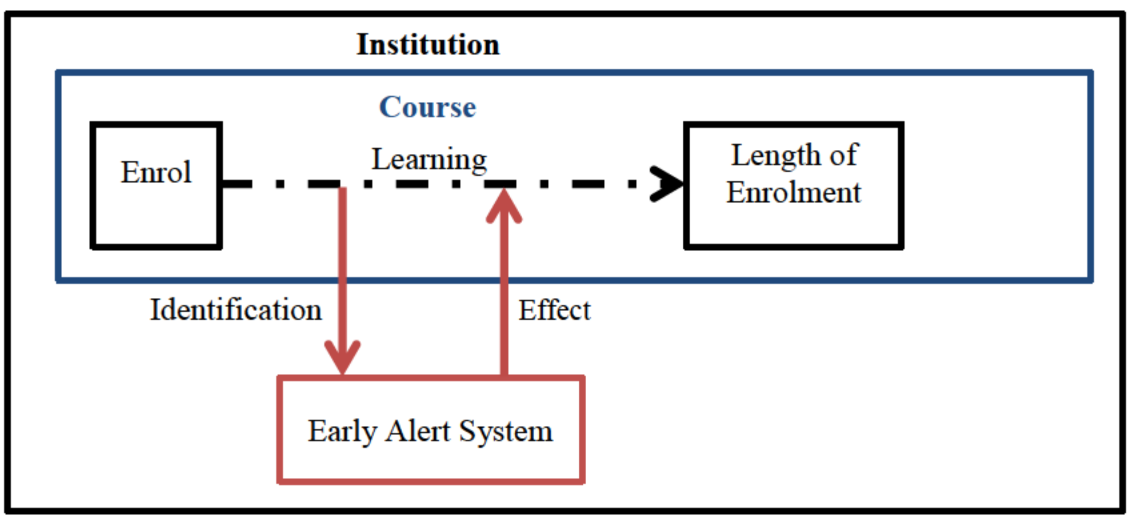


Figure 6.1 – Estimating the length of enrolment

The results from this chapter will assist in identifying the factors that have either positive or negative effects on student enrolment length, which in turn affects student retention.

Outlined in Chapter 2 were many studies which analyse student retention in various methods, however estimating the length of enrolment is not one of them. One explanation for this is the limitations of sample size and the assumptions associated with the statistical methods to perform valid analysis. This presents a unique approach to identifying the factors which affect students' length of enrolment and in turn the retention rate. It is expected that the EAS will improve retention and as such, there should be a significant positive coefficient on the students' length of enrolment, indicating that the EAS has the effect of keeping students enrolled for longer.

6.2 Methods of analysis

Multiple regression analysis using ordinary least squares (OLS) method to estimate the relationship between the dependent variable, length of enrolment, against independent explanatory variables. The model estimated can take on the general mathematical form of

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n + \varepsilon_i \qquad (6.1)$$

where

Y_i is the number of weeks enrolled
β_0 is the regression intercept, a base number of days enrolled
$\beta_1 X_1$ is the first explanatory variable
$\beta_n X_n$ is the nth explanatory variable
ε_i is the error in the model, capturing the difference between observed and predicted observations

The model can be broken up into the same classes of explanatory variables as chapter 5: *demographic*, *institution*, student *performance*, *workload* and *EAS*. The EAS variable is expressed in two ways, Model 6.1 expresses the EAS as a binary variable, where 1 indicates that the student was identified by the EAS during their enrolment. Model 6.2 captures the EAS as the number of times identified, representing the relative severity or risk of the student discontinuing. The low EAS severity corresponds to students identified by the EAS one to four times. The medium EAS severity corresponds to students identified by the EAS five to nine times. The high EAS severity corresponds to students identified by the EAS ten to 19 times. Finally, very high EAS severity corresponds to students identified by the EAS 20 times or more. Based on the results from chapter 5, it is expected that increasing EAS severity will correspond to students being enrolled longer.

6.3 Overall results

Using 16,124 observations, the overall results of model 6 are presented in Table 6.1. There is no statistical difference between models 6.1 and 6.2, so only the estimates for Model 6.1 are presented up to the EAS variable. The full statistical output for both models is presented in Appendix D.

Table 6.1 – OLS approach: overall results

Number of observations	16,124
F(32, 16091)	1693.36
Prob > F	0
R-squared	0.807
Root MSE	18.837

The estimated regression model has an F-test statistic of 1,693, with a corresponding p-value very close to 0. The R^2 value for is 0.807 which indicates around 80.7 per cent of the variation in the students' length of enrolment can be explained by the variation in the *demographic*, *institution*, *performance*, *workload* and *EAS* variables. These measures indicate a highly significant model.

6.4 Demographic effects

Demographics capture the background variables that need to be controlled for in the analysis. Estimating the students' length of enrolment, the demographic coefficients are presented in Table 6.2.

Table 6.2 – OLS approach: demographic results

Variable	Coefficient	Robust Std. Error
Constant	63.004[a]	1.802
Gender	0.319	0.327
Age	0.225[b]	0.092
Age^2	-0.002	0.001
ATSI	0.892	0.823

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The demographic results show that the constant is significant with an estimated value of 63 weeks. This represents over a year of time enrolled at the institution and captures around 3 to 4 teaching periods. With respects to gender, the estimated coefficient of 0.3 weeks is not significant. In the context of the chapter 5 results, while female students are more likely to discontinue or complete, the average length of enrolment for males and female students are similar. So there is no significant difference in the length of enrolment based on gender.

Holding all other variables constant, age shows an increasing positive relationship significant at the 5% level. For each additional year of age, the student is enrolled on average for another 0.225 weeks. The effect for an 18 year old student is being enrolled for between additional 3.45 and 3.96 weeks. At 40 years of age, the effect increases to being enrolled for between 6.05 and 6.89 weeks. This is in line with expectations; older students have a wider life experience to draw on to ensure successful study. Unlike chapter 5, the effect does not change over time either, with the quadratic age term being insignificant. Additionally, students who are older are more likely to be off-campus online students studying in a part-time capacity. This will logically increase the length of enrolment for these students as they balance study, work and family.

The ATSI variable indicates that there is no significant effect on the length of enrolment from identifying as an ATSI student. This varies slightly from the conclusions of chapter 5, where a weakly significant relationship between ATSI students and the likelihood of discontinuing was detected. So while ATSI students may be a little less likely to discontinue, the average length of an ATSI student is not significantly different from a non-ATSI student.

6.5 Institutional effects

Institutional differences such as the course enrolled or fee status is expected to capture the variations in the length of enrolment. The results are presented in Table 6.3 over page.

The *domestic fee* variable compares domestic students who pay university tuition fees up front, versus the base case which is deferring fees using the governments Higher Education Loan Program (HELP). Holding all other variables constant, Table 6.3 indicates that domestic fee paying students are enrolled for around 9.6 weeks longer than HELP students. The effect is significant at the 5% level. This is in line with expectations given fee paying students are likely to be more conscious of the cost of their education than a student who defers their fees.

International fee paying students are enrolled for an additional 3.7 weeks compared to domestic HELP students. The estimated coefficients are significant at the 1% level which is in line with expectations given that international students, like domestic fee paying students, are likely to be more conscious of the cost of their education. This also supports the findings of chapter 5 where international students were both less likely to discontinue and more likely to complete.

Table 6.3 – OLS approach: institutional results

Variable	Coefficient	Robust Std. Error
Domestic Fee	9.665[b]	4.572
International Fee	3.700[a]	0.694
Prior Studies	6.400[a]	0.432
On-campus	-0.819[b]	0.409
Diploma	2.950[c]	1.576
Advanced Diploma	9.005[a]	0.988
Bachelors (Graduate)	1.895[b]	0.745
Bachelors (Honors)	5.984[a]	1.073
School 1	-1.668[b]	0.727
School 2	-1.622[b]	0.670
School 3	-1.113[c]	0.669
School 4	-3.139[a]	0.642
School 5	-3.561[a]	0.754
School 6	-2.656[a]	0.640
School 7	26.011[a]	2.280
School 8	-0.953	0.752
School 9	1.286	0.890

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The *prior studies* variable captures students who have undertaken prior studies of any type. This can include professional certificates or other university courses. Holding all other variables constant, the OLS approach shows students who have undertaken previous studies are enrolled for an additional 6.4 weeks. This is in line with expectations, where students with prior study history should be more prepared for university study, compared to students who have had no such experience.

The *on-campus* variable captures students who study on-campus, compared to the base case which is students being off-campus online students. The results indicate a small but significant effect in both models. Students studying on-campus are enrolled for 0.8 weeks less than their off-campus online counterparts. This result was not expected given that on-campus students have the enriched learning environment of lectures, tutorials and face to face interaction with academics. The result is supported however by the results of chapter 5 which indicate that on-campus students had an increased likelihood of discontinuing. The result may be capturing an interaction effect with the course type, where there is a significant difference between the courses undertaken by students on-campus to students studying off-campus online. In the context of current trends towards online education, this indicates that online off-campus students may actually be better positioned to be retained compared to on-campus students. A more detailed analysis of this effect is conducted in chapter 7 to determine if the results obtained are valid.

The *course type* variable compares students studying diplomas, advanced diplomas, bachelors via graduate entry, and bachelors with honours to the base case of a student undertaking a bachelor level degree. Holding all other variables constant, the results indicate that diploma students are actually enrolled for 2.95 weeks longer than bachelor students. A typical diploma course has 8 units of study to complete, versus a bachelors course which has 24. The result is somewhat unexpected given the difference in the number of units required for the respective qualifications. Advanced diploma results are significant at the 1% level, with students being enrolled for an additional nine weeks. Given advanced diplomas consist of 12 units of study; the progression rate of students within these courses can also explain the increased length of enrolment. Another inference is that since the course is shorter in required units, students are more committed to completion. Both graduate entry and honours students are also enrolled for longer than normal bachelors' students. The result is expected given that students within these categories should be more prepared for the rigours of university study.

The final institution variable of interest is the various *schools* within the university. The estimated coefficients for schools 1 to 7 are significant, indicating that there is a significant difference in the average length of enrolment of students between these schools and the base case school. Schools 1 to 6 show students enrolled on average for between 1.1 and 3.5 weeks less than the base case school. School 7 is significantly different at the 1% level. On average, students enrolled within this school are enrolled for an additional 23 to 26 weeks. Given that school 7 is a relatively small school with very high entry standards, this result is in line with expectations. Furthermore, the magnitude of the differences between schools indicates that no one school has a significant retention issue relative to the other schools.

6.6 Student performance and workload effects

Measures of student performance and workload form important controls when estimating the effects of the EAS. The inclusion of detailed grade information is important as some grades attained do not contribute to course progression. The student performance measures are treated as continuous variables, and as such, the results are interpreted at the margin where attaining an additional grade will have a corresponding effect on the students length of enrolment. In effect, the variables are the grades attained in all previous teaching periods, capturing students' academic record. The estimated effects associated with student performance and workload are presented in Table 6.4.

Table 6.4 – OLS approach: Student workload and performance results

Variable	Coefficient	Robust Std. Error
Withdrawn	0.094[a]	0.004
Withdrawn Early	0.128[a]	0.003
Fail Incomplete	0.100[a]	0.003
Fail	0.087[a]	0.003
Pass	0.078[a]	0.001
Credit	0.081[a]	0.002
Distinction	0.087[a]	0.002
High Distinction	0.099[a]	0.003
Other	0.072[a]	0.004
Workload	-31.091[a]	0.397

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The grades associated with learning environment variables are all significant at the 1% level. For each grade a student attains, model 6 indicates that a student is enrolled for between 0.072 and 0.128 of a week. The effect size is relatively small and one issue with the method of estimation is that it does not differentiate the effects of attaining a negative grade outcome like failure, from the positive grade outcomes like a credit or distinction grade. The results indicate that there is little to no variation between positive and negative grade outcomes. This simply indicates that attaining any grade indicates that the student is enrolled for longer.

The final variable captured in the model is the weighted average workload. The workload variable is an ordinal scale, where 0 represents inactivity, 1 represents part-time workload and 2 represents full-time workload. The results indicate a significant negative relationship which also corresponds to course progression. A student who undertakes full-time study is estimated to be enrolled for around 31 weeks less than a part-time student. The result is in line with expectations given a student who undertakes more units of study each teaching period will complete their qualification quicker than a student undertaking part-time study.

6.7 Early alert system effects

The EAS results reflect the two models estimated, Models 6.1 and 6.2. In Model 6.1, the EAS variable is a binary variable corresponding to identification at some stage within students' enrolment (0 = not identified, 1 = identified). Model 6.2 captures the number of times a student is identified by the system over their enrolment. The results are presented in Table 6.5.

Table 6.5 – OLS approach: EAS severity level

Model	Variable	Coefficient	Robust Std. Error
6.1	EAS Identified	8.956[a]	0.441
6.2	Low EAS severity	8.932[a]	0.441
6.2	Medium EAS severity	13.12[a]	0.512
6.2	High EAS severity	9.384[a]	0.668
6.2	Very high EAS severity	4.988[a]	0.808

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The results indicate a significant effect from the EAS at the 1% level. This contrasts the results from the multinomial logistic regression approach of the previous chapter, which showed limited significance of the EAS variable. Model 6.1 estimates that students who have been identified by the EAS are enrolled for an additional 9 weeks. Model 6.2 shows that there are varying effects within the severity levels. Students identified one to four times by the EAS are enrolled for an additional 8.9 weeks compared to students not identified by the EAS. Students identified five to nine times have the greatest estimated increase in the length of enrolment, with an additional 13.1 weeks of enrolment compared to students not identified by the EAS. Students identified by the EAS 10 to 19 times are estimated to be enrolled for an additional 9.3 weeks. Finally, students identified by the EAS over 20 times have their enrolment increased by five weeks. The results indicate students are retained longer as a result of EAS identification, with students in the medium EAS severity category having the greatest increase. This is the first indication that the

EAS has a positive effect on student outcomes, increasing the length of enrolment. Importantly, as students move into higher EAS severity categories, the additional length of enrolment starts decreasing.

6.8 Chapter summary

The results of the OLS estimates indicate that there are limited demographic effects correlating to the length of students' enrolment. The institution variables were statistically significant throughout the model, highlighting the importance of including variables which control for fee type, prior studies, course type and school enrolled in. The student performance variables captured key information associated with course progression, showing that logically, attaining any grade was associated with increased length of enrolment, but the effect size was minimal.

Testing for the effects of the EAS revealed that students identified by the EAS were on average enrolled for an additional nine weeks compared to students not identified. This shows there is a link between the EAS and student is being enrolled for longer. This can be interpreted as a positive effect on student retention outcomes, which is a significant finding from this model. Not to diminish the finding, one limitation of the method of analysis is that a causal link cannot be established between the EAS and retaining students for longer. It is not possible to conclude that the EAS causes students' length of enrolment to increase. It is therefore important to further analyse the relationship between the EAS and length of student enrolment to understand the nature of the link identified in this model.

The estimated effects presented using an OLS approach provide a solid empirical foundation to develop more complex temporal models in chapter 7. A limitation of the model of the OLS approach is that while the dependent variable captures variations in the length of enrolment, the explanatory variables are essentially a snapshot of three years of data combined. Building from the OLS results, chapter 7 can capture how the different explanatory variables correlate to students' length of enrolment at different times during the students' progression.

Finally, the OLS results presented use a large number of observations, capturing 16,124 student/course combinations over a three year period. The pooled approach conducted by the law of large numbers should give accurate estimates that are relatively unbiased and normally distributed as per the assumptions of the approach. Chapter 7 uses the Cox proportional hazards model in survival analysis to provide a non-parametric estimation of the hazard of discontinuing.

The model handles the complex nature of student enrolment data over time including unbalanced data and censoring. The non-parametric approach of chapter 7 means many of the statistical assumptions associated with the OLS approach are relaxed, meaning that a more accurate and meaningful model can be estimated. The results from chapter 6 provide a solid foundation for the development of complex and detailed statistical models of student retention in chapter 7.

Chapter 7 - Temporal Analysis of Student Retention: A Survival Analysis Approach

7.1 Introduction

Understanding student retention has been a significant area of research within universities for a long time. Understanding retention is complex given the characteristics of the learning environment and the interaction between students, academics and administrators within the higher education system. One limitation of many previous studies are the implicit or explicit assumptions with regards to temporal effects. The factors that affect a student's decision to discontinue their study today may not be the same factors that affect that decision tomorrow should the student stay. As such, this chapter's aim is to provide a comprehensive analysis of retention in a temporal setting. Using the non-parametric form of survival analysis, a detailed understanding is developed as to how factors affecting retention change over time. A review of relevant analysis using survival models is presented to highlight the analysis in the field performed to date. This is followed by a brief review of the survival analysis method with an explanation of how coefficients are interpreted. This chapter will then present the various model specifications used to provide detailed analysis.

In total, there are six different model specifications tested using different variables to capture the effect of the EAS. These were specified and calibrated in Chapter 3. The first model establishes the base model of analysis, identifying the significant *demographic*, *institutional*, *student performance* and *workload* variables. The second model divides students into two groups, those identified by the EAS and those never identified, and tests for significance of variables within groups. This forms the conditional model approach. The third model tests for interaction effects between the demographic, institution and learning environment variables and the EAS function. This is the interaction model. The fourth model tests for a relationship between the triggers used in the EAS and the survival function. This is the EAS trigger model. The fifth model extends this to incorporate the triggers into the base models to test for their significance in the presence of other variables. This is the controlled-trigger model. The sixth and final model uses survival data to use in a treatment effects model. This explores any possible causal link between the EAS and the length of the students' enrolment. The chapter concludes by summarising the key findings from these models, establishing if the EAS is actually affecting student retention.

As discussed in chapter 3, several studies have previously analysed student retention rates using survival analysis (DesJardins, 2003; DesJardins et al., 1999; DesJardins et al., 2002; Ishitani, 2006; Ishitani and DesJardins, 2003). These studies analysed the factors associated with retention in different tertiary settings, taking into account the temporal effects which characterise the university learning environment. Treatment effects modelling was introduced by Rubin (1974), The causal modelling method has been used extensively in literature where causal inference is required on observational data where no experimental control is possible.

7.2 Empirical model and estimation procedure

Survival analysis allows appropriate treatment of variables in a temporal context and allows right censoring. Right censoring occurs when there is variation in the failure times; in this case, when students leave the institution. This is a major limitation on many temporal models, making survival analysis the most appropriate statistical method for student retention analysis. In the context of this study, survival is defined as a student being retained, with the term ‘failure event’ referring to a student leaving university.

Survival analysis can be conducted using both parametric and non-parametric models. For the purposes of this study, it is appropriate to take a non-parametric approach to the estimation of the survival function. This means no assumptions need to be made about the underlying distribution of terms. As stated previously when discussing another approach, “the rationale is that we should go with a statistical model that requires fewer or weaker assumptions” (Liao, 1994, p. 48). While parametric models typically require less data to reach estimated coefficients, given the sample size used for this study, there are no issues around degrees of freedom.

7.2.1 The main assumption of survival analysis

Students exit the model either by discontinuing their studies or through right censoring events, such as completing their studies or the student still being enrolled at the end of the data capture period. The main assumption of survival analysis is that proportional hazards remain constant over time. This variation in time lengths raises a key issue in handling time variables, in that some dependent variables are correlated with time itself, violating the proportional hazards assumption.

7.2.2 Addressing survival analysis assumptions

To correct the issues associated with time varying covariates, models include interaction terms for each time-varying independent variable multiplied with a function of time. This will isolate the effects of the independent variables and provide unbiased estimators. To identify which variables were time variant, the Proportional Hazards assumptions test (PH test) was run for each model. If there was a significant relationship between any of the variables in the model and time itself, the PH test would indicate a significant result using a p-value less than 0.05. By estimating the interaction of significant variables as a function of time, the time varying effects can be isolated, avoiding violations of the proportional hazards assumption. The detailed results for the PH test for each model are presented in Appendix F. According to the PH test specifications, the results should not be significant, indicating no significant interaction with time.

7.3 Base model results

7.3.1 Base survival model description

The first model estimated using survival analysis is the baseline model including *demographic*, *institutional, student performance* and *workload* variables. The purpose of the model is to establish a starting model, from which more complex concepts and ideas can be developed from. Importantly, there are three different versions of the base model, the short-run model, the enduring model and the long-run model. The difference between each model is how the effect of the early alert system is captured in relation to the underlying hazard function.

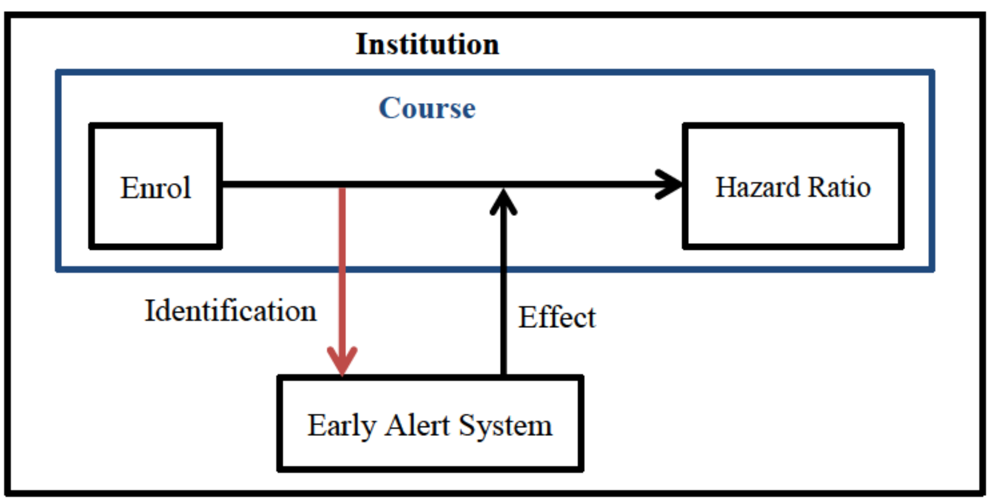


Figure 7.1 – Short run and enduring base models (Authors own contribution)

Figure 7.1 represents the short run and enduring configurations, where the models test whether there is a significant relationship between being identified by the early alert system and the hazard ratio. The short run base model tests the question ‘is there a relationship between being identified by the EAS at a specific moment and the student’s survival rate’. The enduring base model tests the hypothesis ‘is there a relationship between when a student is first identified by the EAS and the student’s survival rate’.

The long run model divides students into two cohorts, those identified at some point during their studies and those never identified during their studies. This is represented in Figure 7.2, where the red path indicates the group of students identified by the system, which is separate from the black path, indicating the students never identified by the early alert system and not interacting with targeted student support services via the EAS.

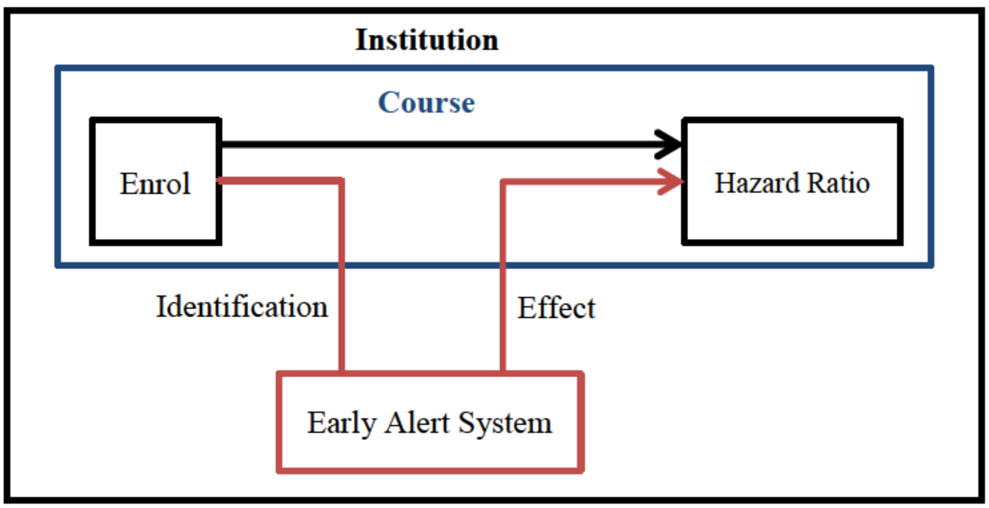


Figure 7.2 – Long run base model
(Authors own contribution)

In dividing students into the two groups, each group will capture different types of errors. For the group of students not identified by the EAS, the group will comprise both students who did not discontinue and did not need targeted student support. It will also comprise of students who did discontinue, needed targeted student support but did not receive it. This is a type 1 error and provides some measure of the effectiveness of the system. If the system was to be running perfectly, all students who were at risk of discontinuing would have been identified.

The second group of students is comprised of those who were identified and will contain both students who needed targeted support and received it and students who did not need targeted student support and received it; the latter representing the type II error of the system. This is much harder to estimate and while captures a measure of inefficiency of the system, is not the main focus of this model. Assuming stability within the models, there should be no significant difference between estimates of the short run, enduring and long run models except for the EAS variable.

7.3.2 Significance of base effects model and assumptions tests

The base model examines the relationship between demographic, institution and learning environment variables. The overall results are presented in Table 7.1

Table 7.1 – Base model: overall model estimates

Short Run		Enduring		Long Run	
LR $\chi^2(50)$	7646.73	LR $\chi^2(49)$	7603.68	LR $\chi^2(51)$	7940.7
Prob > χ^2	0	Prob > χ^2	0	Prob > χ^2	0

PH Test		PH Test		PH Test	
$\chi^2(50)$	39.12	$\chi^2(49)$	37.51	$\chi^2(51)$	39.99
Prob > χ^2	0.8667	Prob > χ^2	0.8843	Prob > χ^2	0.8672

The overall model results show that short-run, ensuring and long-run models are all significant at the 1% level. The Proportional Hazards test (PH test) results show that overall, the models have no significant correlation with time. This indicates that the models have not violated the proportional hazard assumptions associated with survival analysis. Given that the *demographic*, *institutiona*l, *student performance* and *workload* variables are all similar estimates, only the results for the short-run base effects model is presented. The results for all three models are presented in Appendix E for completeness.

7.3.3 Demographic Variables

The estimated hazard ratios for demographic variables associated with the base model are presented in Table 7.2.

Table 7.2 – Base model: demographic results

	Hazard Ratio	Std. Err.
Gender	1.076[b]	0.034
Age	0.969[a]	0.007
Age Squared	1.000[a]	0.000[d]
ATSI	0.837[b]	0.070

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The results in Table 7.2 indicate that there is a significant relationship between gender, age and ATSI status in all three models. This indicates a stable result for demographic variables regardless of how the EAS variable is represented. This is important, showing that demographic variables are correlated to the hazard ratio and subsequently the likelihood of discontinuing their studies.

For interpreting the coefficient for gender, males are coded 0 and females coded 1. The results indicate that females have a hazard ratio around 7.6 per cent higher than their male counterparts. The higher hazard ratio means that in any given week, females face a hazard 7.6 per cent higher than their male counterparts, indicating an increased chance of discontinuing. This gender difference may be attributable to the institution, or it could be systematic within the Australian tertiary sector. This indicates an area for future comparative research between institutions to determine the factors causing the differences in gender hazard ratios.

Age is a significant variable, expressed as a non-linear relationship by including the squared term. Using the estimates provided it is possible to calculate the age at which students face the minimum hazard ratio. This occurs at 51 years of age, showing that as students get older, their hazard ratio declines until they hit 51 years of age, past which, the hazard ratio starts increasing again. This effect is important in the UNE context as the average age of students is around 29 years. This may indicate that the current age profile of the university is helping maximise retention.

The third variable of interest is ATSI which shows that students who have identified as ATSI have a hazard ratio around 14 to 16 per cent lower than students not ATSI identified. This significant result is likely due to the additional on-campus support services provided to ATSI students beyond the normal student support services. This strong finding shows that once the cultural barriers around ATSI admission are removed, ATSI students are more likely in their given awards to continue and complete their studies compared to non-ATSI identified students. This also vindicates the positive services provided by the Oorala Centre, a unique UNE institution that provides support to ATSI students on-campus.

7.3.4 Institution Variables

The institution variables capture five characteristics of the students' enrolment that relate to the institution. These are: *fee type*; *prior studies*; *study mode*; *course type* and *school of enrolment*. The base case for these five categories is a student using the Higher Education Loan Program (HELP) to pay their fees, who has not previously studied and is studying off-campus using online resources. The student is studying a bachelor degree through normal admission paths in a school which offers professional qualifications. The results for the institution variables minus the schools are presented in Table 7.3. Some variables have hazard ratios of 1 and standard errors of 0 as a result of rounding. However the significance of these variables are accurately presented and are important inclusions in the models for completeness.

Table 7.3 – Base model: institutional estimates

	Hazard Ratio	Std. Err.
Domestic Fee	0.605	0.338
Domestic Fee x t	1.010	0.007
International Fee	0.194[a]	0.057
International Fee x t^2	1.000[a]	0.000[d]
International Fee x t^3	1.000[b]	0.000 [d]
Prior Studies	0.587[a]	0.098
Prior Studies x ln(t)	1.133[a]	0.049
On-campus	1.175[a]	0.055
On-campus x t	0.910[a]	0.013
Diploma	1.022	0.114
Advanced Diploma	0.783[b]	0.093
Advanced Diploma x t	1.003	0.002
Bachelors (Graduate)	0.740[a]	0.051
Bachelors (Honours)	0.619[a]	0.078

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The variable fee type is categorical, using domestic HELP students as the base category, of which domestic and international fee paying students are compared to. The results show that there is no significant difference between the domestic fee paying students and domestic HELP students. The model also includes an interaction term between domestic fee paying students and time. This means that the variable initially failed the proportional hazards test. Accounting for the interaction with time, the variable showed no significant difference from HELP fee category students.

There is a significant difference between international students and domestic HELP students. The hazard ratio for international students is not constant over time, and includes two interaction terms with time. The hazard ratio for international students is therefore plotted over time in Figure 7.3 to provide a meaningful interpretation of the true hazard ratio associated with being an international student. The results show that international students start their studies with a significantly lower hazard ratio than HELP fee domestic students. However this increases over time, where there is no significant difference around week 76, or one and a half years into their study.

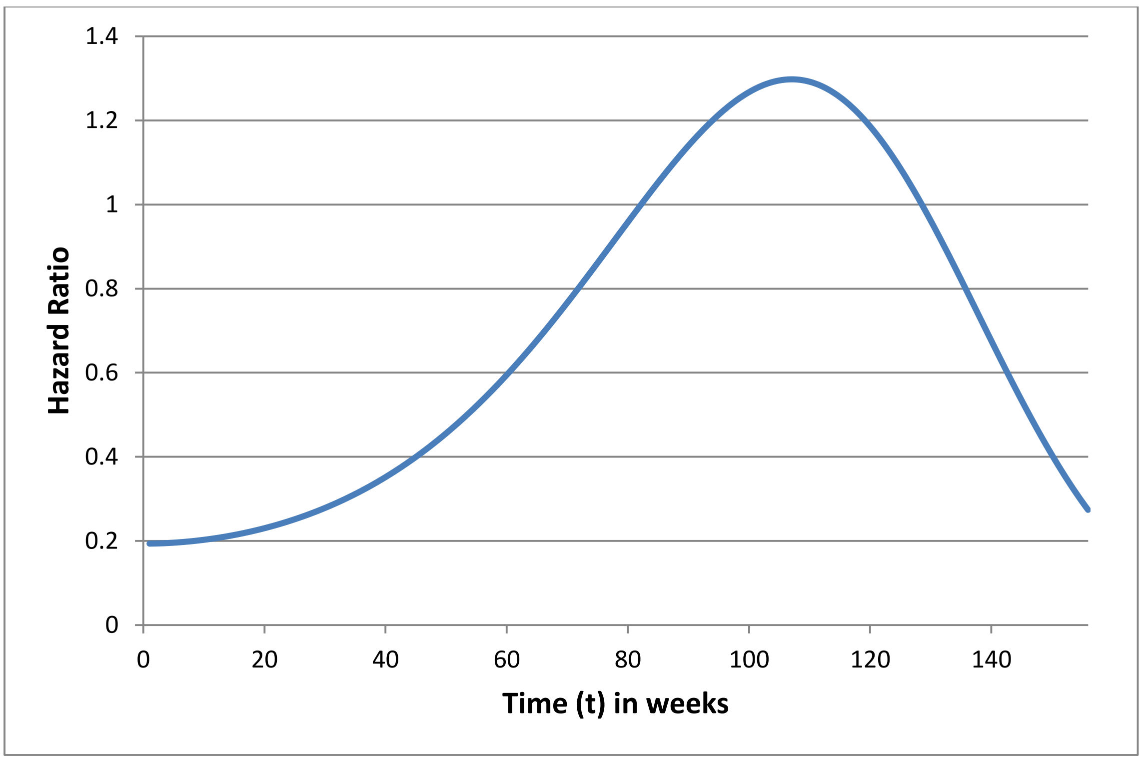


Figure 7.3 – Base model: international fee hazard ratio over time

The hazard ratio continues to rise, peaking at 104 weeks of study. This corresponds to the average length of enrolment for an international student. International students will usually complete two years of their course at the case study institution. As such, the highest risk of discontinuing occurs when students reach the two year mark but have not completed the required units or are not on track to complete on time. In the context of early alert systems, this indicates that focus needs to be placed on international students from weeks 76 to 104 of their enrolment, and possibly before week 76 if the trajectory of the student is to be changed. The results also show that for international students enrolled in longer courses past the two year mark, the hazard ratio decreases again. This contrasts the results of chapters 5 and 6 which aggregated the effect. This is important evidence to support the continued use of survival analysis over the other methods to determine the relationship between variables in the model and the risk associated with discontinuing studies.

The effect of prior study is consistent across all three models, being significant at the 1% level. There is also a significant interaction with time expressed in the logarithmic functional form. The true hazard ratio over time is presented in Figure 7.4.

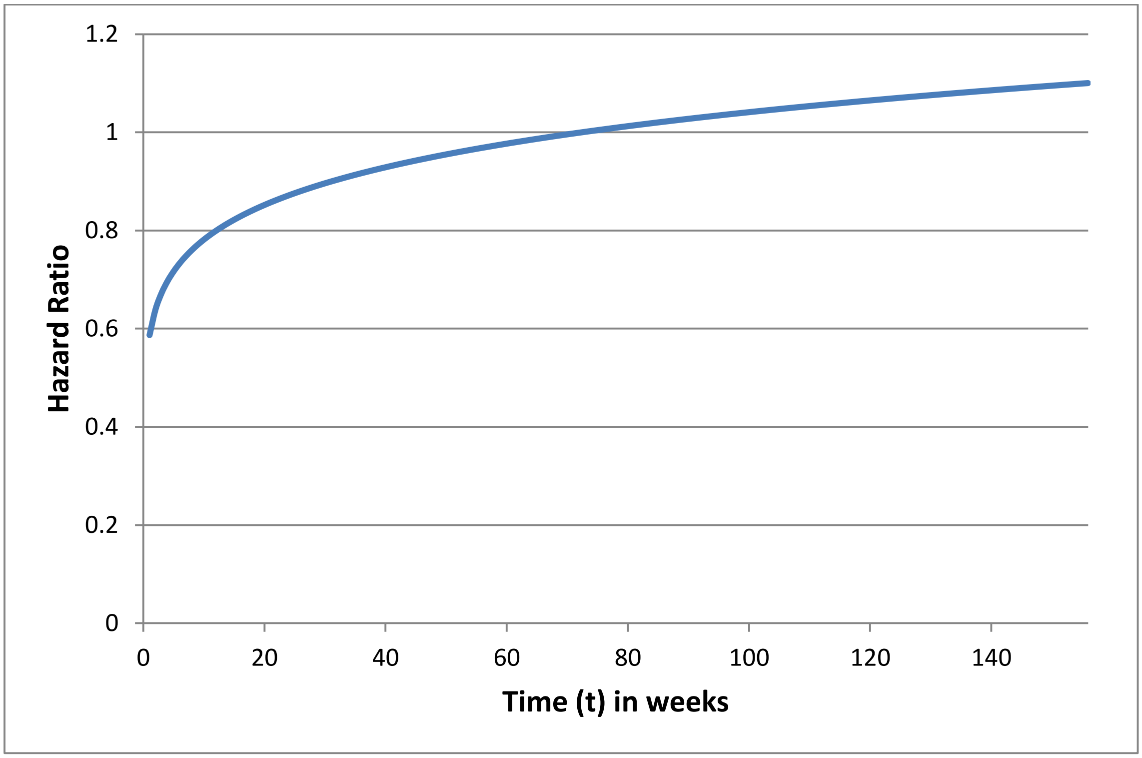


Figure 7.4 – Base model: prior studies hazard ratio over time

The results show that students who have undertaken prior study have a significantly lower hazard ratio at the start of their studies. The hazard ratio increases at a decreasing rate, with there being no significant difference to students without prior studies after the first year. An interesting observation however is that students with prior studies have a higher hazard ratio past around 100 weeks. This indicates that students with prior studies may be at risk of discontinuing later in their courses. This may be a result of already having attained a qualification and as such, may not have the same motivation to complete the studies as someone without a qualification. For EAS systems, this means the effect of prior studies needs to vary depending on how far a student has progressed through their studies.

The on-campus variable compares students who complete their studies attending classes in person at the institution, versus off–campus online students. The results for all three models are significant at the 1% level with time varying effects. The estimated hazard ratio is presented in Figure 7.5.

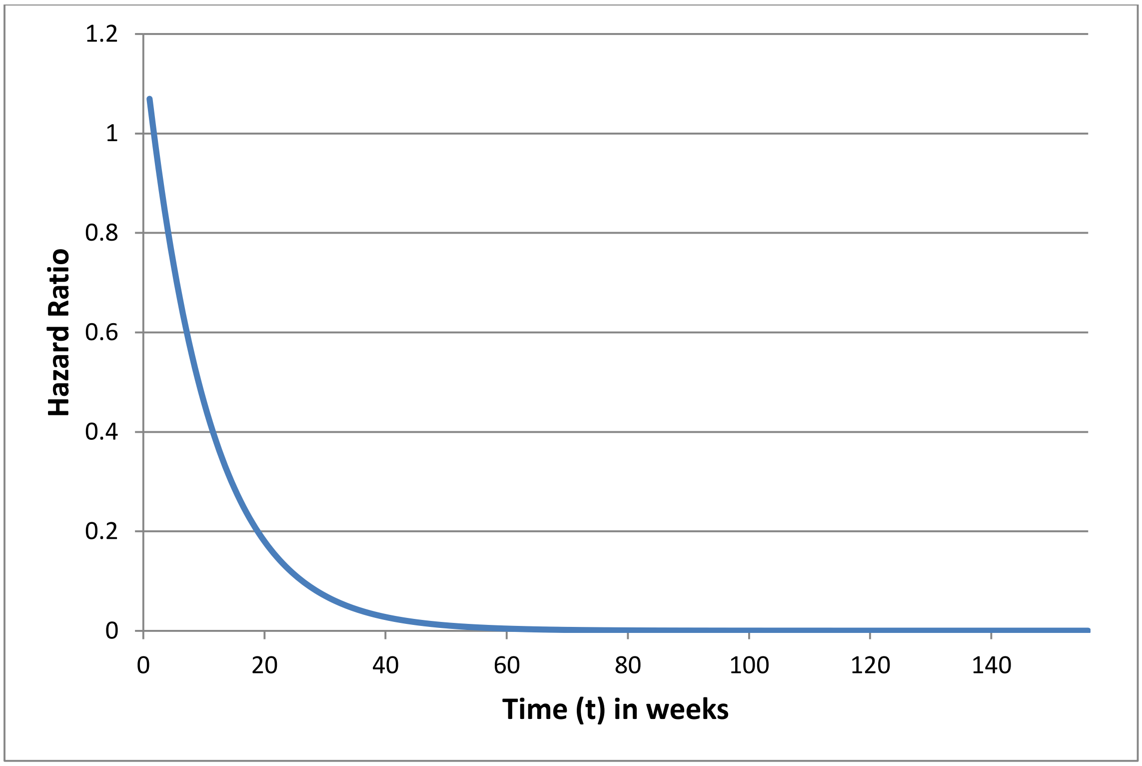


Figure 7.5 – Base model: on-campus hazard ratio over time

The results indicate that initially, on-campus students have a higher hazard ratio than off-campus online students. However, the hazard ratio for on-campus students rapidly decreases after a few weeks, with those students having a significantly lower hazard ratio after three to five weeks. This shows a stark divide between the hazard ratios for students in different modes of study. This raises the important question about how student support can be provided to students in a distributed learning environment such that this divide is reduced.

Comparing the degree types, there is no significant difference in the hazard ratios of diplomas and the base case of the bachelor degree. For advanced diploma courses, there is an interaction effect over time, shown in Figure 7.6.

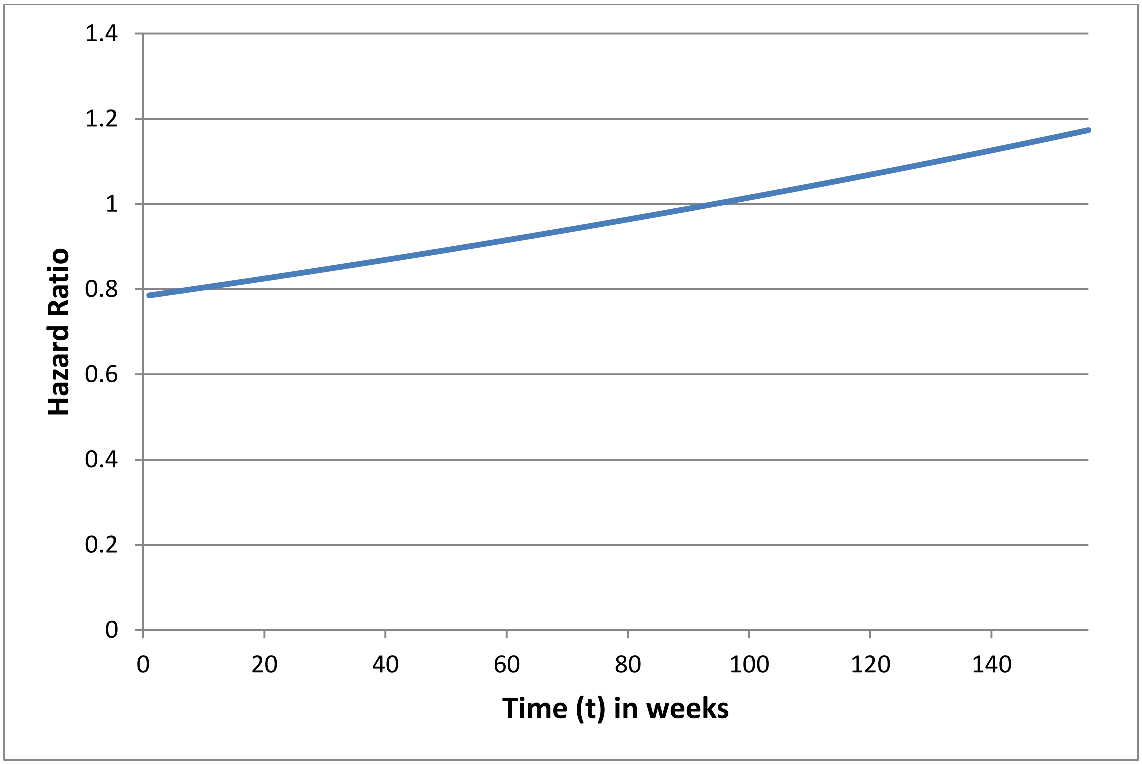


Figure 7.6 – Base model: advanced diploma hazard ratio over time

The base results all show a reduced hazard ratio than normal entry bachelor students at the start of the course, significant at the 5% level. The hazard ratios change over time, where later in the course there is no significant difference in hazard between the two courses. Past around 100 weeks, advanced diploma students have a higher hazard ratio, indicating that prolonged enrolment in this course is not ideal.

There is a significant difference between the normal entry students and bachelor students who are admitted through graduate entry or admitted for honours programs. Students admitted in these cases have a significantly lower hazard ratio at the 1% level of significance. Both results are in line with expectations that students in these categories are at significantly reduced risk of discontinuing their studies. There is also some overlap between variables, with graduate entry students theoretically also being captured in the prior studies indicator. The key difference here is that graduate students would have completed their prior studies, whereas some students in the prior studies indicator may have commenced studies previously, but not completed.

The result for schools presented in Table 7.4, indicate a large degree of variation in the hazard ratios between schools. Additionally, schools 2 and 6 have temporal effects where the hazard ratio is changing over time.

Table 7.4 – Base model: school estimates:

Variable	Hazard Ratio	Std. Err.
School 1	1.127[c]	0.078
School 2	1.272[a]	0.091
School 2 x 1/t	0.006[a]	0.006
School 3	1.181[b]	0.080
School 4	1.297[a]	0.082
School 5	1.295[a]	0.090
School 6	0.860[c]	0.070
School 6 x t	1.004[a]	0.001
School 7	0.778	0.282
School 8	1.045	0.083
School 9	0.958	0.077

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

School 1 has a significantly higher hazard ratio at the 10% level from the base school. School 3 is significantly higher at the 5% level, while schools 4 and 5 are significantly higher at the 1% level. Schools 2 and 6 both have time varying aspects to them, with the results presented in Figure 7.7 and Figure 7.8 respectively over page. For school 2, the hazard ratio starts significantly lower than the base school. However, the hazard quickly increases and by week 20, there is no difference between the two schools. After week 20, the hazard ratio continues to increase at a decreasing rate, reaching a hazard ratio around 20 per cent higher than the base school around week 90.

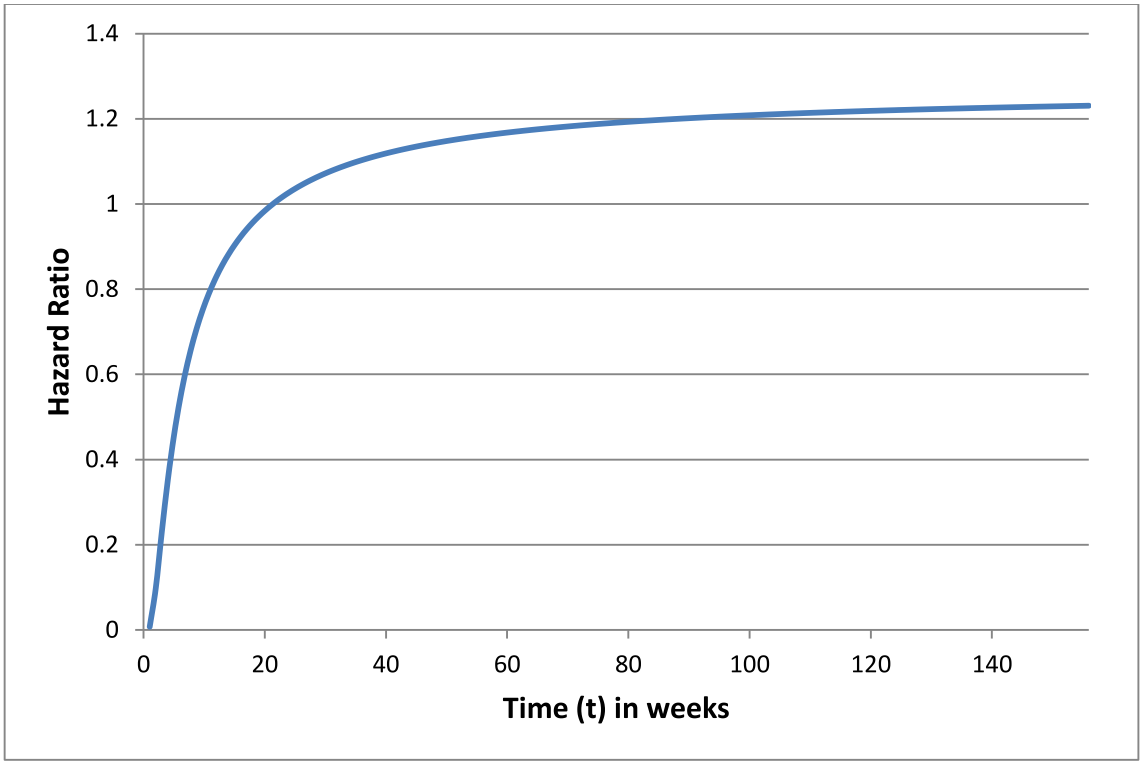


Figure 7.7 – Base mode: school 2 hazard ratio over time

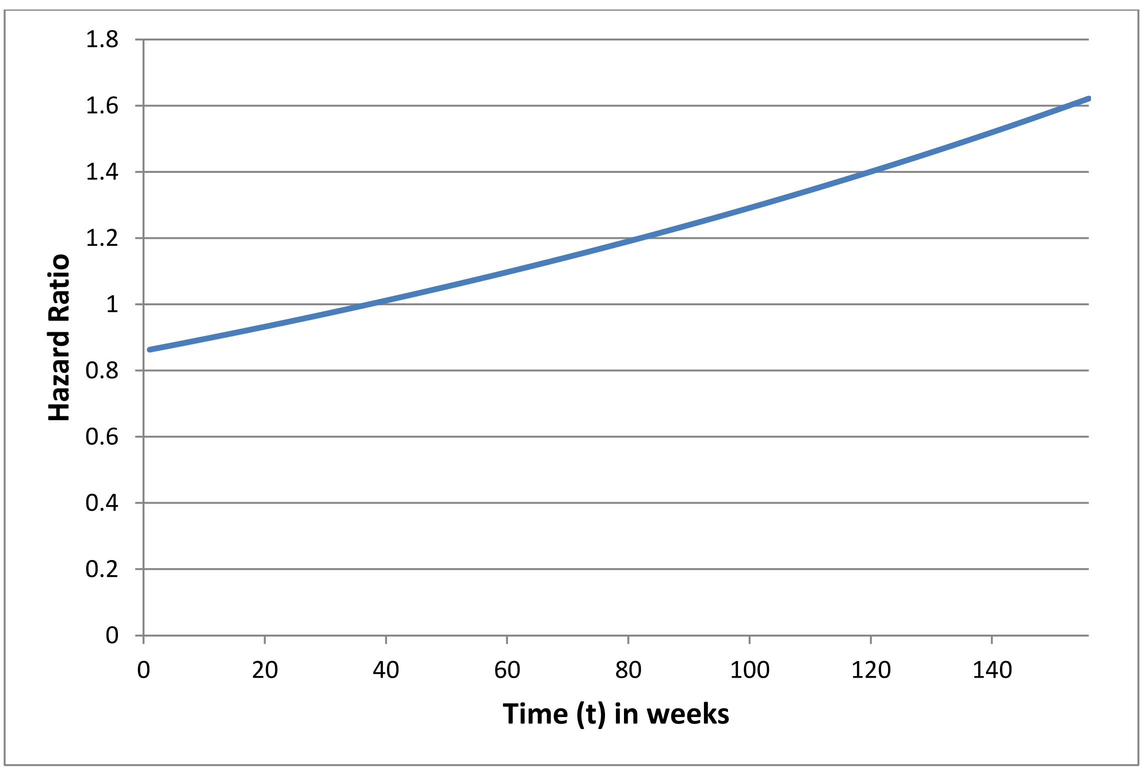


Figure 7.8 – Base model: school 6 hazard ratio over time

In the case of school 6, the hazard ratio increases linearly over time. Initially, students in this school have a lower hazard ratio than the base school. By around week 36 there is no difference between the schools, and past this point, students enrolled in school 6 continue to face an increased hazard ratio. This indicates that the longer students are enrolled in school 6, the higher the risk that the student will discontinue. As such, support efforts within this school need to focus on students who are enrolled for longer. This also raises the question of course progression within the school as a possible issue. An area for future study is to conduct a more detailed analysis within the school to identify possible sources of this problem.

7.3.5 Student performance and workload

The two learning environment variables used in this model include grades of the student and the level of workload undertaken by the student. The grades can be divided into three classes: negative grades (*withdrawn*, *withdrawn early*, *fail incomplete* and *fail*) which do not contribute to a student's progression in the course; positive grades (*pass*, *credit*, *distinction* and *high distinction*) which indicate a level of competency in a particular unit; and *other* administrative grades which capture more complex student situations such as special extension of time or special examinations. The results are presented in Table 7.5 over page.

The interpretation of these grades should be considered as a one unit increase. That is, if a student receives in any given week both a credit and a distinction, the overall changes to the base hazard ratio would be calculated by adding the logarithm of the two coefficients. Additionally, many grades include temporal effects, where the hazard ratio changes over time. The variable for time (t) is indicated along with the mathematical relationship to the estimated coefficients.

The results show relatively consistent estimates across the three models for the grades. *Withdrawn*, *withdrawn early* and *fail incomplete* all interact with time variables which are significant. The hazard ratio over time associated with receiving a *withdrawn* grade is plotted in Figure 7.9. The graph shows the non-linear relationship of the hazard ratio over time, with the initial effects indicating students who withdraw during their first year having a higher hazard ratio. After the first 52 weeks however, the results show either a reduced hazard ratio or no effect resulting from withdrawing. This is logical, given that students who withdraw during their first year are likely to be struggling with study and have invested less time to their course. As such, withdrawing in the first year is a strong indicator of students at risk of not being retained.

Table 7.5 – Base model: student performance and workload estimates

Variable	Hazard Ratio	Std. Err.
Withdrawn	1.371[a]	0.092
Withdrawn x t	0.992[a]	0.002
Withdrawn x t^2	1.000[a]	0.000[d]
Withdrawn Early	1.451[a]	0.081
Withdrawn Early x t	0.992[a]	0.002
Withdrawn Early x t^2	1.000[a]	0.000
Fail Incomplete	1.262[a]	0.024
Fail Incomplete x t	1.000[c]	0.000[d]
Fail	1.141[a]	0.020
Pass	0.847[a]	0.012
Credit	0.891[a]	0.013
Distinction	0.868[a]	0.014
High Distinction	0.874[a]	0.018
Other	0.743[a]	0.052
Inactive	45.447[a]	5.592
Inactive x t (for t <= 16)	0.136[b]	0.106
Inactive x 1/ln(t) (for t > 16)	0.084[a]	0.016
Part-time	2.189[a]	0.362
Part-time x ln(t)	0.976	0.054

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

However, after the first year this may not be a good indicator. The results show that there is either no effect or a reduction in the hazard ratio as a result of withdrawing. This captures students who have progressed further with their degree who are more likely to exhibit grade maximising behaviours, or, reduce workload to continue course progression in the face of adversity.

The *withdrawn early* grade occurs when students withdrew before the financial census date. As such the student is not negatively affected financially or academically with respect to grade point average. The hazard ratio for withdrawn grades depends on time, depicted in Figure 7.10. Like withdrawing, the relationship is non-linear. However, receiving a *withdrawn early* grade increases the hazard ratio for around the first 80 weeks of enrolment. Between around 80 to 100 weeks there is no effect on the hazard ratio, after which the hazard ratio increases again.

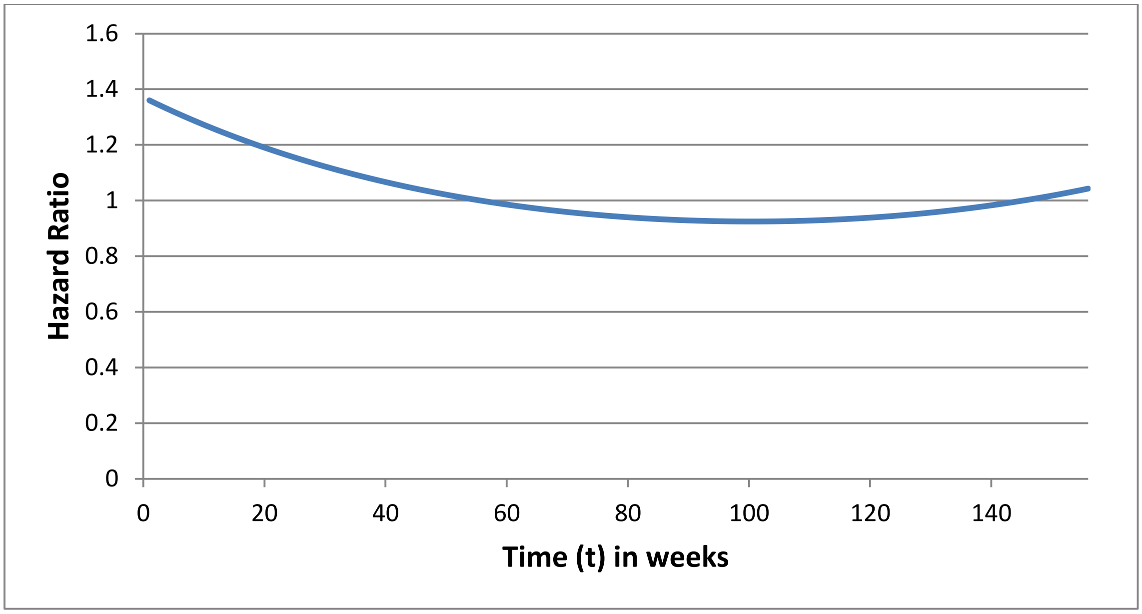


Figure 7.9 – Base model: withdrawn grades hazard ratio over time

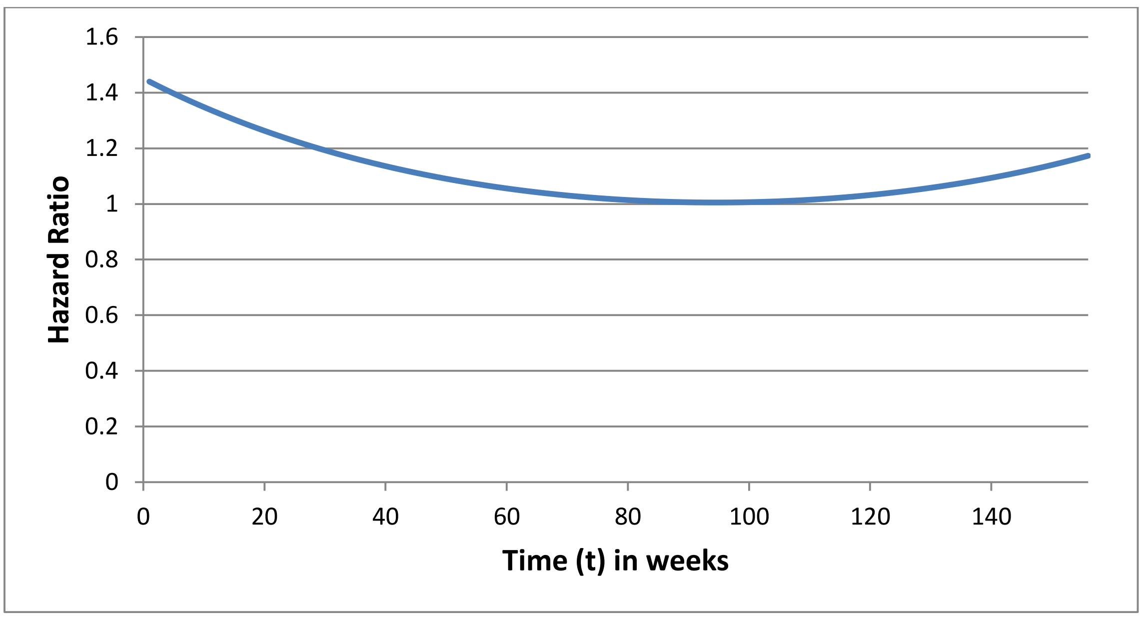


Figure 7.10 – Base model: withdrawn early hazard ratio over time

This shows that students who withdraw early from units within their course do exhibit a higher hazard of discontinuing, especially for the first year and a half of study. The difference between *withdrawn* and *withdrawn early* grades is that, for the latter, the financial cost of the unit is not passed onto the student. Therefore, the difference between the hazard ratios for the two grades may indicate the magnitude of the effect of payment of fees has on retention. This is one area which requires more detailed econometric analysis outside of this study.

The *fail incomplete* grade is also time-dependent with hazard ratios over time plotted in Figure 7.11.

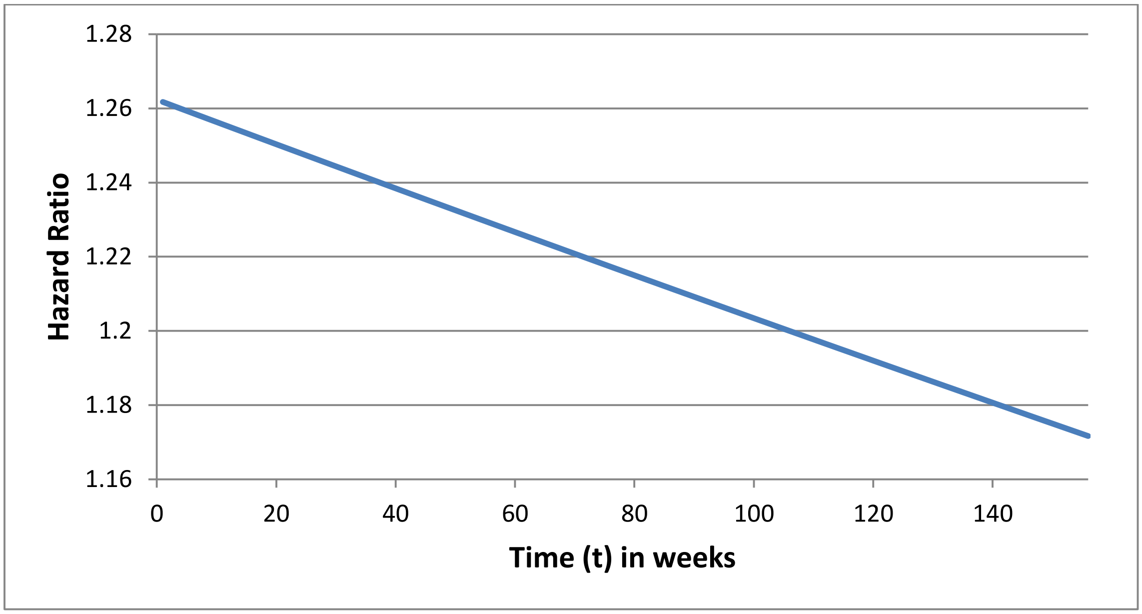


Figure 7.11 – Base model: fail incomplete hazard ratios over time

The graph show that receiving a fail incomplete earlier in the enrolment is associated with a higher hazard ratio. While the risk associated with receiving these grades decreases over time, even after three years the models indicate students will have a hazard ratio 17 per cent higher than students who never received a fail incomplete.

Unlike the previous three unit grades, *failure* grades are far more consistent over time. For each *failure* grade, the students hazard ratio increases by 14 per cent. Given that this effect is relatively constant over time, this variable is a good indicator of students at risk of discontinuing.

The positive grade outcomes of *pass*, *credit*, *distinction* and *high distinction* all have a positive effect in reducing the hazard ratio of the student. Furthermore, the magnitude of the positive effect is relatively constant across all grades. This indicates the positive effect of progressing with a course. Additionally, students who receive a grade in the *other* category also have a reduction in their hazard ratio. This indicates that there are also benefits from students taking advantage of special grade outcomes, such as special exams. Most of these grades also require a degree of administrative interaction, which could assist with academic integration. Therefore it is possible to conclude that any of the grades which help students progress with their course has a positive effect on reducing risk associated with discontinuation.

Workload is another statistically significant variable. A student during any given week can have three possible levels of workload: full-time study which is defined by Australian legislation as three or more units of study per teaching period; part-time study, which is less than three units per teaching period; inactive is where a student undertakes no units of work but is enrolled in a course. The base case is the full-time student. Comparing full-time to inactive students captures the effects associated with students who take a break from their studies and return at a later date. The results in Table 7.5 show that this variable is highly correlated to two variables capturing interactions over time. To understand the complex time varying aspects associated with inactivity, a Kaplan-Meier graph plots the probability of survival for inactive and non-inactive students.

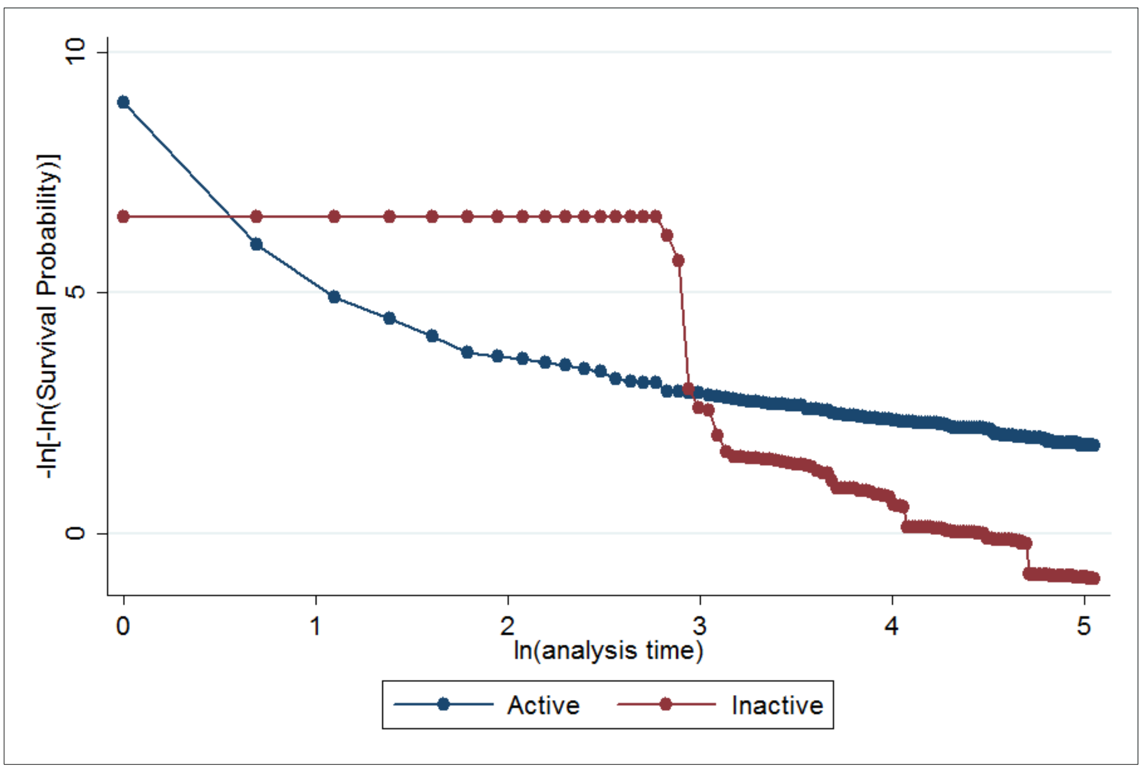


Figure 7.12 – Base model: survival probability of inactive students

If the proportional hazards assumption is upheld in Figure 7.12, the two lines should be parallel. Clearly this is not the case, with a linear function up to week 16, after which the function takes on an inverse logarithmic shape. This indicates a discontinuous function over time, so two functions interacting inactivity over time are used to capture interactions over time. The first function captures the linear effect over the first 16 weeks. The second function captures the inverse logarithmic function from week 17 onwards. The estimated hazard ratio for inactivity is plotted over time incorporating both functions to create Figure 7.13.

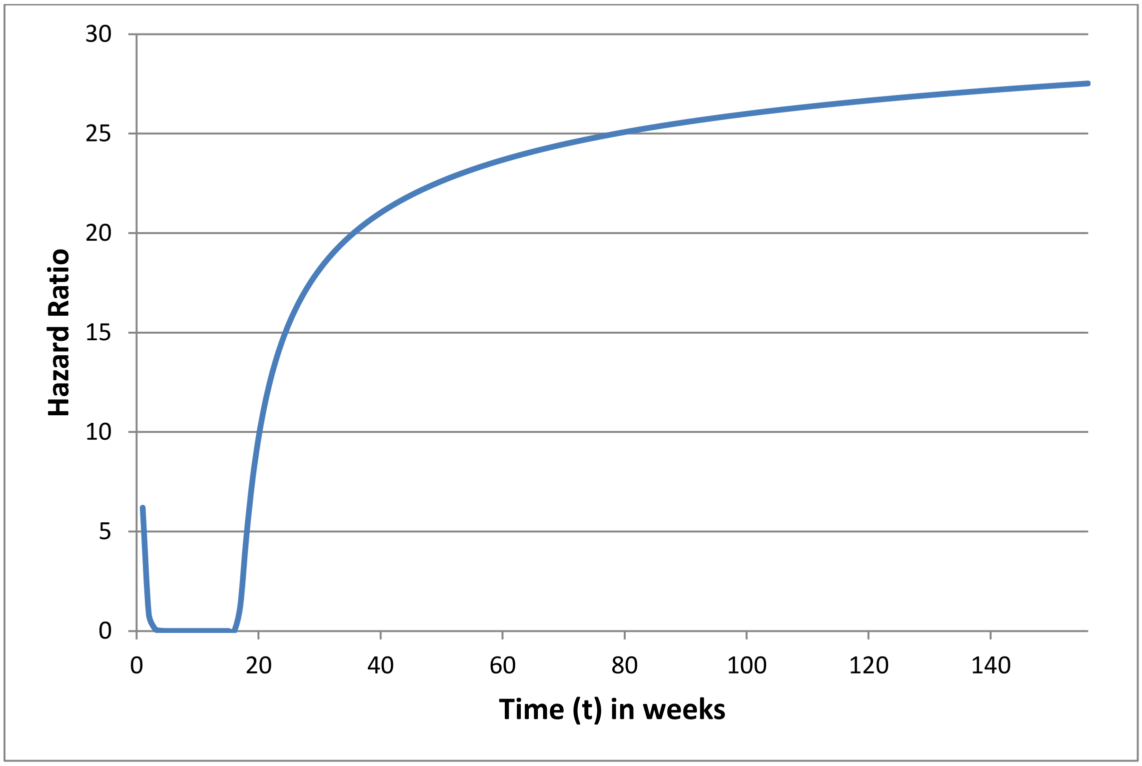


Figure 7.13 – Base model: inactivity hazard ratio over time

The first important note is the magnitude of the scale of the y-axis. Compared to the other estimates, this scale indicates that the magnitude of the effect of inactivity on hazard ratios is very large. The three models initially indicate that inactivity at the start of enrolment increases the hazard ratio significantly. This decreases to near 0 levels until week 18, after which the hazard ratio increases dramatically. The ditch in hazard ratios can be explained by the limited number of inactive observations during these first initial weeks of teaching. After the first teaching period has been completed, then the dramatic increase in the hazard ratios commences. It shows that inactivity is the largest contributing factor to a student's hazard ratio after 18 weeks of enrolment. As such, this is an essential variable that needs to be captured as part of an EAS.

The other mode of study captured by this study is part-time study. Visually, the hazard ratios are presented in Figure 7.14.

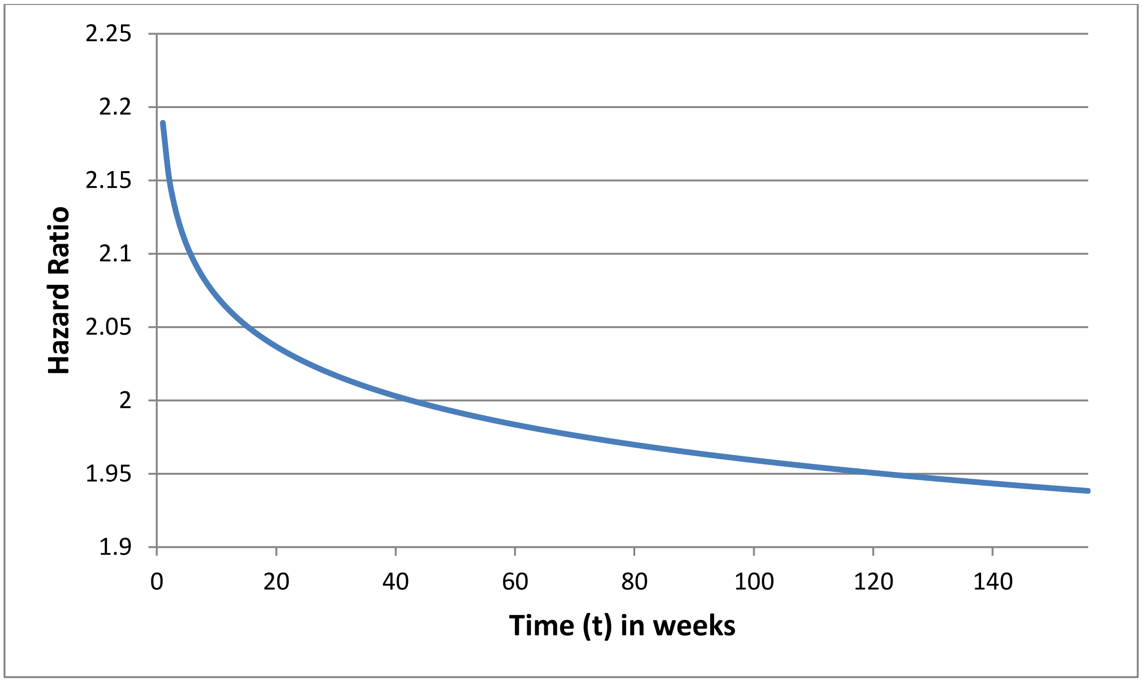


Figure 7.14 – Base model: part-time hazard ratios over time

Figure 7.14 shows that part-time students initially have a significantly higher hazard ratio. This decreases over time, however, after three years, the hazard ratio is still significantly greater than 1. This indicates there is a strong disparity in the risk of discontinuing between part-time and full-time students.

The results show that decreasing the workload of the student increases the hazard ratio and risk associated with discontinuing. This acts as a strong indication that the students' mode of study is a reflection of their level of commitment to their course and remaining enrolled. With respect to EAS design, mode of study becomes an important variable to incorporate. Overall the results indicate that full-time enrolment minimises the risk associated with discontinuing.

7.3.6 Early Alert System effects

The effects associated with the EAS vary relative to the three model configurations. The first captures the short term or immediate effects associated with being identified by the EAS. The second captures the enduring effect of identification, where the EAS variable remains 0 until identification, after which it changes to 1 for the remainder of enrolment. The final model captures the long term effects associated with being identified. The results for the three models are presented in Table 7.6.

Table 7.6 – Base model: Early Alert System effects

	Short-run Effects		Enduring Effects		Long-run Effects	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Early Alert System	2.067[a]	0.214	1.150[a]	0.041	0.533[a]	0.049
Early Alert System x t	0.994[a]	0.002	-	-	-	-
Early Alert System x ln(t) (for t<=18)	-	-	-	-	0.685[a]	0.039
Early Alert System x ln(t - 18) (for t > 18)	-	-	-	-	1.128[a]	0.030

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

All three models show significant effects at the 1% level. Interpreting these effects though depends on the interaction with the variable over time. The short-run and long-run models both interact with different functions of time, while the enduring effects model is independent of time. The plots of the estimated hazard ratios over time are presented in Figure 7.15.

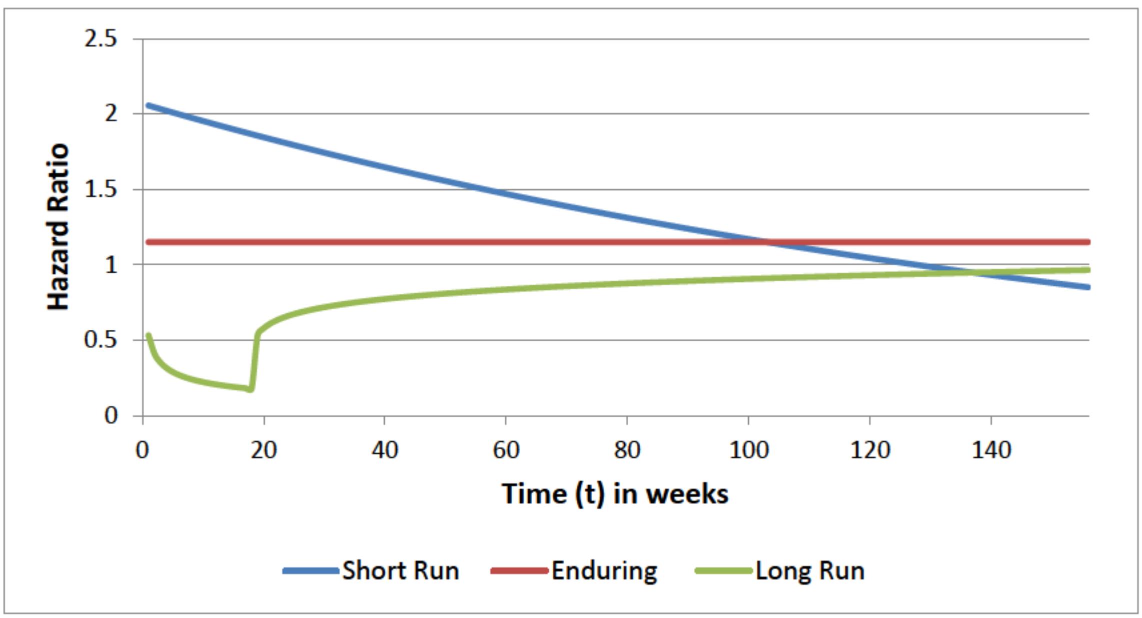


Figure 7.15 – Base model: EAS hazard ratios over time

The short run model shows that students identified by the EAS are most at risk of discontinuing when first enrolled. This is important evidence indicating the efficacy of the system to identify students in need of support. A key caveat, though, is this is a decreasing function, so the longer a student is enrolled, the lower the hazard ratio of the identified student. Figure 7.15 shows that at around 127 weeks, students identified by the EAS have no significant difference in hazard ratio, and past 127 weeks, the students identified by the system actually have a lower hazard ratio than the students not identified by the EAS. This indicates that the EAS may struggle to identify students at risk of discontinuing in the later years of their study.

The enduring effects model shows that students identified by the EAS have an increased hazard ratio 15 per cent higher than students not identified by the EAS. Importantly, this model indicates that once a student is identified by the EAS, they have a significantly higher hazard ratio than students not identified by the EAS. This provides corroborating evidence of the short-run model to indicate students identified have a greater risk of discontinuing.

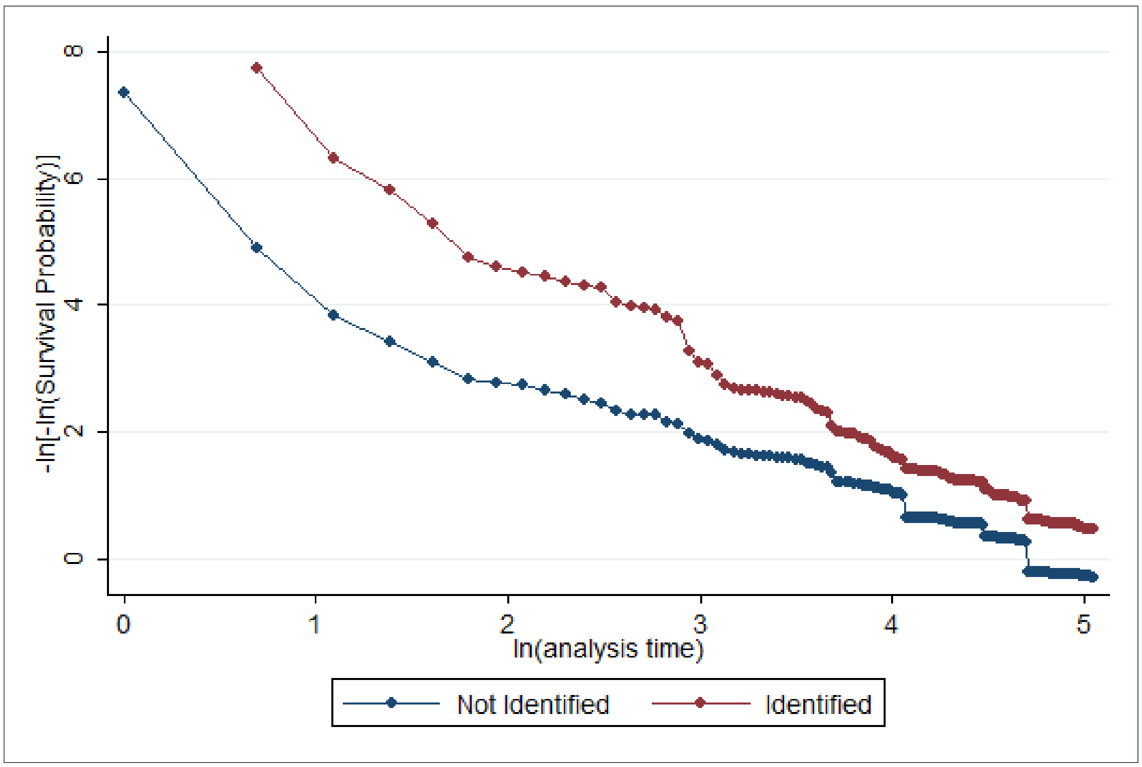


Figure 7.16 – Base model: long run identification hazard ratios over time

The hazard ratio in the long run model captures a discontinuous function over time. Analysing the survival probabilities over time, Figure 7.16 shows the linear relationship for around the first 18 weeks (approx. 2.9 on the logarithmic x-axis scale) with the identified and not identified groups roughly parallel. After 18 weeks there is a rapid decrease for the identified group, bringing the probabilities closer together. The discontinuous nature of the function is reflected in the hazard ratio estimates for the long run model shown in Figure 7.15. The hazard ratio decreases until week 18, after which it quickly reverses its trend and increases at a decreasing rate. The discontinuous nature of the estimates corresponds closely to students completing their first teaching period, and commencing their second teaching period of study. After three years there is close to no significant difference between students identified and not identified by the EAS. The model provides an interesting insight into the function of the EAS. To understand the inner workings of the long-run configuration, conditional analysis within EAS identified and not identified groups is conducted in section 7.4.

7.3.7 Summary of key findings in the base model

In summary, significant *demographic*, *institutional*, *student performance* and *workload* effects were detected using the base survival analysis results. The significant findings show that *gender*, *age* and *ATSI* status (or other racial variables) should be factored in to the design of Early Alert Systems. With respect to the institutional aspects, any EAS should account for hazard ratio changes over time. This is highlighted in the case of *international fee*-paying students, where the hazard ratio has a rapid increase close to completion. This also supports the use of survival analysis and other temporal models to capture the true effect associated with variables over time. When comparing course types, *advanced diploma* students also have a significant difference from normal *bachelor entry* students. The hazard ratio is not constant and increases the longer the student is enrolled. Students who enter through *graduate entry* or directly into *honours* programs have lower hazard ratios than bachelor students admitted through traditional entry. The variation between *schools* shows that there is scope to incorporate school-specific effects within the EAS. In particular, school 6 needs further analysis to establish the significant issues causing the increase in the hazard ratio over time. The between-school differences also represent a level of necessary customisation any EAS system should undertake when deployed at an institution level.

In considering student performance, negative grade outcomes need to be adequately factored in. Given that *withdrawn early* grades occur within the first few weeks of a teaching period, before the financial census date, this should be a major predictor for timely identification of a student who is intending to discontinue their studies. However, timing is also important, with the observed effect of *withdrawn early* grades only occurring within the first year of study. Furthermore, the model shows that the student performance can be treated cumulatively. That is, the total affects associated with a student’s academic performance is the sum of the estimated coefficients. The results show that a student who attained three passes and a fail in a teaching period, would have a reduced hazard ratio overall. Inactivity of students also needs to be factored into the EAS algorithm. Periods of inactivity, regardless of whether the institution is informed of the student's intent to take leave from studying, indicates a significantly higher hazard ratio across most time periods. In terms of magnitude of effect, inactivity contributes the largest increase in the estimated hazard ratio. In designing an EAS, this should be able to capture inactivity as a major predictor of discontinuation.

With respect to an EAS, results indicate that the system is identifying students at risk of discontinuing their studies. The results of the short-run model show that a student identified in any given week have a hazard ratio significantly higher than students not identified. This effect decreases over time however, confirming that students face the greatest risk of discontinuing early within their course. The effect in the enduring model showed a significant effect that once a student was identified, the student remained at an increased hazard ratio for the remainder of their enrolment. The long-run model indicated that overall, students who were identified by the EAS have decreased their hazard ratio earlier on in their studies, however this changes leading into the second teaching period where there is a rapid increase in the hazard. Interestingly, by the third year of enrolment, there is only minimal difference between students identified and not identified by the EAS. This indicates that over time the EAS may be normalising the risk profiles of identified and non-identified student.

A limitation on the interpretation of the result for the long-run base model, and those presented in chapters 5 and 6, comes from the process of dividing the sample into these two groups. The group of students never identified captures those students who did discontinue and did not receive the targeted support they needed (Type 1 error). The group of identified students would include those students identified for support but not in need of targeted student support (Type 2 error). Additionally, the time varying effects associated with the long-run base model require additional probing. As such, the conditional model was developed to test for differences within the two groups. Variations between the two groups will identify relative significant characteristics which can aide in understanding any effect the EAS had on student's hazard ratios. The interaction model analyses the interaction effects associated with the EAS to identify which variables the EAS effects.

7.4 Conditional model results

7.4.1 Conditional survival model description

In this model, the two groups of students used in the long-run model are separately estimated using the same *demographic*, *institution, student performance and workload* variables from the base model. This is reflected in Figure 7.17, where the black pathway shows students who did not interact with the early alert system, the 'No-EAS' group. The red pathway shows the students who were identified and as a result, some effect from the program was induced. This is the 'With-EAS' group.

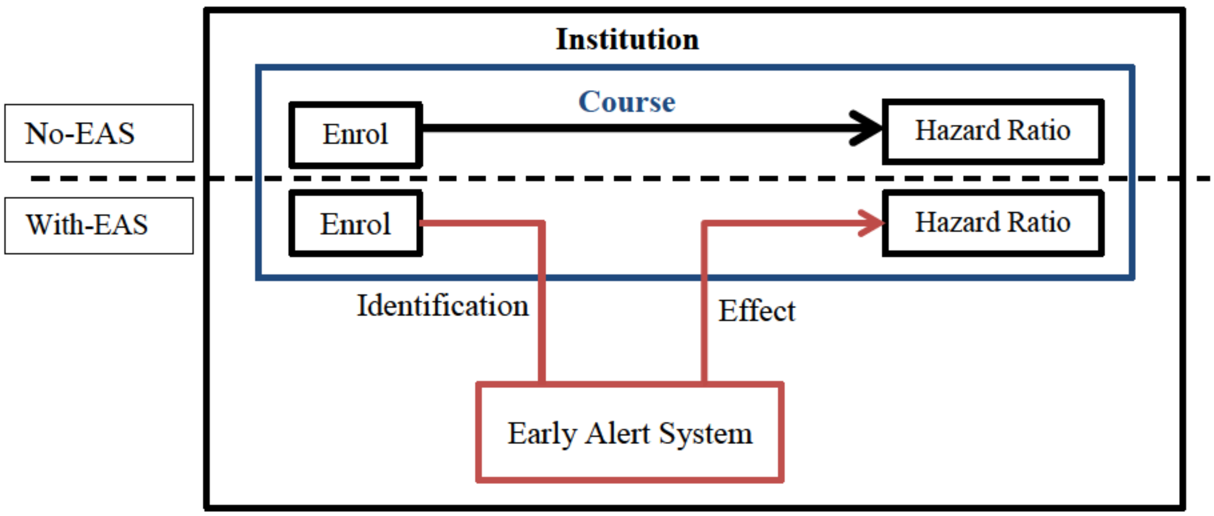


Figure 7.17 – Conditional model: effects under the condition of EAS identification (Authors own contribution)

The purpose of analysing under the condition of EAS identification, is to identify the factors that vary between the identified and non-identified groups. Given the results from the base model, the expected results for conditional model variables could fall into one of two classes. The first possible result of separation of the students into two groups; the group of students under the condition of no EAS identification captures a set of students that are homogenous with respect to their hazard ratio. That is, there should be no difference in the hazard ratio between students within this group. This means the observed effects of the long-run base model would be transferred to all identified students, captured by students under the condition of being identified. The second possible result is where significant factors are constant between the two groups. This would mean that the variables are independent of the identification and effect of the EAS. These two possible results, while providing important information to an EAS, may also indicate that different approaches are needed in how targeted support is conducted.

7.4.2 Significance of the conditional survival model and assumptions tests

The No-EAS and With-EAS overall results are presented in Table 7.7.

Table 7.7 – Conditional model: overall significance and assumptions test

No-EAS		With-EAS	
LR $\chi^2(41)$	1698.85	LR $\chi^2(43)$	5628
Prob > χ^2	0	Prob > χ^2	0

PH Test		PH Test	
$\chi^2(41)$	35.56	$\chi^2(43)$	32.53
Prob > χ^2	0.7107	Prob > χ^2	0.8775

The results indicate that the models are significant at the 1% level. Both models also uphold the proportional assumptions tests with no significant relationship between the base hazard function and time. This indicates that the models provide meaningful survival analysis results.

7.4.3 Demographic variables

Demographic comparison between students identified and not identified by the EAS over the long term reveals similarities and differences in the two groups. The results are presented in Table 7.8. Comparing the two groups of students, *gender* is only significant in the With-EAS group at the 5% level. The results indicate that female students within the identified group had a hazard ratio 9.3 per cent higher than their male counterparts, while no significant effect was observed in the No-EAS group. This is in line with expectations, where by, if the EAS is functioning correctly, those with a higher hazard ratio should be in the identified group and not in the unidentified group.

Age is statistically significant at the 1% level in both groups. The estimated coefficients for the hazard ratios are nearly identical, indicating that the two groups contain similar age profiles and the effect of age is constant between the two. This indicates that despite age being a significant indicator of the likelihood of a student dropping out, the EAS as it currently functions does not discriminate on this factor.

Table 7.8 – Conditional model: demographic variables

	No-EAS		With EAS	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Gender	1.056	0.055	1.093[b]	0.043
Age	0.966[a]	0.011	0.967[a]	0.010
Age Squared	1.000[a]	0.000	1.000[a]	0.000[d]
ATSI	0.433[b]	0.15	0.565	0.198
ATSI x t	1.014[a]	0.005	1.018	0.012
ATSI x t^2	-	-	1.000	0.000[d]

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Finally, *ATSI* status shows a significant effect at the 5% level in the no-EAS group, while no effect is present in the with-EAS group. Additionally, testing against the proportional hazards assumptions showed that this variable for both groups of students varies over time. The results for the No-EAS *ATSI* effect are shown in Figure 7.18.

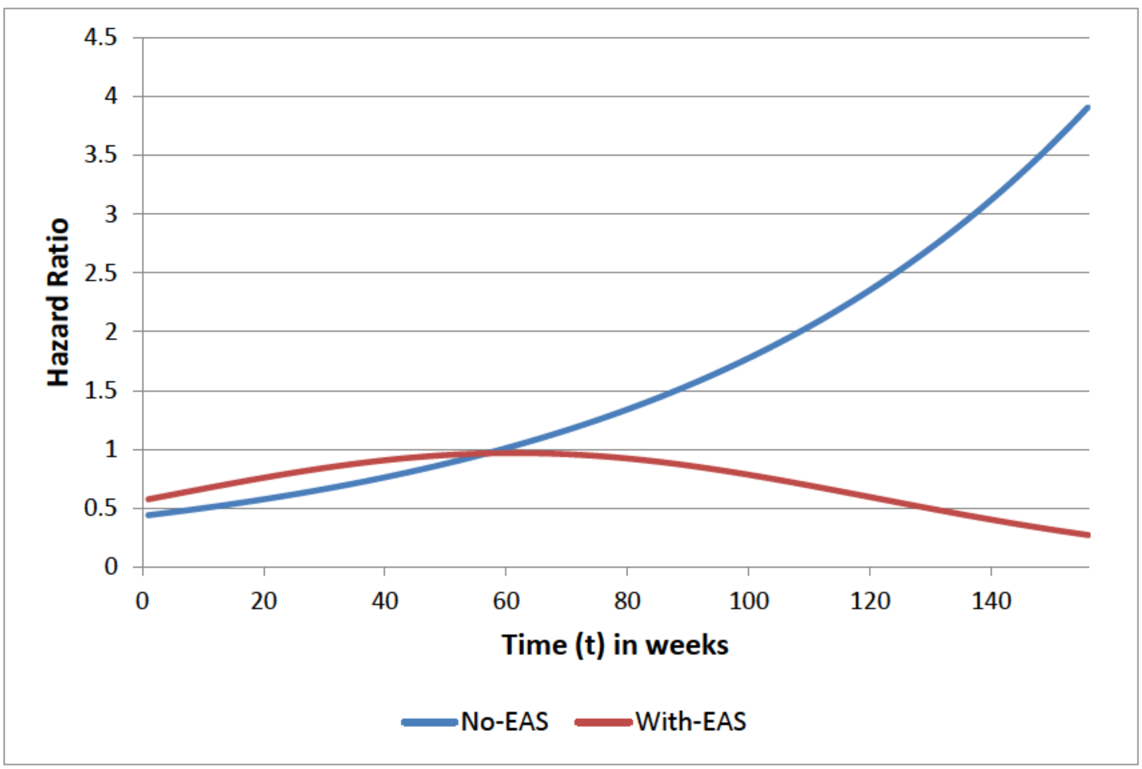


Figure 7.18 – Conditional model: ATSI hazard ratios over time

Initially, ATSI students not identified by the EAS have a significantly lower hazard ratio than non-ATSI students not identified by the EAS. This difference in hazard ratios closes over the first year of study, around which there is no significant difference between ATSI and non ATSI students not identified by the EAS. After the first year of enrolment, the hazard ratio for ATSI students continues to increase significantly. The increased hazard of discontinuing indicates there may be an issue for ATSI students not identified by the EAS, capturing possible EAS model misspecification. For the with-EAS group, there is no significant difference in the hazard ratios of ATSI and non ATSI students. This may indicate that ATSI students within this group have risk profiles which are normalised to be similar to non ATSI students.

7.4.4 Institutional variables

The institutional differences between students identified by the EAS in the long run are presented in two tables. Table 7.9 captures *international fee* paying student, *prior study*, *on-campus* and *course* type effects. The school effects are presented later in Table 7.10.

Table 7.9 – Conditional model: institutional variables

	No EAS		With EAS	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
International Fee	0.728	0.301	0.193[a]	0.064
International Fee x t^2	-	-	1.001[a]	0.000[d]
International Fee x t^3	-	-	1.000[b]	0.000[d]
Prior Studies	0.983	0.06	0.767[a]	0.059
Prior Studies x t^2	-	-	1.000[a]	0.000[d]
Prior Studies x t^3	-	-	1.000[a]	0.000[d]
On-campus	0.918	0.105	1.220[a]	0.065
On-campus x t			0.909[a]	0.016
Diploma	0.943	0.145	1.013	0.169
Advanced Diploma	0.720[b]	0.106	0.855	0.091
Advanced Diploma x t	1.005[b]	0.002	-	-
Bachelors (Graduate)	0.810[b]	0.081	0.717[a]	0.066
Bachelors (Honors)	0.350[a]	0.109	0.637[a]	0.106
Bachelors (Honors) x t	1.011[b]	0.005	-	-

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The institutional variables used in the conditional model excludes domestic fee paying students. This category of students had a small sample size originally and when dividing them into no-EAS and with-EAS, the sample size was too small for the model to reach convergence of a result. However, this variable was also insignificant in the base model, so dropping it from the conditional model is of no significant consequence.

The results for international students show interesting results. There is no significant difference in the hazard ratios of students not identified by the EAS on the basis of fee type. However, within the group of students identified by the EAS, there is significant variations between international and HELP students. The plot of hazards over time shown in Figure 7.19 captures the effects identified in the previous section.

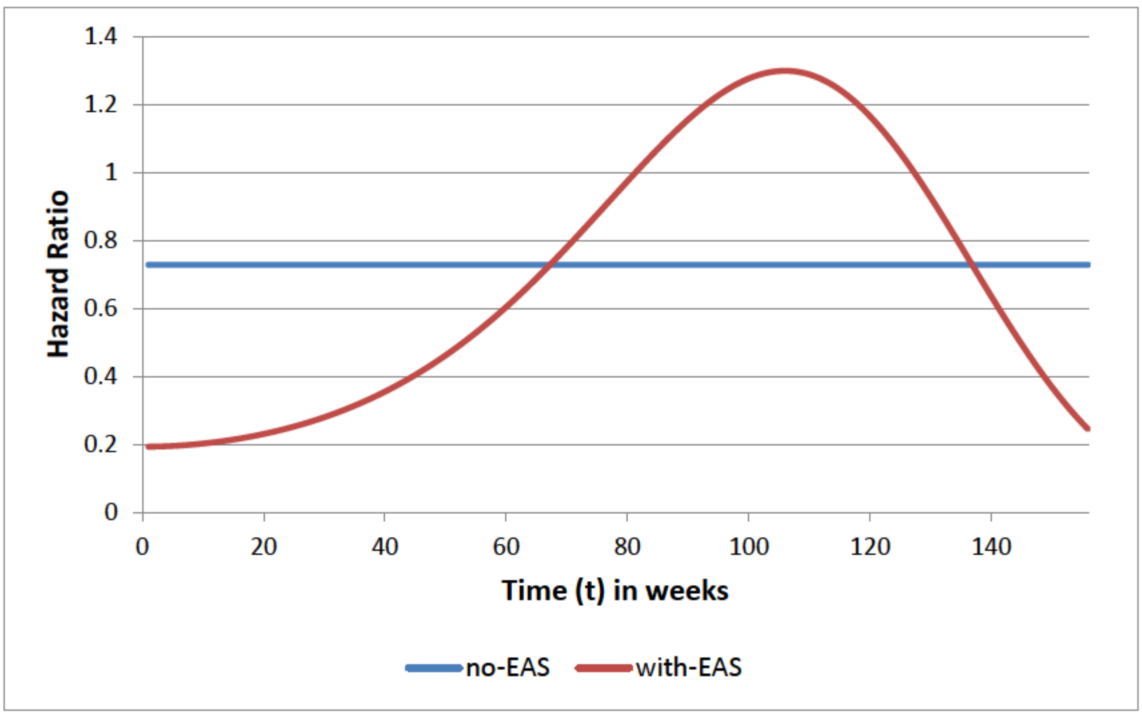


Figure 7.19 – Conditional model: international fee hazard ratios over time

The results show that international students identified by the EAS start with a lower hazard ratio, which increases to peak at the two year mark. After two years, the hazard ratio decreases again. This indicates that even within the identified group of students, there is a difference arising in hazard ratios between international and domestic HELP students. Since this effect occurs within the identified group, this may be more of a systematic effect which is not addressed by the current EAS design. It would be expected that if international students identified by the system were receiving targeted support that the hazard ratio would normalise similar to the non-identified group. That it peaks around the two year point with a significantly higher hazard ratio indicates student support needs additional focusing.

Prior studies show an effect in the with-EAS group at the 1% level, indicating students who have done prior studies and are identified by the EAS have a significantly lower hazard ratio than students without prior studies. There is a time varying effect which represents a minimal increase in hazard over time. After three years the hazard ratio for students with prior study identified by the EAS is 15 per cent lower than students with no prior study identified by the EAS. Compared to the non-identified group, there is no significant difference between the prior study and no prior study groups when not identified by the EAS. This may indicate that students who have undertaken prior studies have a greater willingness to engage with targeted support programs compared to students who have not studied previously. This is an important finding, showing that students respond differently to the EAS depending on past experience.

When comparing on-campus to off-campus online students, there is no significant difference between the groups in the no-EAS category, indicating similar hazard ratios. The with-EAS group shows a significant difference, whereby on-campus students have a hazard ratio 22 per cent higher than off-campus students initially. However, this rapidly changes over time, decreasing where by week four, there is no difference between on-campus and off-campus students identified by the EAS. For on-campus students, the hazard ratio continues to decrease relative to off-campus online students. This is the same effect observed in the base model indicating a major challenge for institutions operating with a significant off-campus online cohort. There are clear differences in the hazards of these students, even when identified by the EAS. How institutions provide support and change the enrolment trajectory of off-campus online students is a major challenge for the future that will require further research.

Course type had varying levels of significance in both models. In the with-EAS group, honours and graduate entry students had a significantly lower hazard ratio than normal entry bachelor level students. Both diploma and advanced diploma students had no significant difference in this model. In the no-EAS group the results were different. Both advanced diploma and honours students not identified by the EAS had hazard ratios that increased over time. The results for all courses in the no-EAS group are plotted in Figure 7.20.

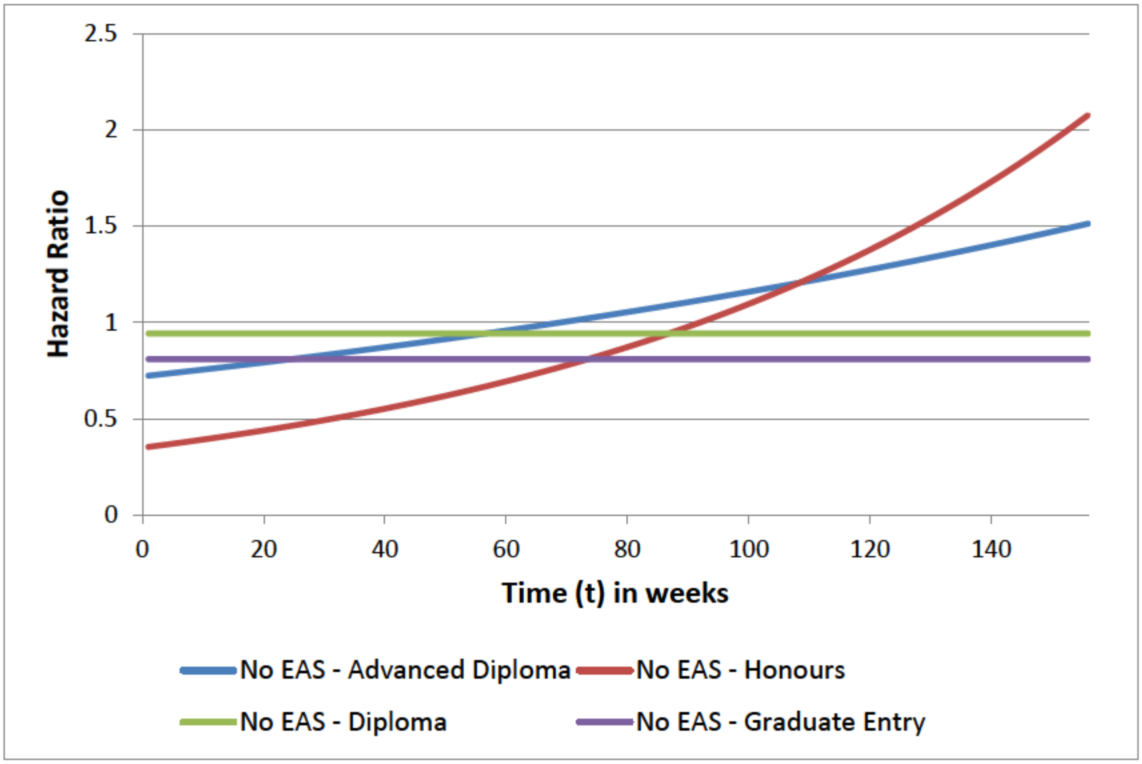


Figure 7.20 – Conditional model: course type hazard ratios over time

The results show that after 60 weeks of enrolment, there is little to no difference in hazards between advanced diploma and bachelor students not identified by the EAS. Over time however, advanced diploma students not identified by the EAS face an increase of the hazard ratio, ending up around 50 per cent higher after three years. An important result is also the effect on honours students. Figure 7.20 shows that at around 90 weeks of study, there is no difference in hazards between bachelor and honours students not identified by the EAS. As time progresses, honours students have a higher hazard ratio than normal entry students. This may indicate a systematic issue associated with the expectations placed on honours students with respect to performance. This warrants a more detailed analysis of the course progression of honours students not identified by the EAS to determine the factors creating this effect.

The no-EAS and with-EAS groups show significant differences and variations between the schools within the university. The results are presented in Table 7.10.

Table 7.10 – Conditional model: institutional variables – schools

	No-EAS		With-EAS	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
School 1	0.888	0.101	1.230[b]	0.106
School 2	2.644[a]	0.611	1.165[c]	0.092
School 2 x 1/t	0.024[a]	0.017	-	-
School 3	1.040	0.119	1.190[b]	0.099
School 4	1.211[c]	0.129	1.322[a]	0.103
School 5	1.154	0.133	1.380[a]	0.118
School 6	0.883	0.092	1.119	0.088
School 7	11.058[a]	8.386	0.717	0.314
School 8	1.035	0.144	1.027	0.100
School 9	0.771[b]	0.099	1.006	0.104

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The results for the no-EAS group show that schools 2, 4, 7 and 9 have significantly different hazard ratios compared to the base case school when students are not identified by the EAS overall. For school 4, the results indicate a higher hazard ratio significant at the 10% level. School 7 has a significantly higher hazard ratio. Further investigation reveals that this school is a small sample with 212 students overall, with only four students not identified by the EAS. As such this result is relatively biased to the small sample and explains why school 7 is omitted from later analysis. The results for school 9 show that of the students not identified by the EAS, the overall hazard ratio was 22.9 per cent lower than the base case school. This highlights school 9 as a possible school of best practice independent of the EAS. An area for future analysis would be to analyse student enrolment patterns within school 9 to determine the sources of the reduced hazard. The hazard ratios for school 2 also vary over time, so the results are plotted in Figure 7.21.

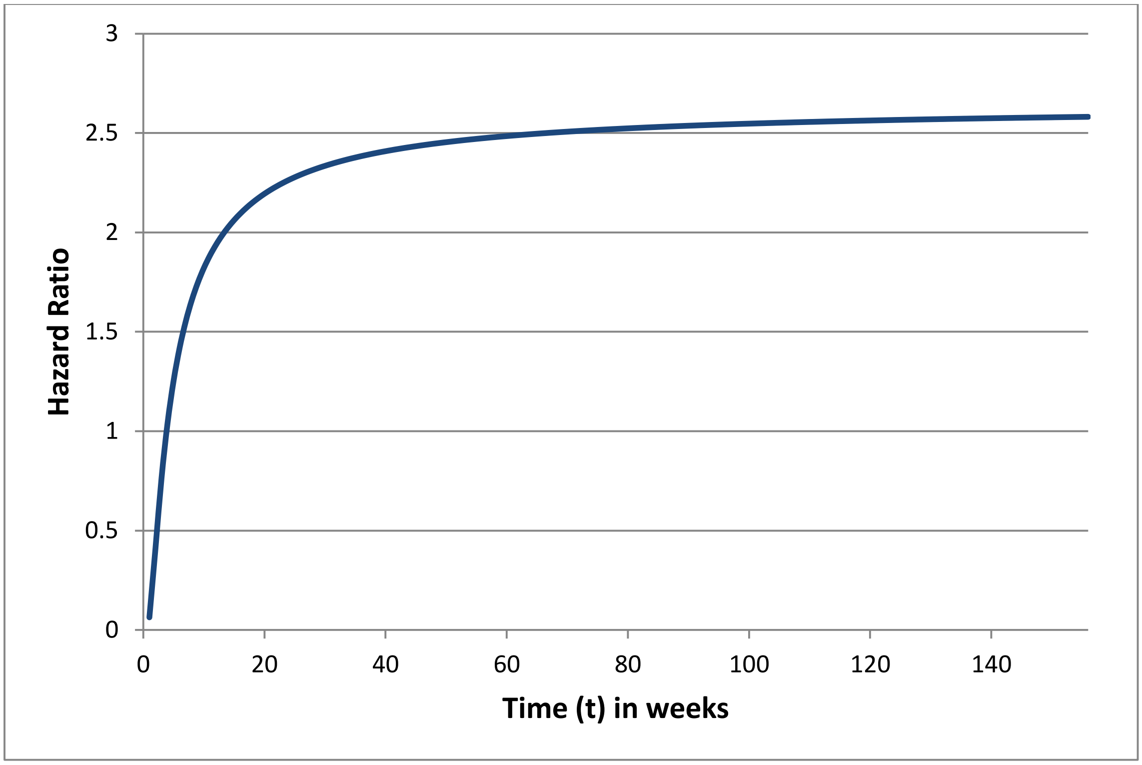


Figure 7.21 – Conditional model: school 2 hazard ratios over time

The results show that students who initially enrol in school 2 and are not identified by the EAS have a significantly lower hazard ratio. This changes after a few weeks and the hazard ratio is around 150% higher in school 2 than the base school after 1 year of study. This only applies for the students not identified by the EAS and as such, this may indicate that there is some level of model misspecification of the EAS not capturing the true risk of discontinuation within school 2.

For the group with EAS, the results are also scattered. Schools 1 to 5 exhibit statistically significant hazard ratios higher than the base case school. The results for school 2 are only significant at the 10% level, however given the results from the no-EAS group, this indicates that school 2 overall has an increased hazard ratio. Schools 4 and 5 indicate students identified by the EAS overall are 32 to 38 per cent higher in hazard ratio than students from the base school.

7.4.5 Student performance and workload variables

The student performance and workload variables show complex interaction with time. As such the results are broken into the two sub groups, grades and workload. The results for grades are presented in Table 7.11.

Table 7.11 – Conditional model: student performance

	No EAS		With EAS	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Withdrawn	1.074	0.050	1.725[a]	0.221
Withdrawn x ln(t)	-	-	0.877[a]	0.027
Withdrawn Early	2.281[a]	0.431	1.063[a]	0.015
Withdrawn Early x ln(t)	0.839[a]	0.039	-	-
Fail Incomplete	2.567[a]	0.896	1.408[a]	0.064
Fail Incomplete x ln(t)	0.864	0.077	-	-
Fail Incomplete x sqrt(t)	-	-	0.984[a]	0.005
Fail	1.363[a]	0.108	1.137[a]	0.021
Pass	0.796[a]	0.042	0.851[a]	0.012
Credit	0.691[a]	0.073	0.880[a]	0.031
Credit x t	1.002[c]	0.001	1.000	0.000[d]
Distinction	0.829[a]	0.033	0.871[a]	0.015
High Distinction	0.643[a]	0.079	0.895[a]	0.020
High Distinction x t	1.002[c]	0.001	-	-
Other	0.298[b]	0.154	0.770[a]	0.056

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The no-EAS group shows that for students not identified by the EAS, withdrawing from a unit has no significant effect on the students' hazard ratio. The with-EAS group is significant with respect to withdrawing, and captures the interaction with time. Plotting the hazard ratios over time in Figure 7.22 shows that for the with-EAS group, the hazard ratio decreases over time when withdrawing from a unit.

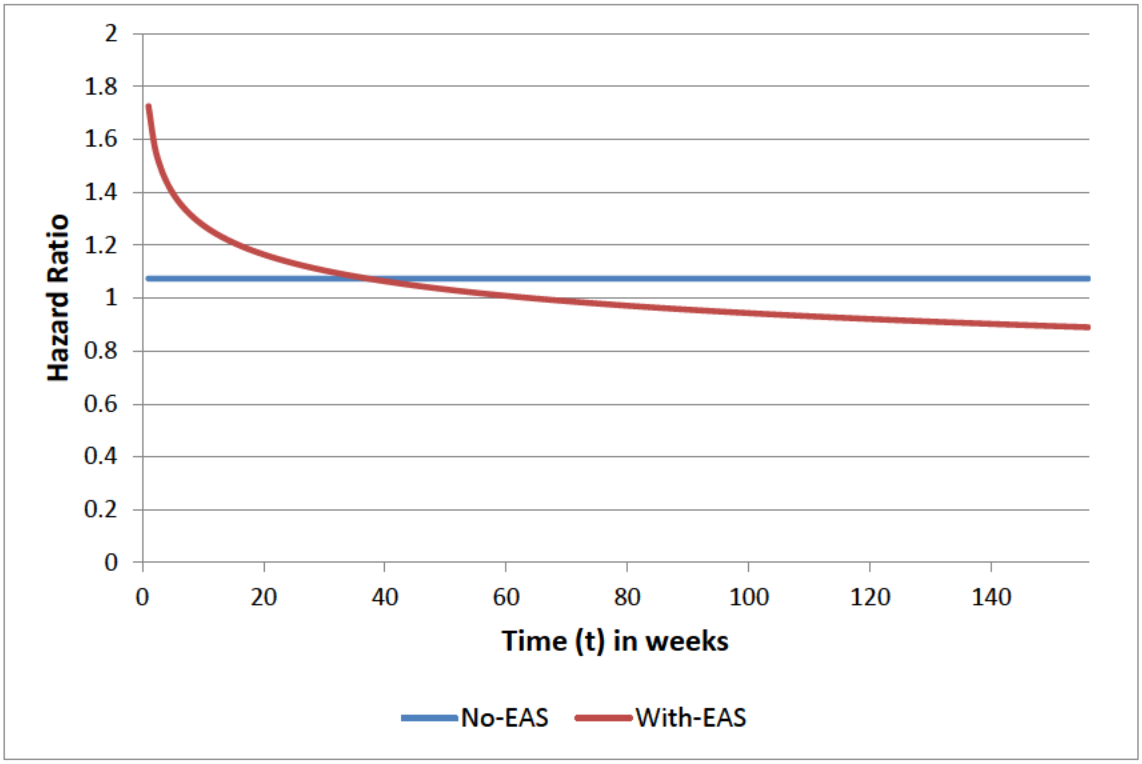


Figure 7.22 – Conditional model: withdrawn grade hazard ratios over time

This indicates that, in the long run model, the only time when students have an increased risk associated from withdrawing is in the first year of studying. After that point, withdrawing from a unit, even if identified by the EAS, will not increase the hazard of discontinuing.

Both models show a significant increase in the hazard ratio associated with *withdrawing early* from a unit. The with-EAS group shows a constant increase independent of time, indicating that students identified by the EAS who withdraw early increase the hazard ratio by 6.3 per cent. This effect is also cumulative, so the more units withdrawn from early, the larger the effect. In the no-EAS group, the effect interacts with time and the hazard ratios are shown in Figure 7.23.

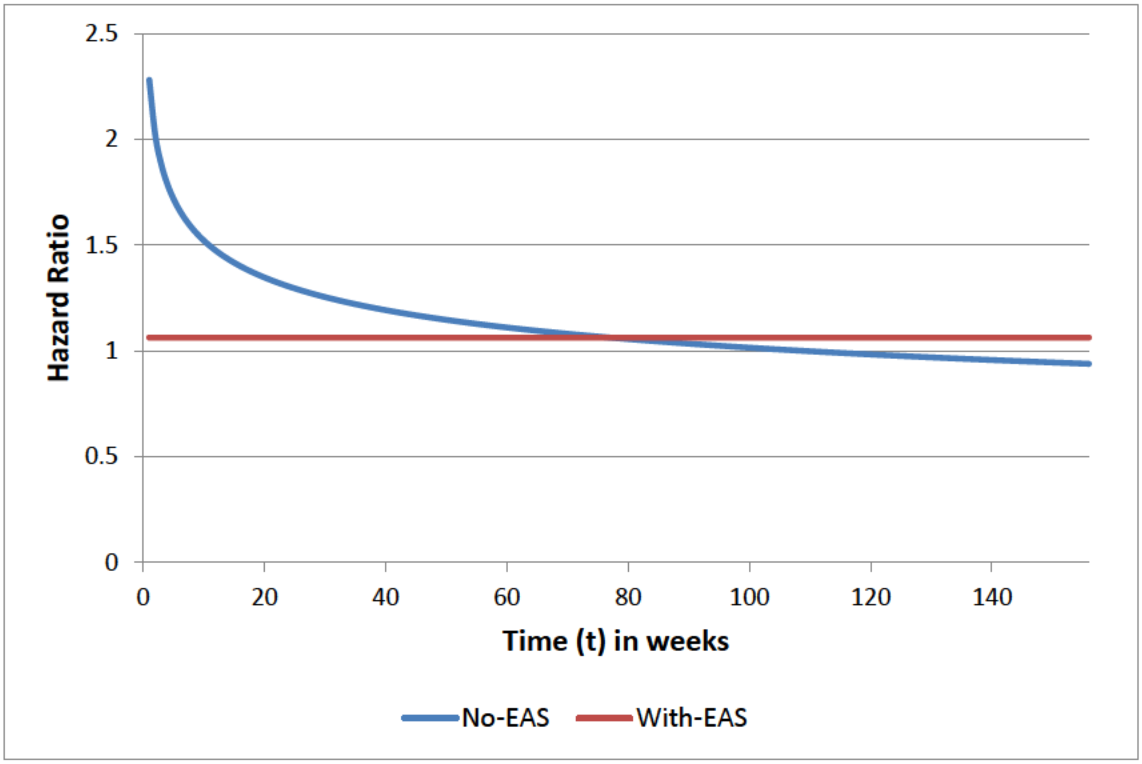


Figure 7.23 – Conditional model: withdrawn early grade hazard ratios over time

The results show that students not identified by the EAS who withdraw early from units, especially in the first year of study, have a significantly higher hazard ratio. The effect decreases over time whereby around week 90 there is no effect. Since withdrawing early is significant in both models irrespective of identification, this is a strong indicator of increased hazard of discontinuing. This should be a fundamental component of any EAS design.

A *failing incomplete* grade is also a complex variable for both models with time interactions. The results for the no-EAS group show that the time interaction variable ln(t) was necessary to capture interactions to satisfy the proportional hazards assumption test, however the interaction is not significant over time. For the with-EAS group, the interaction with time is significant, and the hazard ratios for failing incomplete are presented in Figure 7.24 over page.

The graph shows that there is a slight decrease over time for students identified by the EAS receiving the fail incomplete grade outcome. While both results are significant, there is a large difference between the magnitude of the effects in the no-EAS and with-EAS groups. Overall, students identified by the EAS have a lower hazard ratio when receiving a *fail incomplete* compared to students not identified by the EAS receiving the same grade. This may be a direct effect associated with students receiving support resulting from identification; however without having data from within student support, this is not a strong link.

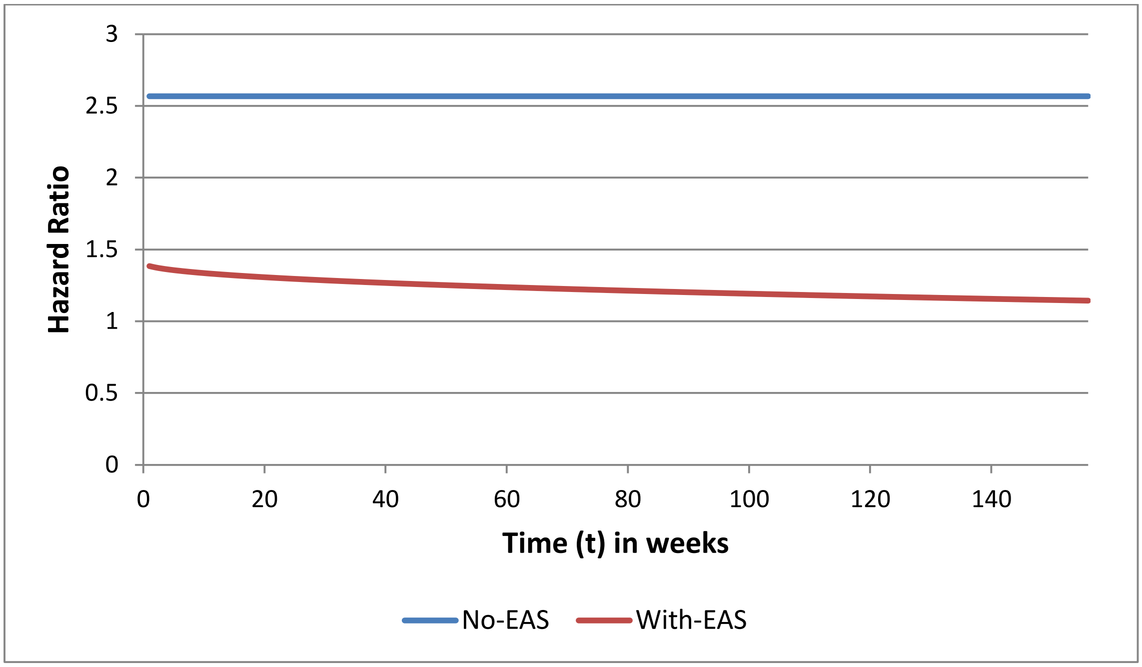


Figure 7.24 – Conditional model: fail incomplete grade hazard ratios over time

Failing a unit is a more constant effect with both models, indicating significant hazard increases that do not change over time. For students not identified by the EAS, the effect of receiving one fail grade is a 36 per cent increase in the hazard ratio. In the case of students identified by the EAS overall, the increase is only 13.7 per cent. Again, the difference in the increase in hazard ratios depending on identification is possible evidence of the positive effect resulting from EAS.

With respect to the positive grades, all show significant positive reductions in the hazard ratio if a student received a *pass*, *credit*, *distinction* or *high distinction*. Both *credit* and *high distinction* have time varying components, indicating a slight increase in the hazard ratio over time. This is relatively small effect and outweighed by the positive effects of attaining these grades. As expected, attaining grades which contribute to progressing through a course decreases the hazard of the student discontinuing.

Workload as identified in previous models is a critical variable that affects the hazard ratio of students. The complex interplay between workload and time means that capturing the effect without violating the proportional hazards assumptions is difficult. As shown in Table 7.12, the functions for both inactivity and part-time have significant interactions over time.

Table 7.12 – Conditional model: workload results

	No EAS		With EAS	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Inactive	14.151[a]	2.857	132.316[a]	55.414
Inactive x t (for t <= 16)	0.126[a]	0.100	-	-
Inactive x 1/ln(t) (for t > 16)	0.046[a]	0.014	-	-
Inactive x 1/ln(t) (for all t)	-	-	0.000[a]	0.001
Part-time	0.596[a]	0.068	0.533[b]	0.150
Part-time x 1/ln(t)	-	-	1.369[a]	0.117

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The estimated hazard ratios for inactivity are plotted in Figure 7.25.

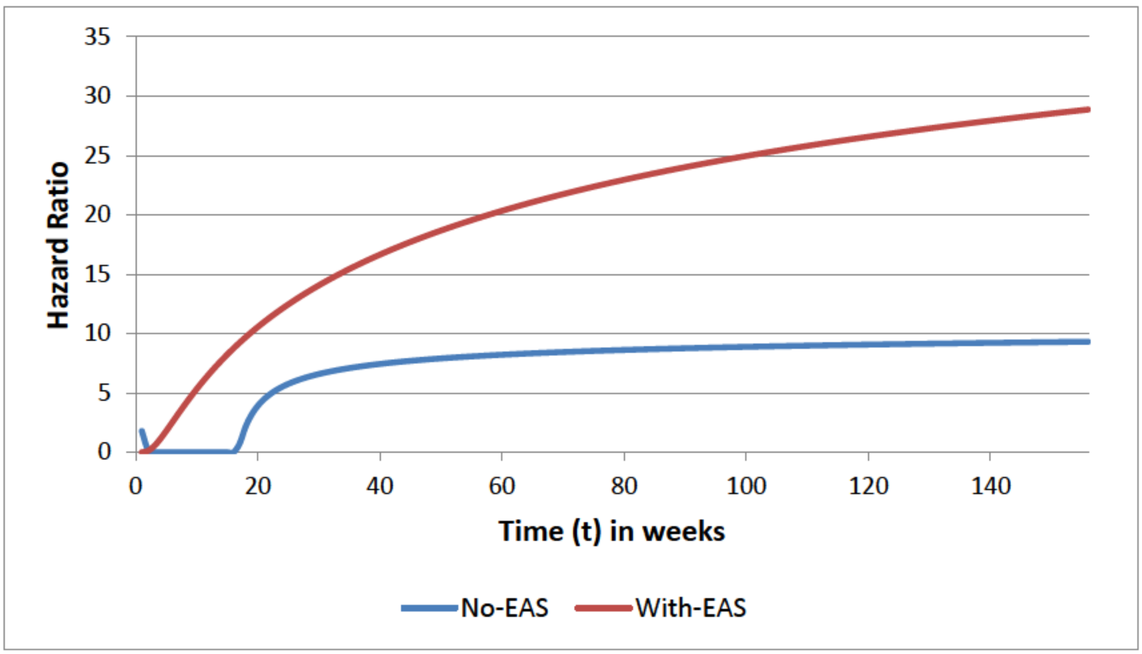


Figure 7.25 – Conditional model: inactivity vs full-time hazard ratios over time

The no-EAS group shows that students overall who are inactive at the start of their enrolment face a higher hazard ratio than full-time students. The results also show that this decreased rapidly over the first 16 weeks of enrolment, to then quickly increase again. This effect is actually a function of the underlying data, where very few students become inactive over the initial teaching period. The sharp increase in hazard does correspond to the completion of the first teaching period. For students identified by the EAS, the increase in hazard occurs much sooner than those not identified. As previously discussed, inactivity is one of the strongest indicators of increased risk of discontinuing. As such, this needs to be captured as part of the EAS, factoring in how long the student has been enrolled.

Comparing part-time students to full-time students is also a complex relationship. Part-time students, while having a lighter workload, are part-time for a reason. This can be competing life priorities which affect the students' capacity to study and progress. The estimated hazard ratio for part-time students is shown in Figure 7.26.

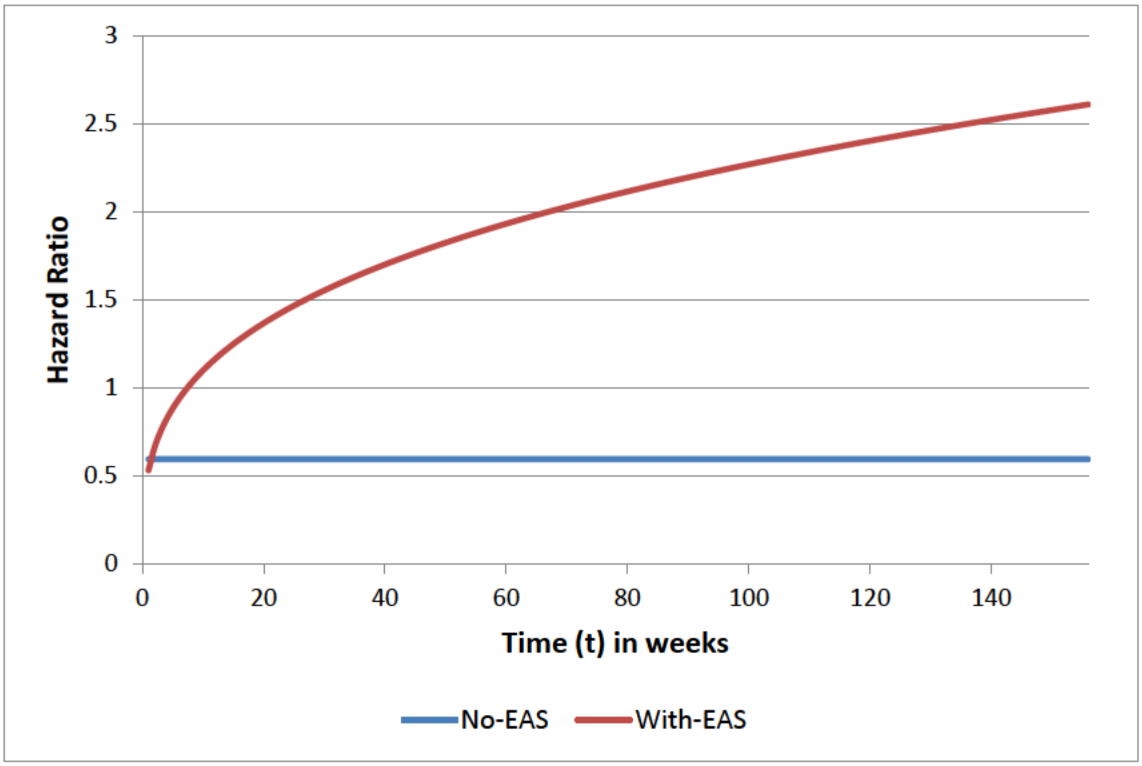


Figure 7.26 – Conditional model: part-time vs full-time hazard ratios over time

Compared to full-time students, part-time students not identified by the EAS have a significantly lower hazard ratio. This indicates that the lighter workload reduces the hazard ratio. However if the student is in the identified category, the risks associated with study part-time increase over time. At six weeks, there is no significant difference between part-time and full-time hazard ratios. After six weeks, the hazard ratio continues to increase for part-time students identified by the EAS. This indicates that if a part-time student is identified by the EAS, there may need to be additional support requirements above that of full-time students.

7.4.6 Summary of findings – Conditional model

In summary, the conditional model expands upon the findings of the long run model, where it was shown students who were identified by the EAS had a significantly lower hazard ratio than students not identified by the system. Many of the significant factors detected in the long run model can be attributed to those students identified by the EAS, such as gender, fee status, prior studies, on-campus attendance and variations between schools. Other effects not detected in the long run became apparent, such that ATSI students not identified by the EAS had a significantly higher hazard ratio than non-ATSI students also not identified by the system. The differences between schools also became pronounced with school 2 also showing a significant difference over time in the hazard ratio for identified students compared to the base school.

Where the effect between models no-EAS and with-EAS models was similar (*age*, *student performance* and *inactivity*) these factors should be treated as independent of the EAS system. They indicate a significant effect which correlates to the students discontinuing their studies, and should be factored in to the development of an EAS. These are also factors which will affect all students regardless of any intervention or targeted support. In the case of the negative grade outcomes, the estimated hazard ratio is lower for those students identified by the EAS than those not identified by the EAS for fail and fail incomplete. This indicates that while the significance of the effect may be independent of identification, the lower hazard ratio for students with EAS compared to no EAS suggests there is some benefit associated with being identified. This provides supporting evidence that students identified by the EAS have a lower hazard ratio.

A key effect in EAS systems may also have been picked up. Students who had prior study and were identified by the EAS had a much lower hazard ratio compared to students with no prior study. This indicates that EAS design may capture the different willingness to engage with support. Students who have studied previously are likely to be more aware of support options and also have a willingness to utilise the services. This introduces an interesting effect associated with the EAS which is akin to the economic concept of willingness to pay. The concept of willingness to engage could provide a useful measure of EAS effectiveness. Blending economic theory on willingness to pay may help understand better the link between students and their willingness to engage with support services.

7.5 Interaction model results

7.5.1 Interaction model description

The interaction model further investigates the relationships between the EAS and the factors that correlate to the hazard ratio. The model is split into two configurations, a short-run and a long run model. Figure 7.27 shows the short-run interaction effect of the early alert program tested back against *demographic*, *institutional, student performance* and *workload* variables.

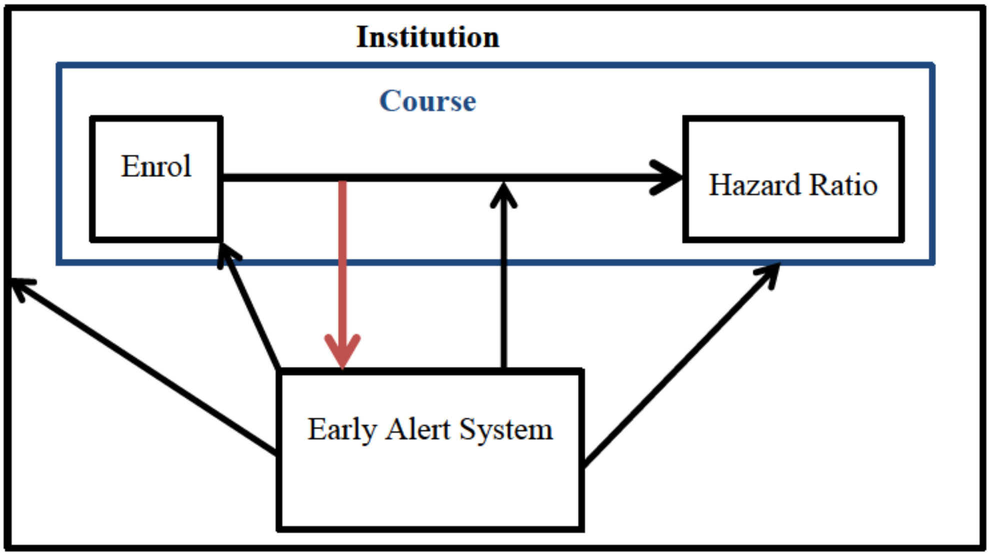


Figure 7.27 – Interaction short-run model
(Authors own contribution)

The red pathway indicates the short-run revolves around the identification process of the EAS at a specific moment of time. The model uses the EAS variable from the base short-run model, which indicated that students identified by the system had a significantly higher hazard ratio than those students not identified by the EAS.

The interactions long-run model tests the same interactions between *demographic*, *institutional, student performance* and *workload* variables, but uses the variable specification from the base long-run model for the EAS effect. This captures the overall effects of the EAS. In the base long-run model, this showed that students identified by the EAS and the resulting interaction with the program, had a lower hazard ratio than those students never identified by the EAS. This is represented diagrammatically in Figure 7.28.

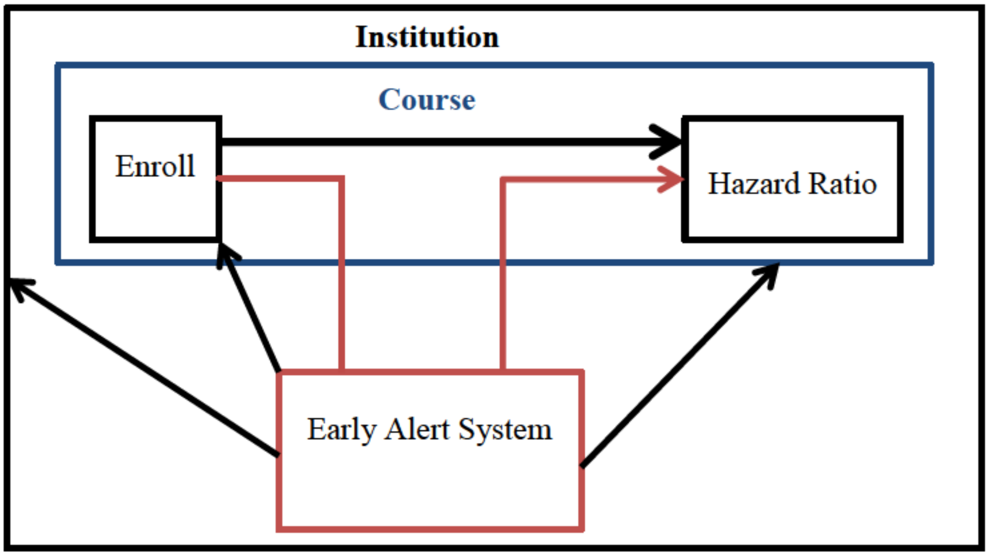


Figure 7.28 – Interaction long-run model
(Authors own contribution)

It is expected that there should be no interaction effects between the EAS and the demographic and institution variables, as these variables are constant throughout a student's enrolment. The effect associated with these variables should not be contingent on identification or any resulting targeted support. If there is a significant effect in the interactions, the interaction coefficient should be greater than 1. If the interaction coefficient is less than 1, this indicates a decrease in the hazard ratio as a result of identification and indicates issues around model misspecification.

The expected result for learning environment variables is different, in that it is expected there should be some interaction effect between the EAS and student grades and the level of workload. This is because these variables change over time and as such, if a student is identified or provided targeted student support, there is scope for students to change their trajectory away from discontinuing to a pathway to course completion.

7.5.2 Significance of interaction model and assumptions tests

The interaction model consists of a battery of interaction terms to test the short run and long run effects associated with identification by the EAS. With 82 and 81 variables for the short-run and the long run models respectively, this presents a relatively complex system. Both models are statistically significant at the 1% level. Testing the assumptions for the models is a complex task whereby not only do the hazards associated with variables fluctuate over time, but so do the interaction effects. Despite the complexity of the systems, both models pass the proportional hazards assumption test. A detailed breakdown of the proportional hazards test is presented in Appendix F. A summary of the results are presented in Table 7.13.

Table 7.13 – Interaction model: overall significance and assumption tests

Short-run		Long-run	
LR χ^2 (82)	7706.61	χ^2 (81)	8006.84
Prob > χ^2	0	Prob > χ^2	0

PH Test		PH Test	
χ^2 (82)	68.22	χ^2 (81)	95.95
Prob > χ^2	0.8622	Prob > χ^2	0.1229

7.5.3 Demographic interactions

The demographic results for model 3 are presented in Table 7.14, and are broken down into base effect and interactions.

Table 7.14 – Interaction model: demographic variables

	Short-run base		Short-run interactions		Long-run base		Long-run interactions	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Gender	1.093[a]	0.035	0.726[b]	0.103	1.058	0.055	1.034	0.068
Age	0.970[a]	0.007	1.020	0.048	0.965[a]	0.011	1.002	0.015
Age Squared	1.000[a]	0.000[d]	1.000	0.001	1.000[a]	0.000[d]	1.000	0.000[d]
ATSI	0.813[b]	0.071	1.552	0.488	1.880[a]	0.454	1.509	0.470
ATSI x 1/t	-	-	-	-	0.000[a]	0.000[d]		
ATSI x t							0.988[a]	0.004

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

As discussed in the overview of results, the models are a complex interaction between variables, time and the EAS. Since the base level effects have been discussed in the base model, the main focus of the analysis is on the interaction effects. From Table 7.14, two significant interaction effects are present. The first occurs in the short run model for female students. The results show that female students have a significantly higher hazard ratio than male students. When the interaction with the EAS is taken into account, the estimated hazard ratio is 0.793. This means female students have a decreased hazard ratio in the short run if identified by the EAS compared to male students not identified by the system. This interaction may indicate that female students are more willing to engage with targeted support in the short term. To make this a solid conclusion, further analysis incorporating student support data would be required.

The other significant interaction effect in the demographic model is associated with ATSI students. Interestingly enough, the interaction effect itself is not significant, however how ATSI students interact with the system over time is. To capture the hazard ratios associated with the base outcome and the interaction effect, the hazard ratios are plotted in Figure 7.29.

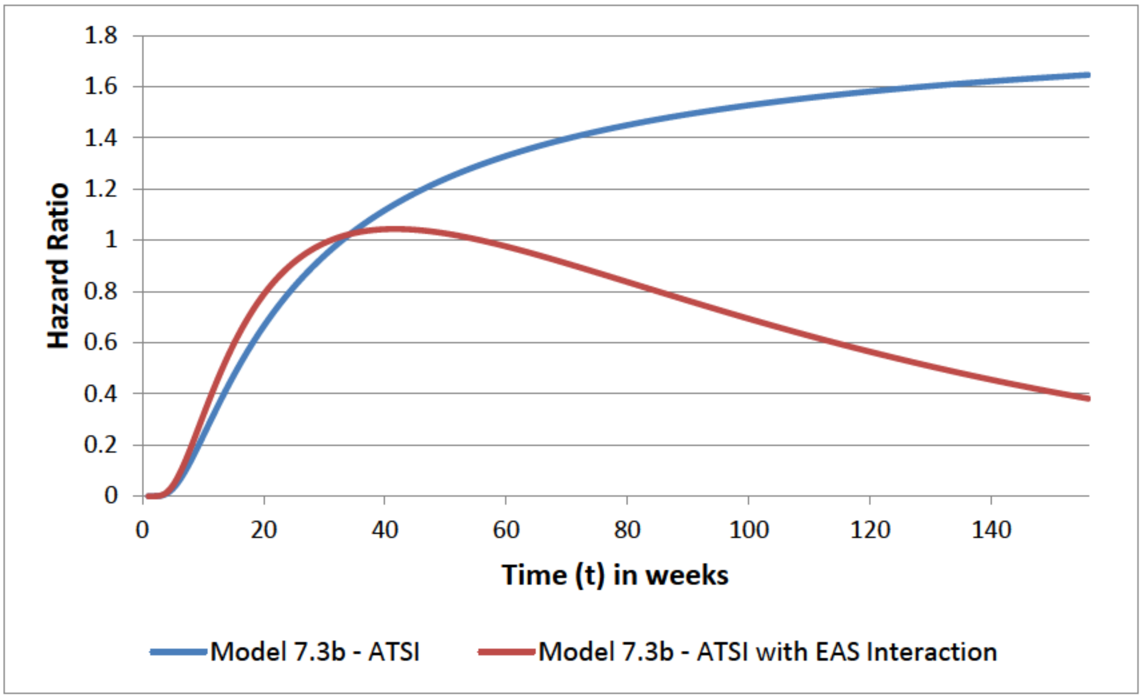


Figure 7.29 – Interaction model: long run ATSI hazard ratios over time

The diagram clearly shows different effects for students interacting with the EAS. Initially, *ATSI* students have a significantly lower hazard ratio than non ATSI students. This increases over time however the hazard ratios diverge depending on the interaction with the EAS. If an ATSI student is not identified by the EAS at any time in their studies, the student will have a higher hazard ratio than non ATSI students after 31 weeks. If the ATSI student is identified at some stage throughout their studies, the results indicate that the hazard ratio will peak around 31 weeks and then start to decrease overall. This is an important finding; the EAS is having a positive effect on the enrolment outcomes of ATSI students overall if they are identified by the system.

7.5.4 Institutional interactions

Testing for interactions against institutional variables reveals that there are no interactions between the EAS in both the short run and long run models, except for within schools. The results are presented in Table 7.15.

Table 7.15 – Interaction model: institutional variables

	Short-run base		Short-run interactions		Long-run base		Long-run interactions	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
International Fee	0.201[a]	0.061	0.778	0.419	0.722	0.298	0.783	0.341
International Fee x t^2	1.000[a]	0.000[d]	-	-	-	-	-	-
International Fee x t^3	1.000[b]	0.000[d]	-	-	-	-	-	-
Prior Studies	0.602[a]	0.102	0.849	0.169	0.988	0.060	0.930	0.070
Prior Studies x ln(t)	1.127[a]	0.048	-	-	-	-	-	-
On-campus	1.153[a]	0.055	1.176	0.219	0.937	0.107	1.217	0.153
On-campus x t	0.911[a]	0.013	-	-	-	-	-	-
Diploma	1.019	0.115	0.795	0.807	0.943	0.145	1.093	0.248
Advanced Diploma	0.792[c]	0.095	1.276	0.513	0.437[a]	0.128	0.859	0.123
Advanced Diploma x t	1.002	0.002	-	-	-	-	-	-
Advanced Diploma x ln(t)	-	-	-	-	1.229[a]	0.090	-	-
Bachelors (Graduate)	0.749[a]	0.052	1.050	0.374	0.800[b]	0.080	0.901	0.123
Bachelors (Honours)	0.616[a]	0.078	1.024	0.755	0.419[a]	0.101	1.062	0.300
Bachelors (Honours) x t	-	-	-	-	1.005	0.003	-	-

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The results for model the interaction short-run model correspond to previously found effects in base short-run model, which indicate: the humped international student hazard distribution over time; the increase in hazards for students who have prior study experience; the decrease in hazard for graduate and honours bachelor students. As stated previously about the expected results, there are no significant interaction effects in the short run model.

The long run model is a little more interesting. Institutional variables for *international fee* paying students, students with *prior study* and *on-campus* students all had significant effects. Taking into account the interactions with the EAS in the long run however, there are no significant effects within these groups. One interpretation of this change is the normalising effect the EAS would have on the risk profile of students identified by the EAS. This suggests that students who had a higher risk profile and subsequently higher hazard ratios, now have estimated hazard ratios that are not significantly different from students not identified by the EAS. To support this conclusion, more detailed analysis of the interaction between the EAS and the student would be required, including information capturing how students were supported after identification.

Analysing the effects within schools, there are significant interaction effects with the EAS. This captures how students from different schools respond to being identified by the EAS. The results are presented in Table 7.16.

Table 7.16 – Interaction model: school variables

	Short-run base		Short-run interactions		Long-run base		Long-run interactions	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
School 1	1.117	0.079	1.089	0.323	0.902	0.102	1.402[b]	0.200
School 2	1.278[a]	0.092	0.737	0.224	1.173	0.138	4.426[a]	1.554
School 2 x 1/t	0.006[a]	0.006	-	-	0.001[a]	0.002	-	-
School 2 x ln(t)	-	-	-	-	-	-	0.720[a]	0.062
School 3	1.196[a]	0.083	0.795	0.227	1.045	0.119	1.142	0.161
School 4	1.347[a]	0.087	0.428[a]	0.130	1.216[c]	0.129	1.100	0.145
School 5	1.295[a]	0.092	0.820	0.245	1.165	0.135	1.204	0.173
School 6	0.856[c]	0.071	0.889	0.244	0.894	0.093	1.273[c]	0.165
School 6 x t	1.004[a]	0.001	-	-	-	-	-	-
School 7	0.818	0.318	0.704	0.799	10.382[a]	7.773	0.061[a]	0.052
School 8	1.038	0.085	1.076	0.373	1.062	0.147	0.995	0.168
School 9	0.946	0.078	1.432	0.494	0.769[b]	0.098	1.331[c]	0.220

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The results for the short-run base model are similar to those from the previous base model in section 7.3. In the short-run model, school 4 has a significant interaction effect associated with students being identified by the EAS. Compared to the base case school, students in school 4 not identified by the EAS have a hazard ratio 34.7 per cent higher. If however the students in school 4 are also identified by the EAS, the hazard ratio drops to 0.577. This means that, compared to students not identified by the EAS in the base school, students in school 4 identified by the EAS in any given week have a hazard 42.3 per cent lower. This is an unexpected result, given that in the short run, no other schools have significant interaction effects. The results suggest there may be model misspecification whereby the EAS is identifying students in school 4 who actually have a lower risk profile. This indicates scope for future analysis within school 4 to identify the causes of this response to identification by the EAS.

In the long run model, schools 1, 2, 6, 7 and 9 all have significant interaction effects, though only schools 2 and 7 are significant at the 1% level. The results show that overall, schools 1, 6 and 9 have an increase in the hazard ratio overall if identified by the EAS. School 7 has a reduction overall which coincides with previous findings. School 2, however, presented a more dynamic problem, with both the hazard ratio and the interaction effect varying over time. The hazard ratios for school 2 are plotted in Figure 7.30.

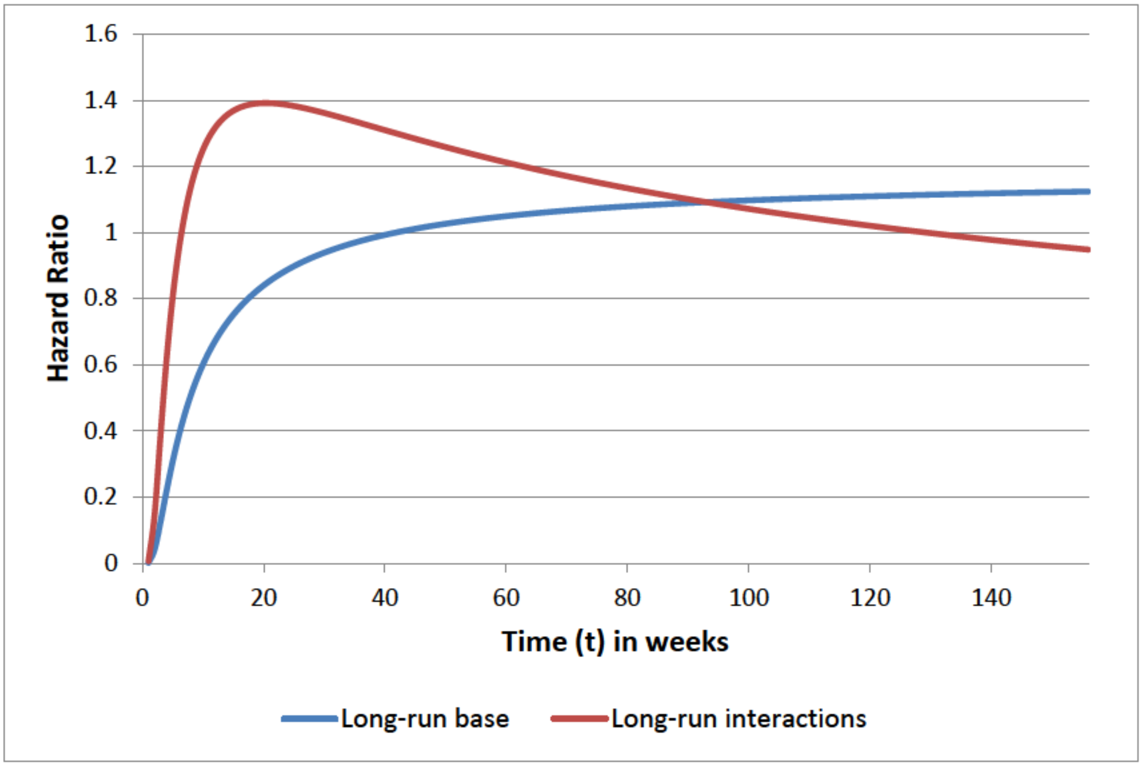


Figure 7.30 – Interaction model: school 2 hazard ratios over time

The results show two different hazard ratios depending on long run identification compared to the base school. Initially, students both have lower hazard ratios compared to students not identified by the EAS. The hazard ratio quickly increases for both, with students identified by the EAS having no significant difference from the base school by week 7. If not identified by the EAS, students have a lower hazard ratio until week 40. A key point is the interaction effect for school 2, where identification by the EAS shows a reversal of the increasing trend. The hazard ratio for identified students becomes equal to students not identified by the EAS around week 90, after which, there is an ongoing effect on the hazard ratios which continues to decrease the longer students are enrolled if identified. This suggests that students in school 2 respond differently to identification compared to both the students within the school and the base case school overall.

7.5.5 Student performance and workload variables

As shown in previous models, grades are significant indicators of the hazard of discontinuing. In the interactions effect model, several complex interactions are captured. The results are presented in Table 7.17.

Table 7.17 – Interaction model: student performance results

	Short-run base		Short-run interactions		Long-run base		Long-run interactions	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Withdrawn	1.378[a]	0.093	0.978	0.061	1.037	0.048	0.955	0.046
Withdrawn x t	0.992[a]	0.002	-	-	-	-	-	-
Withdrawn x t^2	1.000[a]	0.000[d]	-	-	-	-	-	-
Withdrawn Early	1.455[a]	0.082	1.001	0.062	1.367[a]	0.081	0.976	0.030
Withdrawn Early x t	0.992[a]	0.002	-	-	0.994[a]	0.002	-	-
Withdrawn Early x t^2	1.000[a]	0.000[d]	-	-	1.000[a]	0.000	-	-
Fail Incomplete	1.280[a]	0.025	0.762[b]	0.086	1.383[a]	0.069	0.898[b]	0.045
Fail Incomplete x t	0.999[b]	0.000[d]	1.002[c]	0.001	1.000	0.000[d]	-	-
Fail	1.141[a]	0.021	0.736[b]	0.108	1.307[a]	0.103	0.884	0.071
Fail x t	-	-	1.005[a]	0.002	-	-	-	-
Pass	0.847[a]	0.012	1.205[c]	0.136	0.759[a]	0.04	1.125[b]	0.061
Pass x t	-	-	0.997[b]	0.002	-	-	-	-
Credit	0.883[a]	0.013	1.177[a]	0.066	0.614[a]	0.059	1.433[a]	0.143
Credit x t					1.003[a]	0.001	0.997[b]	0.001
Distinction	0.874[a]	0.014	0.884	0.070	0.749[a]	0.039	1.094[b]	0.047
Distinction x t					1.001[c]	0.000[d]	-	-
High Distinction	0.869[a]	0.018	1.169[c]	0.103	0.571[a]	0.065	1.520[a]	0.188
High Distinction x t					1.003[a]	0.001	0.997[b]	0.001
Other	0.769[a]	0.055	0.589	0.200	0.279[b]	0.144	2.832[b]	1.471

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Both withdrawn and withdrawn early grades have no interaction effect associated with the EAS. In the short run model, there is an increase in hazard associated with withdrawing or withdrawing early from a unit. In the long run however, only withdrawing early increases a student's hazard. With respects to fail incomplete grades, the results are more complex to interpret. In the short run, the effect of fail incomplete and its associated interaction with the EAS varies over time. In the long run model, there is an overall interaction effect associated with failing incomplete. The hazard ratios for both the short run and long run model are presented in Figure 7.31, where NI is used to represent the fail incomplete grade.

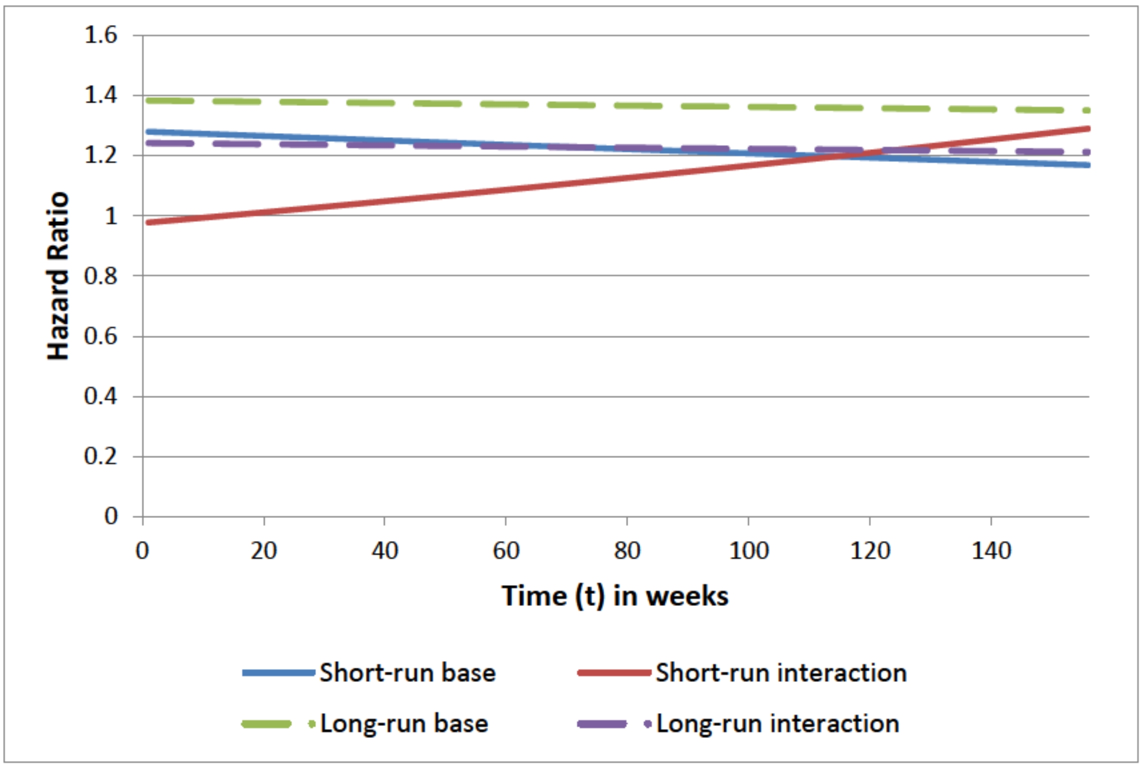


Figure 7.31 – Interaction model: fail incomplete hazard ratios over time

The short term model captured by the dashed lines shows that at the base level, a fail incomplete increases the students' hazard ratio. This decreases slightly over time. However it remains a significant effect over the time captured. The interaction effect is interesting, where students who receive the NI grade in the first teaching period and are identified by the EAS, shows no significant increase in the hazard ratio. However, this effect increases over time leads to a complex interpretation. For example, assume a student receives a fail incomplete grade at the start of their enrolment, and is identified in any given week by the EAS. The true effect on the hazard ratio will depend upon how far the student is through their enrolment. This is an important finding which further suggests EAS design needs to account for temporal effects. The long run model shows that there is also an increased risk associated with attaining a fail incomplete grade. The interaction term shows, however, that students who are identified by the EAS, the increase in hazards are less. This is another piece of positive evidence of the overall efficacy of the EAS in affecting students. Failing a unit in the long run increases a student's hazard by 30.7 per cent. The effect is constant over time and independent of interaction effects with the EAS. In the short-run, the interaction effect is significant at the 1% level and varies over time. Figure 7.32 plots the hazard ratios over time associated with a fail grade in the short run.

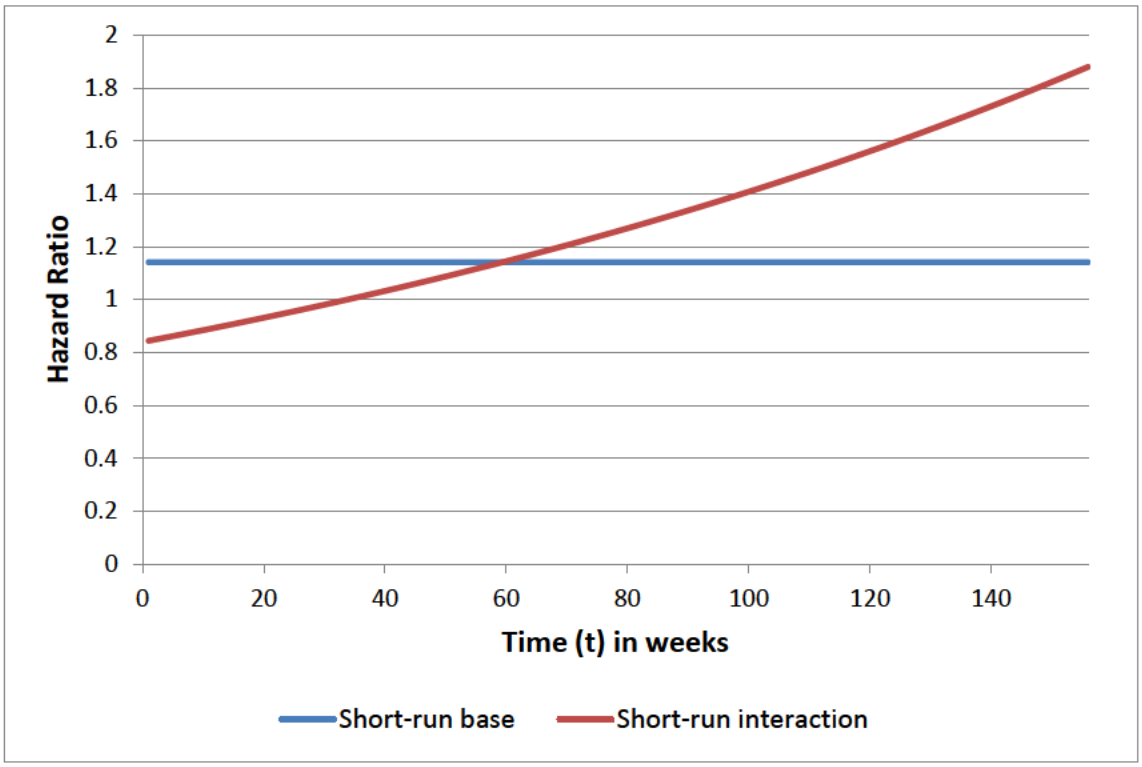


Figure 7.32 – Interaction model: fail grade hazard ratios over time in the short run

The results of the graph show that in the short run, students not identified by the EAS have a constant increase in hazard of 14 per cent. Depending on when students are identified by the EAS depends on the actual effect on the hazard ratio. The results show that where students who fail a unit and are identified within the first teaching period, there is a decrease in the hazard ratio. This important finding captures the link between the EAS and decreased hazards of students discontinuing during that early stage of enrolment upon failing a unit. If however, the student fails a unit and is identified later in their enrolment, the student's expected hazard ratio will be significantly higher.

The positive grades also capture interaction effects in the short run and long run models. For students who pass a unit, there is a decrease in the hazard ratio. However, if the students are also identified by the EAS then it gets more complex, as shown in Figure 7.33.

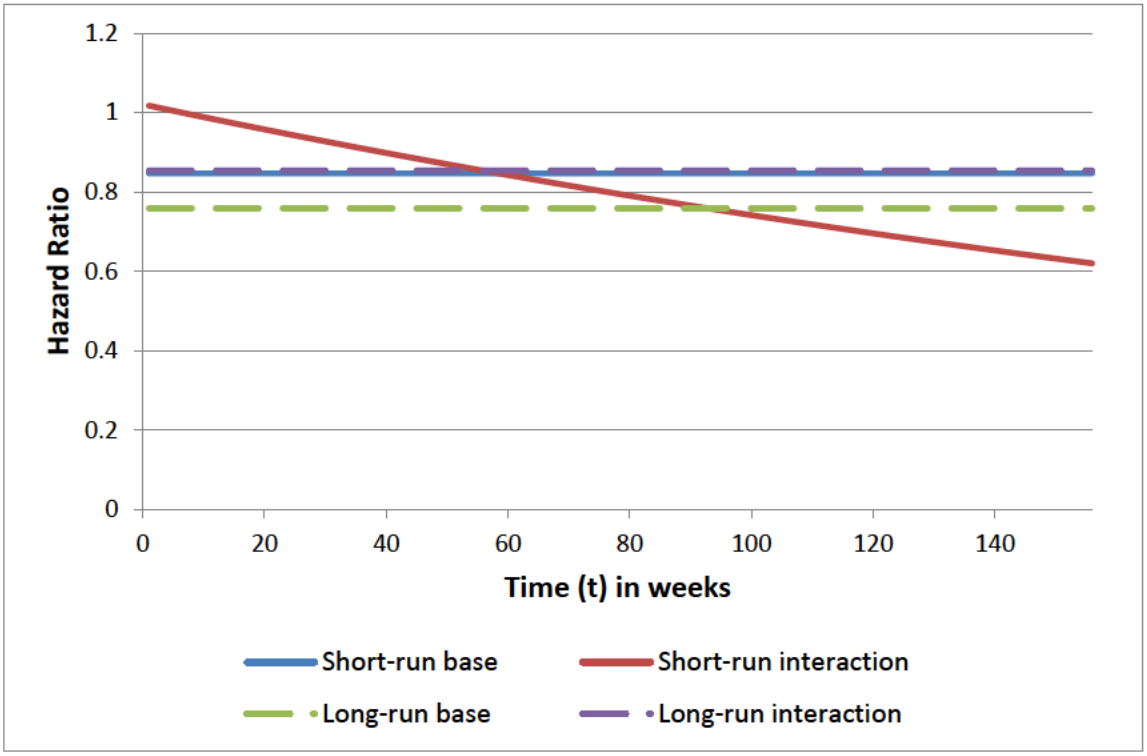


Figure 7.33 – Interaction model: pass grade hazard ratios over times

The diagram shows that in the short run model, students who pass a unit in the short run EAS model have a reduced hazard rate. When students are identified by the EAS, however, the hazard rate is higher early in the first year of enrolment. After the first year, students who pass a unit and are identified by the EAS actually have a decreasing hazard ratio. In the long run, students not identified by the EAS have a lower hazard ratio than those identified by the EAS. This indicates that the system overall is identifying students with a higher risk profile at least where pass grades are concerned.

The positive grades of credit, distinction and high distinction all have time varying and interaction effects in the EAS long run model. Attaining one of these grades has the effect of decreasing the hazard of a student. Taking into account the time varying and interaction effects, attaining a grade later in the student's study will still decrease the hazard ratio, but the effect is less compared to attaining the grade earlier.

The other learning environment variable which has a significant effect on the hazard ratio is the workload of the student. The results are presented in Table 7.18.

Table 7.18 – Interaction model: workload variables

	Short-run base		Short-run interactions		Long-run base		Long-run interactions	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Inactive	34.843[a]	4.206	4.926[a]	1.188	0.746	0.140	7.915[a]	1.080
Inactive x t (for t < 17)	0.138[b]	0.110	-	-	-	-	0.414	0.229
Inactive x 1/sqrt(t) (for t > 16)	0.049[a]	0.012	-	-	-	-	-	-
Inactive x t (for all t)	-	-	0.981[a]	0.004	-	-	-	-
Inactive x ln(t-15) (for t > 16)	-	-	-	-	-	-	1.426[a]	0.080
Part-time	1.900[a]	0.318	1.850[a]	0.385	1.323	0.271	0.511[b]	0.145
Part-time x ln(t)	0.987	0.056	-	-	0.691[a]	0.051	1.926[a]	0.161

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level

The results show significant interactions for both inactive and part-time variables. The base model and the conditional model both showed that inactivity is the most significant variable in terms of magnitude affecting the hazard ratio. To understand the complex interplay between inactivity, time and the EAS, the hazard ratios are plotted in Figure 7.34.

The results indicate similar effects as previously identified in models 7.1 and 7.2. The short run base models represented with the dashed lines show that students who become inactive within the first few weeks of enrolment have a significantly higher hazard ratio. In the case of short run inactivity without interaction effects, there is a sharp increase in the hazard ratio tapering off with an estimated hazard coefficient of around 28. Taking into account the EAS short run interactions, the increase in hazard is significantly higher. This hazard ratio peaks at a coefficient of 60 before tapering off over time. The reduction over time may represent some effect associated with the benefits of the EAS, but in context, inactivity still significantly increases the hazard ratio of the student.

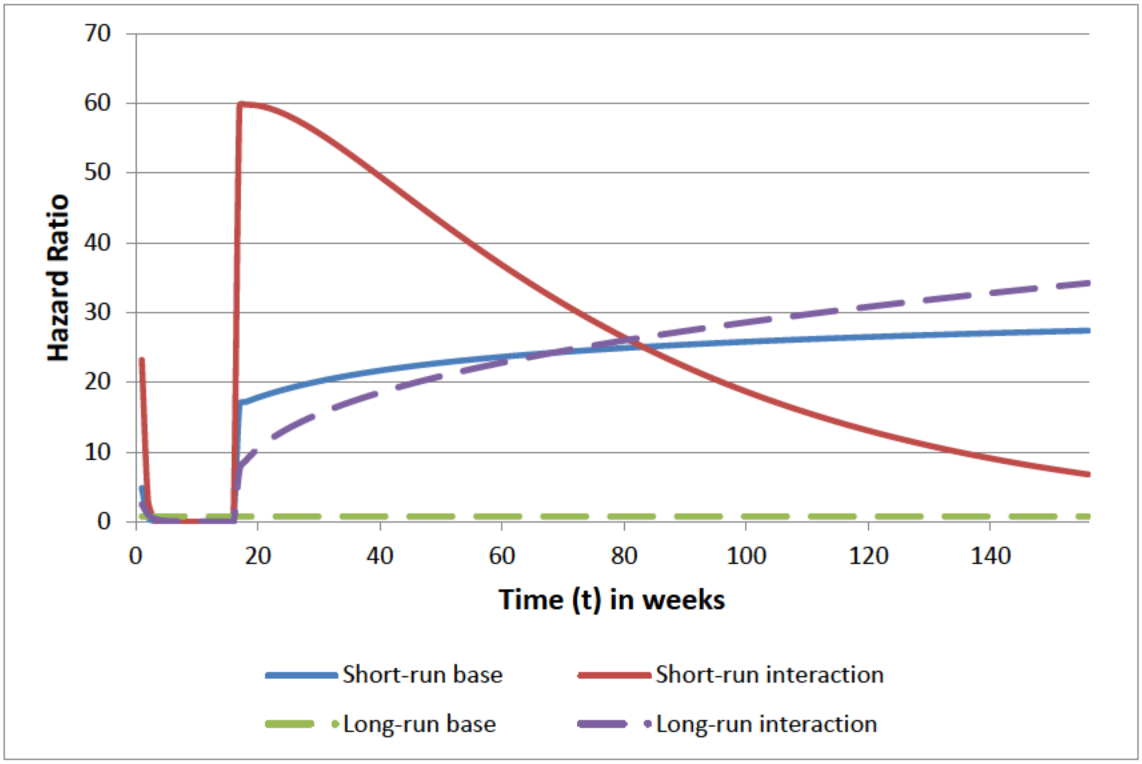


Figure 7.34 – Interaction model: inactivity hazard ratios over time

The long run EAS model however tells a different story. The green line in Figure 7.34 captures the base effect of inactivity overall. For students who were inactive, there is no significant difference in the hazard ratios to students studying full-time. This effect is independent of time completely and does not change. Introducing the interaction effect where the student is both inactive and identified by the EAS in the long run, the hazard ratio quickly increases over time. This rapid increase may be evidence of correctly identifying students who have a higher risk profile by the EAS. As such it is safe to conclude that any EAS design should incorporate temporal effects associated with inactivity.

7.5.6 Summary

In summary, the results for the interactions model highlight the complex interplay between variables, time and the EAS. The results indicate areas where the EAS system is functioning correctly, and other areas where the system is not correctly specified. For example, the system showed that ATSI students who were identified by the system overall had a change in hazard ratio trajectories, improving the hazard ratio significantly. The results also showed that students enrolled in different schools responded in different ways to being identified by the EAS. In the case of school 2, while there was a rapid initial increase in the hazard ratio over time, those students who were identified by the EAS had a reduced hazard at the start of the third year of study.

The results show learning environment variables to be the most complex to capture in the interactions survival model. To capture all the necessary time varying elements, the model incorporated interactions with time in various functional forms. This highlights a challenge for modelling complex systems. Incorporating more variables from the learning environment into the model increases the complexity of the model.

7.6 Early Alert System trigger model results

7.6.1 Early Alert System trigger model description

The EAS was designed to use 34 triggers to identify students in need of student support. The data corresponding to EAS triggers 15 to 18 revealed that these four triggers were never implemented, meaning the triggers never activated over the data capture period. Additionally, trigger seven had limited number of activations over the capture period, resulting in a sample size too small to be incorporated into the analysis. As such, while 34 triggers have been defined, on 29 are tested in the analysis. The EAS triggers fall into two temporal categories: teaching period variables and daily variables. The teaching period variables are defined as variables which remain constant for at least one teaching period. These variables can change between teaching periods. The daily variables change regularly throughout the teaching period. The full list of the 34 triggers can be found in appendix B.

A relationship between identification and the hazard ratio was established in the base model. Following from this, in theory, there should be a relationship between the underlying triggers involved in the identification process and the hazard of students discontinuing. Each trigger in the EAS has an accompanying positive or negative weight. The negative weights represent a level of hazard or risk associated with activating that trigger. Negatively weighted triggers should correspond to the estimated increases in hazard ratios if functioning within expectations.

The EAS-trigger model estimates the hazard ratio of students using the EAS triggers as explanatory variables. Since triggers can be activated on a daily basis, any given student can have multiple instances of a trigger within one week. This indicates there are several ways of describing the triggers in the model. One option is to consider if the number of times a student activates a trigger is important. Another option is to consider if triggers activated by all students should be considered, or if only those students identified by the EAS should form the set of triggers analysed. These options create four model configurations for analysing the triggers. The first is to only analyse if a trigger was activated in a given week by any student. This is referred to as the All-Students-Base (ASB) model. The second model is then to test if the number of times in a given week is significant for all students. This is the All-Student-Multiple (ASM) model. The third model restricts the triggers to only those students identified by the EAS, and considers only if a trigger was activated in a week. This is the Identified-Student-Base (ISB) model. The final model considers only the triggers activated by identified students, but tests if the number of times the trigger is activated is signification. The model is referred to as the Identified-Student-Multiple model. It is expected that testing the EAS triggers in these alternate configurations will reveal which triggers have significant relationships to the hazard function, and which triggers are sensitive to model specification.

ASB and ASM models focus on the relationship between the triggers and the hazard ratio of all students, irrespective of whether the student was identified or not. This is depicted in Figure 7.35 where the red pathway shows the direct correlation with the hazard ratio.

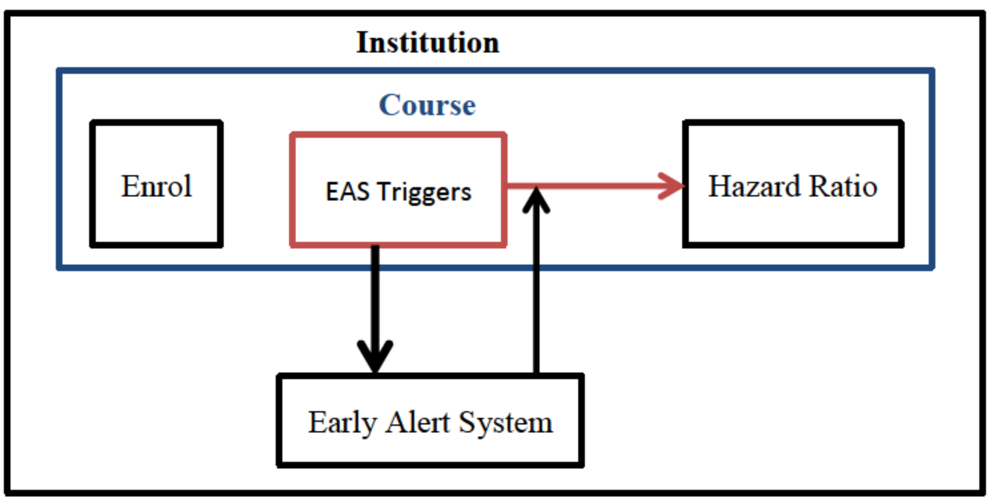


Figure 7.35 – EAS-trigger model: ASB and ASM configurations (Authors own contribution)

The ISM and ISB models use the information from the base long-run model to limit the triggers to only those students who were identified. This creates conditional triggers that effectively removes noise from the data so only triggers which correlate with identification by the EAS are used. The ISB and ISM models form a conditional logic approach to the interpretation of the results, placing more emphasis on the triggers which most strongly contribute to the identification process. This is represented in Figure 7.36 where the red pathway represents the identified students relationship to the hazard ratio.

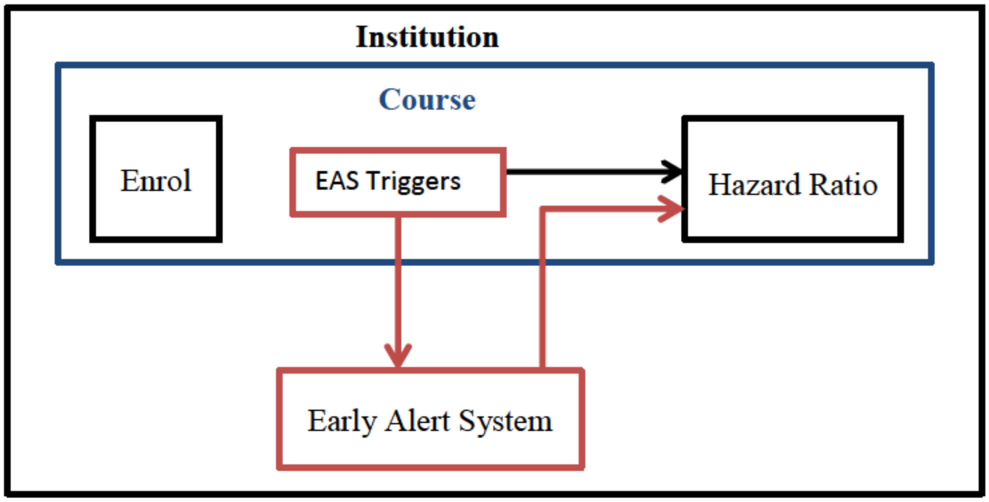


Figure 7.36 – EAS-trigger model: ISB and ISM configurations (Authors own contribution)

The relationship between the triggers can be broken up into five distinct classes. The first are significant triggers constant over time. Variables in this class have a constant effect on the hazard ratio independent over time for all models. The second class of triggers is the significant triggers which decrease over time. Variables in this class indicate that the hazard associated with the trigger decreases as a student progresses for all models. The third class of triggers is significant, increasing over time. The hazard ratio early in a student's enrolment is less than the hazard later in the enrolment for all models. The fourth class of triggers is the inconsistent over time category. Variables in this class capture increasing, decreasing and constant effects between models which indicate sensitivity to model specification. The final class of triggers are the insignificant triggers, which show no relationship between the trigger and the estimated hazard ratio of the student for all models.

7.6.2 Significance of the EAS-trigger model and assumptions test

The overall significance of the models presented in Table 7.19 show models significant at the 1% level. In the broadest sense, there is a significant relationship between the EAS and the hazard ratio of students. This also means that given the EAS was designed to identify students at risk of disengaging, it also means that the system is identifying students at risk of discontinuing.

Assuming the system is actually identifying students disengaging from their studies, this provides an empirical link between disengagement and discontinuation.

Table 7.19 – EAS-trigger model: overall significance and assumption tests

ASB		ASM		ISB		ISM	
LR χ^2 (44)	1830.6	LR χ^2 (44)	1943.69	LR χ^2 (40)	1627.8	LR χ^2 (40)	1712.35
Prob > χ^2	0	Prob > χ^2	0	Prob > χ^2	0	Prob > χ^2	0

PH Test		PH Test		PH Test		PH Test	
χ^2 (44)	39.16	χ^2 (44)	40.1	χ^2 (40)	27.85	χ^2 (40)	21.62
Prob > χ^2	0.6789	Prob > χ^2	0.6396	Prob > χ^2	0.9265	Prob > χ^2	0.9922

The proportional hazard assumptions test shows that the models have no significant relationship with time variables. This means that the models do not violate the proportional hazards assumption of the Cox model. A full detailed summary of the test is provided in Appendix F.

7.6.3 Triggers with constant effects over time

With respect to EAS design, ideally all triggers included in the system should have constant effects over time. This makes system design simple and parsimonious and represents the current assumptions of the EAS design. In testing the triggers of the case study EAS, four of 34 triggers showed constant significant effects independent of time. The estimated coefficients for the four triggers are presented in Table 7.20.

Table 7.20 – EAS-trigger model: triggers with constant effects over time

	ASB		ASM		ISB		ISM	
Trigger	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
21	0.085[a]	0.061	0.133[a]	0.090	0.091[a]	0.065	0.130[a]	0.087
24	0.408[a]	0.126	0.444[a]	0.123	0.358[a]	0.129	0.419[a]	0.137
27	0.252[a]	0.127	0.276[a]	0.129	0.267[a]	0.134	0.284[a]	0.132
33	0.488[a]	0.120	0.482[a]	0.104	0.517[a]	0.128	0.547[a]	0.118

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The four triggers shown are all significant at the 1% level across all four model specifications. The estimated effects vary slightly between models but this is only of mild concern.

Trigger 21 corresponds to the student being an international student. According to EAS specification, international students have a negatively weighted trigger reflecting increased risk of discontinuation. The EAS-trigger model indicates, however, these students have a reduced hazard ratio throughout their studies, between 87 per cent and 91 per cent lower than non-international students. International student status was effectively captured in the base model in section 7.3, in the fee status variable. The results from these models showed an increasing then decreasing hazard over time. This contradictory finding indicates that while international student status is a significant variable that is constant in the EAS, there may be a misspecification of the EAS compared to the base models analysed.

Trigger 24 corresponds to a student who was previously enrolled in a pathways course. The results in Table 7.20 show that students who activate this trigger have a significantly lower hazard than students not activating the trigger. This result is incongruous to the EAS specification, which weights the trigger negatively, indicating a higher risk of disengaging. This means two possible effects are occurring. Students who activate this trigger have a higher risk of disengaging but a lower risk of discontinuing. The other explanation is that the EAS is not capturing what it intends to capture, and that some students are being identified through this trigger who actually have a lower hazard ratio. If the captured effect is the former, then this shows a point of difference between disengaging and discontinuation. However, the latter is more likely, which indicates that the EAS may need to be refined to ensure the trigger is functioning correctly.

Trigger 27 corresponds to students carrying over a special extension of time (SET) exam into the current teaching period. This is negatively weighted according to EAS design, meaning it is expected that students activating this trigger are at a higher risk. The results, however, show that students activating trigger 27 have a lower hazard ratio. This result is in line with the results from the base model in section 7.3 where SET exams are captured as part of the “other” grade outcome. This grade outcome consistently corresponded to a lower hazard ratio. Like trigger 24, this may indicate that there is either a disparity between disengagement and discontinuation with respect to SET exams, or this shows EAS misspecification with respect to the trigger not capturing the proposed relationship indicated by the trigger weighting.

The last constant trigger over time was trigger 33. This trigger corresponds to a student receiving a fail grade in the prior teaching period. Logically, this trigger should have a positive relationship to the hazard ratio. The results presented in Table 7.20 indicate students activating this trigger have a lower hazard ratio than students who do not activate this trigger. The result is in contrast to the grade results of previous models. One possible reason for a lower hazard ratio is due to the time delay of the trigger. The trigger captures what occurred in the previous teaching period. It is possible that as a result of the student receiving the grade in the previous teaching period, they have since sought out support. As such, the trigger is actually capturing the combined effect of both the initial fail grade and any subsequent support the student has received. This requires further investigation but highlights a potential issue with the triggers design.

7.6.4 Triggers with decreasing effects over time

Seven triggers were found to have declining hazard ratios over time. These results were consistent across all four models. The coefficients and corresponding interactions with time are presented in Table 7.21 over page.

Table 7.21 – EAS-trigger model: EAS triggers with decreasing hazard ratios over time

	ASB		ASM		ISB		ISM	
Trigger	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
3	0.091[a]	0.013	0.594[a]	0.024	0.058[a]	0.014	0.624[a]	0.029
3 x 1/ln(t)	58.462[a]	21.372	2.590[a]	0.281	206.136[a]	141.4	2.184[a]	0.290
4	0.800	0.157	0.932	0.147	0.646[c]	0.152	0.753	0.149
4 x t	0.992[b]	0.004	0.990[a]	0.004	0.996	0.004	0.993[c]	0.004
11	1.361[b]	0.202	1.443[b]	0.207	1.199	0.216	1.248	0.217
11 x t	0.982[a]	0.004	0.981[a]	0.004	0.981[a]	0.004	0.981[a]	0.004
13	1.132	0.152	1.365[b]	0.168	0.925	0.152	1.133	0.165
13 x t	0.990[a]	0.003	0.988[a]	0.003	0.992[b]	0.003	0.991[a]	0.003
19	4.285[a]	2.013	4.047[a]	1.801	4.833[a]	2.568	4.253[a]	2.147
19 x t	0.950[b]	0.022	0.952[b]	0.020	0.947[b]	0.024	0.951[b]	0.022
28	3.085[a]	0.935	3.325[a]	0.962	24.831[a]	16.608	27.136[a]	16.644
28 x t	0.966[a]	0.009	0.965[a]	0.009	-	-	-	-
28 x ln(t)	-	-	-	-	0.337[a]	0.079	0.335[a]	0.073
32	0.820	0.421	1.193	0.515	0.250[a]	0.092	1.039	0.548
32 x t	0.978[c]	0.012	0.973[b]	0.011	0.982[b]	0.009	0.972[c]	0.014

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Trigger 3 captures that the student is enrolled in a currently high attrition unit. To capture the effect of this trigger over time, the hazard ratio relative to the week of enrolment is presented in Figure 7.37.

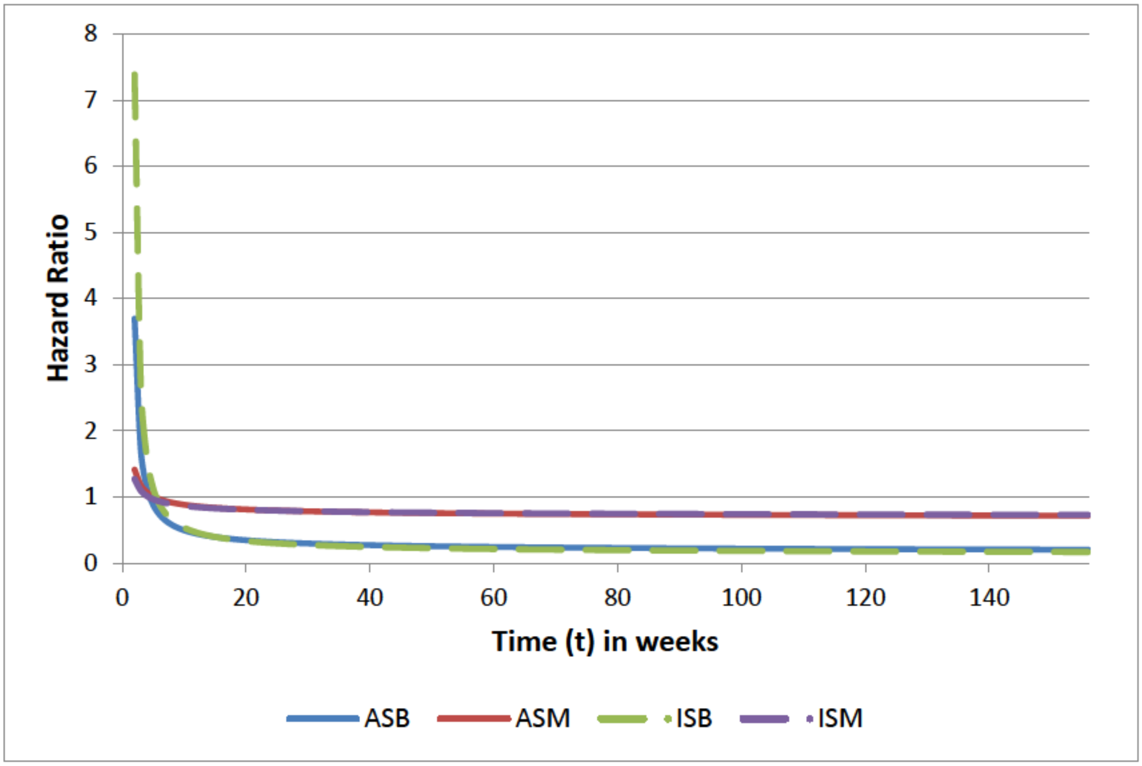


Figure 7.37 – EAS-trigger model: trigger 3 hazard ratio over time

The results for trigger 3 show that there is a significantly higher hazard ratio in the first four weeks of enrolment. The magnitude of the effect is estimated to be higher in the ASB and ISB models, however the hazard ratios for the four models converge at week four, upon which there is no significant difference resulting from this trigger. After week 4, the trigger has a decreased hazard ratio below 1. This indicates the trigger may be a valid trigger to include in the EAS, but only for the first four weeks of a student's enrolment. After which, the trigger may actually cause students to be wrongly identified.

Trigger 4 is similar to trigger 3. However, it is activated when a student enrols in a unit that historically had high levels of attrition. The hazard ratio for trigger 4 is plotted in Figure 7.38 over page.

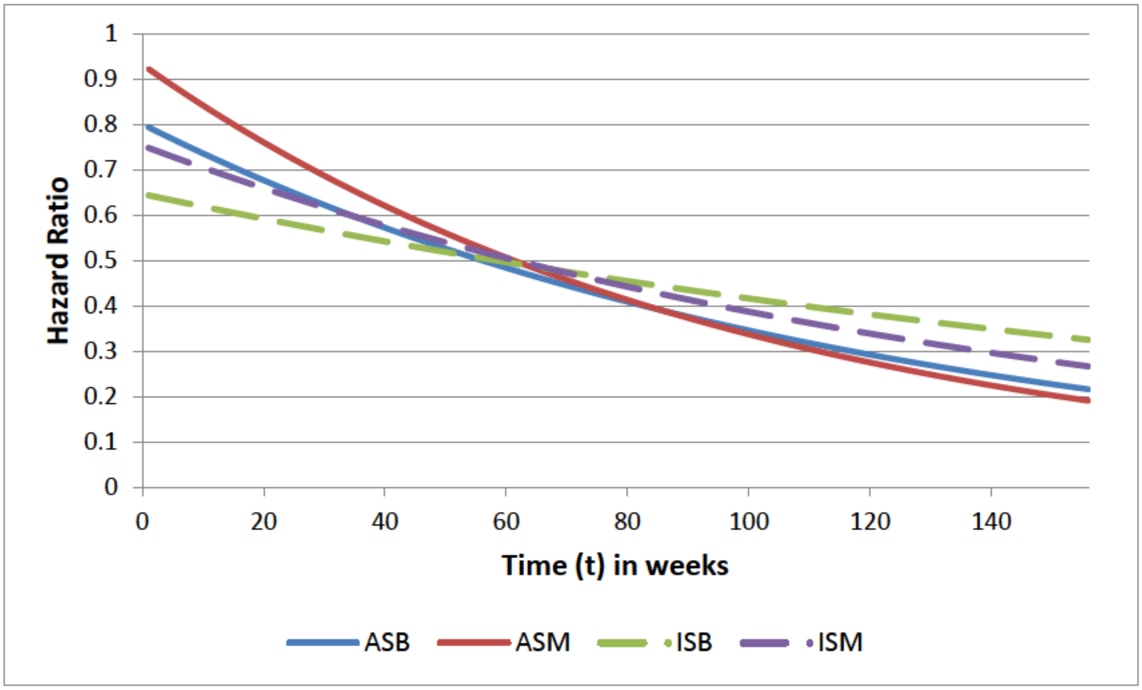


Figure 7.38 – EAS-trigger model: trigger 4 hazard ratio over time

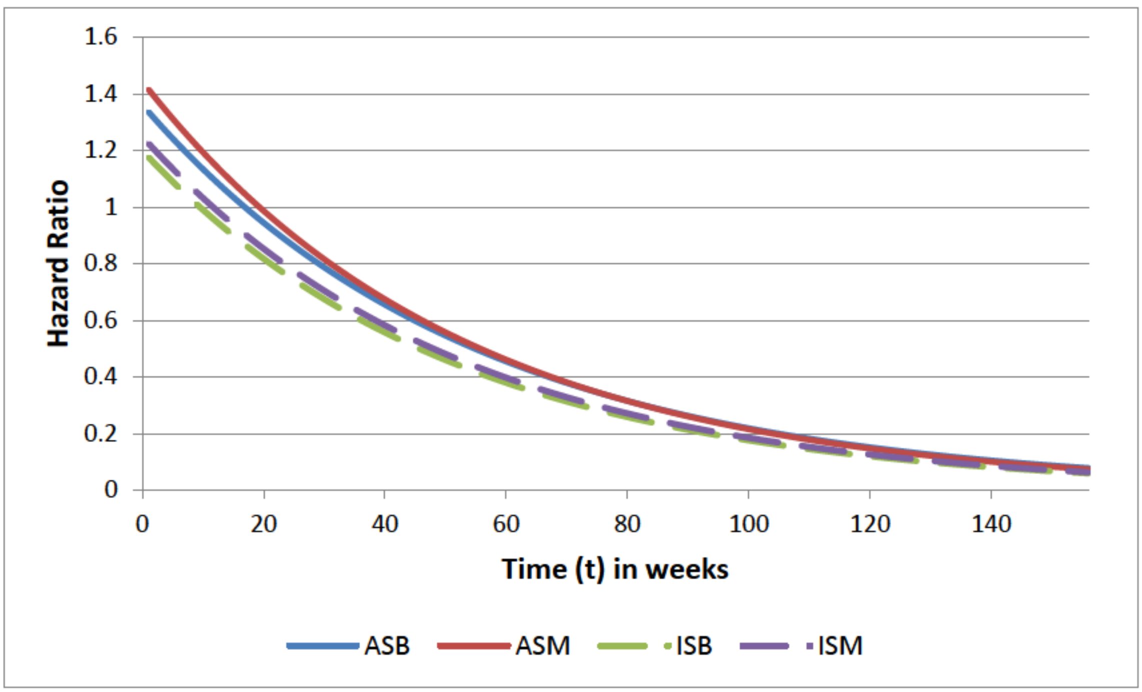


Figure 7.39 – EAS-trigger model: trigger 11 hazard ratio over time

The results indicate that the hazard ratio at initial enrolment is either close to or below 1. As students' progress with their enrolment, the hazard ratio associated with the trigger decreases. The trigger never captures students with a hazard ratio above 1, indicating current EAS specification of the trigger may not be correctly identifying students in need of assistance. Trigger 11 captures students who have had high periods of e-reserve inactivity. This is where the student fails to log into the library's e-reserve portal to access study materials. A student needs to have not logged in for 31 to 40 days to set off the trigger. The estimated hazard ratio is plotted in Figure 7.39. Logically, a student cannot activate this trigger within the first four weeks of study. The ASB and ASM models have a hazard ratio of 1 at week 11, while the ISB and ISM has a hazard ratio of 1 at week 17.

This means if a student is identified by the EAS between weeks four and 11, the student has a higher hazard ratio. Overall, any student who activates this trigger has a higher hazard ratio between weeks 4 and 17. After these respective points in time however, the student will have a lower hazard ratio. This indicates that the trigger may be a valid identifier of students in need of additional support but only during the first teaching period.

Trigger 13 is a related trigger, capturing students with medium levels of e-reserve inactivity. This trigger is activated if the student does not log into the e-reserve portal for between 21 and 30 days. The results are presented in Figure 7.40 over page. The results show that for ASB, ASM and ISM, there is an initial higher hazard ratio than students who do not activate this trigger. This decreases over time, with ASB and ISM having a hazard ratio of 1 around week 11. For the ASM model, the trigger equals 1 in week 26.

The model shows some evidence indicating that the trigger is useful during the initial stages of students' enrolment. However, past week 11, this trigger continues to decrease, capturing students with a hazard ratio lower than students who do not activate the trigger. As such, this trigger is useful in the EAS design during initial enrolment, however is less effective later in the students' enrolment.

Trigger 19 captures a complex conditional statement focused around slower student progression. For trigger 19 to be set off, student enrolment has involved greater than double their number of currently enrolled units in the current teaching period. For example, if a student has already completed six units of study but is only currently enrolled in two units at the start of the teaching period, the trigger will activate. The estimated hazard ratio associated with trigger 19 is presented in Figure 7.41.

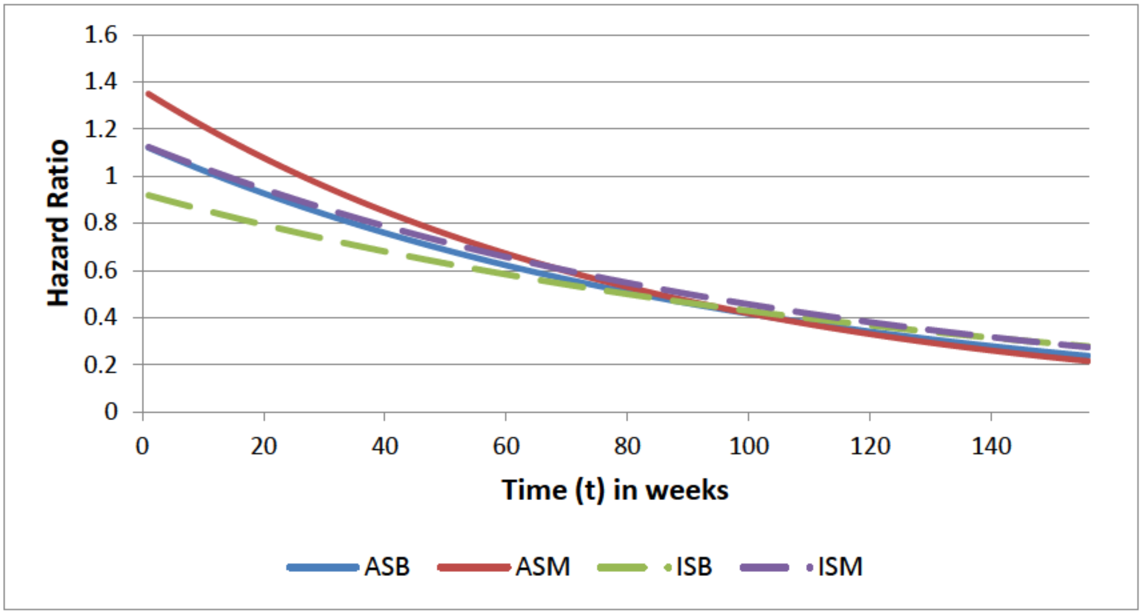


Figure 7.40 – EAS-trigger model: trigger 13 hazard ratio over time

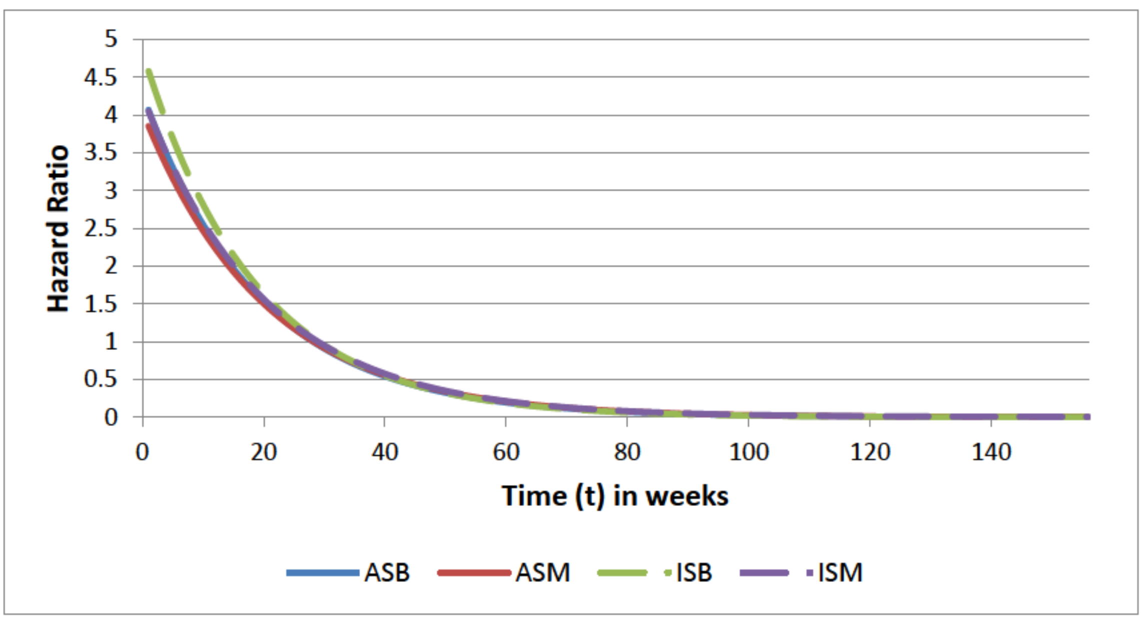


Figure 7.41 – EAS-trigger model: trigger 19 hazard ratio over time

The results show that initially, this trigger captures students with a significantly higher hazard ratio than students who do not activate the trigger. This declines consistently across the four sub-models, converging on parity of 1 in week 28. At this point, there is no significant difference in hazard of students activating trigger 19 and those students not activating 19. As such, trigger 19 can function as an excellent trigger for the early initial phase of a student's enrolment. With respect to EAS design, it has limited use and should not apply after week 28 of enrolment.

Trigger 28 captures students who have high portal inactivity where the student has failed to log into the MyUNE portal for the last 31 to 40 days. The hazard ratio is plotted over time in Figure 7.42.

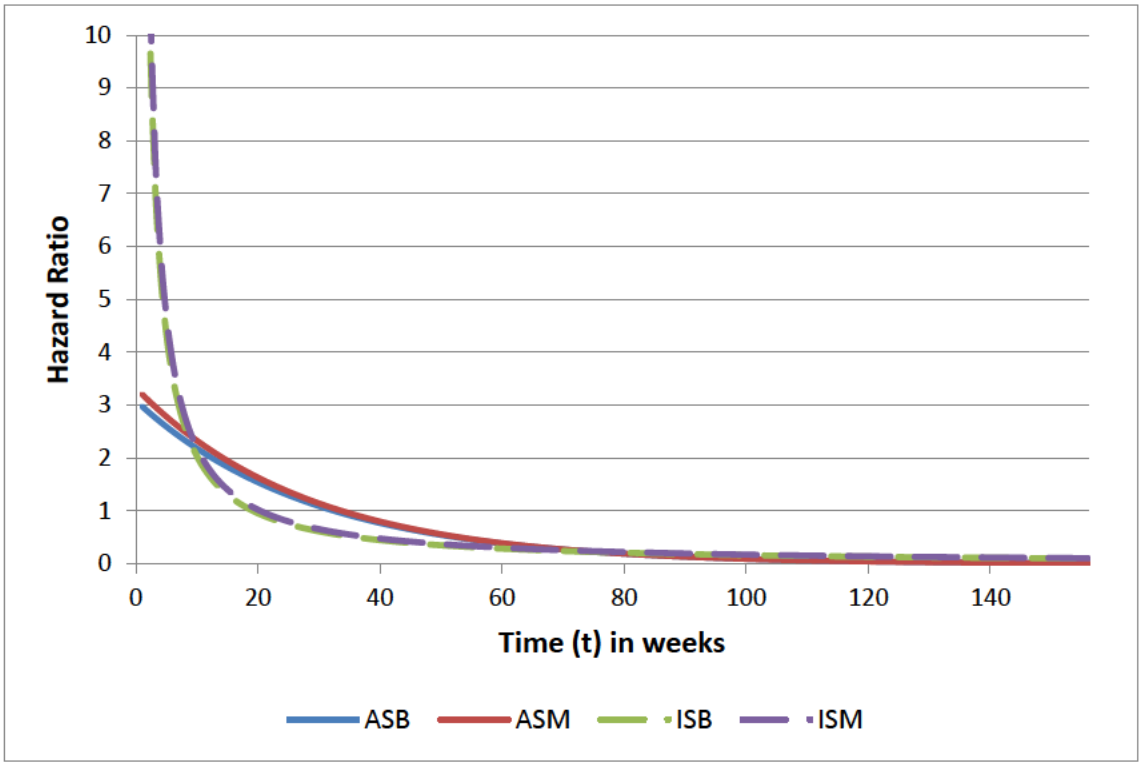


Figure 7.42 – EAS-trigger model: trigger 28 hazard ratio over time

As can be observed, there is a distinct difference between the ASB and ASM models to the ISB and ISM models. The ASB and ASM models capture all students who activated this trigger regardless of identification. It shows that initially students have a hazard ratio around 3, decreasing to 1 in around week 33. This means that over that period of time, if any student activated this trigger, there was an increased hazard of discontinuing. In the ISB and ISM models, the initial hazard is much greater, indicating that if the student was actually identified by the EAS, the risk of discontinuing is greater. It decreases more rapidly than the ASB and ASM models, with it reaching 1 around week 20. This means that the trigger effectively becomes invalid past this point in time. It is possible to conclude that this is an excellent trigger to capture students who are at risk of discontinuing early in their study.

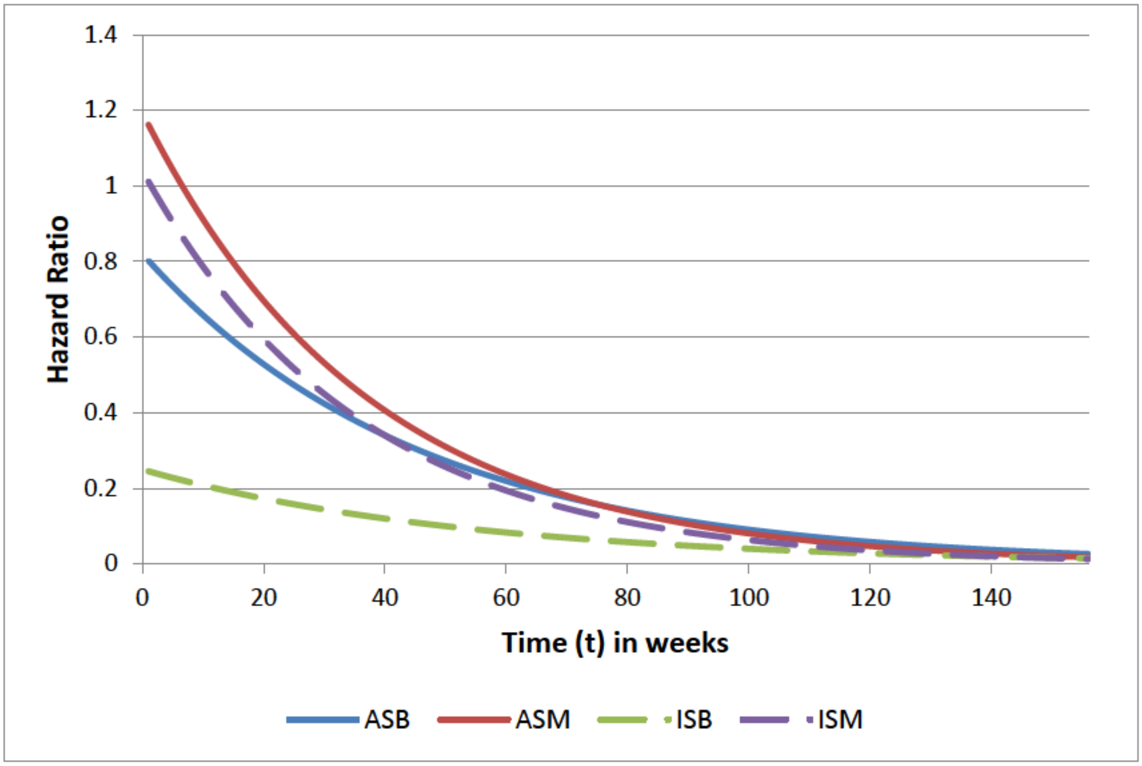


Figure 7.43 – EAS-trigger model: trigger 32 hazard ratio over time

Trigger 32 captures students enrolled in a teacher enabling course. This course is designed to provide students who enter teaching degrees with basic maths and English skills. Initially, the four models indicate different levels of starting hazard, with only the ASM model showing an above average hazard. This however decreases and eventually converges. The trigger in its current model specification does not capture students with a significantly higher hazard ratio except in the ASM model. Even then, the increased hazard is only observed for the first seven weeks before decreasing past 1. As such, this trigger may not be useful in identifying students at risk of discontinuing. It may however still be useful in capturing students at risk of disengaging.

7.6.5 Triggers with increasing effects over time

Triggers that increase in hazard over time may be useful in identifying students at risk of discontinuing their studies later in their enrolment. Only one trigger was identified as increasing over time, with the estimated coefficients presented in Table 7.22 and the chart of hazard ratios over time in Figure 7.44.

Table 7.22 – EAS-trigger model: EAS triggers with increasing over time

	ASB		ASM		ISB		ISM	
Trigger	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
20	0.007[a]	0.01	0.016[a]	0.022	0.004[a]	0.006	0.009[a]	0.014
20 x t	1.058[a]	0.016	1.044[a]	0.014	1.066[a]	0.017	1.053[a]	0.016

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

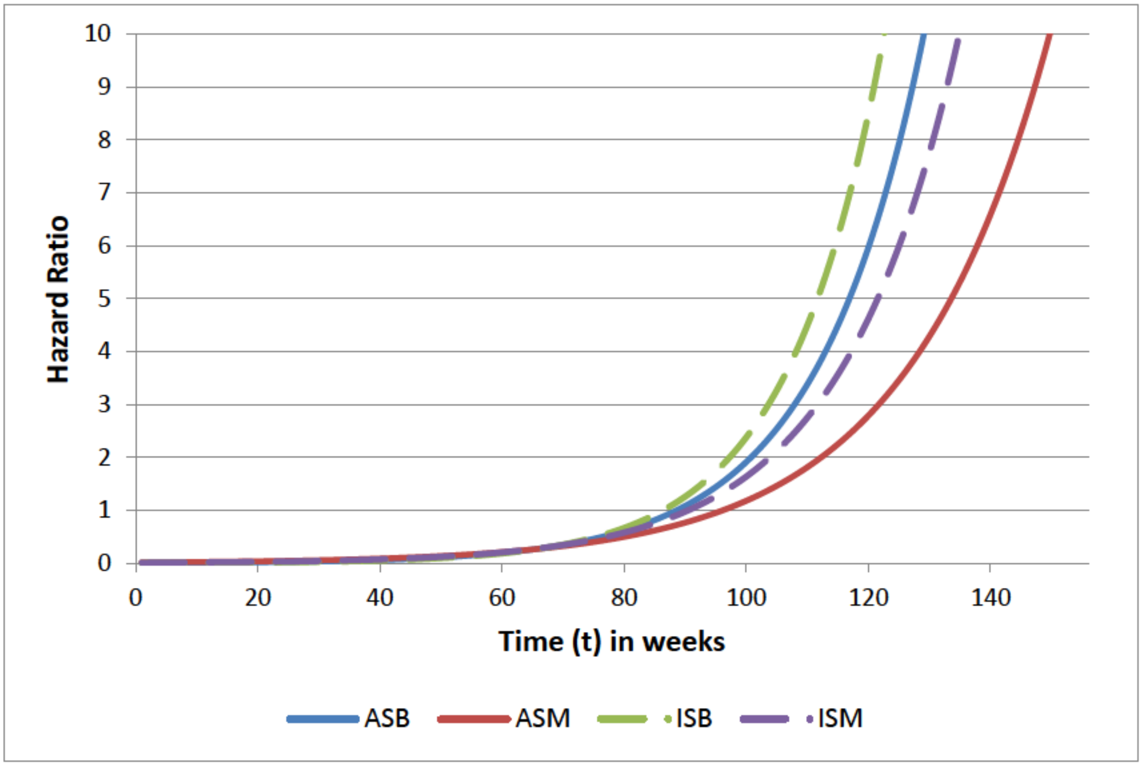


Figure 7.44 – EAS-trigger model: trigger 20 hazard ratio over time

Trigger 20 corresponds to the student appearing in a high risk category in a previous teaching period. The results are significant at the 1% level indicating that they are significantly different from 1. Figure 7.44 shows trigger 20 significantly increasing in hazard over time. This is a sensible result since the trigger is effectively delayed, where very few if any students would be activating this trigger early in their studies. By around week 85, the four models have reached parity, after which activating this trigger genuinely indicates an increased hazard ratio. Given that the trigger is innately designed to capture students at higher risk, this is a logical trigger to include in an EAS. However, the weighting associated with the trigger should change over time.

7.6.6 *Triggers with inconsistent effects over time*

Many of the EAS triggers showed inconsistent results amongst models. This may be where all but one trigger had time varying interactions, where the results of the EAS-trigger model do not align or where all models seem to show conflicting hazard estimates. The results for the inconsistent triggers are presented in Table 7.23.

Table 7.23 – EAS-trigger model: EAS triggers with inconsistent effects over time

	ASB		ASM		ISB		ISM	
Trigger	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
5	0.384[a]	0.036	0.822[a]	0.010	0.278[a]	0.050	0.821[a]	0.010
5 x sqrt(t)	1.000[a]	0.000	-	-	-	-	-	-
5 x t	-	-	-	-	1.010[c]	0.005	-	-
5 x t^3	-	-	-	-	1.000[a]	0.000	-	-
6	0.579[c]	0.178	0.547[b]	0.163	0.699	0.226	0.665	0.208
12	0.821	0.125	1.050	0.145	0.483[a]	0.056	0.783	0.135
12 x t	0.992[b]	0.003	0.989[a]	0.003	-	-	0.994[b]	0.003
14	0.868[c]	0.074	0.942	0.071	0.848[c]	0.080	0.975	0.081
22	0.063[a]	0.012	0.577	0.197	0.059[a]	0.012	0.733	0.408
22 x ln(t)	-	-	0.538[a]	0.066	-	-	0.519[a]	0.093
25	2.572[a]	0.304	2.530[a]	0.256	1.478[a]	0.131	1.363[a]	0.112
25 x t	0.991[a]	0.003	0.990[a]	0.002	-	-	-	-
26	10.21[a]	1.379	3.969[a]	0.295	11.100[a]	1.250	5.864[a]	0.499
26 x t	0.986[a]	0.003	-	-	-	-	-	-
26 x t^2	-	-	-	-	1.000[a]	0.000	1.000[a]	0.000
29	1.086	0.189	1.150	0.198	0.472[a]	0.066		
29 x t	0.986[a]	0.004	0.985[a]	0.004	-	-	0.487[a]	0.068
30	6.243[a]	2.525	5.348[a]	2.167	0.772	0.128	0.776	0.127
30 x ln(t)	0.550[a]	0.070	0.578[a]	0.074	-	-	-	-
31	0.904	0.195	5.045[b]	3.223	0.731	0.190	0.778	0.200
31 x ln(t)	-	-	0.601[b]	0.122	-	-	-	-
34	12.53[a]	11.165	1.985[b]	0.603	1.805	0.691	0.825	0.166
34 x ln(t)	0.477[a]	0.125	-	-	-	-	-	-
34 x t	-	-	0.980[a]	0.007	-	-	-	-

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Trigger 5 corresponds to students enrolled in a residential college on-campus. The varying hazard estimates are captured in Figure 7.45.

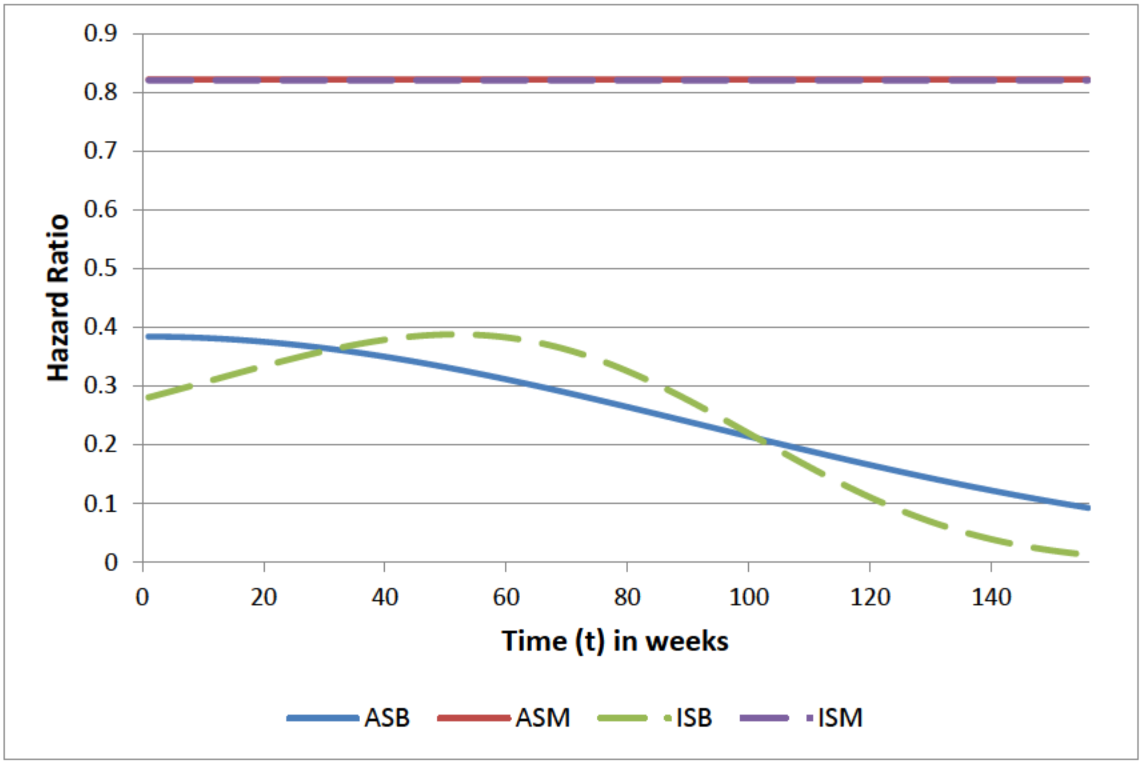


Figure 7.45 – EAS-trigger model: trigger 5 hazard ratio over time

The results show that the weighted models produced constant hazards independent of time. According to the weighted model, each time a student activates this trigger, the hazard ratio of the student is estimated to be around 18 per cent lower than a non-college student. In the unweighted model, the hazard fluctuates with increases and decreases over time. Overall, however, it can be interpreted that students have a lower hazard ratio. Given that the four models do not align, this may indicate sensitivity within the trigger and may have confounding variables affecting hazard ratios.

Trigger 6 captures students who, using the 'e-motion' tool, indicated that they were currently happy with a unit of study. This trigger was only significant at the 10% in the ASB and ASM models, showing a decreased hazard ratio. While the result is in line with expectations of happy students being less likely to discontinue, the overall effect is weak or non-existent as in ISB and ISM models. As such, this does not make it a strong trigger usable for an EAS.

Trigger 12 captures students with low e-reserve inactivity, not logging into the library portal for between 10 and 20 days. The results are plotted in Figure 7.46.

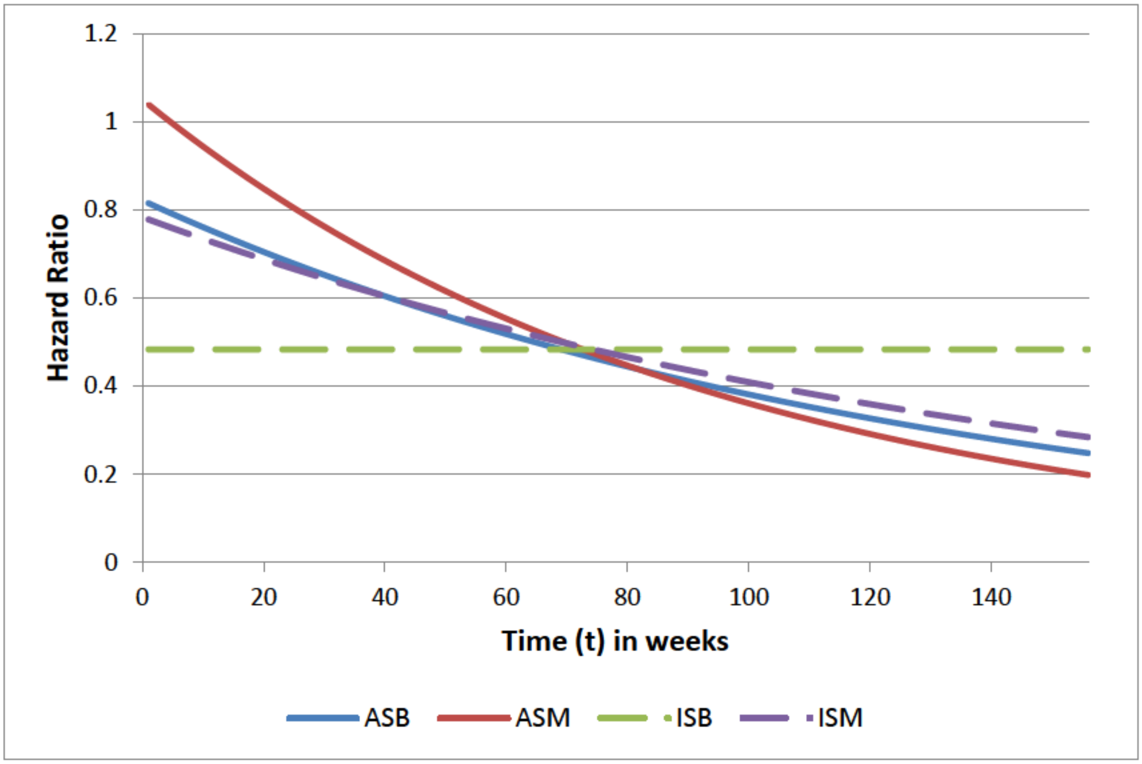


Figure 7.46 – EAS-trigger model: trigger 12 hazard ratio over time

Trigger 12 hazard ratios show both decreasing hazards over time and a constant hazard over time for the ISB model. The results indicate that over time a student is likely to have a reduced hazard ratio than students who do not activate the trigger. As such, this trigger is producing nonsensical results, and is likely to result in identifying students who do not have an increased hazard of discontinuing.

Trigger 14 captures students with very high e-reserve inactivity. The trigger, however, only had significant results at the 10% level for the ASB and ISB models respectively. Additionally, the triggers show a decreased hazard ratio, which is nonsensical and against expectations. As such, the trigger is unlikely to be making a positive contribution to the identification process of students with higher hazard of discontinuing.

Trigger 22 captures students who have had no prior enrolment history within UNE. The hazard ratio estimates are presented in Figure 7.47. The estimated hazard ratios are significantly less than 1, indicating a student activating this trigger has a lower hazard ratio compared to those not activating the trigger. The expected result would have been students with no prior UNE enrolment having either no significant difference or higher hazard ratio.

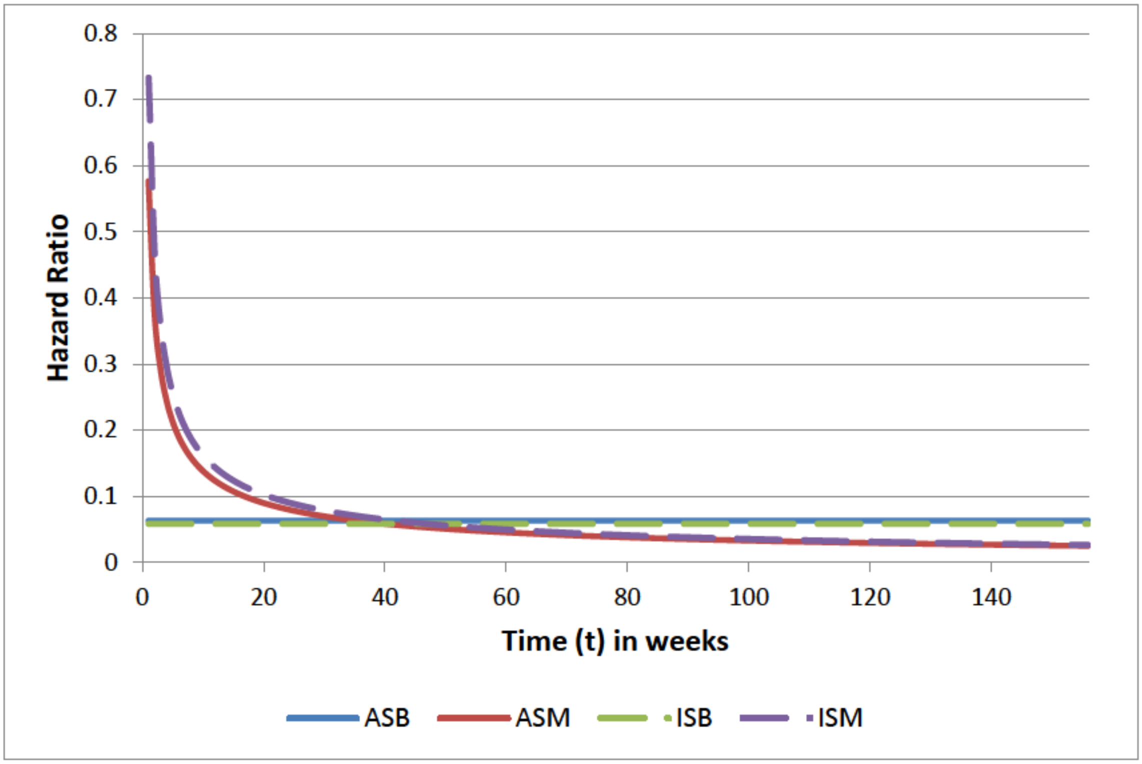


Figure 7.47 – EAS-trigger model: trigger 22 hazard ratio over time

The results show that the ASM and ISM models decrease over time, while the ASB and ISB models are constant. Despite this difference, the variation indicates that activation of this trigger correlates to a decreased hazard ratio. This is not an expected outcome. The inconsistent estimates against expectation indicate that this trigger may not be capturing students with genuine increased hazard.

Trigger 25 captures whether the student has been flagged for contact by the retention team in the current teaching period. It is expected that students already identified for follow-up contact would have a higher hazard ratio. The results for trigger 25, presented in Figure 7.48 over page, confirm this is the case.

Trigger 25 is considered as an inconsistent trigger as the ASB and ASM models show different estimates of the hazard ratio to the ISB and ISM models. For the ASB and ASM models, students identified for follow up support in the current teaching period have a higher hazard at the start of their enrolment. This decreases over time until around week 90 and 107 for the ASB and ASM models respectively, the hazard ratio is equal to 1. This means that this trigger, when applied to all students, indicates increased hazard for nearly the first two years of enrolment. According to the ISB and ISM models, students identified by the EAS who activate this trigger are at an increased hazard by around 36 to 47 per cent. While the results were inconsistent between models, they show this is an important trigger in identifying students with higher hazard ratios.

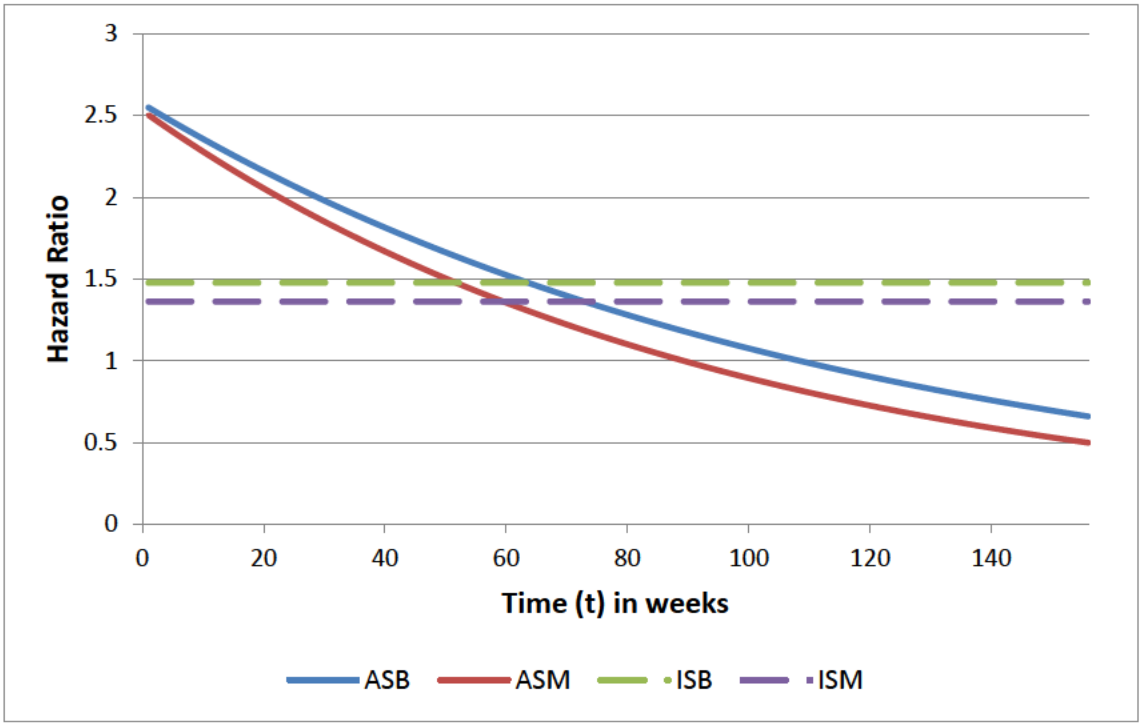


Figure 7.48 – EAS-trigger model: trigger 25 hazard ratio over time

Trigger 26 is more sporadic in the estimated hazard ratio over time, as shown in Figure 7.49. It captures students flagged for contact by the retention team in the previous teaching period. The results show the ASB. ISB and ISM models decreasing over time. Overall, the effect indicates that students previously flagged for contact by the student support team have a higher hazard ratio. For three of the models, this effect persists into the third year of study where the hazard ratio approaches 1. In the ASM model, the effect is constant over time, with an increased hazard ratio of around 300%. For the most part, this trigger provides a strong indication of increased hazard, demonstrating the value of recursive triggers in identifying students. However, the inconsistent and somewhat unstable estimates indicate that this variable may have unaccounted confounding effects with other variables.

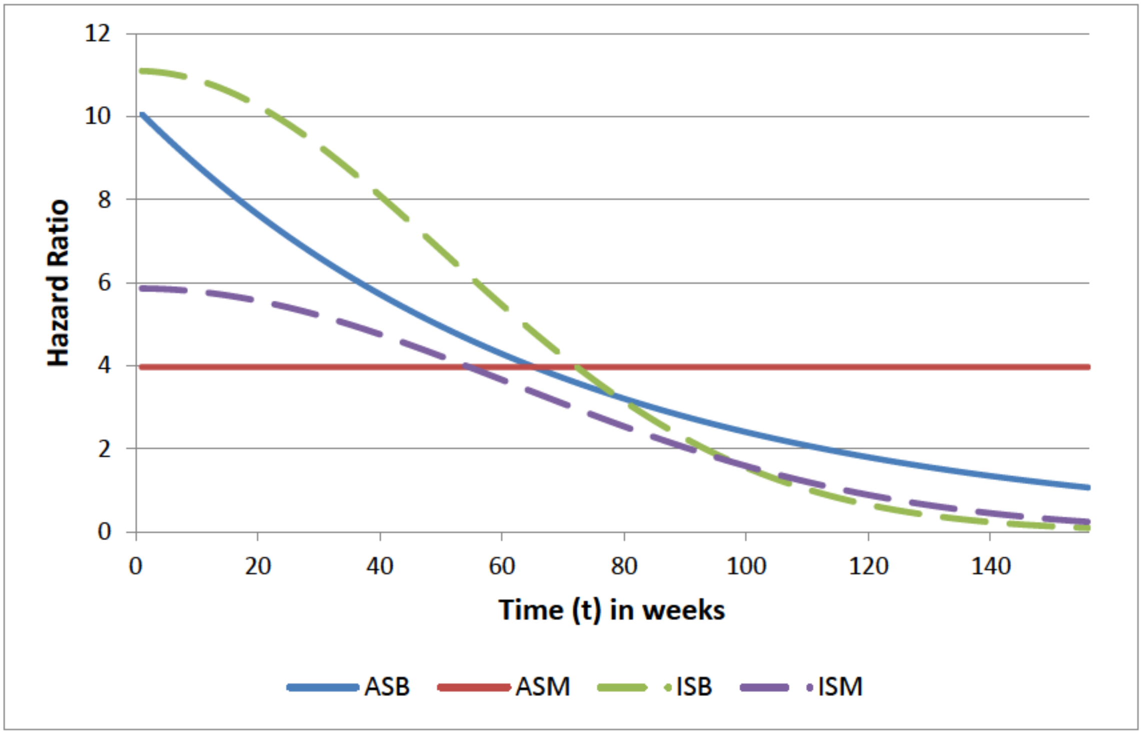


Figure 7.49 – EAS-trigger model: trigger 26 hazard ratio over time

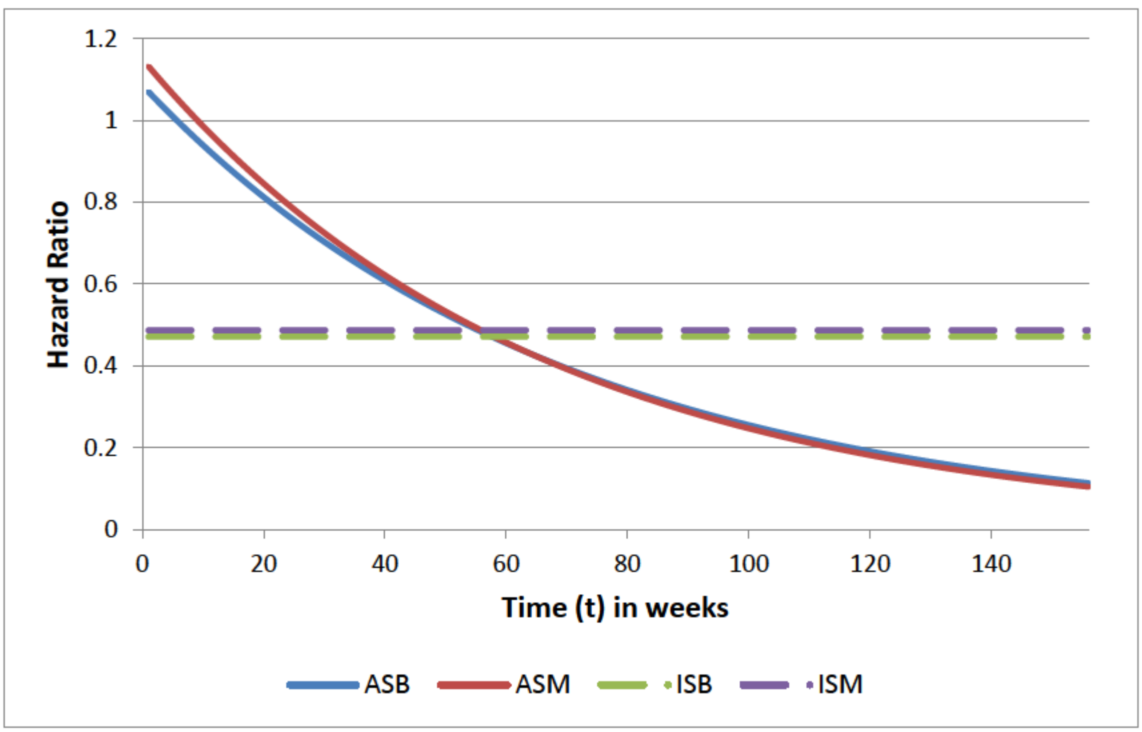


Figure 7.50 – EAS-trigger model: trigger 29 hazard ratio over time

Trigger 29 captures students who have low levels of portal inactivity, failing to access the online study portal for between 10 to 20 days. The estimated hazard ratios are plotted in Figure 7.50. The results indicate varying hazards where model ASB and ASM models are decreasing over time and ISB and ISM models are constant over time. The ASB and ASM models indicate that students who do not access the portal have an increased hazard initially, around 10 to 15 per cent higher than normal students. However, this increased hazard diminishes to parity by week eight or nine, after which the hazard for students activating this trigger is actually less than 1. The results indicate that this trigger is not stable and may not be an optimal predictor of increased risk of discontinuing.

Trigger 30 activates when a student has medium portal inactivity, not logging into the student portal for between 21 and 30 days. The results for this trigger are more pronounced than trigger 29. However, the estimated hazard over time indicates instability. The hazard ratio estimates are shown in Figure 7.51.

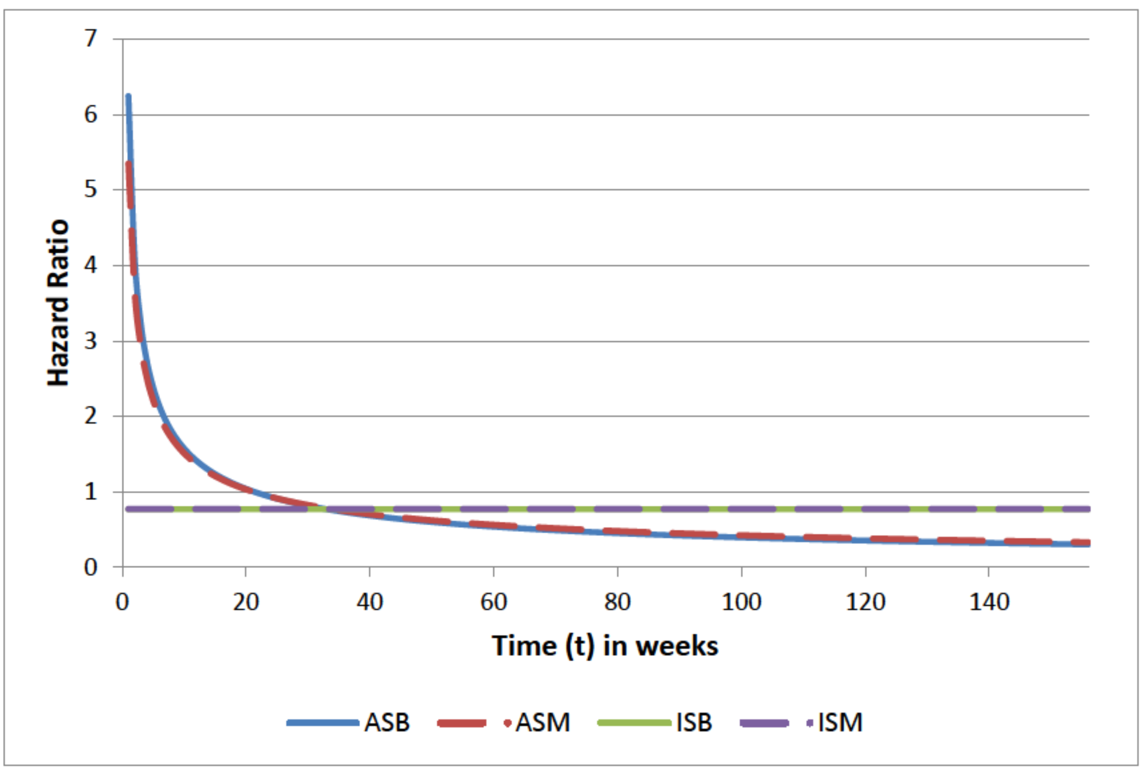


Figure 7.51 – EAS-trigger model: trigger 30 hazard ratio over time

The ASB and ASM models show an increased hazard ratio for students from weeks 1 to 20. This indicates that the trigger overall captures students with an increased hazard of discontinuing. However, within the ISB and ISM models, the results show that students identified by the EAS who activate this trigger have a slightly reduced hazard. This conflicting result indicates inconsistent hazards for students over time.

Trigger 31 captures very high portal inactivity, where students have failed to log into the student portal for 41 days or more. Estimated hazard ratios plotted in Figure 7.53 show that only the ASM model has an increased hazard of discontinuing over the first 24 weeks of enrolment.

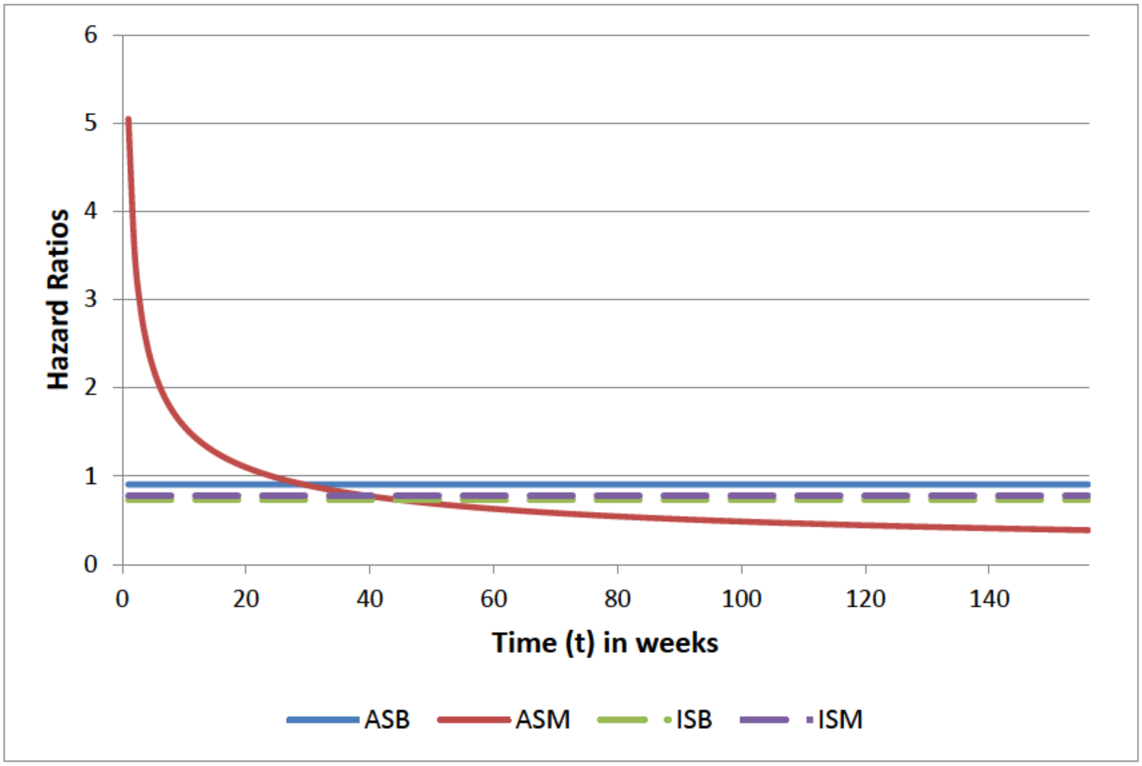


Figure 7.52 – EAS-trigger model: trigger 31 hazard ratio over time

As is the case for triggers 29 and 30, trigger 31 is logical in its inclusion within the EAS, capturing student portal inactivity; however, it may be affected by confounding variables causing inconsistent results. This instability may be rectified by adding appropriate control variables which is explored in the controlled-trigger model in section 7.7.

The final trigger with conflicting results is trigger 34. This trigger is activated if the student received a fail incomplete grade in the previous teaching period. As shown in preceding sections 7.3 to 7.5, attaining a fail incomplete grade increases the hazard ratio of a student, decreasing over time. The results for trigger 34 though are nonsensical as shown in Figure 7.53.

The ASB and ASM models decrease over time, starting with higher hazard ratios and reaching a hazard ratio of 1 around week 34. While the magnitude of the effect varies between the ASB and ASM models, they appear to reflect the overall trends discovered in previous models.

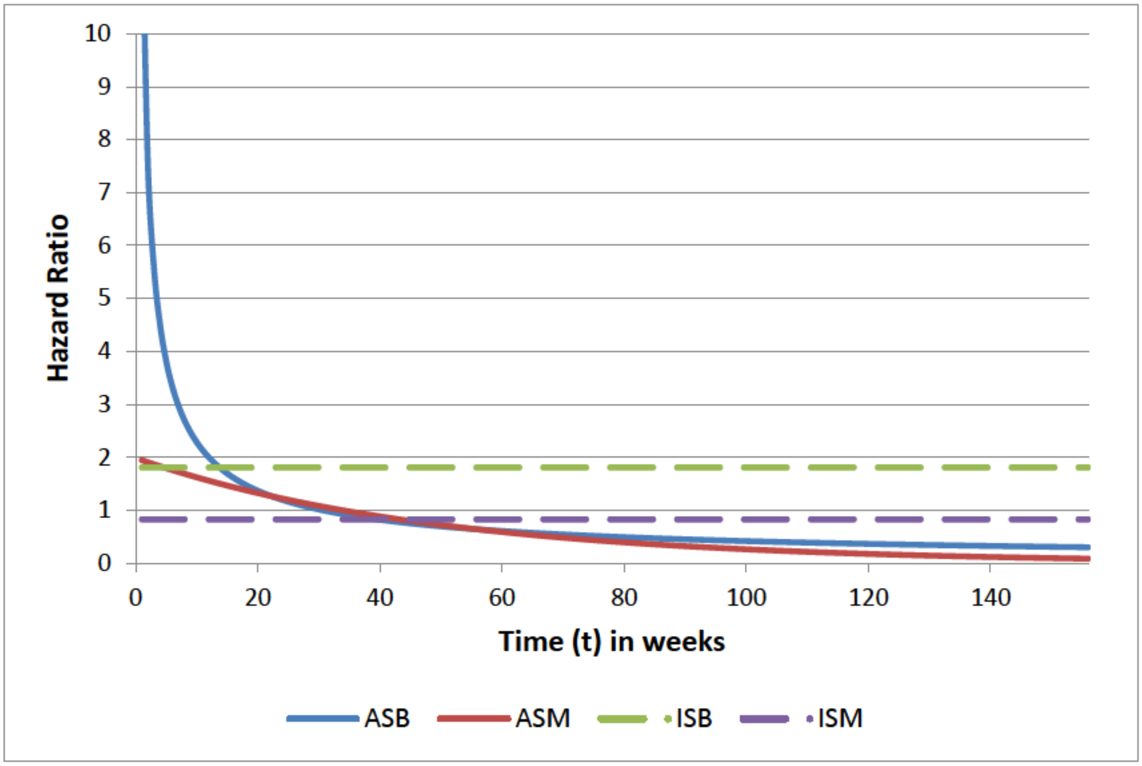


Figure 7.53 – EAS-trigger model: trigger 34 hazard ratio over time

The ISB and ISM models produced conflicting results, with the ISB model indicating a constant increase in hazard. The ISM indicates a constant decrease in hazard. However, both of these estimates are not significant even at the 10% level. As such, for a student identified by the EAS, activating this trigger actually indicates no change in hazard.

7.6.7 Triggers with no significant effects over times

In total, 11 out of the 34 triggers had no significant relationship with the hazard ratio of students. The results are presented in Table 7.24.

Trigger 1 indicates a student who was admitted through an alternative entry pathway. It is expected that students admitted through alternative entry are less prepared for university study and as such have a higher hazard ratio. The results show that students who are admitted this way have no significant difference in hazard. This either means that the assumed relationship does not exist, or the EAS has misspecified the trigger such that it does not actually capture students with increased hazard of discontinuing.

Table 7.24 – EAS-trigger model: EAS triggers with no significant effects

	ASB		ASM		ISB		ISM	
Trigger	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
1	0.341	0.243	0.354	0.247	0.405	0.289	0.454	0.311
2	0.715	0.296	0.851	0.287	0.706	0.32	0.729	0.295
7	-	-	-	-	-	-	-	-
8	0.950	0.371	0.934	0.352	0.935	0.431	0.915	0.409
9	0.780	0.358	0.937	0.390	0.632	0.373	0.840	0.429
10	1.011	0.516	1.045	0.515	0.883	0.520	0.974	0.555
15	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	-
17	-	-	-	-	-	-	-	-
18	-	-	-	-	-	-	-	-
23	1.290	0.403	0.878	0.234	1.365	0.445	1.058	0.298

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Trigger 2 captures Aboriginal or Torres Strait Islander students. As shown in previous models, ATSI students identified by the EAS usually have a significantly lower hazard than non-ATSI students. Trigger 2, being insignificant, indicates that there is model misspecification for ATSI status and may need updating to accurately reflect the underlying hazard for ATSI students.

Triggers 7, 8, 9 and 10 correspond to the e-motion tool. The emotion tool allows students to reflect how they are feeling about a unit of study be setting an emoticon to varying levels of happiness. Trigger 7 reflects students who “do not want to say” how they are feeling about a unit. This trigger was excluded from the model due to insufficient sample size, causing the model to not reach convergence. Trigger 8 corresponds to the neutral feeling option, trigger 9 corresponds to feeling unhappy about a unit and trigger 10 corresponds to feeling very unhappy about a unit. Logically, unhappy feelings towards the units students are studying should capture increased hazard ratio. However the results for triggers 8, 9 and 10 are insignificant. This may be a result of insufficient sample size and overall underutilisation of the tool. As such, it may be a strong indicator in a specific case. However, in the context of the survival mode, these triggers do not contribute to accurately identifying students with increased hazard of discontinuing.

Triggers 15 to 18 are meant to capture data on students receiving assignment extensions in both the current and previous teaching periods. After sorting the data for analysis, triggers 15 through to 18 recorded no observations. These variables were not included in the model as it appears the triggers are yet actually to be implemented in the EAS. Theoretically, these seem to be logical inclusions in an EAS design, however, these will need to be empirically tested for validity in a future study.

Trigger 23 captures students who enrolled in five or more units in a single teaching period. While students undertaking additional workload above the full-time four units per teaching period, this trigger hardly captures students at risk of disengaging or discontinuing. As such, there is no change in the hazard ratio of students who attempt above full-time workload. This trigger may actually be causing students to be incorrectly identified as at risk and may be better excluded from the EAS overall.

7.6.8 Summary of effects and implication for EAS design

The original design parameters for the EAS were to identify students at risk of disengaging from their studies. Analysing the current EAS trigger specification with respect to discontinuation, reveals the degree currently defined triggers capture students with increased hazard of discontinuation. To summarise the results, Table 7.25 indicates which triggers reflect an increased hazard ratio and when the hazard ratio is above 1. This shows a potential parsimonious approach to identifying students at risk of discontinuing using the already existing EAS triggers for disengaging.

Table 7.25 – EAS-trigger model: recommended valid EAS triggers over time

Trigger	Valid Time Frame
3	Weeks 1 to 4
11	Weeks 1 to 9
13	Weeks 1 to 11
19	Weeks 1 to 28
20	Weeks 91 +
25	Weeks 1 to 89
26	Weeks 1 to 110
28	Weeks 1 to 19

While the eight triggers capture increased hazard ratios over different time periods, many of the other triggers not included capture important theoretical relationships. Given that models from previous sections capture important relationships between expected variables, one way to test the validity of the triggers is to the base model as control variables. As indicated in 7.7.6, many triggers were sensitive to model specification, showing inconsistent results. Controlling for other factors may change the behaviour of the triggers to align with expectation. As such, the controlled-trigger model joins the base model from section 7.3 and the EAS-trigger model to analyse the triggers, controlling for *demographic*, *institutional*, *student performance* and *workload* variables.

7.7 Controlled-trigger model results

7.7.1 Controlled-trigger model description

The fifth model tested using survival analysis extends the EAS-trigger model by combining it with the base model. This allows two types of inferences to be made throughout the section. The first is how triggers behave in the presence of demographic, institution and learning environment control variables. The second inference is around EAS design, where the model can identify additional variables beyond the EAS triggers which should be included in the EAS. The controlled-trigger model divides triggers into the same configurations as the previous section. To distinguish the four configurations, they will be termed as follows: Controlled All-Student-Base (CASB); Controlled All-Student-Multiple (CASM); Controlled Identified-Student-Base (CISB); Controlled Identified-Student-Multiple (CISM). The CASB and CASM models are represented in Figure 7.54 over page, where the red pathway only focuses on the identification aspect.

The CISB and CISM models, however, filter the triggers to only include triggers that were activated by students who, at some point in their enrolment, were identified by the EAS for targeted student support. This is diagrammatically represented in Figure 7.55 over page, where the red pathway is only for those students identified by the EAS, while the black pathway is students who never were identified.

The separation of students into the two classes in the CISB and CISM models allow the effect of the triggers to be limited to only those students identified and de-noises the data to help clarify the real effects for those students identified by the system.

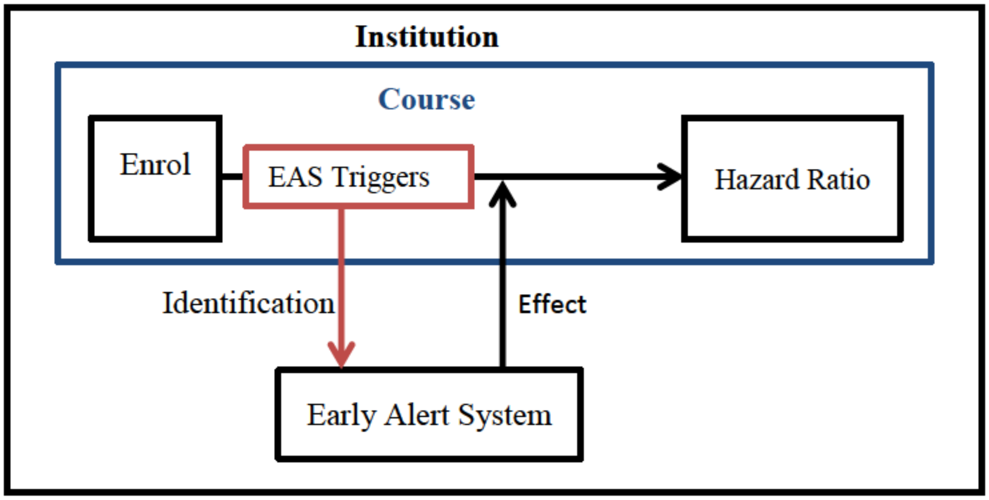


Figure 7.54 – Controlled-trigger model: CASB and CASM configurations
(Authors own contribution)

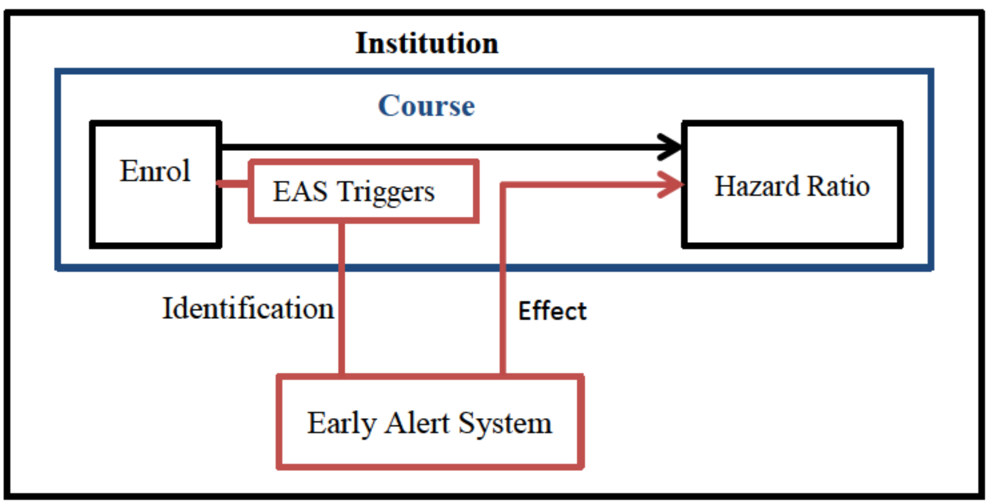


Figure 7.55 – Controlled-trigger model: CISB and CISM configurations
(Authors own contribution)

7.7.2 Significance of controlled-trigger model and assumptions tests

The controlled-trigger model is the most complex model presented in this study. The sub-models contain between 89 and 95 explanatory variables. The results presented in Table 7.26 show all models are significant at the 1% level.

Table 7.26 – Controlled-trigger model: overall significance and assumption tests

CASB		CASM		CISB		CISM	
LR χ^2 (93)	1830.6	χ^2 (95)	1943.69	LR χ^2 (89)	1627.8	χ^2 (90)	1712.35
Prob > χ^2	0	Prob > χ^2	0	Prob > χ^2	0	Prob > χ^2	0

PH Test		PH Test		PH Test		PH Test	
χ^2 (93)	39.16	χ^2 (95)	40.1	χ^2 (89)	27.85	χ^2 (90)	21.62
Prob > χ^2	0.6789	Prob > χ^2	0.6396	Prob > χ^2	0.9265	Prob > χ^2	0.9922

The proportional hazards test for each model show no significant correlation with time. As such, the hazard ratio inferences are valid for all models. The detailed breakdown of the proportional hazards assumption test is presented in Appendix B.

7.7.3 Demographic variables

The demographic results presented are stable across all four versions of the controlled-trigger model and are within the expected values given the results of the underlying base model. The results are presented in Table 7.27.

Table 7.27 – Controlled-trigger model: demographic variables

	CASB		CASM		CISB		CISM	
Variable	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Gender	1.078[b]	0.034	1.076[b]	0.034	1.078[b]	0.034	1.077[b]	0.034
Age	0.971[a]	0.007	0.969[a]	0.007	0.970[a]	0.007	0.969[a]	0.007
Age Squared	1.000[a]	0.000	1.000[a]	0.000	1.000[a]	0.000	1.000[a]	0.000
ATSI	0.850[c]	0.073	0.849[c]	0.072	0.854[c]	0.073	0.855[c]	0.073

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Gender indicates that female students are around 7.6 to 7.8 per cent higher in hazard ratio and is significant at the 5% level. Age has a significant relationship at the 1% level and points to a decreasing hazard ratio over time until the age of 44 where it turns to increase again. This is in line with the expectations developed in the base model, albeit the turning point occurs a few years earlier. Finally, ATSI students have a hazard ratio that is around 15 per cent lower than non-ATSI students. This result is stable across the model specifications when including the triggers, indicating that the demographic variables should form part of the EAS. In an additive EAS model, the weights associated with the relevant triggers should be the hazard ratios.

7.7.4 Institutional variables

Using institutional variables to control for EAS trigger effects shows similar estimated hazard coefficients to those obtained the base model in section 7.3. Results for institutional variables are presented in Table 7.28 and school variables presented in Table 7.29.

Table 7.28 – Controlled-trigger model: institutional variables

Variable	CASB		CASM		CISB		CISM	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Domestic Fee	0.628	0.352	0.602	0.336	0.590	0.33	0.579	0.323
Domestic Fee x t	1.01	0.007	1.010	0.007	1.011	0.007	1.011	0.007
International Fee	0.191[a]	0.058	0.188[a]	0.057	0.184[a]	0.056	0.187[a]	0.056
International Fee x t^2	1.001[a]	0.000	1.001[a]	0.000	1.001[a]	0.000	1.001[a]	0.000
International Fee x t^3	1.000[b]	0.000	1.000[b]	0.000	1.000[a]	0.000	1.000[b]	0.000
Prior Studies	0.411[a]	0.072	0.422[a]	0.074	0.442[a]	0.076	0.468[a]	0.08
Prior Studies x ln(t)	1.207[a]	0.054	1.201[a]	0.053	1.191[a]	0.053	1.175[a]	0.052
On-campus	1.113[b]	0.053	1.115[b]	0.053	1.113[b]	0.053	1.120[b]	0.053
On-campus x t	0.919[a]	0.012	0.920[a]	0.012	0.920[a]	0.012	0.922[a]	0.012
Diploma	0.995	0.111	0.992	0.111	0.996	0.111	0.997	0.112
Advanced Diploma	0.764[b]	0.091	0.741[b]	0.088	0.748[b]	0.089	0.741[b]	0.088
Advanced Diploma x t	1.003[c]	0.002	1.003[b]	0.002	1.003[c]	0.002	1.003[b]	0.002
Bachelors (Graduate)	0.751[a]	0.051	0.752[a]	0.051	0.746[a]	0.051	0.746[a]	0.051
Bachelors (Honours)	0.613[a]	0.077	0.603[a]	0.076	0.618[a]	0.077	0.607[a]	0.076

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The institution variables are within expectations formed in the base model and remain relatively constant across the four models with the coefficients in close proximity to each other. International fee students have a hazard which varies over time, peaking around the end of the second year of study, before decreasing again.

Students having undertaken prior studies are still significant at the 1% level, however the benefit of prior studies is more pronounced. In the base model, students with prior study had a decreased hazard ratio up to week 68. In the controlled-trigger model, the hazard ratio is less than 1 until week 107 as shown in Figure 7.56. Only the CASB model is shown as the other configurations were nearly identical to this.

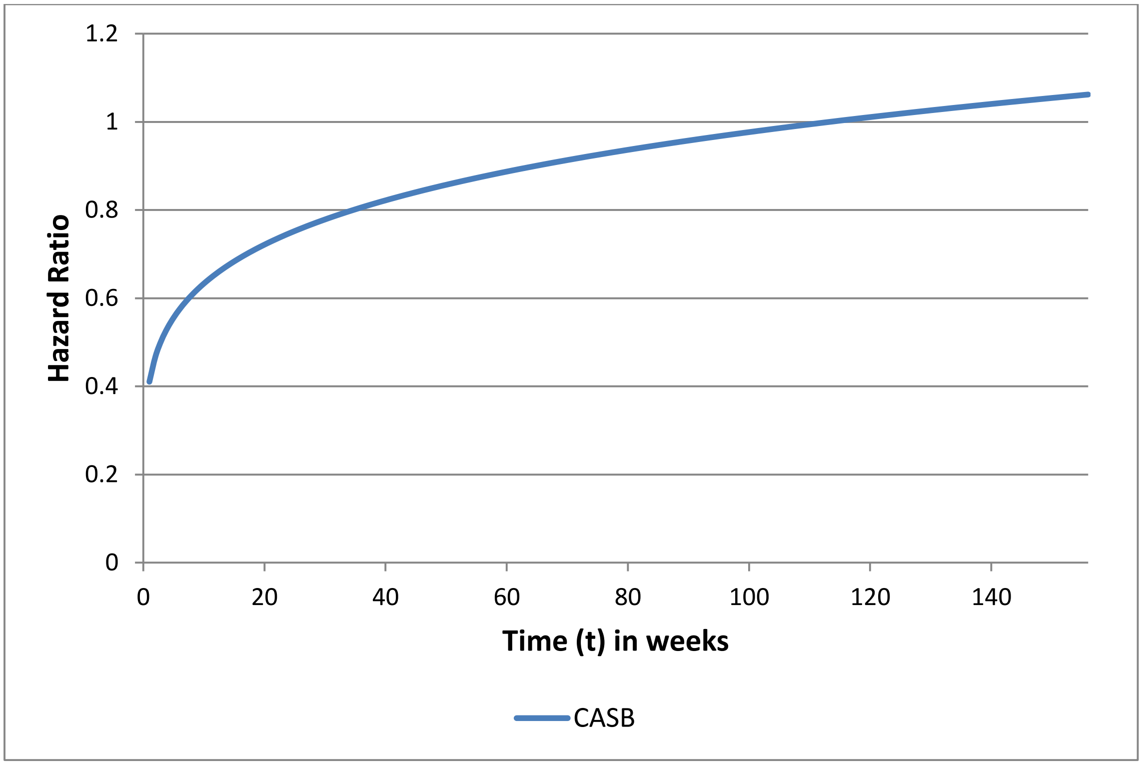


Figure 7.56 – Controlled-trigger model: prior study hazard ratios over time

The change in the effect of prior studies between the base model and the controlled-trigger model is only minor. Overall, this indicates that a student with some history of prior studies has a lower hazard, especially over the first year of study.

The effects of living on-campus are similar to the base model. On-campus students initially start with a hazard higher than off-campus online students, significant at the 5% level. Over time this rapidly decreases indicating that the initial hazard of living on-campus is soon outweighed by the benefits of on-campus learning. This indicates that a strong focus needs to be placed in off-campus online student support.

Degree type indicates the expected result that students who are graduate students or honours students enrolled in a bachelor's degree have a significantly lower hazard ratio than students admitted through normal entry. The time varying effects associated with an advanced diploma also hold to earlier estimates from the base model. This indicates a high level of stability between the models.

Controlling for school variables, Table 7.29 presents the estimated hazard ratios for each school within the institution.

Table 7.29 – Controlled-trigger model: school variables

	CASB		CASM		CISB		CISM	
Variable	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
School 1	1.105	0.076	1.100	0.076	1.104	0.076	1.099	0.076
School 2	1.276[a]	0.091	1.270[a]	0.091	1.277[a]	0.091	1.272[a]	0.091
School 2 x 1/t	0.005[a]	0.005	0.005[a]	0.005	0.005[a]	0.005	0.005[a]	0.005
School 3	1.193[a]	0.081	1.188[b]	0.080	1.194[a]	0.081	1.189[b]	0.08
School 4	1.299[a]	0.082	1.298[a]	0.082	1.304[a]	0.083	1.300[a]	0.082
School 5	1.290[a]	0.089	1.297[a]	0.090	1.297[a]	0.090	1.300[a]	0.09
School 6	0.834[b]	0.068	0.824[b]	0.067	0.841[b]	0.068	0.831[b]	0.068
School 6 x t	1.005[a]	0.001	1.005[a]	0.001	1.004[a]	0.001	1.005[a]	0.001
School 7	0.830	0.315	0.846	0.321	0.828	0.314	0.844	0.321
School 8	1.030	0.082	1.030	0.082	1.031	0.082	1.029	0.082
School 9	0.963	0.077	0.960	0.077	0.962	0.077	0.957	0.077

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The results provided are similar to the base model, with minor changes in the level of significance. School 1 has no significant difference in hazard from the base case school in the controlled-trigger model. In the base model, school 1 had a hazard ratio higher than the base school significant at the 10% level for models short-run and enduring effects models. School 6 had a base hazard ratio significant at the 10% level in the base model, where in the controlled-trigger model, the hazard ratio is significant at the 5% level. Overall, the hazard ratio differences between schools remain relatively constant.

7.7.5 Student performance and workload variables

Student performance and workload are significant variables that affect the hazard ratio. The results for controlled-trigger model are presented in Table 7.30.

Table 7.30 – Controlled-trigger model: student performance and workload

	CASB		CASM		CISB		CISM	
Variable	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Withdrawn	1.361[a]	0.094	1.344[a]	0.093	1.335[a]	0.092	1.340[a]	0.092
Withdrawn x t	0.993[a]	0.002	0.993[a]	0.002	0.993[a]	0.002	0.993[a]	0.002
Withdrawn x t^2	1.000[a]	0.000	1.000[a]	0.000	1.000[b]	0.000	1.000[a]	0.000
Withdrawn Early	1.474[a]	0.083	1.430[a]	0.080	1.449[a]	0.081	1.433[a]	0.08
Withdrawn Early x t	0.992[a]	0.002	0.993[a]	0.002	0.992[a]	0.002	0.993[a]	0.002
Withdrawn Early x t^2	1.000[a]	0.000	1.000[a]	0.000	1.000[a]	0.000	1.000[a]	0.000
Fail Incomplete	0.993	0.076	1.274[a]	0.025	1.270[a]	0.025	1.276[a]	0.025
Fail Incomplete x t	2.484[a]	0.768	1.000[c]	0.000	1.000	0.000	1.000[c]	0.000
Fail	1.154[a]	0.021	1.153[a]	0.021	1.152[a]	0.021	1.153[a]	0.021
Pass	0.848[a]	0.012	0.848[a]	0.012	0.847[a]	0.012	0.848[a]	0.012
Credit	0.894[a]	0.013	0.892[a]	0.013	0.893[a]	0.013	0.892[a]	0.013
Distinction	0.870[a]	0.014	0.870[a]	0.014	0.870[a]	0.014	0.871[a]	0.014
High Distinction	0.876[a]	0.018	0.876[a]	0.018	0.875[a]	0.018	0.876[a]	0.018
Other	0.743[a]	0.052	0.744[a]	0.053	0.742[a]	0.052	0.745[a]	0.053
Inactive	56.401[a]	7.260	53.600[a]	6.928	52.352[a]	6.707	50.372[a]	6.472
Inactive x t (for t < 17)	0.057[b]	0.064	0.034[a]	0.042	0.090[a]	0.081	0.067[b]	0.070
Inactive x 1/ln(t) (for t > 16)	0.056[a]	0.011	0.060[a]	0.012	0.061[a]	0.012	0.061[a]	0.012
Part-time	1.491[b]	0.246	1.451[b]	0.238	1.393[b]	0.228	1.399[b]	0.228
Part-time x ln(t)	1.097[c]	0.061	1.096	0.061	1.109[c]	0.062	1.097[c]	0.061

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Taking the grade results in order, both withdrawn and withdrawn early estimates are in line with the base model. Both show quadratic relationships whereby the effect decreases then increases over time. Students withdrawing have an increased hazard ratio only over the first year of study. Students withdrawing early show a decreasing hazard ratio. However, it remains above 1 for all 156 weeks captured in the model. This indicates that withdrawn early should be included in EAS design for the whole time, while withdrawing only need be factored in if the student is still within the first year of study.

Fail incomplete results vary slightly from the base model. The estimated hazard ratios are presented in Figure 7.57 for the CASB and CASM models. The CASM model is nearly identical to the CISB and CISM models.

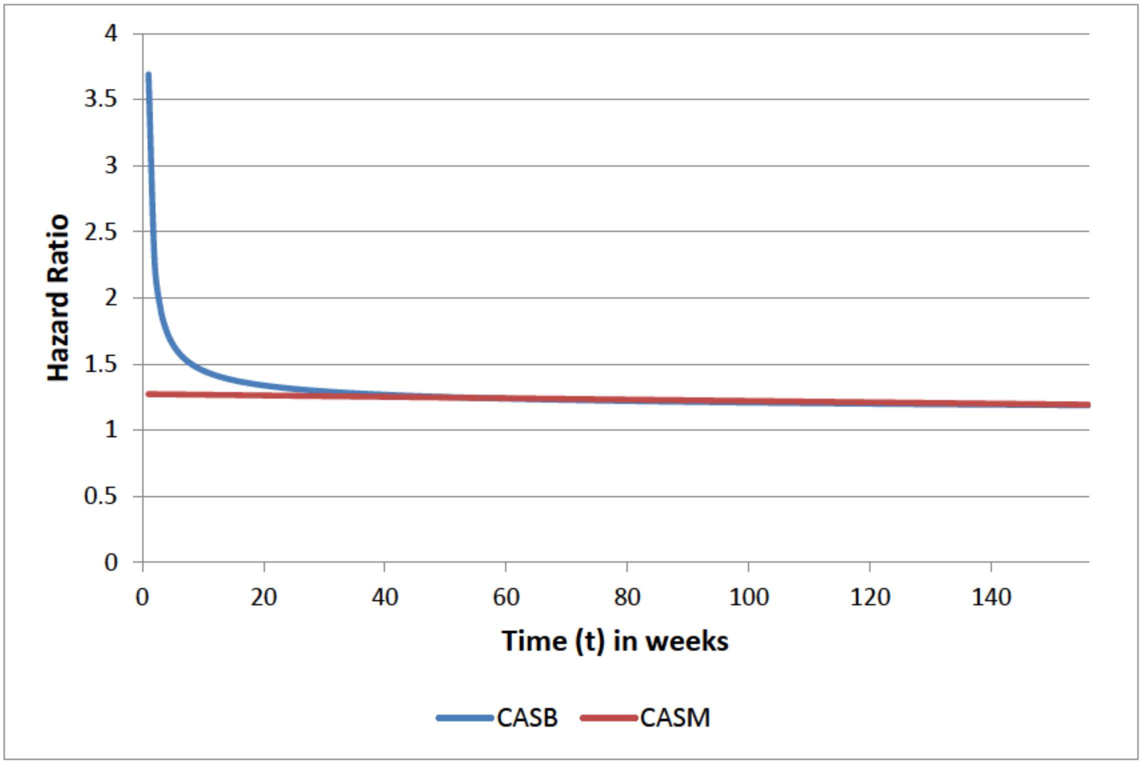


Figure 7.57 – Controlled-trigger model: fail incomplete hazard ratios over time

The results show a decreasing hazard over time; however CASB initially starts with a significantly higher hazard estimate than the other models. This decreases over time and eventually provides a similar estimate of hazard past week 26. Overall, the effect indicates that each fail incomplete grade attained significantly increases the estimated hazard of a student.

All of the remaining grades have constant effects over time. Each time a student fails a unit, the hazard ratio increases by around 15 per cent. Each time a student attains a grade that contributes to their progression, the hazard ratio decreases by 10 to 15 per cent. If a student receives an "other" grade, capturing the full range of administrative grades, the hazard decreases further.

Students who are inactive still have one of the greatest hazard effects in the model. The estimates indicate that inactive students at the very beginning of their enrolment will have a significantly higher hazard ratio. This decreases close to 0 between weeks 2 and 16, indicating full-time students have a significantly higher hazard ratio over this time. This likely occurs due to limited inactive observations occurring over this time period. After week 16, the hazard ratio associated with inactivity rapidly increases, indicating students who are inactive after the first teaching period have a higher hazard of discontinuing.

The hazard estimates for part-time students are different to the base model, showing more consistent results between configurations. The hazard ratios for part-time students are plotted in Figure 7.58.

The results indicate that part-time students start with a higher initial hazard ratio compared to full-time students. The hazard increases over time, indicating that compared to full-time students, part-time students have increasing difficulty throughout their studies. The results indicate that studying part-time is not optimal. Furthermore, full-time students struggling with studies may transfer to part-time study as a solution to their issues, without understanding that studying part-time also increases their hazard ratio. This provides an important result that part-time study should be a variable captured within an EAS and factored into the advice given to students about workload.

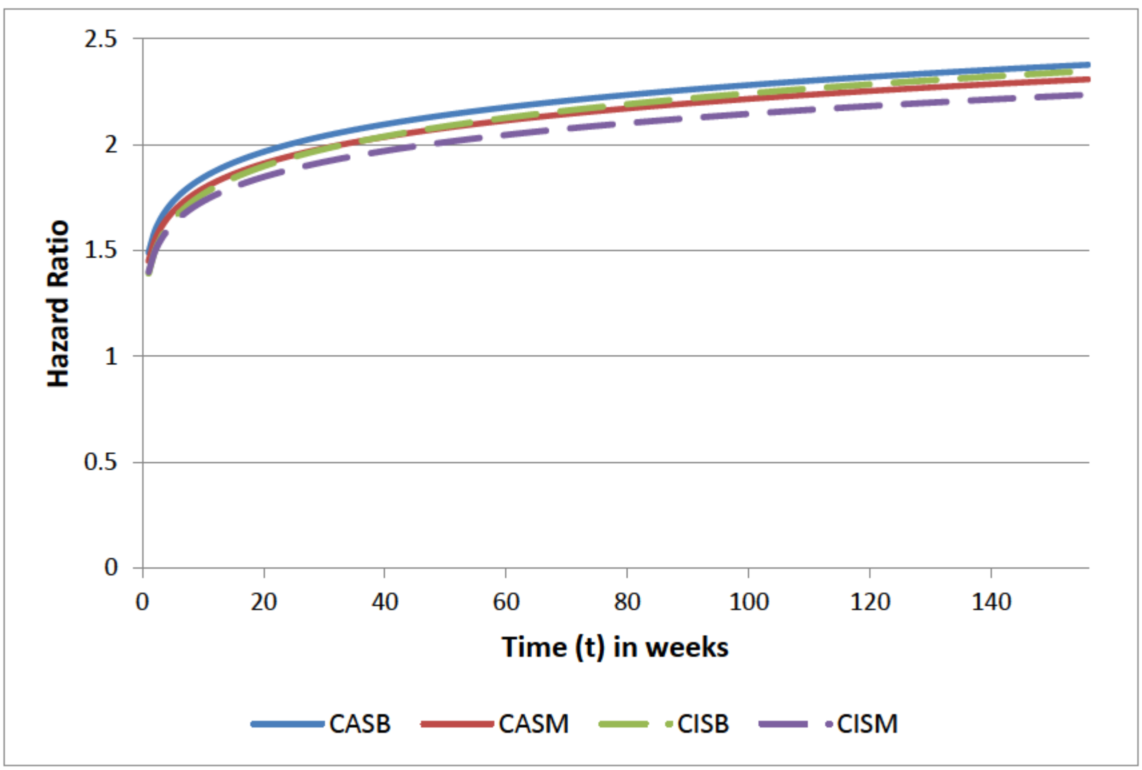


Figure 7.58 – Controlled-trigger model: part-time hazard ratios over time

7.7.6 *Triggers with constant effects over time*

Using model the base model to control for *demographic*, *institutional*, *student performance* and *workload* variables, the controlled-trigger model yields more triggers with constant effects than the EAS-trigger model. The estimated hazard ratios are presented in Table 7.31. Many of the variables include interactions with time that are not significant. This corresponds to ensuring the variables do not violate the proportional assumptions tests for survival analysis.

Table 7.31 – Controlled-trigger model: triggers with constant effects over time

	CASB		CASM		CISB		CISM	
Trigger	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
3	0.660[a]	0.071	0.876[a]	0.016	0.550[a]	0.067	0.858[a]	0.018
3 x t	0.999	0.003	1.000	0.000	1.001	0.003	1.001	0.000
11	1.490[a]	0.211	1.608[a]	0.218	1.364[c]	0.235	1.511[b]	0.248
11 x t	0.998	0.003	0.997	0.003	0.996	0.004	0.995	0.004
19	3.822[b]	2.153	3.645[b]	1.934	3.771[b]	2.441	3.548[b]	2.079
19 x t	0.956	0.030	0.96	0.028	0.956	0.033	0.963	0.028
24	0.373[a]	0.115	0.425[a]	0.122	0.327[a]	0.118	0.384[a]	0.131
25	2.680[a]	0.324	2.303[a]	0.261	2.119[a]	0.202	1.865[a]	0.164
25 x t	1.001	0.003	1.002	0.003	-	-	-	-
27	0.298[b]	0.151	0.337[b]	0.145	0.333[b]	0.169	0.336[b]	0.148
33	0.515[a]	0.130	0.543[a]	0.120	0.56[b]	0.143	0.595[b]	0.135

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Trigger 3 was decreasing in the previous EAS-trigger model, changing to a significant constant effect in the controlled-trigger model. The trigger corresponds to students being enrolled in a high attrition unit. The results indicate that students activating this trigger have a significantly lower hazard ratio than students not activating this trigger. This is a nonsensical response given the control variables in the model. This suggests misspecification of the EAS trigger as the hazard ratio is not within expectations of an increased hazard ratio.

Trigger 11 corresponds to students having high e-reserve usage inactivity, indicating a student has not accessed the online library resources for between 31 and 40 days. The EAS-trigger model, this trigger was only valid for the first few weeks of a student's enrolment. With the control variables in place, the trigger now provides a sensible result. Students activating this trigger in the controlled-trigger model have an increase in the hazard ratio of between 36 to 60 per cent.

Trigger 19 corresponds to slower student progression where students who have already completed a number of units enrol in less than half the number of completed units. In many instances, this trigger is activated during periods of inactivity or when students change from full-time to part-time workload between teaching periods. The results indicate that students who activate this trigger have a significantly higher hazard ratio than students not activating this trigger. This indicates that under the presence of the control variables, the trigger is a logical inclusion in the EAS design.

Trigger 24 corresponds to a student previously enrolled in a pathways-enabling course. These courses provide support to students whose academic skills may not be sufficient to start university immediately. The trigger indicates that students activating this trigger have a significantly lower hazard ratio than students not activating the trigger. This goes against expectations of increased hazard. As such, trigger 24 is misspecified in the presence of the control variables and may actually be identifying students incorrectly.

Trigger 25 corresponds to students flagged for contact by the retention team in the current teaching period. The EAS-trigger model, this trigger had some instability between configurations however in the controlled-trigger model, this trigger is relatively stable between configurations. In the presence of the control variables, it shows students already flagged for support have a hazard ratio 80 to 168 per cent higher than students not already flagged for support. In many ways this trigger indicates that the right students are being identified for support. The recursive nature of the trigger may also mean that students identified for support will continue to be identified for support. This may be problematic as the EAS could be populating the list of "most at risk" students with students who have already been identified as at risk. As such, while good at identifying students with increased hazard, it may actually be causing new students to miss out on being identified for support because they have not already been identified.

Trigger 27 captures students carrying over a special extension of time exam. In the EAS-trigger model, the trigger showed a significant reduction in hazard which confirmed findings of previous models. This remains the case with the controlled-trigger model; students activating this trigger have a significantly lower hazard ratio. This indicates that while SET exams may indicate students disengaging from their studies, it is not indicative of students who are going to discontinue their studies. This trigger may need revising to ensure it corresponds to the EAS objective.

Finally, trigger 33 corresponds to a student who received a fail grade in the last teaching period. Fail grades are captured in the control variables and indicate an increased hazard if a fail grade is attained. Trigger 33 captures the lag effect of failing a unit. In this case, the estimated hazard for a student who attained a fail in the previous teaching period is actually more than 50 per cent below that of students who did not fail a unit in the previous teaching period. This trigger may actually be capturing students who have since corrected their enrolment trajectory post fail grade. The student might have engaged with student support, re-dedicated themselves to their studies and be more motivated to succeed after failing a unit. The results indicate that this trigger may not be capturing students at risk of discontinuing and may need to be reframed as part of EAS design to capture more immediate effects, such as failing a unit in the last four weeks.

7.7.7 Triggers with decreasing effects over time

In the previous EAS-trigger model, there were seven triggers which had decreasing hazard ratios over time. Some triggers showed significantly higher hazard at the start of the students' enrolment before tapering off. In the controlled-trigger model, only two triggers have decreasing hazard ratios over time. These are presented in Table 7.32.

Table 7.32 – Controlled-trigger model: triggers with decreasing effects over time

	CASB		CASM		CISB		CISM	
Trigger	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
26	6.895[a]	1.084	3.935[a]	0.512	8.568[a]	1.437	4.508[a]	0.597
26 x t	0.973[a]	0.003	0.978[a]	0.003	0.972[a]	0.003	0.979[a]	0.003
28	2.249[a]	0.663	2.54[a]	0.718	11.265[a]	8.474	11.899[a]	8.198
28 x t	0.986	0.009	0.984[c]	0.009	-	-	-	-
28 x ln(t)	-	-	-	-	0.478[a]	0.127	0.486[a]	0.121

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Trigger 26 captures students flagged for contact with the student support team in the previous teaching period with results presented in Figure 7.59.

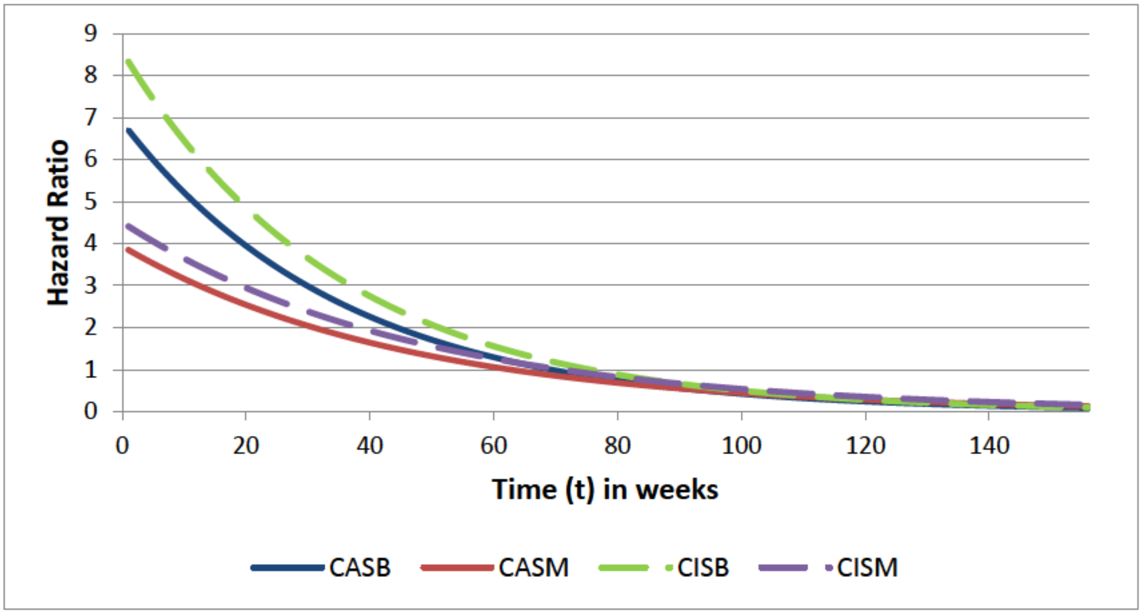


Figure 7.59 – Controlled-trigger model: trigger 26 hazard ratio over time

The hazard estimates show varying levels of initial hazard at the beginning of the students' enrolment. All four sub-models decrease at varying rates, but converge to similar hazard ratios from around week 70 onwards. Figure 7.59 shows that students activating this trigger have a higher hazard ratio for the first year of study. Given that the trigger cannot be activated in the first teaching period of enrolment due to its lag effect, this limits the trigger's usefulness to a narrow time period. This does not preclude the trigger from EAS inclusion, but the trigger should be conditional on the number of weeks studied.

Trigger 28 captures students with high portal inactivity, not logging onto the student portal for between 31 to 40 days. The estimated hazard ratios are plotted in Figure 7.60 over page. The CISM model only uses triggers by identified students. The hazard ratio is above 1 between initial enrolment and week 26. Given this trigger requires 31 days of portal inactivity to have occurred first, this indicates that the trigger is only valid between weeks 5 and 26. As such, the trigger is a useful indicator of possible early discontinuation; however, it is not a valid trigger after the first half a year of study.

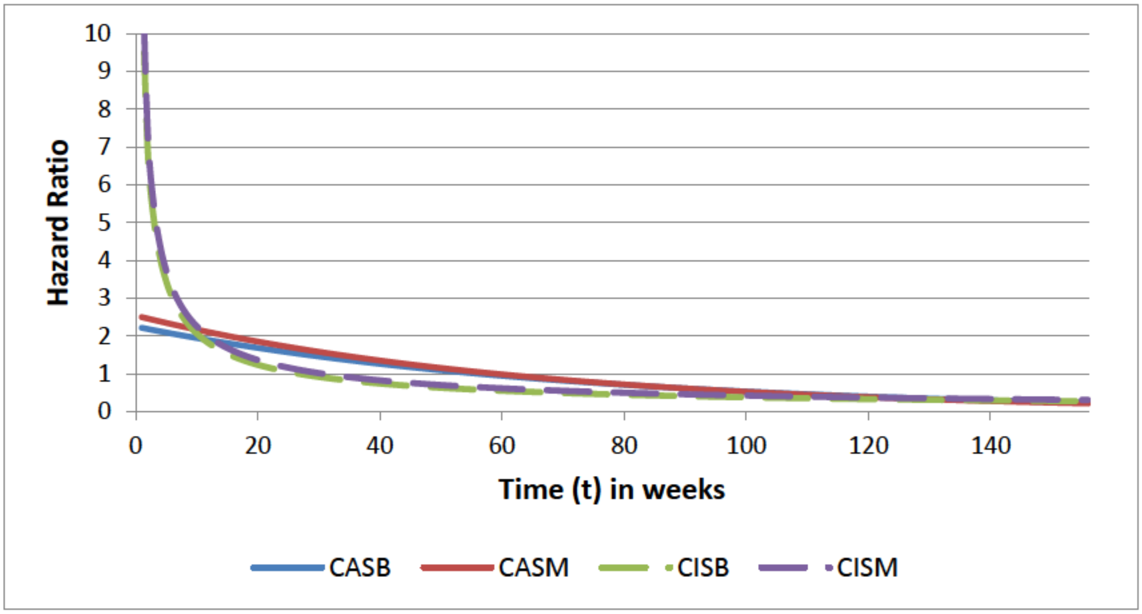


Figure 7.60 – Controlled-trigger model: trigger 28 hazard ratio over time

7.7.8 Triggers with increasing effects over time

Triggers with increasing effects over time mean that students activating those triggers have an increased hazard ratio as students' enrolment progresses. Such triggers can form an important tool in identifying students at higher risk of discontinuing later in their enrolment. The estimated hazard ratios for triggers with increasing effects are presented in Table 7.33.

Table 7.33 – Controlled-trigger model: triggers with increasing effects over time

'	CASB		CASM		CISB		CISM	
Trigger	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
5	2.319[a]	0.266	1.131[a]	0.022	2.412[a]	0.290	1.137[a]	0.023
5 x 1/t	0.000[a]	0.000	0.127[a]	0.044	0.000[a]	0.000	0.097[a]	0.038
12	0.919	0.140	1.105	0.152	0.668[b]	0.126	0.873	0.147
12 x t	1.007[b]	0.003	1.005[c]	0.003	1.010[a]	0.003	1.007[b]	0.003
13	1.197	0.160	1.292[b]	0.156	0.990	0.160	1.136	0.161
13 x t	1.007[b]	0.003	1.006[b]	0.003	1.009[a]	0.003	1.008[a]	0.003
14	0.508	0.277	0.391[a]	0.135	0.217[a]	0.099	0.316[a]	0.126
14 x ln(t)	1.415[b]	0.215	1.552[a]	0.155	1.785[a]	0.226	1.643[a]	0.186
20	0.008[a]	0.012	0.009[a]	0.015	0.008[a]	0.012	0.013[a]	0.02
20 x t	1.050[a]	0.016	1.048[a]	0.017	1.051[a]	0.016	1.041[a]	0.016

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Trigger 5 corresponds to students who are enrolled in college. In the prior EAS-trigger model with no control variables, this trigger produced inconsistent results between configurations, overall indicating a decrease in hazard ratio which was against expectations. In the controlled-trigger model, however, the hazard ratio estimates show the trigger behaving in line with expectations of increased hazard over time. The estimated hazard ratios for trigger 5 are presented in Figure 7.61.

The results show that students who attend college have a significantly lower hazard ratio when first commencing university. This reflects the initial induction period where college students are encouraged to be part of the college community. The hazard ratio increases over time however, with the controlled-trigger model indicating there is no significant difference between the hazard of college and non-college students by around week 16. After week 16, the hazard ratio associated with living in a residential college increases.

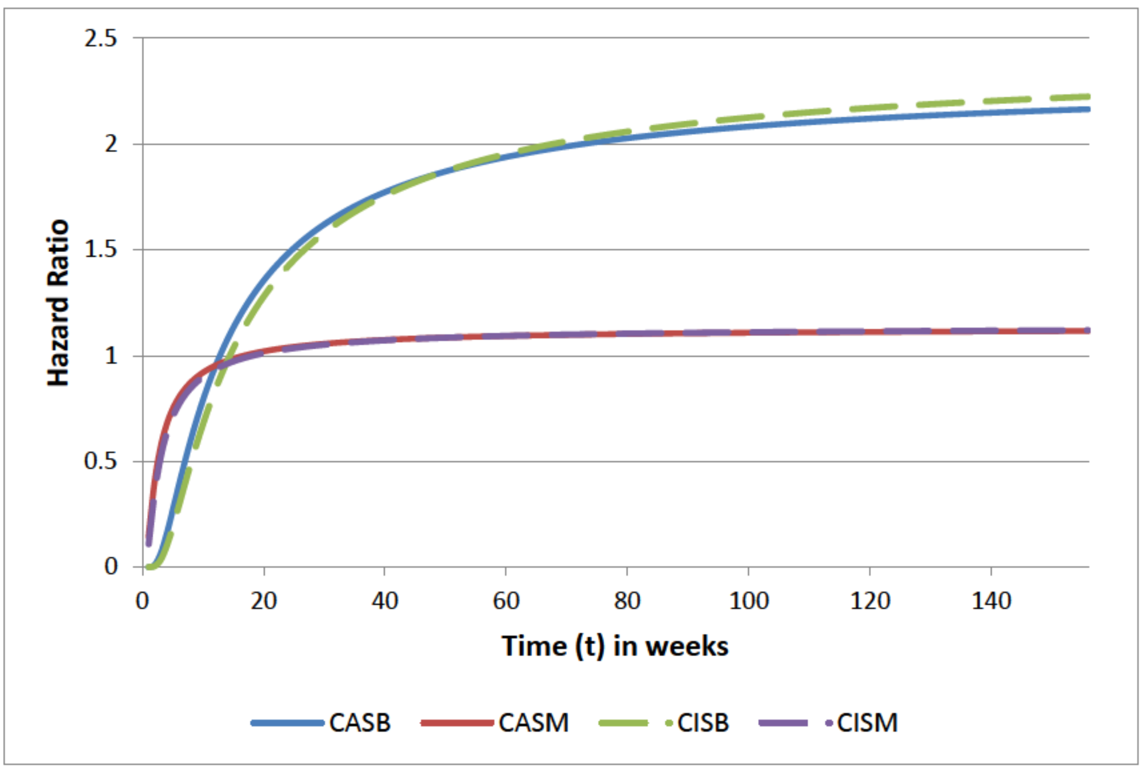


Figure 7.61 – Controlled-trigger model: trigger 5 hazard ratio over time

This time period corresponds to the first teaching period students undertake. As such it is possible to conclude that college students have a lower hazard ratio over their first teaching period at university. After the first teaching period, however, the hazard ratio increases over time. The unweighted models indicate that the hazard ratio for college students is 100% higher by around week 70, corresponding to the end of first teaching period of the second year. The CASM and CISM models show less of an increase in the hazard ratio; however, the trigger still plays an important role in the true estimated hazard. Given that most students are enrolled in a residential college for seven days of the week, this means that most college students would activate this trigger on a daily basis causing the frequency weights for the EAS trigger would also be around seven. As such, the true hazard ratio for college students under the weighted model should be multiplied by seven, which brings the estimated hazards in line with the CASB and CISB models.

Triggers 12, 13 and 14 reflect varying levels of student e-reserve portal inactivity. Trigger 12 corresponds to low inactivity, whereby a student fails to log into the e-reserve portal for between 10 and 20 days. Trigger 13 corresponds to medium e-reserve inactivity, not logging into the e-reserve portal for 21 to 30 days. Trigger 14 corresponds to very high e-reserve inactivity, not logging into the e-reserve portal for 41 days or more. In all cases, it is expected that students who do not log into e-reserve portal should exhibit a higher than normal hazard ratio, corresponding to disengagement from the learning process.

Trigger 12 in the controlled-trigger model provides estimates that are in line with expectations. The estimated hazard ratios for trigger 12 are presented in Figure 7.62 over page. The estimated hazard ratio for students at commencement indicates activating this trigger has no significant difference from not activating the trigger. The four sub-models show varying levels of increasing hazard over time, but after the first year of study, all four sub-models show a significantly higher hazard as a result of low e-reserve inactivity. Furthermore, the hazard ratio continues to increase as students' progress further with their studies.

Trigger 13 shows similar hazard ratio estimates, presented in Figure 7.63 over page. The estimated hazards for trigger 13 are either equal or above 1 at the start of enrolment. This indicates students activating this trigger at the earliest possible time (after 20 days of enrolment with no e-reserve log in), already have a significantly higher hazard of discontinuing. This persists throughout enrolment and increases in hazard over time.

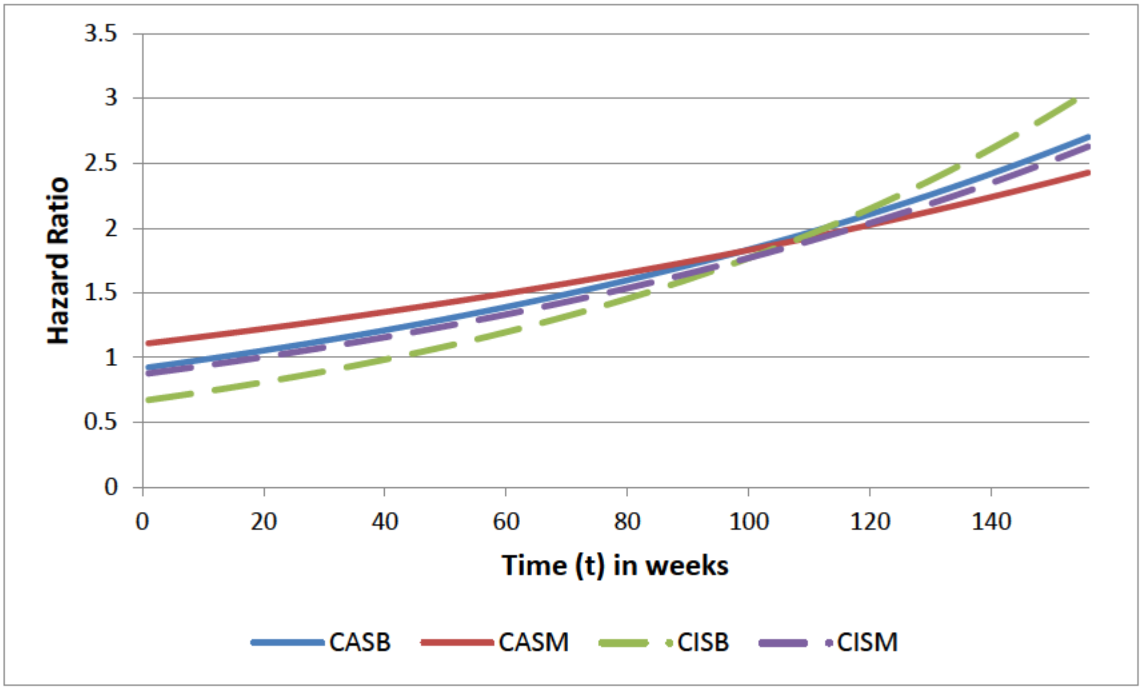


Figure 7.62 – Controlled-trigger model: trigger 12 hazard ratio over time

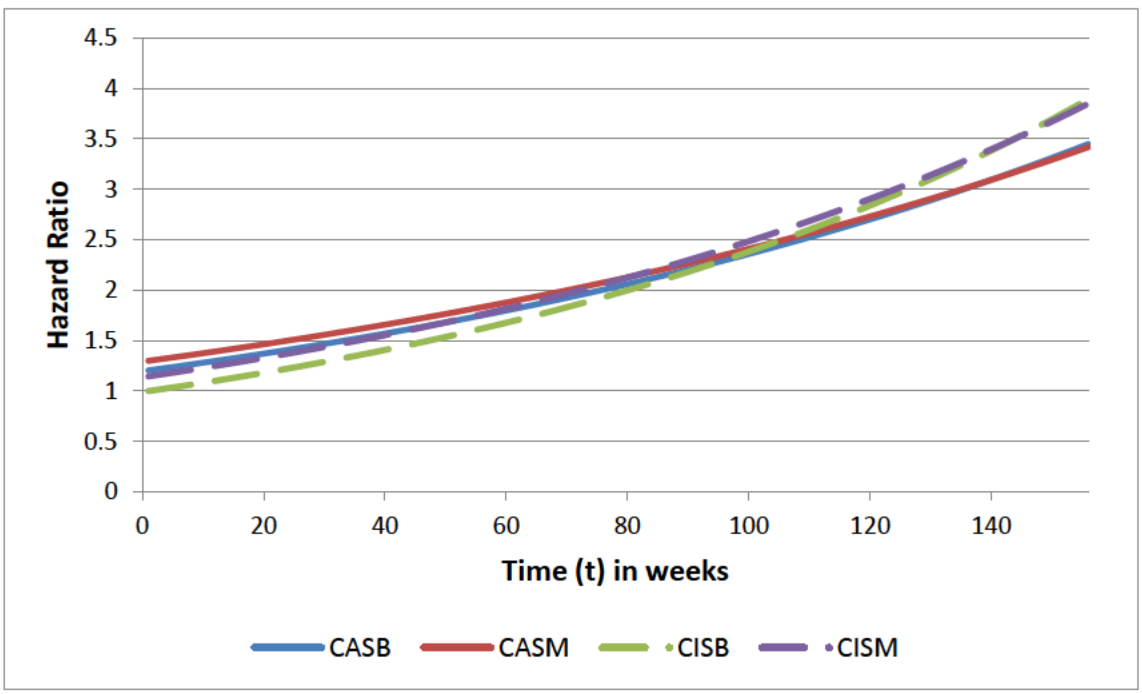


Figure 7.63 – Controlled-trigger model: trigger 13 hazard ratio over time

The results for trigger 14 are similar, however, the model has slight variations on the effects over time due to the delay in activating this trigger. The estimated hazard ratios for trigger 14 are presented in Figure 7.64.

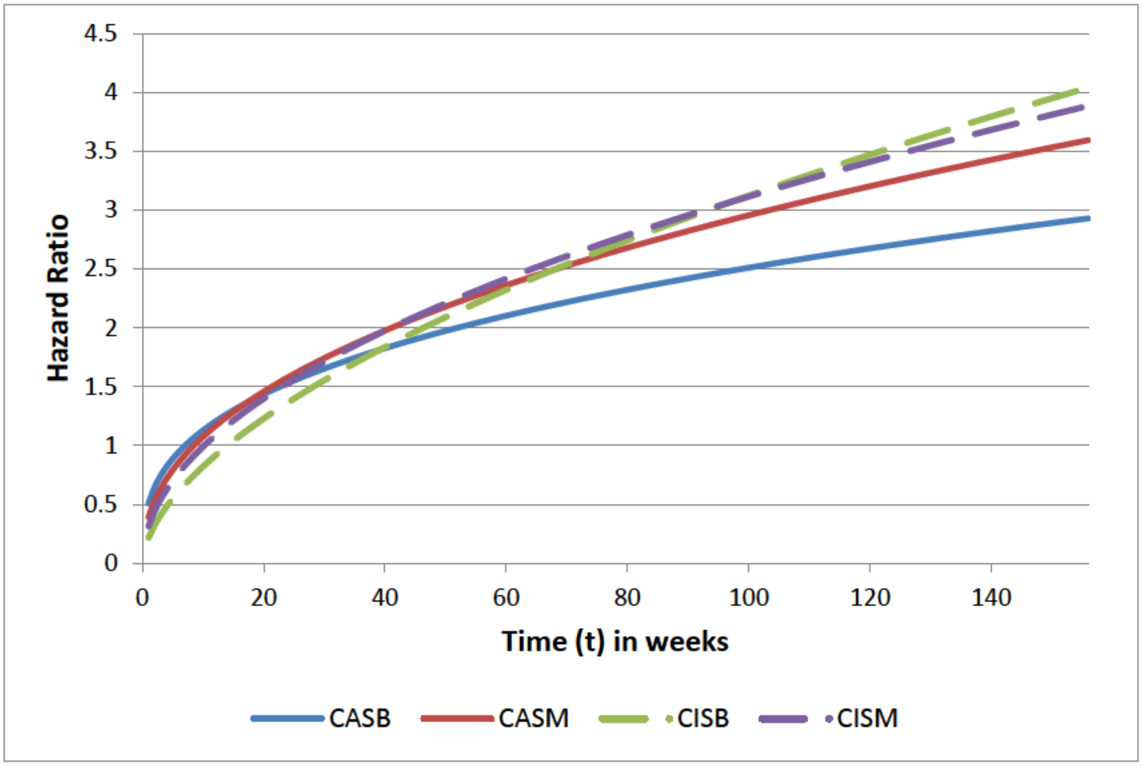


Figure 7.64 – Controlled-trigger model: trigger 14 hazard ratio over time

The estimated hazard ratio starts initially below 1 indicating a decreased hazard associated with the trigger. Given it takes at least 41 days to activate this trigger, the first six weeks of enrolment provide no valid inference. After week 15, all sub-models estimate an increased hazard ratio.

For triggers 12, 13 and 14, the increasing hazard over time captures an important relationship between academic progression and study resources required to assist in progression. Units undertaken throughout a course generally increase in complexity and difficulty over time. As units increase in difficulty and complexity, students have an increased imperative to utilise the e-reserve portal for accessing study material related to their course. Triggers 12, 13 and 14 capture the effect where e-reserve inactivity later in course progression corresponds to an increased hazard of discontinuation. This indicates student e-reserve portal inactivity makes an important contribution to identifying students with increased hazard of discontinuation. This means that e-reserve portal inactivity should be incorporated into EAS design, when control variables are also included.

Trigger 20 corresponds to students who appeared in the high risk category in a previous teaching period. In the prior EAS-trigger model, this trigger showed a significant increase in hazard if activating this trigger past week 91. The results for the controlled-trigger model are consistent with the previous results, with the estimated hazard ratios plotted in Figure 7.65. The trigger shows some variation between sub-models. Students activating this trigger have an increased hazard ratio after week 98 for the CISB model, and up to week 108 in the CISM model. Despite the variation between the models, this trigger serves as a strong identifier of students at risk of discontinuing their enrolment, especially towards the end of their course.

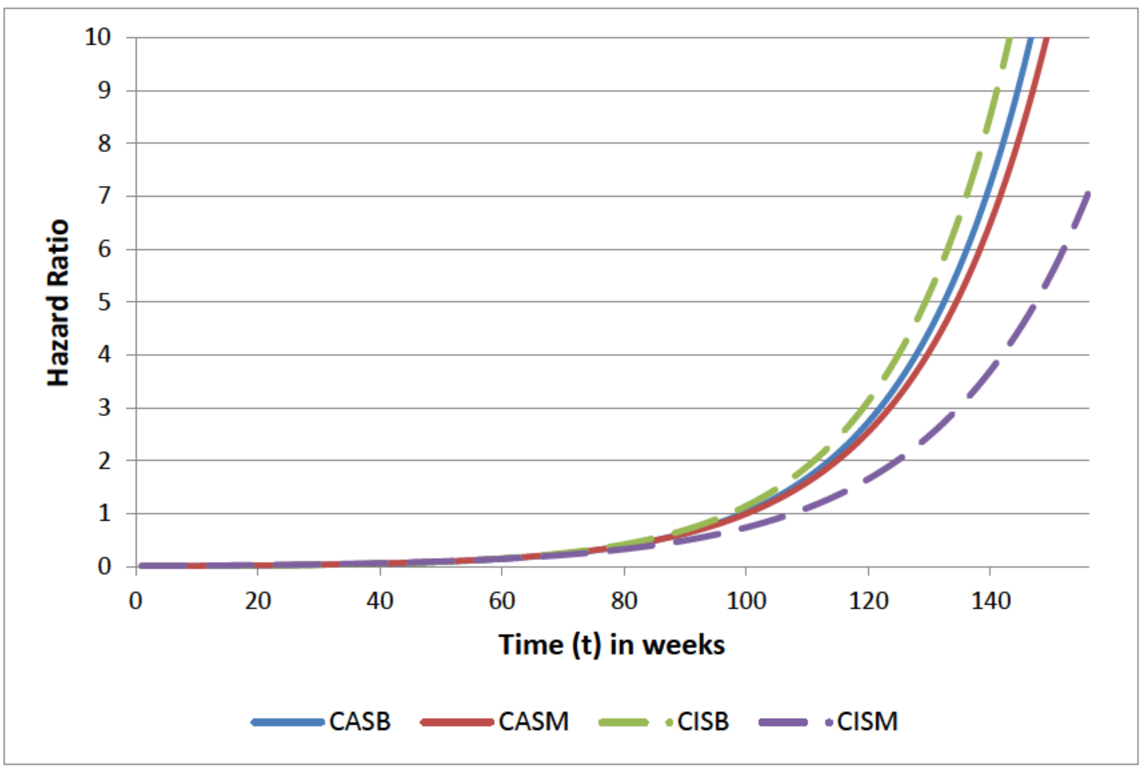


Figure 7.65 – Controlled-trigger model: trigger 20 hazard ratio over time

7.7.9 Triggers with inconsistent effects over time

Triggers with inconsistent effects over time have varying hazard ratios between the configurations of models. Controlling for *demographic*, *institution*, *student performance and workload* variables, six triggers have inconsistent effects between models. These are presented in Table 7.34.

Table 7.34 – Controlled-trigger model: triggers with inconsistent effects over time

	CASB		CASM		CISB		CISM	
Trigger	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
22	0.058[a]	0.011	0.060[a]	0.012	0.055[a]	0.012	0.798	0.517
22 x 1/t	-	-	11.709[a]	8.288	-	-	-	-
22 x ln(t)	-	-	-	-	-	-	0.475[a]	0.099
29	1.085	0.186	1.118	0.188	0.690[b]	0.099	0.722[b]	0.102
29 x t	0.995	0.004	0.995	0.004	-	-	-	-
30	1.190	0.212	1.183	0.211	1.319	0.222	1.332[c]	0.219
30 x 1/t	9.972[b]	10.781	13.106[b]	14.398	-	-	-	-
31	0.990	0.219	0.919	0.224	0.801	0.213	0.874	0.230
31 x 1/t	-	-	22.314[c]	40.476	-	-	-	-
32	1.218	0.646	1.374	0.615	0.428[b]	0.156	1.421	0.721
32 x t	0.980	0.013	0.979[c]	0.012	0.976[a]	0.009	0.974[c]	0.014
34	16.861[a]	15.037	1.757[c]	0.581	1.972[c]	0.767	0.661[c]	0.139
34 x t	-	-	0.978[a]	0.008	-	-	-	-
34 x ln(t)	0.418[a]	0.110	-	-	-	-	-	-

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Trigger 22 corresponds to students with no prior enrolment history at the institution. It is expected that this trigger would capture students underprepared for university study and as such, who have a higher hazard ratio. In the presence of the control variables, the trigger has varying results. Unweighted models 7.5a and 7.b have a constantly lower hazard ratio; weighted models 7.5a and 7.5b decrease in hazard over time. Overall, students activating this trigger have a decreased hazard ratio. This result goes against expectation and may need to be removed from the EAS to prevent student misidentification.

Triggers 29, 30 and 31 correspond to low, medium and very high levels of student portal inactivity. Students activate these triggers when they fail to log into the online student portal for 10 to 20 days, 21 to 30 days and 41 days or more respectively. Trigger 29 shows that there is no significant effect in the CASB and CASM models, while CISB and CSIM models show a consistent decrease in the hazard ratio. Trigger 30 has initially high hazard ratios as shown in Figure 7.66.

The estimated hazard ratios for the CASB and CASM models converge with the CISB and CISM models around week 15. Importantly, CISB and CISM models indicate an increased hazard ratio but is only significant at the 10% level. The inconsistent results for trigger 30 show that medium student portal inactivity is not a strong indicator of students with increased hazard of discontinuing.

Trigger 31 also shows inconsistent results, with only the CASM model having a significant effect. The results clearly show that very high student portal inactivity does not correspond to increased hazard of discontinuing. Given the results from triggers 29, 30 and 31, the use of portal inactivity as triggers for the EAS should be reconsidered. One issue that may be affecting the

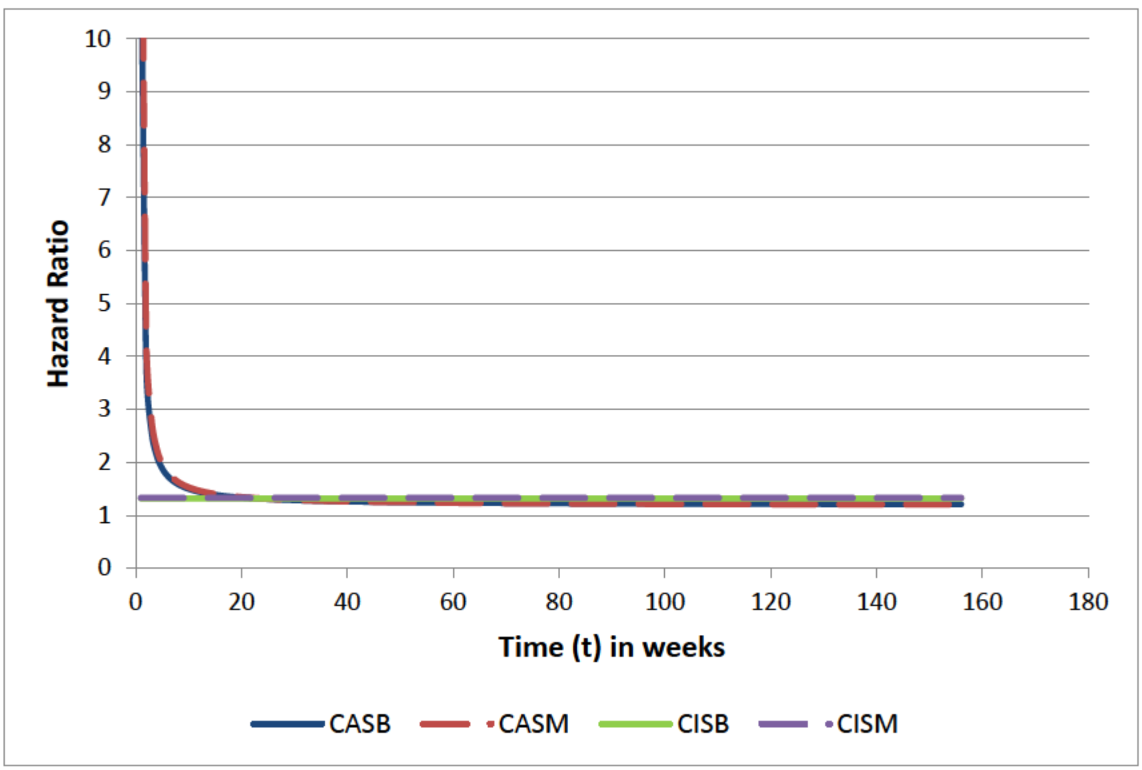


Figure 7.66 – Controlled-trigger model: trigger 30 hazard ratio over time

Triggers' capacity to correctly function is the assumptions made around student portal activity. The portal is the key administrative point, where students can update enrolment, personal details and access results. Additionally, the portal provides a link to the LMS for each unit currently enrolled. The issue is that students can also access the LMS directly without logging in to the student portal. As such, student portal login data may not be correlating to LMS login data. The student portal may not be capturing students at risk of disengaging, whereas the LMS login data could be capturing such data.

Trigger 32 captures students enrolled in a teaching enabling course. The results from Table 7.34 show the CASB model is insignificant; CASM and CISM models have effects only significant at the 10% level; only the CISB model has significant effects at the 1% level, estimating a decreased hazard ratio over time. The inconsistent results associated with this trigger in the presence of control variables indicate that trigger 32 does not capture students with a higher hazard of discontinuing.

Finally, trigger 34 corresponds to students who received a fail incomplete in the previous teaching period. The results for the controlled-trigger model are similar to the prior EAS-trigger model, showing conflicting estimates over time. As such it can be concluded that while theoretically the trigger should be included in the EAS, empirically the results do not support its inclusion. Including trigger 34 within the EAS is likely to cause students to be wrongly identified for support.

7.7.10 Triggers with no significant effects over time

Triggers with no significant effects over time should be excluded from the EAS. The estimated hazard ratios for the controlled-trigger model are presented in Table 7.35 over page. Trigger 1 corresponds to students admitted through alternative entry pathways. The results indicate that there is no change in the students' hazard ratio activating this trigger. As such it should be excluded from the EAS to minimise misidentification of students in need of support.

Trigger 2 corresponds to ATSI status. While ATSI status is captured as part of the control variables, its use as a trigger in this model produces the same results as the prior EAS-trigger model. As such, the trigger for ATSI should be replaced with a variable similar to that of the control variable, which captures time varying effects as well as the interaction effect with the EAS when a student is identified as per the interactions model.

Table 7.35 – Controlled-trigger model: triggers with no significant effects over time

	CASB		CASM		CISB		CISM	
Trigger	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
1	0.389	0.285	0.407	0.281	0.489	0.36	0.524	0.358
2	0.704	0.297	0.817	0.283	0.662	0.306	0.675	0.282
4	1.004	0.399	1.046	0.319	0.633	0.294	0.641	0.248
4 x ln(t)	1.051	0.134	1.042	0.103	1.181	0.166	1.182	0.137
6	0.783	0.243	0.758	0.228	0.983	0.324	0.921	0.291
7	-	-	-	-	-	-	-	-
8	1.052	0.414	1.056	0.393	1.061	0.495	1.059	0.467
9	0.807	0.372	1.015	0.421	0.649	0.384	0.873	0.446
10	1.159	0.594	1.159	0.58	1.022	0.604	1.063	0.613
15	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	-
17	-	-	-	-	-	-	-	-
18	-	-	-	-	-	-	-	-
21	0.528	0.387	0.539	0.31	0.566	0.415	0.494	0.307
23	1.179	0.368	0.933	0.241	1.151	0.379	1.011	0.292

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

Trigger 4 captures students in historically high attrition units. Controlling for demographic, institution and learning environment variables, this variable has changed from decreasing hazards over time in the previous EAS-trigger model to no significant effect in the controlled-trigger model.

Triggers 6 to 10 capture the varying levels of "e-motion" associated with units of study. The e-motion tool may be useful in capturing students' feelings about their unit of study. However, the results show no link between the e-motion tool and changes in the hazards of discontinuing. As such, its inclusion in the EAS may be superfluous.

Trigger 21 activates if the student is an international student. In the EAS-trigger model, this showed a constant decreased hazard which is somewhat in line with the previous models. However, given that the international fee paying variable was captured in the control variables, it is logical that the trigger has no effect. While international students should be captured by the EAS, it is important that the trigger specification corresponds to the changes in hazard over time as demonstrated in the base model.

Finally, trigger 23 captures students who are enrolled in five or more units in a teaching period. As suggested in the EAS-trigger model, this trigger is unlikely to capture students at risk of either disengaging or discontinuing as the students have opted to increase their workload. As such, inclusion of this trigger may actually be identifying students who do not require additional support. If the trigger is to be included, it should be done in light of the students' previous academic performance, suggesting the EAS include interaction effects with the academic record.

7.7.11 Summary

Including *demographic*, *institutional*, *student performance* and *workload* variables as control variables caused many of the triggers estimate hazard ratios in line with expectations. For example, the college student trigger changed to reflect increased hazard over time. This validates its inclusion within the EAS design. An important caveat is that the effect was only identified in the presence of the control variables. This means the current EAS design may not be capturing students genuinely at risk of discontinuing if other demographic, institution and learning environment variables are not included.

From the results of the controlled-trigger model, triggers that correlate to higher hazard ratios should be included in the EAS design. Table 7.36 compares the valid triggers from the EAS-trigger model and the controlled-trigger model.

Table 7.36 – Recommended valid EAS triggers over time

	EAS-trigger model	Controlled-trigger Model
Trigger	Valid Time Frame	Valid Time Frame
3	Weeks 1 to 4	-
5	-	All
11	Weeks 1 to 9	All
12	-	All
13	Weeks 1 to 11	All
14	-	All
19	Weeks 1 to 28	All
20	Weeks 91 +	Week 98 onwards
25	Weeks 1 to 89	All
26	Weeks 1 to 110	Weeks 1 to 52
28	Weeks 1 to 19	Weeks 1 to 26

The controlled-trigger model captures more valid triggers than the EAS-trigger model. Additionally, the valid time frame indicates that 7 of the 10 valid triggers are suitable for use over the length of the students' enrolment. Overall, the results of the controlled-trigger model show for the EAS to function optimally, *demographic*, *institutional*, *student performance* and *workload* variables should be included in the EAS design. Despite the criticisms of the individual triggers used in the identification process, overall the EAS appears to identify students with increased hazard (short-run base model) and have an enduring benefit of reducing hazard overall (long-run base model). To fully test EAS effectiveness, a treatment effects approach is needed to prove causality.

7.8 Survival treatment effects

7.8.1 Model specification

Previous models in this chapter focused on estimating the hazard function of students using survival analysis. It was found that the EAS system had an effect on improving the length of enrolment and reducing the hazard ratio of identified students. However, inferring an effect is not the same as concluding the EAS causes increased student retention. Inferring causation requires more rigorous statistical methods. One such method is treatment effects modelling, where the impact of a particular treatment is assessed using observational data.

Treatment effects models were designed for environments where experimental control groups are not possible due to ethical, experimental or other reasons. The model works by separating observations into two groups based on the treatment variable. The first group is a treatment group; the second is an effective control group which acts as a proxy for an experimental control group. In this case, students identified by the EAS are the treatment group and those not identified by the EAS are the effective control. The model estimates the effect on the dependent variable for the two groups, imputing potential alternative outcomes using data from the other group. In this case, using the identified student's data, the dependent variable for the non-identified group is estimated as if it were identified. The counterfactual is also estimated. That is, using the non-identified student's data, the dependent variable is estimated for the identified group as if it were not identified. This then allows the estimations to be compared to test for causal effects associated with the treatment regime. This makes it an ideal model for evaluating the treatment effects of the EAS, where it is not possible to exclude non-identified students from receiving support during their studies in some form.

When estimating treatment effects models, there are three key parameters of interest. The first is the potential outcome means (POM). The POM provides an estimate of the potential outcomes should the identified students not have been identified. In the context of the EAS estimation with length of enrolment in weeks as the dependent variable, this can be interpreted as the average number of weeks identified students would be enrolled if they were not identified. The second outcome of interest is the average treatment effects (ATE). The ATE is provides an estimate of the treatment effect in the whole population. For example, what is the average effect of being identified by the EAS in terms of length of enrolment? The third and final outcome is the average treatment effect on the treated (ATET). The ATET captures the effect of the treatment only for those who received the treatment. In this case, it captures the effect of the EAS on the subgroup of identified students only. These three measures form the basis of treatment effects modelling.

Using STATA version 14, a survival treatment effects model was estimated. This allows consistent estimation of effect using temporal data used throughout chapter 7. One limitation of this model was for the model to converge, GPA was substituted for the grade distribution. While using GPA reduces the level of detail within the model, it acts as a reasonable proxy for the grades of a student. In this case a 7-point GPA was calculated, where a *fail* or *fail-incomplete* = 0, *pass* = 4, *credit* = 5, *distinction* = 6 and *high distinction* = 7. The other variable, *length of enrolment*, is a total length of time, including holidays and non-teaching weeks. As such any estimated effect is a measure of total additional time enrolled at the institution, not just teaching weeks.

Finally, there are two different approaches to using the survival treatment effects model. The first is to estimate the overall effect the EAS has on the student cohort. This approach divides the students into the identified/not identified sub groups used previously. The second approach is to estimate the effects associated with different treatment levels. This is the approach taken in Chapter 5. By separating students into varying treatment levels, this approach captures the effect of being identified by the EAS multiple times.

7.8.2 Overall treatment effects

The survival treatment effect model uses the regression adjustment method to estimate the effect of the EAS on the time to discontinuation. The model includes ancillary variables, the demographic, institution and learning environment variables used in previous models with the exception of grade distribution. The outcome variable is assumed to have a Weibull distribution.

The model does not provide estimated coefficients as to the contribution of the individual ancillary variables. The results of the ATE estimation are presented in Table 7.37.

Table 7.37 – Survival treatment effects results: Overall ATE

	Coefficient	Robust Std. Err.	z	P>z	[95% Conf.	Interval]
ATE	13.512[a]	1.382	9.78	0	10.8027	16.2205
POM	67.090[a]	1.283	52.28	0	64.575	69.6056

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The POM indicates the estimated average number of weeks for which identified students would have been enrolled for if they were not identified by the EAS. In this case, the model predicts identified students would have been enrolled for 67 weeks if not identified. This estimate is over a year long, indicating the average length of enrolment is a year. The average treatment effect (ATE) is estimated to be 13.5 weeks, a significant result at the 1% level. This indicates that, taking into account the ancillary demographic, institution and learning environment variables, identification by the EAS causes students to be enrolled for an average additional 13.5 weeks. With 95% confidence, the true impact of the EAS causes students to be enrolled for between 10.8 weeks and 16.2 weeks.

The results for the ATET estimation are presented in Table 7.38

Table 7.38 - Survival treatment effects results: Overall ATET

	Coefficient	Robust Std. Err.	z	P>z	[95% Conf.	Interval]
ATET	13.777[a]	1.431	9.63	0	10.9715	16.5815
POM	67.634[a]	1.340	50.46	0	65.0071	70.2607

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The ATET results focus only on the group of students identified by the EAS. The ATET results are similar to the ATE result, with the results significant at the 1% level. The potential outcome means for the identified group only is 67.6 weeks and the ATET estimate is 13.7 weeks. Both of these estimates are slightly higher than the ATE results, however given the estimates capture overlapping confidence intervals, there is no significant difference between the ATE and ATET results.

This leads to an important finding: the EAS system is causing students to be enrolled for longer. This fundamental empirical result is important in establishing the efficacy of the system at the implementing institution. To date, there is no other literature with respect to EAS that can make such a claim with this level of statistical rigour.

7.8.3 Varying levels of severity

Establishing that identification and the resulting interaction with student support services has an effect on retaining students, this section analyses the varying levels of identification. To refresh, the EAS identifies the top 200 students on a daily basis who are deemed at risk of disengagement. This list of students then forms the basis for targeted support from the student support team. In many instances, a student will be identified more than once, with some students identified on more than 20 days during their enrolment. As such, students are divided up into five categories of severity, in line with chapter 5. The frequency table is presented below in Table 7.39.

Table 7.39 – Survival treatment effects: frequencies of treatment variable

Severity Level	Times Identified	Frequency	Per cent (%)
0	0	4,830	29.96
1	1 – 4	5,582	34.62
2	5 – 9	2,829	17.55
3	10 – 19	1,858	11.52
4	20+	1,025	6.36
	Total	16,124	100

The frequency table shows that approximately 30 per cent of students are never identified by the EAS. 34.6 percent of students were identified between one to four times, with only 6.36 per cent of students identified more than 20 times. This model compares severity levels 1 to 4 with the base treatment group 0.

Table 7.40 – Survival treatment effects results: severity level 1

Severity Level 1	Estimate	Coefficient	Robust Std. Error	Z-Score	P-Value	Lower 95%	Upper 95%
ATE	ATE	1.201	1.520	0.79	0.43	-1.7778	4.17884
	POM	64.618[a]	1.200	53.86	0	62.2666	66.9697
ATET	ATET	0.964	1.588	0.61	0.544	-2.1489	4.07709
	POM	65.205[a]	1.281	50.91	0	62.6952	67.7156

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The severity level 1 results compares those students identified by the EAS between 1 to 4 times to students not identified. The predicted outcome means of the ATE and ATET models are 64.6 and 65.2 weeks respectively. The estimates are significant at the 1% level. These estimates are similar to the overall model estimated in the previous section, capturing a level of stability in the estimates of the POM.

At severity level 1, the ATE and ATET estimates are not statistically significant. This means there is no significant effect on the length of the student's enrolment given they were identified by the EAS one to four times. Under the causal inferences, this indicates that the initial identification process and resulting student support has no effect on the outcomes as measured in the length of enrolment. This is an important finding as it indicates that there may be issues associated with the EAS either identifying students who do not need support, or the support provided to students identified one to four times is insufficient to have an impact on student outcomes. However, given that the overall model indicated an effect of increasing student's enrolment by 13.5 weeks, it is more likely that the EAS effect at this level has the effect of normalising student outcomes. This means students identified by the EAS who were at risk of either disengaging or discontinuing, modify their behaviour such that their expected outcomes are now no different to the control group.

Severity level 2 results are presented in Table 7.41. As with severity level 1, the POM estimates show that students are expected to be enrolled for between 64.6 and 65.8 weeks for the ATE and ATET models respectively. The overlapping nature of these confidence intervals with both the overall model and severity level 1 results further supports the stability of the POM estimates.

Table 7.41 – Survival treatment effects results: severity level 2

Severity Level 2	Estimate	Coefficient	Robust Std. Error	Z-Score	P-Value	Lower 95%	Upper 95%
ATE	ATE	13.985[a]	1.694	8.26	0	10.6648	17.3047
	POM	64.636[a]	1.170	55.27	0	62.3438	66.9283
ATET	ATET	14.377[a]	1.774	8.1	0	10.9001	17.8543
	POM	65.825[a]	1.289	51.05	0	63.2982	68.3525

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The ATE estimate, which captures the population level effect of the EAS, estimates a significant impact of increasing enrolment by 14 weeks when a student receives severity level 2. This corresponds to a student being identified five to nine times by the EAS. Within the identified group of students, the ATET estimate is slightly higher at 14.3 weeks. Both estimates are significant at the 1% level. This result shows that the EAS is having an impact on students who have been identified 5 or more times. This raises important questions about student support associated with the EAS. Is the support offered to students constant throughout their enrolment, or are students who are identified five or more times offered increased support that does not become available until a higher number of identifications are recorded? The evidence from the severity levels 3 and 4 answer this further.

Increasing the treatment variable further, students identified between 10 and 19 times are compared to not identified students in Table 7.42

Table 7.42 – Survival treatment effects result: severity level 3

Severity Level 3	Estimate	Coefficient	Robust Std. Error	Z-Score	P-Value	Lower 95%	Upper 95%
ATE	ATE	41.078[a]	1.554	26.43	0	38.0306	44.1239
	POM	64.592[a]	1.132	57.06	0	62.3731	66.8105
ATET	ATET	43.366[a]	1.590	27.29	0	40.2509	46.4801
	POM	66.270[a]	1.250	53.04	0	63.8207	68.7181

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The model for treatment level 3 provides significant results at the 1% level for all the POM, ATE and ATET estimates. The estimated POM for the ATE and ATET models are still around the 65-week mark, confirming the results from the previous models.

The ATE and ATET estimates are significantly greater than the previous treatment levels. At the population level, the ATE estimates show that being identified by the EAS between 10 and 19 times causes students to be enrolled for an additional 41 weeks. The ATET estimate is similarly large, with the average effect on treated students being an additional 43.4 weeks of enrolment. This is a massive effect, and given that 1,858 students fell into this treatment category from the sample data, this is a huge benefit to students and the institution. This provides the most important evidence so far with respect to the system's efficacy. For the student, adding over three quarters of a year to an enrolment allows time for significant progress on the course to be made, capturing two additional teaching periods at least. From the institutions perspective, this is around two teaching periods of tuition fees it otherwise wouldn't have had.

When analysing severity level 4, similar results are observed. Presented in Table 7.43, the results show the POM's for ATE and ATET to be 64.6 and 67.7 weeks respectively.

Table 7.43 – Survival treatment effects result: severity level 4

Severity Level 4	Estimate	Coefficient	Robust Std. Error	Z-Score	P-Value	Lower 95%	Upper 95%
ATE	ATE	44.374[a]	1.762	25.18	0	40.9201	47.8275
	POM	64.596[a]	1.151	56.13	0	62.3399	66.8511
ATET	ATET	47.495[a]	1.973	24.07	0	43.628	51.3619
	POM	67.655[a]	1.390	48.7	0	64.9311	70.3771

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The estimated ATE is 44.4 weeks and ATET of 47.5 weeks. All are significant at the 1% level. This means on average, a student identified more than 20 times by the EAS will be enrolled for 44.4 weeks longer than a student who has not been identified. Within the group of students with this level of treatment, the average effect was an increased length of enrolment of 47.5 weeks.

7.8.4 Summary

A significant causal relationship between identification by the EAS and the length of the students' enrolment was established. This important finding provides evidence that the system is having a positive effect on student outcomes. While no significant effect was found at the first treatment level, the more intense treatments resulted in better student outcomes. It is clear that for the system to have long-term impacts on student outcomes, students need to be identified five or more times. While it is not possible to see inside the black box of student support post identification, the results do indicate that students repeatedly identified by the system attain support that affects their outcomes. This may reflect varying levels of support offered by student services when a student is identified multiple times.

One limitation on this finding is that there is no benchmark data enabling conclusions on the efficacy of the EAS. It is not possible to conclude the EAS system is better than the previous system of general student support. However, the results presented here do form a benchmark, against which future analyses can be measured. Future enhancements and improvements to the EAS should be detected using the same models, allowing for performance measurement to be conducted.

Overall, students who are identified by the system gain an additional 13.5 weeks of enrolment. This constitutes an extra teaching period for students to improve their outcomes and progress towards completing their degree. It also represents an additional teaching period of revenue for the institution. Students identified more than 10 times benefited the most from the system, indicating that high levels of identification are correlated to increased support.

7.9 Chapter summary

This chapter introduced temporal effects to accurately model student outcomes. Using survival analysis in varying functional forms, the factors that affect student retention are identified. Furthermore, the temporal estimates allow the factors to change over time, with some variables being more significant early in a student’s enrolment, while having little effect later on. This flexibility allowed a detailed treatment of the EAS data to clearly identify how EAS relate to student retention.

The base model used varying specification for the EAS to test variables and their effect on student retention. The hazard ratio estimates showed a significant relationship between all three demographic variables, gender, age and ATSI status. Institution variables had differing levels of significance, however, with international students having a significantly lower hazard ratio than domestic HELP students most of the time. The hazard ratio for international students did peak around the two years of enrolment mark, indicating this is an area on which to focus support for international students. The results show that bachelor students of graduate entry and honours had a significantly lower hazard ratio, in line with theory. Variations between schools were found, where comparing to the base school, three schools consistently had statistically higher hazard ratios than the base school. Two schools had time varying effects, where the hazard of discontinuing while enrolled within these schools increased over time. Finally grade distributions showed the negative grades have significantly higher hazard ratios, while positive grades reduced students' hazard ratio. Periods of inactivity also increased students' hazard ratios. Overall, the base model showed interesting effects on the EAS system which reflected the underlying model designs. The base short-run model, showed that in any given week a student identified by the EAS had a significantly higher hazard ratio. This reflects the EAS system identifying students at risk of discontinuing. The base long-run model indicated students who had been identified by the system had a significantly lower hazard ratio than students not identified by the EAS. This is the first evidence in the temporal model as to the effects of the EAS system.

The conditional model showed overall, demographic, institution and learning environment variables affected identified students more than the non-identified students. Females identified by the EAS were at a higher hazard ratio than male also identified by the EAS. However, this was not the case in the group of students not identified by the EAS. Within the non-identified category, ATSI students had a significantly difference hazard ratio from non-ATSI students, Age, graduate entry and honours were constant across both models. There was significant variation within the schools section where, within the identified students subgroup, schools 1 – 5 all had statistically higher hazard ratios compared to the base school. Withdrawing from a unit had a significant increase in the hazard ratio for students identified by the EAS, while withdrawn early had a significant increase in hazard if the student was not identified by the EAS. A critical finding was the willingness to engage especially from students who had prior studies. This result suggests that students with prior studies have a lower hazard ratio than other students identified by the EAS, attributable to an increased willingness to engage with support services.

The interactions model used an interactions effect approach with a short run and long run model. In the short run, females were more likely to discontinue than their male counterparts. If, however, females were also identified by the EAS, the hazard ratio was significantly lowered. There was no long run difference between male and females upon interacting with the EAS. The long run model showed that ATSI students had a lower hazard ratio over time when interacting with the EAS. This is a strong indication that ATSI students benefit in the long run from the EAS interaction. School 2 had varying hazard effects in the long run model, where students identified by the EAS initially had a hazard ratio above the not identified group. This changed over time where eventually the identified group had a reduction in hazard. School 4 saw a decrease in the hazard ratio in the short run. Positive grades all showed a decrease in the hazard ratio both in the short run and long run. However, identification by the EAS did indicate an increase in the hazard ratio of students attaining positive grades. Overall, the interactions model was in line with expected results.

The EAS-trigger model introduced the triggers used in the identification process of the EAS. This allowed aspects of the inner workings of the EAS to be analysed in effect on the students' hazard ratio. The results were varying, showing a wide range of constant, positive, negative and inconsistent and insignificant relationships. Evaluating the triggers showed that 8 triggers reflected increased hazard ratios; however these were not consistent over time. Trigger 3 which captures students enrolled in a high-attrition unit remained valid only for the first four weeks of enrolment. Many of the triggers showed inconsistent or insignificant relationships to the hazard ratio.

The controlled-trigger combined the base and EAS-trigger models to allow estimation of the effects of the EAS triggers while controlling for *demographic*, *institutional, student performance* and *workload* variables. The models consisted of 89 variables or more, capturing the complex interplay of hazard ratios over time. The model revealed that demographic effects remain constant in both the short run and long run. This validates their inclusion within the model. Furthermore, this may suggest a fuller range of demographic and background variables need to be tested in future studies should more data become available. With institutional variables, the results remained relatively constant from the base model. Schools 3, 4 and 5 had significantly higher hazard ratio across the models, schools 2 and 6 continued to have varying hazards over time. This indicates scope for detailed analysis within each school to identify the contributing factors to this variation. Negative grades were stable across models, estimating an increase in the hazard ratio. Ten of the EAS triggers showed statistically significant effects in line with expectations. The results for the controlled-trigger model highlight two key issues. The EAS functions better with the inclusion of control *demographic*, *institutional*, *student performance* and *workload* variables. Secondly, the current EAS model may be over specified leading to incorrect student identification. A more parsimonious approach may be recommended, supported by the empirical results presented in this study.

Finally, the survival treatment effects model provided causal analysis of the survival data using treatment effects modelling. This established that overall, the EAS extends the students' enrolment by 13.5 weeks, enough time for a full teaching period. Analysing the level of treatment, the majority of this effect lies with students who have been identified by the system 10 or more times. These students, once identified, will continue to be enrolled for over an extra 40 weeks at least. This is a significant finding, showing not only the causal relationship, but the magnitude of the benefit the EAS has.

In summary, based on survival analysis of time to discontinuing, the EAS has a positive effect on student outcomes. It extends into a causal model, where the EAS is sufficient to identify students at risk of disengaging, and assist them to better outcomes. Like a match striking the fire, the EAS is the source of ignition for students interacting with the support services available at university.

Chapter 8 - Financial Implications of an Early Alert System

8.1 Introduction

An increased number of universities are implementing EAS, and increased numbers of vendors are offering systems. Proactively identifying students in need of support comes at a cost in terms of implementing and maintaining such systems. For administrators at institutions implementing such systems, hard questions about financial resources need to be asked. How much should an institution spend to attain an improvement in student outcomes? Do early alert systems have a positive return on investment and by how much? If changes are made to the system, can the changes be quantified?

This chapter analyses the data set to value the effect of the EAS using a financial metric: student tuition fees. From the intuition's perspective, these tuition fees can be termed as revenue. The interchangeability of these terms depend on perspective; however, throughout this chapter the term 'revenue' is used. Several revenue models are tested using a causal treatment effects model. The first model analyses the financial implications of student retention rates, estimating the overall cost of students discontinuing. The second model estimates the overall effect EAS had on revenue from students. The third model analyses the difference in revenue for continuing and completing students versus discontinuing students, under the conditions of EAS identification. The fourth model analyses the variation in revenue between schools within the university. The final model concludes by looking at the revenue effects associated with the timing of identification.

Research into the financial implications and effectiveness of EAS is limited due to the field still developing and maturing. As part of a dissertation, Simons (2011) conducted a survey of 529 four-year higher education providers in the United States. A key research question of the study focused on the effectiveness of early alert programs: how did institutions measure effectiveness and the overall impact of the program on students (Simons, 2011, p. 88)? Two key measures of overall retention and between teaching periods persistence were the most frequent responses. In evaluating the program effectiveness,

> of the almost 40% that noted retention as the ultimate goal, most did not clarify a specific program outcome that precipitated retention. It is nearly impossible to link student retention in general terms directly back to a service area (Simons, 2011, pp. 116 - 117).

While some time has passed since this survey was conducted, the fact still holds true that there has been little reflective assessment on program effectiveness.

Arnold and Pistilli (2012) estimated the benefits of the Course Signals program at Purdue. It was concluded that students taking Signals courses improved graduation rates by 21 per cent. This study, however, was criticised by Caufield (2013b) for not making it clear if the number of courses were controlled for. As such, it was not possible to disaggregate the effects of "students taking more Course Signals courses because they persist, … [compared to] persisting because they are taking more Signals courses" (Caufield, 2013a).

Marrington et al. (2010) analysed the benefits of an in-house program for first year students at Queensland University of Technology. The Student Success Program (SSP) monitored first year student data identifying students at risk of attrition, allowing targeted interventions to take place. A critical part of this study was to take the estimated effects of the program and to turn this into tangible financial benefits of the program. Using EFTSL estimates of student tuition fees paid, around $1,740,000 of student tuition fees were retained, taking into account program costs. It was argued that this was positive evidence for the economic case of EAS.

The main issue with EAS evaluation is establishing a causal link between the system and improved retention outcomes. It is fundamental to all institutions implementing EAS that they can prove the system is causing an improvement in outcomes. In chapter 7, a causal link was established where being identified by the case study EAS caused students length of enrolment to increase. The objective of this section is to then quantify financially the additional benefit derived from this improvement in retention.

8.2 Empirical analysis

8.2.1 Calculating revenue

Revenue is generally calculated as the price per unit of a good or service, multiplied by the quantity of units. Mathematically, this is represented as follows:

$$Revenue = Price\ Per\ Unit\ \times Quantity\ of\ Units \tag{8.1}$$

An ideal situation is that tuition fees for each student are provided as part of the data set. However, in lieu of this data not being presented, it is possible to create a close approximation on the student fees paid by each student, using the fee schedule of the institution and matching this with the student's academic record. The fee schedule comes in three distinct brackets which were used in previous statistical models. These are domestic Higher Education Loan Program (HELP) students, domestic fee-paying students and international students. Intersecting with domestic student fee categories, university courses are assigned "bands", an Australian government regulated fee structure for study undertaken in a given area. The domestic student fee schedule is provided in Table 8.1.

Table 8.1 - Domestic Student Fee Schedule

	Domestic HELP	Domestic Fee-Paying
	Per 6 CP	Per 6 CP
Band 1	$755.00	$679.50
Band 2	$1,076.00	$968.40
Band 3	$1,260.00	$1,134.00
Band 4	$755.00	$679.50
Band 5	$1,260.00	$1,134.00
Band 6	$1,076.00	$968.40
Band 7	$1,076.00	$968.40

International student fees are set independently from the regulated pricing scheme and charged on an annual basis. A full year of study typically consists of eight units of study, and as such, the annual fee is divided by eight, proving the estimated cost per unit of study for an international student. The cost per unit of study was imputed based on the international student fee schedule for 2013 (University of New England, 2013) and matched with the corresponding bands of study for domestic students. The fee schedule used in this study links schools to fee category, presented in Table 8.2. The revenue generated from each student is estimated by multiplying the student's fee category for school of enrolment, with the number of units undertaken. This excludes units where a *withdrawn early* or *other* grade was attained. The "other" grade option captures a variety of administrative grades. As such some students may have actually been charged fees for units with these grades. This means that the revenue estimates are likely to be biased to provide an underestimate of the fees actually paid by a particular student.

Table 8.2 - Fee Schedule by School

School	Band	Domestic Help	Domestic Upfront	International Fee
1	Band 1	$755	$680	$2,228
2	Band 1	$755	$680	$2,228
3	Band 2	$1,076	$968	$2,621
4	Band 6	$1,076	$968	$2,621
5	Band 3	$1,260	$1,134	$2,231
6	Band 4	$755	$680	$2,228
7	Band 3	$1,260	$1,134	$2,621
8	Band 2	$1,076	$968	$2,296
9	Band 1	$755	$680	$2,228
Base	Band 5	$1,260	$1,134	$1,966

This allows for a conservative estimation of the financial effect of the EAS. Another limitation on this method is that it fails to capture the effect on fees when a student undertakes a double degree in different schools with different bands. The model categorises students in the school where the majority of study was undertaken. This is another source of variation between the estimated tuition fees and the actual tuition fees. However, it can be argued that on aggregate across the large data set, the tuition fees associated with double degrees will balance out.

8.2.2 Treatment effects modelling

Treatment effects models refer to a family of statistical models which allow causal inferences to be made using observational data. The models were originally designed for use in medicine, where ethical reasons and study design limitations prevented the use of control groups. The model is applicable for the evaluation of EAS, where it is difficult to have a true control group of students with limited, controlled or no access to student support.

There exist many different treatment effects models to suit varying situations. Propensity score matching (PSM) is just one of the models, with frequent use in economic applications. The propensity score refers to the likelihood of being in the treated or untreated groups based on observed explanatory variables. Bryson (2002, p. 22) used PSM to test if being a union member caused employees to have higher wages, finding that this only occurred in specific cases. Brand and Halaby (2006, p. 768) used a combination of regression and PSM approaches to analyse the effects of elite college attendance and career outcomes, with mixed results when analysing wage premiums. Caliendo, Hujer, and Thomsen (2008, p. 412) used PSM to measure the effects of job creation schemes in Germany, with sub-group analysis revealing that only long-term unemployed women from East Germany have a significant benefit from the programs.

As outlined in Chapter 3, PSM functions on matching observations from the treatment group to the non-treatment group to develop the counterfactual analysis required in causal analysis. A contextualised diagram depicting the matching process is presented in Figure 8.1. A student identified by the EAS is matched with a similar student with exactly the same gender (M/F) and school (Sx). Comparing the two students, the treatment effects model imputes how much the identified student would pay in tuition fees based on not being identified, estimated from the non-identified students outcome. The opposite applies, the non-identified students' tuition fees are imputed based on being identified by the EAS. The imputed results are then compared, allowing an estimation of the treatment effect. The method further allows matching on multiple Nearest-Neighbours (NN), which increases the precision of effect estimation (Caliendo and Kopeinig, 2008, p. 45).

Two characteristics of the matching process are: first, some students will be excluded due to no suitable match being found. This results in varying sample sizes for models. Second, excluding some students from the sample means some explanatory variables may be excluded due to collinearity. While these characteristics of the process are not ideal, the estimated effects remain robust due to the model focusing on the overall average effect of the treatment, and not the individual effect of any one explanatory variable.

Caliendo and Kopeinig (2008) outlined the practical considerations of implementing PSM methods, including selection of the matching algorithm. Four matching algorithms are commonly used to pair observations for analysis. Firstly, NN algorithm identifies the nearest neighbour based on a distance function measure of propensity score.

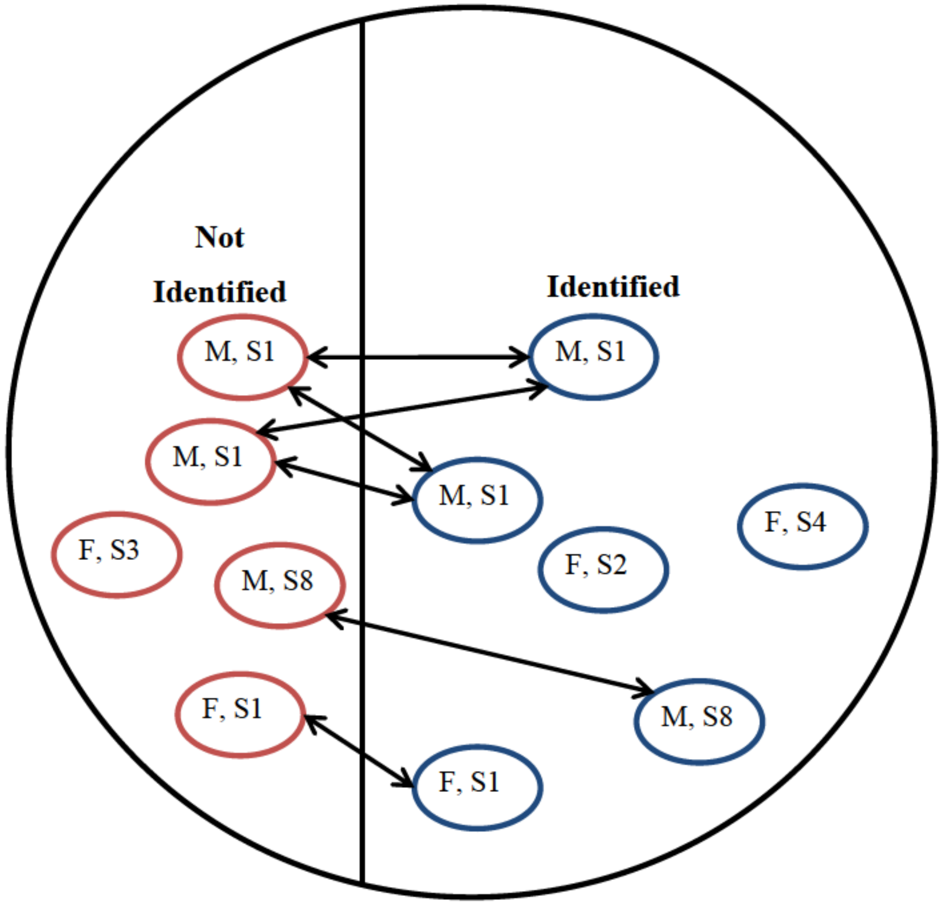


Figure 8.1 – Diagram of treatment effects nearest neighbour matching

The second algorithm, calliper and radius, limits the distance to the nearest neighbour to ensure that matches representative of the alternative outcomes of treatment. Third, stratification and interval matching allows analysis within sub-groups, comparing mean differences in outcomes between treated and control observations. Finally, kernel and local linear matching are "non-parametric estimators that use weighted averages on (nearly) all, depending on the choice of kernel function, individuals in the control group to construct the counterfactual outcome" (Caliendo and Kopeinig, 2008, p. 43).

For the purposes of this analysis, the nearest neighbour method was selected. This algorithm allows observations to be exactly matched based on theoretical underpinnings of the model estimated. The models will continue to use *demographic*, *institutional and workload* variables. For determining financial effects in this chapter, it is important to exactly match students in the treatment group to the control group, on *gender, age, ATSI status, domestic fee, international fee, course type, school of study* and *workload.* Previous chapters have shown that the demographic variables have a significant effect on retention and therefore need to be included in the model. The fee type of the students is important as the fee schedule varies between these groups, while course type and school of study also reflect varying fee structures. Finally, workload is important to include to ensure that students undertaking similar number of units each teaching period are compared. Additionally, students are matched “with replacement” with a minimum of four matches per student. The result is the “average quality of matching will increase and the bias will decrease” (Caliendo and Kopeinig, 2008, p. 41).

This study uses five models to estimate varying effects on revenue.

- Model 8.1 estimates the effects of discontinuing study.
- Model 8.2 estimates the effects of being identified by the EAS.
- Model 8.3 estimates the effects of discontinuing within the EAS sub-groups.
- Model 8.4 estimates the effects of being identified by the EAS within schools.
- Model 8.5 estimates the effects of being identified by the EAS at times of identification.

Model 8.1 uses the students’ enrolment status as the treatment variable, with revenue as the dependent variable, estimating the financial effects associated with discontinuing. By comparing discontinued students with all other students, it is possible to estimate the cost to the institution when a student discontinues. Comparing discontinued students to completed students estimates the cost to the institution of a discontinuing student not completing their qualification. These two measures indicate the magnitude of financial loss associated when students decide to discontinue. The remaining models use EAS identification as the treatment variable, estimating the effect of the EAS on revenue under different settings.

Diagrammatically, the relationship between revenue and the discontinuation is depicted in Figure 8.2, while the relationship between revenue and the EAS is depicted in Figure 8.3. A major difference from previous chapters is the removal of student grades. When estimating the effect of a given treatment, it is important not to include variables which can be affected by the treatment itself (Caliendo and Kopeinig, 2008, p. 38). It is plausible and expected that student performance will change as a result of the EAS identification and resulting support. As such, its exclusion from the model is justified.

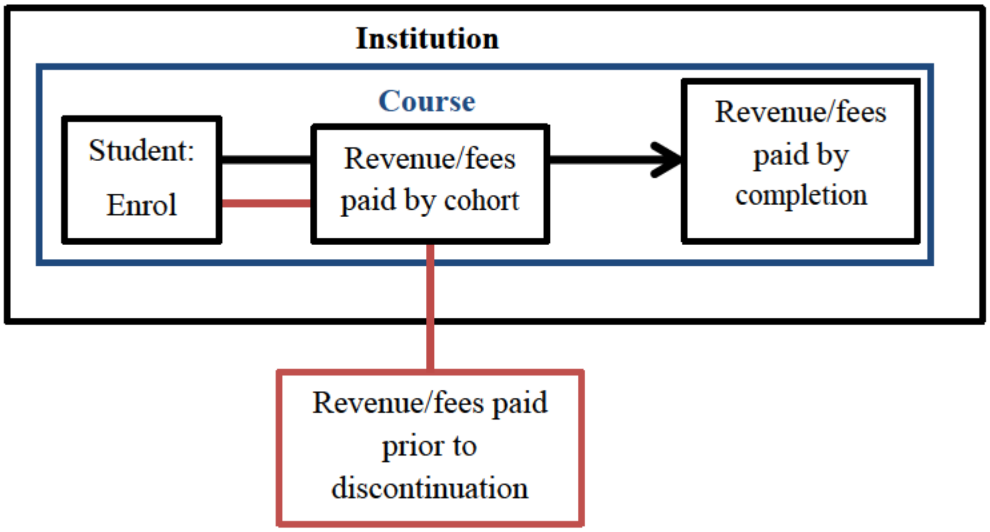


Figure 8.2 –Discontinuation Treatment Effects Model

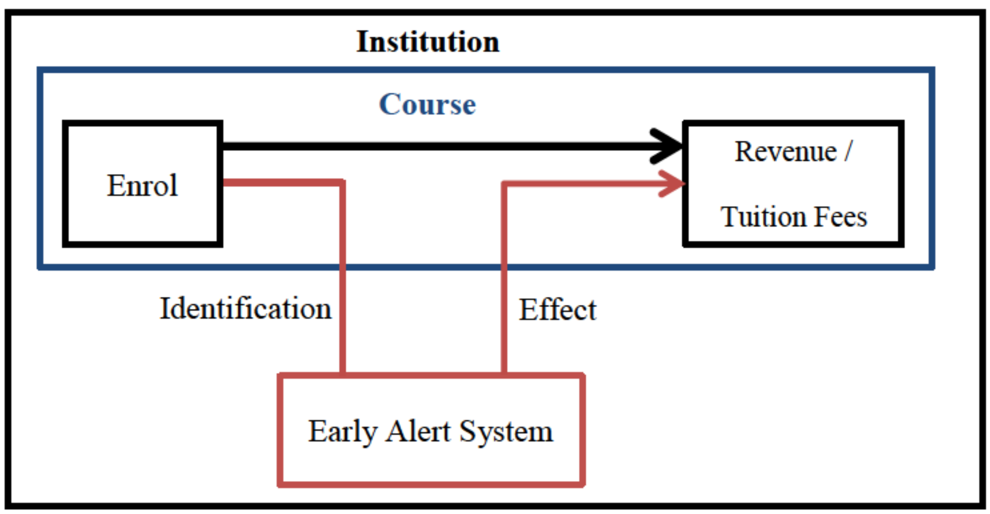


Figure 8.3 –EAS Treatment Effects Model

The data set used for the analysis is a pooled data set containing data across multiple teaching periods with students commencing at various times. A potential source of bias with the model is students who started in 2011 and identified by the EAS are matched and compared to students who start in 2013 and not identified by the EAS. To address this issue, an additional variable is included in the model, *teaching period of commencement*. This codes the students' commencement date as a categorical variable allowing exact matching of students based on when they first started classes. This prevents temporal mismatching occurring and ensures robust estimation of the effect.

Treatment effects modelling has two parameters of interest which help in identifying the causal relationship. The first parameter is the Average Treatment Effects (ATE), which compares the effect to the entire population. This provides a broad estimate of overall causal effect of the EAS. The second parameter is the Average Treatment Effects on the Treated (ATET), which only captures the effect on those people affected by a particular treatment. This provides a more specific estimate of what the causal effect would be under the condition of being identified by the EAS. These two parameters of interest are estimated for all models.. Additionally, summaries of the matching covariate estimates which indicate the validity of the matching process are presented in Appendix I.

8.3 The cost of discontinuation results

The first model estimated analysed the effect on revenue resulting from the decision to discontinue. One approach to quantifying this effect compared discontinued students to the entire student cohort, consisting of enrolled and completed students. This captured the loss of revenue at the moment when a student discontinues. Another approach compared discontinued students only to those who have completed their qualifications. The second approach provided an estimate of the overall loss of revenue. The results for the two approaches are presented in Table 8.3.

Table 8.3 – The cost of discontinuation and not completing results

Model	Estimate	Coefficient ($)	Robust Standard Error ($)	Z-value	Sample Size
Cost of discontinuing	ATE	4,687[a]	65.60	71.45	13,690
	ATET	4,231[a]	70.06	60.39	13,690
Cost of not completing	ATE	7,170[a]	255.66	28.04	2,460
	ATET	7,307[a]	339.99	21.49	2,460

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level
[d] rounded to three decimal places but not equal to zero

The results for both models are significant at the 1% level. The cost of discontinuing indicates that for each student discontinued, the university on average loses $4,687 overall, with the decision for an individual student costing $4,231 on average. Comparing discontinued and completed students, the institution loses $7,170 overall when a student discontinues instead of completing. In any particular case, the cost is higher at around $7,307.

Table 8.4 indicates the undergraduate student numbers at UNE over the period captured by the data set (Department of Education, 2014a). Using this data, the benefit of increasing student retention by 1% is calculated. As such, Table 8.4 also includes the number of students who would be retained from a 1% increase in the retention rate.

Table 8.4 – Undergraduate Student Numbers

	Number of Undergraduate Students	Dept. Education Retention Rate	Number of students retained from 1% increase in retention
2011	12,120	73.47%	121.2
2012	12,857	73.26%	128.57
2013	13,504	71.61%	135.04

By combining the ATE estimated from Table 8.3 with the undergraduate student numbers in Table 8.4, estimates of the cost of discontinuing and cost of not completing are estimated in Table 8.5. The total cost of discontinuing is calculated by the number of undergraduate student numbers multiplied by the attrition rate multiplied by the ATE estimate. The benefit of a 1% increase in retention is calculated by the number of undergraduate students multiplied by 1% multiplied by the ATE.

Table 8.5 – Estimated overall costs and benefits from 1% increase in retention rate

	Cost of discontinuing	Cost of not graduating
	Benefit of 1% Retention Increase ($)	Benefit of 1% Retention Increase ($)
2011	568,100	868,947
2012	602,646	921,786
2013	632,972	968,173

Comparing discontinued students to the remaining student body, the benefits of increasing student retention by 1% range from $568,100 in 2011 to $632,972 in 2013. The benefits of increasing student retention and seeing these students through to graduation results in a financial benefit of between $868,947 in 2011 to $968,173 in 2013. Given that student numbers have been increasing over time, this means that the financial benefits of increasing retention will also increase.

In summary, the results show there is a significant financial effect associated with students discontinuing. This is not a surprising outcome. However, what is important is the magnitude of the problem. Being able to accurately estimate the size of the financial implications associated with retention yields important information on potential benefits that can be gained from increasing student retention. Furthermore, this provides an important benchmark from which to measure new programs to be introduced to affect student retention.

8.4 Revenue effects of the EAS

Measuring the effects of an EAS is an important process all institutions need to undertake to validate the efficacy of the system. One way the efficacy of the EAS system can be measured is the additional revenue from students staying enrolled for longer. Using treatment effects models, students were divided into two groups, identified by the EAS and not identified by the EAS as done in previous chapters. Ideally, having more information on the student support process after identification would allow more detailed analysis of the individual aspects of the program and their contribution to the revenue function. In lieu of this detailed data not being present, it is still applicable to treat the EAS as a black box process and estimate meaningful effects resulting from the EAS and resulting student outcomes. The estimated ATE and ATET are presented in Table 8.6.

Table 8.6 – Overall effect of EAS on revenue

Model 8.2	Coefficient ($)	Robust Standard Error ($)	Z-value	Sample Size
ATE	4,004[a]	80.87	49.51	14,012
ATET	5,058[a]	102.56	49.31	14,012

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level

The effect of the EAS on revenue is significant at the 1% level for both the ATE and ATET estimates. The EAS is estimated to increase overall student tuition fee spending by around $4,004 per student. Individual students identified by the program on average continued to spend an extra $5,058 in fees compared to students not identified by the program. This is a significant finding which highlights the financial benefits of the EAS. This positive result corroborates the results found from previous chapters, that there is a significant positive effect associated from the program.

The estimate, however, does not come without limitations and issues. No similar benchmark estimate of the student support system were available before the introduction of the EAS. As such, it is not possible to estimate the benefits of changing from the prior support system to the EAS. This is an important measure in determining the benefit/cost ratio of installing the EAS and would help validate the EAS further. On the positive side however, the estimate provided can be used as a benchmark for future changes to the system. If the system is enhanced or changed in any manner, then the same process can be repeated to allow comparisons of the program valuations and estimation of benefit/cost ratios can be calculated for the program revision.

8.5 Conditional analysis within the Early Alert System

The previous two sections estimated the cost of students discontinuing and the benefit of the EAS overall. Another way to analyse the financial effect of the EAS is to estimate the additional tuition fees a continuing or completing student spends as opposed to a discontinuing student, under the condition they were identified or not identified by the EAS. Only the ATE is shown for simplicity; however, the ATET estimates are also presented in Appendix H.

Table 8.8 – Revenue of student outcomes under the condition of EAS identification

	Sub-group of students	Coefficient ($)	Robust Standard Error ($)	Z-value	Sample Size
Discontinued/ Continue	Not Identified	2,263[a]	69.77	32.43	3,142
	Identified	5,138[a]	89.33	57.52	9,099
Discontinued/ Complete	Not Identified	4,960[a]	112.37	44.14	183
	Identified	7,528[a]	321.97	23.38	1,623

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level

The results for each groups are significant at the 1% level. The results show if a student isn't identified by the EAS and continues, they will spend around $2,263 more than a student who discontinues. If, however, the student is identified by the EAS, the student will spend $5,138 extra on tuition fees compared to a student who discontinues. A t-test for the difference of two means results in a t-test statistic of 25.36. This means the difference between the two groups is significantly different at the 1% level. The additional $2,875 in student tuition fees associated with EAS identification further supports the benefits of the EAS system.

In the second model, if a student was not identified by the EAS and completed, on average the student would spend an additional $4,960 compared to a discontinuing student. If a student was identified and completed, then the additional amount of revenue spent compared to a discontinuing student was $7,528. A t-test for differences in means attains a test statistic of 7.53, a value indicating the difference is significant at the 1% level. Again, the statistical difference between these two groups indicates a financial benefit of the EAS. Importantly, this financial benefit arises due to an increased length of enrolment. As such this indicates that the EAS is increasing student persistence within their chosen course.

8.6 Early Alert System effects within schools

Variation between schools yields important information about level of identification within each school and can assist with areas of targeted resource allocation. For this section, the data set of students is divided into individual schools. The treatment effects model estimates the effect of the EAS on revenue within each school.

Table 8.9 – EAS effects within schools: results

School	ATE Coefficient ($)	ATE Standard Error	ATET Coefficient ($)	ATET Standard Error	Sample Size
0 (Base)	5,112[a]	307	6,098[a]	377	1,101
1	2,881[a]	206	3,771[a]	265	1,409
2	2,263[a]	134	2,992[a]	180	2,086
3	6,507[a]	304	7,685[a]	359	1,320
4	4,446[a]	251	5,143[a]	305	1,677
5	4,931[a]	296	6,017[a]	366	1,753
6	3,394[a]	145	4,363[a]	186	2,940
7	-	-	-	-	-
8	5,627[a]	460	6,816[a]	543	807
9	2,269[a]	196	3,189[a]	288	919

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level

The results show significant effects in all schools at the 1% level. The estimated treatment effects on identified students range from school 2, the smallest identification effect of $2,269, through to school 3 with an identification effect of $6,507. The range of variation between schools indicates that the program has a larger benefit in some school versus others. This supports the case that the differences between schools need to be factored into EAS design. The most important conclusion that can be drawn from the results is that all schools have a significant positive effect associated with the EAS. This supports the case of EAS being implemented at an institutional level where a unified approach is taken to offering support.

8.7 Early Alert System revenue effects over time

One final way of looking at the EAS effect is to introduce the temporal effect. It is expected that students identified earlier in their studies will have a greater revenue effect than students identified later in their studies. Categorising students within the data set on which year they are first identified, Table 8.12 shows that the majority is first identified by the EAS in their first year.

Table 8.12 – Year students are identified

Year Identified	Number of Students
0	4,830
1	10,619
2	596
3	79
Total	16,124

Using the treatment effects model, students identified in years 1, 2 and 3 are compared to the base group of students not identified. The results of treatment effect model 8.5 are presented in Table 8.13.

Table 8.13 – Revenue effects over time: temporal effects of the EAS

Year	Estimate	Coefficient ($)	Standard Error	Z-Score	Sample Size
1	ATE	4,000[a]	83.16	48.10	13,343
	ATET	5,151[a]	106.25	48.48	13,343
2	ATE	2,675[a]	213.67	12.52	1,755
	ATET	3,911[a]	259.86	15.05	1,755
3	ATE	3,479[a]	545.98	6.37	187
	ATET	3,256[a]	675.19	4.82	187

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level

The results for the ATET show that students identified by the EAS in their first year contribute $4,000 more in revenue than students not identified by the EAS. By the second year, this had reduced to $2,675, but in third year increased to $3,479. One limitation of the data is the diminishing sample size available for analysis in the third year. The number of observations within the treatment group in the third year was only 28 students, matched with 195 control students. This is relatively small and as such, the matching process only occurs on a few variables. However, despite the diminished sample size, the results are still statistically significant. This indicates that the EAS maintains significant value for all students, regardless of when they were identified. This is an important finding as it shows that there is value of an EAS beyond the first year of enrolment.

8.8 Chapter Summary

Revenue from student tuition fees is a fundamental source of income for institutions. The size of the retention problem was estimated using revenue from tuition fees as a baseline metric. The treatment effects method of estimating the benefits of the EAS yielded statistically significant results in all models estimated. Using *demographic*, *institutional* and *workload* variables to match observations, valid comparisons were made between students when a control group is not possible. This supports the use of treatment effects modelling as an appropriate method of analysis and evaluation for EAS.

Comparing discontinued students to the general student population, the first model indicated that the cost of students not being retained was approximately $4,687 per student. If discontinued students are compared to graduating students, the estimate is significantly higher, costing the institution approximately $7,170 per student. These estimates show the magnitude of financial loss associated with current student discontinuation. A modest 1% increase in the undergraduate student retention rate can yield significant financial benefits. With a 1% increase in student retention, the university could have gained an additional $568,000 in 2011, up to $633,000 in 2013. If the 1% of undergraduate students retained continues onto graduation, this increases further to a benefit of $868,000 in 2011 to $968,000 in 2013. The estimated cost of discontinuation is an important measure for any institution. For the case study institution, since approximately $4,700 is lost per student who discontinues, then theoretically this also acts as an upper estimate of the additional amount of funds that can be spent per student to improve retention. This forms an important baseline measurement for funding student support initiatives for the case study institution.

Estimating the effect of the EAS overall showed that students identified by the EAS end up paying more tuition fees on average than students not identified by the system. This correlates to being enrolled longer and undertaking more units. The revenue linked to the EAS was quantified, with students identified by the EAS. This is an important significant finding. Given that the majority of student support services at the institutional level function through the EAS, this provides an overall estimate of the value of student support at the case study institution. For administrators, having this information can greatly enhance understanding and the importance of adequately funded student support services. Furthermore, this estimated value allows a benchmark to be taken. Future changes and enhancements to the EAS can be quantified allowing benefit/cost estimates to be calculated, allowing administrators to make evidence based decisions to be made.

Analysing students under the condition of EAS identification, the results supported the previous findings. Within the identified group of students, the revenue from continuing students was on average $5,138 more than discontinuing students, whereas in the not identified group, the revenue from continuing students was only on average $2,263 more than discontinuing students. The difference between these two amounts was statistically significant, indicating that there is a significant increase in revenue associated with EAS identification. When comparing discontinuing students to graduating students not identified by the EAS, the revenue from graduating students was on average $4,960 more than discontinuing students. When students were identified by the EAS, the revenue from graduating students was $7,528 more than those discontinuing students. The difference between the two estimated effects was statistically significant at the 1% level, further supporting the conclusion that the EAS has significant financial benefits in terms of revenue, which translates to increased student enrolment and improved retention.

Analysing within schools, the results showed an amount of variation of EAS effects on revenue. School 9 had the smallest ATE estimate of $2207 with school 3 having the largest ATE of $6617. Overall, all schools (where possible to estimate) had a benefit from the EAS, supporting an institution wide approach to EAS design. This is an important significant finding, as it shows that a demonstrated institution wide approach to EAS design can have benefits for all schools within the institution. This will vary depending on institutions; however the results here indicate that for the case study institution, the EAS is a benefit to all schools. The results indicate that some schools may also benefit from additional support programs within schools, and that funding allocated to these programs should reflect both the benefits and the need for programs based on retention rates.

The final model analysed the value of the EAS based on when students were first identified. It shows that students identified in their first year of study corresponded to an additional $4,000 in revenue. If the student was identified in their second or third year of study, the additional amount of revenue was $2,675 and $3,479, respectively.

All of these estimates were significant at the 1% level, indicating that the EAS has significant value for all students, independent of course progression. This is a significant finding in the context of student retention, given that the major focus in the past has been on first year student retention. The results show that EAS can have significant value to students at later stages of study. This further supports that EAS should not only be implemented at the school level, but capture all students within the institution. Overall, the results show that there is significant financial benefit to implementing an institution wide EAS which captures all students irrespective of course progression.

Chapter 9 - Summary, General Discussion and Conclusion

9.1 Introduction

This chapter summarises and discusses the major results and their implications. The chapter is organised as follows. Section 9.2 presents a general overview of the thesis. The main findings are summarised and discussed in section 9.3 to 9.5. Section 9.6 discusses research implications. The chapter is concluded in section 9.7 with a reflection on the future of Early Alert Systems.

9.2 Brief overview of the study

Maximising student retention is always going to be a central objective for universities globally. There will always be students who discontinue their studies. However, universities can develop programs which minimise the number of students discontinuing, while maximising student outcomes and welfare. Early Alert Systems (EAS) implementing learning analytic solutions to identify students at risk hold great promise. Combined with other support initiatives, universities are being technologically revolutionised to help achieve improved outcomes for students, institutions and the sector as a whole.

The main objective of this study is to employ a microeconometric approach to evaluate the relationship between Early Alert Systems and student retention. Using the information and data collected from a case university, the following research questions were considered:

1. What variables affect the student retention rate of undergraduate students?
2. What is the effect of EAS on student retention rates?
3. What is the relationship between the variables affecting student engagement and variables affecting student retention?
4. What are the financial implications of improving student retention rates?

The thesis is organised in nine chapters. Chapters 1 to 4 provided the contextual and background information, Chapters 5 to 8 presented the empirical and analytical results, with the general discussion, summary and conclusions presented in the final chapter. More specifically, chapter 1 introduced the main research questions and developed the need for the study undertaken. Chapter 2 reviewed student retention theory and learning analytics literature on early alert systems. The chapter outlined the importance of evaluating EAS for effectiveness to support the business case of EAS being implemented within institutions. Chapter 3 reviewed the empirical approaches used to estimate effects associated with student retention. It concluded with the development of the approaches used to estimate the effects of an EAS. Chapter 4 covered the case study institution and the data set provided for analysis. A critical part of the data provided, was the data used by the EAS for identification of students. Chapter 5 used a likelihood analysis approach to identifying the variables that affect the likelihood of student outcomes and how this correlates to the overall student retention rate. Chapter 6 used an Ordinary Least Squares (OLS) approach to measure changes in the length of students' enrolment. This provided as a means of quantifying the magnitude of improvements in retention associated with the EAS and other factors affecting retention. Chapter 7 analysed retention in a temporal setting using survival analysis. This approach allowed a detailed quantitative analysis of the hazard function associated with students discontinuing from their studies. Finally, chapter 8 used a treatment effects approach to analyse the financial implications of the EAS with causal inferences. The chapter discusses the significant findings from these chapters by answering the research questions in order. It then proposes areas where the study could be further enhanced and scope for future research. This chapter concludes by contextualising the research findings in the learning analytics community and provides commentary on the future of EAS.

The major findings from the empirical analyses are summarised in a combined manner in order to address the main research questions. These summaries are presented in the ensuing four sections below.

9.3 Variables affecting student retention of undergraduate students

It is important to establish the significance and magnitude of relationships between different aspects of the university environment. This builds the foundation for complex models used in this study. Four categories of variables are used to capture the factors affecting student retention and control for effects outside of the EAS. These are the *demographic*, *institutional*, *student performance* and *workload* variables. *Demographic* variables capture the background factors students bring with them when enrolling to study. *Institutional* variables capture students' interaction with the institution, including fee category, course type and school of enrolment. *Student performance* variables use the academic record of the students to factor in varying outcomes in the learning environment. *Workload* adjusts for the rate a student progresses on a chosen course of study. Each class of variable is summarised by the three exploratory models of chapters 5, 6 and 7, which used different approaches to understand student retention.

To understand the complexity of student retention, heat maps summarising the results of the microeconometric models are presented below, along with the key for understanding the heat maps presented in Table 9.1. The colour scheme indicates the varying levels of significance, whereby those in blue show significant positive effects and red suggests negative effects. Those with white shaded parts of the table indicate no relationship.

Table 9.1 – Heat map legend

Significance		Colour
Positive Effect	1%	
	5%	
	10%	
No significant Effect		
Negative Effect	10%	
	5%	
	1%	

Key elements of the heat maps should be interpreted as follows:

- For the likelihood approach, a positive effect either decreases the likelihood of discontinuing, or increases the likelihood of completing. In contrast, a negative effect either increased the likelihood of discontinuing, or decreases the likelihood of completing.
- Using the OLS approach, a positive effect is one that increases the length of students' enrolment, whereas a negative effect will shorten students' enrolment.
- For the survival analysis approach, a positive effect decreases students' hazard ratio and a negative effect increases the hazard ratio.

9.3.1 Demographic variables

Three demographic factors: *gender*, *age* and Aboriginal and Torres Strait Islander *(ATSI)* status were used by this study. These variables control for factors that may affect a student's outcomes with respect to retention that are distinct characteristics of the student. Table 9.2 summarises the significant variables for the various approaches used to understand retention.

Table 9.2 – Summary of effects for demographic variables

Approach	Liklihood Approach		OLS Approach	Survival Analysis Approach		
Dependant Variable	Probability of Discontinue	Probability of Completing	Length of Enrolment	Hazard Ratio		
Year of effect	Pooled		Pooled	1	2	3
Gender (M=0, F=1)						
Age						
Age Squared						
ATSI						

Significance	Positive Effect			No significant Effect	Negative Effect		
	1%	5%	10%		10%	5%	1%
Colour							

Gender is an important variable from a social perspective. Ideally there should be no significant difference in student outcomes based on gender. Using multinomial logistic regression to estimate the likelihood of discontinuing and completing, the results indicated that female students were 24% more likely to discontinue their studies than their male counterparts. However, the results also showed that the female student cohort was more likely to complete their qualification. Using an OLS approach, there was no significant effect due to gender on the students' length of enrolment. Using a survival analysis approach, there was a significant effect, where female students had a significantly higher hazard ratio. This was consistent at all times throughout students' enrolment. These results show that while ideally there should be no significant difference between student outcomes based on gender, this is not the case. This finding highlights the importance of EAS systems to test and target gender differences. Furthermore, within the suite of university support options, there need to be services which address gender issues.

The second demographic variable that affects student retention is age. This variable was consistently significant throughout the study. Most instances of the variable showed a non-linear relationship. Using a likelihood approach, this relationship showed the probability of completing increased up to the age of 43, and then decreased afterwards. Using OLS, the results indicated a linear relationship on the length of a student's enrolment, with each additional year increasing the length of enrolment by 0.22 weeks. The effect was significant at the 5% level. Therefore, the difference in expected length of enrolment for a 28 year old student (university average) and an 18 year old student is 2.2 weeks of extra enrolment. Using survival analysis, the hazard ratio of a student decreased up to 51 years of age, after which the hazard ratio started increasing again. The results converge on similar conclusions: age is a significant factor that affects student outcomes and, in turn, retention. It needs to be factored into EAS design, as older students will have different life experience and circumstances that differ greatly from high school graduate students. This will in turn affect the likelihood of discontinuing their studies. This is important, especially for institutions with diverse student cohorts like the case study institution.

The third factor that affects student retention is identifying as ATSI. The likelihood approach showed a weakly significant decrease in the likelihood of discontinuing from ATSI students. The OLS approach attained no significant effect of identifying as ATSI, indicating no significant difference in the length of enrolment between ATSI and non-ATSI students. Survival analysis approach however revealed a more detailed picture. Overall, ATSI students had a significantly lower hazard ratio than non-ATSI students. Assessing the short run and long run effects, ATSI students had a significantly lower hazard ratio than non-ATSI students in the short run. In the long run, ATSI students started with a significantly lower hazard ratio that increased over time, where after 36 weeks, ATSI students were at a higher hazard of discontinuing than non-ATSI students.

If the ATSI student was also identified by the EAS, there was a significant reduction in the hazard ratio after the 36 week mark over time. This indicates that there is a strong link between EAS identification and an improvement in ATSI student outcomes over time. Analysis of EAS trigger for ATSI showed no significant relationship between the ATSI trigger and the survival function. This then raises an interesting question given that the interaction effects model showed a decrease in hazard if identified. One plausible explanation for this is that ATSI students are being identified by the EAS on other triggers independent of the ATSI trigger. Analysing the EAS triggers and controlling for *demographic, institutional, student performance and workload*, ATSI students showed a reduced hazard ratio overall. This then indicates that the EAS trigger may be able to identify ATSI students in need of support, but the trigger may not be functioning correctly until other factors are controlled for in the EAS.

In summary, the results show ATSI students have a lower hazard function than non-ATSI students, making it an important factor that affects student retention for the case study institution.

One explanation for the decreased hazard ratio is additional support ATSI students receive from a specialist centre for ATSI students on-campus. These positive results show the case study institution has appropriate policies in place to retain ATSI students. There is perhaps scope within the institution to further pursue increased ATSI student numbers, using the current approach to ATSI student support.

9.3.2 Institutional variables

The institutional setting for students varies greatly, with a range of courses to choose from. Therefore it is important to capture the nuances of the students' institutional variables to adequately model retention. A heat map of effects the institutional variables have on retention measures is provided in Table 9.3.

Fee categories of students were used to divide students into three clear groups: *domestic HELP*, *domestic upfront* and *international fee* students. *HELP* refers to the Higher Education Loan Program, a system designed to allow Australian students to finance study through a no-interest government loan. Comparing domestic *HELP* students to students who paid for tuition fees upfront, the models revealed no difference between the two groups. In part this is due to the small sample size of students who pay fees upfront. As such, while there may be a difference between *HELP* and up-front fee paying students, any effect is too small to make a significant difference on the retention rate.

Table 9.3 - Summary of effects for institution variables

Approach	Liklihood Approach		OLS Approach	Survival Analysis Approach		
Dependant Variable	Probability of Discontinue	Probability of Completing	Length of Enrolment	Hazard Ratio		
Year of effect	Pooled	Pooled	Pooled	1	2	3
Domestic Fee						
International Fee						
Prior Studies						
On Campus						
Diploma						
Advanced Diploma						
Bachelors (Graduate)						
Bachelors (Honors)						
School 1						
School 2						
School 3						
School 4						
School 5						
School 6						
School 7						
School 8						
School 9						

Significance	Positive Effect			No significant Effect	Negative Effect		
	1%	5%	10%		10%	5%	1%
Colour							

International fee students on the other hand had a significant effect in all models. International fee- paying students were significantly less likely to discontinue from their studies than domestic students. This corresponds to similar findings of the research by Olsen (2008, p. 5) and the current trends in higher education data (Department of Education, 2014b). The likelihood approach showed international students were far more likely to complete their qualification than domestic HELP students. The OLS results showed that international students are enrolled for less time. Technically this is shown as a negative effect, however, this effect is due to most of the international students already completing a year of study before entry into the institution. As such, it should not be seen in the same light as other negative effects from the model. Using survival analysis to capture temporal effects, the models continually showed international students initially had a hazard ratio less than domestic HELP students. The estimates changed over time however, indicating that international students may need additional focused support closer to the completion of their qualifications. The results imply that EAS design needs to factor in fee status, to accurately identify students in need of support.

Prior studies is an institutional variable; however, it could also be considered part of the demographic or background variables as it captures whether the student has undertaken any previous study after high school. Regardless of classification, it was expected that students who had undertaken previous study would be more prepared for university, reducing their hazard ratio. Using the likelihood approach, students with prior studies were more likely to both discontinue and complete their studies. This indicates that there are two contrasting effects associated with prior studies as an explanatory variable in retention analysis. The first is that having undertaken prior studies, some students have a higher opportunity cost associated with studying and may not need to undertake further study. This would explain the increased likelihood of discontinuing. Conversely, students with prior studies have had experience with studying and will be more likely to complete their chosen courses. In measuring the length of students' enrolment, students who had undertaken prior studies were enrolled for 6.4 weeks more than students who had not done prior studies. Using survival analysis, the effect was less pronounced, with only an initial reduction in the hazard ratio over the first weeks of study. The results indicate that overall, students with prior studies are enrolled for longer with a lower initial hazard ratio, but including prior studies as an explanatory variable may need to be treated with caution especially when using a likelihood approach. If prior studies are to be used in an EAS, it may be important to have information that allows the separation of the contrasting effects identified with the likelihood approach. A key limitation of the variable is what constitutes prior study. In this study, there was no way to disaggregate different levels of prior study. It is expected that some prior courses will better prepare students than others, and for the institution it may be important to capture more detailed data with respect to prior study.

The mode of enrolment showed significant effects in all models. The *on-campus* variable was significant in all of the models in some form. The likelihood approach showed on-campus students were more likely to discontinue their studies than off-campus students. Furthermore, on-campus students were less likely to complete their qualifications. Using the OLS approach, these results were supported, with on-campus students on average being enrolled for between 0.8 and 1.4 of a week less than off-campus online students. The conclusions from these models, however, were not supported when other variables in the model were controlled for. The survival analysis approach showed that on-campus students did have a higher hazard ratio than off-campus online students, but only for the first few weeks of teaching. Four weeks after commencement, the hazard ratio for an on-campus student was lower than that of off-campus online students. Furthermore, the hazard ratio continues to decline for on-campus students over time, indicating a strong disparity between on-campus and off-campus online students with respects to retention. An important finding is it highlighting of the limitations of the approaches used in chapters 5 and 6, which indicated the opposite effect. Without adequately accounting for temporal effects, the pooled results would give a false indicator of success in supporting online students. Whilst the case study institution has a strong history of supporting students studying off-campus online, the results from the survival analysis show that there is a gap in retention between on-campus and off-campus online students.

The *course* a student undertakes influences student outcomes greatly. The likelihood approach showed students undertaking a diploma were highly likely to complete their qualification compared with bachelor level students. This should not come as a surprise given the number of units required to complete a diploma is one third that of a bachelor course. Advanced diploma students were also more likely to complete their qualifications. However, they were also more likely to discontinue than bachelor level students. This may reflect the level of commitment students have to completing advanced diplomas. On the other hand, students admitted both via graduate entry and honours programs are less likely to discontinue. They are also less likely to complete. However, this is more likely to be a limitation of the data set itself. Honours courses generally require four years of study. Capturing only three years of data means that most honours students in the data set would still be enrolled, with only a few completing. In chapter 6 where the length of enrolment was estimated, the results between courses vary in both significance and magnitude. Diploma students are estimated to be enrolled for 2.95 to 3.3 weeks longer than bachelor students. This result likely captures the interaction with the workload variable. While diploma students do fewer units than bachelor students, they take more time to complete their qualifications due to lower workload levels. The effect is more pronounced in the advanced diploma group, where students are enrolled for an additional 9 to 9.2 weeks. Graduate students are enrolled for an extra 1.8 to 1.9 weeks and honours students for an extra 5.9 to 6.3 weeks.

These results indicate that course variations should be accounted for in combination with the workload variable to accurately determine the effect of different courses on the length of enrolment. Using the survival analysis approach where the hazard ratio is estimated, results consistently showed no significant difference between diploma and bachelor students. Advanced diploma students had a lower hazard ratio, but only for the first few weeks of study. With respect to graduate entry and honours students, there was a significantly lower hazard ratio. This indicates that the course level is a strong indicator of student retention. However, it needs to be adjusted for the varying workload levels of students.

The final institutional variable is the *schools* within the institution. The base case school offers professional qualifications. It is expected that there will be variations between schools due to the different objectives and courses taught. The likelihood analysis showed that compared to the base case school, students in schools 3, 4 and 8 were more likely to discontinue their studies. School 7 however showed that students were far less likely to discontinue. With respect to the likelihood of completion, schools 1 to 6 and 9 were less likely to complete their qualifications than students in the base school. The OLS approach revealed length of enrolment estimates that support the conclusions of chapter 5. Students enrolled in schools 1 to 7 were enrolled for a shorter period of time, varying between 1.1 and 3.5 weeks less than students in the base case school. Students enrolled in school 7 were enrolled for an additional 24 to 26 weeks, which is to be expected given the courses taught within this small school. The hazard ratio estimates from the survival analysis showed that students in schools 3 to 5 had a much higher hazard ratio at all points of time, while school 1 had a less significant increase in the hazard ratio. School 6 only exhibited an increase in the hazard ratio in the second year of study. This means that schools themselves are a contributing factor to student retention and need to be considered when identifying students at risk with an EAS.

9.3.3 Student performance and workload variables

Student performance and *workload* are significant variables required when conducting retention analysis. A summary of the results are presented in Table 9.4 for the various models. Grade distribution provided a detailed breakdown of student academic records over time. This was preferred to GPA alone as it gives a more accurate representation of the students' performance within the learning environment.

Table 9.4 - Summary of effects for student performance and workload variables

Approach	Liklihood Approach		OLS Approach	Survival Analysis Approach		
Dependant Variable	Probability of Discontinue	Probability of Completing	Length of Enrolment	Hazard Ratio		
Year of effect	Pooled		Pooled	1	2	3
Withdrawn						
Withdrawn Early						
Fail Incomplete						
Fail						
Pass						
Credit						
Distinction						
High Distinction						
Other						
Increasing Workload						
Inactive						
Part Time						

Significance	Positive Effect			No significant Effect	Negative Effect		
	1%	5%	10%		10%	5%	1%
Colour							

The likelihood approach results showed that the negative grade outcome of failing incomplete increased the likelihood of discontinuing. Interestingly, students who withdrew, withdrew early or failed a unit were less likely to discontinue. This indicates that despite attaining these negative grades, students who continue to attempt units lower the chances of discontinuing. This in essence captures the importance of students persevering with studies. With respect to the likelihood of completing, all of the negative grades were associated with decreased odds of completing. The OLS results were significant and positive with values between 0.078 and 0.128. This means that for each additional grade a student attained, they were enrolled for an additional 0.078 up to 0.128 weeks. The magnitude of the effect is minimal and in many ways only helps to control for student performance within the model. In the survival analysis approach, the effect size was more pronounced. Attaining a *fail* or *fail incomplete* grade had more consistent negative effect, being constantly associated with a higher hazard ratio. The positive grade outcomes of *pass*, *credit*, *distinction* and *high distinction* all affected the students' hazard ratio by lowering them for each instance these grades were attained.

An important finding was that only in the first year of study, receiving a *withdrawn* or *withdrawn early* grade was associated with a significantly higher hazard ratio. This indicates that withdrawing from units can work as an early indicator of behaviour that leads to choosing to discontinue. However, it is only valid for the first year of study. This shows there is a clear distinction between the effect of withdrawing from a unit versus failing a unit. This effect may be attributable to students engaging with university administration to formally withdraw from a unit. This may assist the student in being more confident to engage with the institution and better understand what is required to progress and complete qualifications. In the context of providing student support, advising students to withdraw from a unit of study may be the best option if sufficient progress has been made, where there is no significant impact on the hazard of discontinuing.

The final variable of interest is *workload*. This continuous variable takes on values between 0 and 2. In the likelihood and OLS approaches, this was a weighted average capturing the number of units a student was enrolled in, in any given week. In the survival approach, this variable was expressed as a categorical variable, alternating between inactive (0), part-time (1) and full-time (2) in any given week. This allowed a more flexible model to reflect how a student can take on varying levels of workload at different stages of their course. As seen in the likelihood approach, increasing the workload of a student increased the likelihood of discontinuing and increased the likelihood of completing. This captures a critical issue for students when deciding the amount of work to undergo during each teaching period. If the student takes on more units of study, they will progress through their chosen course quicker, corresponding to the increased likelihood of completing. With increased workload also comes more stress and greater opportunity cost, which also explains why the likelihood of discontinuing also increases. This is an important finding for student support services in general, as it demonstrates that one method for increasing student retention is for the student to take on a decreased workload.

Using the OLS approach, increasing the workload decreased the length of enrolment. For example, if a student was studying part-time, this decreased the length of enrolment by 31 weeks. It is estimated that a full-time student would see a reduction in the time enrolled by double this. This is a logical outcome as students who undertake more workload will complete their courses sooner. The survival approach allowed for the hazard ratios for inactivity to vary over time. Generally, students who were inactive had the largest increase in the magnitude of the hazard ratio compared to any other variable modelled. This is an expected result; it highlights the importance of retaining some degree of engagement or interaction with the learning environment. The model did not differentiate between formal inactivity and informal inactivity, and this might be an important area of research to investigate further.

Focusing on part-time students, chapter 7 results showed in most instances part-time students had a higher hazard ratio than full-time students. The results show that the workload variable has a significant relationship with student retention. Importantly, workload needs to reflect the actual work students are undertaking. In Chapter 7, this was the number of units undertaken during any given week. In reality, courses have varying workloads within the units of study. In the future, this may be an important variable to capture at a more detailed level, for example, the number of hours of study each week. This may provide more meaningful interpretation of workload.

9.3.4 The effect of time

One way to measure improvements in retention associated with student support is to measure the increased time a student remains enrolled after support is accessed. As such, one of the most important factors of student retention is time itself. The survival analysis approach allowed temporal effects to be captured, with some variables like *workload* having complex changes over time. Importantly, many of the explanatory variables interacted with time as a variable. This indicates that factors affecting student retention fluctuate throughout the students' enrolment, with some factors being important during the early stages of enrolment like attaining a withdrawn grade. While other variables show significant changes later in enrolment, like the increase in hazard for international students at the two year mark. It can be generally concluded that any time itself is a significant factor that affects student retention. For the case study institution, temporal effects were significantly pronounced in the triggers used by the EAS to identify students at risk. More of this is discussed in section 9.3, but it highlights that EAS must factor in temporal changes to adequately predict students at-risk of not being retained.

9.3.5 Summary

To address the research question, variables that affect student retention include *demographic*, *institutional*, *student performance* and *workload*. All three demographic variables, *gender*, *age* and *ATSI* status showed significant relationships to measures of retention. Of the institutional variables, *international fee* paying, *on-campus*, *course type* and *school* all showed consistent significant relationships to the various measures of student retention. Other institutional variables of *domestic fee* paying and *prior studies* showed varying levels of significance through the study. These variables may not affect retention; however it can be argued that they need to be included in models on student retention as controlling variables. Performance of students measured by the grade distribution formed an important component of the models. These variables were significant in all models. Finally, *workload* was a significant contributing factor to students and the decision to discontinue.

The results presented have important implications for student retention analysis within the learning analytics community. Firstly, the factors that affect student retention are complex, with many variables having temporal effects. The results demonstrate the need to have a temporal approach to analysing retention which is only possible if suitably granular data is available. This raises a second point: the need for comprehensive data capturing the many aspects of the learning environment. The data set provided by the case study institution allowed analysis to take place with unprecedented detail in the survival analysis approach. This supports other research which indicated “teachers, students, faculty, support staff and administrators can all benefit through the application of data to understand what’s happening in classrooms and how to improve and optimise learning” (Siemens et al., 2013, p. 5). The results demonstrate the benefits of cooperative data collection and allowing analysis to take place independent of data silos.

9.4 The effect of the EAS on student retention rates

A major focus of this study was on the effect of the EAS on student retention. Using three years of observational data, the effect of the EAS was tested using a variety of statistical approaches. Importantly, the data relating to the EAS only captured the identification process. No data was obtained about the resulting student support services accessed by the students identified by the EAS. A black-box approach was taken in analysing the results, where the identification of students served as the input to the support system, and student outcomes the result of the process. This still allows valid inferences to be made about effects associated with the EAS, which are summarised using a heat map in Table 9.5

Table 9.5 – Summary of effects for the EAS

Approach	Liklihood Approach		OLS Approach	Short Run Survival Approach			Long Run Survival Approach			Survival Treatment Effects Approach
Dependant Variable	Probability of Discontinue	Probability of Completing	Length of Enrolment	Hazard Ratio			Hazard Ratio			Length of Enrolment
Year of Enrolment	Pooled	Pooled	Pooled	1	2	3	1	2	3	Pooled
EAS Identified										
Severity Level 1				NA			NA			
Severity Level 2										
Severity Level 3										
Severity Level 4										

Significance	Positive Effect			No significant Effect	Negative Effect		
	1%	5%	10%		10%	5%	1%
Colour							

The likelihood approach expressed the EAS in two different configurations. The first was to take an overall effect by splitting students into identified and not identified sub groups. The second configuration was to have varying severity levels based on the number of times a student was identified. Severity level 1 indicates students identified one to four times; severity level 2 indicates students identified five to nine times; severity level 3 indicates students identified ten to nineteen times; severity level 4 indicates students identified by the EAS more than twenty times during their enrolment.

The results show that the EAS had a significant positive effect on the probability of completing. The effect was significant at the 5% level for students identified between one and four times, and significant at the 1% level for all other instances. Estimating the likelihood of discontinuing, only students identified between five to nine and ten to nineteen times had a significant negative effects at the 10% and 1% levels respectively. This means that only students identified in these categories had a higher likelihood of discontinuing. The varying results from the model provides limited capacity to infer a strong relationship between the EAS and student retention. From the likelihood approach, it appears the EAS is more closely linked to completion than discontinuation.

The OLS results were expected to positively affect the length of enrolment, translating to a positive increase in student retention. Using both the overall effect and varying treatment levels approaches, the model yielded significant effects in both models at the 1% level. Overall, the increased length of enrolment resulting from identification was an additional 8.9 weeks of enrolment. With the varying levels of treatment, students identified one to four times displayed the same 8.9 week increased in length of enrolment. If the student was identified five to nine times, this increased further to an additional 13.1 weeks of enrolment time. Being identified ten to nineteen times showed students being enrolled for an additional 9.4 weeks and students identified more than twenty times had an additional 5 weeks added to their enrolment. This is an important finding, as it shows the effects of the EAS are constantly significant at the 1% level.

The survival analysis results capture a range of significant effects associated with the EAS. Whilst a number of models were tested, they can be generally broken down into short run and long run models. The short run approach estimates the hazard ratio of a student at the moment of identification. It was expected that at the moment a student is identified by the EAS, the student has a significantly higher hazard ratio. The results in Table 9.5 indicated that students identified in the short run had a higher hazard ratio for the first two years of study. After this time, the students identified by the EAS did not have a significantly higher hazard ratio.

The long run approach estimates the hazard ratio of the students overall, broadly separating students into EAS identified and not EAS identified groups. No expectation was formed with respects to this model, however, it was found that students in the identified group had a significantly lower hazard ratio than students not identified by the EAS. The positive effect was only valid for the first two years of study before tapering off and becoming insignificant in the third year. Given that both the short run and long run approaches taper to have insignificant effects in the third year, this indicates that the EAS may struggle with identifying students correctly later in the students enrolment.

This result was supported by analysing the survival treatment effects approach. This approach blends survival analysis and treatment effects approaches to estimate the length of the students enrolment. The model shows overall, being identified by the EAS caused students to be enrolled for an extra 13.5 weeks. The magnitude of the effect was in line with the OLS approach used to estimate the length of enrolment. When varying level of identification were taken into account, the average treatment effect on the treated for identification between one and four times was insignificant. This indicated that there was no significant effect when initially identified by the EAS. Being identified five to nine times caused students to be enrolled for an extra 14.4 weeks on average, a significant result at the 1% level. Being identified ten to nineteen times caused students to be enrolled for an extra 43 weeks and being identified more than twenty times caused students to be enrolled for an additional 47.5 weeks. Both of these were significant at the 1% level. While it is not possible to see inside the black box of student support after identification, these results show that there is a causal effect on the students' length of enrolment associated with increased levels of identification.

The results for survival treatment effects approach are important, suggesting a causal link exists between the identification process and students' length of enrolment. Furthermore, it reveals important details about the function of the EAS. Currently, the most pronounced effect occurs when a student is identified more than ten times. This means, on ten or more days, the student ranked in the top 200 list of at-risk of disengaging. While the EAS is having a positive effect, it is unlikely that a new student who is commencing a course is going to produce sufficient data within the EAS triggers to be identified as at risk, let alone enough data to be identified more than ten times. This is further supported by the Kaplan Meier diagram presented in Figure 9.1, where the initial divergence between survival functions occurs within the first teaching period.

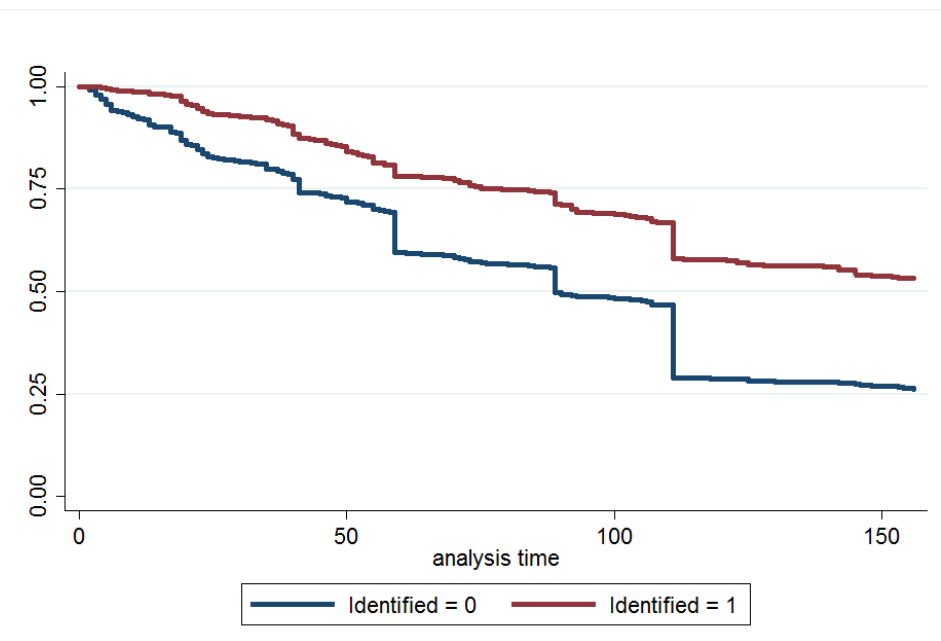


Figure 9.1 – Kaplan-Meier survival estimates

This indicates one area of improvement for the EAS is training the system to better identify those students who are going to discontinue early in their enrolment. If the system was to incorporate both a predictive element and the current reactive triggers, this may be one way of addressing the issue.

9.5 Relationship between student engagement and retention

The third research question posed in this study was whether a relationship exists between the variables affecting student engagement and variables affecting student retention? The original EAS design objective was improving student engagement using 34 triggers to identify students (Appendix B) at risk of disengagement. Assuming the EAS is correctly identifying students at risk of disengaging, overall the relationship between engagement and retention is supported with empirical evidence.

Two models tested the EAS triggers. The first was a standalone model which used just the EAS triggers. The second model incorporated the base survival model, using *demographic*, *institutional*, *student performance* and *workload* variables as controls. For simplicity, these models will be referred to as the stand alone and control models respectively.

The standalone model concluded that only seven triggers were valid in identifying students with increased hazards of discontinuing. The other triggers either captured no effect, or the effect was sensitive to model specification. The valid triggers identified had varying effects over time and as such, needed a temporal approach to accurately capture the link between engagement and retention. The control model showed that when other factors were controlled for in the model, ten triggers were able to capture the relationship between engagement and retention.

From these results, two key conclusions can be made. First, under the assumption that the EAS is correctly identifying students at risk of disengaging, there is a link between student engagement and retention. This is demonstrated both in the overall effect of the EAS system on student retention, and by the trigger analysis which breaks down the identification process into individual variables. The second conclusion is that the assumption the system is identifying students at risk of disengagement is a huge assumption to be making in the absence of quantitative results. A robust analysis of student engagement should be conducted using models similar to chapter 7, where the failure event is defined using a metric or combination of metrics that capture genuine student disengagement.

The first significant trigger was college attendance which was expected to increase students' chances of disengaging, but actually found students who attended college had a lower hazard ratio. It is easy to stereotype college students and college culture, however the evidence suggests there is more to college life than is being captured by the EAS. While college students may appear to be disengaging from study, the informal and unobserved college environment may actually mean the student is far more engaged than originally assumed.

The second significant trigger was that the student was already flagged for contact by the retention team in the current teaching period. Students setting off this trigger had a significantly higher hazard ratio in the short term model. However, this was not a significant variable in the long term model, meaning that overall this trigger did not affect the hazard ratio. As such, the question needs to be asked if this actually captures disengaging behaviour or if this is just a recursive trigger.

The third significant trigger was that the student was already flagged for contact by the retention team in the previous teaching period. Like the previous trigger, this is a recursive trigger where identification in one period increases the likelihood of identification in the following period. However, again, it has to be asked if this actually represents students exhibiting disengaging behaviour.

The next three triggers with significant hazard ratios relate to the student online portal. Students with low, medium and high portal inactivity had significantly higher hazard ratios than students who did not set off these triggers. This variable does capture student disengagement, and provides the link to increased hazard ratios as a result of this behaviour. This shows empirical evidence of a link between student engagement and retention.

The final two triggers relate to fail and fail incomplete grades attained in the prior teaching period. If you assume that students who fail or fail incomplete become disenchanted and disengaged upon receiving these results, then the increased hazard ratios support the link.

In conclusion, the results show some of the triggers that capture student engagement also affect student retention. One of the major limitations of linking engagement to retention is that the EAS actually only contains three direct measures of student engagement; the e-motion tool, the e-reserve activity and the student portal access. The remaining triggers included in the EAS capture indirect effects of student engagement, or they attempt to control for other variables that affect engagement. Given the quantity of triggers used and the reactive nature of the system, there is scope to improve the function of the system if it was to objectively target student retention.

The wider implications of the analysis indicate the importance of using triggers in EAS that are representative of student behaviour. Regardless of whether the system is designed for student engagement or retention, the triggers need to take valid measurement of the learning environment. The development of applications like the e-motion tool can assist in capturing complex aspects of learning, such as how a student feels towards a particular unit. However, tools also need to generate sufficiently useful information that allow support services and advisors to act on. Underutilisation and biased usage of tools can invalidate the identification process for an EAS, and may increase the rate of misidentification.

9.6 Financial implications of student retention

The results of chapter 8 establish financial benchmarks using 2013 values to ensure parity. A key finding of this study was that there is a casual relationship between the EAS and increased student tuition fees and in turn, retention. Being able to control for *demographic*, *institutional* and *workload* variables through nearest neighbour matching, the estimated value of the EAS was around $4,000 per student. This demonstrates that the program is of significant financial value to the implementing institution. The result demonstrates the magnitude of the link between the EAS and the institution’s revenue function and builds a strong case for institutions globally to move towards implementing an EAS.

A major benefit of benchmarking the value of the EAS is that it enables the case study institution to value changes made to the black box that is student support. It would be expected that new initiatives or major changes to the student support process should be reflected in the valuation of the program overall. The result therefore supports the development of evidence-based approaches to the provision of student support. Future enhancements to the EAS can be measured against the benchmarks provided to capture return on investment estimates of the changes made. This provides a critical decision-making tool for the institution for analysis in the future. Furthermore, it demonstrates the importance of creating valid benchmarks, be it financial or using some other measure, to evaluate the effectiveness of an EAS.

Another finding of the study is that EAS has institution-wide financial benefits. The effect of the EAS within schools varied, but all schools within the institution had a significant positive benefit from the program. Furthermore, the EAS retained a positive value when identifying students beyond their first year of study. This is a critical conclusion, as it demonstrates empirically the need for institution-wide approaches to implementing EAS encompassing all students. Additionally, there has been significant focus on early identification and the development of first year programs (Bevitt, Baldwin, and Calvert, 2010; Jamelske, 2009; Kemlo and Ryan, 2012; Kuh, Cruce, Shoup, Kinzie, and Gonyea, 2008; Marrington et al., 2010; Nelson, Duncan, and Clarke, 2009) to address student retention. However the results indicate that EAS have significant value in identifying students in need of support at all stages of study. The results strongly support EAS being implemented institution-wide, encompassing all students. This also has significant benefits for implementing institutions as it allows a more focused and evidence based approach to the provision of student support.

For the learning analytics community, this study provides significant findings in identifying an approach to valuing EAS. One of the major challenges learning analytic programs have in attaining institution-wide acceptance, implementation and support, is the ability to quantify the benefits of initiatives. While likelihood, OLS and survival analysis approaches yield important technical information on how systems function, interpreting and understanding these results require a degree of statistical literacy. Estimating the effect of an EAS in terms of revenue makes it easier for the business case to be made. Furthermore, the treatment effects approach provides the statistical certainty that comes with causal inference.

The results of this study provide important information that administrators within institutions can use to support learning analytic initiatives. The approach to measuring financial implications of the EAS demonstrate a method that can be used for valuing other initiatives. With more detailed data, such as graduate survey information capturing graduate wage and employment information, it would be possible to undertake full economic analyses on the effect of learning analytic initiatives. This opens up a major avenue for future analysis, allowing evidence based valuation of initiatives. For learning analytics practitioners, the approach allows microeconometric analysis of effects, which can support the business case for an initiative which has positive effects on the learning environment. For institutions, this approach will be paramount in helping inform decision makers on what initiatives to implement to best support students, and the benefits and costs of doing so. For students, this has the flow on potential to ensure that institutions are maximising retention with the best learning support available. Overall, the study is a significant step in supporting the wider adoption of learning analytics and Early Alert Systems.

9.7 Areas for enhancement– research implications

The EAS literature forms an important component of the field of learning analytics literature. Figure 2.4 on page 23 conceptualised the relationship between many variables that are used throughout the learning analytics literature to understand relationships in the learning environment and student retention. One limitation of this study that comes from this framework is the highly interconnected and complex nature of the learning process. For example, factors that affect student admission are also going to have some effect on course design, which in turn can affect how students are identified for support using an EAS. For the estimation of effects on retention, this resulted in the models failing the link-test. This is a statistical test which identifies if a model is sufficiently specified. The results in Appendix G show that whilst many of the models were comprehensive in their estimation of effects, they still didn't contain all the variables required to properly capture the learning process that affects retention. An area for future analysis is to develop an empirical model which can accurately reflect the modern learning process. This then leads into the first area for enhancement.

The first improvement comes from the range and scope of variables available. Despite the significant and expansive data set provided, several key variables could have the potential to enhance the study further. Given that many students are either located off-campus, or move to campus to study, postcode information is important. This would allow for some degree of socioeconomic control of outcomes. Second, data on student admission scores would allow control of prior academic performance to be incorporated into the model. Assignment and examination grades would allow an even finer measurement of academic performance than the overall grade results themselves. As demonstrated with the trigger analysis, student portal access is a significant variable affecting the hazard ratio. Log data has been shown (Gobert, Baker, and Wixon, 2015, p. 53) to provide valuable information in understanding student engagement and in turn, retention. Results through the EAS triggers showed this could be a useful source of information for identifying students at risk of discontinuing their studies. Finally, the actual fees paid by students rather than an estimation of fees would improve the accuracy of the financial estimates in chapter 8. The availability of data is always going to be critical to learning analytic projects. One benefit of this study is that it has demonstrated the level of detail that can be obtained when comprehensive data sets are made available. Estimating student retention effects on a weekly level breaks new ground and shows how important granular data is.

From the data limitations stems the impacts on the statistical approaches used to analyse student retention. The approaches used were not exhaustive of all avenues available to identify the relationship between retention and an EAS. For example, within the treatment effects family of statistical models, propensity score matching is just one approach with many algorithms other than nearest neighbour which can estimate effects. Caliendo and Kopeinig (2008) provide an excellent guide on implementing these causal models which could be used to analyse many other learning analytic initiatives for causal effects. The sequence analysis approaches used by Méndez et al. (2014) linking course design and retention, provided another approach to analysing retention at the course level.

Another key area for enhancement is opening the black box that is student support. Throughout the study, only the beginning and the end of this process was known. How students were contacted, how often, what services were accessed and the level of support given are all unknown. While student privacy and confidentiality are important, the capacity to produce anonymised data sets for analysis is possible. This represents a critical area for future research if different support strategies are ever to be evaluated and quantified in an econometric framework. Furthermore, understanding what occurs within the student support process has important financial implications. Given the limited resources institutions have to implement support programs, it will become increasingly important to evaluate and validate the efficacy of all aspects of student support. This shouldn’t be seen in the light of justifying expenditure on support services, but in the context it is intended, to identify how best to support students with the limited resources that are available.

The final area for enhancement is the provision of benchmarks. A key limitation of this study was having no benchmark estimates of the effects associated with student support prior to the EAS introduction. All data used for this analysis was taken from the start of 2011 to the end of 2013. While it is possible to conclude the system works as far as it impacts on student outcomes, it cannot be concluded that it is an improvement on what it replaced or displaced.

9.8 Reflections of findings and the future of Early Alert Systems

This study has made significant contributions to understanding the relationship between EAS and student retention. It is a complex temporal relationship, but one that has been shown to be positive overall. This study has shown that there are many variables that affect student outcomes, however demographic, institutional, student performance and workload form an important base to EAS design. While there is increasing information collected from the learning environment, a base set of variables are required to control for characteristics of the institution. In the case study institution, only a few of these base characteristics were included in the EAS design. It was shown that only once the base variables were included to control for institutional characteristics, did more of the EAS triggers function as expected. Many of the base variables were also selected on the theoretical expectations of effects. For EAS designers, this highlights the importance of not just selecting all available data, but taking an evidence based approach to EAS design, based on theoretical and empirical understanding.

One area of complexity captured in this study was the temporal affects. Many of the variables used to estimate aspects of student retention were not constant over time. Furthermore, the mathematical functions used to describe the temporal relationships included polynomial functions, inverse logarithmic and discontinuous functions. This demonstrates the need for EAS to have a dynamic approach to identifying students in need of support. The factors that are influencing a student's decision to discontinue one day, will not be the same factors the next day. For student support teams, temporal effects should also be factored in to the advice provided to students. As shown in section 9.1.3, withdrawing from a unit of study only affects a student's decision to discontinue in the first year of study. As such, it is important for institutions to explore what factors affect student retention at different stages of learning. Lessons from econometrics show that while more variables can assist with greater explanatory power in a statistical model, issues around endogeneity, multicollinearity and heteroscedasticity will persist. It will be increasingly important to ensure that the complex EAS design is statistically robust and valid. The field of econometrics has a great deal to offer the learning analytics community in this area.

An interesting area for future study is comparison between EAS. This study provides an important contribution to the learning analytics field, showing how one EAS has affected student retention at a case study institution. Once the field has matured to the point where data sets can be compiled from many institutions, a key question needs to be addressed: which EAS works best for which environment? It is expected that the diversity between institutions themselves will yield many EAS solutions that are optimised for different environments. To what degree will these systems converge in the future through comparative analysis and how will they diverge to capture the variations between the environments in which they are used? To enable cross-institutional analysis of EAS, there also needs to be formalised measures of EAS effects. Given that financial effects are going to be institution and country specific, EAS effect should be measured on a more tangible variable, length of enrolment and changes in the likelihood of outcomes. Using a suite of measures to comprise an EAS performance index may be an important step to maximising EAS performance.

Further analysis is required to determine the value of EAS at larger scale institutions with different student profiles. However, this study provides promising evidence that institution wide EAS initiatives work. This supports the business case for institutions to implement EAS, with potential for positive returns on investment. The results also have important implications for the development and wider implementation of learning analytics within other institutions. In the case of the University of Wisconsin, "learning analytics has not diffused broadly throughout the organisation" (Siemens et al., 2013, p. 13). The results presented here show the benefits of one application of learning analytics and how it assisted an institution in improving retention. Other institutions can build from these results to promote evidence based enhancement of the learning environment. Using treatment effects approaches, causal inferences can be constructed to understand exactly how changing a parameter in the learning environment affect student performance, well-being or other important measures.

One key caveat to EAS valuation is determining the value of the system to the student. The valuation in this instance has taken a strong institutional perspective. However, it is important to also consider the economic impact on students, especially in situations where prolonging the students enrolment may not be in their best interests. Another key caveat to future EAS going forward is asking whether they actually represent fair value for investment. The results of the study suggest yes, with possible causal effects on tuition fees alone being in the thousands of dollars per student identified. In the future it is expected more complex valuations of the systems will show the true economic impact of EAS. It is expected that the wider effects of more students attaining higher degrees will translate to increased future earnings and taxes paid, decreased health expenditure and welfare dependence. It would also capture the positive synergies for the institution in terms of reputation and enhanced student experience. Based on the results of the study, it can be concluded that EAS have a bright future as the foundation for student retention maximisation.

Bibliography

Adams, T., Banks, M., Davis, D., and Dickson, J. (2010). *The Hobsons retention project: Context and factor analysis report.* Melbourne. Tony Adams and Associates.

Aguilar, S., Lonn, S., and Teasley, S. D. (2014). *Perceptions and use of an early warning system during a higher education transition program.* In Proceedings of the Fourth International Conference on Learning Analytics and Knowledge. ACM. 113-117.

Arnold, K. E. (2010). *Signals: Applying Academic Analytics.* Educause Quarterly, 33(1), 1.

Arnold, K. E., and Pistilli, M. D. (2012). *Course signals at Purdue: using learning analytics to increase student success.* In Proceedings of the Second International Conference on Learning Analytics and Knowledge. ACM. 267-270.

Arnold, K. E., Tanes, Z., and King, A. S. (2010). *Administrative perceptions of data-mining software Signals: Promoting student success and retention.* The Journal of Academic Administration in Higher Education, 29-40.

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

Austin, J. T., Yaffee, R. A., and Hinkle, D. E. (1992). *Logistic regression for research in higher education.* Higher Education: Handbook of Theory and Research, 8, 379-410.

Australian Bureau of Statistics. (2013). *Hitting the books: Characteristics of higher education students.* Australian Social Trends - cat. no. 4102.0. Retrieved 10/07/2015, 2015, from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4102.0Main+Features20July+2013#p10

Australian Bureau of Statistics. (2014). *Government Finance Statistics, Education, Australia, 2012-13.* Canberra: Australian Bureau of Statistics, Retrieved from http://www.abs.gov.au/ausstats/abs@.nsf/mf/5518.0.55.001.

Australian Education Network. (2014a). *Groupings of Australian Universities.* Retrieved 05/05/2014, from http://www.australianuniversities.com.au/directory/australian-university-groupings/

Australian Education Network. (2014b). *List of universities in Australia.* Retrieved 09/12/2014 from http://www.australianuniversities.com.au/list/

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., and Graesser, A. C. (2010). *Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments.* International Journal of Human-Computer Studies. 68(4). 223-241.

Baker, R. S., Lindrum, D., Lindrum, M. J., and Perkowski, D. (2015). *Analyzing Early At-Risk Factors in Higher Education e-Learning Courses.* Retrieved 12/09/2015 from http://www.columbia.edu/~rsb2162/2015paper41.pdf

Barber,R. S., and Sharkey, M. (2012). *Course correction: using analytics to predict course success.* In the Proceedings of the Second International Conference on Learning Analytics and Knowledge . ACM. 259-262.

Bean, J. (1980). *Dropouts and turnover: The synthesis of a causal model of student attrition.* Research in Higher Education, 12(2), 155-187.

Bean, J., and Metzner, B. (1985). *A Conceptual Model of Nontraditional Undergraduate Student Attrition.* Review of Educational Research, 55(4), 485-540.

Berry, W. D., and Feldman, S. (1985). *Multiple regression in practice.* Beverly Hills and London: Sage.

Bevitt, D., Baldwin, C., and Calvert, J. (2010). *Intervening early: Attendance and performance monitoring as a trigger for first year support in the biosciences.* Bioscience Education 15(1), 1-14, doi: 10.3108/beej.15.4

Blackboard. (2014, March 13). *Using the Retention Center.* Blackboard Help. Retrieved on 03/07/2015 from https://help.blackboard.com/en-us/Learn/9.1_SP_10_and_SP_11/Instructor/040_Student_Course_Experience/Student_Performance/Using_the_Retention_Center

Bradley, D., Noonan, P., Nugent, H., and Scales, B. (2008). *Review of Australian higher education: final report [Bradley review]*. Canberra. Department of Education, Employment and Workplace Relations.

Brand, J. E., and Halaby, C. N. (2006). *Regression and matching estimates of the effects of elite college attendance on educational and career achievement.* Social Science Research, 35(3), 749-770.

Bryson, A. (2002). *The union membership wage premium: an analysis using propensity score matching.* London. London School of Economics and Political Science. 1-34.

Cabrera, A. F. (1994). *Logistic regression analysis in higher education: An applied perspective.* Higher Education: Handbook of Theory and Research, 10, 225-256.

Cabrera, A. F., Nora, A., Castaneda, M. B., and Hengstler, D. (1992). *The convergence between two theories of college persistence.* Journal of Higher Education, 63(2), 143-164.

Caliendo, M., Hujer, R., and Thomsen, S. (2008). *The Employment Effects of Job Creation Schemes in Germany-A Microeconometric Evaluation.* Advances in Econometrics, 21, 383-430. ISBN: 978-1-84950-523-9

Caliendo, M., and Kopeinig, S. (2008). *Some practical guidance for the implementation of propensity score matching.* Journal of Economic Surveys, 22(1), 31-72.

Campbell, J. (2007). *Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study*. ProQuest LLC.

Campbell, J., DeBlois, P., and Oblinger, D. (2007). *Academic Analytics: a new tool for a new era.* Educause Review, 42(4), 40-57.

Caufield, M. (2013a). *A simple, less mathematical wat to understand the course signals issue.* Hapgood. Retrieved 10/07/2015, from http://hapgood.us/2013/09/26/a-simple-less-mathematical-way-to-understand-the-course-signals-issue/

Caufield, M. (2013b). *Why the course signals math does not add up.* Hapgood. Retrieved 08/12/2014 from http://hapgood.us/2013/09/26/why-the-course-signals-math-does-not-add-up/

Cox, D. R. (1972). *Regression models and life-tables.* Journal of the Royal Statistical Society. Series B (Methodological), 187-220.

Cox, D. R. and Oakes, D (1984). *Analysis of survival data*. Chapman and Hall, London ; New York

Department of Education. (2014a). *2013 Student Summary.* Department of Education. Canberra: Retrieved 20/06/2015 from https://education.gov.au/selected-higher-education-statistics-2013-student-data.

Department of Education. (2014b). *Selected Higher Education Statistics – 2013 Student Data, All Students Table.* Department of Education. Canberra. Retrieved 20/06/2015 from http://docs.education.gov.au/system/files/doc/other/2013allstudents_0.xls.

Department of Industry. (2012). *Appendix 4 - attrition, progress and retention.* Canberra. Department of Industry. Retrieved 04/8/2013 from: http://www.innovation.gov.au/highereducation/HigherEducationStatistics/StatisticsPublications/Pages/Students12FullYear.aspx

Department of Industry. (2013). *Summary of the 2012 full year higher education student statistics.* Canberra: Department of Industry.

Desire2Learn. (2013). *Desire2Learn.* Desire2Learn. Retrieved from http://www.desire2learn.com/

DesJardins, S. L. (2003). *Event history methods: conceptual issues and an application to student departure from college.* Higher Education: Handbook of Theory and Research, Springer. 421-471

DesJardins, S. L., Ahlburg, D. A., and McCall, B. P. (1999). *An event history model of student departure.* Economics of Education Review, 18(3), 375-390.

DesJardins, S. L., Ahlburg, D. A., and McCall, B. P.. (2002). *A Temporal Investigation of Factors Related to Timely Degree Completion*. The Journal of Higher Education, 73(5), 555–581. Retrieved from http://www.jstor.org/stable/1558433

Dey, E. L., and Astin, A. W. (1993). *Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression.* Research in Higher Education, 34(5), 569-581.

Essa, A., and Ayad, H. (2012). *Student success system: risk analytics and data visualization using ensembles of predictive models.* Proceedings of the Second International Conference on Learning Analytics and Knowledge. . ACM. 158-161.

Ferguson, R., Clow, D., Macfadyen, L., Tynan, B., Dawson, S., and Alexander, S. (2015). *Setting learning analytics in context: overcoming the barriers to large-scale adoption.* Journal of Learning Analytics, 1(3), 120-144.

Fike, D. S., and Fike, R. (2008). *Predictors of first-year student retention in the community college.* Community College Review, 36(2), 68-88.

Gobert, J. D., Baker, R. S., and Wixon, M. B. (2015). *Operationalizing and detecting disengagement within online science microworlds.* Educational Psychologist, 50(1), 43-57.

Hinkle, D., Austin, J., and McLaughlin, G. (1989). *Log-linear models: applications in higher education research.* Higher Education: Handbook of Theory and Research, 5, 323-353.

Ishitani, T. T. (2003). *A longtitudinal approach to assessing attrition behaviour amoung first-generation students.* Research in Higher Education, 44(4), 433-449.

Ishitani, T. T.. (2006). *Studying Attrition and Degree Completion Behavior among First-Generation College Students in the United States.* The Journal of Higher Education, 77(5), 861–885. Retrieved from http://www.jstor.org/stable/3838790

Ishitani, T. T., and DesJardins, S. L. (2003). *A longtitudinal investigation of dropout from college in the United States.* Journal of College Student Retention. 4(2). 173 - 201.

Jamelske, E. (2009). *Measuring the impact of a university first-year experience program on student gpa and retention. higher education*, 57(3), 373-391. doi: 10.2307/40269128

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., and Baron, J. D. (2014). *Early alert of academically at-risk students: An open source analytics initiative.* Journal of Learning Analytics. 1(1). 6-47.

Jones-White, D. R., Radcliffe, P. M., Huesman Jr., L. R., and Kellogg, J. P. (2010). *Redefining student success: Applying different multinomial regression techniques for the study of student graduation across institutions of higher education.* Higher Education. 51(2). 154–174.

Jones, K. M., Thomson, J., and Arnold, K. (2014). *Questions of Data Ownership on-campus.* Educause Review. Accessible online at http://www. educause. edu/ero/article/questions-data-ownership-campus.

Kemlo, L., and Ryan, J. (2012). *Student Transition: Implementing an effective first year experience model for the common core courses in the College of Business.* Melbourne: RMIT University. 1-5. Retrieved from http://www.fyhe.com.au/past_papers/papers12/Papers/12B.pdf

Khamis, C., and Kiernan, F. (2013). *Track and Connect: A tailored individual support program for at-risk students at the University of Sydney.* Sydney: University of Sydney. Retrieved from http://fyhe.com.au/past_papers/papers13/6F.pdf

Krumm, A. E., Waddington, R. J., Teasley, S. D., and Lonn, S. (2014). *A learning management system-based early warning system for academic advising in undergraduate engineering*. In Learning Analytics. Springer New York. 103-119.

Kuh, G., Cruce, T., Shoup, R., Kinzie, J., and Gonyea, R. (2008). *Unmasking the effects of student engagement on first-year college grades and persistence.* The Journal of Higher Education, 79(5), 540 - 563.

Kuh, G., Kinzie, J., Buckley, J., Bridges, B., and Hayek, J. (2007). *Piecing together the student success puzzle: Research, propositions and recommendations.* ASHE Higher Education Report, 32(5).

Liao, T. F. (1994) *Interpreting probability models - logit, probit and other generalized linear models.* No. 101. In Lewis-Beck, M. S. (Series Ed.), Quantitative Applications in the Social Sciences (101). Thousand Oaks, California. Sage Publications.

Lin, D. Y., and Wei, L.-J. (1989). *The robust inference for the Cox proportional hazards model.* Journal of the American Statistical Association, 84(408), 1074-1078.

Lin, T. C., Yu, W. W. C., and Chen, Y. C. (2012). *Determinants and probability prediction of college student retention: new evidence from the Probit model.* International Journal of Education Economics and Development, 3(3), 217-236.

Marrington, A. D., Nelson, K. J., and Clarke, J. A. (2010). *An economic case for systematic student monitoring and intervention in the first year in higher education.* Paper presented at the Proceedings of 13th Pacific Rim First Year in Higher Education Conference, Adelaide. Retieved from http://eprints.qut.edu.au/33231/1/c33231.pdf

Méndez, G., Ochoa, X., and Chiluiza, K. (2014). *Techniques for data-driven curriculum analysis.* Proceedings of the Fourth International Conference on Learning Analytics and Knowledge. ACM. pp. 148-157

Milem, J. F., and Berger, J. B. (1997). *A modified model of college student persistence: exploring the relationship between Astin's theory of involvement and Tinto's theory of student departure.* Journal of College Student Developement, 387 - 400.

Nelson, K. J., and Creagh, T. A. (2013). *A good practice guide: safegaurding student learning engagement.* Brisbane: Queensland University of Technology. Retrieved 08/06/2015 from http://eprints.qut.edu.au/59189/1/LTU_Good-practice-guide_eBook_20130320.pdf

Nelson, K. J., Duncan, M. E., and Clarke, J. A. (2009). *Student success: The identification and support of first year university students at risk of attrition.* Studies in Learning, Evaluation, Innovation and Development, 6(1), 1-15.

Organisation for Economic Co-operation and Development. (2011). *Education at a Glance 2010: OECD Indicators.* OECD Pub. ISBN 978-92-64-07566-5

Organisation for Economic Co-operation and Development. (2014). *Education at a Glance 2013: OECD Indicators.* OECD Pub. ISBN 978-92-64-20105-7

Organisation for Economic Co-operation and Development. (2015). *Education at a Glance 2014: OECD Indicators.* OECD Pub. ISBN 978-92-64-21505-4

Olsen, A. (2008). *Staying the course: retention and attrition in Australian universities.* Strategy Policy and Research in Education Limited. Retrieved from http://www.spre.com.au/download/AUIDFRetentionResultsFindings.pdf

Pascarella, E., and Terenzini, P. (1980). *Predicting Freshman Persistence and Voluntary Dropout Decisions from a Theoretical Model.* The Journal of Higher Education, 51(1), 60-75.

Peng, C.-Y. J., So, T.-S. H., Stage, F. K., and John, E. P. S. (2002). *The use and interpretation of logistic regression in higher education journals: 1988–1999*. Research in Higher Education. 43(3). 259-293.

Purdue University Information Technology (2013) *Course Signals* Retrieved 11/02/2015 from http://www.itap.purdue.edu/learning/tools/signals/

Radcliffe, P., Huesman, R., and Kellogg, J. (2006). *Modeling the incidence and timing of student attrition: A survival analysis approach to retention analysis.* Paper presented at the annual meeting of the Association for Institutional Research in the Upper Midwest (AIRUM).

Romero-Zaldivar, V.-A., Pardo, A., Burgos, D., and Kloos, C. D. (2012). *Monitoring student progress using virtual appliances: A case study*. Computers and Education, 58(4), 1058-1067.

Rubin, D. B. (1974). *Estimating causal effects of treatments in randomized and nonrandomized studies.* Journal of educational Psychology, 66(5), 688-701.

Schudde, L. T. (2011). *The causal effect of campus residency on college student retention*. The Review of Higher Education. 34(4). 581-610.

Siemens, G. (2012, April). *Learning analytics: envisioning a research discipline and a domain of practice.* In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. ACM. 4-8.

Siemens, G. (2013). *Learning analytics: the emergence of a discipline*. American Behavioral Scientist. 57(10). 1380-1400. DOI: 10.1177/0002764213498851

Siemens, G., Dawson, S., and Lynch, G. (2013). *Improving the quality and productivity of the higher education sector: Policy and strategy for systems-level deployment of learning analytics.* Sydney. Office for Learning and Teaching. Retrieved from http://www.olt.gov.au/system/files/resources/SoLAR_Report_2014.pdf

Simons, J. M. (2011). *A National Study of Student Early Alert Models at Four-Year Institutions of Higher Education*. ProQuest LLC.

Singell, L. D., and Waddell, G. R. (2010). *Modeling retention at a large public university: can at-risk students be identified early enough to treat?* Research in Higher Education. 51(6). 546-572.

Society of Learning Analytics Research. (2012). *About SoLAR*. SoLAR - Society of Learning Analytics Research. Retrieved August/10/2013, from http://www.solaresearch.org/mission/about/

Stratton, L. S., O'Toole, D. M., and Wetzel, J. N. (2005). *A multinomial logit model of college stopout and dropout behavior.* Economics of Education Review. 27(3). 319-331.

Stratton, L. S., O'Toole, D. M., and Wetzel, J. N. (2007). *Are the factors affecting dropout behaviour related to initial enrolment intensity for college undergraduates?* Research in higher education. 48(4). 453 - 485.

Straumsheim, C. (2013). *Mixed Signals*. Inside Higher Ed. Retrieved from http://www.insidehighered.com/news/2013/11/06/researchers-cast-doubt-about-early-warning-systems-effect-retention

Tally, S. (2009). *Signals tells students how they're doing even before the test*. Purdue University news. Retrieved 23/11/2014 from http://www.purdue.edu/uns/x/2009b/090827ArnoldSignals.html

Tanes, Z., Arnold, K. E., King, A. S., and Remnet, M. A. (2011). *Using Signals for appropriate feedback: Perceptions and practices.* Computers and Education. 57(4). 2414-2422.

Tinto, V. (1975). *Dropout from higher education: A theoretical synthesis of recent research.* Review of Educational Research. 45(1). 89-125.

Tinto, V. (2006). *Research and practice of student retention: what next?* Journal of college student retention. 8(1). 1-19.

Tynan, B., and Buckingham Shum, S. (2013). *Designing systematic learning analytics at the Open University.* SoLAR Open Course, Strategy and policy for systematic learning. Retrieved from http://www.slideshare.net/sbs/designing-systemic-learning-analytics-at-the-open-university

University of New England. (2012a). *University of New England Annual Report 2011* Armidale. University of New England.

Unviersity of New England. (2012b). *Assessment Policy.* Armidale. University of New England. Retrieved 17/07/2014. 2014. from http://www.une.edu.au/__data/assets/pdf_file/0008/37439/assessment-policy-2012.pdf

University of New England. (2013). *International Prospectus.* Armidale. University of New England .

University of New England. (2014a). *Amalgamation and its aftermath.* Armidale. University of New England Retrieved 09/12/2014 from http://www.une.edu.au/about-une/a-world-of-learning/the-une-story

University of New England. (2014b). *Student Fees.* Armidale. University of New England Retrieved 25/05/2014, from http://www.une.edu.au/study/fees/student-fees

University of New England. (2014c). *Description table of AWE triggers.* Armidale. University of New England. Unpublished raw data. Retrieved 14/03/2014.

Waddington, R. J., and Nam, S. (2014, March). *Practice exams make perfect: incorporating course resource use into an early warning system.* In Proceedings of the Fourth International Conference on Learning Analytics And Knowledge. ACM. 188-192.

Wolff, A., Zdrahal, Z., Nikolov, A., and Pantucek, M. (2013, April). *Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment.* In Proceedings of the Third International Conference on Learning Analytics and Knowledge. ACM. 145-149.

Zaki, M. J. (2001). *SPADE: An efficient algorithm for mining frequent sequences.* Machine Learning, 42(1-2), 31-60.

Appendix A: University Grades

The grading systems used by UNE are defined in the assessment policy of the university. An extract of section 21 is provided from the most recent assessment policy (Unviersity of New England, 2012, p. 15)

HD - High Distinction

Equivalent to Honours Class I (H1) in intent, 7 on the GPA scale. Excellent performance indicating complete and comprehensive understanding and/or application of the subject matter; achieves all basic and higher order in tended unit objectives and graduate attributes linked to the assessment tasks; minimal or no errors of fact, omission and/or application present; clear and unambiguous evidence of possession of a very high level of required skills; demonstrated very high level of interpretive and/or analytical ability and intellectual initiative; very high level of competence.

(Numerical conversion: scores and/or aggregate marks of 85% or above)

D - Distinction

Equivalent to Honours Class II, Division 1 (H2A) in intent, 6 on the GPA scale. Very good performance indicating reasonably complete and comprehensive understanding and/or application of the subject matter; achieves all basic and most higher-order unit objectives and graduate attributes linked to the assessment tasks; some minor flaws; clear and unambiguous evidence of possession of a high level of required skills; demonstrated high level of interpretive and/or analytical ability and intellectual initiative; high level of competence.

(Numerical conversion: scores and/or aggregate marks between 75% and 84%)

C - Credit

Equivalent to Honours Class II, Division 2 (H2B) in intent, 5 on the GPA scale. Good performance indicating reasonable and well-rounded understanding and/or application of the subject matter; achieves all basic but only a few higher-order intended unit objectives and graduate attributes linked to the tasks; a few more serious flaws or several minor ones; clear and unambiguous evidence of possession of a reasonable level of most required skills; demonstrated reasonable level of interpretive and/or analytical ability and intellectual initiative; reasonable level of competence.

(Numerical conversion: scores and/or aggregate marks between 65% and 74%)

P - Pass

Equivalent to Honours Class III (H3) in intent, 4 on the GPA scale Satisfactory performance indicating adequate but incomplete or less well-rounded understanding and/or application of the subject matter; achieves many basic but very few or none of the higher-order intended unit objectives and graduate attributes linked to the assessment tasks; several serious flaws or many minor ones; clear and unambiguous evidence of possession of an adequate level of an acceptable number of required skills; demonstrated adequate level of interpretive and/or analytical ability and intellectual initiative; adequate level of competence.

(Numerical conversion: scores and/or aggregate marks between 50% to 64%)

N - Fail

0 on the GPA scale. Unsatisfactory performance indicating inadequate and insufficient understanding and/or application of the subject matter; achieves few or none of the basic and higher order intended unit objectives and graduate attributes linked to the assessment tasks; numerous substantive errors of fact, omission and/or application present; clear and unambiguous evidence of non-possession of most or all required skills; insufficiently demonstrated level of interpretive and/or analytical ability and intellectual initiative; fails to address the specific criteria; inadequate level of competence.

(Numerical convers ion: scores and/or aggregate marks of less than 50%)

NC - Compulsory Fail

Failed an assessment component that must be passed in order to pass the unit. This grade is used when an assessment task, such as a final examination, that must be passed in order to

pass the unit (as detailed in the Unit Requirements) has not been passed (resulting in a fail in the unit), but where the overall mark is 50% or higher.

NI – Fail did not satisfy unit requirements

(0 on the GPA scale) One or more mandatory requirements for the completion of the unit (as detailed in the unit requirements) were not fulfilled.

S or US Satisfactory or Unsatisfactory

(S = 4, US = 0 on the GPA scale) In some units, the grading system is organised on a satisfactory/unsatisfactory (pass/fail) basis. When this grading system is used the appropriate interpretive descriptors to apply will be those for the grade of at least Pass or Fail.

W - Withdrawn

The student withdrew from the unit without academic penalty.

Appendix B: EAS Triggers

Trigger	Description
1	Student admitted through alternate entry pathway
2	Student is Aboriginal or Torres Strait Islander
3	Unit is a currently high attrition unit
4	Unit is a historically high attrition unit
5	Student is a college resident
6	Student registered "Happy" in e-Motion
7	Student registered "I do not want to say" in e-Motion
8	Student registered "Neutral" in e-Motion
9	Student registered "Unhappy" in e-Motion
10	Student registered "Very Unhappy" in e-Motion
11	Student has high e-Reserve usage inactivity (31-40 days)
12	Student has low e-Reserve usage inactivity (10-20 days)
13	Student has medium e-Reserve usage inactivity (21-30 days)
14	Student has very high e-Reserve usage inactivity (41+ days)
15	Student has been granted 1-2 assignment extension in current teaching period
16	Student has been granted more than 2 assignment extension in current teaching period
17	Student has submitted 1-2 assignments late in current teaching period
18	Student has submitted more than 2 assignments late in current teaching period
19	Student enrolment has involved > double their number of currently enrolled units in current teaching period, post start of teaching
20	Student has appeared in High Risk Category in a previous teaching period
21	Student is an international student
22	Student has no prior enrolment at UNE
23	Student is enrolled in 5 or more units in a single teaching period
24	Student was previously enrolled in a pathways enabling course
25	Student has been flagged for contact by the retention team in current teaching period
26	Student has been flagged for contact by the retention team in a previous teaching period
27	Student is carrying over Special Extension of Time (SET) exams from a previous teaching period, and is enrolled in current teaching period
28	Student has high portal usage inactivity (31-40 days)
29	Student has low portal usage inactivity (10-20 days)
30	Student has medium portal usage inactivity (21-30 days)
31	Student has very high portal usage inactivity (41+ days)
32	Student was enrolled in the Teacher Enabling course
33	Student received a fail in a unit in a prior teaching period
34	Student received a fail incomplete in a unit in a prior teaching period

Appendix C: Likelihood Analysis Statistical Output

	Model 5.1				Model 5.2			
Number of observations	16124				16124			
LR χ 2	7431.43				7500.5			
Prob > χ 2	0				0			
Pseudo R2	0.2747				0.2772			
	Probability of Outcome				Probability of Outcome			
	Discontinue		Complete		Discontinue		Complete	
Variable	RRR	Std. Error	RRR	Std. Error	RRR	Std. Error	RRR	Std. Error
Constant	1.253[c]	0.146	0.007[a]	0.002	1.273[b]	0.148	0.006[a]	0.002
Gender	1.104[b]	0.048	1.203[b]	0.109	1.104[b]	0.048	1.214[b]	0.111
Age	0.960[a]	0.007	1.062[a]	0.017	0.96[a]	0.007	1.053[a]	0.017
Age^2	1.001[a]	0	0.999[a]	0	1.001[a]	0	0.999[a]	0
ATSI	0.789[c]	0.096	0.796	0.213	0.791[c]	0.096	0.8	0.217
Domestic Fee	2.138[b]	0.805	0.497	0.41	2.14[b]	0.808	0.488	0.412
International Fee	0.421[a]	0.072	4.161[a]	0.717	0.413[a]	0.071	4.682[a]	0.816
Prior Studies	3.010[a]	0.175	6.526[a]	0.589	3.011[a]	0.176	6.48[a]	0.586
On-campus	1.495[a]	0.101	0.531[a]	0.066	1.479[a]	0.101	0.524[a]	0.066
Diploma	1.254	0.213	7.813[a]	2.254	1.249	0.212	7.76[a]	2.251
Advanced Diploma	1.491[a]	0.146	4.889[a]	0.982	1.48[a]	0.145	4.836[a]	0.973
Bachelors (Graduate)	0.672[a]	0.063	7.436[a]	1.396	0.668[a]	0.062	7.016[a]	1.328
Bachelors (Honors)	0.385[a]	0.063	33.491[a]	6.344	0.386[a]	0.063	34.456[a]	6.562
School 1	1.196[c]	0.114	0.146[a]	0.029	1.203[c]	0.115	0.133[a]	0.027
School 2	0.995	0.088	0.444[a]	0.073	0.996	0.088	0.444[a]	0.073
School 3	1.300[a]	0.124	0.383[a]	0.07	1.295[a]	0.123	0.379[a]	0.07
School 4	1.250[b]	0.111	0.344[a]	0.059	1.244[b]	0.111	0.338[a]	0.058
School 5	1.219[b]	0.121	0.031[a]	0.009	1.216[b]	0.121	0.032[a]	0.009
School 6	0.966	0.083	0.341[a]	0.055	0.966	0.083	0.334[a]	0.054
School 7	0.909	0.51	1.068	0.883	0.862	0.484	1.232	1.024
School 8	1.407[a]	0.161	0.852	0.143	1.411[a]	0.162	0.734[c]	0.126
School 9	0.908	0.102	0.518[a]	0.108	0.904	0.102	0.512[a]	0.107

Variable	Model 5.1				Model 5.2			
	Discontinue		Complete		Discontinue		Complete	
	RRR	Std. Error	RRR	Std. Error	RRR	Std. Error	RRR	Std. Error
Withdrawn	0.739[a]	0.015	0.488[a]	0.047	0.729[a]	0.015	0.509[a]	0.049
Withdrawn Early	0.873[a]	0.014	0.592[a]	0.033	0.869[a]	0.014	0.596[a]	0.033
Fail Incomplete	1.118[a]	0.019	0.57[a]	0.068	1.099[a]	0.02	0.612[a]	0.073
Fail	0.941[b]	0.023	0.553[a]	0.039	0.925[a]	0.024	0.601[a]	0.043
Pass	0.705[a]	0.012	1.141[a]	0.021	0.697[a]	0.012	1.188[a]	0.023
Credit	0.722[a]	0.013	1.123[a]	0.019	0.716[a]	0.013	1.16[a]	0.021
Distinction	0.707[a]	0.013	1.094[a]	0.018	0.701[a]	0.013	1.135[a]	0.019
High Distinction	0.725[a]	0.017	1.134[a]	0.021	0.719[a]	0.017	1.173[a]	0.022
Other Grade	0.545[a]	0.05	0.716[a]	0.047	0.537[a]	0.049	0.745[a]	0.049
Workload	1.334[a]	0.036	2.736[a]	0.167	1.332[a]	0.037	2.854[a]	0.175
EAS Identified	1.064	0.052	0.705[a]	0.084				
Low Severity					1.054	0.053	0.752[b]	0.09
Medium Severity					1.155[c]	0.087	0.441[a]	0.071
High Severity					1.538[a]	0.171	0.305[a]	0.059
Very High Severity					1.141	0.157	0.202[a]	0.047

Appendix D: OLS Regression Statistical Output

	Model 6.1	Model 6.2
Number of observations	16,124	16,124
F(32, 16091)	1693.36	1709.39
Prob > F	0	0
R-squared	0.807	0.8102
Root MSE	18.837	18.68

Variable	Model 6.1		Model 6.2	
	Coefficient	Robust Standard Error	Coefficient	Robust Standard Error
Gender	0.319	0.327	0.332	0.324
Age	0.225[b]	0.092	0.26[a]	0.092
Age^2	-0.002	0.001	-0.002[c]	0.001
ATSI	0.892	0.823	0.597	0.816
Domestic Fee	9.665[b]	4.572	9.553[b]	4.593
International Fee	3.7[a]	0.694	3.681[a]	0.692
Prior Studies	6.4[a]	0.432	6.388[a]	0.428
On-campus	-0.819[b]	0.409	-1.393[a]	0.406
Diploma	2.95[c]	1.576	3.301[b]	1.547
Advanced Diploma	9.005[a]	0.988	9.201[a]	0.983
Bachelors (Graduate)	1.895[b]	0.745	1.804[b]	0.741
Bachelors (Honors)	5.984[a]	1.073	6.356[a]	1.066
School 1	-1.668[b]	0.727	-1.9[a]	0.723
School 2	-1.622[b]	0.67	-1.525[b]	0.665
School 3	-1.113[c]	0.669	-1.214[c]	0.663
School 4	-3.139[a]	0.642	-3.085[a]	0.635
School 5	-3.561[a]	0.754	-3.535[a]	0.746
School 6	-2.656[a]	0.64	-2.659[a]	0.634
School 7	26.011[a]	2.28	23.987[a]	2.226
School 8	-0.953	0.752	-1.263[c]	0.75
School 9	1.286	0.89	1.172	0.883

Variable	Model 6.1		Model 6.2	
	Coefficient	Robust Standard Error	Coefficient	Robust Standard Error
Withdrawn Early	0.128[a]	0.003	0.127[a]	0.003
Fail Incomplete	0.1[a]	0.003	0.1[a]	0.003
Fail	0.087[a]	0.003	0.09[a]	0.003
Pass	0.078[a]	0.001	0.079[a]	0.001
Credit	0.081[a]	0.002	0.081[a]	0.002
Distinction	0.087[a]	0.002	0.087[a]	0.002
High Distinction	0.099[a]	0.003	0.099[a]	0.003
Other	0.072[a]	0.004	0.077[a]	0.004
Workload	-31.091[a]	0.397	-31.31[a]	0.396
EAS Identified	8.956[a]	0.441	-	-
Low EAS severity	-	-	8.932[a]	0.441
Medium EAS severity	-	-	13.12[a]	0.512
High EAS severity	-	-	9.384[a]	0.668
Very High EAS severity	-	-	4.988[a]	0.808

Appendix E: Survival Analysis Statistical Output

Short Run		Enduring effects		Long Run	
LR $\chi^2(50)$	7646.73	LR $\chi^2(49)$	7603.68	LR $\chi^2(51)$	7940.7
Prob > χ^2	0	Prob > χ^2	0	Prob > χ^2	0

PH Test		PH Test		PH Test	
$\chi^2(50)$	39.12	$\chi^2(49)$	37.51	$\chi^2(51)$	39.99
Prob > χ^2	0.8667	Prob > χ^2	0.8843	Prob > χ^2	0.8672

	Short Run		Enduring effects		Long Run	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
Gender	1.076[b]	0.034	1.076[b]	0.034	1.076[b]	0.034
Age	0.969[a]	0.007	0.969[a]	0.007	0.965[a]	0.007
Age Squared	1[a]	0	1[a]	0	1[a]	0
ATSI	0.837[b]	0.07	0.836[b]	0.07	0.858[c]	0.072
Domestic Fee	0.605	0.338	0.6	0.335	0.507	0.281
Domestic Fee x t	1.01	0.007	1.01	0.007	1.012[c]	0.007
International Fee	0.194[a]	0.057	0.195[a]	0.058	0.204[a]	0.061
International Fee x t^2	1[a]	0	1[a]	0	1[a]	0
International Fee x t^3	1[b]	0	1[b]	0	1[b]	0
Prior Studies	0.587[a]	0.098	0.579[a]	0.097	0.617[a]	0.103
Prior Studies x ln(t)	1.133[a]	0.049	1.136[a]	0.049	1.122[a]	0.048
On-campus	1.175[a]	0.055	1.183[a]	0.055	1.187[a]	0.056
On-campus x t	0.91[a]	0.013	0.912[a]	0.013	0.939[a]	0.013
Diploma	1.022	0.114	1.021	0.114	0.979	0.11
Advanced Diploma	0.783[b]	0.093	0.778[b]	0.092	0.653[a]	0.077
Advanced Diploma x t	1.003	0.002	1.003[c]	0.002	1.005[a]	0.002
Bachelors (Graduate)	0.74[a]	0.051	0.739[a]	0.051	0.737[a]	0.05
Bachelors (Honours)	0.619[a]	0.078	0.618[a]	0.077	0.595[a]	0.074

	Short Run		Enduring effects		Long Run	
	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.	Hazard Ratio	Std. Err.
School 1	1.127[c]	0.078	1.126[c]	0.078	1.103	0.076
School 2	1.272[a]	0.091	1.273[a]	0.091	1.264[a]	0.09
School 2 x 1/t	0.006[a]	0.006	0.006[a]	0.006	0.006[a]	0.005
School 3	1.181[b]	0.08	1.183[b]	0.08	1.156[b]	0.078
School 4	1.297[a]	0.082	1.298[a]	0.082	1.296[a]	0.082
School 5	1.295[a]	0.09	1.3[a]	0.09	1.32[a]	0.091
School 6	0.86[c]	0.07	0.859[c]	0.07	0.855[c]	0.069
School 6 x t	1.004[a]	0.001	1.004[a]	0.001	1.004[a]	0.001
School 7	0.778	0.282	0.799	0.289	0.833	0.301
School 8	1.045	0.083	1.044	0.083	1.031	0.082
School 9	0.958	0.077	0.957	0.077	0.924	0.074
Early Alert System	2.067[a]	0.214	1.15[a]	0.041	0.533[a]	0.049
Early Alert System x t	0.994[a]	0.002	-	-	-	-
Early Alert System x ln(t) (for t<=18)	-	-	-	-	0.685[a]	0.039
Early Alert System x ln(t - 18) (for t > 18)	-	-	-	-	1.128[a]	0.03

Appendix F: Survival Analysis Assumptions Tests In Detail

Base Short-Run Model – Proportional Hazards Test

	rho	chi2	df	Prob>chi2
Gender	-0.00663	0.23	1	0.6329
Age	0.00294	0.04	1	0.8351
Age Squared	-0.00955	0.45	1	0.5026
ATSI	0.01818	1.67	1	0.1957
Domestic Fee	0.00422	0.09	1	0.7676
Domestic Fee x t	-0.00472	0.11	1	0.7436
International Fee	-0.01273	0.92	1	0.3385
International Fee x t^2	0.01486	1.2	1	0.2724
International Fee x t^3	-0.0143	1.09	1	0.2969
Prior Studies	0.02341	2.65	1	0.1038
Prior Studies x ln(t)	-0.02556	3.15	1	0.0762
On-campus	0.00649	0.21	1	0.6486
On-campus x t	0.00175	0.02	1	0.8886
Diploma	-0.00408	0.09	1	0.7688
Advanced Diploma	0.01863	1.73	1	0.1883
Advanced Diploma x t	-0.01452	1.03	1	0.3107
Bachelors (Graduate)	-0.00409	0.08	1	0.7753
Bachelors (Honors)	0.00976	0.49	1	0.4857
School 1	0.00969	0.48	1	0.4898
School 2	-0.00967	0.48	1	0.4894
School 2 x 1/t	0.00897	0.4	1	0.5287
School 3	-0.00141	0.01	1	0.9188
School 4	-0.00963	0.47	1	0.491
School 5	0.00571	0.16	1	0.6881
School 6	-0.00964	0.48	1	0.4883
School 6 x t	0.016	1.32	1	0.2499
School 7	0.00097	0	1	0.9439
School 8	0.01186	0.72	1	0.3957
School 9	-0.00682	0.24	1	0.6212
Withdrawn	-0.01199	0.62	1	0.4294
Withdrawn x t	0.01366	0.78	1	0.3768
Withdrawn x t^2	-0.01358	0.72	1	0.3977
Withdrawn Early	-0.0047	0.11	1	0.7378
Withdrawn Early x t	0.00714	0.26	1	0.6067
Withdrawn Early x t^2	-0.00845	0.38	1	0.5362
Fail Incomplete	0.00451	0.1	1	0.7532
Fail Incomplete x t	-0.02504	2.82	1	0.0934
Fail	-0.01159	0.76	1	0.3834
Pass	-0.00366	0.07	1	0.7926
Credit	0.00225	0.03	1	0.8692
Distinction	0.00424	0.1	1	0.7556
High Distinction	0.01221	0.86	1	0.3539
Other	-0.00247	0.03	1	0.8724

Continued from previous page

	rho	chi2	df	Prob>chi2
Inactive	-0.02321	2.84	1	0.0922
Inactive x t (for t <= 16)	0.00311	0.02	1	0.8817
Inactive x 1/ln(t) (for t > 16)	0.02099	1.69	1	0.1939
Part-time	0.00191	0.02	1	0.891
Part-time x ln(t)	-0.00375	0.07	1	0.7878
Early Alert System	0.00099	0.01	1	0.9421
Early Alert System x t	-0.0053	0.15	1	0.7031
global test		39.12	50	0.8667

Base Enduring Model – Proportional Hazards Test

	rho	chi2	df	Prob>chi2
Gender	-0.00648	0.22	1	0.6404
Age	0.00381	0.07	1	0.7879
Age Squared	-0.01018	0.51	1	0.475
ATSI	0.01823	1.68	1	0.1943
Domestic Fee	0.00421	0.09	1	0.7682
Domestic Fee x t	-0.00484	0.11	1	0.7374
International Fee	-0.01254	0.89	1	0.3458
International Fee x t^2	0.01502	1.23	1	0.2676
International Fee x t^3	-0.01448	1.11	1	0.2918
Prior Studies	0.02358	2.69	1	0.1013
Prior Studies x ln(t)	-0.02589	3.23	1	0.0724
On-campus	0.00465	0.11	1	0.7437
On-campus x t	0.0023	0.03	1	0.8548
Diploma	-0.00391	0.08	1	0.7783
Advanced Diploma	0.01902	1.8	1	0.1798
Advanced Diploma x t	-0.01567	1.19	1	0.2749
Bachelors (Graduate)	-0.00419	0.09	1	0.7697
Bachelors (Honors)	0.01016	0.53	1	0.4685
School 1	0.00991	0.5	1	0.4798
School 2	-0.00965	0.48	1	0.49
School 2 x 1/t	0.0089	0.39	1	0.5316
School 3	-0.0014	0.01	1	0.9196
School 4	-0.00939	0.45	1	0.5022
School 5	0.00544	0.15	1	0.7023
School 6	-0.00912	0.43	1	0.5119
School 6 x t	0.01507	1.17	1	0.2785
School 7	0.00142	0.01	1	0.9183
School 8	0.01236	0.78	1	0.3763
School 9	-0.00642	0.22	1	0.6418

	rho	chi2	df	Prob>chi2
Withdrawn	-0.01145	0.57	1	0.4509
Withdrawn x t	0.01304	0.71	1	0.3995
Withdrawn x t^2	-0.0129	0.65	1	0.4216
Withdrawn Early	-0.00436	0.1	1	0.7565
Withdrawn Early x t	0.00651	0.22	1	0.6396
Withdrawn Early x t^2	-0.00778	0.32	1	0.5695
Fail Incomplete	0.00432	0.09	1	0.7635
Fail Incomplete x t	-0.02503	2.81	1	0.0937
Fail	-0.0121	0.83	1	0.3622
Pass	-0.00381	0.08	1	0.7842
Credit	0.00222	0.03	1	0.8708
Distinction	0.00395	0.08	1	0.7722
High Distinction	0.01232	0.88	1	0.3492
Other	-0.00349	0.05	1	0.8214
Inactive	-0.02022	2.12	1	0.1458
Inactive x t (for t <= 16)	0.00262	0.02	1	0.9002
Inactive x 1/ln(t) (for t > 16)	0.0178	1.19	1	0.2763
Part-time	0.00165	0.01	1	0.9061
Part-time x ln(t)	-0.00354	0.06	1	0.7995
Early Alert System	-0.01314	0.47	1	0.4948
global test		37.51	49	0.8843

Base Long-Run Model – Proportional Hazards Test

	rho	chi2	df	Prob>chi2
Gender	-0.00592	0.18	1	0.6696
Age	0.01043	0.54	1	0.4614
Age Squared	-0.01419	0.99	1	0.3199
ATSI	0.01446	1.06	1	0.3031
Domestic Fee	0.00579	0.16	1	0.6877
Domestic Fee x t	-0.00782	0.29	1	0.5924
International Fee	-0.01372	1.07	1	0.3019
International Fee x t^2	0.01779	1.7	1	0.1921
International Fee x t^3	-0.01715	1.53	1	0.2164
Prior Studies	0.02112	2.16	1	0.1413
Prior Studies x ln(t)	-0.0241	2.81	1	0.0939
On-campus	0.00885	0.39	1	0.5334
On-campus x t	0.0031	0.07	1	0.7974
Diploma	0.00124	0.01	1	0.9289
Advanced Diploma	0.02728	3.66	1	0.0559
Advanced Diploma x t	-0.03034	4.36	1	0.0369
Bachelors (Graduate)	-0.00274	0.04	1	0.8488
Bachelors (Honors)	0.0156	1.24	1	0.2658
School 1	0.01185	0.71	1	0.399
School 2	-0.00869	0.39	1	0.5339
School 2 x 1/t	0.00774	0.3	1	0.583
School 3	0.00123	0.01	1	0.929
School 4	-0.00972	0.48	1	0.4871
School 5	0.00126	0.01	1	0.9294
School 6	-0.00786	0.32	1	0.5717
School 6 x t	0.01385	1	1	0.3181
School 7	0.00194	0.02	1	0.8885
School 8	0.01471	1.11	1	0.2925
School 9	-0.00363	0.07	1	0.7922

	rho	chi2	df	Prob>chi2
Withdrawn	-0.01573	1.08	1	0.2985
Withdrawn x t	0.01754	1.3	1	0.2545
Withdrawn x t^2	-0.01744	1.19	1	0.2752
Withdrawn Early	-0.00103	0.01	1	0.9414
Withdrawn Early x t	0.00182	0.02	1	0.8954
Withdrawn Early x t^2	-0.00284	0.04	1	0.8352
Fail Incomplete	-0.00594	0.17	1	0.6781
Fail Incomplete x t	-0.01247	0.71	1	0.4004
Fail	-0.01611	1.47	1	0.2255
Pass	-0.00865	0.38	1	0.5369
Credit	-0.00267	0.04	1	0.8457
Distinction	-0.00167	0.01	1	0.9031
High Distinction	0.00835	0.4	1	0.5277
Other	-0.00466	0.09	1	0.7655
Inactive	-0.0052	0.13	1	0.7201
Inactive x t (for t <= 16)	0.00188	0.01	1	0.9261
Inactive x 1/ln(t) (for t > 16)	0.01004	0.36	1	0.5487
Part-time	0.00792	0.33	1	0.5684
Part-time x ln(t)	-0.00356	0.06	1	0.8022
Early Alert System	0.03011	4.62	1	0.0315
Early Alert System x ln(t) (for t<=18)	-0.01092	0.68	1	0.4095
Early Alert System x ln(t - 18) (for t > 18)	-0.03194	5.18	1	0.0229
global test		39.99	51	0.8672

Conditional Model: No-EAS group - Proportional Hazards Test

	rho	chi2	df	Prob>chi2
Gender	-0	0.45	1	0.501
Age	0.01	0.17	1	0.6807
Age Squared	-0	0.6	1	0.4397
ATSI	0.02	0.66	1	0.4149
ATSI x t	-0	0.72	1	0.3955
International Fee	0.01	0.08	1	0.7786
Prior Studies	0	0	1	0.9796
On-campus	0.01	0.16	1	0.688
Diploma	0.01	0.14	1	0.7091
Advanced Diploma	0.04	2.31	1	0.1287
Advanced Diploma x t	-0	2.69	1	0.1012
Bachelors (Graduate)	0.01	0.11	1	0.7433
Bachelors (Honors)	0.03	1.24	1	0.2654
Bachelors (Honors) x t	-0	2.65	1	0.1036
School 1	0.02	1.1	1	0.2946
School 2	-0	0.87	1	0.352
School 2 x 1/t	0.02	0.69	1	0.4062
School 3	-0	0.03	1	0.8666
School 4	-0	1.52	1	0.2181
School 5	-0	0.5	1	0.4797
School 6	0.02	0.48	1	0.4871
School 7	-0	0.35	1	0.5548
School 8	0.02	0.64	1	0.4239
School 9	-0	0.05	1	0.8251

	rho	chi2	df	Prob>chi2
Withdrawn	-0.03767	2.53	1	0.1114
Withdrawn Early	-0.00716	0.11	1	0.7368
Withdrawn Early x ln(t)	0.00598	0.08	1	0.7792
Fail Incomplete	0.03116	1.68	1	0.1944
Fail Incomplete x ln(t)	-0.03564	2.31	1	0.1282
Fail	-0.03245	2.11	1	0.1462
Pass	-0.04281	3.24	1	0.0717
Credit	-0.04316	3.93	1	0.0475
Credit x t	0.03984	3.62	1	0.0571
Distinction	0.00171	0.01	1	0.9387
High Distinction	-0.02453	1.17	1	0.2801
High Distinction x t	0.02231	0.87	1	0.3503
Other	0.01256	0.3	1	0.5839
Inactive	0.00881	0.13	1	0.7137
Inactive x t (for t <= 16)	-0.00206	0	1	0.9522
Inactive x 1/ln(t) (for t > 16)	-0.00692	0.05	1	0.8173
Part-time	0.00066	0	1	0.9773
Global test		35.56	41	0.7107

Conditional Model: With-EAS group - Proportional Hazards Test

	rho	chi2	df	Prob>chi2
Gender	-0.00075	0	1	0.9656
Age	0.01667	0.8	1	0.37
Age Squared	-0.01715	0.82	1	0.3649
ATSI	0.02427	1.59	1	0.2068
ATSI x t	-0.02809	2.24	1	0.1341
ATSI x t^2	0.02813	2.47	1	0.1158
International Fee	-0.0211	1.7	1	0.1923
International Fee x t^2	0.02446	2.21	1	0.1375
International Fee x t^3	-0.02407	2.06	1	0.1515
Prior Studies	0.01701	0.92	1	0.3367
Prior Studies x t^2	-0.01969	1.2	1	0.2734
Prior Studies x t^3	0.01858	1.07	1	0.3011
On-campus	0.00625	0.12	1	0.7263
On-campus x t	0.01003	0.38	1	0.5379
Diploma	-0.00987	0.32	1	0.5733
Advanced Diploma	0.00928	0.28	1	0.5938
Bachelors (Graduate)	0.00936	0.27	1	0.6021
Bachelors (Honors)	0.0064	0.13	1	0.7152
School 1	-0.00957	0.3	1	0.5852
School 2	-0.0062	0.13	1	0.7229
School 3	-0.00844	0.24	1	0.6274
School 4	-0.0123	0.49	1	0.4835
School 5	-0.0059	0.11	1	0.74
School 6	0.02778	2.55	1	0.1103
School 7	0.00343	0.04	1	0.8365
School 8	0.01254	0.52	1	0.4714
School 9	-0.0116	0.45	1	0.5046

	rho	chi2	df	Prob>chi2
Withdrawn	-0.02803	2.41	1	0.1209
Withdrawn x ln(t)	0.02935	2.68	1	0.1015
Withdrawn Early	-0.02709	2.46	1	0.1169
Fail Incomplete	0.00681	0.15	1	0.701
Fail Incomplete x sqrt(t)	-0.00879	0.24	1	0.6242
Fail	-0.00931	0.31	1	0.5757
Pass	0.00592	0.11	1	0.7356
Credit	0.00024	0	1	0.9894
Credit x t	-0.00586	0.1	1	0.7478
Distinction	0.00562	0.11	1	0.7441
High Distinction	0.00823	0.24	1	0.6212
Other	-0.01179	0.36	1	0.5458
Inactive	-0.02693	2.12	1	0.1455
Inactive x 1/ln(t) (for all t)	0.02518	1.75	1	0.1859
Part-time	0.02589	2.32	1	0.1274
Part-time x 1/ln(t)	-0.03055	3.19	1	0.074
global test		32.53	43	0.8775

Interaction Short-Run Model - Proportional Hazards Test

	rho	chi2	df	Prob>chi2
Gender	-0.00922	0.44	1	0.5073
Age	0.00822	0.34	1	0.56
Age Squared	-0.01424	1	1	0.3165
ATSI	0.02397	2.91	1	0.0883
International Fee	-0.01418	1.17	1	0.2786
International Fee x t^2	0.01492	1.24	1	0.2654
International Fee x t^3	-0.01395	1.05	1	0.3045
Prior Studies	0.02153	2.25	1	0.1335
Prior Studies x ln(t)	-0.02424	2.84	1	0.0921
On-campus	0.01036	0.53	1	0.4672
On-campus x t	0.00146	0.01	1	0.9074
Diploma	-0.00489	0.12	1	0.7243
Advanced Diploma	0.018	1.6	1	0.2055
Advanced Diploma x t	0.00347	0.06	1	0.8085
Bachelors (Graduate)	0.0092	0.43	1	0.5111
Bachelors (Honors)	-0.01429	0.99	1	0.319
School 1	0.00773	0.31	1	0.5806
School 2	-0.01084	0.6	1	0.4368
School 2 x 1/t	-0.0034	0.06	1	0.8061
School 3	-0.01817	1.69	1	0.1939
School 4	0.00266	0.04	1	0.8512
School 5	-0.01087	0.61	1	0.4351
School 6	0.00026	0	1	0.985
School 6 x t	0.01163	0.69	1	0.4058
School 7	-0.00674	0.24	1	0.6268
School 8	0.01424	1.05	1	0.3066
School 9	0.01161	0.65	1	0.4198

	rho	chi2	df	Prob>chi2
Withdrawn	-0.0132	0.77	1	0.3809
Withdrawn x t	0.01485	0.94	1	0.3327
Withdrawn x t^2	-0.01483	0.87	1	0.3514
Withdrawn Early	-0.0063	0.2	1	0.6537
Withdrawn Early x t	0.0093	0.45	1	0.5042
Withdrawn Early x t^2	-0.01108	0.65	1	0.4218
Fail Incomplete	0.00663	0.21	1	0.6455
Fail Incomplete x t	-0.02801	3.49	1	0.0618
Fail	-0.01994	2.21	1	0.137
Pass	0.00666	0.23	1	0.6337
Credit	0.0063	0.21	1	0.647
Distinction	0.002	0.02	1	0.8837
High Distinction	0.01413	1.16	1	0.2815
Other	-0.00556	0.12	1	0.7277
Inactive	-0.01538	1.32	1	0.2499
Inactive x t (for t <= 16)	-0.00171	0.01	1	0.9031
Inactive x 1/sqrt(t) (for t > 16)	0.00257	0.02	1	0.9005
Part-time	0.01601	1.21	1	0.2721
Part-time x ln(t)	0.0013	0.01	1	0.9253
Gender x EAS	-0.00189	0.02	1	0.8907
Age x EAS	-0.01877	1.37	1	0.2416
Age x EAS	0.01437	0.78	1	0.3779
ATSI x EAS	-0.0113	0.67	1	0.4126
International Fee x EAS	0.00501	0.12	1	0.7318
Prior Studies x EAS	0.00911	0.39	1	0.5301
On-campus x EAS	-0.0072	0.26	1	0.6099
Diploma x EAS	0.00185	0.02	1	0.8963
Advanced Diploma x EAS	-0.00156	0.01	1	0.9066
Bachelors (Graduate) x EAS	-0.01786	1.39	1	0.2379
Bachelors (Honors) x EAS	0.00699	0.25	1	0.6153

	rho	chi2	df	Prob>chi2
School 1 x EAS	-0.00034	0	1	0.9809
School 2 x EAS	-0.0181	1.63	1	0.2019
School 3 x EAS	-0.01006	0.52	1	0.4706
School 4 x EAS	0.01688	1.4	1	0.2369
School 5 x EAS	-0.00389	0.07	1	0.7921
School 6 x EAS	-0.00088	0	1	0.9498
School 7 x EAS	-0.00282	0.05	1	0.83
School 8 x EAS	-0.00695	0.26	1	0.609
School 9 x EAS	-0.00053	0	1	0.9689
Withdrawn x EAS	0.0085	0.31	1	0.5787
Withdrawn Early x EAS	0.00993	0.51	1	0.473
Fail Incomplete x EAS	-0.00009	0	1	0.9946
Fail Incomplete x EAS x t	0.0031	0.05	1	0.8189
Fail x EAS	0.02241	1.94	1	0.1639
Fail x EAS x t	-0.0265	2.1	1	0.1472
Pass x EAS	-0.0015	0.01	1	0.921
Pass x EAS x t	-0.00197	0.01	1	0.9031
Credit x EAS	0.00211	0.02	1	0.8765
Distinction x EAS	0.01477	1.23	1	0.2668
High Distinction x EAS	0.00226	0.02	1	0.8771
Other x EAS	-0.00594	0.21	1	0.6449
Inactive x EAS	-0.00763	0.3	1	0.5869
Inactive x EAS x t	0.01091	0.63	1	0.4282
Part-time x EAS	0.00557	0.19	1	0.6632
global test		68.22	82	0.8622

Interaction Long-Run Model - Proportional Hazards Test

	rho	chi2	df	Prob>chi2
Gender	-0.00896	0.42	1	0.5194
Age	0.00482	0.12	1	0.7239
Age Squared	-0.01015	0.55	1	0.459
ATSI	0.00152	0.01	1	0.9126
ATSI x t	0.00486	0.1	1	0.7467
International Fee	0.0044	0.1	1	0.7545
Prior Studies	-0.00019	0	1	0.9888
On-campus	0.00469	0.11	1	0.7356
Diploma	0.00421	0.1	1	0.7566
Advanced Diploma	0.00773	0.31	1	0.5795
Advanced Diploma x ln(t)	-0.00636	0.21	1	0.647
Bachelors (Graduate)	0.0037	0.07	1	0.7975
Bachelors (Honors)	0.01775	1.56	1	0.2114
Bachelors (Honors) x t	-0.0127	0.84	1	0.3597
School 1	0.01677	1.42	1	0.2333
School 2	-0.00027	0	1	0.984
School 2 x 1/t	0.00476	0.14	1	0.7094
School 3	-0.00126	0.01	1	0.9269
School 4	-0.01588	1.32	1	0.2513
School 5	-0.00884	0.39	1	0.5345
School 6	0.01002	0.52	1	0.4727
School 7	-0.00804	0.35	1	0.5524
School 8	0.01072	0.57	1	0.4491
School 9	-0.0037	0.07	1	0.7873

	rho	chi2	df	Prob>chi2
Withdrawn	-0.01646	1.31	1	0.2526
Withdrawn Early	-0.00319	0.05	1	0.8186
Withdrawn Early x t	-0.00626	0.21	1	0.6498
Withdrawn Early x t^2	0.00456	0.11	1	0.7347
Fail Incomplete	-0.01776	1.78	1	0.1817
Fail Incomplete x t	-0.03016	4.02	1	0.0449
Fail	-0.01396	1.07	1	0.3015
Pass	-0.01652	1.32	1	0.2514
Credit	-0.00834	0.38	1	0.5396
Credit x t	0.00226	0.03	1	0.8634
Distinction	0.00651	0.23	1	0.6346
Distinction x t	-0.00828	0.35	1	0.5522
High Distinction	0.00347	0.06	1	0.8064
High Distinction x t	-0.01049	0.47	1	0.4936
Other	0.0045	0.11	1	0.7444
Inactive	0.02358	3.08	1	0.0793
Inactive x t (for t <= 16)	-0.00296	0.01	1	0.9109
Inactive x ln(t-15) (for t > 16)	-0.0223	2.83	1	0.0923
Part-time	0.01606	1.46	1	0.227
Part-time x ln(t)	-0.01181	0.72	1	0.3968
Gender x EAS	0.0077	0.31	1	0.5789
Age x EAS	0.00364	0.07	1	0.7972
Age^2 x EAS	-0.00021	0	1	0.9884
ATSI x EAS	-0.01402	0.93	1	0.3356
ATSI x EAS x t	0.01304	0.86	1	0.3539
International Fee x EAS	0.01235	0.77	1	0.3796
Prior Studies x EAS	0.01816	1.71	1	0.1907
On-campus x EAS	0.00786	0.32	1	0.5729
Diploma x EAS	-0.00713	0.27	1	0.6048
Advanced Diploma x EAS	-0.00728	0.28	1	0.5977
Bachelors (Graduate) x EAS	0.00242	0.03	1	0.8663
Bachelors (Honors) x EAS	-0.01044	0.65	1	0.4196

	rho	chi2	df	Prob>chi2
School 1 x EAS	-0.01928	1.89	1	0.1695
School 2 x EAS	-0.01346	1.01	1	0.315
School 2 x EAS x ln(t)	0.01478	1.23	1	0.2668
School 3 x EAS	-0.00296	0.05	1	0.8292
School 4 x EAS	0.00676	0.24	1	0.6265
School 5 x EAS	0.00427	0.09	1	0.7636
School 6 x EAS	0.00429	0.1	1	0.7579
School 7 x EAS	0.01019	0.56	1	0.4544
School 8 x EAS	-0.00364	0.07	1	0.7958
School 9 x EAS	-0.00383	0.08	1	0.781
Withdrawn x EAS	0.00017	0	1	0.9904
Withdrawn Early x EAS	0.02143	2.32	1	0.1277
Fail Incomplete x EAS	0.01952	2.26	1	0.1324
Fail x EAS	0.01036	0.59	1	0.4425
Pass x EAS	0.01615	1.25	1	0.2632
Credit x EAS	0.00376	0.08	1	0.7839
Credit x EAS x t	0.00043	0	1	0.9743
Distinction x EAS	-0.00845	0.38	1	0.5355
High Distinction x EAS	-0.00077	0	1	0.9563
High Distinction x EAS x t	0.00532	0.13	1	0.7196
Other x EAS	-0.00633	0.21	1	0.6475
Inactive x EAS	-0.01932	1.77	1	0.1829
Part-time x EAS	0.00623	0.21	1	0.6486
Part-time x EAS x t	-0.01356	0.97	1	0.3244
Global test		95.95	81	0.1229

EAS-Trigger ASB Model - Proportional Hazards Test

Trigger	rho	chi2	df	Prob>chi2
1	-0.01322	0.87	1	0.3497
2	0.00577	0.17	1	0.6821
3	-0.02833	2.61	1	0.1065
3 x 1/ln(t)	0.01191	0.49	1	0.4861
4	0.01346	0.99	1	0.3192
4 x t	-0.01722	1.29	1	0.2554
5	0.01689	1.38	1	0.2396
5 x sqrt(t)	-0.01971	1.75	1	0.1864
6	0.01066	0.54	1	0.462
8	0.01526	1.18	1	0.2777
9	-0.01627	1.37	1	0.2418
10	-0.0112	0.67	1	0.4134
11	-0.01402	1.07	1	0.3005
11 x t	0.01695	1.67	1	0.1958
12	-0.00928	0.45	1	0.5016
12 x t	0.00988	0.52	1	0.4718
13	-0.00177	0.01	1	0.9029
13 x t	-0.00123	0.01	1	0.9318
14	-0.01057	0.52	1	0.4708
19	0.00982	0.31	1	0.5762
19 x t	-0.02003	0.66	1	0.4175
20	0.01239	0.49	1	0.4836
20 x t	-0.00899	0.26	1	0.61
21	0.01517	1.18	1	0.2767
22	-0.01494	1.14	1	0.2865
23	0.00093	0	1	0.9459
24	0.00431	0.1	1	0.7577
25	0.00652	0.23	1	0.6326
25 x t	-0.00986	0.43	1	0.5138

Trigger	rho	chi2	df	Prob>chi2
26	0.02052	2.22	1	0.1362
26 x t	-0.02193	2.47	1	0.1161
27	-0.00252	0.03	1	0.8577
28	-0.01939	2.7	1	0.1007
28 x t	0.02332	4.71	1	0.0299
29	-0.01401	1.1	1	0.2934
29 x t	0.01787	1.87	1	0.172
30	-0.02137	2.71	1	0.0999
30 x ln(t)	0.02676	4.31	1	0.0379
31	-0.02147	2.36	1	0.1243
32	0.00322	0.05	1	0.8318
32 x t	-0.00615	0.15	1	0.7021
33	-0.00453	0.1	1	0.7516
34	0.00643	0.12	1	0.7257
34 x ln(t)	-0.00743	0.15	1	0.6961
Global test		39.16	44	0.6789

EAS-Trigger ASM Model - Proportional Hazards Test

Trigger	rho	chi2	df	Prob>chi2
1	-0.01101	0.58	1	0.4451
2	0.00536	0.14	1	0.7048
3	0.01307	0.59	1	0.4429
3 x 1/ln(t)	-0.01255	0.52	1	0.4714
4	0.00335	0.06	1	0.8072
4 x t	-0.00485	0.14	1	0.7115
5	-0.01648	1.29	1	0.2555
6	0.01143	0.6	1	0.4387
8	0.01575	1.19	1	0.2743
9	-0.01595	1.76	1	0.1842
10	-0.01182	0.7	1	0.4029
11	-0.01108	0.67	1	0.4115
11 x t	0.01519	1.34	1	0.2472
12	-0.00216	0.03	1	0.8737
12 x t	0.0037	0.07	1	0.7906
13	0.00286	0.04	1	0.8359
13 x t	-0.00367	0.07	1	0.7889
14	-0.02772	3.43	1	0.064
19	0.01417	0.61	1	0.4342
19 x t	-0.0252	0.87	1	0.3501
20	0.01255	0.46	1	0.4994
20 x t	-0.01161	0.38	1	0.538
21	0.01632	1.21	1	0.2715
22	-0.0192	1.01	1	0.3154
22 x ln(t)	0.02191	1.65	1	0.1993
23	0.01324	0.78	1	0.3785
24	0.00328	0.05	1	0.8317
25	0.02953	4.18	1	0.0408
25 x t	-0.02382	2.41	1	0.1202

Trigger	rho	chi2	df	Prob>chi2
26	-0.02907	3.86	1	0.0494
27	-0.00238	0.03	1	0.8743
28	-0.01879	2.65	1	0.1034
28 x t	0.02286	4.58	1	0.0324
29	-0.01226	0.82	1	0.3664
29 x t	0.01687	1.62	1	0.2032
30	-0.02553	3.28	1	0.07
30 x ln(t)	0.02838	4.27	1	0.0387
31	-0.01984	2.08	1	0.1497
31 x ln(t)	0.02216	2.66	1	0.1029
32	0.00616	0.15	1	0.6946
32 x t	-0.00671	0.17	1	0.6831
33	-0.00555	0.13	1	0.714
34	-0.0072	0.27	1	0.6012
34 x t	0.01378	1.09	1	0.2971
Global test		40.1	44	0.6396

EAS-Trigger ISB Model - Proportional Hazards Test

Trigger	rho	chi2	df	Prob>chi2
1	-0.01319	0.87	1	0.3499
2	0.00707	0.25	1	0.6164
3	0.0082	0.3	1	0.587
3 x 1/ln(t)	-0.00951	0.42	1	0.5194
4	0.02045	2.09	1	0.1486
4 x t	-0.02967	3.41	1	0.0646
5	0.01866	1.5	1	0.2207
5 x t	-0.02255	2.18	1	0.1395
5 x t^3	0.02108	2.14	1	0.1439
6	0.01167	0.62	1	0.4302
8	0.01566	1.17	1	0.2802
9	-0.01295	0.83	1	0.3609
10	-0.00787	0.33	1	0.567
11	-0.00229	0.03	1	0.8704
11 x t	0.00297	0.05	1	0.8297
12	-0.0187	1.81	1	0.1781
13	0.00507	0.11	1	0.735
13 x t	-0.00572	0.15	1	0.6952
14	-0.0064	0.19	1	0.6648
19	0.01217	0.46	1	0.4954
19 x t	-0.0217	0.7	1	0.4027
20	0.0106	0.38	1	0.5391
20 x t	-0.01216	0.5	1	0.479
21	0.01535	1.21	1	0.2717
22	-0.016	1.29	1	0.2566
23	0.00051	0	1	0.9709
24	-0.00026	0	1	0.9851
25	-0.0129	0.8	1	0.3724

Trigger	rho	chi2	df	Prob>chi2
26	0.00846	0.36	1	0.5493
26 x t^2	0.01195	1.05	1	0.3057
27	-0.00343	0.06	1	0.8072
28	-0.00078	0	1	0.9583
28 x ln(t)	0.00103	0	1	0.9453
29	-0.0212	2.31	1	0.1287
30	-0.02214	2.46	1	0.1167
31	-0.00253	0.03	1	0.8563
32	-0.02338	3.02	1	0.0821
32 x t	0.01595	1.78	1	0.1821
33	-0.00843	0.35	1	0.5551
34	-0.01306	1.14	1	0.2852
Global test		27.85	40	0.9265

EAS-Trigger ISM Model - Proportional Hazards Test

Trigger	rho	chi2	df	Prob>chi2
1	-0.01179	0.64	1	0.4244
2	0.00696	0.2	1	0.6572
3	0.01117	0.45	1	0.5023
3 x 1/ln(t)	-0.01056	0.39	1	0.5308
4	0.00888	0.42	1	0.5177
4 x t	-0.01142	0.71	1	0.4009
5	-0.0117	0.65	1	0.421
6	0.01261	0.7	1	0.4038
8	0.01653	1.26	1	0.2622
9	-0.01286	1.28	1	0.2575
10	-0.0078	0.3	1	0.5809
11	-0.00196	0.02	1	0.8867
11 x t	0.0022	0.03	1	0.8727
12	0.00921	0.46	1	0.4998
12 x t	-0.01501	1.07	1	0.3004
13	0.00264	0.04	1	0.8489
13 x t	-0.0046	0.11	1	0.74
14	-0.01469	0.97	1	0.3254
19	0.01522	0.63	1	0.427
19 x t	-0.02961	0.93	1	0.3336
20	0.00916	0.28	1	0.5983
20 x t	-0.00769	0.2	1	0.6535
21	0.01386	0.88	1	0.3491
22	-0.01274	0.64	1	0.4241
22 x ln(t)	0.01442	0.88	1	0.3483
23	-0.00111	1	1	0.9444
24	-0.00254	0.03	1	0.8679
25	-0.00851	0.33	1	0.5678

Trigger	rho	chi2	df	Prob>chi2
26	0.01299	0.6	1	0.4398
26 x t^2	0.00676	0.28	1	0.5962
27	-0.00282	0.03	1	0.8521
28	0.00145	0.01	1	0.9223
28 x ln(t)	-0.0015	0.01	1	0.9216
29	-0.02258	2.58	1	0.1085
30	-0.02296	2.61	1	0.1065
31	-0.00659	0.22	1	0.6405
32	0.01404	0.65	1	0.4189
32 x t	-0.02297	1.15	1	0.2842
33	-0.00689	0.2	1	0.653
34	-0.02256	2.08	1	0.149
Global test		21.62	40	0.9922

EAS-Trigger CASB Model - Proportional Hazards Test

Variable	rho	chi2	df	Prob>chi2
Gender	-0.00376	0.07	1	0.7865
Age	0.00769	0.3	1	0.5865
Age Squared	-0.01289	0.82	1	0.3659
ATSI	0.01746	1.55	1	0.2138
Domestic Fee	0.00458	0.1	1	0.748
Domestic Fee x t	-0.00476	0.11	1	0.7405
International Fee	-0.01379	1.11	1	0.2913
International Fee x t^2	0.0164	1.52	1	0.2181
International Fee x t^3	-0.01609	1.41	1	0.2344
Prior Studies	0.01407	1.08	1	0.2997
Prior Studies x ln(t)	-0.01546	1.27	1	0.2588
On-campus	0.00561	0.15	1	0.6962
On-campus x t	0.00228	0.03	1	0.8595
Diploma	-0.00526	0.14	1	0.7047
Advanced Diploma	0.01854	1.72	1	0.1902
Advnaced Diploma x t	-0.01552	1.18	1	0.2781
Bachelors (Graduate)	-0.00751	0.27	1	0.6008
Bachelors (Honors)	0.01273	0.83	1	0.3628
School 1	0.01432	1.05	1	0.3056
School 2	-0.00728	0.27	1	0.6023
School 2 x 1/t	0.00729	0.26	1	0.6085
School 3	-0.00086	0	1	0.9501
School 4	-0.0108	0.6	1	0.4393
School 5	0.01025	0.52	1	0.4714
School 6	-0.00663	0.23	1	0.6334
School 6 x t	0.01177	0.71	1	0.3987
School 7	-0.00262	0.04	1	0.8438
School 8	0.012	0.74	1	0.3896
School 9	-0.00471	0.12	1	0.7331

Variable	rho	chi2	df	Prob>chi2
Withdrawn	-0.01157	0.6	1	0.4368
Withdrawn x t	0.01453	0.93	1	0.3357
Withdrawn x t^2	-0.01525	0.95	1	0.3309
Withdrawn Early	-0.00309	0.05	1	0.8269
Withdrawn Early x t	0.00679	0.24	1	0.6274
Withdrawn Early x t^2	-0.0086	0.39	1	0.5335
Fail Incomplete	-0.04723	9.84	1	0.0017
Fail Incomplete x t	0.04497	9.03	1	0.0027
Fail	-0.01037	0.6	1	0.4378
Pass	0.00581	0.17	1	0.6784
Credit	0.01037	0.57	1	0.4494
Distinction	0.00983	0.52	1	0.4722
High Distinction	0.01603	1.48	1	0.2234
Other	-0.00016	0	1	0.9917
Inactive	-0.01357	1.08	1	0.2988
Inactive x t (for t <= 16)	0.00449	0.09	1	0.7635
Inactive x 1/ln(t) (for t > 16)	0.00679	0.18	1	0.6754
Part-time	0.01454	1.1	1	0.2946
Part-time x ln(t)	-0.01695	1.55	1	0.2134
Trigger 1	-0.02114	2.43	1	0.1189
Trigger 2	0.0005	0	1	0.972
Trigger 3	-0.01805	1.68	1	0.1951
Trigger 3 x t	0.02787	4.27	1	0.0389
Trigger 4	0.02607	3.24	1	0.0717
Trigger 4 x ln(t)	-0.03031	4.41	1	0.0356
Trigger 5	-0.0091	0.58	1	0.4464
Trigger 5 x 1/t	0.00256	0.05	1	0.8186
Trigger 6	0.021	2.13	1	0.1442
Trigger 8	0.0215	2.38	1	0.1226

Variable	rho	chi2	df	Prob>chi2
Trigger 9	-0.01422	1.05	1	0.3055
Trigger 10	-0.00985	0.52	1	0.4691
Trigger 11	0.01073	0.57	1	0.4492
Trigger 11 x t	-0.01424	1.01	1	0.3148
Trigger 12	0.00995	0.52	1	0.4702
Trigger 12 x t	-0.01381	1.03	1	0.3101
Trigger 13	0.02213	2.21	1	0.1371
Trigger 13 x t	-0.03152	4.44	1	0.035
Trigger 14	0.02473	2.89	1	0.089
Trigger 14 x ln(t)	-0.03587	5.67	1	0.0172
Trigger 19	0.00563	0.11	1	0.7346
Trigger 19 x t	-0.00709	0.15	1	0.7018
Trigger 20	0.00807	0.22	1	0.6403
Trigger 20 x t	-0.00322	0.04	1	0.8509
Trigger 21	0.00558	0.16	1	0.6852
Trigger 22	-0.00679	0.24	1	0.6236
Trigger 23	0.00166	0.01	1	0.9048
Trigger 24	0.00662	0.22	1	0.6363
Trigger 25	0.00867	0.43	1	0.5119
Trigger 25 x t	-0.01395	0.92	1	0.3372
Trigger 26	-0.00646	0.26	1	0.6109
Trigger 26 x t	0.01068	0.69	1	0.4048
Trigger 27	-0.00104	0.01	1	0.9402
Trigger 28	-0.01811	2.18	1	0.1394
Trigger 28 x t	0.0205	3.27	1	0.0707
Trigger 29	-0.00734	0.29	1	0.589
Trigger 29 x t	0.01012	0.55	1	0.4584
Trigger 30	0.0149	1.02	1	0.3117
Trigger 30 x 1/t	-0.0103	0.44	1	0.5063
Trigger 31	0.00068	0	1	0.9597
Trigger 32	0.00495	0.11	1	0.7402
Trigger 32 x t	-0.0068	0.19	1	0.6604
Trigger 33	-0.01013	0.52	1	0.4689
Trigger 34	0.00322	0.03	1	0.8635
Trigger 34 x ln(t)	-0.00558	0.08	1	0.7736
Global test		73.42	93	0.9334

EAS-Trigger CASM Model - Proportional Hazards Test

Variable	rho	chi2	df	Prob>chi2
Gender	-0.00279	0.04	1	0.8404
Age	0.00959	0.46	1	0.4977
Age Squared	-0.01477	1.07	1	0.3004
ATSI	0.01871	1.77	1	0.1828
Domestic Fee	0.00479	0.11	1	0.7377
Domestic Fee x t	-0.00506	0.12	1	0.726
International Fee	-0.01351	1.07	1	0.3016
International Fee x t^2	0.01593	1.43	1	0.231
International Fee x t^3	-0.01558	1.33	1	0.2489
Prior Studies	0.01428	1.09	1	0.2957
Prior Studies x ln(t)	-0.0156	1.29	1	0.2568
On-campus	0.00534	0.14	1	0.7089
On-campus x t	0.00241	0.03	1	0.8543
Diploma	-0.00455	0.11	1	0.7428
Advanced Diploma	0.01944	1.89	1	0.1697
Advanced Diploma x t	-0.01746	1.49	1	0.2229
Bachelors (Graduate)	-0.00882	0.38	1	0.5386
Bachelors (Honors)	0.01344	0.92	1	0.3365
School 1	0.01502	1.15	1	0.283
School 2	-0.00729	0.27	1	0.6012
School 2 x 1/t	0.00748	0.28	1	0.5985
School 3	-0.00146	0.01	1	0.9157
School 4	-0.01099	0.62	1	0.4306
School 5	0.00932	0.43	1	0.5128
School 6	-0.00515	0.14	1	0.7114
School 6 x t	0.00893	0.41	1	0.5221
School 7	-0.00261	0.04	1	0.8439
School 8	0.01219	0.76	1	0.3821
School 9	-0.00481	0.12	1	0.7272

Variable	rho	chi2	df	Prob>chi2
Withdrawn	-0.01117	0.56	1	0.4544
Withdrawn x t	0.01326	0.77	1	0.3804
Withdrawn x t^2	-0.01341	0.73	1	0.3929
Withdrawn Early	-0.00197	0.02	1	0.8905
Withdrawn Early x t	0.00436	0.1	1	0.7573
Withdrawn Early x t^2	-0.00569	0.17	1	0.6817
Fail Incomplete	0.00989	0.47	1	0.4917
Fail Incomplete x t	-0.03135	4.42	1	0.0355
Fail	-0.01122	0.71	1	0.4006
Pass	0.00478	0.12	1	0.7329
Credit	0.01142	0.7	1	0.4031
Distinction	0.00906	0.44	1	0.5073
High Distinction	0.01706	1.68	1	0.195
Other	-0.00023	0	1	0.9878
Inactive	-0.01253	0.96	1	0.3276
Inactive x t (for t <= 16)	0.0026	0.04	1	0.8505
Inactive x 1/ln(t) (for t > 16)	0.00761	0.23	1	0.6344
Part-time	0.01323	0.9	1	0.342
Part-time x ln(t)	-0.01543	1.29	1	0.2554
Trigger 1	-0.01953	1.79	1	0.1805
Trigger 2	-0.00252	0.03	1	0.8577
Trigger 3	-0.01381	0.84	1	0.3586
Trigger 3 x t	0.02132	2.39	1	0.1225
Trigger 4	0.02027	2.06	1	0.1516
Trigger 4 x ln(t)	-0.02461	3.12	1	0.0775
Trigger 5	-0.00941	0.57	1	0.4516
Trigger 5 x 1/t	0.00318	0.06	1	0.807
Trigger 6	0.02177	2.23	1	0.1354
Trigger 8	0.02207	2.28	1	0.1306
Trigger 9	-0.01156	0.92	1	0.3379
Trigger 10	-0.01213	0.77	1	0.3815
Trigger 11	0.01031	0.53	1	0.4671
Trigger 11 x t	-0.01529	1.11	1	0.2928

Variable	rho	chi2	df	Prob>chi2
Trigger 12	0.00839	0.38	1	0.5368
Trigger 12 x t	-0.01245	0.81	1	0.3694
Trigger 13	0.01709	1.35	1	0.2458
Trigger 13 x t	-0.02576	3.05	1	0.0806
Trigger 14	0.00933	0.43	1	0.5122
Trigger 14 x ln(t)	-0.01043	0.55	1	0.4579
Trigger 19	0.0079	0.2	1	0.6568
Trigger 19 x t	-0.00921	0.2	1	0.6548
Trigger 20	0.00579	0.12	1	0.7299
Trigger 20 x t	-0.00497	0.09	1	0.7705
Trigger 21	0.00545	0.1	1	0.7552
Trigger 22	-0.0048	0.12	1	0.7298
Trigger 22 x 1/t	0.00204	0.01	1	0.9072
Trigger 23	-0.00125	0.01	1	0.9359
Trigger 24	0.00477	0.1	1	0.7493
Trigger 25	0.00743	0.31	1	0.5764
Trigger 25 x t	-0.00984	0.44	1	0.5049
Trigger 26	0.00134	0.01	1	0.9208
Trigger 26 x t	-0.00036	0	1	0.9791
Trigger 27	-0.0008	0	1	0.9606
Trigger 28	-0.01584	1.77	1	0.183
Trigger 28 x t	0.01932	2.98	1	0.0842
Trigger 29	-0.00628	0.21	1	0.6505
Trigger 29 x t	0.00828	0.36	1	0.5479
Trigger 30	0.01556	1.08	1	0.2978
Trigger 30 x 1/t	-0.01062	0.43	1	0.512
Trigger 31	0.0065	0.18	1	0.6718
Trigger 31 x 1/t	-0.00357	0.03	1	0.8654
Trigger 32	0.00109	0.01	1	0.9432
Trigger 32 x t	-0.00299	0.04	1	0.8497
Trigger 33	-0.01237	0.71	1	0.4007
Trigger 34	-0.01665	1.64	1	0.2006
Trigger 34 x t	0.02081	2.73	1	0.0984

EAS-Trigger CISB Model - Proportional Hazards Test

Variable	rho	chi2	df	Prob>chi2
Gender	-0.00346	0.06	1	0.8027
Age	0.00815	0.33	1	0.5637
Age Squared	-0.01344	0.89	1	0.3449
ATSI	0.01722	1.5	1	0.2203
Domestic Fee	0.00468	0.11	1	0.7432
Domestic Fee x t	-0.00519	0.13	1	0.7191
International Fee	-0.01267	0.94	1	0.3318
International Fee x t^2	0.01482	1.25	1	0.2634
International Fee x t^3	-0.01447	1.16	1	0.282
Prior Studies	0.0161	1.37	1	0.2421
Prior Studies x ln(t)	-0.01794	1.67	1	0.1957
On-campus	0.00611	0.18	1	0.6706
On-campus x t	0.00339	0.07	1	0.7916
Diploma	-0.00499	0.13	1	0.7192
Advanced Diploma	0.01923	1.84	1	0.1745
Advanced Diploma x t	-0.01709	1.42	1	0.233
Bachelors (Graduate)	-0.00725	0.26	1	0.6132
Bachelors (Honors)	0.0118	0.71	1	0.3994
School 1	0.01424	1.04	1	0.3084
School 2	-0.00753	0.29	1	0.5901
School 2 x 1/t	0.00795	0.31	1	0.5775
School 3	-0.00176	0.02	1	0.8989
School 4	-0.01144	0.67	1	0.4126
School 5	0.00947	0.44	1	0.5062
School 6	-0.00505	0.13	1	0.7165
School 6 x t	0.01026	0.54	1	0.4622
School 7	-0.00287	0.05	1	0.8293
School 8	0.01341	0.92	1	0.3366
School 9	-0.00465	0.11	1	0.7356

Variable	rho	chi2	df	Prob>chi2
Withdrawn	-0.01023	0.47	1	0.4931
Withdrawn x t	0.012	0.63	1	0.4272
Withdrawn x t^2	-0.01187	0.57	1	0.4487
Withdrawn Early	-0.00231	0.03	1	0.87
Withdrawn Early x t	0.00499	0.13	1	0.7215
Withdrawn Early x t^2	-0.00644	0.22	1	0.6413
Fail Incomplete	0.01057	0.54	1	0.4631
Fail Incomplete x t	-0.03343	4.97	1	0.0258
Fail	-0.01012	0.57	1	0.4486
Pass	0.00478	0.12	1	0.7328
Credit	0.00986	0.52	1	0.4711
Distinction	0.00913	0.45	1	0.5035
High Distinction	0.01628	1.54	1	0.2153
Other	-0.00014	0	1	0.9928
Inactive	-0.01173	0.8	1	0.3709
Inactive x t (for t <= 16)	0.00311	0.03	1	0.8679
Inactive x 1/ln(t) (for t > 16)	0.00786	0.23	1	0.6298
Part-time	0.01598	1.31	1	0.2524
Part-time x ln(t)	-0.01702	1.55	1	0.2135
Trigger 1	-0.02146	2.49	1	0.1143
Trigger 2	0.0024	0.03	1	0.866
Trigger 3	-0.00376	0.07	1	0.7934
Trigger 3 x t	0.00858	0.36	1	0.5476
Trigger 4	0.02102	1.93	1	0.1643
Trigger 4 x ln(t)	-0.02684	3.14	1	0.0764
Trigger 5	-0.00994	0.71	1	0.3992
Trigger 5 x 1/t	0.00251	0.05	1	0.8181
Trigger 6	0.02069	1.97	1	0.1605
Trigger 8	0.0212	2.16	1	0.1417
Trigger 9	-0.01336	0.89	1	0.3447
Trigger 10	-0.00838	0.37	1	0.5416
Trigger 11	0.0137	0.88	1	0.3472
Trigger 11 x t	-0.02015	1.82	1	0.1773

Variable	rho	chi2	df	Prob>chi2
Trigger 12	0.02015	1.92	1	0.1656
Trigger 12 x t	-0.02849	3.68	1	0.0552
Trigger 13	0.02013	1.68	1	0.1945
Trigger 13 x t	-0.02875	3.42	1	0.0642
Trigger 14	0.01086	0.62	1	0.4297
Trigger 14 x ln(t)	-0.013	0.91	1	0.3392
Trigger 19	0.00546	0.1	1	0.7479
Trigger 19 x t	-0.00613	0.1	1	0.7475
Trigger 20	0.00611	0.13	1	0.7166
Trigger 20 x t	-0.00493	0.09	1	0.7694
Trigger 21	0.0029	0.04	1	0.8323
Trigger 22	-0.00805	0.33	1	0.5665
Trigger 23	0.0031	0.05	1	0.8237
Trigger 24	0.00242	0.03	1	0.8631
Trigger 25	0.02373	3.12	1	0.0774
Trigger 26	-0.00875	0.43	1	0.5126
Trigger 26 x t	0.00568	0.21	1	0.6488
Trigger 27	-0.00232	0.03	1	0.8662
Trigger 28	-0.01266	0.92	1	0.3369
Trigger 28 x t	0.01386	1.12	1	0.2889
Trigger 29	0.00425	0.1	1	0.7528
Trigger 30	0.01035	0.55	1	0.4594
Trigger 31	0.0166	1.49	1	0.2223
Trigger 32	-0.02031	2.17	1	0.1404
Trigger 32 x t	0.02086	3.09	1	0.0789
Trigger 33	-0.01195	0.72	1	0.3946
Trigger 34	-0.01638	1.81	1	0.1788
Global test		74.56	89	0.8635

EAS-Trigger CISM Model - Proportional Hazards Test

Variables	rho	chi2	df	Prob>chi2
Gender	-0.00302	0.05	1	0.8277
Age	0.00979	0.48	1	0.4881
Age Squared	-0.01495	1.1	1	0.2937
ATSI	0.01706	1.47	1	0.2256
Domestic Fee	0.0049	0.12	1	0.7319
Domestic Fee x t	-0.00529	0.13	1	0.7143
International Fee	-0.01309	0.99	1	0.3203
International Fee x t^2	0.01608	1.44	1	0.2296
International Fee x t^3	-0.01577	1.35	1	0.2457
Prior Studies	0.01659	1.42	1	0.2328
Prior Studies x ln(t)	-0.01775	1.61	1	0.2038
On-campus	0.00593	0.17	1	0.6792
On-campus x t	0.00293	0.05	1	0.8229
Diploma	-0.00484	0.12	1	0.7267
Advanced Diploma	0.02022	2.04	1	0.1532
Advanced Diploma x t	-0.01836	1.64	1	0.2003
Bachelors (Graduate)	-0.0079	0.3	1	0.582
Bachelors (Honors)	0.0134	0.92	1	0.3383
School 1	0.01509	1.16	1	0.2807
School 2	-0.00701	0.25	1	0.6158
School 2 x 1/t	0.00735	0.27	1	0.6051
School 3	-0.00108	0.01	1	0.9378
School 4	-0.01017	0.53	1	0.4657
School 5	0.00934	0.43	1	0.5121
School 6	-0.00479	0.12	1	0.7305
School 6 x t	0.00928	0.44	1	0.5057
School 7	-0.0031	0.05	1	0.8148
School 8	0.01333	0.91	1	0.3394
School 9	-0.00411	0.09	1	0.7657

Variables	rho	chi2	df	Prob>chi2
Withdrawn	-0.01025	0.47	1	0.493
Withdrawn x t	0.01263	0.7	1	0.4041
Withdrawn x t^2	-0.01279	0.66	1	0.4152
Withdrawn Early	-0.00108	0.01	1	0.9394
Withdrawn Early x t	0.00341	0.06	1	0.8074
Withdrawn Early x t^2	-0.00473	0.12	1	0.7318
Fail Incomplete	0.00924	0.41	1	0.5212
Fail Incomplete x t	-0.02988	4.01	1	0.0452
Fail	-0.011	0.68	1	0.4099
Pass	0.00417	0.09	1	0.7663
Credit	0.01007	0.54	1	0.4612
Distinction	0.00841	0.38	1	0.5378
High Distinction	0.01571	1.42	1	0.2326
Other	0.00036	0	1	0.981
Inactive	-0.01071	0.69	1	0.4069
Inactive x t (for t <= 16)	-0.00041	0	1	0.9801
Inactive x 1/ln(t) (for t > 16)	0.00688	0.18	1	0.6711
Part-time	0.01307	0.87	1	0.3511
Part-time x ln(t)	-0.0136	0.99	1	0.3197
Trigger 1	-0.01756	1.41	1	0.2349
Trigger 2	0.00228	0.02	1	0.8825
Trigger 3	-0.00186	0.01	1	0.9035
Trigger 3 x t	0.00585	0.16	1	0.6881
Trigger 4	0.01831	1.62	1	0.2034
Trigger 4 x ln(t)	-0.02368	2.71	1	0.0998
Trigger 5	-0.01342	1.1	1	0.2948
Trigger 5 x 1/t	0.00575	0.18	1	0.6711
Trigger 6	0.02113	2.04	1	0.153
Trigger 8	0.02212	2.19	1	0.1391
Trigger 9	-0.01109	0.96	1	0.3282
Trigger 10	-0.0094	0.45	1	0.5003
Trigger 11	0.01209	0.7	1	0.4015
Trigger 11 x t	-0.0198	1.69	1	0.193

Variables	rho	chi2	df	Prob>chi2
Trigger 12	0.01727	1.56	1	0.2112
Trigger 12 x t	-0.02688	3.28	1	0.0702
Trigger 13	0.0158	1.11	1	0.2912
Trigger 13 x t	-0.02552	2.83	1	0.0923
Trigger 14	0.01507	1.04	1	0.3088
Trigger 14 x ln(t)	-0.01676	1.35	1	0.2457
Trigger 19	0.01234	0.44	1	0.5069
Trigger 19 x t	-0.01405	0.37	1	0.5444
Trigger 20	0.00758	0.2	1	0.6541
Trigger 20 x t	-0.00302	0.03	1	0.8602
Trigger 21	0.00387	0.06	1	0.8113
Trigger 22	-0.02321	2.83	1	0.0924
Trigger 22 x ln(t)	0.02563	3.69	1	0.0548
Trigger 23	-0.00034	0	1	0.9829
Trigger 24	-0.00058	0	1	0.9687
Trigger 25	0.02571	3.47	1	0.0625
Trigger 26	-0.006	0.16	1	0.6925
Trigger 26 x t	0.00136	0.01	1	0.9218
Trigger 27	-0.00137	0.01	1	0.9311
Trigger 28	-0.01012	0.58	1	0.4463
Trigger 28 x ln(t)	0.01121	0.7	1	0.4021
Trigger 29	0.0037	0.07	1	0.7869
Trigger 30	0.00968	0.46	1	0.4967
Trigger 31	0.01302	0.9	1	0.3432
Trigger 32	0.01036	0.35	1	0.5527
Trigger 32 x t	-0.01893	0.73	1	0.3941
Trigger 33	-0.01417	0.96	1	0.3272
Trigger 34	-0.02331	2.44	1	0.1182
Global test		74.55	90	0.88

Appendix G: Survival Analysis Link-test results

The link test results are significant. This indicates that the survival models are underspecified.

No. of subjects	16,124	Number of obs	1,119,710
No. of failures	5,072		
Time at risk	1119710		
		LR chi2(2)	7688.4
Log likelihood	-42100.719	Prob > chi2	0

t	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
_hat	1.1626	0.0303	38.34	0	1.1031	1.222
hatsq	-0.0522	0.0084	-6.23	0	-0.0686	-0.0358

Appendix H: Treatment Effects Analysis

Conditional analysis within the Early Alert System: ATET Estimates

	Sub-group of students	Coefficient ($)	Robust Standard Error ($)	Z-value	Sample Size
Discontinued/ Continue	Not Identified	1879[a]	63.06	29.79	3,142
	Identified	4589[a]	96.59	47.51	9,099
Discontinued/ Complete	Not Identified	5288[a]	150.21	35.21	183
	Identified	7540[a]	408.47	18.46	1,623

Appendix I: Treatment Effects Match Summaries

Cost of discontinuing

	Raw	Matched
Number of observations	14012	28024
Treated observations	9698	14012
Control observations	4314	14012

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0356	0	1.0302	1
Age	-0.6706	-0.1656	0.8188	1.0358
Age^2	-0.5872	-0.1349	0.6347	0.9718
ATSI	0.0234	0	1.5795	1
Domestic Fee	-0.038	0	0.2545	1
Prior Study	0.0221	0	1.0489	1
Diploma	-0.0404	0	0.4904	1
Advanced Diploma	-0.3172	0	0.2308	1
Bachelor (Graduate Entry)	-0.1299	0	0.6778	1
Bachelor (Honours)	-0.0663	0	0.5835	1
School 1	-0.0268	0	0.9319	1
School 2	-0.0914	0	0.8396	1
School 3	0.1092	0	1.3774	1
School 4	0.0653	0	1.1702	1
School 5	0.0323	0	1.0769	1
School 6	0.0249	0	1.0364	1
School 8	0.0595	0	1.2626	1
School 9	-0.1695	0	0.5666	1
Units Enrolled	1.1896	0.4377	1.4785	1.1267
Starting Period 2	-0.0187	0	0.9441	1
Starting Period 4	0.1599	0	1.2533	1
Starting Period 5	-0.061	0	0.8199	1
Starting Period 6	-0.0058	0	0.972	1
Starting Period 7	-0.21	0	0.8849	1
Starting Period 8	-0.0898	0	0.6542	1

Cost of not completing

	Raw	Matched
Number of observations	2460	4920
Treated observations	581	2460
Control observations	1879	2460

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0545	0	1.0676	1
Age	0.0055	-0.1317	1.0718	0.8298
Age^2	0.015	-0.1374	1.062	0.7717
ATSI	0.1184	0	5.6101	1
International Fee	0.3993	0	8.2656	1
Prior Study	0.4178	0	1.4379	1
Advanced Diploma	0.1059	0	1.7128	1
Bachelor (Graduate Entry)	0.208	0	2.9362	1
Bachelor (Honours)	0.2123	0	5.9932	1
School 1	-0.0348	0	0.9164	1
School 2	-0.1326	0	0.7674	1
School 3	-0.0937	0	0.7288	1
School 4	-0.2228	0	0.3555	1
School 5	0.0275	0	1.4361	1
School 6	-0.1653	0	0.8253	1
School 8	0.2231	0	1.5676	1
School 9	0.101	0	1.471	1
Units Enrolled	0.6576	0.5642	0.6396	0.6445
Starting Period 2	0.1874	0	1.5422	1
Starting Period 4	0.1095	0	1.1858	1
Starting Period 5	0.0521	0	2.1506	1
Starting Period 7	-0.3818	0	0.292	1

Overall effect of EAS on revenue

	Raw	Matched
Number of observations	14012	28024
Treated observations	9698	14012
Control observations	4314	14012

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0356	0	1.0302	1
Age	-0.6706	-0.1656	0.8188	1.0358
Age^2	-0.5872	-0.1349	0.6347	0.9718
ATSI	0.0234	0	1.5795	1
Domestic Fee	-0.038	0	0.2545	1
Prior Study	0.0221	0	1.0489	1
Diploma	-0.0404	0	0.4904	1
Advanced Diploma	-0.3172	0	0.2308	1
Bachelor (Graduate Entry)	-0.1299	0	0.6778	1
Bachelor (Honours)	-0.0663	0	0.5835	1
School 1	-0.0268	0	0.9319	1
School 2	-0.0914	0	0.8396	1
School 3	0.1092	0	1.3774	1
School 4	0.0653	0	1.1702	1
School 5	0.0323	0	1.0769	1
School 6	0.0249	0	1.0364	1
School 8	0.0595	0	1.2626	1
School 9	-0.1695	0	0.5666	1
Units Enrolled	1.1896	0.4377	1.4785	1.1267
Starting Period 2	-0.0187	0	0.9441	1
Starting Period 4	0.1599	0	1.2533	1
Starting Period 5	-0.061	0	0.8199	1
Starting Period 6	-0.0058	0	0.972	1
Starting Period 7	-0.21	0	0.8849	1
Starting Period 8	-0.0898	0	0.6542	1

Cost of discontinuing given not EAS identified

	Raw	Matched
Number of observations	3142	6284
Treated observations	1961	3142
Control observations	1181	3142

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0431	0	1.0539	1
Age	0.1521	0.1234	0.9601	1.079
Age2	0.1273	0.1197	0.9915	1.1517
Prior Study	-0.1515	0	0.6505	1
Diploma	-0.0154	0	0.7532	1
Advanced Diploma	-0.082	0	0.8167	1
Bachelor (Graduate Entry)	0.1146	0	1.4284	1
School 1	-0.074	0	0.814	1
School 2	-0.0329	0	0.9442	1
School 3	-0.1525	0	0.6133	1
School 4	-0.02	0	0.9443	1
School 5	0.0919	0	1.2892	1
School 6	0.1141	0	1.1679	1
School 8	-0.0095	0	0.9637	1
School 9	-0.0136	0	0.969	1
Units Enrolled	-0.0819	-0.0754	0.4456	0.6342
Starting Period 2	-0.1698	0	0.4248	1
Starting Period 4	-0.342	0	0.6374	1
Starting Period 5	-0.1867	0	0.5295	1
Starting Period 6	0.0056	0	1.031	1
Starting Period 7	0.841	0	1.143	1

Cost of discontinuing given EAS identified

	Raw	Matched
Number of observations	9099	18198
Treated observations	6357	9099
Control observations	2742	9099

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	0.0185	0	0.986	1
Age	0.0274	0.0796	1.05	1.2026
Age2	0.0286	0.0892	1.0572	1.3117
ATSI	-0.0268	0	0.6046	1
International Fee	0.0422	0	1.5337	1
Prior Study	-0.0859	0	0.8432	1
Advanced Diploma	0.0022	0	1.022	1
Bachelor (Graduate Entry)	0.0695	0	1.3355	1
Bachelor (Honours)	0.0019	0	1.0241	1
School 1	-0.0184	0	0.9502	1
School 2	-0.0582	0	0.8838	1
School 3	-0.0174	0	0.9559	1
School 4	-0.091	0	0.8228	1
School 5	0.028	0	1.0641	1
School 6	0.1021	0	1.164	1
School 8	0.0243	0	1.0886	1
School 9	0.0027	0	1.0115	1
Units Enrolled	0.1168	0.0093	0.5928	0.7973
Starting Period 2	-0.1838	0	0.5943	1
Starting Period 4	-0.0528	0	0.942	1
Starting Period 5	-0.0383	0	0.8711	1
Starting Period 6	-0.0297	0	0.8677	1
Starting Period 7	0.5647	0	1.9018	1
Starting Period 8	0.0545	0	3.0098	1

Cost of not completing given not EAS identified

	Raw	Matched
Number of observations	183	366
Treated observations	40	183
Control observations	143	183

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Age	0.0685	-0.0298	0.9556	0.8307
Age2	0.0534	-0.05	0.9083	0.7729
Advanced Diploma	-0.3049	0	0.5529	1
Bachelor (Graduate Entry)	0.3473	0	2.0109	1
Units Enrolled	0.5129	0.3658	0.2207	0.1225
Starting Period 2	-0.1028	0	0.8478	1
Starting Period 4	0.2627	0	1.3053	1

Cost of not completing given EAS identified

	Raw	Matched
Number of observations	1623	3246
Treated observations	451	1623
Control observations	1172	1623

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0626	0	1.0737	1
Age	0.1141	-0.046	1.1794	0.8825
Age^2	0.1158	-0.0548	1.2071	0.8445
International Fee	0.4402	0	6.4705	1
Prior Study	0.4223	0	1.4464	1
Advanced Diploma	0.2453	0	7.9338	1
Bachelor (Graduate Entry)	0.1087	0	3.8646	1
Bachelor (Honours)	0.1782	0	4.4429	1
School 1	-0.0281	0	0.9309	1
School 2	-0.1336	0	0.7691	1
School 3	-0.1183	0	0.6916	1
School 4	-0.2245	0	0.3968	1
School 5	0.045	0	1.7283	1
School 6	-0.3197	0	0.635	1
School 8	0.2634	0	1.6273	1
School 9	0.1994	0	2.4892	1
Units Enrolled	0.5237	0.4658	0.5946	0.6
Starting Period 2	0.2345	0	1.8239	1
Starting Period 4	0.0217	0	1.0375	1
Starting Period 7	-0.2664	0	0.4226	1

Effect of EAS on revenue within schools

School – Base Case

	Raw	Matched
Number of observations	1101	2202
Treated observations	760	1101
Control observations	341	1101

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0819	0	1.0047	1
Age	-0.6536	-0.152	0.7392	0.9877
Age2	-0.5958	-0.1324	0.5688	0.8782
Domestic Fee	-0.1354	0	0.26	1
Prior Study	0.0582	0	1.13	1
Bachelor (Graduate Entry)	-0.1092	0	0.6196	1
Units Enrolled	1.0496	0.5071	1.4028	1.1327
Starting Period 2	0.0135	0	1.0293	1
Starting Period 4	0.0898	0	1.1337	1
Starting Period 5	-0.0457	0	0.8526	1
Starting Period 6	0.0632	0	1.2949	1
Starting Period 7	-0.2391	0	0.8256	1
Starting Period 8	-0.0159	0	0.9289	1

School 1

	Raw	Matched
Number of observations	1409	2818
Treated observations	951	1409
Control observations	458	1409

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0239	0	1.0283	1
Age	-0.6954	-0.1244	0.9757	1.135
Age2	-0.5972	-0.0895	0.7887	1.0836
Prior Study	0.0462	0	1.0843	1
Advanced Diploma	-0.1436	0	0.2446	1
Bachelor (Honours)	-0.0673	0	0.8262	1
Units Enrolled	1.2785	0.6232	1.6326	1.5452
Starting Period 2	-0.015	0	0.9537	1
Starting Period 4	0.2078	0	1.3236	1
Starting Period 5	-0.1126	0	0.6905	1
Starting Period 6	0.0119	0	1.051	1
Starting Period 7	-0.1529	0	0.8802	1
Starting Period 8	-0.1639	0	0.5152	1

School 2

	Raw	Matched
Number of observations	2086	4172
Treated observations	1345	2086
Control observations	741	2086

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0397	0	1.0394	1
Age	-0.5886	-0.0857	0.8622	1.078
Age2	-0.5181	-0.0608	0.7067	1.0328
Prior Study	-0.0786	0	0.8859	1
Advanced Diploma	-0.3046	0	0.4432	1
Bachelor (Honours)	-0.1472	0	0.4561	1
Units Enrolled	1.2502	0.4549	2.0548	1.3271
Starting Period 2	-0.0275	0	0.9195	1
Starting Period 4	0.1193	0	1.202	1
Starting Period 5	0.0678	0	1.2419	1
Starting Period 6	0.0252	0	1.111	1
Starting Period 7	-0.2512	0	0.8396	1
Starting Period 8	-0.031	0	0.8877	1

School 3

	Raw	Matched
Number of observations	14012	28024
Treated observations	9698	14012
Control observations	4314	14012

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0356	0	1.0302	1
Age	-0.6706	-0.1656	0.8188	1.0358
Age2	-0.5872	-0.1349	0.6347	0.9718
ATSI	0.0234	0	1.5795	1
Domestic Fee	-0.038	0	0.2545	1
Prior Study	0.0221	0	1.0489	1
Diploma	-0.0404	0	0.4904	1
Advanced Diploma	-0.3172	0	0.2308	1
Bachelor (Graduate Entry)	-0.1299	0	0.6778	1
Bachelor (Honours)	-0.0663	0	0.5835	1
School 1	-0.0268	0	0.9319	1
School 2	-0.0914	0	0.8396	1
School 3	0.1092	0	1.3774	1
School 4	0.0653	0	1.1702	1
School 5	0.0323	0	1.0769	1
School 6	0.0249	0	1.0364	1
School 8	0.0595	0	1.2626	1
School 9	-0.1695	0	0.5666	1
Units Enrolled	1.1896	0.4377	1.4785	1.1267
Starting Period 2	-0.0187	0	0.9441	1
Starting Period 4	0.1599	0	1.2533	1
Starting Period 5	-0.061	0	0.8199	1
Starting Period 6	-0.0058	0	0.972	1
Starting Period 7	-0.21	0	0.8849	1
Starting Period 8	-0.0898	0	0.6542	1

School 4

	Raw	Matched
Number of observations	1677	3354
Treated observations	1223	1677
Control observations	454	1677

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.1091	0	0.9925	1
Age	-0.5786	-0.1519	0.8202	1.0523
Age2	-0.5042	-0.1209	0.6568	0.9816
Prior Study	0.0993	0	1.2209	1
0.course type	-0.1398	0	0.3787	1
Units Enrolled	0.9797	0.3681	1.0268	0.8856
Starting Period 2	-0.0564	0	0.8387	1
Starting Period 4	0.2566	0	1.4154	1
Starting Period 5	-0.1738	0	0.5628	1
Starting Period 6	-0.0729	0	0.6957	1
Starting Period 7	-0.2031	0	0.896	1
Starting Period 8	-0.1231	0	0.571	1

School 5

	Raw	Matched
Number of observations	1753	3506
Treated observations	1245	1753
Control observations	508	1753

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	0.0378	0	0.9823	1
Age	-0.6242	-0.16	0.9964	1.1267
Age2	-0.5404	-0.123	0.777	1.0342
Prior Study	0.0463	0	1.1395	1
Bachelor (Graduate Entry)	-0.2929	0	0.9694	1
Units Enrolled	1.0388	0.4815	1.4637	1.0748
Starting Period 2	0.0678	0	1.2171	1
Starting Period 4	0.1965	0	1.3996	1
Starting Period 5	-0.0576	0	0.8466	1
Starting Period 6	-0.0194	0	0.9397	1
Starting Period 7	-0.2507	0	0.8593	1
Starting Period 8	-0.1642	0	0.5976	1

School 6

	Raw	Matched
Number of observations	2940	5880
Treated observations	2065	2940
Control observations	875	2940

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0693	0	1.1805	1
Age	-0.6096	-0.0764	0.9996	1.0888
Age2	-0.545	-0.0566	0.8114	1.0434
ATSI	0.0462	0	1.4977	1
Prior Study	0.0874	0	1.2304	1
Bachelor (Graduate Entry)	-0.346	0	0.4137	1
Units Enrolled	1.1814	0.3303	1.8446	1.1789
Starting Period 2	-0.0039	0	0.9872	1
Starting Period 4	0.172	0	1.3293	1
Starting Period 5	-0.0934	0	0.7502	1
Starting Period 6	-0.0122	0	0.9366	1
Starting Period 7	-0.249	0	0.9106	1
Starting Period 8	-0.0851	0	0.6098	1

School 8

	Raw	Matched
Number of observations	807	1614
Treated observations	599	807
Control observations	208	807

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	0.062	0	0.8578	1
Age	-0.7966	-0.1002	1.1999	1.1115
Age2	-0.7173	-0.0743	1.0258	1.1233
Prior Study	0.1089	0	1.3581	1
4.course_type	0	0	0	0
Units Enrolled	0.9101	0.2042	0.7377	0.6913
Starting Period 2	-0.1253	0	0.6256	1
Starting Period 4	0.2111	0	1.2551	1
Starting Period 5	-0.0223	0	0.9044	1
Starting Period 7	-0.336	0	0.8781	1

School 9

	Raw	Matched
Number of observations	919	1838
Treated observations	504	919
Control observations	415	919

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.01	0	1.0077	1
Age	-0.827	-0.2283	0.6641	0.8433
Age2	-0.7347	-0.2143	0.5471	0.7775
Prior Study	0.1442	0	1.3919	1
Advanced Diploma	-1.0085	0	0.6929	1
Units Enrolled	1.4169	0.463	2.3058	1.5758
Starting Period 2	0.0502	0	1.2199	1
Starting Period 4	0.0669	0	1.0665	1
Starting Period 5	0.1239	0	2.4249	1
Starting Period 7	-0.2003	0	0.8746	1
Starting Period 8	0.0219	0	1.2318	1

Effect of EAS on revenue by years

Year 1

	Raw	Matched
Number of observations	13343	26686
Treated observations	9085	13343
Control observations	4258	13343

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0577	0	1.0502	1
Age	-0.7079	-0.1783	0.8019	1.0144
Age2	-0.6191	-0.1483	0.6172	0.9441
ATSI	0.0247	0	1.6121	1
Prior Study	0.0117	0	1.0264	1
Diploma	-0.0382	0	0.5167	1
Advanced Diploma	-0.3231	0	0.1864	1
Bachelor (Graduate Entry)	-0.1378	0	0.6585	1
Bachelor (Honours)	-0.0662	0	0.5864	1
School 1	-0.0335	0	0.9158	1
School 2	-0.0989	0	0.8265	1
School 3	0.1162	0	1.4024	1
School 4	0.0698	0	1.1803	1
School 5	0.0311	0	1.0731	1
School 6	0.0203	0	1.0293	1
School 8	0.0574	0	1.2509	1
School 9	-0.1622	0	0.5657	1
Units Enrolled	1.2591	0.4577	1.4409	1.1061
Starting Period 2	-0.0284	0	0.9131	1
Starting Period 4	0.1581	0	1.2576	1
Starting Period 5	-0.0668	0	0.805	1
Starting Period 6	-0.0051	0	0.9754	1
Starting Period 7	-0.1831	0	0.9076	1
Starting Period 8	-0.0809	0	0.6882	1

Year 2

	Raw	Matched
Number of observations	2407	4814
Treated observations	414	2407
Control observations	1993	2407

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Gender	-0.0741	0	1.1364	1
Age	-0.1251	-0.2488	0.98	0.7613
Age2	-0.1114	-0.2529	0.9876	0.6806
Prior Study	0.266	0	1.649	1
Advanced Diploma	-0.0239	0	0.9321	1
Bachelor (Graduate Entry)	0.0359	0	1.1253	1
School 1	0.1161	0	1.3814	1
School 2	-0.0646	0	0.9076	1
School 3	0.0424	0	1.1817	1
School 4	0.0434	0	1.1816	1
School 5	-0.0049	0	0.9893	1
School 6	-0.1708	0	0.8484	1
School 8	0.1284	0	1.7401	1
School 9	0.0074	0	1.0221	1
Units Enrolled	0.4015	0.1377	0.7441	0.5822
Starting Period 2	0.1933	0	1.7376	1
Starting Period 4	0.1815	0	1.1103	1
Starting Period 5	0.0367	0	1.1506	1
Starting Period 6	-0.0555	0	0.7181	1
Starting Period 7	-0.7505	0	0.1577	1

Year 3

	Raw	Matched
Number of observations	223	446
Treated observations	28	223
Control observations	195	223

	Standard Difference		Ratio	
	Raw	Matched	Raw	Matched
Age	-0.1622	-0.095	0.6366	0.5041
Age2	-0.2131	-0.1744	0.4932	0.3771
Advanced Diploma	-0.1557	0	0.915	1
School 2	0.3111	0	1.1938	1
School 3	0.3226	0	2.2709	1
School 6	-0.3792	0	0.671	1
Units Enrolled	-0.1243	-0.2794	0.2106	0.1532
Starting Period 2	0.1423	0	1.5862	1