A statistical investigation into the contextual background of Scottish undergraduates

Nathan Patrick Luke Burns

Department of Mathematics and Statistics

University of Strathclyde

A thesis submitted for the degree of

Doctor of Philosophy

2025

Author's Declaration

This thesis is submitted to the University of Strathclyde for the degree of Doctor of Philosophy in the Faculty of Science.

I, Nathan Patrick Luke Burns, confirm that the contents of this thesis is the result of my own original research. It has been composed by myself and has not been previously submitted for examination which has led to the award of a degree.

Date: 10/06/2025

Signed:

Parts of the work contained within this thesis have been published or presented at national or international conferences:

Publications

N. P. Burns, D. Young, A. Sherriff, P. Black, A. Blackshaw, L. A. Kelly, "Tracking the success of contextual offer students at one Scottish Higher Education Institution", *Higher Education Quarterly*, 2025 [accepted with revisions].

National Conferences

N. P. Burns, A. Sherriff, L. A. Kelly, D. Young, "The postcode lottery: how access to Scotland's most advanced mathematics qualification at school-level impacts students' degree-level performance", *British Society for Research into Learning Mathematics*, Online, March 2024.

N. P. Burns, A. Sherriff, D. Young, L. A. Kelly, "A comparison of generalised linear models to determine the academic success of university students from traditionally disadvantaged backgrounds", *Royal Statistical Society*, Brighton, United Kingdom, September 2024.

N. P. Burns, A. Sherriff, L. A. Kelly, D. Young, "A comparison of semi-parametric and parametric survival methods on student dropouts from higher education.", *Royal Statistical Society*, Brighton, United Kingdom, September 2024.

International Conferences

N. P. Burns, A. Sherriff, L. A. Kelly, D. Young, "The impact of student dropout prediction on Scotland's Widening Access Agenda", *7th World Conference on Research in Teaching and Education*, Munich, Germany, May 2024.

This book and all its contents are dedicated to Aida Ferrer Aguilar.

You are the reason this exists. *Te amo siempre mi pulpo.*

Acknowledgements

I would like to thank Pat and Mike Bennett for their kind funding contributions that has supported this work (and my life) for the last 4.5 years. Thanks to Kayleigh Kerr who started this research off and passed me the baton.

I could not have possibly carried on without the ceaseless support and endless patience of my supervisors: Louise Kelly, David Young, and Andrea Sherriff. Thank you for your wonderful wisdom, you are all simply divine.

A warm and sincere thank you to the Widening Access and Strategy and Policy teams: Al Blackshaw, Peter Black, and Stephanie McKendry. A special thank you to Neil Shaw who combed through eight years worth of the University's prospectus data and manually recorded each programmes entry requirements - you have patience I cannot fathom. Special thanks to the contextual offer research group, led by Vikki Boliver and Jonathan Schulte, who gave me my first big break in publishing research.

Special thanks also goes to the PhD crew that helped keep this thesis afloat, as well as my own sanity from falling overboard. In order of appearance: Ainsley, Jon, Florence, Kate, Ryan Stewart, Tunde, Connor, Matthew, Natalia, all of the Michaels, Hollie, Ann, Matty, Dan, Claire and the 5s team. Thank you to my fellow office mates Sophie, Gyu Ho, Kamel, Ryan Kirkpatrick, Asmaa and to Hasan for your spare socket. Thank you to all my friends and family who supported me emotionally, there are too many to name but you know who you are.

Lastly, I would like to thank my family. You have all supported me so much in my life. To Mum and Dad, you have always pushed for us to achieve our best, while never failing to have our backs. I hope you can be as proud of this as I am.

Abstract

Background

Since 2016, Scotland's "Widening Access Agenda" has primarily focused on increasing the proportion of those from the 20% most deprived areas in Scotland who gain admission to higher education. In 2024, this conversation shifted towards not just disadvantaged students' access to university, but their academic performance once they are on-programme. This thesis aims to address the increased interest in students' academic outcomes by analysing these alongside students' socio-economic and demographic backgrounds at one Scottish higher education institution - the University of Strathclyde. This thesis is unique in that it is the first known in-depth, temporal analysis of student registration records at the population-level within the United Kingdom. In addition to being of policy importance to wider-Scotland, the results of this thesis are of also operational importance to the University of Strathclyde, which has a target of 90-95% retention for first-year undergraduates by 2030.

Aims

To explore 10 years' worth of registration records on Scottish-school leavers at the University of Strathclyde in general. To determine whether or not contextual offer students are achieving similar levels of academic success as their standard offer peers. To measure the association between students' academic outcomes and their prior attainment, demographics, and socio-economic background. To identify which statistical modelling techniques most appropriately fit the data.

Data

This thesis analyses 10 years' worth of registration records at the University of Strathclyde (2012/13 - 2021/22). These data are provided by the Strategy & Policy team with some additional data from the University of Strathclyde's Widening Access team. Area-level deprivation is measured using the Scottish Index of Multiple Deprivation (SIMD), where the 20% most deprived areas were denoted "SIMD Quintile 1". The reproducibility of the results is of vital importance, hence this thesis details precisely how datasets were gathered, cleaned and joined. The data were filtered to only consider "Scottish-school leavers" - the population of interest - which contained 18,988 unique students.

Methods

This thesis presents the theory behind regression and survival modelling techniques. Three generalised linear regression models were examined: the Logistic, Modified Poisson, and Log-Binomial. Similarly, three survival models were examined: the Logit Discrete Time-to-Event, Cox-Proportional Hazards, and Parametric Weibull. Models explored the associations between a successful/unsuccessful outcome at university and students' prior attainment from secondary education, their socio-economic background, and demographic background. Relevant models were compared to one another on the basis of their estimated effects and goodness-of-fit to identify the most appropriate modelling techniques.

Results

Between 2012/13 and 2021/22, around 90% of school-leavers were retained after the end of their first academic session, around 74% of school-leavers completed their Bachelor's with Honours degree within four years and around 9% of school-leavers dropped-out of the University. The majority of these drop-outs occurred in the first academic session (6%). Each of these academic outcomes were significantly affected by a student's SIMD Quintile and prior attainment from secondary education. School-leavers from SIMD Quintile 1 had significantly lower chances of a successful outcome at the University compared to their peers from SIMD Quintiles 2-5, even when they had the same levels of prior attainment. Students who likely received a standard offer to the University of Strathclyde were 8.3% more likely to be retained at the end of first year and 18.6% more likely to complete their Bachelor's with Honours degree within four years, compared to students who likely received a contextual offer. Both regression and survival methods adequately fit the data, although the regression models had various issues related to the interpretation of estimated effects. These effects were mitigated when using an academic outcome that was rarer, i.e. drop-out rather than retention or completion. Discrete survival methods were the most appropriate model fits, however, if regression methods were to be applied it seems unlikely that incorrect conclusions would be drawn.

Future analyses

Gaps between the academic outcomes of students from different socio-economic backgrounds (measured using SIMD) have been identified, even when they had similar levels of prior attainment. There is huge potential for future research into student registration data that could assist the university and wider-Scotland to achieve targets on Widening Access and the academic outcomes of students more generally. Future analyses should examine the associations with other key explanatory variables, such as university-level attainment and other measures of socio-economic background. These data could be used to develop early-risk prediction models to assist the university in making more targetted interventions. Data from more recent cohorts could be examined to measure any potential impact from the COVID-19 pandemic. It could also be explored whether the current entry requirement thresholds are appropriate or could be adjusted to reflect what is now known about the relationship between prior attainment and a successful outcome at the university. It is hoped that the results of this thesis provide a blueprint for analysing student outcomes for teams at the University of Strathclyde as well as at other institutions across the United Kingdom.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Widening Access (and Participation)" is an umbrella term popular in the United Kingdom for policies and interventions that aim to reduce inequalities in education. For example, increasing the number of students represented in higher education from traditionally under-represented groups, targetting support towards socio-economically disadvantaged students, or the closing of attainment gaps. This thesis will primarily focus on Widening Access efforts within the Scottish higher education system which has received increasing levels of attention from the Scottish Government, higher education institutions and other stakeholders in the education sector, both in Scotland and in the Rest of the United Kingdom (RUK).

Since 2016, the Scottish Government and Scottish universities have been working towards their goals on Widening Access to higher education for students from socio-economically deprived backgrounds [3–5]. The primary objective is to, by 2030, increase the proportion of full-time, first-degree undergraduate students registered at Scottish higher education institutions from the 20% most deprived areas in Scotland such that they are equally represented (i.e. equal to 20%) [3, 6]. Area-level deprivation in Scotland is measured using the Scottish Index of Multiple Deprivation (SIMD) where the 20% most deprived areas are defined as those that come under SIMD Quintile 1 (also known as "SIMD20" in some literature). One of the many tools used to increase the proportion of SIMD Quintile 1 and other disadvantaged students in higher education is the practice of "contextualised admissions" (also known as "contextual offers"), where the entry requirements to a degree programme are marginally lowered for students from deprived backgrounds [3]. Contextual offers will be a key topic that is explored in Chapter 8.

As of 2024, the proportion of SIMD Quintile 1 students in Scottish institutions stood at 16.5% [7]. While much of the emphasis on measuring progress for Widening Access has been focused on admissions, comparatively less emphasis has been placed on the performance of those students once they are on-programme. There is, however, some indication that this is changing after the new Commissioner for Fair Access (2024) recommended that equal weight be given to disadvantaged students' academic outcomes [7]. This thesis aims to contribute to the research and literature on the outcomes of Widening Access students in higher education. It will do so by analysing ten years' worth of registration records (2012/13 - 2021/22) at one Scottish higher education institution: the University of Strathclyde.

The results from this thesis are also of operational importance to the University of Strathclyde. The University has set Key-Performance Indicators (KPI) including that, by 2030, it will achieve a target of 90-95% retention for first-year undergraduates [KPI 2 - [8]]. Understanding which factors affect whether or not a student will drop out of university or continue their studies is therefore of vital importance. The data used in this thesis primarily comes from the registration records provided by the University of Strathclyde's Strategy & Policy team with some additional data from the University of Strathclyde's Widening Access team. Both teams have expressed keen interest in the results to assist with their operations. As a result, a key aim of this thesis was the reproducibility of the results by detailing precisely how datasets were gathered, cleaned and joined. It is hoped that the work provides a blueprint for analysing student outcomes for teams at the University of Strathclyde as well as at other institutions across the United Kingdom.

1.1 Thesis Outline

Chapter 2 will summarise the Scottish education system, how disadvantage is measured, the inequality present within the UK's education systems (which forms the rationale behind the Widening Access agenda) and the progress that has been made towards Scotland's Widening Access targets. Chapter 3 will detail how data on the target population (school-leavers) were gathered, joined, and cleaned to derive the **School-leavers dataset**. It will also define subsets of the **School-leavers dataset** used for each analysis chapter (7, 8, and 9). Chapter 4 will define the relevant outcome and explanatory variables, some of which will be used in the analyses chapters and some of which are planned to be used in future publications. Chapter 5 will briefly explore the variables contained within the **School-leavers dataset**, and their relationships with one another, such that the reader can understand the make-up of students in the dataset. These explorations will also aid the interpretation of analyses chapters. Chapter 6 will contain the relevant theory behind the regression and survival methods applied to the **School-leavers dataset**. Chapter 7 will measure the association between holding a qualification in Advanced Higher Mathematics and students' chances of a successful outcome at the University. It will also measure whether or not this effect is moderated by the recommendation (or lack thereof) of Advanced Higher Mathematics prior to entry. Chapter 8 will compare the academic outcomes of standard and contextual offer students. Chapter 9 will attempt to build a survival framework for analysing the time until students drop out of the University. All analyses chapters (7, 8, and 9) will also measure the effect that prior attainment, socio-economic background, and other factors, have on students' academic outcomes. Relevant literature will be referenced throughout the thesis. In particular, Chapter 2 will highlight previous research on measuring the academic outcomes of students at schools and in higher education. The beginning of each analysis chapter will also contain a short motivational section which will cite literature relevant to the area of analysis.

1.2 Aims of the Thesis

The aims of the thesis are:

1. To explore the demographic, socio-economic and prior attainment information of school-leavers at the University of Strathclyde.
2. To determine whether or not contextual offer students are achieving similar levels of academic success as their standard offer peers.

3. To measure how much academic success/failure is affected by a student's:
   (a) Socio-economic background,
   (b) Prior attainment from secondary education,
   (c) Demographics,
   (d) Choice of degree programme.
4. To determine what the most appropriate method is for modelling the effects on academic outcomes.

Aims (1), and (3) will be addressed throughout the thesis, primarily in Chapters 5, 7, 8 and 9. Aim (2) will be specifically addressed in 8. The final aim (4) will be addressed in the conclusions of Chapter 10 which summarises the findings from Chapters 5, 7, 8 and 9. Future modelling approaches and research questions will be proposed based on these findings.

Chapter 2

Scottish Education and Widening Access in Context

Before analysing the University of Strathclyde data, it is necessary to understand the context behind Widening Access efforts in Scotland. This chapter will summarise the Scottish education system (Section 2.1), how those who are disadvantaged can be identified (Section 2.3), the inequalities present at each stage of Scottish/UK education (Section 2.4), and finally, what progress has been made towards the Scottish Widening Access targets since their introduction in 2017 (Section 2.5). Frequent reference to SIMD is made throughout the chapter and is defined in Section 2.3.1. While this chapter focuses primarily on inequality and under-representation found within Scottish education, relevant examples from the rest of the United Kingdom (RUK) are also highlighted here where appropriate. This chapter also touches upon the impact of the COVID-19 pandemic on Scottish and University of Strathclyde students.

2.1 The Scottish Educational Context

Education is devolved within Scotland, meaning it is governed by the Scottish Government rather than the UK Government. As such, there are differences in the structure and policies of the educational systems in Scotland when compared to England, Wales, or Northern Ireland. In Scotland, there are four stages of education that a student can progress through: Pre-school, which covers those aged 2-4 years old; Primary education, of which there are seven years – P1 (typically aged 5 at the beginning of the year) to P7 (aged 11); Secondary education (or high-school), which has four compulsory years – S1 (aged 12) to S4 (aged 15) – and two optional years – S5 (aged 16) to S6 (aged 17); Tertiary education which covers both colleges and universities for those typically aged 18 or older. Tertiary education delivered at a university is commonly referred to as higher education, while at a college it is referred to as further education.

The majority of young people in Scotland choose stay on until the late stages of secondary education (S5 and S6) [9]. After secondary school, students can choose to apply to college or university depending on whether or not they satisfy the relevant entry requirements. College studies typically are two years in duration, where in the first year a student completes their Higher National Certificate (HNC) and in the second year a student completes their Higher National Diploma (HND). Students from secondary school or college may choose to apply to university where they can sit their Bachelor's with Honours degree (four years duration) followed by a Master's degree (one year duration). This system is similar to the rest of the UK with some slight differences in the structure and length of each stage, as well as the ages of students that enter each stage. For example, students in Scotland leave secondary education aged 17-18, whereas in the RUK this is 18-19. In Scotland, Bachelor's with Honours degrees typically last four years but last three years in the rest of the UK, meaning that all UK students typically leave higher education at the same age (21-22).

2.1.1 Attainment Required for Entry to Higher Education

To gain access to higher education institutions within the UK, students must gain the relevant attainment in Scottish Credit and Qualifications Framework (SCQF) level 6+, or UK level 3+, qualifications which are taken in the later years of secondary education [10, 11]. In Scotland, the relevant qualifications are, typically, Highers and Advanced Highers, while in the rest of the UK these are A-levels. The Scottish Qualifications Authority (SQA) is the awarding body for Highers and Advanced Highers as well as other secondary and vocational examinations [12].

This thesis will exclusively focus on the Higher and Advanced Higher examinations which most Scottish-domiciled students take across Scotland. Highers are taken over one academic year, with five subjects being the usual maximum, and typically in S5 and S6. This means that at the point at which most Scottish students apply to higher education, admissions services can rely on the results from formal examinations rather than just predicted grades. This differs from the RUK where students sit A-levels over two academic years [13].

In the final year of secondary education, S6, students have the option of taking Advanced Highers. Students who sit Advanced Highers can sometimes be eligible for second year entry to some undergraduate degree programmes [14–16]. These qualifications are typically aimed at students who wish to apply to the most competitive programmes and institutions or wish to better prepare themselves for higher education level of study [15, 17]. Advanced Highers add a complicated layer to admission processes, since they may not be as accessible nor even necessary for all degree programmes. Advanced Highers and their effect on degree outcomes is examined in Chapter 7.

2.1.2 Applying to Higher Education

Students are required to submit applications to the Universities and Colleges Admissions Service (UCAS) – a centralised organisation for processing higher education applications – to their institutions of choice. Students are encouraged to apply in their final year of secondary education, although they can apply earlier than this.

Deadlines for applications are in October for competitive programmes (such as those at Oxford, Cambridge, or in medicine) and January for most other programmes [18]. Students can make five choices in their application, where a "choice" is for a specific programme at a specific institution. Students can apply for as many or as few programmes as they like at a single institution, up to the limit of five [18]. If an applicant is unsuccessful in all choices listed on their October/January application, they are permitted one final application submitted between February and July [19].

Following the deadlines, institutions process applications to their programmes and issue decisions (offers or rejections); applicants receive these via UCAS as they are processed and thus at any time they may have one or more outstanding decisions. The deadline for institutions to make decisions is typically in May for applications received in January, and July for those received after February [20]. There are two types of offer issued, conditional and unconditional.

Conditions on an offer can include attainment of certain qualifications, attending a summer school or widening access programme, or sitting an additional test prior to entry (normally in the August/November directly prior to registration) [21].

2.1.3 Scotland's Higher Education Institutions

Scotland has 19 officially recognised universities, with the Principals from each forming the group: "Universities Scotland" [22]. Each of the Scottish universities can be informally grouped into either the four "ancient" universities founded prior to 1600, the four "old" universities founded prior to 1992, or the remaining "new" universities founded post-1992. Note that the Open University in Scotland does not fit into any of these groups, and is sometimes not included in Scotland's official count of institutions in certain contexts. Two of Scotland's ancient universities: the University of Glasgow and the University of Edinburgh; are founding members of the Russell Group which describes itself as a collection of the United Kingdom's most world-class and research-intensive universities [23]. Though not recognised in any official capacity, the Russell Group is very much acknowledged in public and political discourse.

Being founded in 1796 and awarded university status in 1964, the University of Strathclyde belongs to Scotland's group of Old universities. The University of Strathclyde has an international reputation for being socially-progressive but highly competitive in terms of its entry requirements (see Chapter 5 Figure 5.12). It is Scotland's third largest university by total population of UK students [24] and regularly ranks within the top 50% of UK universities [25].

Since 2005, the Scottish Government has allocated funds to Scotland's higher education institutions through the Scottish Funding Council (SFC) [26]. Institutions are held accountable for the funds they are allocated by the SFC through Outcome Agreements [27]. The SFC also determine the number of funded places for students at institutions. Students that can prove residence in Scotland for at least three years at the time of registration (classified as "Home" students) are eligible for their tuition fees to be paid [28]. The SFC determine the number of students for whom fees will be available and higher education institutions are set an annual target via their Outcome Agreement with the SFC; they should not exceed these targets meaning there is a limit on the number of Home students that can be registered. Note that targets for Widening Access students (see Section 2.5.1) are included within overall target numbers but will have sub-targets, and these are set in line with meeting the Commission on Widening Access's targets. In 2017/18 the Scottish Government allocated around £1.1 billion to the SFC for institutions and around £213 million for tuition fees to Scottish and EU domiciled students [29, 30].

2.2 Impact of the COVID-19 Pandemic on Education

The onset of the coronavirus pandemic in March 2020 brought significant disruption to all stages of education in the UK, the full impact of which will likely persist for years to come. In secondary-education, the cancellation of in-person exams [31, 32], the use of teacher-assessed grades and the reversal of the Scottish Government's "alternative certification model" [33–35], led to significant "grade inflation" [36] that affected the number of qualified applicants to higher education [37]. Higher education institutions had to cancel face-to-face teaching and examinations [38, 39], adapt to "open-book" assessments [40–42], and adopt a "blended-learning" approach which mixed in-person and online delivery of teaching [43–46]. In particular, a "no-detriment policy" was enacted at the University of Strathclyde, which gave extra favour to students to compensate for the disruption to their education [47–49]. There was particular concern that students from more socio-economically deprived backgrounds may have been disproportionately affected by the disruption [33, 35, 50]. Care should therefore be taken when interpreting any analyses on the academic outcomes of students who were affected/unaffected by the pandemic. A more detailed summary of the impact of the disruption is given in Appendix A Section A.1.

2.3 Defining Disadvantage using Contextual Indicators

In Widening Access literature, a "contextual indicator" is a discriminator between groups that can be used to identify individuals who are advantaged or disadvantaged. It is impossible that an indicator perfectly discriminates between groups of individuals given that "there is no hard boundary between the disadvantaged and advantaged" [51]. Ideally, the aim is to use indicators which result in the fewest possible false positives and false negatives. Here, a false-positive is an individual identified as disadvantaged when they are not, and a false-negative an individual identified as advantaged when they are not. What constitutes a suitable indicator, or an acceptable error-rate, is by its nature subjective and political.

Boliver and Gorard (2017-2022) [51–55] have provided the most in-depth analyses on the suitability of contextual indicators. They conceptualised suitability based on their definition of "reliability" and "validity". They proposed that a "valid" indicator is one which accurately identifies disadvantage – i.e. optimises the number of false-positives and false-negatives to an acceptable standard – whereas a "reliable" indicator is one which relies on verifiable information and has few missing data. For example, while coming from a rural area as an indicator is reliable – rural status can be verified from postcode data easily – its validity is questionable, since every individual that comes from a rural area is not necessarily disadvantaged, hence there are likely to be a large number of false-negatives and false-positives [52]. Boliver and Gorard [51, 52, 55] argued that reducing the number of false positives was more important than doing so for false negatives. In their view, the damage done to the "status quo" by the latter is relatively little, while by the former would "...at best [be] giving a misleading picture of how much good is being done, and at worst [be] doing more harm than good if most of the limited resources end up going to advantaged individuals rather than disadvantaged ones." [51].

The remainder of this section will identify the subset of the indicators which are used by the University of Strathclyde in its contextualised admissions policies [56]. It will end with a summary of the debate around which indicators are the most suitable in practice. A full list of the contextual indicators used in higher education institutions across Scotland is provided in Appendix A.2.4.

2.3.1 The Scottish Index of Multiple Deprivation (SIMD)

The Scottish Index of Multiple Deprivation (SIMD) is used as a contextual indicator at all Scottish higher education institutions and is the measure used to frame the each of the targets set by the Commission on Widening Access [3] (also see Section 2.5.1).

SIMD is a multi-layered measure of how deprived an area – known as a "data zone" – is within Scotland [57–59]. Roughly 700-800 people live within each data zone. SIMD ranks the 6976 data zones in a weighted calculation using seven measures (or "domains") of deprivation: income (28%); employment (28%); education, skills and training (14%); health (14%); geographic access to services (9%); crime (5%); and housing (2%) [58]. This aggregation means that SIMD is an indicator of multiple deprivation and not a measure of how poor or rich a given area is [58]. The ranked areas can be grouped together, often into deciles (ten groups) or quintiles (five groups). The latter is the grouping used most often in practice and in literature, where SIMD Quintile 1 (also referred to as "SIMD20" in some cases) refers to the 20% most deprived areas and SIMD Quintile 5 refers to the 20% least deprived areas. Visual examples are given in Figures 2.1 and 2.2. The University of Strathclyde considers anyone from SIMD Quintiles 1 or 2 as eligible for a contextual offer [56] as part of its outcome agreement with the SFC (for example, see the 2021 outcome agreement [60]). In other words, anyone from the 20% or 40% most deprived areas in Scotland.

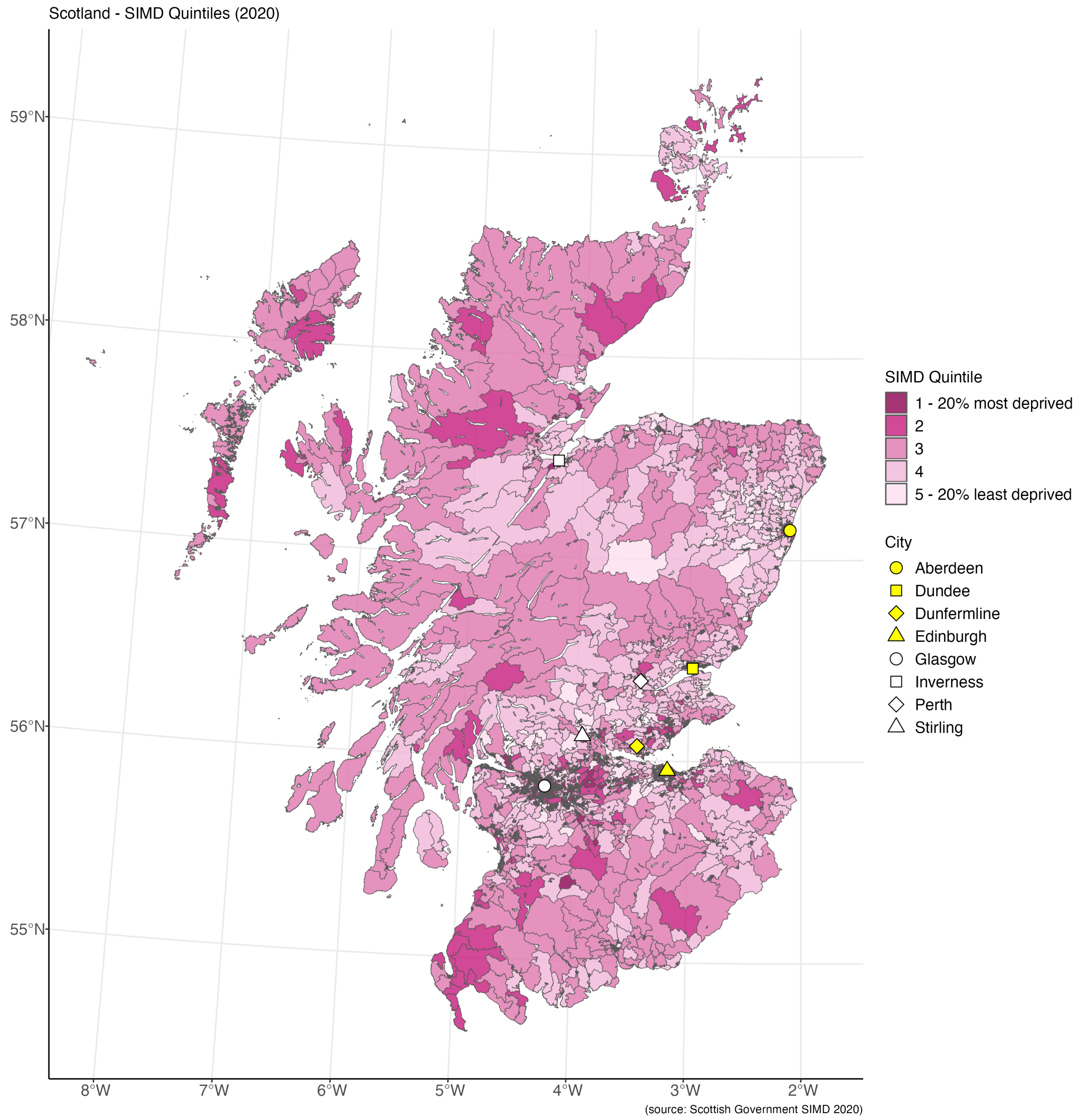


Figure 2.1: SIMD Quintiles (2020) across Scotland (not including the Shetland Islands). Scotland's 8 major cities are indicated with dots. Plot created using "shapefiles" provided by the Scottish Government [1].

SIMD receives irregular updates, with the most recent editions being in 2020 [59], and other historic editions used from 2004-2016 [61, 62]. For this reason, careful interpretation of literature that uses SIMD is advised, where the year the publication and SIMD updates were released should be noted. The Scottish

Government provides a postcode-linking tool for SIMD [59, 61, 62], making it an attractive measure for higher education institutions since it can be linked to applying/registered students consistently and reliably.

SIMD does have some drawbacks, however. It is not a measure of how deprived an individual is from a given area. An important statistic in understanding SIMD is that two out of three people on low-income do not live in a deprived area, and only one out of three people that do live in a deprived area are on low-income [58]. It should be used with caution in rural areas; living in a rural area can mean longer distances travelled to jobs, schools, and services like the post office or a GP. These factors play a large role in rural deprivation but are given smaller weightings in the calculation of SIMD rankings than the domains of employment or income [58]. In the technical documentation [58] it is stated that "[identifying people experiencing disadvantage] will not work as well in rural areas, and we need to look at other ways of assessing need and making decisions about allocating resources". SIMD has a bias towards large urban areas, particularly within Scotland's central belt where for example nearly one quarter of all SIMD Quintile 1 areas in Scotland can be found within Glasgow City Council [63] (Figure 2.2). Boliver and Gorard (2017-2022) [51–55] have frequently voiced their criticisms regarding area-level indicators, including SIMD, stating that they are not suitable for use as contextual indicators.

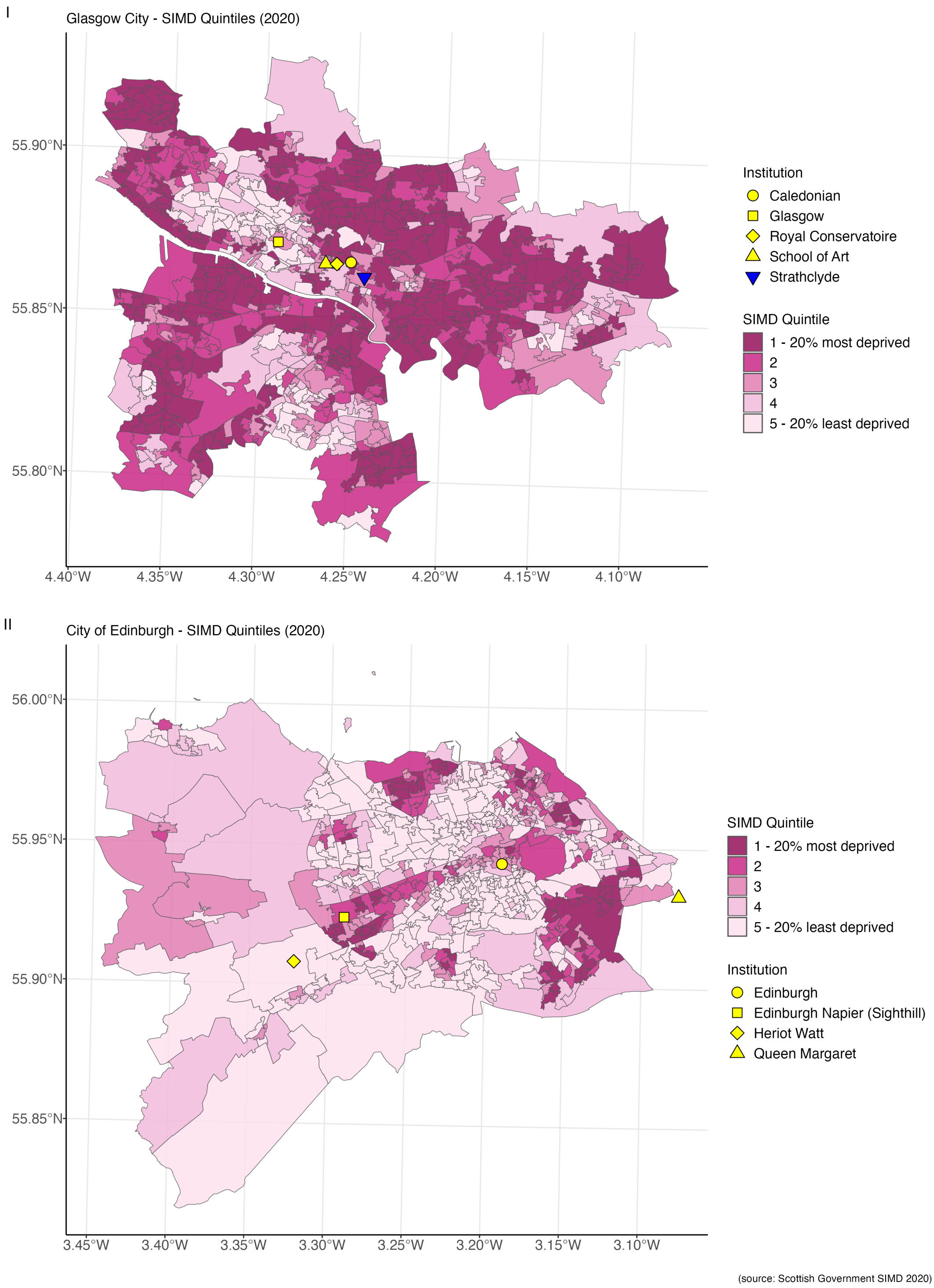


Figure 2.2: SIMD Quintiles (2020) within the "Glasgow City" and "City of Edinburgh" local authorities. Each city's higher education institutions are indicated with dots. Plot created using "shapefiles" provided by the Scottish Government [1].

2.3.2 Care-experience / Caring responsibilities

Care-experience and caring responsibilities are both considered contextual indicators by the University of Strathclyde [56]. Examples of what is considered care-experience include: adopted children who were previously looked-after, foster care, and residential care [64]. The Commission on Widening Access [3] recommended that all Scottish institutions should guarantee any qualified care-experienced student a place as well as provide non-refundable bursaries to support their studies. Students can disclose any caring responsibilities in their UCAS applications to institutions [65]. UCAS considers caring responsibilities to include "unpaid care to a family member, partner or friend who could not cope without their support..." due to "a long-term illness, disability, a mental health condition, or an addiction" [65]. The University of Strathclyde uses the following information to verify caring responsibility status: "an NHS carers card; a letter from a carers centre; a confirmation letter from a GP, medical professional, or teacher; or evidence of being in receipt of carers allowance. If your caring duties are mentioned in your teacher's reference, then no further evidence is required" [56].

Boliver et al. [52] considered care-experience to be a highly valid and reliable indicator of disadvantage. "School attainment levels and higher education progression rates are markedly poorer for young people who have spent time in registered care relative to other young people" [52]. Caring responsibilities was also considered by Boliver et al. [51] as a valid indicator of disadvantage. While there were no recommendations from the Commission on Widening Access [3] for institutions to adopt caring responsibilities as a contextual indicator, it is in wide-spread use across institutions (see Tables A.1 and A.2).

2.3.3 Low-progression/priority schools

Low-progression/priority schools are defined at the University of Strathclyde as Scottish secondary schools that had low proportions of students progressing to higher education and/or whether it was highlighted as a priority school by the Schools for Higher Education Programme [56, 66]. At the University of Strathclyde, the list of priority schools has changed over the years, with the latest update to Strathclyde's policy occurring for 2022/23 entry [66]. This translated to roughly 145 schools (out of 347) at the time the policy was implemented [66], which may not reflect the current number due to openings, closures, or mergers. While highly reliable, indicators based on the progression/attainment rates of students in schools have been criticised as being "of limited validity when it comes to measuring the socio-economic circumstances of specific individuals" [51].

2.3.4 The Indicator Debate

As a result of data access and protection issues, Scotland's Widening Access approach has predominantly focussed on contextual indicators that are aggregated on the area-level or school-level such as SIMD and the school attended. These indicators are much more reliable since they can easily be verified through postcode and local authority information that are already collected and disclosed to institutions. However, these indicators are not viewed as favourably by Boliver and Gorard [51–55]. Where initially in 2017, Boliver et al. deemed area- and school-level indicators acceptable if used in conjunction with other indicators [52], their stance hardened in 2022 towards replacing these indicators with the aforementioned more viable individual measures "...if the ambition is to equalise access to higher education within a generation is to be achieved" [51]. They also argued that these indicators could "do more harm than good" given that their use "... will inevitably result in widening access initiatives being poorly targeted and generate widening access statistics that are highly misleading" [51].

Other groups haven taken issue with the widespread use of aggregated indicators in British institutions. Universities Scotland acknowledged the ease of use and access that comes with SIMD [4], but asserted that "... it is still no substitute for the development of more sensitive ways of measuring individual disadvantage" and that "Scotland cannot afford to wait until 2030 to get it right for groups of under-represented students who do not fit into the [SIMD Quintile 1] category." [4]. Criticisms have been pointed towards SIMD's big-city bias [51, 53, 67] meaning that those in rural deprivation can often be looked over and that rural institutions can struggle to recruit when SIMD students are less likely to move away from home [68]. More recently, Robert Gordon University asked that other indicators be considered alongside SIMD in achieving Widening Access targets to account for rural factors [67]. In response to criticisms, the Commissioner for Fair Access recommended in 2024 that SIMD should continue to be used as the primary measure for progress towards Widening Access targets, but that institutions also collectively agree upon which other indicators should be used to "demonstrate their wider work promoting fair access" [7].

The current reality in Scotland is that SIMD and indicators like it, are applied on an individual basis to decide applicants' eligibility for contextual offers, which was not their intended operation. Despite their contentious nature, area and school-level indicators remain relevant in current and future Widening Access discussions, particularly the latter given their role in Widening Access targets

(Section 2.5.1). Meanwhile, the roles of more valid and reliable individual measures, such as eligibility for free-school meals, are limited so long as data access remains an issue for higher education institutions (see Appendix A.2.4).

2.4 Attainment Gaps

Inequality in the academic outcomes between groups of students (for example males and females) are referred to as "attainment gaps". Attainment gaps can manifest themselves at the very early stages of education. For example, in Aberdeen City council, a study which measured the assessment scores of pupils in Primary 1 found attainment gaps at the beginning of the school year between those who were younger, received clothing grants, or had English as a second language, compared to their peers who were older, did not receive clothing grants, and were native English speakers [69]. Additionally, they found that the gap between high and low attaining pupils had widened by the end of the school year, and that while girls had made more progress than boys in reading, they had made comparatively less progress in mathematics [69].

2.4.1 Attainment Gaps in Primary and Secondary Education

The Scottish Government report "Achievement of Curriculum for Excellence (CfE) Levels 2021/22" looked at the attainment levels of primary and early secondary school students from 2016/17 to 2021/22 [70]. It found that in 2021/22, the attainment gap between P1 to P7 pupils from the most and least deprived areas was

21.3 percentage-points for literacy levels and 17.8 percentage-points for numeracy levels [70]. While these were decreases compared to the year prior (which was affected by the coronavirus pandemic), the gap over time had barely changed since 2016/17. For S3 students in 2021/22, there was a gap of 16.3 percentage-points in literacy skills which had widened by 2.7 percentage-points since 2016/17 [70]. Meanwhile for their numeracy skills, there was a gap of 15.0 percentage-points which was relatively unchanged compared to the gap in 2015/16 [70].

These attainment gaps persist into the later stages of secondary education. A multi-generational cross-sectional study was published in 2009 that looked at the attainment of students at the late stages of secondary education (S4-S6) against characteristics such as sex, ethnicity, social class, family structure, parents' backgrounds, and other school-level and socio-economic factors, between the years 1985-2005 [71]. This occurred during major reforms to Scottish education at the time, including the introduction of the Standard Grade qualifications [72]. Amongst the Croxford's findings [71] were: that there was an upward trend in attainment and participation rates amongst 16 and 18 year olds; a widening gap between the attainment of females over males aged 16-18; that those from independent schools had consistently higher UCAS Tariff points than those who came from state schools; that factors related to parents' occupation and education had "additional effects on attainment". Of all the characteristic considered, it was found that "social class [was] the greatest source of inequality"; at age 18 those from managerial/professional backgrounds had "substantially" higher attainment than those from intermediate or working-class backgrounds, and this gap widened over time [71].

This was not the only record of attainment gaps present in the late stages of secondary school. Boliver [52] showed attainment gaps from 2007-2009 between secondary school students receiving/not receiving free-school meals (FSM) and those from more/less deprived areas (measured using SIMD – see Section 2.3.1). For example, 13.3% of students on FSM had left school with five or more Highers compared to 47.6% of those not on FSM, and 20.6% of students from SIMD Quintile 1 areas had left school with five or more Highers compared to 53.6% of those from SIMD Quintiles 3-5 [52]. More recently, the Scottish Government has conducted an analysis on the attainment of secondary school students from each SIMD Quintile in the years 2011-2016 [73]. It showed that in 2015/16, 33.6% of students from the 20% most deprived areas left school with at least one SQA Higher versus 77.9% of those from the 20% least deprived areas; an attainment gap of 44.3 percentage-points [73]. However, this did decrease to 38.5 percentage-points in 2015/16 [73].

2.4.2 Inequality and "Fairness" in Applications and Entry Requirements to Higher Education

When looking at the application and offer rates of English students to the prestigious Russell Group of universities between 1996 to 2006, Boliver [74] found that access to these universities were "far from fair". They demonstrated that even when controlling for secondary school attainment, students from lower social classes and state schools were less likely to apply to Russell Group universities than their peers from higher social classes and private schools [74]. Boliver [74] also found that once controlling for prior attainment at A-levels, students from lower social classes, state schools, and various ethnic minority backgrounds, were much less likely to be given an offer than their peers who were from a higher social class, private schools, or were white.

In 2018, the application to higher education rate for young people from the least deprived areas in Scotland was 49.9%, nearly three times the percentage of those from the most deprived areas at 16.9% [75]. However, the offer rates for students from more deprived areas was far higher than other quintiles, showing a commitment from Scottish institutions to admit "disadvantaged" but capable applicants [75]. The gap between the acceptance rates of students from the least and most deprived areas (SIMD Quintiles 1 and 5) also shrunk from 11.7 percentage-points in 2016 to just 3.5 percentage-points in 2018 [75].

To add to concerns, it is becoming increasingly competitive for Scottish domiciled students to enter higher education due to the growth in the number of applications outstretching the growth in funded places [29]. The Commission on Widening Access [3] claimed that institutions had raised entry requirements in response but to a level that had "risen well beyond what is required to succeed in degree level study". Boliver et al. [52] argued this was evident in the rising number of qualifications students were entering Scottish institutions with from 2006-2015 (from roughly AAABB at Higher in 2006 to "in excess" of AAAAA at Higher by 2015). Boliver et al. [52] also showed that within Scotland's most highly selective institutions, an 80%+ probability of progressing from first year to second year, and a 65%+ probability of achieving a "good pass" at honours (first-class or upper second-class) given completion of a degree programme, could be maintained while also lowering entry requirements by two grades (to ABBBB) for science programmes, and five grades (to BBBCC) for arts programmes. This was based upon the assumption that the 80%+ and 65%+ figures were "high bars" for degree-level success, though Boliver et al. [52] acknowledged that what constituted an acceptable level of success would be subjective.

2.4.3 Attainment Gaps in Degree-level Performance

In a discussion paper from the Commissioner for Fair Access [37], it was found that in the three academic sessions, 2013/14 to 2015/16, SIMD Quintile 1 students had lower rates of progression from first year to second year, completion of an honours degree, and attainment of a "good pass" (either first-class or upper second-class) at honours, than students from SIMD Quintiles 2-5. The size of the gaps between the groups were large: 5 percentage-points for progression, 9 percentage-points for completion, and 15 percentage-points for a good pass [37]. At the University of St. Andrews, Lasselle et al. [76] found that coming from the 40% most deprived areas in Scotland had a significant and negative association with achieving a First or Upper Second class degree. Elsewhere in the UK, Crawford [77] found similar differences between English pupils who had attended schools with the lowest and highest proportions of pupils eligible for free-school meals. The size of these gaps were 5.4 percentage-points for dropping-out, 11.0 percentage-points for completion, and 21.8 percentage-points for attaining a good pass at honours [77]. However, when Crawford [77] compared students with similar levels of prior attainment, those from more advantaged backgrounds became less likely to achieve positive outcomes such as completion or high degree classification, and more likely to drop out than their more disadvantaged peers.

Crawford [77] was not the only study to have found that students from disadvantaged backgrounds can perform as well as, if not better, than their non-disadvantaged peers. Lasselle et al. [76] similarly showed how "students with three A grades [at Higher] from below average [attainment] schools perform equally as well as those with four A grades [at Higher] from above average [attainment] schools". At the University of Bristol, Hoare et al. [78] found that "students who attended independent schools performed better in A-level examinations than those who attended state schools" although they "did not outperform students from state schools in their university degree programmes". This was true regardless of whether other factors were or were not accounted for, such as A-level attainment prior to university [78]. More recently, Cameron et al. [79] found that at Abertay University, "those from [Widening Participation] backgrounds were equally as likely to gain a good degree as their non-[Widening Participation] counterparts and to be in graduate and/or sports employment."

To add to the evidence, a study on students' degree performances at the University of Edinburgh between 2004-2006 found that most Widening Participation-indicated students were just as likely to complete their Honours degree compared to their non-indicated peers [69]. In contrast to the findings from Crawford [77] however, Croxford et al. [69] found that '[Widening Participation] students were less likely to achieve a [good degree classification] even after taking account of prior qualifications." Croxford et al. [69] expanded that while "prior qualifications were the main factor determining degree outcomes", the differences in the degree classifications observed between the groups were "only partly explained by prior qualifications" and that this may reflect the fact that disadvantaged students at higher education suffer the same hardships as those in secondary education.

Notably, Croxford et al. [69] found that once accounting for prior attainment, students from independent schools did not perform as well as their peers from state schools, in accordance with previous findings [76–78].

2.4.4 Differences in Graduate Outcomes and Life Chances

In 2023, Universities Scotland [80] claimed that 73.8% of Scottish graduates believed that their university experience helped them find the type of job they wanted and that 69.4% believed that going to university helped them to do so faster. A study conducted at a post-1992 UK university [81], found that those from "working-class" backgrounds[1] were very aware that a "good" degree could improve their social mobility, and yet 64.8% of the surveyed students questioned their pursuit of higher education and whether it would help them to achieve the social mobility they desired. Audit Scotland also found that of the 90% of graduates from Scottish higher education institutions who found employment or further study in 2013/14, 59% of these graduates stated that their degree was necessary or advantageous in securing the job [29].

Whilst there is evidence that a degree in higher education can lead to better medium-to-long-term earnings for most UK students [82], there also exists gaps between the graduate outcomes of different socio-economic groups. However, these differences vary depending on whether the student was domiciled in England, Scotland, Wales or Northern Ireland. For example, Macmillan and Vignoles [83] found significant socio-economic gaps (measured using indicators on

[1]Measured as those who were first-generation to attend higher education and those from POLAR Quintile 1 areas. These are areas which have low participation in higher education rates amongst young people. See Appendix A.2.1 for full definition of POLAR.

attendance of state/private school , POLAR (see A.2.1 for definition), and highest earning parent's most recent occupation) amongst English and Welsh-domiciled students in terms of securing employment 6 months and 3 years post-graduation, though could not find similarly significant gaps amongst the Scottish-domiciled students. They posited that perhaps there are some inherent differences in the Scottish education system and/or labour market that require further investigation [83]. Vignoles and Crawford [84] also found evidence of gaps between students in terms of their post-graduation earnings at a single UK institution depending on the social-class of their parents or whether they attended a state or private school.

Other empirical data at English higher education institutions showed that amongst UK domiciled students (including Scottish) those who came from POLAR Quintile 1 areas (see A.2.1 for definition) were less likely to be in professional employment 6 and 40 months after graduation than those from Quintile 5 [85]. Belfield et al. [86] also found gaps in post-graduation earnings between English-domiciled students from different social-classification groups (based on free-school meals and postcode data).

2.4.5 The Case for Contextualisation

Given the decades-worth of empirical evidence on education inequalities, at every stage of education, it is understandable why there is pressure to ensure a fairer education system for all. This is particularly the case for higher education, since degrees can improve graduates' opportunities and life earnings in the labour market [29, 80–82].

Efforts to address these gaps in higher education come under the umbrella of Widening Access, the justifications of which are often rooted in both moral and economic arguments. Scotland's Commissioner for Fair Access, whose job is to lead the national strategy on Widening Access, has tied its progress to efforts that address child poverty and gaps in the labour market [7]. The Commissioner argued that "Making lower offers to applicants from deprived backgrounds is not 'dumbing down' entry standards. Not all applicants have the same advantages, in terms of family support or school experience. Making the same offer to everyone is not only unfair; it fails to identify students with the greatest potential. Universities need to make much bolder use of contextual admissions" [6]. Elsewhere, government-commissioned reports agree that there is an economic imperative in recruiting more students to higher education from deprived or under-represented backgrounds [3, 87, 88]. For example, the Commission on Widening Access stated that "Scotland is missing out on the economic potential of some of our finest talents." [3].

2.5 Scotland's Widening Access Agenda

In 2014, the Scottish Government formed the Commission on Widening Access, whose remit was to identify where and how Scottish education could improve to be more inclusive and fairer to all people of Scotland, no matter their background. Their aim was to determine how Scotland could achieve the First Minister's pledge that "...a child born today in one of our most deprived communities will, by the time he or she leaves school, have the same chance of going to university as a child born in one of our least deprived communities" [89]. The Commission's philosophy was that the institutional inequality present in the Scottish education system was "unfair, damaging and unsustainable" and that Scotland had a "moral, social, and economic duty to tackle this inequality" [90]. Their interim report [90] reviewed all evidence on access and outcomes amongst different backgrounds, which was followed by a final report, "The Blueprint for Fairness" [3], which gave 34 recommendations to the Scottish Government, higher education institutions, and other stakeholders within the education system. This report informs much of the discussion and direction of Widening Access efforts within Scotland and is considered a pivotal moment in the attitudes of Government and institutions towards Widening Access.

2.5.1 Relevant Recommendations and Targets

The Commission [3, 90] interpreted the pledge by the First Minister into a headline target for 2030 which would need to be achieved alongside several interim targets by higher education institutions. These were that:

- By 2021, 16% of entrants to higher education in Scotland should come from the 20% most deprived backgrounds [3] (achieved in 2019/20 [91, p. 10]).
- By 2026, 18% of entrants to higher education in Scotland should come from the 20% most deprived backgrounds [3].
- 2030, 20% of entrants to higher education in Scotland should come from the 20% most deprived backgrounds [3].

The targets were all defined according to the Scottish Index of Multiple Deprivation (SIMD) [3] (see Section 2.3.1 for definition of SIMD), such that the "most deprived backgrounds" were areas that fell under the categorisation of SIMD Quintile 1 (also known as "SIMD20"). Additionally, the Commission [3] recommended a target for each individual higher education institution to satisfy:

- By 2021, 10% of entrants to each individual higher education institution should come from the 20% most deprived backgrounds [3].

Though this target was scrapped in 2024 [7] (see Section 2.5.3 for more details.) Other recommendations from the Commission [3] included the installation of a Commissioner For Fair Access (Sir Peter Scott 2017-2024; John McKendrick 2024-present), whose responsibilities were to lead the national conversation, publish annual reports and research, and be an advocate for disadvantaged learners and Widening Access efforts in Scotland. The development of a Unique Leaner Number (ULN) was also recommended [3], with the idea that it could track all students (not just Widening Access beneficiaries) through all stages of Scottish education, including if and when they changed degree programme or institution. This recommendation was also supported by Universities Scotland [4]. The Commission [3] also recommended that all care experienced students be given bursary support from the Scottish Government and that they be entitled to a place at any Scottish higher education institution should they meet the standard entry requirements (SERs) of the degree programme they wish to apply for. They did not give any recommendations for those with caring responsibilities [3].

Finally, to aid in achieving the interim targets, a key recommendation from the Commission [3] was the uptake of "contextualised admissions" – the practice of considering a student's contextual background alongside their prior attainment in admissions decisions – across all Scottish higher education institutions. Students who were eligible for contextual offers would then be assessed against "Minimum Entry Requirements" (MERs) rather than the Standard Entry Requirements (SERs). The Commission [3] did not stipulate how much institutions should lower SERs by, though in practice this became one or two grades. The Commission [3] recommended that MERs should be implemented across all degree programmes in Scotland and publicly viewable by potential applicants by 2019. In addition to anyone from an SIMD Quintile 1 area within Scotland being eligible for a contextual offer, the Commission [3] stated that all care-experienced students should be guaranteed a place of study should they satisfy the MERs. Outwith these criteria, the Commission [3] did not explicitly state which students should be eligible, leaving this to the discretion of each individual institution. As a result, each institution has made use of a range of "contextual indicators" used to identify individuals who, in their opinion, were disadvantaged (see Section 2.3).

2.5.2 Response from Institutions and Stakeholders

In 2015/16, prior to the final publication from the Commission on Widening Access [3], a study was conducted [92] that interviewed all 18 higher education institutions across Scotland[1] to determine their attitudes towards progressive admissions policies. They found that at the most competitive institutions, there was a view that contextualised admissions were at times unfair towards highly

[1]Not including the Open University.

qualified applicants from advantaged backgrounds [92]. Acknowledgment of how structural inequality led to differences in prior attainment was only acknowledged where a student had achieved high grades despite their disadvantaged background [92]. The prevailing attitude at the most competitive universities was that the goal of admissions processes was to identify the most academically capable, or "brightest and best", as measured by prior attainment [92]. Boliver et al. [92] concluded in stating that these attitudes, while not unique to Scotland, formed barriers to the nation's Widening Access ambitions. More recently, Boliver et al. [53] noted that ambitious use of contextualised admissions would be necessary if the Commission on Widening Access's [3] targets were to be achieved.

Despite these institutional attitudes, the Commission's [3] recommendations for educational reform were accepted, in full, by both the Scottish Government [5] and by Universities Scotland [93]. Universities Scotland [93] responded positively to the recommendations, supporting the notion that action should be taken "...at all levels of education and by all relevant partners and stakeholders". They agreed to implement contextualised admissions and to the timeline for disclosing minimum entry requirements across all degree programmes by 2019 [93]. However, they expressed concerns over the reliance on SIMD as the sole measure of deprivation, acknowledging its advantages and weaknesses, "Scotland cannot afford to wait until 2030 to get it right for groups of under-represented students who do not fit into the SIMD20 category" [93]. They were not alone in their criticism of SIMD as the chosen measure for Widening Access targets (see Section 2.3.4 for more details). The endorsements from institutions and government demonstrates that there is political will within Scotland to achieve the targets set by the Commission on Widening Access [3].

While consensus has developed around the need for the contextualised admissions in Scotland and the UK, some take umbrage with the notion that it can address what they see as structural inequalities in wider society, warning that instead that contextualised admissions could preserve them. Mountford-Zimdars and Moore [94] encouraged the critical engagement with the accepted theory of contextualised admissions, pointing out that the evidence surrounding what constitutes the "potential" to succeed at higher education is "relatively weak" and "requires re-evaluation in the light of continuous changes in young people's qualifications." They impart that the attractiveness of contextualised admissions may be in its ability to "make small progress while not upsetting the existing system" [94]. Meanwhile, Boliver [95] takes issue with the system of "meritocracy" that prevails within the United Kingdom – that institutions should seek to admit the "brightest and best" as measured by prior academic achievement. They argue that such a system cannot deliver the social mobility desired without addressing the societal inequalities that exist outside of education. "In a more equal society, not only would it be easier for those from relatively disadvantaged backgrounds to get to university, ... it would also matter much less... whether they went to university or not." [95].

2.5.3 Progress and the Shifting of the Widening Access Agenda

In 2015/16, the proportion of SIMD Quintile 1 students in Scottish higher education institutions stood at 14% [6]. In 2021/22, the interim target of 16% across Scotland was achieved [68, 96], though in 2023/24, the proportion stood at only 16.5% [7]. The 2021 target of 10% for each individual institution was not reached by five universities – Aberdeen, Robert Gordon, Edinburgh, Highlands and Islands, and St Andrews [68]. In contrast, five other universities had reached the 2030 target of 20% ten years early, all located within the Greater Glasgow metropolitan area – Glasgow Caledonian, Glasgow School of Art, Royal Conservatoire of Scotland, Strathclyde, and West of Scotland [68, 96].

The response from some universities was to criticise using SIMD as the primary measure of disadvantage (Section 2.3.4). The outgoing Commissioner for Fair Access (Sir Peter Scott) warned in 2022 that some institutions had a "free pass" when it came to reaching the 2030 target given their geography and history of recruitment, while others struggled despite their best efforts [68]. The Commissioner [68] recommended re-defining the institutional target to account for other measures of disadvantage, though remained firm that the goal for 2030 should be framed in terms of SIMD remained unchanged in order to preserve a cohesive national strategy. This recommendation was echoed by his successor, Commissioner John McKendrick, in 2024 [7]. The new Commissioner also recommended that institutions should be expected to commit to raising the proportion of SIMD Quintile 1 students to 20%, or if 20% was evidently not possible, raising the proportion to at least match the institution's previous high [7].

Perhaps the most consequential recommendation from the new Commissioner for Fair Access was that "equal weight" be given to not only students entering higher education, but their academic outcomes and the "student experience" [7]. The Commissioner stated their belief that fair access was not only about gaining access into university but allowing students to "thrive" at university and to help them "achieve positive outcomes after graduation" [7]. They argued that the retention rates of target groups had seen little progress since 2016 and committed to investigating these in 2024 to determine why this was the case [7].

This constitutes a significant shift from equality of opportunity to also considering equality of outcome. It is true that while most of the focus in Scotland's Widening Access Agenda since 2016 has been to increase the proportion of those from SIMD Quintile 1 to higher education, less emphasis has been put on how these students perform once admitted to higher education. For example, in the University of Strathclyde's strategy for 2030 [8], a target of 90-95% average retention from year 1 to year 2 was proposed for all undergraduates, with no retention target given specifically for Widening Access students. Indeed, there has been concern over a lack of support structures at institutions for struggling students with lesser attainment due to their disadvantaged backgrounds [97]. Not adequately measuring the degree-level performance of Widening Access students has been a blind spot for Widening Access efforts in Scotland.

2.6 Summary

The conclusions that can be drawn on educational inequality from the empirical evidence in the UK are as follows. For decades, there have been stubbornly high attainment gaps from primary to the late secondary stages of education that divides the sexes, social classes, schools, and the deprived from everyone else [52, 70, 71, 73]. There is "unfairness" baked into application and acceptance rates of students from different socio-economic backgrounds [74]. Entry requirements have inflated beyond what is necessary to achieve success at degree-level [52].

There is evidence of gaps in degree-level outcomes between those from the most and least advantaged backgrounds [37, 69]. Although there is some evidence to suggest that success at higher education cannot be predicted based on prior attainment alone and that contextual information may also necessary to better understand a student's "potential" to succeed [69, 76–79]. All of this leads to the conclusion, that the education system at every stage is not fair nor equal.

Since 2016, the Scottish Government and Scottish higher education institutions have responded to these inequalities by attempting to increase the proportion of students from disadvantaged backgrounds in higher education. In particular, they have set a goal that by 2030, those from the 20% most deprived areas in Scotland should be equally represented. However, less focus has been put on the academic outcomes of these students once they are admitted. This thesis therefore aims to contribute to the research and literature on the outcomes of Widening Access students in higher education. In Chapter 7, the topic of access to Advanced Highers and their effect on students' academic outcomes in higher education will be explored. Chapter 8, will explore the academic outcomes of students who received/did not receive a contextual offer to the University. Finally, Chapter 9 will examine the differences in dropout rates across socio-economic groups. Each of these chapters will also examine other factors that may affect students' chances of a successful outcome, such as sex, ethnicity, and prior attainment from secondary education. They will also identify the most appropriate statistical methods for measuring the academic outcomes of students more generally. Prior to fitting any models however, data on the academic outcomes of students is required. The next two chapters (3 and 4) will detail how these data were gathered, joined and cleaned. They will also define the population of interest: Scottish school-leavers at the University of Strathclyde.

Chapter 3

Deriving the Relevant Datasets

The target population of interest for this thesis is Scottish “school-leavers”, who are studying their first full-time undergraduate degree at the University of Strathclyde. “School-leavers” shall be defined as students who “enter” (are admitted/join) the University of Strathclyde in the year directly following their completion of secondary education. Note that the Widening Access targets relate to school-leavers, college entrants and mature students. The school-leaver population was chosen specifically because they are the largest sub-group of students, they directly relate to Scotland’s Widening Access target for 2030 and because data on these school-leavers were the most readily-available. Data held by the Strategy and Planning team on college and mature students were a work-in-progress and thus not available for analyses. Furthermore, the relationships which affect students’ trajectories through higher education may be distinct for each of these sub-groups of students. Thus, by focussing only on school-leavers potential sources of variation could be reduced.

This chapter will outline how the primary dataset of interest, the **School-leavers dataset**, was derived from linkage of the University's registration records to many other datasets. In this chapter and throughout the thesis, the names of key datasets/subsets are highlighted in **bold**, while variables are highlighted in *italics*, where necessary to achieve clarity (e.g. **School-leavers dataset** and *Academic Session*). Note that it was necessary to reference some variables in Section 3.2 on data linkage, prior to their formal definition in Chapter 4 (e.g. *Academic Cohort* and *Postcode*). Appropriate signposting to variable definitions will be given where needed.

3.1 The Nomenclature of Stage versus Year

The length of time a student has spent registered at the University does not necessarily match the "year" of the degree programme they are registered with. For example, if student A is repeating "1st year" of "BSc Mathematics", then they will be in the their "2nd year" of registration. Similarly, if student B enters directly into "2nd year" of "BSc Mathematics", then they will only be in their "1st year" of registration. To avoid confusion, the "stage of a degree programme" will be the preferred terminology, while "year" will be used to refer to the amount of time that has passed since a student's first registration session. Going back to the examples, student A would be in stage 1 of "BSc Mathematics" but in their 2nd year, while student B would be in stage 2 of "BSc Mathematics" but in their 1st year.

3.2 The Core Student Record (CSR)

The data used in this thesis were primarily derived from the University of Strathclyde's Core Student Record (CSR) provided by the Strategy & Policy team (the data owners). A Data Protection Impact Assessment (DPIA) was required to be completed for the Stratgey & Policy team to approve access to the data (see attached documents in Appendix B). The CSR is a database which tracks a student's entire journey through the University, from entry to graduation. For example, progression to the next academic year, repeating a year, voluntary suspension, exiting, or graduation. The CSR also provides details on demographics, such as sex, ethnicity, disability status, and on attainment achieved prior to entry to the university. The version of the CSR provided was the subset of "funded"[1] students who entered the University between academic sessions 2012/13 and 2021/22 (ten years' worth of registration records). A flowchart of the data linkage process for the CSR data tables and other datasets is highlighted in Figures 3.1 to 3.7.

The CSR is made up of five core data tables: (i) the **Retention, Progression and Outcome** table, (ii) the **Registration and Applicant Codes** table, (iii) the **Demographics** table, (iv) the **Date of Births** table, and (v) the **Attainment on Entry** table. Each table contained information that was either collected by the University while the student was registered (degree programme, academic outcome, registration status etc.), or from information that was disclosed to the University by the student in their UCAS application (demographics, attainment, date of birth etc.). Each of these tables are linked to one another via students'

[1]Students whose tuition was paid for by the government.

Registration Number and/or UCAS *Applicant Number*. These numbers are linked together via a unique identifier variable, *HPLE Person Code*, which is contained within the **Registration and Applicant Codes** table.

Tables (i-iv) were cleaned and linked to one another via each student's unique *HPLE Person Code* or *Registration Number* or *Applicant Number* (see Figure 3.1). The resultant dataset was then filtered to only consider "**H**ome" [1], **U**ndergraduate students studying their **F**irst-degree at the university **F**ull-time, known as the **HUFF dataset** for short (see Figure 3.1). This dataset contains 91,381 observations which corresponded to a unique "instance" of registration for a single student. For example, if a student was registered for four academic sessions, then the **HUFF dataset** would have four observations, or instances, for that student. There are 224 students who had multiple "first registration" records in the CSR. For each student, the instance with the earliest date was taken as the their first registration session and all subsequent sessions were deemed as continuations.

[1] In Scotland these are Scottish-domiciled students. This definition is distinct from "Home" students at institutions in England, Wales, and Northern Ireland, which refers to anyone domiciled in the UK.

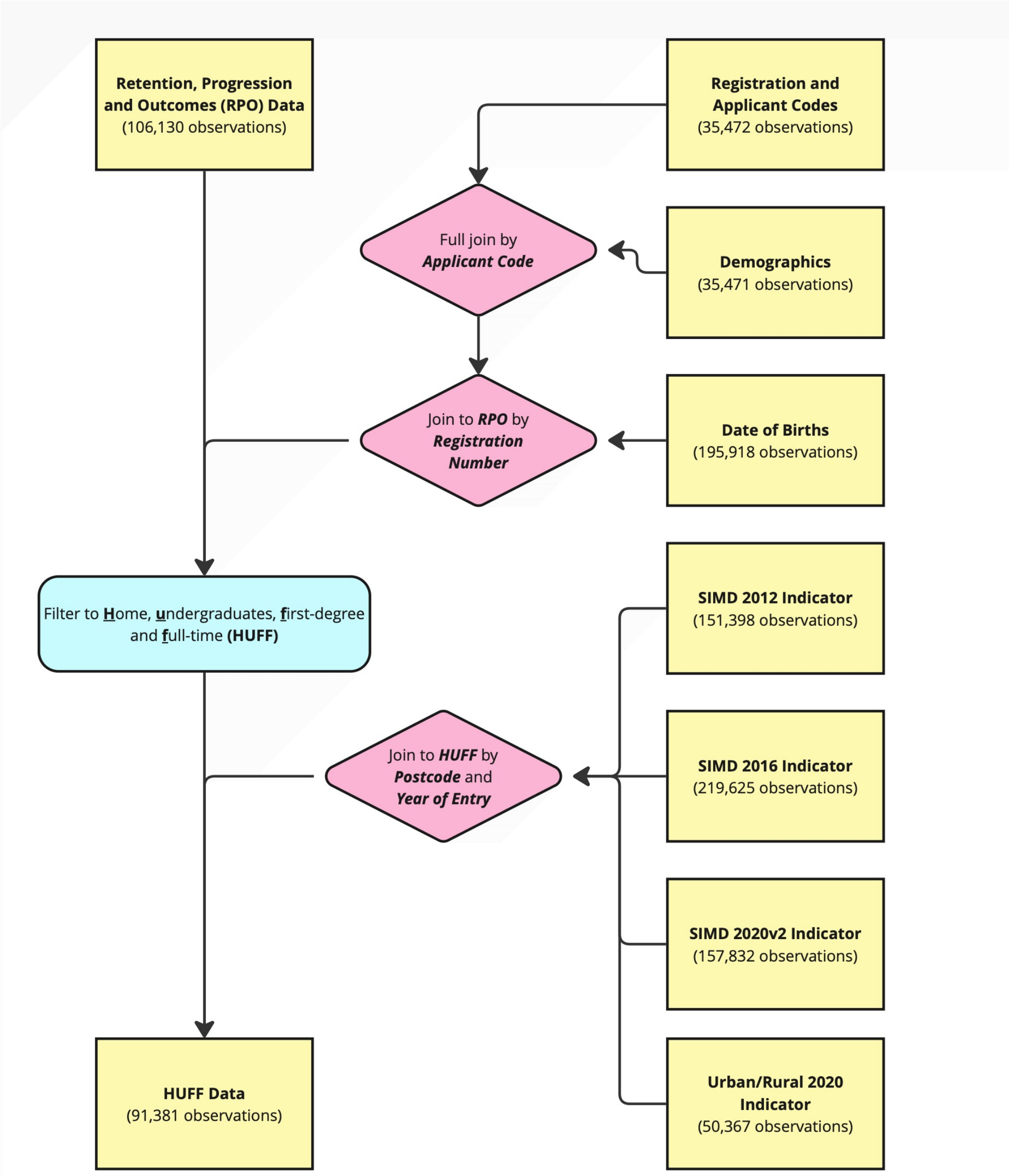


Figure 3.1: Cleaning and linking the Core Student Record Data.

3.3 Linkage to the Scottish Government Datasets

The **HUFF dataset** was then linked to the relevant "SIMD Postcode Look-up Tool" provided by the Scottish Government [59, 61, 62] via the *Academic Cohort* and *Postcode* variables (for definitions, see Sections 4.1 and 4.3, respectively). For example, the **SIMD 2012 Indicator** look-up tool was used for students from *Academic Cohorts* 2012/13 - 2016/17, while the **SIMD 2016 Indicator** tool was used for those from *Academic Cohorts* 2017/18 - 2020/21, and the **SIMD 2020v2 Indicator** tool was used for the 2021/22 *Academic Cohort* (Figure 3.1). It was assumed that these were the versions of SIMD that each entrant would have been assessed against while their UCAS application was being processed. There were errors when linking the **HUFF dataset** to the Scottish Government's 3-Fold **Urban/Rural 2020 Indicator** [98] via each student's *Postcode* variable. This was due to some postcodes being retired from use and because some postcodes lay on the boundary of multiple datazones. This required the data to instead be linked via the *Datazone 2001* and *Datazone 2011* variables, which were derived from the linkage of the SIMD datasets to the **HUFF dataset**. The "3-fold" version of the **Urban/Rural 2020 Indicator** (used three classification groups: "Urban", "Accessible Rural" and "Remote Rural") was connected to all observations in the data regardless of *Academic Cohort*, since this indicator was the most up-to-date at the time. Despite this, 25% of all students in the final dataset had missing data in the Urban/Rural classification. The large amount of missing data presents a problem for robust interpretations of this variable within models. It is recommended that future updates try to address these issues prior to use in model fits.

3.4 The Attainment and Prospectus Data

The **Attainment on Entry** data table contained the entire attainment history of students that was disclosed to the University. This included all attainment awarded prior to the student's first registration (disclosed to the University via UCAS or other application route) and the classifications/qualifications awarded while the student was registered at the University (e.g. "Bachelor's with Honours First Class"). Specific university-level attainment, such as module marks and credits awarded, were not included in the **Attainment on Entry** data table. Each observation within the **Attainment on Entry** table corresponded to a unique attainment record, in a given year, for a single student. For example, if a student had attained five Highers in 2016 and three Highers in 2017, then the **Attainment on Entry** table would have eight observations for that student. This table was filtered to only consider attainment that was awarded prior to the student first registering at the University (Figure 3.2). To more easily interpret each students' attainment record from secondary education, only their grades at Higher and Advanced Higher were considered. These were then grouped into a "Grade String" (e.g. AAABB at Higher, BC at Advanced Higher, ...etc.), to create the **Grouped Attainment at Entry** data table. Each unique student in this table had at most two associated observations: one for their attainment at Higher, and one for their attainment in Advanced Higher. As a precaution, anyone who had attainment prior to entry that was of a Scottish Credit and Qualifications Framework Level greater than Advanced Higher (Level 7) [10] was removed from the dataset. This was to ensure the final output would have school-leavers and not college entrants. A full list of all attainment codes removed that resulted in a student's removal is given in Appendix C Table C.5.

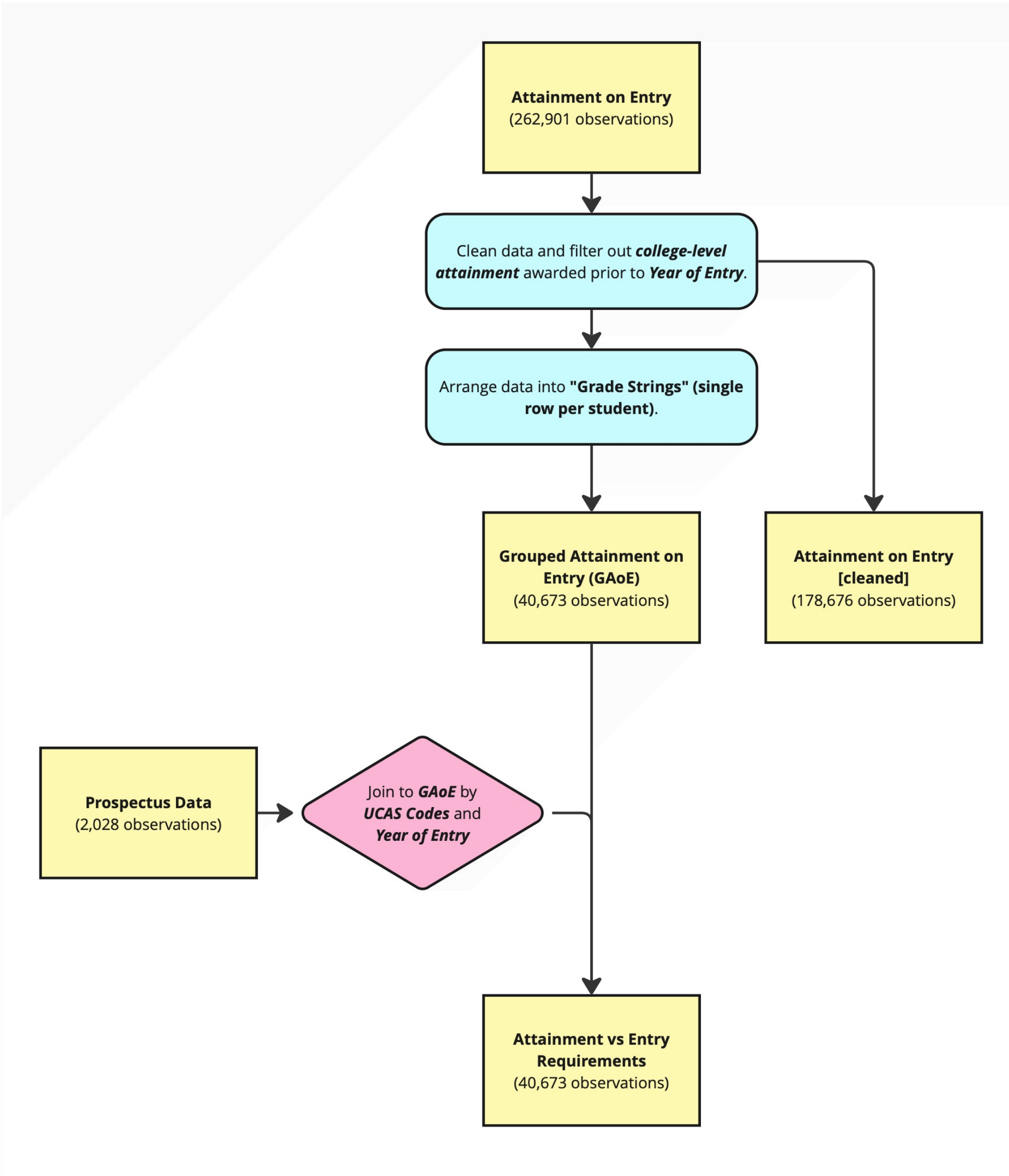


Figure 3.2: Cleaning and linking the Attainment and Prospectus Data.

The University of Strathclyde Prospectuses from 2013/14 - 2023/24 were made available for study. The **Prospectus dataset** developed by the Widening Access Team (University of Strathclyde) was created through manual recording on an excel spreadsheet of the entry requirements shown on each prospectus release from 2015/16 - 2023/24. Budget constraints meant that the entry requirements for academic sessions 2013/14 and 2014/15 could not be manually recorded. While the **Prospectus dataset** did not contain the specific entry requirements for every degree programme, it was sufficient for most programmes. Degree programmes not included in the **Prospectus dataset** included many of the joint honours degrees offered by the Faculty of Humanities and Social Sciences, since the magnitude of subject combinations was too great to capture. A summary of the **Prospectus dataset** and its variable descriptions can be seen in Appendix C Table C.1. The **Prospectus dataset** was linked to the **Grouped Attainment on Entry** data table via each student's *Academic Cohort* and each degree programme's unique *UCAS Code* (Figure 3.2). The resultant dataset was named the **Attainment vs Entry Requirements** dataset.

3.5 Deriving the School-Leavers Dataset

There were 22,936 total students in the CSR from 2012/13 to 2021/22, or around 2,200 per academic session. The population of interest is students who come directly to the University from secondary education, or "school-leavers". School-leavers could not be directly identified in the CSR, so were assumed to be anyone aged 18 or under at the point of their first registration (Figure 3.4). Trivially, this excluded anyone from the same stage of secondary education but aged 19, or those who had taken a gap year before coming to the University. However, it was not possible to distinguish these students in the data from mature students or college entrants, so unfortunately 1,976 of these students had to be removed (around 198 per year).

The HUFF dataset was filtered to only consider school-leavers and then linked to the **Attainment vs Prospectus** dataset via each student's unique *HPLE Person Code* (Figure 3.4). 258 observations which had missing or unknown *Ethnicity* data were removed. The inclusion or removal of these students had no effect on the model estimates derived in Chapters 7, 8 and 9. The resultant subset was referred to as the **School-leavers person-period dataset** and contained 18,988 unique students (Figure 3.4).

3.6 Person-Period versus Person-Level Format

There are two formats that describe how data are arranged. The first is a "person-level" format, which organises each row of a dataset such that they correspond to a unique individual. The alternative is a "person-period" format, which organises each row of a dataset such that they correspond to a unique individual at a unique time-period. A visual example of these two data formats is provided in Figure 3.3. The **HUFF dataset** is in a person-period format, where each observation represents an instance of registration. For most of the statistical models used in this thesis, a person-level format was required. Therefore, the **School-leavers dataset** was converted to a person-level format where required. For brevity, both formats of the data may be referred to simultaneously as the **School-leavers dataset**. Clarifications will be given on the specific format where it is necessary.

Person-Level Format

Student	**Academic Cohort**	**Max Reg. Year Count**	**Drop-out Status**
A	2012/13	4	0
B	2020/21	1	1
C	2014/15	5	0
D	2014/15	5	1
E	2016/17	6	0
F	2012/13	7	1

Person-Period Format

Student	**Academic Session**	**Reg. Year Count**	**Drop-out Status**
A	2012/13	1	0
A	2013/14	2	0
A	2014/15	3	0
A	2015/16	4	0
B	2021/22	1	1
C	2014/15	1	0
C	2015/16	2	0
C	2016/17	3	0
C	2017/18	4	0
C	2018/19	5	0
D	2014/15	1	0
D	2015/16	2	0
D	2016/17	3	0
D	2017/18	4	0
D	2018/19	5	1
E	2016/17	1	0
E	2017/18	2	0
E	2018/19	3	0
E	2019/20	4	0
E	2020/21	5	0
E	2021/22	6	0
F	2012/13	1	0
F	2013/14	2	0
F	2014/15	3	0
F	2015/16	4	0
F	2016/17	5	0
F	2017/18	6	0
F	2018/19	7	1

Figure 3.3: Comparison of person-level format (left) to person-period format (right) for school-leavers dataset.

3.7 Direct Entry Students

"Direct-Entry" students are those who skip stage 1 of a degree programme and enter directly into a higher stage of the degree programme. This practice is most often seen with college entrants or students who have Advanced Highers. A direct-entry student's particular pathway through higher education will therefore be different to the traditional Scottish school-leaver who enters into stage 1. The presence of these students could bias the interpretation of any effects on academic outcomes, since direct-entry students will have spent a reduced amount of time to achieve similar outcomes compared to their peers. Thus, $1,095$ direct-entry students were removed from the **School-leavers dataset**.

Of the $1,095$ direct-entry students removed, 782 of these were students registered with the "MPharm Pharmacy" programme. This programme is, unusually, a four-year Integrated Masters programme where most students are expected to enter stage 2. While there were 294 "MPharm Pharmacy" who entered into stage 1, it was decided to remove all $1,076$ "MPharm Pharmacy" students given that this degree programme was delivered differently to other undergraduate degree programmes at the University. For similar reasons, any student registered on the "MPhys Physics with Advanced Research" was also removed. Future analyses could examine whether these programmes had similar relationships between covariates and academic outcomes compared to the rest of the University's school-leaver population.

3.8 Widening Access Degree Programmes

The University runs two Widening Access programmes: the "Engineering Academy" and "Natural Sciences". Both of these are bespoke programmes are aimed at students from deprived backgrounds with lower prior attainment. Students may transfer onto a traditional Bachelor's with Honours degree after one or two years dependent upon academic performance. It was decided to remove the 782 students registered to these programmes from the **School-leavers dataset** since their particular pathways through higher education (and potentially their academic outcomes) are different to Scottish school-leaver on traditional degree programmes. Future analyses of these students' academic outcomes would be of great interest to Widening Access literature.

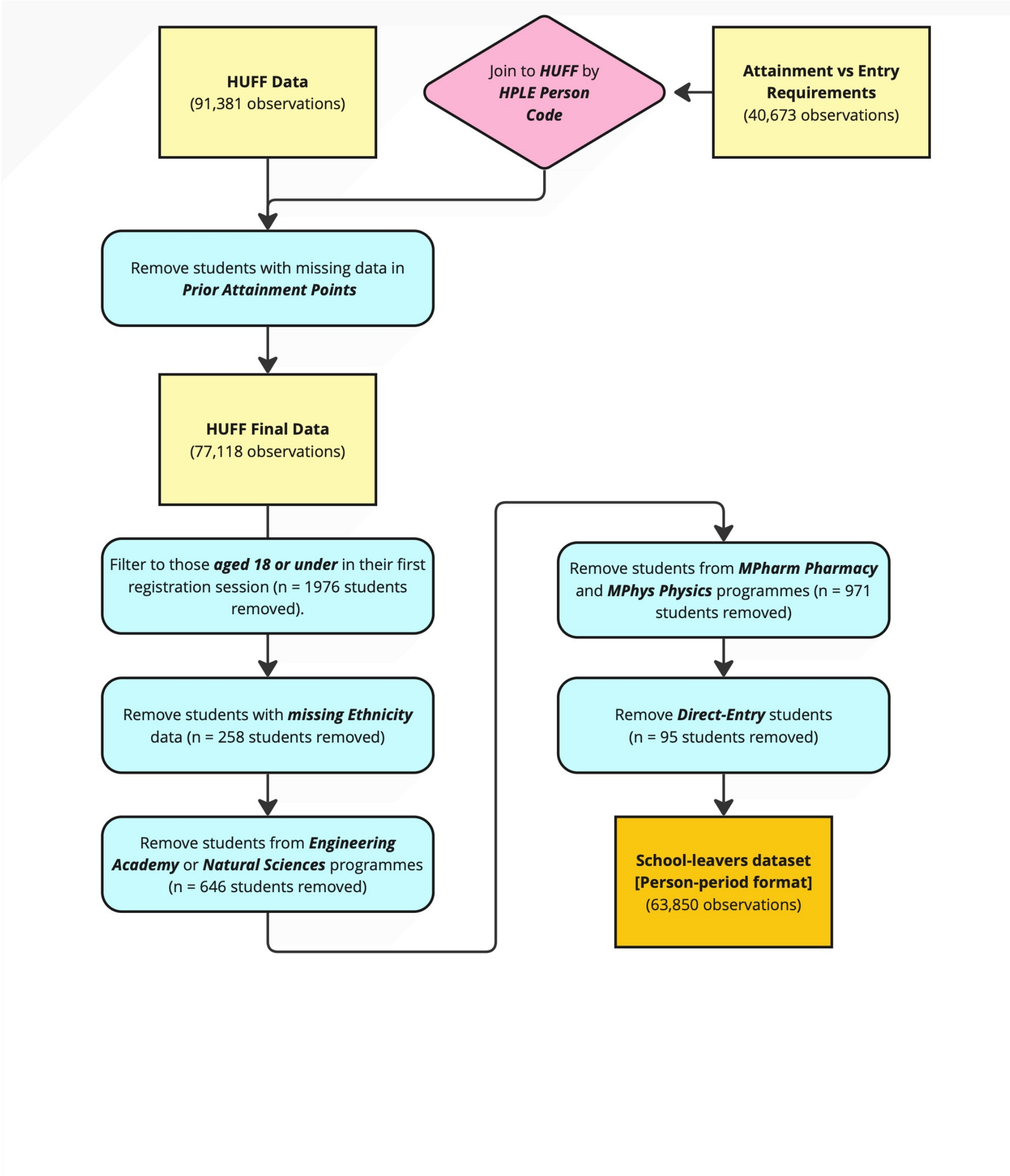


Figure 3.4: Deriving the School-Leavers Dataset.

3.9 Deriving the Analyses Subsets

The **School-leavers dataset** is the starting point for all analyses in this thesis and was also used to generate the descriptive statistics shown in Chapter 5. However, prior to the specific analyses in Chapters 7, 8 and 9 it was necessary to filter the **School-leavers dataset** to particular subsets to answer the relevant research question.

In Chapter 7, the **Advanced Higher subset** (Figure 3.5) only included school-leavers from the Faculties of Science and Engineering and from *Academic Cohorts* 2012/13 - 2018/19. It also removed all students who had missing data in the *Best Mathematics Qualification* variable (see Section 4.2 for definition). The **Advanced Higher subset** was split further (for reasons explained in Section 7.3) into the **Maths subset** which only looked at school-leavers from the Department of Mathematics and Statistics and the **SciEng subset** which looked at school-leavers from the remaining departments in the Faculties of Science and Engineering (i.e. not including those from the Department of Mathematics and Statistics).

The **Contextual Offer subset** (Figure 3.6) used in Chapter 8 only considered school-leavers from *Academic Cohorts* 2015/16 - 2018/19 and removed all students who had missing data in the *Offer Received* variable (see Section 4.5 for definition).

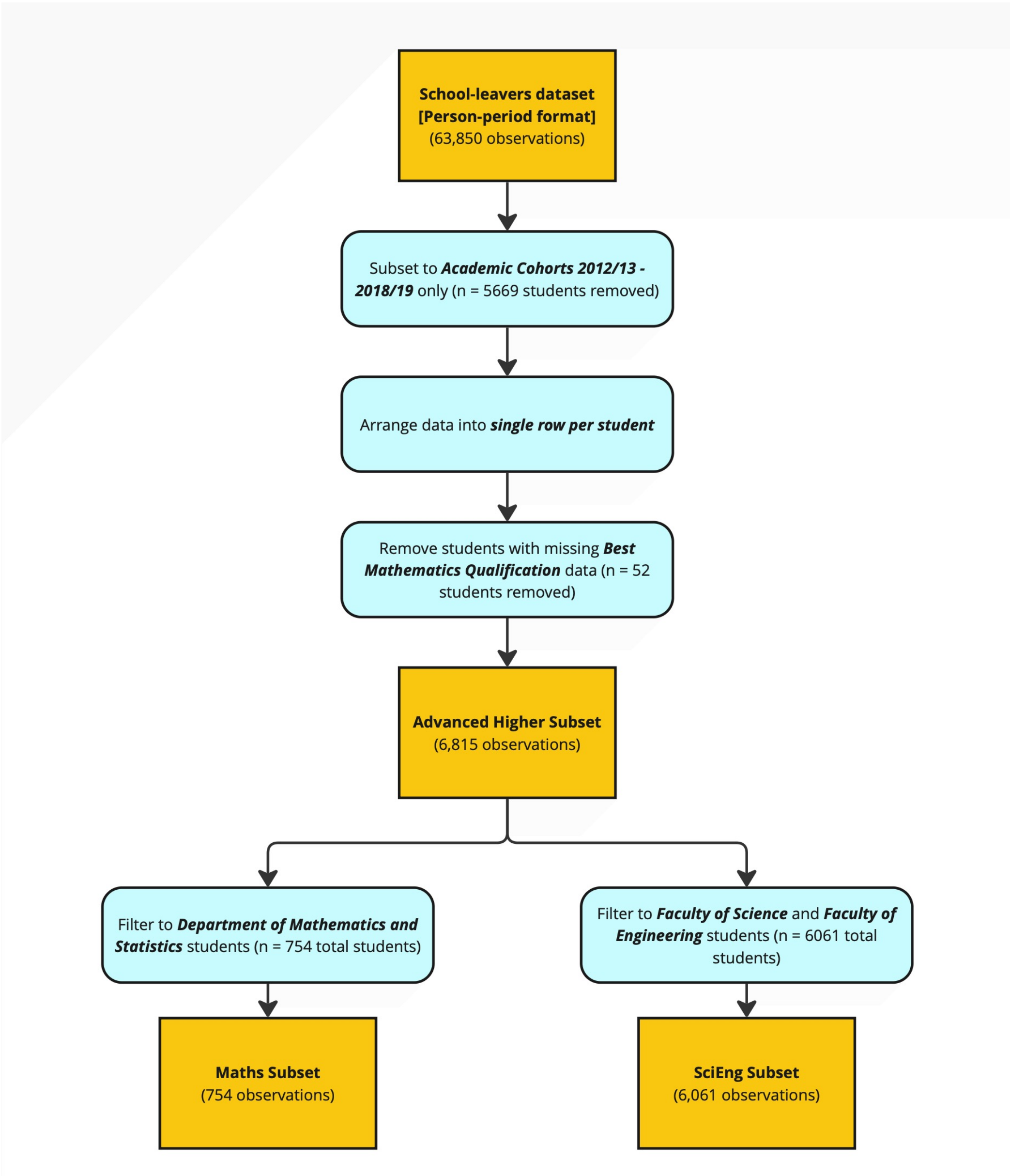


Figure 3.5: Deriving the Advanced Higher subsets used in Chapter 7.

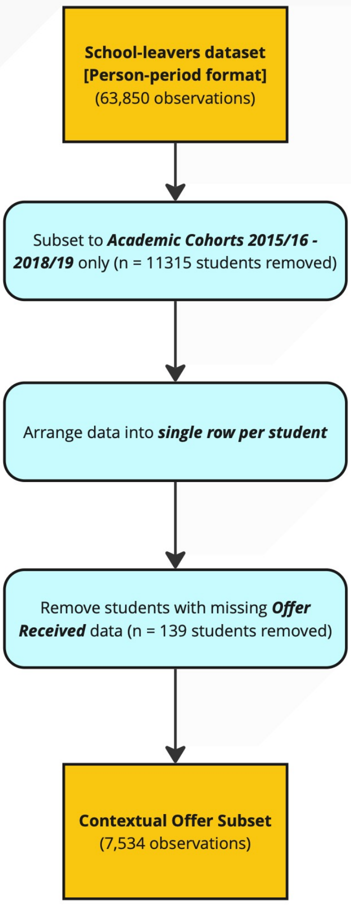


Figure 3.6: Deriving the Contextual Offer subset used in Chapter 8.

Finally, two subsets were used in Chapter 9: the **Drop-outs person-level subset** and the **Drop-outs person-period subset** (Figure 3.7). Both of these subsets contained the same school-leavers, specifically those from *Academic Cohorts* 2012/13 - 2018/19. Both subsets were required in order to fit the continuous and discrete survival models outlined in Chapter 6. Again, both formats of the data may be referred to simultaneously as the **Drop-outs subset** for brevity. Clarifications will be given on the specific format where it is necessary.

3.10 Summary

The **School-leavers dataset** was derived via a combination of the University of Strathclyde's student registration and attainment records, prospectus data and the Scottish Government datasets. It specifically contains Scottish school-leavers who began their registration between the *Academic Sessions* of 2012/13 - 2021/22. The following chapter will now define the explanatory variables and outcomes contained within the **School-leavers dataset**. Knowledge of these will aid the reader's interpretations of the results obtained in the analyses chapters.

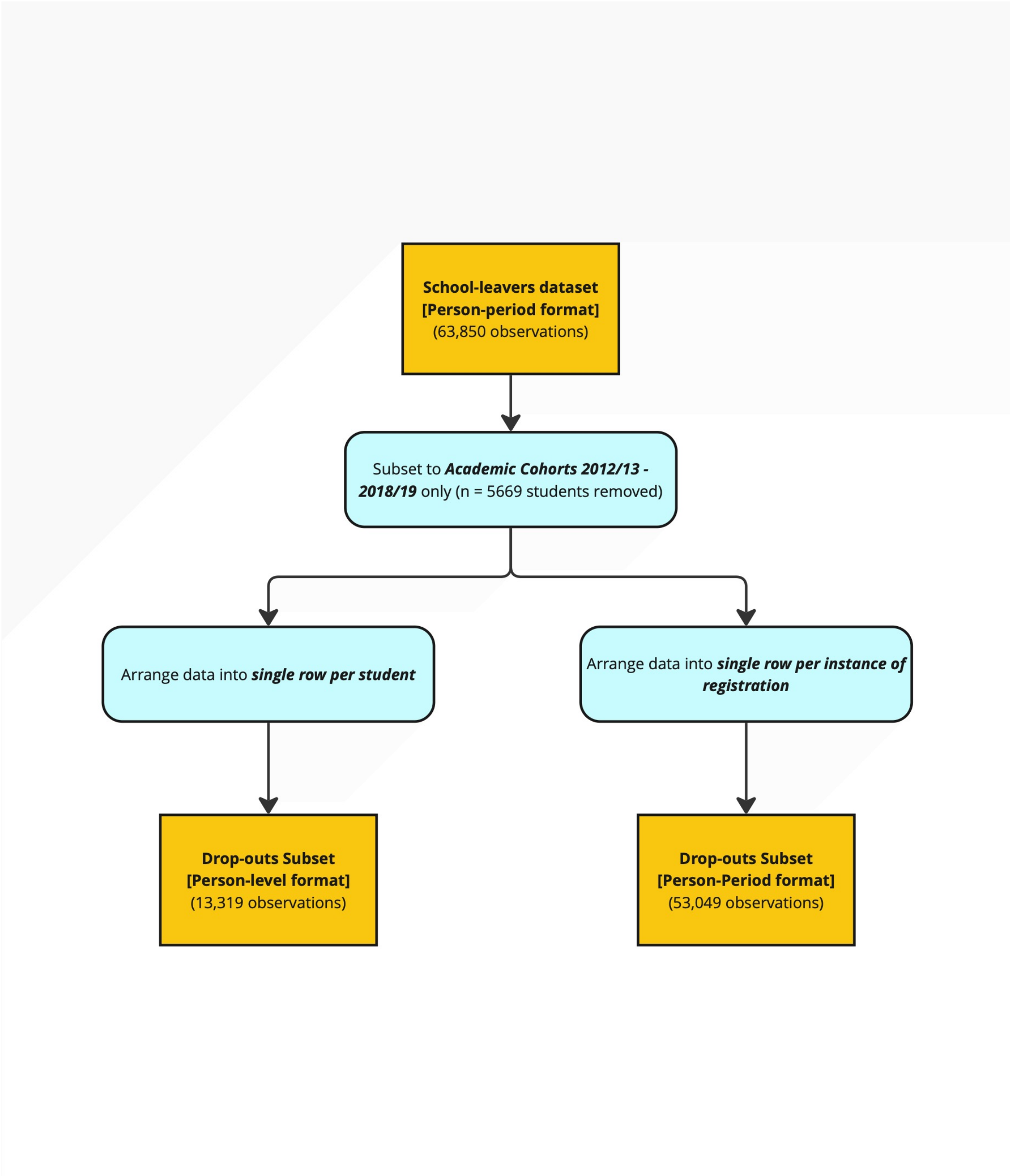


Figure 3.7: Deriving the Drop-outs subsets used in Chapter 9.

Chapter 4

The School-Leavers Dataset

This chapter will define each of the variables contained within the **School-leavers dataset** and where applicable, which parent datasets the variables were derived from. It is broken up into several sub-sections which summarise similar groups of variables. Descriptive summaries of the **School-leavers dataset** are explored in Chapter 5, but may be referenced in this chapter. This chapter also will detail whether or not each variable was "time-dependent" (i.e. could change value from instance-to-instance) or "fixed" (i.e. had the same value over all of a student's instances). As detailed in Section 3.6, there are both person-period and person-level formats for the **School-leavers dataset**. When deriving the **School-leavers person-level dataset** all time-dependent variables were fixed at the value that was true during the student's first registration session, unless stated otherwise. Finally, this chapter will highlight the limitations in the current release of the data and what future releases could improve upon (Section 4.7).

4.1 The Registration Variables

There are 13 variables related to the registration status of students (Table 4.1). The *HPLE Person Code* is an identifier that unites all registration instances for a given student, thus is used as a unique ID variable for each student.

There are ten unique *Academic Sessions* (or "sessions"), within the **School-leavers dataset**, spanning 2012/13 until 2021/22. The first instance of a student registered at the University was assumed to be the point they "entered" (joined or enrolled at) the University for the first time. The session a student entered the University is defined as the student's *Academic Cohort*. For example, every student who entered in *Academic Session* 2012/13 was considered part of the "2012/13" academic cohort. Students who entered at the same time were part of the same cohort regardless of how long they stayed registered at the University, or whether they repeated an academic stage. This makes comparisons between students over time more straightforward since their starting point can be identified. *Academic Session* is a time-dependent variable, while *Academic Cohort* is a fixed variable. There are roughly $1,800$ to $2,000$ school-leavers per *Academic Cohort* in the **School-leavers dataset** (Section 5.1 Table 5.1).

The *Reg. Year Count* is a time-dependent variable that records the number of *Academic Sessions* a student had spent registered at the university. For example, if student A entered in 2012/13 and spent three sessions registered at the university, then their value for *Reg. Year Count* in 2012/13 would be 1, in 2013/14 would be 2, and in 2014/15 would be 3. The time-dependent variable *Stage of Programme* refers to the stage of a degree programme for the student's relevant instance. Note that *Reg. Year Count* and *Stage of Programme* are not always equal (see Section 3.1). For the **School-leavers person-level dataset**, the total number of sessions a student had spent registered at the University was recorded using the fixed variable *Max Reg. Year Count*. Using the same example as before, the value of *Max Reg. Year Count* for student A would be 3.

Table 4.1: Excerpt of registration variables within the School-leavers dataset.

No.	**Variable**	**Variable Type**	**Origin**	**Description**
1	HPLE Person Code	Primary Key/ID	CSR	Unique person-tracking code
2	Academic Cohort	Fixed	CSR	Session entered (2012/13, 2013/14, ...)
3	Academic Session	Time-dependent	CSR	Current session (2012/13, 2013/14, ...)
4	Reg. Year Count	Time-dependent	Derived	Count of sessions passed since entered
5	Max Reg. Year Count	Fixed	Derived	Maximum sessions registered (1, 2, 3, ... etc.)
6	Stage of Programme	Time-dependent	CSR	1, 2, 3, 4, 5
7	UCAS Code	Fixed	CSR	UCAS Code for degree programme
8	Programme Title	Time-dependent	CSR	Title of degree programme
9	Department	Time-dependent	CSR	Department registered with
10	Faculty	Time-dependent	CSR	Faculty registered with
11	Changed Prog. Title	Fixed	Derived	Ever changed? (Yes/No)
12	Changed Dept.	Fixed	Derived	Ever changed? (Yes/No)
13	Changed Faculty	Fixed	Derived	Ever changed? (Yes/No)
14	Repeated Stage	Fixed	Derived	Ever repeated? (Yes/No)
15	Break	Fixed	CSR	Ever took break? (Yes/No)

There are four faculties at the University of Strathclyde: Faculty of Engineering, Faculty of Science, Faculty of Humanities and Social Sciences (HaSS), and Strathclyde Business School. For brevity and consistency, the latter two faculties will be referred to as the "Faculty of HaSS" and the "Faculty of Business", and all faculties may be referred to as simply "Engineering", "Science", "HaSS", or "Business". More details on the specific departments and degree programmes at the University of Strathclyde are explored in Appendix C.5. The *Programme Title*, *Department*, and *Faculty* variables are time-dependent. In the **School-leavers person-level dataset**, if a student ever changed *Programme Title*, *Department*, and *Faculty* this was recorded using the fixed variables: *Changed Prog. Title*, *Changed Dept.*, *Changed Faculty*; with binary responses “Yes/No”. Similarly, whether or not a given student ever took a break in their studies is given by the fixed variable *Break*. A more detailed breakdown of these variables is given in Appendix C.6.

Strictly speaking, a degree programme is one which has a unique *UCAS Code*. For example, the *UCAS Code* “G100” corresponds to the degree programme “BSc Honours Mathematics” and “G101” to “MMath Mathematics”. However, each student’s *UCAS Code* remains fixed at the code for the degree programme they entered in their first registration session. Additionally, the *Programme Title*, which is time-dependent, does not contain the level of the programme, (e.g. “BSc”, “MEng”). This makes it impossible to distinguish between students who are registered on the BSc Honours versus Integrated Masters version of a degree programme past their second registration year (see Appendix C.6).

4.2 The Retention-Progression-Outcome Variables

The group of Retention-Progression-Outcome (RPO) variables (Table 4.2) allows for the tracking of a student’s entire journey through the University. This includes (but is not limited to) progression to the next academic year, repeating a year, voluntary suspension, exiting, or graduation. The RPO variables are time-dependent, and organised into four “Levels” (0, 1, 2, 3) that become progressively more detailed about the registration status of a student at the end of a particular instance. A reminder that an "instance" is a unique registration record in a given year for a given student (introduced in Section 3.2). For example, *RPO Level 0* details whether, by the end of the relevant instance, the student was still actively studying, taking a break in their studies, had exited with a qualification or had exited without a qualification. In the most detailed level, *RPO Level 3*, there are 31 unique responses ranging from the degree classification obtained to the reason for a student's exit. When information from each of the levels are combined, they provide a holistic picture of a student's registration status for a particular instance. In the **School-leavers dataset** there are 87 unique combinations of RPO Levels, each representing a unique registration status. The RPO Levels are too detailed and complicated for analyses. Thus, binary outcome variables of interest were derived from the RPO Levels. These are the fixed variables *Ever Dropped Out*, *Retention Status*, and *Completion Status* and the time-dependent variable *Drop-out Status*.

For the fixed outcome *Ever Dropped Out*, a student was considered to have "dropped out" if one of their instances had ever recorded an *RPO Level 1* value of "Exit with no award" and was considered to have not dropped out otherwise. For the time-dependent outcome *Drop-out Status*, a student was considered to have "dropped out" in the specific instance where an *RPO Level 1* value of "Exit with no award" was recorded, and was considered to have not dropped out, or "censored", in the remaining instances. An exception was made in that students who had been registered for 6 academic sessions or longer were defined as "censored" even if they had been recorded as a drop-out. More details on the reasons for doing so and the definition of censoring are explained in Section 6.9.3. Between $7 - 11\%$ of each *Academic Cohort* were school-leavers who dropped out of the University.

A student was considered to have been "retained" at the end of their first session *(Retention Status)* if in their first instance they recorded an *RPO Level 1* of "Progressed" and was considered as "Exited/Re-registered" otherwise. This definition of retention is consistent with that used in the University's Key-performance indicators [8]. It is also consistent with a discussion paper on retention rates by the Commissioner for Fair Access 2018 [37]. Around 90% of students were successfully retained within the **School-leavers dataset**.

A student was considered to have achieved "completion" if they had been awarded at least a Bachelor's with Honours Degree (of any classification) within four *Academic Sessions*. More precisely, a *Completion Status* of "Yes" was satisfied if any of the following conditions were satisfied:

(i) the student was recorded as achieving a classifiable Bachelor's with Honours degree prior to their fifth instance at the University, or

(ii) the student was recorded as having "Progressed" into the fifth stage of their degree programme prior to their fifth instance, or

(iii) the student was recorded as achieving a classifiable or unclassifiable Masters Degree or a Masters with Honours Degree award prior to their fifth instance.

Conditions (ii-iii) captures all students who had achieved their Integrated Masters Degree within four years, or had progressed into the "5th stage" of their Integrated Masters Degree within four sessions. These are considered successful completions since even if the student were to fail the final stage of their Integrated Masters

Degree, they would have at least exited the University with a Bachelor's with Honours degree. Around 74% of students successfully achieved completion within the **School-leavers dataset**.

The no-detriment policy enacted for *Academic Sessions* 2019/20 and 2020/21 [47–49], likely had an effect on the academic outcomes of school-leavers. Hence, two binary (Yes/No) variables were derived which grouped together cohorts affected by the no detriment policy. Anyone who belonged to the *Academic Cohorts*: 2019/20 or 2021/22, was recorded as "Yes" for the *No Detriment Retention* variable, since these cohorts would have had their 1st *Stage of Programme* affected by the policy. Anyone in the 2016/17, 2017/18, 2018/19 cohorts, was recorded as "Yes" for the *No Detriment Completion* variable, since these cohorts would have had their 2nd, 3rd, 4th, or 5th *Stage of Programme* affected by the policy.

The precise academic session a student exited the university (dropped out) or achieved completion is denoted by the fixed variables *Session Exited* and *Session Achieved Completion*, respectively. The *Academic Session* of a given student's most recent instance is recorded by the fixed variable *Most Recent Session.*

Table 4.2: Excerpt of Retention-Progression-Outcome variables within the School-leavers dataset.

No.	Variable	Variable Type	Origin	Description
16	RPO Level 0	Time-dependent	CSR	4 unique registration statuses
17	RPO Level 1	Time-dependent	CSR	4 unique registration statuses
18	RPO Level 2	Time-dependent	CSR	26 unique registration statuses
19	RPO Level 3	Time-dependent	CSR	31 unique registration statuses
20	Ever Dropped Out	Outcome/Fixed	Derived	Eventually dropped out? (Yes/No)
21	Drop-out Status	Outcome/Time-dependent	Derived	Dropped out of relevant session? (Yes/No)
22	Retention Status	Outcome/Fixed	Derived	Progressed at end of first session? (Yes/No)
23	Completion Status	Outcome/Fixed	Derived	Bachelor's with Honours within four years? (Yes/No)
24	No Detriment Retention	Fixed	Derived	Yes if in 2019/20 or 2021/22 cohort, No otherwise.
25	No Detriment Completion	Fixed	Derived	Yes if in 2016/17, 2017/18 or 2018/19 cohort, No otherwise.
26	Session Exited	Fixed	Derived	2012/3, 2013/14, ...
27	Most Recent Session	Fixed	Derived	2012/3, 2013/14, ...
28	Session Achieved Completion	Fixed	Derived	2012/3, 2013/14, ...

4.3 The Demographic and Socio-economic Variables

There are six demographic variables that are included in, or are derived from, the University's CSR: *Sex*, *Age*, *Ethnicity*, *Disability Status*, *Postcode*, and *Local*. *Sex* is a fixed binary variable (Male or Female) defined as the sex a student had disclosed on their UCAS application. *Age* is a time-dependent variable defined as the age (in years) of the student as of 1st August of the relevant instance. For example, if the academic session of a given student's instance was 2012/13, then the student's *Age* would be their age as of the 1st August 2012. *Ethnicity* is a fixed binary variable ("White" or "Ethnic-minority") that groups together various ethnicities that were disclosed in the students' UCAS applications. All recorded ethnicities and the binary grouping they were assigned to are shown in Appendix C Table C.3. Similarly, *Disability Status* is a fixed binary variable that groups together anyone who identified a disability in their UCAS application. All recorded disabilities - whether mental-health or physical - are shown in Appendix C Table C.4 as well as how they were classified into the binary grouping. The *Postcode* variable refers to the student's home postcode provided in their UCAS application to the University. The *Local* variable was derived using *Postcode* and refers to whether the student's home postcode was "local" to the University. This was defined as any postcode that contained a leading "G", such that it was assumed to fall within the "Greater-Glasgow" geographical region.

The *SIMD Quintile* and *Urban/Rural Status* variables were linked to the data from the Scottish Government datasets (see Section 3.3). Each student's *SIMD Quintile* was derived from the version of the SIMD indicator that was in use at the time of their application to the University (see Section 3.3). Thus, students from the same *Postcode* but different *Academic Cohorts* may have different values for *SIMD Quintile*. Visualisations of SIMD Quintiles across Scotland and Glasgow are shown in Figures 2.1 and 2.2, respectively, from Chapter 2. Only the 2020 version of the Urban/Rural Indicator was used so students from the same *Postcode* will have the same *Urban/Rural Status*. The response for the 3-Fold indicator is "Urban"[1], "Accessible Rural" and "Remote Rural" [2]. The precise definition of Urban/Rural status can be found in the supporting documentation on the Scottish Government websites [2, 98]. A visualisation of the Urban/Rural 3-Fold Indicator for data zones across Scotland is provided in Figure D.3.

[1]Updated for the 2020 release to now be called "Rest of Scotland" [98], though "Urban" will be preferred here.

Table 4.3: Excerpt of demographic and socio-economic variables within the School-leavers dataset.

No.	**Variable**	**Variable Type**	**Origin**	**Description**
29	Sex	Fixed	CSR	Male/Female
30	Ethnicity	Fixed	Derived	"White" or "Ethnic-minority"
31	Disability Status	Fixed	Derived	"Disabled" or "None disclosed"
32	Age	Time-dependent	CSR	Age as of 1st August of relevant academic session
33	Postcode	Fixed	CSR	Student's postcode
34	Local	Fixed	Derived	Postcode within Glasgow area? (Yes/No)
35	SIMD Quintile	Fixed	ScotGov	SIMD Quintile of Postcode
36	Urban/Rural Status	Fixed	ScotGov	Status of Postcode

4.4 The Prior Attainment Variables

As stated in Section 3.4, students' prior attainment from secondary education was summarised using "Grade Strings". Since comparing students' Grade Strings to one another would be too computationally expensive once all combinations were considered, it was decided to convert these strings into "simple" points: where for Highers, A - 3 points, B - 2 points, and C - 1 point, and similarly for Advanced Highers. The simple points total for each student was calculated separately for their Highers *(All Highers Reg. Points)* and Advanced Highers *(All Adv. Highers Points)*. These were then summed together to obtain their *Prior Attainment Points* totals which represents the potential of a student based on their entire academic performance prior to entry. A summary of these *Prior Attainment Points* per *Academic Cohort* is provided in Section 5.2. The total points achieved in Highers at the "point of application" was also calculated (*All*

Highers Appl. Points). This was defined as the simple points conversion of all Higher grades awarded prior to the $(n_0 - 1)$ *Academic Session*, where n_0 is the student's first academic session at the University (*Academic Cohort*).

The "simple" method of calculating *Prior Attainment Points* was considered easier to interpret than, for example, the UCAS tariff points system (A – 33 points, B – 27 points, C – 21 points for Highers) in which many UK and international qualifications are compared, and therefore the definition of a single tariff point is not clear. Under this simple system, a single point increase corresponds to an increase in grade, for example a C to a B, or a B to an A. UCAS Tariff Points also had issues when comparing prior attainment to entry requirements. These issues are detailed further in Appendix C Section C.3.

4.4.1 Prior Attainment Quintile

For some analyses, it was preferable to express student's prior attainment as a categorical variable. The *Prior Attainment Quintile* groups together students who attained similar levels of attainment within each *Academic Cohort*. First, students were ranked according to their *Prior Attainment Points* compared to their peers within the same *Academic Cohort*. Tie-breaker rules were used to decide the ranking order between students with the same *Prior Attainment Points* which are outlined in Appendix C Section C.4.1. Students were then grouped together into five quintiles; those within the top 20% were classified as "Quintile 5", followed by those in the top 20-40% as "Quintile 4", and so on until the bottom 20% point totals were classified as "Quintile 1". The quintile a student belonged to was named their *Prior Attainment Quintile*. The appropriateness of this interpretation was verified using Figure C.1 from Appendix C.4.

Table 4.4: Excerpt of attainment variables within the School-leavers dataset.

No.	**Variable**	**Variable Type**	**Origin**	**Description**
37	All Highers Reg. Points	Fixed	Derived	Sum total at registration
38	All Highers Appl. Points	Fixed	Derived	Sum total at application
39	All Adv. Highers Points	Fixed	Derived	Sum total at registration
40	Prior Attainment Points	Fixed	Derived	Highers + Advanced Highers
41	Prior Attainment Quintile	Cohort-dependent	Derived	Total Points converted to Quintiles

4.5 The Entry Requirement Variables

There were six variables (Table 4.5) derived using information from the *Prospectus dataset* and the *Attainment on Entry* data contained in the CSR.

The standard entry requirements for the degree programme a student entered in their first registration session (*Std. Entry Req.*) came in the form of a "Grade String" as described previously (see Section 3.4). These Grade Strings only cited Highers (and not Advanced Highers) and were typically four-to-five grades in length (e.g. AABBB, AAAA, BBBBB etc.). These were converted into tariff points (*Std. Entry Tariffs*) using the "simple" system described previously (see Section 3.4). There were some changes to the standard entry requirements over time; these are detailed in the data explorations from Section 5.5. While the prospectus data also had "2nd sitting" standard entry requirements and minimum entry requirements, these were ignored since not every programme cited these requirements. Minimum entry requirements could also be simply derived from standard entry requirements by subtracting one grade, e.g. AAAB becomes AABB.

To make comparisons between entry requirements and each student's prior attainment mimic real-world practice, each student's Grade String at Higher was modified such that the only the best five Highers they had been awarded at the point of application were considered. Since the population of interest was school-leavers, it was assumed that this would represent the attainment they had achieved in S5. This modified Grade String was then converted into "simple" points (see Section 3.4) to derive the *Best 5 Highers Appl. Points* variable. For example, if student A had obtained AAABBCCC at Higher across S5 and S6, then their *Best 5 Highers Appl. Points* would be equal to AAABB = 13 points. Advanced Highers were not considered here since these are typically aimed at S6 students. For each student, it was recorded whether or not they, at the point of application, had met/exceeded, or had not met, the standard entry requirements for the degree programme they entered in their first registration session (*Met Std. Entry Req.*).

The **School-leavers dataset** did not contain information on the conditions or type of offer a student received. Thus, a proxy indicator was necessary to identify whether a student had likely received a standard or contextual offer (*Offer Received*). As highlighted in Section 2.3, the University of Strathclyde considers students eligible for a contextual offer to be those from SIMD Quintiles 1 or 2, those who attended a low-progression school, and those with care-experience or caring responsibilities [56]. However, in the **School-leavers dataset**, only the student's SIMD Quintile was available; no other eligibility criteria were known.

Thus, the *Offer Received* variable was defined such that students were classified as “Contextual Offer” if they were from SIMD Quintiles 1 or 2 and had achieved below the standard entry requirements at the time of application *Met Std. Entry Req.* = “No”, otherwise they were classified as “Standard Offer” students.

Table 4.5: Excerpt of variables within the School-leavers dataset that were derived from prospectus data.

	Variable	**Variable Type**	**Origin**	**Description**
42	Standard Entry Req.	Fixed	Prospectus	Lettered grades (e.g. AABBC)
43	Standard Entry Tariffs	Fixed	Derived	Grades converted to points (continuous)
44	Met Std. Entry Req.	Fixed	Derived	Compares to best 5 Highers (Yes/No)
45	Best 5 Highers Appl. Points	Fixed	Derived	Sum total at application
46	Offer Received	Fixed	Derived	Contextual Offer/Standard Offer

4.6 The Advanced Higher Variables

A student’s *Best Maths Qualification* was defined as whether a student had attained “Higher” or “Advanced Higher” in “Mathematics”, “Statistics”, or “Mathematics of Mechanics”. A student’s *Best English Qualification*, *Best Biology Qualification*, *Best Chemistry Qualification*, and *Best Physics Qualification* are similarly defined for their respective secondary-school subjects (Table 4.6).

The University of Strathclyde Prospectuses from 2013/14 - 2021/22 were examined to determine whether or not degree programmes within the faculties of Science and Engineering recommended Advanced Higher Mathematics. The Prospectus for 2012/13 session could not be located, although it was assumed that recommendations on Advanced Highers were similar to the 2013/14 session. An exploration of these recommendations, and the particular language used, is discussed in Appendix C.7. From this information, several binary explanatory variables were created: *AH Mathematics Recommended, AH Biology Recommended, AH Physics Recommended, AH Chemistry Recommended*; each with responses "Yes" or "No" (Table 4.6). If a programme had ever encouraged students entering into stage 1 to take up Advanced Higher Mathematics, the programme was recorded as a "Yes", regardless of how strong the language used to encourage students. Otherwise, a programme was said to have not recommended Advanced Higher Mathematics, i.e. was recorded as "No".

Table 4.6: Excerpt of variables related to Advanced Highers within the School-leavers dataset.

No.	**Variable**	**Variable Type**	**Origin**	**Description**
47	Best English Qualification	Fixed	Derived	Higher, Advanced Higher
48	Best Maths Qualification	Fixed	Derived	Higher, Advanced Higher
49	Best Biology Qualification	Fixed	Derived	Higher, Advanced Higher
50	Best Physics Qualification	Fixed	Derived	Higher, Advanced Higher
51	Best Chemistry Qualification	Fixed	Derived	Higher, Advanced Higher
52	AH Mathematics Recommended	Fixed	Prospectus	Advanced Higher (Yes/No).
53	AH Biology Recommended	Fixed	Prospectus	Advanced Higher (Yes/No).
54	AH Physics Recommended	Fixed	Prospectus	Advanced Higher (Yes/No).
55	AH Chemistry Recommended	Fixed	Prospectus	Advanced Higher (Yes/No).

4.7 Limitations and Potential Improvements

The data collected for the CSR was re-examined and finalised after the end of each academic session. If there were multiple possible values that could be true for each instance, then the final value was taken as the value that was true by the end of that instance. For example, if a student had changed degree programme from "MEng Electrical and Mechanical Engineering" in semester 1 of 2016/17 to "BSc Mathematics" in semester 2 of 2016/17, then the student's value for *Programme Title* will be "Mathematics" for the whole of 2016/17. Similarly, the time a student has spent registered at the university (*Reg. Year Count*) is rounded up to the nearest whole academic session.

Currently, the definition of *Offer Received* means that those not from SIMD Quintiles 1 or 2, but satisfying the other criteria [56] for a contextual offer, will be misclassified as standard offer students. However, the true misclassification rate is expected to be low given that anecdotal evidence from 2024/2025 applications showed significant overlap between SIMD Quintile and the other eligibility criteria. Thus, there is confidence that most students will be correctly classified as either standard or contextual offer. The classification rate could be improved with access to secondary school attended and care-experience/caring-responsibilities, although access to offer data could remove the need for the proxy indicator, *Offer Received*, entirely. The Strategy & Policy team at the University of Strathclyde is currently working towards inclusion of offer data in future updates of the CSR.

The analysis could also have been improved with more complete and robust records of entry requirements. The lack of prospectus data prior to 2015/16 means that Academic Cohorts 2012/13 – 2014/15 cannot not be considered for the contextual offers analysis (Chapter 8). Missing data in the standard entry requirements also meant that around 338 (2.54%) students from 2015/16 – 2018/19 had to be removed, albeit this number was very small (around 1.8% of all observations from 2015/16 – 2018/19). Of the 338 students, 225 were from the Faculty of HaSS, 77 were from Science, 36 were from Business, and 0 were from Engineering. The missing data introduces a negligible amount of bias towards/against faculties.

Care should be taken when interpreting a student's *Prior Attainment Points*. For example, these do not account for the fact that some subjects may be more relevant than others for a particular degree programme. Nor do they account for a student's other qualifications outwith Higher and Advanced Higher. Given that the analyses within this thesis only consider Scottish school-leavers however, the relevance of these other qualifications should be minor.

Grouping together students using the binary classifications: *Disability Status* and *Ethnicity*, was necessary to have enough observations within each level of the variables such that the statistical models could be fit to the data. Future analyses could make use of the more granular disability and ethnicity data, since the student experience of ethnicities within the "Ethnic-minority" category are likely to be varied, and likewise for "Disabled" students.

The **School-leavers dataset** did not contain an exhaustive list of the information collected on students. For example, it is not currently possible to discriminate between undergraduates who were registered on a Bachelor's with Honours Degree versus an Integrated Masters Degree. The expected length of a student's degree programme is also unknown. The only socio-economic indicators within the **School-leavers dataset** were *SIMD Quintile* and *Urban/Rural Status*, future analyses could attempt to construct more area and school-level indicators using the school attended or postcode a student is resident in. Given the current state of data collection in Scotland, it is unlikely that individual characteristics such as low-income and free-school meals data could be made available in the near future (see Chapter 2 Section 2.3 for more details). Finally, a feasible improvement to the **School-leavers dataset** would be incorporating a measure of attainment at the University-level, for example marks in specific modules or credit-weighted averages per academic stage. This information would be useful for improving upon the models constructed in Chapters 7, 8 and 9.

Chapter 5

Exploring the School-Leavers Dataset

This chapter descriptively summarises the variables within the **School-leavers dataset**. The aim of this chapter is to identify which additional variables may be appropriate to investigate within multivariable model fits to the academic outcomes of interest: *Retention Status*, *Completion Status*, and *Drop-out Status*. The explorations may also identify potential associations for future avenues of research. Some additional visualisations can be found in Appendix D.

Note that this chapter analyses the **School-leavers dataset** which includes all students from *Academic Cohorts* 2012/13 - 2021/22. This differs from the analyses chapters which examine their own subsets. These subsets either contain students from *Academic Cohorts* 2012/13 - 2018/19 (Chapters 7 and 9) or *Academic Cohorts* 2015/16 - 2018/19 (Chapter 8).

5.1 Summary of Categorical Variables

There are roughly $1,800$ to $2,000$ school-leavers per *Academic Cohort* in the **School-leavers dataset** (Table 5.1). There is a rough 50:50 split in the proportion of school-leavers who: are male and female, were aged 18 and aged 17 or under at first registration, and those whose home postcodes are within/outside Glasgow. The vast majority of school-leavers come from "Urban" areas (61%), although the *Urban/Rural Status* is unknown/missing for 26% of school-leavers. A small proportion of school-leavers (7%) identify as an ethnicity classified as "Ethnic-minority". Similarly, only a small proportion (6%) had disclosed a disability to the University upon their first registration. The Faculty of HaSS has the largest proportion of school-leavers, followed by Engineering, Science, and then Business.

The proportions of school-leavers entering the University from *SIMD Quintiles* 1 and 2 have increased over time (Figure 5.1). This is unsurprising given the national strategy to increase representation from these quintiles [3] and the University's own targets set out in its outcome agreement with the SFC. In 2020, the University of Strathclyde reached the national target of 20% of SIMD Quintile 1 for full-time, first-degree undergraduates entrants, 10 years early [68], although this proportion was 16.3% within the **School-leavers dataset** (Figure 5.1). This difference is likely because many of the SIMD Quintile 1 students counted towards the University's targets include college-entrants and mature students which were removed from the **School-leavers dataset**.

Table 5.1: Summary of variables examined in the school-leavers dataset.

Variables	Levels	Count	Proportion
Academic Cohort	2012/13	1827	0.10
	2013/14	1866	0.10
	2014/15	1953	0.10
	2015/16	1985	0.10
	2016/17	1972	0.10
	2017/18	1893	0.10
	2018/19	1823	0.10
	2019/20	1848	0.10
	2020/21	1813	0.10
	2021/22	2008	0.11
Age at Entry	17 or under	9397	0.49
	18	9591	0.51
Disability Status	Disabled	1086	0.06
	None	17902	0.94
Ethnicity	Ethnic-minority	1386	0.07
	White	17602	0.93
Faculty	Business	3311	0.17
	Engineering	5336	0.28
	HaSS	5895	0.31
	Science	4446	0.23
Local to Glasgow	Glasgow-based	9204	0.48
	Outside Glasgow	9784	0.52
SIMD Quintile	1	2194	0.12
	2	2797	0.15
	3	3038	0.16
	4	4073	0.21
	5	6886	0.36
Sex	Female	9470	0.50
	Male	9518	0.50
Urban/Rural Status	Accessible Rural	1896	0.10
	Remote Rural	575	0.03
	Urban	11662	0.61
	Unknown	4855	0.26
Overall	**-**	**18988**	**1.00**

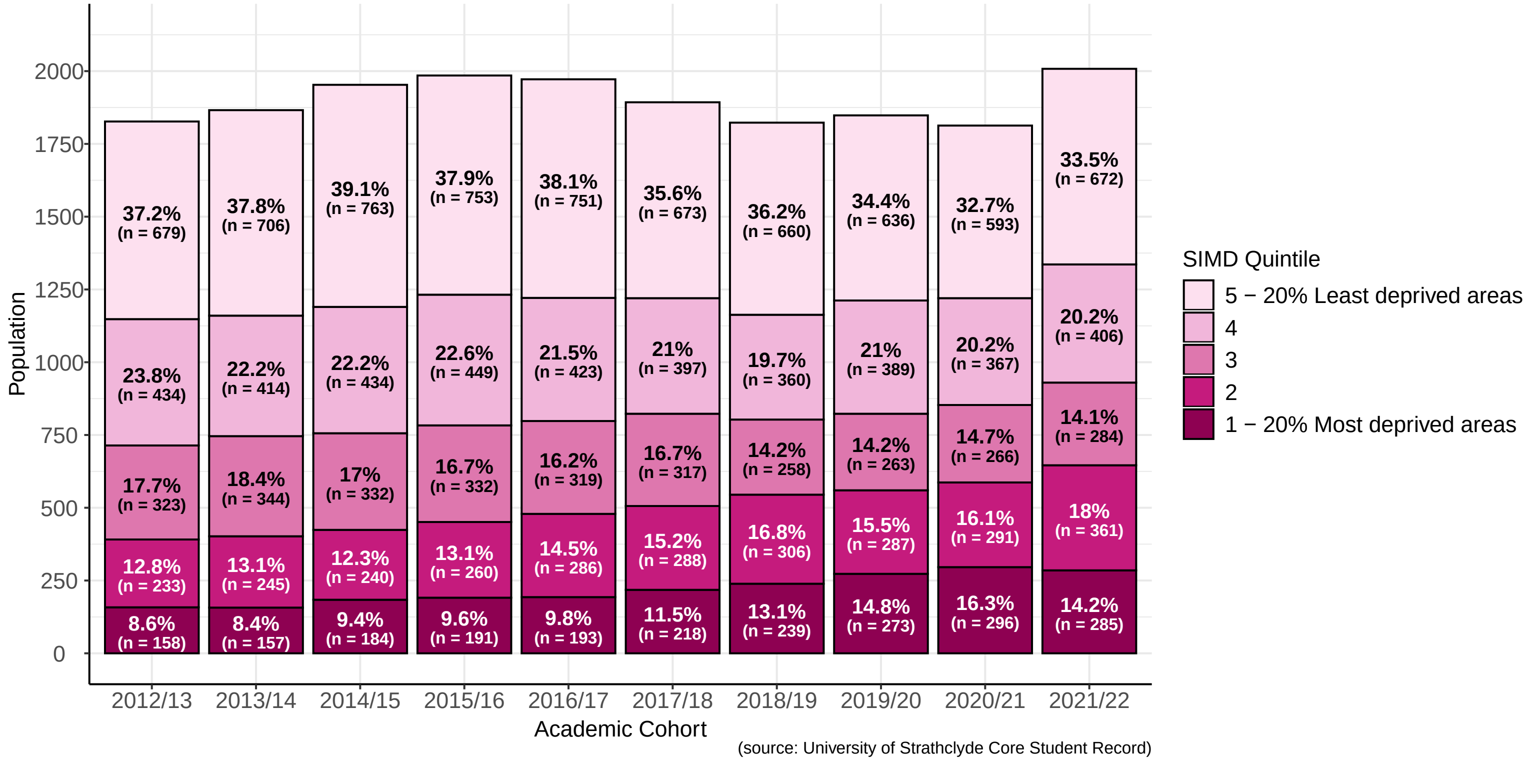


Figure 5.1: Population of school leavers and their SIMD Quintile from 2012/13 to 2021/22.

5.2 Summary of Prior Attainment

Boxplots were created to compare the *Prior Attainment Points* across different groups of students (Figures 5.2, 5.3, 5.4 and 5.5), though outlier points were hidden in these plots to preserve anonymity. The *Prior Attainment Points* of school-leavers appears to have increased with each successive *Academic Cohort* (Figure 5.2). Students who achieved successful academic outcomes at University tended to have higher *Prior Attainment Points* compared to those who did not (Figure 5.3). Students from *SIMD Quintiles* 1 and 2 had median prior attainment that was 2 points lower (equivalent to a Higher B grade) than Quintiles 3, 4, and 5 (Figure 5.4). The Faculty of Engineering had the highest median points, followed by Business, HaSS, and Science (Figure 5.4). The variance of these points was larger within Engineering and Science than Business and HaSS. This might be because students in STEM programmes are more likely to have Advanced Highers, though this would need to be confirmed through further exploration of the data. Unsurprisingly, standard offer students had higher median *Prior Attainment Points* compared to contextual offer students (5 points higher - Figure 5.5), and similarly for those whose *Best Mathematics Qualification* was an Advanced Higher versus Higher (4 points higher - Figure 5.5). There did not appear to be substantial differences in the variance nor median *Prior Attainment Points* between the levels of *Sex, Ethnicity, Disability Status, Age at Entry, Local to Glasgow, Urban/Rural Status, No Detriment Retention* and *No Detriment Completion* (Figures 5.4, 5.5).

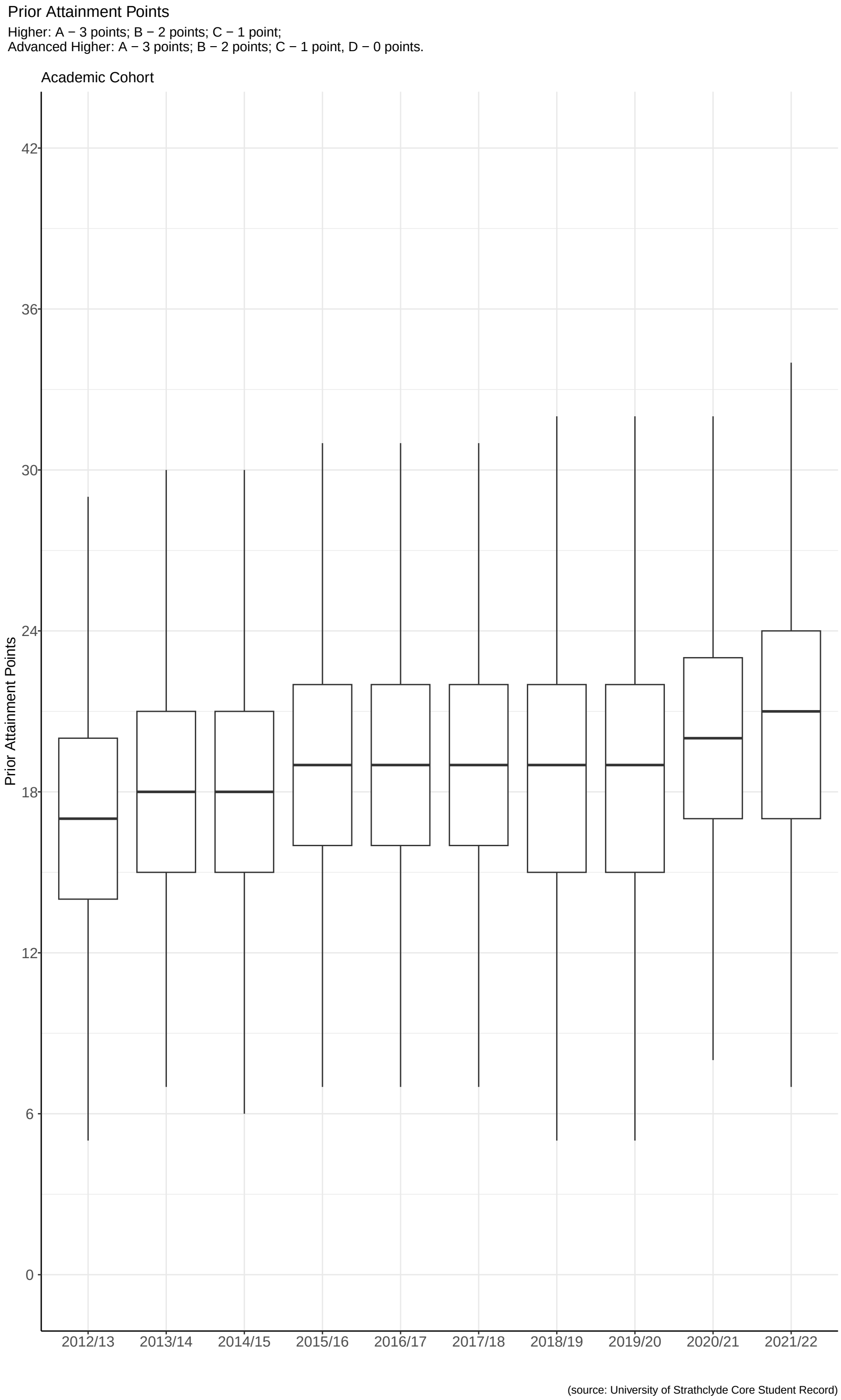


Figure 5.2: Boxplot of Prior Attainment Points versus Academic Cohort (1 of 3). Outlier points have been hidden to prevent identification of students when comparing outliers across the other boxplot figures.

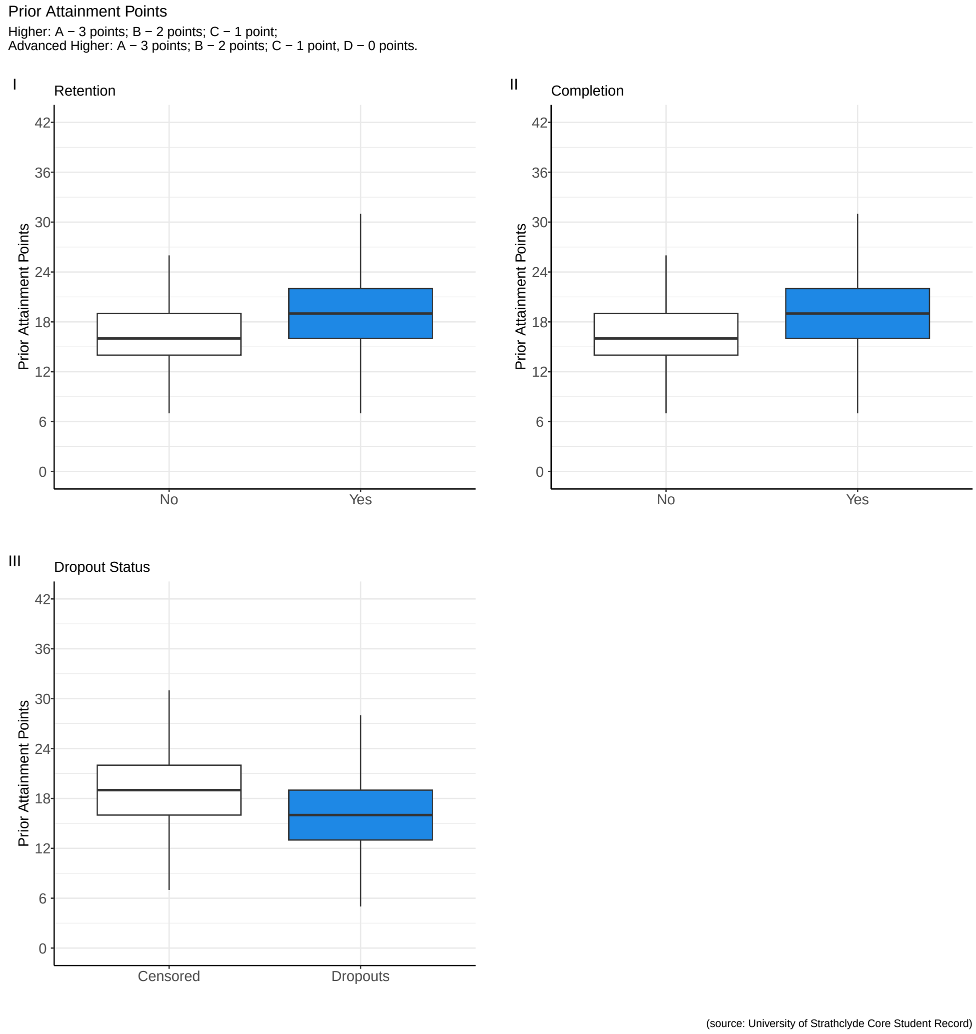


Figure 5.3: Boxplot of Prior Attainment Points versus outcome variables. Outlier points have been hidden to prevent identification of students when comparing outliers across the other boxplot figures.

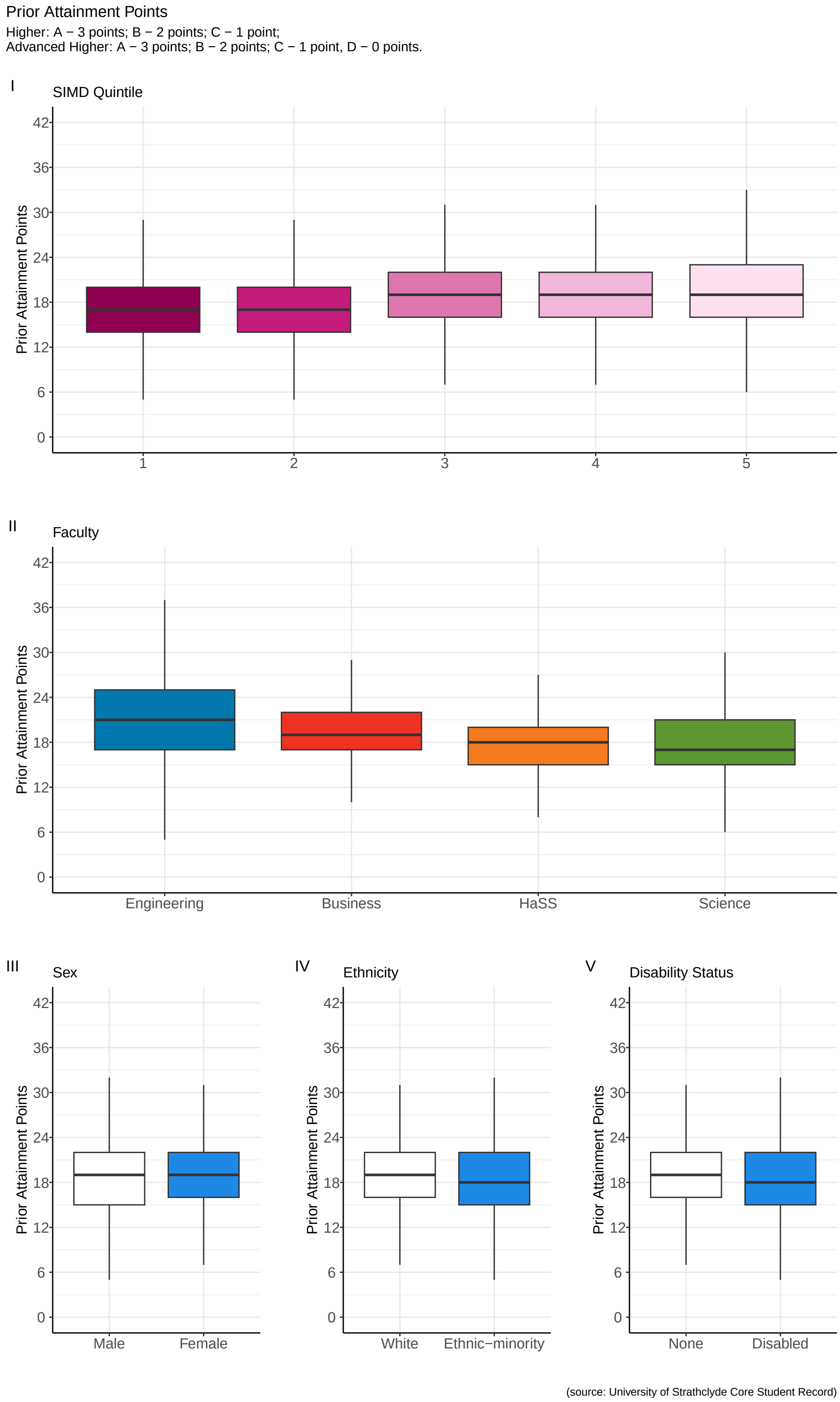


Figure 5.4: Boxplot of Prior Attainment Points versus explanatory variables (2 of 3). Outlier points have been hidden to prevent identification of students when comparing outliers across the other boxplot figures.

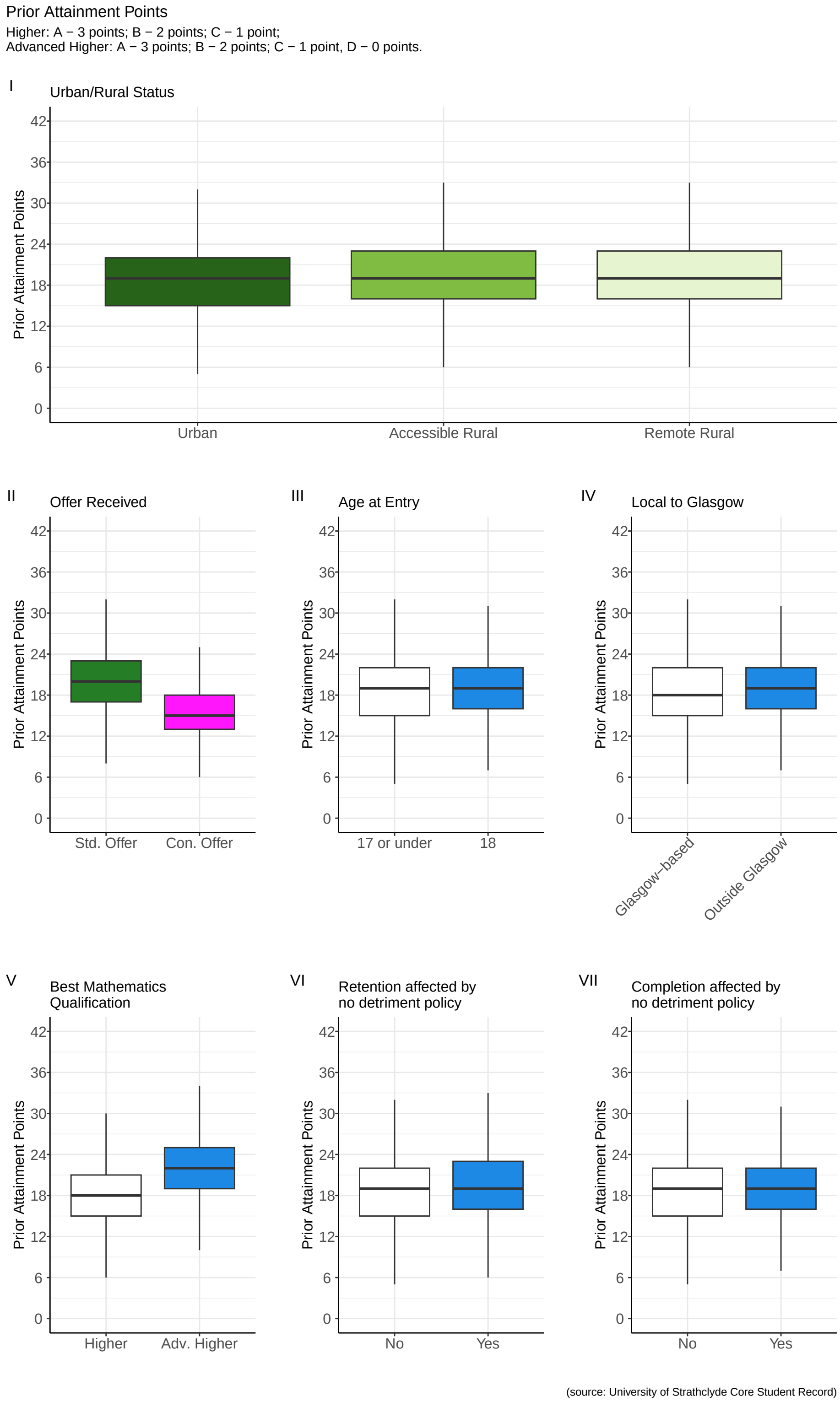


Figure 5.5: Boxplot of Prior Attainment Points versus explanatory variables (3 of 3). Outlier points have been hidden to prevent identification of students when comparing outliers across the other boxplot figures.

5.2.1 Prior Attainment Quintile

A reminder that *Prior Attainment Quintile* creates five groups of students within each *Academic Cohort* that have similar attainment profiles (a more precise definition is given in Section 4.4.1). Hence, the range of *Prior Attainment Points* within each *Prior Attainment Quintile* can vary from cohort-to-cohort (for example, due to grade inflation over time). A summary of these ranges are shown in Table 5.2. From this, it can be seen that those with "average" attainment compared to their peers (*Prior Attainment Quintile 3*) had around 18.7 points, which is roughly the equivalent of six As at Higher. *Prior Attainment Points* appear to be approximately normally distributed within each *Academic Cohort* (Figure 5.6). The *Prior Attainment Points* of school-leavers appear to increase with each successive Academic Cohort (Table 5.3). The mean increased by 1.3 points over the 3-year period between 2012/13 and 2015/16 and increased by 1.2 points in the single year between 2019/20 and 2020/21. The latter increase coincided with the cohorts of students who received teacher assessed grades at secondary education (see Section 2.2 for more details).

Table 5.2: Summary of Prior Attainment Points by Prior Attainment Quintile.

Prior Attainment Quintile	**Min**	**Q1**	**Median**	**Q3**	**Max**	**Mean**	**S.D.**
1	2.00	12.00	13.00	14.00	16.00	12.8	1.86
2	14.00	15.00	16.00	17.00	19.00	16.3	1.14
3	16.00	18.00	19.00	19.00	22.00	18.7	1.23
4	18.00	20.00	21.00	22.00	25.00	21.3	1.38
5	21.00	24.00	25.00	27.00	42.00	25.4	2.36

Table 5.3: Summary of Prior Attainment Points grouped by Academic Cohort.

Academic Cohort	N	Min	Q1	Median	Q3	Max	Mean	SD
2012/13	1827	5.00	14.00	17.00	20.00	34.00	17.5	4.38
2013/14	1866	5.00	15.00	18.00	21.00	38.00	18.1	4.29
2014/15	1953	2.00	15.00	18.00	21.00	42.00	18.4	4.28
2015/16	1985	6.00	16.00	19.00	22.00	38.00	18.8	4.43
2016/17	1972	3.00	16.00	19.00	22.00	33.00	18.9	4.44
2017/18	1893	7.00	16.00	19.00	22.00	37.00	18.9	4.42
2018/19	1823	3.00	15.00	19.00	22.00	36.00	18.8	4.74
2019/20	1848	2.00	15.00	19.00	22.00	34.00	18.7	4.72
2020/21	1813	3.00	17.00	20.00	23.00	36.00	19.9	4.78
2021/22	2008	7.00	17.00	21.00	24.00	38.00	20.6	4.78
Overall	**18988**	**2.00**	**15.00**	**19.00**	**22.00**	**42.00**	**18.9**	**4.61**

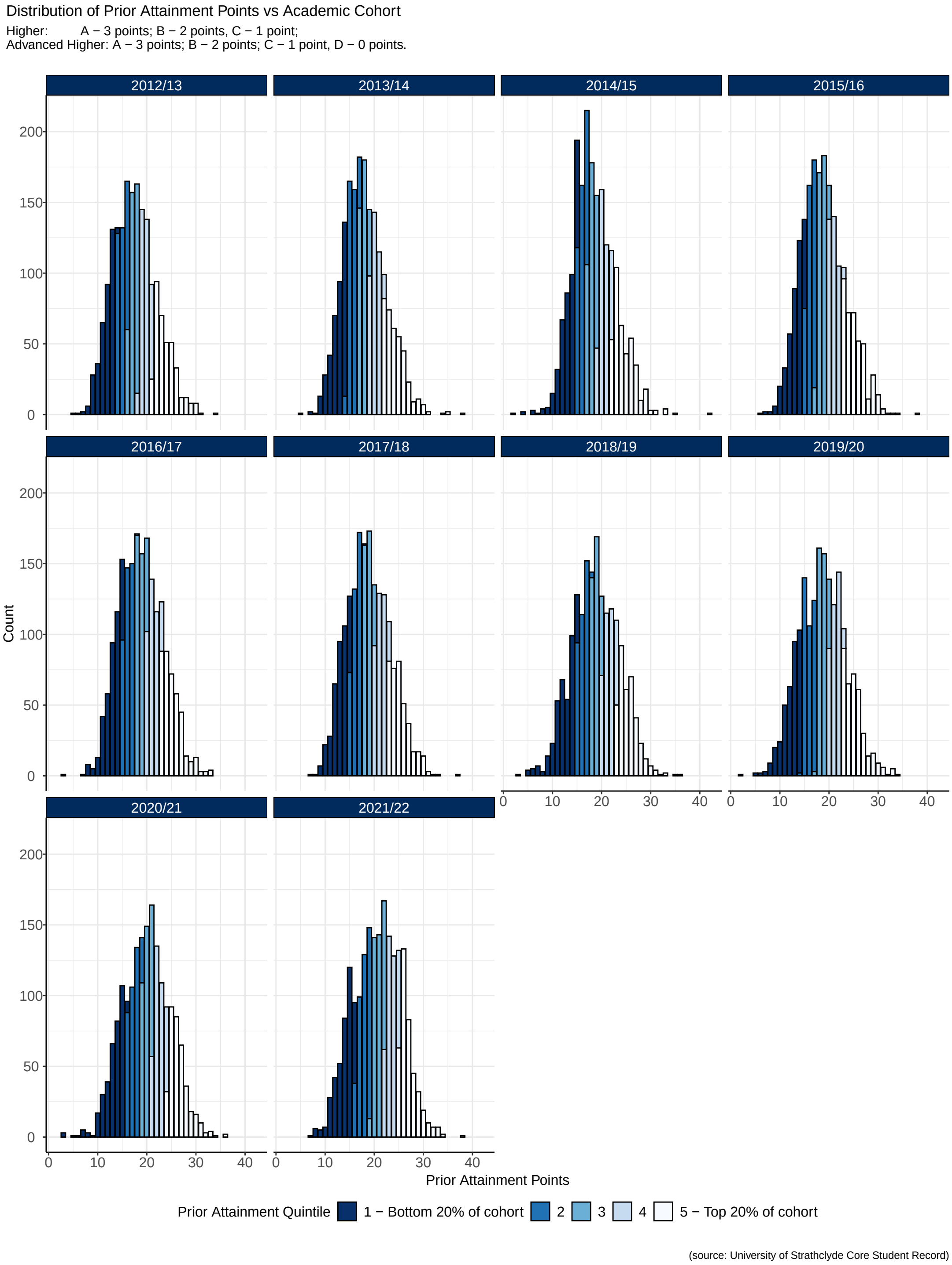


Figure 5.6: Distribution of Prior Attainment Points grouped by Academic Cohort. Prior Attainment Quintiles within each cohort have been highlighted.

5.3 Academic Outcomes

The following interpretations come from Tables D.1 - D.6 in Appendix D. In the **School-leavers dataset**, roughly 90% of students were retained at the end of their first academic session (Table D.1), around 74% achieved completion within four sessions (Table D.3), and approximately 9% of students dropped out of their degree programme (Table D.5). The retention rate per *Academic Cohort* was relatively consistent (Table D.1), although it peaked in 2019/20 and reached its lowest in 2021/22, coinciding with when the no-detriment policy was introduced and repealed (see Sections 2.2 and A.1 for more information on the no-detriment policy). The completion rate appears to have increased slightly over time (Table D.3). Drop-out rates appear relatively unchanged between *Academic Cohorts* (Table D.5). Students from the Faculty of Business had the highest retention and completion rates, as well as the lowest drop-out rate, of all the four faculties. This was followed by Engineering, HaSS, and then Science which had the lowest rates of success of each of the faculties. Students from higher *SIMD Quintiles* and *Prior Attainment Quintiles* appear to have higher rates of retention and completion, as well as lower rates of drop-outs compared to those from lower quintiles. Females appear to have higher completion rates than males (Table D.3), and similarly for those who did not disclose a disability versus those who did.

5.4 Other Associations Between Variables

Faculty and *Sex* appeared to be strongly associated and the proportion of males to females within each faculty remained more or less consistent over time (Figure 5.7). There was a male-to-female ratio of roughly 55:45 in Science, 45:55 in Business, 75:25 in Engineering and 25:75 in Humanities and Social Sciences (HaSS). Ethnic-minorities were more represented in Engineering and Science (between 7% and 20% depending on the academic cohort) than in Business and HaSS (less than 5% per academic cohort - Figure D.1). There did not appear to be an association between *Faculty* and *Disability Status* (Figure D.2).

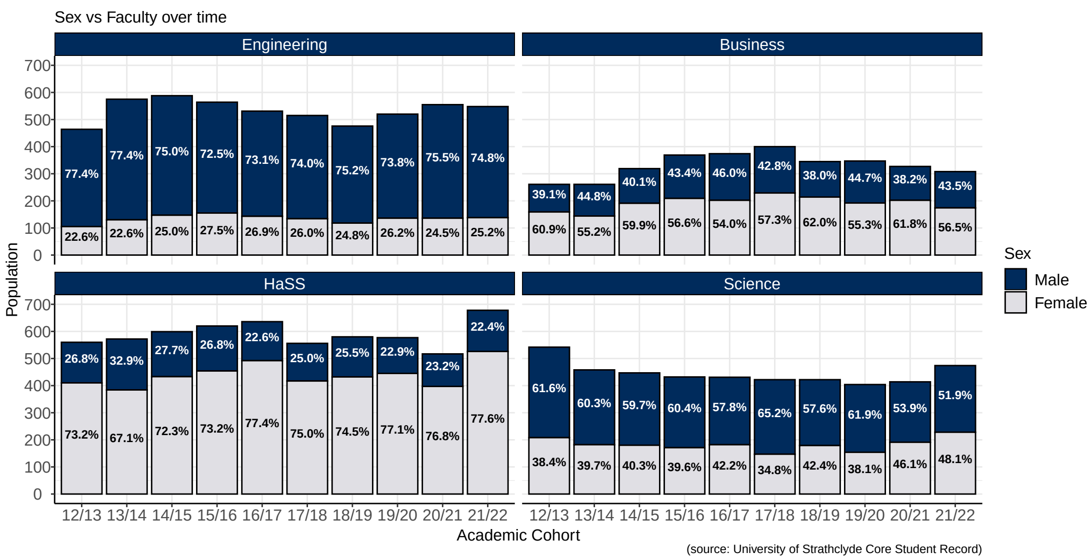


Figure 5.7: Count/proportion of male/female school-leavers who entered each faculty from 2012/13 to 2021/22.

Mosaic plots measure the association between two categorical variables by comparing the proportion of students within the crossed-levels of each variable. From the mosaic plots for *SIMD Quintile* versus the other explanatory variables (Figures 5.8 and 5.9), there appears to be a strong association between *SIMD Quintile* and *Ethnicity*, since ethnic-minorities were more represented in *SIMD Quintiles* 1 and 2 than Quintiles 3, 4 and 5. *SIMD Quintile* also appeared to have some association with *Faculty*; Science and HaSS had larger proportions of *SIMD Quintile* 1 students than Engineering and Business, with the inverse being true for *SIMD Quintile* 5. Females appeared to be marginally more represented in the lower *SIMD Quintiles*, while *Disability Status* did not look to be associated with *SIMD Quintile*. Mosaic plots were also created for the association between *Faculty* and the other explanatory variables (Figures D.4 and D.5), and similarly for *Prior Attainment Quintile* (Figures D.6 and D.7), however no additional noteworthy associations were found.

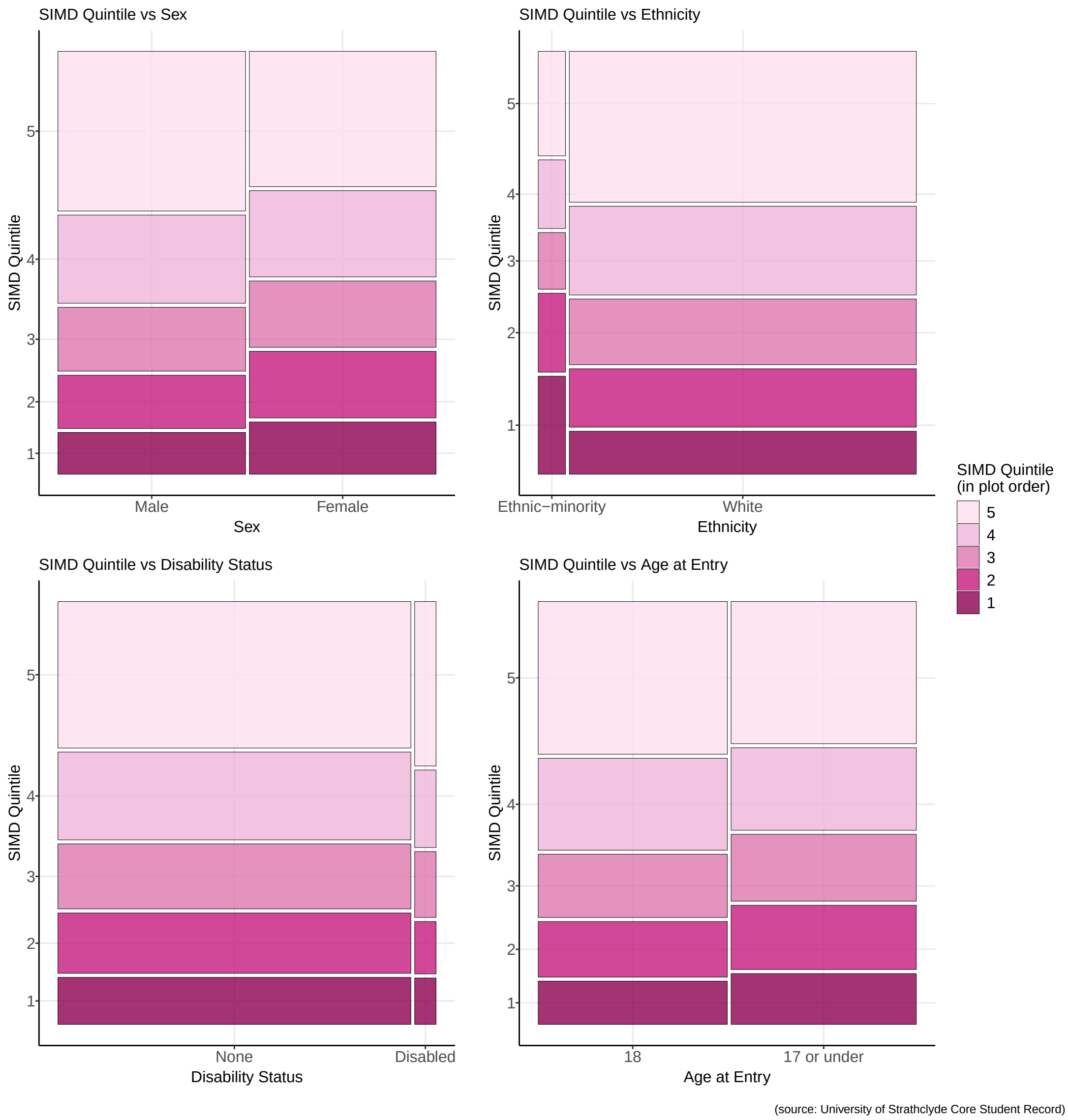


Figure 5.8: Mosaic plots showing proportion of school-leavers within each strata: SIMD Quintile versus explanatory variables (1 of 2).

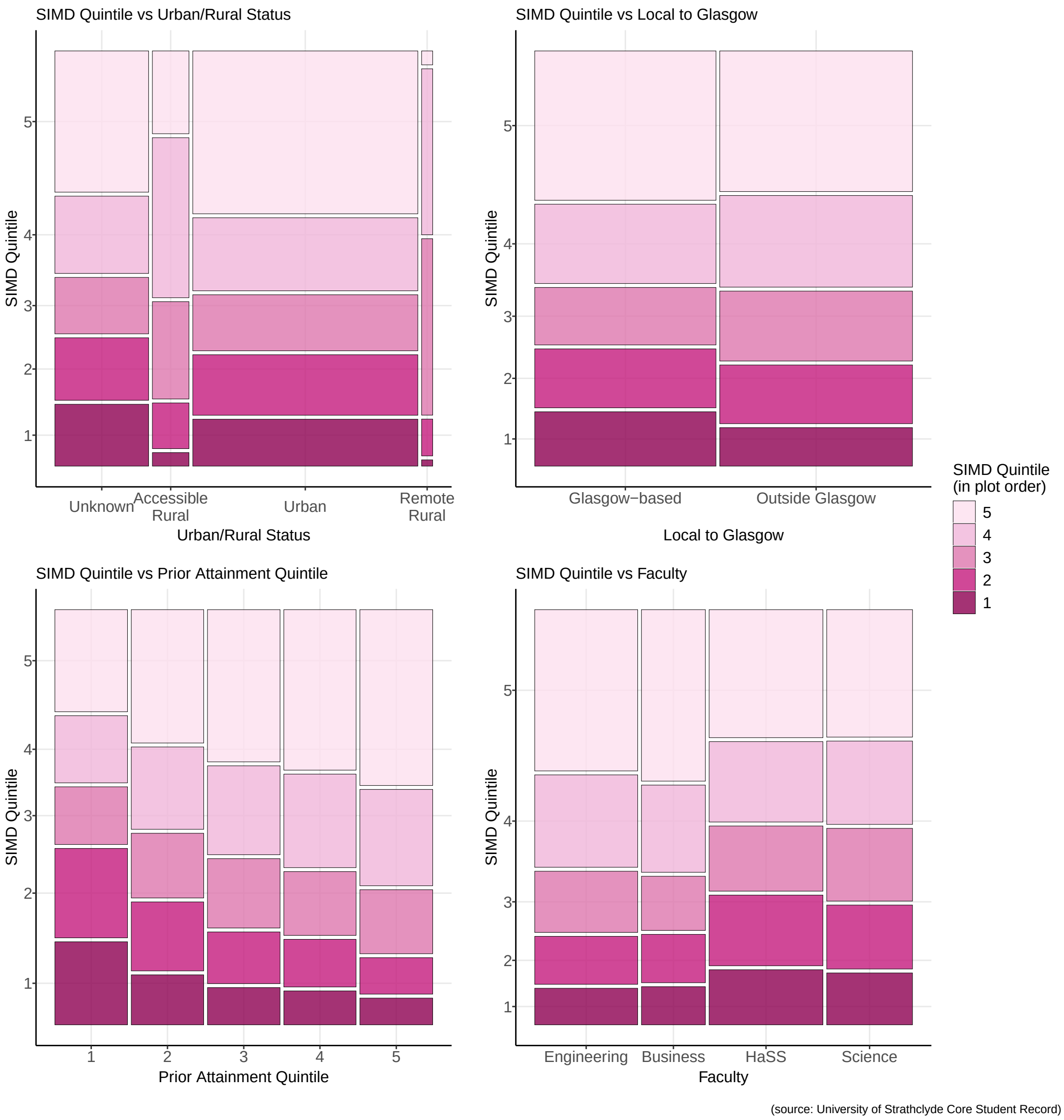


Figure 5.9: Mosaic plots showing proportion of school-leavers within each strata: SIMD Quintile versus explanatory variables (2 of 2).

5.4.1 Interaction Plots

Retention rates follow a clear ordinal pattern with respect to *Prior Attainment Quintile* that remains relatively consistent over time (Figure 5.10). The completion rates of *Prior Attainment Quintiles* 2-5 appear to increase with each successive academic cohort, whereas the completion rates within *Prior Attainment Quintile* 1 remain flat over time (Figure 5.11). The retention and completion rates across *SIMD Quintiles* appear to change across *Academic Cohorts*. For example, the retention gap between *SIMD Quintiles* is very small between 2012/13 - 2016/17 (Figure 5.10), but then an ordinal pattern begins to appear from 2017/18 onwards that persists through to 2021/22. The beginning of this ordinal pattern coincides with the publication of the "Blueprint for Fairness" in 2016 and its recommendation that contextual offers be taken up across all Scottish universities [3]. The completion rates of *SIMD Quintile* 1 students do not appear to change over time and are consistently the lowest of all the quintiles (Figure 5.11). In contrast, the completion rates for *SIMD Quintile* 5 students increase over time and are consistently the highest of all the quintiles. The completion rates of *SIMD Quintiles* 2-4 fluctuate over time. This variation could be due to sampling variability since *SIMD Quintiles* 2-4 have fewer students and lower completion rates than *SIMD Quintile* 5, which has more students and a higher completion rate.

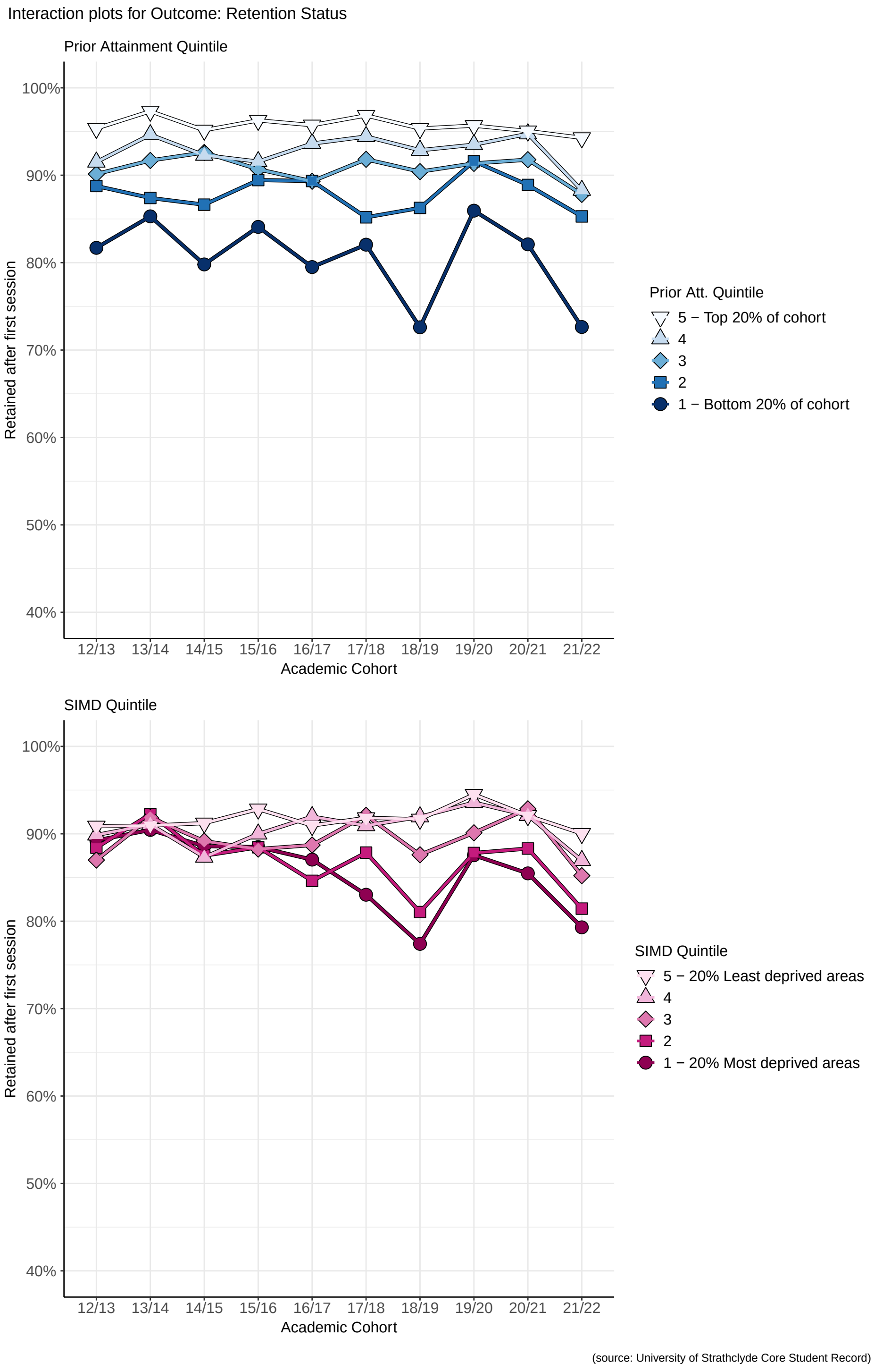


Figure 5.10: Interaction plots showing proportion of successful retentions within each SIMD/Prior Attainment Quintile, per Academic Cohort (1 of 2).

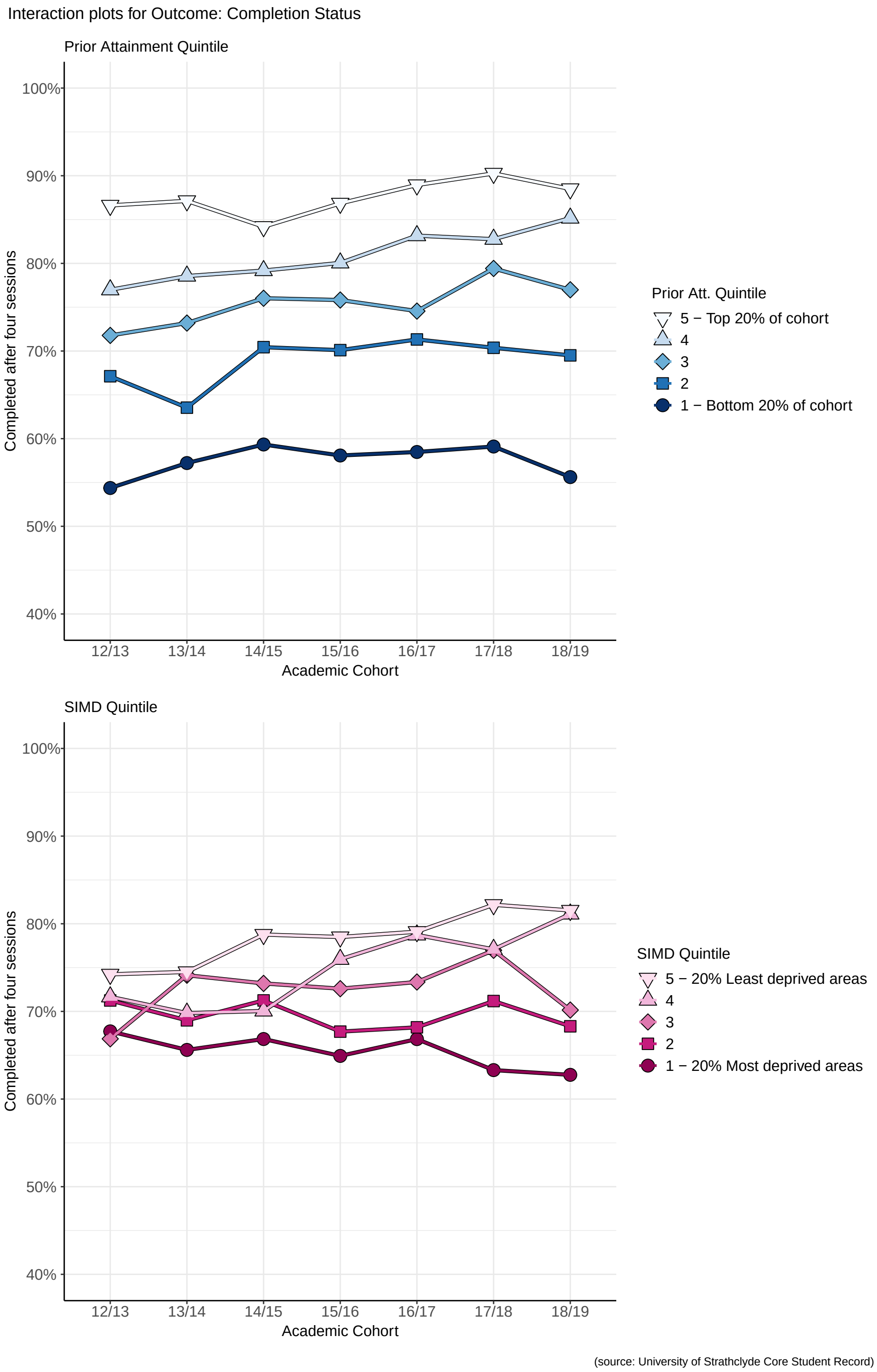


Figure 5.11: Interaction plots showing proportion of successful completions within each SIMD/Prior Attainment Quintile, per Academic Cohort (2 of 2).

5.5 Entry Requirements

The **Prospectus dataset** showed that for entry in 2015/16 the most common entry requirement was three subjects at grade A and one subject in grade B (AAAB) in Scottish Higher qualifications (Figure 5.12). By 2023/24, the most common entry requirement was AAAA in Higher qualifications. In total, there were 138 programmes which changed their entry requirements; 53.0% of these changes occurred in academic intake session 2017/18 while 30.6% occurred in 2021/22. A total of 122 programmes increased their entry requirements (116 by one grade, 6 by two grades) while only 12 decreased them (5 by one grade, 5 by two grades, 2 by three grades). There were 9 programmes that changed their entry requirements on multiple occasions.

Students from *Academic Cohorts* 2012/13 - 2014/15 had no entry requirements information since these data could not be recorded (see Section 3.4). There were 49 programmes in the **School-leavers dataset** that had missing entry requirements information. This meant that of the 13, 342 students from *Academic Cohorts* 2015/16 - 2021/22, there were 338 (2.54%) students of who had missing entry requirements data; 225 students from the Faculty of HaSS, 77 from Science, 36 from Business, and 0 from Engineering. This is unsurprising given that the **Prospectus dataset** did not contain an exhaustive list of all degree programmes offered by the University.

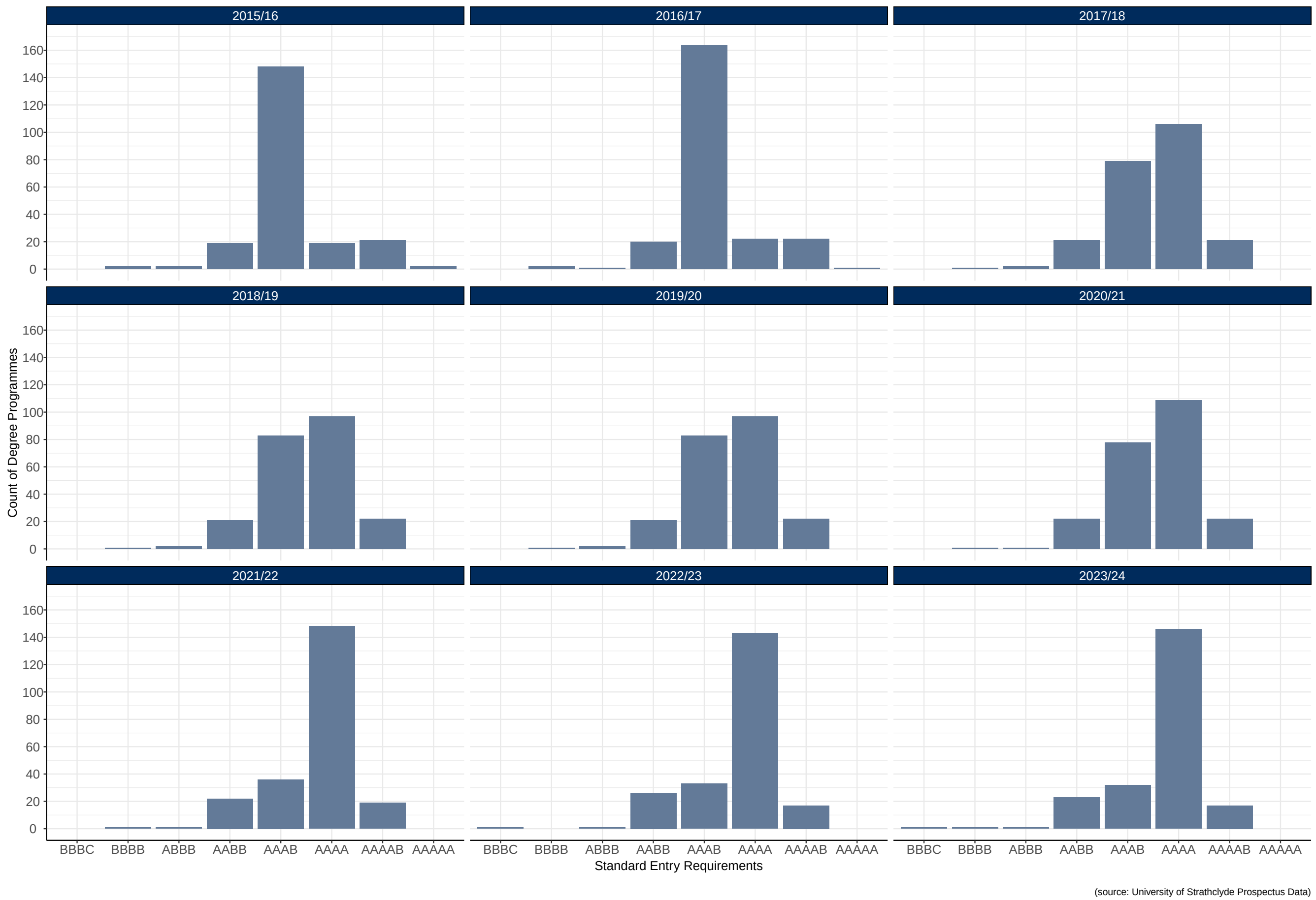


Figure 5.12: Bar-charts of standard entry requirements grouped by academic intake session. Requirements are for students studying Scottish Highers seeking entry to full-time, undergraduate degree programmes. Data was taken from University of Strathclyde prospectuses 2015/16 - 2023/24.

5.6 Summary

In addition to *SIMD Quintile* and *Prior Attainment Points/Prior Attainment Quintile*, it may be appropriate to consider the following variables in multivariable model fits: *Academic Cohort*, *Sex*, *Faculty*, *Ethnicity*, *Disability Status*, *Urban/Rural Status*, *Local to Glasgow*. Care should be taken with the *Disability Status* and *Ethnicity* variables given the low number of observations in the "Disabled" and "Ethnic-minority" categories. The analyses in Chapters 8 and 9 use the *Ethnicity* variable. While *Disability Status* is not considered in any of the multivariable models of this thesis, it is recommended that this variable be investigated in future analyses. *Urban/Rural Status* and *Local to Glasgow* may be appropriate to investigate in future analyses given they potentially had some association with the academic outcomes. These variable were not of primary concern for this thesis however, and were dropped from consideration. The registration status variables (*Changed Prog. Title*, *Changed Dept.*, *Changed Faculty*, *Repeated Stage*, *Break*) were not investigated in-depth here. This was because their inclusion to any model would complicate interpretations. Furthermore, these variables are best interpreted as time-dependent covariates which are not considered within this thesis, but are recommended for future analyses (see 10.2).

Chapter 6

Methods

This Chapter explains the mathematical theory behind the methods applied in Chapters 7, 8, and 9. It will focus on regression methods which model binary outcomes and using survival methods to model the time until an event occurs. This chapter borrows equations and notation from Dobson and Barnett 2018 [99], Moore 2016 [100], James et al. 2021 [101], Vittinghof et al. 2012 [102], Singer and Willett 2003 [103], Tutz and Schmid 2016 [104], and Therneau et al. 1990 [105], and re-contextualises these in terms of students and their academic outcomes. Uppercase letters represent random variables (e.g. Y, X) and lowercase letters represent observed values (e.g. $y_1, y_2, \ldots, y_i$ where $i = 1, 2, 3, \ldots$ represents the i-th observation). Variables denoted in **bold** represent vectors when in lower case (e.g. $\boldsymbol{y}$), and matrices when in upper case (e.g. $\boldsymbol{X}$). The vector $\boldsymbol{y}$ can either represent a vector of random variables ($\boldsymbol{y}_i = [Y_1, Y_2, \ldots, Y_i]^T$, where T denotes the transpose) or a vector of observed values ($\boldsymbol{y}_i = [y_1, y_2, \ldots, y_i]^T$), which will be made clear from context.

6.1 Modelling Binary Outcomes

For many of the problems considered in this thesis, the outcome of interest is a binary response variable. For example, whether or not a student was retained after their first registration session *(Retention Status)* or whether or not a student completed their Bachelor's with Honours degree within four years *(Completion Status)*. The definitions for these outcomes can be seen in Section 4.2. Traditional linear regression, which is a member of the family of generalised linear models, is not sufficient for binary outcomes (for reasons explained in later sections). Instead, other generalised linear regression models are required, three of which (Logistic, Log-binomial, Modified Poisson) will be explained within this chapter.

Binary outcomes can be represented by the random variable Y, where $Y = 1$ for observations that achieve the outcome (i.e. success), and $Y = 0$ for observations that do not achieve the outcome (i.e. failure). Let the probability of a successful outcome be denoted by $\pi_x = \mathrm{P}(Y = 1|X)$ for a given predictor variable, X. To illustrate an example, consider the outcome that is completion of a Bachelor's with honours degree, where $Y = 1$ if the student is successful and $Y = 0$ if unsuccessful. Let X be a predictor categorical variable where $X = 1$ represents females and $X = 0$ represents males. If there were 80 females, and 60 completed their degree, then the probability of completion for females would be estimated as $\pi_1 = 60/80 = 0.75$. Likewise, if there were 120 males and 60 completed their degree, then the probability of completion for males would be estimated as $\pi_0 = 60/120 = 0.5$.

The effect of any predictor, X, on the probability of success, π_x, can be expressed in one of two ways. The more intuitive approach, is to express this in terms of risk-ratios (RRs), also known as relative-risks in some literature. This can be expressed as (Equation 6.1)

$$\mathrm{RR} = \frac{\mathrm{P}(Y = 1|X = 1)}{P(Y = 1|X = 0)} = \frac{\pi_1}{\pi_0} \tag{6.1}$$

where π_1 is the probability of experiencing the outcome for individuals who are exposed to the characteristic, $X = 1$, over the probability π_0 of experiencing the outcome for those who were not exposed to the characteristic, $X = 0$. Using the same example as before, the risk-ratio associated with females (the exposure) achieving the outcome (completion) would be $\pi_1/\pi_0 = 0.75/0.5 = 1.5$. In other words, females would be 1.5 times as likely as males to complete their degree.

An alternative to risk-ratios are odds-ratios (ORs) which can be expressed as (Equation 6.2)

$$\mathrm{OR} = \frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)} \tag{6.2}$$

where π_1 and π_0 are defined as previously. Odds ratios are harder to interpret than risk-ratios but are common in literature (for example, [76, 78]) due to the popularity of the logistic regression model (Section 6.4) that naturally derives odds-ratios. Care should be taken when interpreting odds-ratios since language such as “risk” and “more likely” are associated with describing probabilities and risk-ratios. In fact, odds-ratios are frequently misinterpreted as risk-ratios by researchers and the lay audience. This is problematic since “the odds-ratio will always exaggerate the size of the effect when compared to the relative risk" [108]. Estimation is only appropriate when the prevalence of an outcome is less than 10% such that the error between odds- and risk-ratios is small [106–111].

Using the previous example again, the odds-ratio associated with females (the exposure) achieving the outcome (completion) would be

$$\frac{\pi_1/(1-\pi_1)}{\pi_0/(1-\pi_0)} = \frac{0.75/(0.25)}{0.5/(0.5)} = 3$$

In other words, females would have 3 times the odds of males to complete their degree. If odds-ratios were used to estimate risk-ratios here, they would over-exaggerate the effect of being female on completion by a factor of 2.

6.2 Linear Regression

To consider the effect of many covariates on the outcome, regression techniques can be applied. The most simple regression technique is linear regression (Equation 6.3 [99]).

$$\mathrm{E}(Y_i) = \mu_i = \boldsymbol{x}_i^T\boldsymbol{\beta}; \qquad Y_i \sim \mathrm{N}(\mu_i, \sigma^2) \tag{6.3}$$

where $\mathrm{E}(Y_i) = \mu_i$ is the expected value. Given a set of observations, $\boldsymbol{y} = [y_1, y_2, \ldots, y_i]^T$, Equation 6.3 becomes Equation 6.4.

$$y_i = \beta_0 + \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{\epsilon_i} \tag{6.4}$$

where y_i are the observed values which are assumed to be independent, continuous and normally distributed, $\mathrm{N}(\mu_i, \sigma^2)$, β_0 is the intercept term, the vector of covariates is represented by $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]^T$ where p is the number of covariates, and $\boldsymbol{\epsilon}_i = [\epsilon_1, \epsilon_2, \ldots, \epsilon_i]^T$ is the vector of random error terms that are assumed to be independently and identically distributed, as well as normally distributed.

One cannot model binary outcomes using the linear regression model for several reasons. The first is that, trivially, binary outcomes are not normally distributed [102]. The second is that response for linear regression, $\boldsymbol{y}$ may predict probabilities that fall outside of the $[0, 1]$ interval. The third reason, is that for continuous predictors, it may not be the case that the risk associated with the binary outcome is linear [102]. Instead, alternative methods from the family of generalised linear models must be considered.

6.3 Generalised Linear Models

Linear regression is a member of the family generalised linear models (GLMs). All members of this family can be expressed as follows (Equation 6.5)

$$f(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta} \tag{6.5}$$

where $f(\mu_i)$ is known as the "link function". Linear regression is the special case of Equation 6.5 where the link function is the identity function, $f(\mu_i) = \mu_i$ [99]. The link function within GLMs allow for the modelling of a variety of outcomes. Three models are considered: Logistic, Log-binomial, and Poisson regression; each of which have their own link function that allow for the modelling of binary outcomes. Each model has its own assumptions that are required to be satisfied. However, all GLMs share the assumption that continuous covariates are linearly associated with the (transformed) outcome. This assumption is always satisfied for categorical variables, since they are modelled as binary covariates which will always be linear. The linearity assumption can be assessed by examining a plot of the predicted probabilities versus the values of the continuous covariate.

6.4 Logistic Regression

Logistic regression is one of the most popular classification techniques for modelling a categorical outcome (Equation 6.6 [99]).

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \boldsymbol{x}_i^T\boldsymbol{\beta} \tag{6.6}$$

where the associated link function is the log-logistic, or "logit", transformation $f(\pi_i) = \log(\pi_i/[1-\pi_i])$. Note that the term for the errors, $\boldsymbol{\epsilon}$, is deliberately omitted for reasons that are examined in Section 6.4.3. The intercept term, β_0, is interpreted as the log odds-ratio for a successful outcome when all p covariates for the i-th observation, $\boldsymbol{x}_i^T = [x_1, x_2, \ldots, x_p]^T$, are held constant (i.e. are equal to zero). The coefficients, $\boldsymbol{\beta}$, are estimated using maximum likelihood estimation (see Section 6.6). Each coefficient, $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]^T$, is interpreted as the change in log odds associated with a unit increase in the p-th covariate, when all other $p-1$ covariates are held constant (i.e. are equal to zero or equal to their mean values).

6.4.1 Example interpretation

Consider the following Logistic regression fit (Equation 6.7)

$$\text{logit}(\pi_i) = \left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{6.7}$$

where the outcome, $\text{logit}(\pi_i)$, describes the log odds for completion of a Bachelor's with Honours degree, with a continuous covariate for Age (in years), x_1, and a categorical covariate for Sex, $x_2 = 0$ if the i-th observation is male and $x_2 = 1$ if female. Then, the β_1 coefficient represents the change in the log odds-ratio that is associated with every unit increase in age (one year) when all other covariates are held constant. Similarly, β_2 is the change in log odds-ratio associated with being female ($x_2 = 1$), when all observations are held constant. Finally, β_0 is the log odds of successful completion of a degree when all covariates are held at the value of zero. Notice that the interpretation of age ($x_1 = 0$) does not make sense since no one registered at university can be aged 0. For this reason, it is common to centre continuous variables at their mean values to improve their interpretation at the intercept [102]. This does not affect the point estimate of the other covariates in the model [102]. Both the intercept and coefficients can be exponentiated to obtain odds-ratios which are easier to interpret than log odds-ratios.

6.4.2 Assumptions

In Logistic regression it is assumed that the covariates are linearly associated with the logit of the outcome. This can be assessed visually through plotting the data points for each continuous predictor against the fitted values. A smoothed estimate to the data points can be compared to the residuals of the fitted regression model to determine whether the assumption is satisfied [102]. This could be followed up by fitting a model with polynomial terms and determining their significance in a Wald's test [102]. If the linear term is significant but the higher order terms are non-significant, then this suggests that the linear term adequately fits the model.

It is also assumed that the observed outcomes, y_i, follow a Binomial distribution, that the outcomes for each observation, y_i, are independent from one another, and that the continuous covariates are not collinear with one another (multicollinearity assumption). The first two assumptions should be trivial to assess prior to fitting the model to the data. Multicollinearity can be assessed through robust exploration of variables prior to model fitting.

6.4.3 Error Terms/Limitations

Note that Equation 6.6 omits any error terms, ϵ_i. These are in fact present, but are omitted by most textbooks and published research [112]. This is because the standard error terms cannot be explicitly derived in Logistic regression since the logit transform of the observed outcome can only equal 0 or 1 [112]. To address this, one can scale the standard error and log-odds terms through division by σ, where σ is the standard deviation of the error term. Norton et al. [112] warned that this has several consequences. The first is that the standard error is sensitive to changes in the model specification, i.e., which explanatory variables are included in the model fit, even if these variables are independent of one another. This means that estimated odds-ratios are not unique, since they are dependent upon the chosen model specification. Consequently, it is not possible to compare odds-ratio between datasets, samples of the same dataset, or even between different models fit to the same data. It is also possible to add/remove variables to determine the robustness of a model fit. This, combined with the knowledge that odds-ratios are poor estimates of risk-ratios when the prevalence of the outcome is greater than 10% (see Section 6.1), has led many to recommend models that derive risk-ratios over odds-ratios [106, 107, 111–113].

6.5 Risk Ratio Regression Models

The Log-binomial and Poisson regression models are alternatives to the odds-based Logistic regression model in that they derive risk-ratios. Namely, The Log-binomial model (or log-linear model) can be expressed as (Equation 6.8 [99])

$$\log(\mu_i) = \beta_0 + \boldsymbol{x}_i^T \boldsymbol{\beta} \tag{6.8}$$

Where the link function $f(\mu_i) = \log(\mu_i)$. The coefficients, β_p, are estimated using the maximum likelihood method (see Section 6.6) and are interpreted as the risk-ratios associated with a unit increase in the covariates, $\boldsymbol{x}_i^T$. The assumptions of the Log-binomial model are more or less identical to those laid out for Logistic regression. One exception is that all covariates are assumed to be linearly associated with the log of the outcome, rather than the logit of the outcome. Hence, all assumptions are assessed in a similar fashion to Logistic regression (see Section 6.4.2).

It is not uncommon for the Log-binomial model to run into numerical errors during fitting. This is related to the the log-link function which constrains all linear predictors to be negative [114]. While there are approaches to dealing with convergence issues (see, for example, Williamson et al. [114]) an alternative risk model can be considered, the Poisson regression model (Equation 6.9 [99]).

$$\log \mathrm{E}(Y_i) = \log(\mu_i) = \log(n_i) + \boldsymbol{x}_i^T \boldsymbol{\beta}; \qquad Y_i \sim \mathrm{Po}(\mu_i) \tag{6.9}$$

where n_i is the number of exposed observations and, like the Log-binomial model, the link function $f(\mu_i) = \log(\mu_i)$. Unlike Logistic and Log-binomial regression, the outcome variable in Poisson regression is assumed to be count data, $Y_i \sim \mathrm{Po}(\mu_i)$. Hence, it is assumed that the mean equals the variance, $\mathrm{E}(Y_i) = \mathrm{Var}(Y_i) = \mu_i$ [101]. All covariates are assumed to be linearly associated with the log of the outcome, and are estimated using the method of maximum likelihood estimation (see Section 6.10). Chen et al. [115] found that Log-binomial models was more sensitive than Poisson regression to biased estimates if the model was misspecified. Although they also stated that when correctly specified, Log-binomial models should be preferred since the Poisson model may be more likely to reject the null hypothesis when the alternative is true [115].

Poisson regression can be used to model binary outcomes, where the outcome and the coefficients, β_p, are interpreted as rough estimates of risk-ratios. However, such models suffer from "under-dispersion", where the mean is greater than the variance, which is a violation of the Poisson assumption. This results in overly conservative standard errors [116], which means that true significant effects may be modelled as non-significant. This can be counter-acted using a

Modified Poisson regression model with robust variance terms, also known as "sandwich estimation" [117]. Robust error terms can be implemented in `R` using the `sandwich` package [118].

6.6 The Likelihood Function (Regression)

In Logistic, Log-binomial, and Poisson regression models, estimation of the coefficients is achieved through maximising the likelihood function (Equation 6.10).

$$L(\beta_p|y_i) = \prod_{i=1}^{n} f(y_i|\beta_p) \tag{6.10}$$

The log-likelihood function (Equation 6.11) can be maximised instead since it is computationally less expensive to maximise the log-likelihood function.

$$l(\beta_p|y_i) = \sum_{i=1}^{n} \log(f(y_i|\beta_p)) \tag{6.11}$$

For logistic and poisson regression with grouped data, the log-likelihood function is a measure of goodness-of-fit though a deviance test. For ungrouped data, the value of the log-likelihood on its own has no interpretation [102]. However, if the likelihood of one model is compared to a nested model that was trained on the same sample data, it can be used to determine which model is the better fit [102]. The test that determines whether the model with the extra variable(s) fits the data significantly better than the nested model is known as the likelihood ratio test. Under the null hypothesis, the larger model does not improve upon the nested model. The test statistic is calculated as follows (Equation 6.12)

$$LR = 2[\log(L_1) - \log(L_2)] \tag{6.12}$$

Where L_1 is the likelihood derived from the larger of the two models. It can be shown that this follows a chi-square distribution, χ^2, with degrees of freedom equal to the number of additional variables in the larger model, L_1 [102]. The likelihood ratio test is useful in determining whether the addition of extra explanatory variables lead to a significantly better model fit. However, given that the likelihood functions are different between Logistic, Log-binomial, and Poisson regression, these cannot be compared to one another.

6.6.1 Wald's test

The Wald's test is used to assess the significance of an estimated effect, $\hat{\beta}$, in a model. The null hypothesis is $H_0 : \beta = 0$, and the test statistic is calculated as $Z_w = \hat{\beta}/s.e.(\hat{\beta})$, where "s.e." is the standard error of the coefficient [100]. The test statistic, Z_w, is assumed to follow either a standard normal distribution where H_0 is rejected when $Z_w > |z_{\alpha/2}|$, or, if squared Z_w^2, a chi-square distribution where H_0 is rejected when $Z_w > \chi^2_{\alpha,1}$ [100]. Confidence intervals can also be constructed around the estimate, $\hat{\beta} \pm z_{\alpha/2}[s.e(\hat{\beta}])$ [100]. The results from Wald's tests and likelihood-ratio tests are close estimates of one another when the sample size is sufficiently large, though the latter is considered more reliable [100].

6.7 Goodness-of-fit

Dobson and Barnett [99] show that deviance residuals can be explicitly derived from Binomial and Poisson distributions and can therefore be used as estimates for the goodness-of-fit for the logistic, Log-binomial, and Poisson regression models. The derivation of deviance residuals for each of these models will not be shown here, for further details consult Dobson and Barnett [99, p. 87-92]. A plot of these residuals should be symmetrical around the zero line, and outlier observations can be identified.

6.8 Interaction terms

In some cases, the effect of a given variable on achieving a successful academic outcome may be affected by the levels of another variable. For example, in Chapter 5 it was highlighted that females generally had higher retention and completion rates than males. It was also highlighted, that there were apparent differences in the proportion of males and females within each faculty (e.g. more males in Engineering, more females in HaSS). It could be the case that the effect of *Sex* on retention is affected by the *Faculty* the students are registered with. To address this, an interaction term between these two variables can be included in a regression model. In this context, a significant interaction term would indicate that the differences in retention rates between males and females are not the same across faculties. Variables which have interactive effects on another variable's relationship with the outcome are classified into one of the three categories: a moderator, a mediator, or a confounder.

A moderator is a variable that influences the magnitude and direction of an effect another variable has on the outcome variable. In the example highlighted previously, *Sex* would be a moderator for the effect between *Faculty* and retention. Similarly, it could be said that *Faculty* is a moderator for the effect between *Sex* and retention.

A mediator is a variable that explains the process by which another variable and the outcome are related. An example of a mediator in the **School-leavers dataset** would be *SIMD Quintile*, which partially explains the relationship between *Offer Received* and retention; all contextual offer students must be from SIMD Quintile 1 or 2 (see Section 4.5 for full definition of *Offer Received*). Similarly, *Prior Attainment Points* and *Met Std. Entry Req* would also be partial-mediators of the relationship between *Offer Received* and retention; all contextual offer students must attain below the standard entry requirements in the year prior to their registration.

A confounder is a variable whose presence affects the relationship between another variable and the outcome, when the relationship is not true. A potential confounder in the **School-leavers dataset** is the COVID-19 pandemic, which affected all students registered in 2019/20 and 2020/21 (and potentially in later academic cohorts). Accounting for the effect of COVID-19 pandemic is tricky, since the only proxy for this is *Academic Cohort*. The reality is that this effect cannot be completely controlled for, thus care must be taken when interpreting results which include the affected cohorts.

6.9 Survival Methods

Survival methods are used to model the time until an event occurs. Their advantage comes from being able to account for observations which never experience the event of interest, known as "censored" observations. An example would be cancer trials, where the effect of a particular drug on patients' time-to-death is measured. Here, those patients who do not die before the end of the observation period, or those who go into remission, would be censored. Survival methods can be applied to student registration data to model the the time until a student drops out of their degree programme (event), whilst also accounting for students who never drop out due to completing their degree programme or because they are still registered by the end of the observation period (censored). The remaining sections discusses such applications to the **School-leavers dataset**.

6.9.1 Setting Up the Study

The outcome of interest is the time until a student drops out of university. Let the random variable T denote the time until dropout, or the "survival time", of a student. Let the "observation period" be defined as the fixed period of time: 2012/13 - 2021/22, the academic sessions which are included in the **School-leavers dataset**. Let the "study period" for each student be defined as the period of time each student was registered at the University. The observation period is fixed while the study period is relative to each student, beginning in their first registration session and ending in their last registration session.

6.9.2 Discrete versus Continuous Time

In reality, a student can decide to drop out at any point in time after they register, $T = t$, where $t \in \mathbb{R} > 0$, which could be measured in years, months, days, or even hours. In the **School-leavers dataset**, registration data are collected and updated each academic session, i.e. every year (see Section 4.7). Intervals of time that are larger (and finite), such as years, tend to be described as discrete time intervals, while smaller units of time such as months or days are described as continuous time intervals [103].

In the **School-leavers dataset**, time is measured in academic sessions, or years, which is not a continuous variable. To account for this, the continuous time range must be broken up into $q+1$ discrete intervals $[0, a_1), [a_1, a_2), \ldots, [a_{q-1}, a_q), [a_q, a_\infty)$ where a student who drops out in the interval $[a_{t-1}, a_t)$ is said to have dropped out at time, $T = t$ [104]. For example, a student who drops out in their first academic session, represented by interval $[0, a_1)$, would have survival time $T = 1$. Similarly, a student who drops out in their second academic session, represented by the interval $[a_1, a_2)$, would have survival time $T = 2$, etc. This means that in the **School-leavers dataset**, time until drop-out is strictly-speaking a discrete random variable, $T \in [1, 2, 3, 4, 5, \ldots, q]$.

6.9.3 Censored Observations

Trivially, students will not drop out if they successfully complete their degree or are still continuing their studies by the end of the observation period. To account for these observations, the technique of "censoring" is employed. Let the random variable, C, denote the time at which an observation becomes censored. A subject is said to be "right-censored" if they had not yet experienced the outcome by the end of their study period. A subject is said to be "left-censored" if they were at risk of experiencing the outcome prior to the beginning of their study period. Since no student can be at risk of dropping out prior to beginning their registration, there are no left-censored observations. Observations can also be "interval-censored", though this type of censoring is not considered here (for more on interval censoring, see Moore et al. [100, p. 187]). Since only right-censored observations are present in the **School-leavers dataset**, the remainder of this chapter will refer to them as "censored" observations for brevity.

Survival methods assume that all censoring is "non-informative". If a student is coded as censored when in reality they have dropped out, then their censoring would be known as "informative". Trivially, informative censoring introduces bias to any model estimates. To address informative censoring, one could apply competing-risks models, though these will not be explored in this thesis.

6.9.4 Modelling Survival Time

To account for right-censoring, the observed time for each student becomes (Equation 6.13 - [104, p. 52] and [100, p. 16])

$$Y = \min(T, C) \tag{6.13}$$

where the random variable Y is the time at which a student drops out, T, or is censored C, whichever happens first, and the random variables T and C are assumed to be independent. Let the following indicator variable, δ, be defined as (Equation 6.14 - [104, p. 52] and [100, p. 16])

$$\delta = \begin{cases} 1, & \text{if } T \leq C; \\ 0, & \text{if } T > C \end{cases} \tag{6.14}$$

This indicator function will be useful later when modelling the time until drop-out, T, whilst also accounting for censored observations, C. It will also be used when defining the likelihood function (Section 6.12.1).

6.9.5 End of the study period/Truncated Survival Times

Students can begin their registration at the University on any of the academic sessions within the observation period (2012/13 - 2021/22). However, the maximum amount of time a student can be followed is until the end of the observation period (2021/22). This presents two problems for the modelling of survival times.

The first problem is that students from different *Academic Cohorts* are examined for different lengths of time. While theoretically the survival time, T, for each student could be infinite, it was very rare to see any individual registered for longer than 5 academic sessions. Exploratory analyses conducted on the **School-leavers dataset** (not shown to preserve anonymity of students) revealed that students with unusually long registrations ($>$ 5 years) were actually those who had entered into some form of voluntary or academic suspension. Most of these students became censored anyway, either due to completing their degree, still being registered or still being in suspension by the end of the observation period. Thus, it was decided to truncate students who had survival times $T > 5$. In other words, all 305 students who were registered for 6 sessions or longer were coded as censored at 5 years. This had the advantage of standardising the length of follow-up time but the disadvantage of coding some students who did drop out as censored, albeit only 5 in total. This informative censoring (Section 6.9.3) will therefore introduce a negligible amount of bias to any model fit to the **School-leavers dataset**.

The second problem is that it is not possible for students from *Academic Cohorts* 2018/19 - 2021/22 to have completed their degree by the end of the observation period. This is only possible for direct-entry students who have been removed from the **School-leavers dataset** (see Section 3.7). To address this problem, only students from *Academic Cohorts* 2012/13 to 2018/19 were considered when applying survival models.

6.9.6 Person-level versus Person-Period data

In general, it is important to determine whether the time interval in a dataset is discrete or continuous prior to any analyses, since this will inform the type of model that is applied to the data and which format the data is required to be in. Discrete time intervals require a person-level format, while continuous time intervals require a person-period format (Figure 6.1). If time-dependent covariates (TDCs) are desired, then the data would need to be in a person-period format for both types of models; no TDCs were used in the analyses of these data, but are recommended for future analyses (see the discussion in Section 9.8).

6.9.7 An Example of Survival Time and Censoring

A visual example of the survival time and censoring set-up for the **School-leavers dataset** is illustrated in Figure 6.2. These example students match with those presented in Figure 6.1. Students A and B show the typical cases of censored and dropout observations, respectively. Student C represents those who take longer than expected to complete their degree, having done so at the end of the 5th academic session. Student D drops out despite being registered for 5 sessions. This is perhaps due to repeating many stages of their degree programme and/or entering into academic suspensions before eventually dropping out. Students E and F are examples of the truncation applied to observations; both are censored due to being registered for longer than 5 academic sessions. This is despite Student F eventually dropping out at some point after 5 sessions.

Person-Level Format

Student	Academic Cohort	Max Reg. Year Count	Dropout Status
A	2012/13	4	0
B	2020/21	1	1
C	2014/15	5	0
D	2014/15	5	1
E	2016/17	6	0
F	2012/13	7	1

Person-Period Format

Student	Academic Session	Reg. Year Count	Dropout Status
A	2012/13	1	0
A	2013/14	2	0
A	2014/15	3	0
A	2015/16	4	0
B	2021/22	1	1
C	2014/15	1	0
C	2015/16	2	0
C	2016/17	3	0
C	2017/18	4	0
C	2018/19	5	0
D	2014/15	1	0
D	2015/16	2	0
D	2016/17	3	0
D	2017/18	4	0
D	2018/19	5	1
E	2016/17	1	0
E	2017/18	2	0
E	2018/19	3	0
E	2019/20	4	0
E	2020/21	5	0
E	2021/22	6	0
F	2012/13	1	0
F	2013/14	2	0
F	2014/15	3	0
F	2015/16	4	0
F	2016/17	5	0
F	2017/18	6	0
F	2018/19	7	1

Figure 6.1: Comparison of person-level format (left) to person-period format (right) for school-leavers dataset.

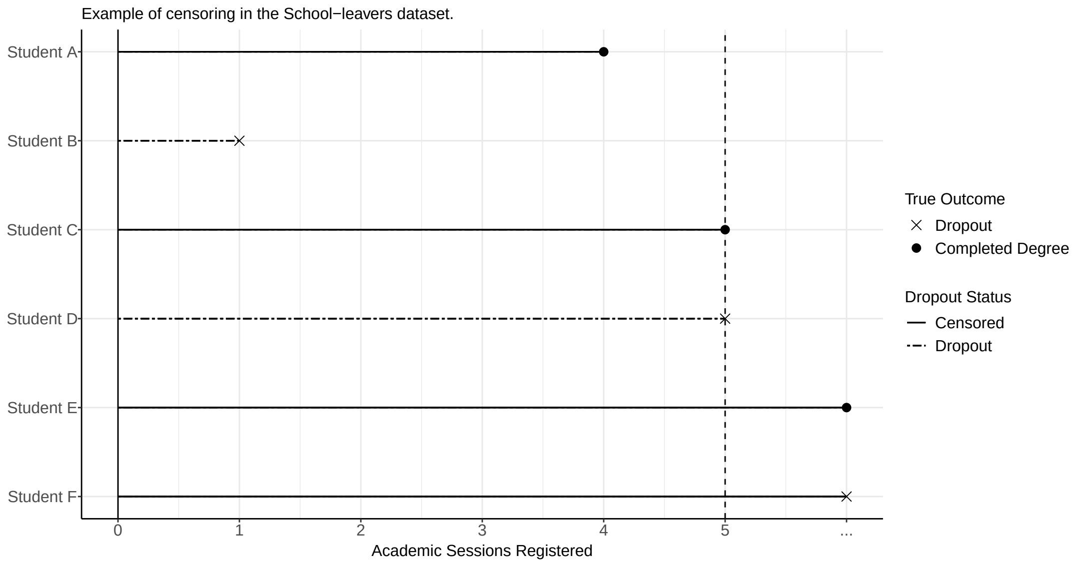


Figure 6.2: Example of censoring setup for the school-leavers dataset.

6.10 The Survival and Hazard Functions

Assume that the time until a student drops out in the **School-leavers dataset**, T, is discrete such that $T \in [1, 2, 3, 4, 5, \ldots, q]$ for q discrete intervals. The hazard function (Equation 6.15 - [104]) can be defined as the risk of dropout at time, t, conditional upon surviving all time periods previous to t

$$h_t = P(T = t | T \geq t) \quad t = 1, 2, \ldots, q \tag{6.15}$$

The cumulative hazard function is simply the sum of the hazards up until time, t

$$H_t = \sum_{k=1}^{t} h_k \tag{6.16}$$

Similarly, the survival function (Equation 6.17 - [104]) denotes the probability of survival until time t, given that a student has not dropped out prior to t.

$$S_t = P(T \geq t) = \prod_{k=1}^{t}(1 - h_k) \tag{6.17}$$

Now assume that T is a continuous random variable such that students can drop out at any point in time, $T = t \in \mathbb{R} > 0$. The hazard function now denotes the risk of drop-out at an instantaneous moment in time, Δt (Equation 6.18)

$$h(t) = \lim_{\Delta t \to 0+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \tag{6.18}$$

The continuous hazard function does not represent a probability, but it does have the condition that $h(t) \geq 0$ [119]. The cumulative hazard function (Equation 6.19), is simply the area under the hazard curve up to time, t.

$$H(t) = \int_0^t h(u)\, du \tag{6.19}$$

Where the interpretation of the cumulative hazard function is the total amount of risk accumulated up until time t [100].

6.11 The Kaplan-Meier Estimate

The Kaplan-Meier Estimate of the survival function is useful tool for inference in both the continuous- and discrete-time frameworks. Order all discrete survival times, $T_1, T_2, \ldots, T_j$, let d_j denote the number of students that dropped out at time T_j and let r_j represent the number of students at risk of dropping out at time, t [99–101, 104]. Then the Kaplan-Meier Estimate can be derived as (Equation 6.21)

$$\hat{S}(t) = \prod_{j:T_j \leq t} \frac{r_j - d_j}{r_j} \tag{6.20}$$

If survival times are continuous, then order these into discrete intervals $[0, T_1), [T_1, T_2), \ldots, [T_j, \infty)$. The Kaplan-Meier estimate becomes

$$\hat{S}(t) = \prod_{j:T_j \leq t} 1 - \frac{1}{r_j} \tag{6.21}$$

where r_j represent the number of students at risk of dropping out at time, t. For the continuous Kaplan-Meier estimate, it is assumed that there are no tied survival times (see 6.13.2), and that all survival times are independent of one another.

A "survival curve" is a plot of the Kaplan-Meier Estimates versus time. The survival curve appears as a smoother function in continuous time than in discrete time. This is because the likelihood of students sharing the same survival times, $T_1, T_2, \ldots, T_j$, is higher under a discrete-time framework.

6.11.1 Log-rank test

Due to the presence of censored data and differing lengths of follow-up, one cannot simply estimate the difference between two groups of students' survival times (e.g. males versus females). Instead, one may apply the Log-rank test, where the null hypothesis is that the difference between the survival curves of two groups is equal to zero.

Order all observed survival times in the data, $t_1 < \cdots < t_k$, assuming there are k drop-outs. To calculate the test statistic, a 2x2 table of counts is required at each time period, t_k, from the each group of students, $i = 1, 2$. This table includes the number of students who were at risk of drop-out, r_{ik}, and the number of students who actually dropped out, d_{ik}, (Table 6.1).

Table 6.1: The number of students at risk of drop-out r_k, and the number of students who actually dropped out d_k, within two groups ($i = 1, 2$) of students.

	Group 1	**Group 2**	**Total**
Dropped out	d_{1k}	d_{2k}	d_k
Survived	$r_{1k} - d_{1k}$	$r_{2k} - d_{2k}$	$r_k - d_k$
Total	r_{1k}	r_{2k}	r_k

James et al. 2021 [101, p. 467-468] then construct the log-rank test statistic, W, as follows (Equation 6.22)

$$W = \frac{X - \mu}{\sqrt{\mathrm{Var}(X)}} \tag{6.22}$$

where $X = \sum_{k=1}^{K} d_{1k}$ is the total number of students who dropped out across all time periods (d_{1k} comes from Table 6.1). The test statistic, W, can be shown to approximately represent the following (Equation 6.23 - [101, p. 467-468])

$$W = \frac{\sum_{j=1}^{K} d_{1k} - \mu_k}{\sqrt{\sum_{k=1}^{K} \mathrm{Var}(d_{1k})}} \tag{6.23}$$

Where the mean at each time period, μ_k, can be calculated as (Equation 6.24)

$$\mu_k = \frac{r_{1k} d_k}{r_k} \tag{6.24}$$

and the variance can be calculated as (Equation 6.25)

$$\mathrm{Var}(d_{1k}) = \frac{d_k (r_{1k}/r_k)(1 - r_{1k}/r_k)(r_k - d_k)}{(r_k - 1)} \tag{6.25}$$

For a sufficiently large sample size, n, the test statistic, W, follows an approximately standard normal distribution [101], or, W can be squared to obtain a chi-square random variable with one degree of freedom [100]. The latter is the approach adopted within the `survival` package in R [120, 121].

In survival analysis, the goal is often to estimate the effect a particular covariate, or group of covariates, have on the hazard function over time. Kaplan-Meier Estimates and Log-rank tests are useful for comparing the hazard of drop-out or survival time between groups of students, for example males versus females. To control for more than one explanatory variable, then multivariable survival modelling techniques are required. This can be achieved through various statistical models which assume that follow-up time is a discrete or continuous random variable. Both will now be examined, starting with the discrete models.

6.12 Discrete Time-to-Event Regression Models

Assume that the time until dropout random variable is discrete, $T \in [1, 2, 3, 4, 5, \ldots, q]$. The Logit Discrete Time-to-Event (DTE) model can used to model the hazard of drop-out over discrete time intervals and is expressed as (Equation 6.26 - [104, p. 37])

$$\log\left(\frac{h(t|\boldsymbol{x})}{1 - h(t|\boldsymbol{x})}\right) = \beta_{0t} + \boldsymbol{x}_i^T\boldsymbol{\beta} \tag{6.26}$$

where $0 < h(t|\boldsymbol{x} < 1$ at each discrete interval and the associated link function is the log-logistic, or "logit", transformation $f(\pi_i) = \log(\pi_i/[1 - \pi_i])$. This is very similar to the traditional logistic regression model see previously (Section 6.4), where now the intercept term, β_{0t}, depends on time while the coefficients, $\boldsymbol{\beta}$, remain fixed. The intercept term, β_{0t}, is referred to as the "baseline hazard function" and represents the hazard of dropping out over time, t, when all p covariates for the i-th student, $\boldsymbol{x}_i^T = [x_1, x_2, \ldots, x_p]^T$, are held constant at zero. The coefficients, $\boldsymbol{\beta_p}$, are interpreted as the change in log odds of dropping out across all time periods, that is associated with a unit increase in the p-th covariate, when all other $p - 1$ covariates are held constant. The change in log odds is assumed to be the same at all time periods, a property known as the "proportional odds" assumption. This assumption can be assessed by examining the plot of the logit survival (or hazard) function over time (Equation 6.26) grouped by each categorical explanatory variable; if the lines for each level of the variable are roughly parallel, then the proportional odds assumption holds.

6.12.1 The Likelihood Function (Survival)

Similar to the traditional logistic regression model (Section 6.6), the method of maximum likelihood estimation can be applied to the Logit DTE model to derive the estimated effects of the covariates, $\boldsymbol{\beta}$. However, the censored observations must also be accounted for when deriving the likelihood function.

Assume that time until drop-out is a discrete random variable, T, and let δ_i be defined as in Equation 6.14. The likelihood function of the survival data can be described as (Equation 6.27 - [104, p. 52-53] and [101, p. 470])

$$L(\beta) = \prod_{i=1}^{n} h(t_i)^{\delta_i} S(t_i) \tag{6.27}$$

where n is the number of observations and $h(t_i)$ and and $S(t_i)$ are the values of the hazard and survival functions, respectively, for the i-th observation. This is often expressed in terms of the log-likelihood (Equation 6.28) since its maximum likelihood estimates are less computationally expensive to derive.

$$l(\beta) = \sum_{i=1}^{n} \delta_i \log h(t_i) + \log S(t_i) \tag{6.28}$$

6.13 The Cox-Proportional Hazards Model

Assume that the time until dropout random variable is continuous $T = t$, where $t \in \mathbb{R} > 0$. The Cox-Proportional Hazards (CPH) model, also known as the Cox Regression model, is a semi-parametric method that can be used to model the hazard of drop-out over continuous time and can be expressed as (Equation 6.29 - [104, p. 50])

$$h(t|\boldsymbol{x}) = h_0(t) \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) \tag{6.29}$$

where $h_0(t)$ is the baseline hazard function and $\boldsymbol{x}_i^T$ and $\boldsymbol{\beta_p}$ are similarly defined as before. The coefficients, $\boldsymbol{\beta}$, are interpreted as the change in hazard of dropping out across all time periods, that is associated with a unit increase in the p-th covariate, when all other $p-1$ covariates are held constant (i.e. are equal to zero). This is known as the proportional hazards assumption and is a key property in how the CPH model functions. This assumption can be assessed by examining the plot of the complementary-log-log (cloglog) transformation of the survival (or hazard) function over time (Equation 6.30) against each explanatory variable; if the lines for each level of the variable are roughly parallel, then the proportional hazards assumption holds.

$$\log[-\log S_1(t)] = \beta + \log[-\log S_0(t)] \tag{6.30}$$

When the risk of experiencing the outcome is sufficiently rare (say $< 10\%$), then cloglog transformations of the survival (or hazard) functions are approximately the equivalent of the logit transformations of the survival (or hazard) functions [103, p. 422]. An alternative approach is to apply a Schoenfeld residuals test of proportional hazards, this will not be explained in this chapter but further reading can be found here [100, p. 96-100].

In the CPH model, no form for the baseline hazard function, $h_0(t)$, is assumed. Hence, the values of the coefficients, $\boldsymbol{\beta}$, must be estimated through maximisation of the partial likelihood function (Equation 6.31).

6.13.1 The Partial-Likelihood Function

In discrete-time, the value of the baseline hazard function, $h_0(t)$, can be explicitly derived for each time period, t. In continuous-time, since there are infinitely many points in time a student could dropout, the baseline hazard function, $h_0(t)$, cannot be explicitly derived. Therefore, it must be assumed to follow some shape, for example the Weibull distribution (see Section 6.14). Another alternative, proposed by Cox in 1972 [122], is to assume no form for the baseline hazard function, and instead maximise the partial-likelihood function.

Assuming that time until dropout is a continuous random variable, $T = t$ where $t \in \mathbb{R} > 0$, and that there are no tied survival times, the partial likelihood function can be expressed as (Equation 6.31 - [104, p. 64] and [101, p. 473])

$$PL(\beta) = \prod_{i=1}^{k} \frac{h_0(t_i)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})}{h_0(t_i)\sum_{j\in r_{t_i}}\exp(\boldsymbol{x}_j^T\boldsymbol{\beta})} \tag{6.31}$$

where $t_1 < \cdots < t_k$ are the ordered survival times that are observed in the data, assuming there are k drop-outs and r_{t_i} for $i = 1, \ldots, k$ are the sets of all students that are at risk of dropping out at time, t_i. Notice that the baseline hazard function, $h_0(t)$ appears on both the numerator and the denominator of Equation 6.31. This means that the partial-likelihood function can be maximised without assuming any form for the baseline hazard function, $h_0(t)$. This only works when the proportional hazards assumption is valid (Equation 6.29), such that the baseline hazard of dropping out across all time periods is the same across all covariates.

6.13.2 Tied Survival Times

Maximisation of the partial-likelihood function for a continuous time-to-event random variable, T, assumes that there are no tied survival times. In other words, if $t_1 < \cdots < t_k$ are the ordered survival times that are observed in the data, then $t_{i-1} \neq t_i$ for all $i = 1, \ldots, k$. Yet as established previously (Section 6.9.2), student dropouts will occur in naturally discrete intervals in the **School-leavers dataset**. Hence, there will be many students who share the same time until dropout, T.

To handle these tied survival times, one can use the method of exact partial likelihood [104, p. 64]. This works by replacing the risk factors within the partial-likelihood function (Equation 6.31) with the conditional risk an observation drops out given the number of students that dropped out at the same time, t_i [104, p. 64]. The exact partial likelihood method is computationally expensive [104, p. 64]. Alternatives include the Breslow or Efron approximations. These are not explained here, but further reading on these can be found here [100, p.65-69] and here [123, p. 563-564]. The Efron approximation is the default method used in the `cox.zph()` function from the `survival` package in R [120, 121], which are used in Chapter 9.

6.13.3 Stratified Cox-Proportional Hazards

Stratification is an alternative option for fitting a CPH model when certain variables violate the proportional-hazards assumption [102]. A different baseline hazard function is assumed within each level of the stratification variable [102**?**], such that (Equation 6.32)

$$h_j(t|\boldsymbol{x}) = h_{0j}(t)\exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) \tag{6.32}$$

Where $h_j(t|\boldsymbol{x})$ is the hazard function within the j-th stratum. It is assumed that the effect of the covariates, $\boldsymbol{\beta}$ is the same across each stratum. The likelihood function then becomes the product of the likelihoods derived within each stratum (Equation 6.33 - [102, p. 228] and [**?**])

$$L(\beta) = \prod_j L_j(\beta) \tag{6.33}$$

And the log-likelihood is the summed contributions within each stratum [?]. Stratifying by one or more variables means that the effect of the stratified variables on the outcome cannot be estimated. Stratification also requires a sufficient sample size within each strata, otherwise the statistical power may be inadequate [102].

6.14 Parametric Survival Models

An alternative to CPH models are parametric survival models which assume that the form of the baseline hazard function follows some probability distribution. While this could be any number of different distributions, this thesis focuses on only one, the Weibull distribution. Assuming that the baseline hazard function, $h_0(t)$ is specified by a Weibull distribution, the survival function can be expressed as (Equation 6.34 - [100, p. 138])

$$S(t) = \exp(-\exp(-\mu/\sigma)t^{1/\sigma}) \tag{6.34}$$

with hazard function (Equation 6.35)

$$h(t) = \frac{1}{\sigma}\exp(\frac{-\mu}{\sigma})t^{\frac{1}{\sigma}-1} \tag{6.35}$$

here $\sigma = 1/\alpha$ and $\mu = -\log(\lambda)$, for some $\alpha \in \mathbb{R}$. The exponential distribution is the special case of the Weibull distribution where the scale parameter $\sigma = 1$. The effects of covariates in parametric models are estimated by maximising the value of the likelihood function (Equation 6.28).

The Weibull model is an example of an "Accelerated Failure Time" (AFT) model, where the covariates estimate the effect on survival time, S, rather than time until drop-out, T. Let $\gamma_j = (\gamma_1, \ldots, \gamma_p)^T$ denote the set of unknown coefficients associated with the set of explanatory variables, $x_{ij} = (x_{i1}, \ldots, x_{ip})^T$, for the i-th student. The interpretation of the estimates of AFT models, γ_j, is that a unit change in the j-th covariate is associated with change in the survival time of a student by a factor of γ_j. Values of $\exp(\gamma) > 1$ are associated with longer survival times and values of $\exp(\gamma) < 1$ are associated with shorter survival times. The Weibull model is unique in that it is the only AFT model which is the equivalent of a proportional hazards model where $\exp(\beta) = \exp(-\gamma/\sigma)$ [100, p. 146].

6.15 Martingale residuals

The martingale residuals determine the functional form of a continuous covariate within the fit of a particular survival model. Assuming only right-censored observations (which is the case for our student dropout problem) and that there are no time-dependent covariates, then these can be expressed as follows (Equation 6.36 - [100, p. 87])

$$m_i = \delta_i - \hat{H}_0(t_i) \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) \tag{6.36}$$

Where $\hat{H}_0(t_i)$ is the estimate of the cumulative baseline hazard function at time, $t = t_i$, $m_i \in (-\infty, 1]$ and $\sum_{i=1}^{n} m_i = 0$. The interpretation of the martingale residuals is the difference between the observed value of the censoring indicator, δ_i, and the expected value of this indicator, $E(\delta_i) = \hat{H}_0(t_i) \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})$, assuming proportional-hazards. For example, if $m_i > 0$ this indicates that the individual, i, survived for longer than expected, while $m_i < 0$ implies the individual survived for a shorter period than expected.

6.16 Deviance residuals

The deviance residuals in survival analyses are a transformation of the martingale residuals that gauge a model's goodness-of-fit to the data (Equation 6.37 - [100])

$$d_i = \text{sign}(m_i) - 2[m_i + \delta_i \log(\delta_i - m_i)]^{1/2} \tag{6.37}$$

Where sign(m_i) is the signed square root of the martingale residual, m_i. The advantage of deviance residuals are that they make the martingale residuals, $m_i \in (-\infty, 1]$, symmetrical around zero, though this is assuming that the model is a good fit for the data. Therneau et al. [105] note that when censoring is greater than 40% (as is often the case in the student dropout problem if successful outcomes are censored), "...a large bolus with residuals near 0 distorts the normal approximation, but the transform is still helpful in symmetrizing the set of residuals".

6.17 Summary

In Chapters 7 and 8, the Logistic, Modified Poisson, Log-Binomial, will be applied to the **School-leavers dataset** to determine the associations between the two binary outcomes: retention and completion, and students' contextual background and prior attainment profiles. Regression methods are commonly used for measuring such associations in Widening Access literature [52, 69, 76–78]. Comparing the estimates between these three models will aid researchers in Widening Access understand the interpretation of regression methods in these scenarios. However, given that the prevalence of the retention and completion outcomes are so common (90% and 70%, respectively - see Section 5.3), it is anticipated that some of the issues highlighted in this chapter will be encountered. For example, the exaggeration of the logistic estimates (Section 6.4.3) and the convergence of the Log-Binomial model (Section 6.5). Chapter 9 will then attempt to address some of these limitations by instead applying survival methods to model the time until a student drops out of the University, which should be sufficiently rare ($< 10\%$).

Chapter 7

Influence of Advanced Higher Mathematics on Outcomes for Students Registered on STEM Programmes

In Chapter 6, three regression methods were introduced from the family of generalised linear models that could be applied where there is a binary outcome variable: Logistic, Log-Binomial and the Modified Poisson models. In this chapter, each of these regression methods will be compared on their fit to the data for the binary outcomes of retention and completion. It will also answer research questions on the effect that Advanced Higher Mathematics qualifications have on academic outcomes at the University. Before proceeding however, some background information on the topic of Advanced Highers is required.

Parts of the analysis contained within this chapter were submitted as a journal article to *MSOR Connections* and is currently under peer-review. This chapter contains some additional detail that was not included in the original submission, such as the estimates and fit of the Logistic and Log-Binomial models, the deviance and linearity plots for all models, and an expanded background and discussion.

7.1 Background

To gain access to higher education degree programmes, Scottish domiciled students must obtain the relevant grades in Scottish Higher qualifications typically achieved in either of the final two years of secondary school (S5 and S6). However, for S6 students who have already obtained a Higher in a given subject but wish to study further, the opportunity exists to sit a more advanced qualification, the Advanced Higher.

Advanced Highers were introduced in 1999 as a replacement for Certificate of Sixth Year Studies [124]. Students may take Advanced Highers in a range of subjects, for example, English, Mathematics, Statistics, Physics, Chemistry, Biology, etc. [125]. Typically, students are expected to have at least passed the relevant Higher as a pre-requisite to sitting an Advanced Higher, though ultimately presentation for the exam is at the discretion of the school. Advanced Higher learners are encouraged to be more “pro-active” and “independent” in their studies to bridge the gap between secondary and higher education [14]. Advanced Highers rank as Level 7 on the Scottish Credit and Qualifications Framework (roughly equivalent to UK Level 4), the same level as a Higher National Certificate achieved at college [10].

Unlike Highers, Advanced Highers are not compulsory subjects for university entry. Yet, for admission to some of the most competitive degree programmes at higher education institutions in the UK, Advanced Highers are sometimes required or recommended. For example, the University of Oxford expects pupils to achieve at least AAB at Advanced Higher unless there is sufficient evidence from the applicant that their school was unable to provide these qualifications [17]. In contrast, the Universities of Glasgow and Edinburgh do not require Advanced Highers for most programmes except medicine [126, 127]. The University of Strathclyde generally recommends, but does not require, Advanced Highers for entry to its Science, Engineering, Business and Law degrees [128]. In some Scottish institutions, students with Advanced Higher qualifications may be able to forgo certain examinations from the first year of their degree programme or skip the first year entirely and enter directly into second year. This is because the content from an Advanced Higher may overlap with the content taught at the first stage of the typical four-year Scottish degree programme.

Disconcertingly, empirical evidence on Scottish students' access to Advanced Highers is scant. Despite this, it appears to be generally accepted by stakeholders that an inequality does exist. It is implicitly acknowledged by the existence of Widening Access programmes which aim to improve access to Advanced Highers for those from disadvantaged backgrounds. Examples include the Regional Improvement Collaboratives in Tayside and in the West of Scotland [129], the University of Dundee's City Campus Project [130], and Glasgow Caledonian University's "Advanced Higher Hub" . The latter was available to those from the 40% most deprived areas across Glasgow [131]. The topic of access to Advanced Highers has received attention from British media outlets [132, 133]. Some have even referred to what they see as unequal access to Advanced Highers as a "postcode lottery" [134]. In 2018, the Herald online newspaper published an article in which the general secretary of the Educational Institute of Scotland teaching union warned that "Staffing limitations often mean that Advanced Higher classes can't be offered." [133]. In January 2024, the closure of the Advanced Higher Hub was announced [135] resulting in backlash from some members of the media [136–138]. Glasgow City Council defended this decision, justifying that it had plans to develop a new, more decentralised approach that could offer Advanced Highers to an expanded pool of learners [138, 139]. This announcement of the Hub's closure came at roughly the same time the new Commissioner for Fair Access argued that any cuts to fair access programmes should come alongside an impact assessment on the affected students [7]. These reports serve to highlight the contentious nature that unequal access to Advanced Highers has on public discourse, regardless of whether or not gaps in access truly exist.

A critical question arises from this context: do Advanced Highers improve Scottish students' chances of success at university? Yet again, however, little published literature exists in this area. In 2014, Croxford et al. [69] found that students with "more Advanced Highers and/or A-levels achieve better degree outcomes on average". In 2018, students who had attended Glasgow Caledonian's Advanced Higher Hub associated their positive early-experience of university with attendance at the hub [131]. It should be noted however, that participants in this study were self-selected, meaning that the results from this study [131] are not generalisable.

7.2 Aims of the Chapter

The aims of this chapter are to contribute to the literature on Advanced Highers and their effect on student's academic outcomes at higher education. It will do so by focussing on one qualification, Advanced Higher Mathematics, and a subset of degree programmes in the fields of Mathematics, Statistics, Science and Engineering. This approach was taken since Mathematics is the most commonly recommended Advanced Higher at the University of Strathclyde (see Section C.7), thus it provides the largest sample of students to draw from.

The first aim is to determine if there are any associations between the number/proportion of students who had Higher/Advanced Higher Mathematics across *SIMD Quintiles*. While any significant association does not imply that there is unequal access to Advanced Higher Mathematics across SIMD Quintiles, it could add to the evidence on the contextual backgrounds of students if Advanced Higher Mathematics is found to have a positive effect on successful outcomes at university.

The second aim of this chapter is to determine what effect, if any, having an Advanced Higher Mathematics qualification has on students' retention and completion outcomes. If Advanced Higher Mathematics is found to have a positive effect on student's academic outcomes at university, then this is problematic if access to Advanced Highers is not equal (as believed by various Scottish institutions and stakeholders).

The third and final aim of this chapter is to determine the most appropriate regression technique for modelling the binary outcomes of retention and completion, whether Logistic, Log-Binomial or Modified Poisson. Understanding the advantages/disadvantages of each would better inform future modelling approaches on the **School-leavers dataset**.

7.3 Data

The subset of the **School-leavers dataset** used for this analysis was the **Advanced Higher subset** (outlined in Section 3.9). It has $6,815$ observations from the Faculties of Science and Engineering across *Academic Cohorts* 2012/13-2018/19. For brevity, those whose *Best Mathematics Qualification* was at Higher may be referred to as "Higher Mathematics students" and similarly for "Advanced Higher Mathematics students". The Department of Mathematics & Statistics had the highest proportion (78.8%) of Advanced Higher Mathematics students within its school-leaver population (Figure 7.1). Departments which tended to recommend Advanced Higher Mathematics for their programmes had higher proportions of Advanced Higher Mathematics students, and vice versa (see Section C.7 Table C.19 for Advanced Highers that were recommended by each department).

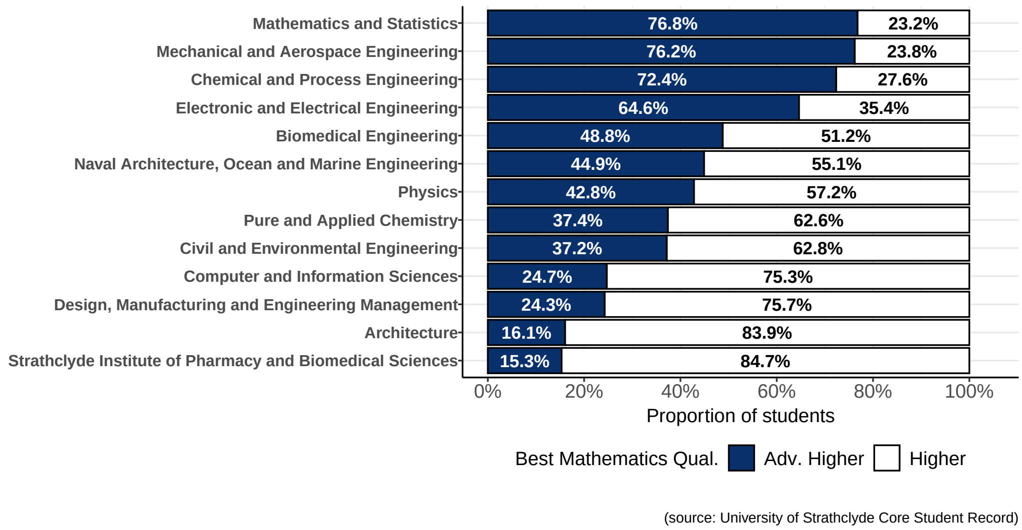


Figure 7.1: Proportion of school-leavers with Higher/Advanced Higher Mathematics within each department.

The outcomes of interest for this analysis are *Retention Status* and *Completion Status* (see Section 4.2 for outcome definitions), which are informally referred to here as the retention and completion outcomes. Prior to fitting any models, the **Advanced Higher subset** was subset into two groups: students who were registered with the Department of Mathematics and Statistics (denoted the **Maths subset**, $n = 754$) and students who were registered across all other departments within the Faculties of Science and Engineering (denoted the **SciEng subset**, $n = 6,061$). This approach was taken as it was assumed that Advanced Higher Mathematics would potentially have a different effect on students registered with mathematics and statistics programmes (**Maths subset**) than students registered with other science degree programmes (**SciEng subset**). Furthermore, Advanced Higher Mathematics was recommended for all of the Department of Mathematics and Statistics programmes, whereas within other departments this could vary from programme-to-programme. Around 63.9% of programmes in the **SciEng subset** recommended Advanced Higher Mathematics and around 64.7% of all students in this subset were registered with such programmes. In the **Maths subset**, 85.4% of students were successfully retained at the end of first year and 57.8% successfully completed their degree programme within four years. In the **SciEng subset**, 90.2% of students were successfully retained and 73.0% successfully completed their degree.

7.4 Methods

The first aim was addressed through a Chi-square test for trend between *SIMD Quintile* and *Best Mathematics Qualification* across the whole **Advanced Higher subset**. To address the second and third aims, Logistic, Modified Poisson, and Log-Binomial regression models were fit to both the **Maths subset** and the **SciEng subset**, for each academic outcome (*Retention Status* and *Completion Status*). The models applied to the **Maths subset** controlled for the following effects: whether or not a student held a Higher or Advanced Higher Mathematics qualification (*Best Mathematics Qualification*), *Prior Attainment Points*, *Sex*, *Ethnicity*, *SIMD Quintile* and *Academic Cohort*. The models applied to the **SciEng subset** used the same explanatory variables but also considered an interaction term between *Best Mathematics Qualification*, and whether or not a student's programme recommended Advanced Higher Mathematics (*AH Maths Recommended*), since not all programmes in the **SciEng subset** did so. A list of the explanatory variables used and their descriptions can be seen in Table 7.1.

Given that the retention and completion outcomes were common ($> 10\%$) in the **Maths subset** and the **SciEng subset**, it was expected that the derived odds-ratios from the Logistic regression models would be larger than the derived risk-ratios from the Modified Poisson and Log-Binomial models (see Section 6.1). Therefore, model estimates and goodness-of-fit were compared across models prior to interpreting the estimated effects of each covariate. For the Log-Binomial models, initial values for the covariates had to be specified to aid convergence. Each parameter's starting value was set to zero except the intercept, which was set to $\log(\mathrm{E}[Y])$ where $\mathrm{E}[Y]$ was the expected value of the retention or completion outcome. The *Prior Attainment Points* variable was mean-centred (mean value of 19.02, the equivalent of AAAAA-AC at Higher). The theory underpinning the regression methods is outlined in Chapter 6. For more details on the definitions of explanatory variables, see Section 4.6.

Table 7.1: List of explanatory variables used in the regression models for the retention and completion outcomes.

Variable	**Type**	**Values**	**Reference Level**
Academic Cohort	Categorical	2012/13 - 2018/19	2012/13
AH Mathematics Recommended	Categorical	Yes, No	No
Best Mathematics Qualification	Categorical	Higher, Adv. Higher	Higher
Ethnicity	Categorical	White, Ethnic-Minority	White
Prior Attainment Points	Continuous	$(-12, 13)$	0 (mean-centred)
Sex	Categorical	Male, Female	Female
SIMD Quintile	Categorical	1, 2, 3, 4, 5	1

7.4.1 Software Used for Analysis

All analyses were conducted using the statistical software `R` (version 4.3.1) [140]. Logistic, Poisson and Log-Binomial regression models were fit using the `glm()` function from the `stats` package [140]. Robust variances for the Modified Poisson Regression model were derived using the `sandwich (3.1-0)` package [118]. Effects plots were created using the `sjPlot (v.2.8.16)` package [141]. Additional packages used for general cleaning and visualisations included: `tidyverse (2.0.0)` [142], `patchwork (1.2.0)` [143], and `xtable (1.8-4)` [144].

7.5 Results - Chi-Square Test

A significant association (χ^2 = 86.6, d.o.f = 1, p-value < 0.001) was found between *SIMD Quintile* and *Best Mathematics Qualification*. There was a lower than expected number of school-leavers in *SIMD Quintiles* 1-2 with an Advanced Higher Mathematics qualification (Table 7.2). The proportion of school-leavers entering the Faculties of Science and Engineering with an Advanced Higher Mathematics qualification was roughly equal in *SIMD Quintile* 3, and greater than 50% in *SIMD Quintiles* 4-5 (Figure 7.2).

Table 7.2: Observed and Expected Counts for Higher/Advanced Higher Mathematics students within each SIMD Quintile.

SIMD Quintile	Observed (Expected) Counts	
	Higher	Adv. Higher
1	627 (519)	418 (526)
2	754 (666)	589 (677)
3	807 (799)	803 (811)
4	985 (1067)	1166 (1084)
5	1641 (1763)	1912 (1790)

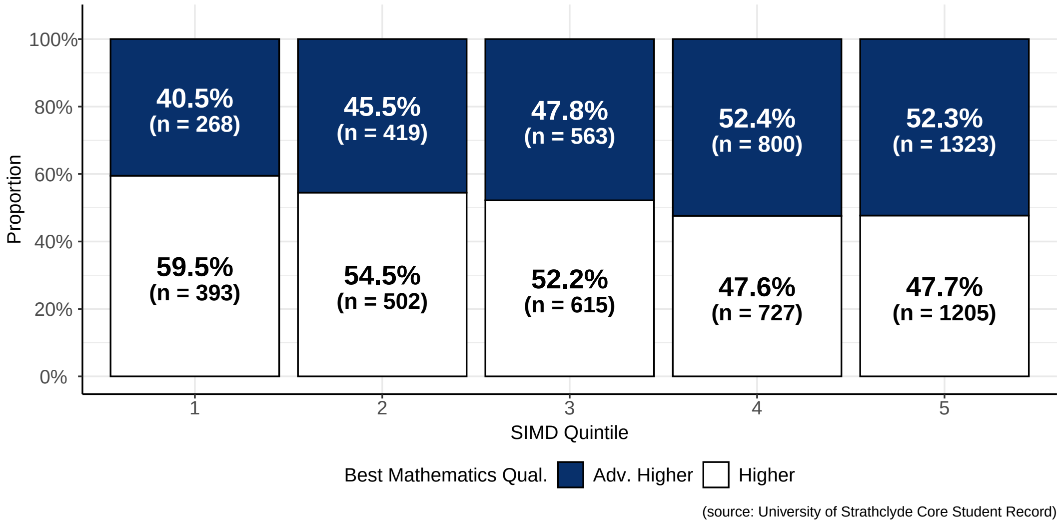


Figure 7.2: Proportion of school-leavers with Higher/Advanced Higher Mathematics within each SIMD Quintile.

7.6 Results - Regression Models

7.6.1 Mathematics and Statistics (Maths) Subset

For the retention outcome (Table 7.3), the Logistic, Modified Poisson, and Log-Binomial models each had different estimated risk/odds ratios when compared to one another. For example, compared to Higher Mathematics students, Advanced Higher Mathematics students had 4.383 times the odds of being retained (Logistic), or had 1.365 times the risk of being retained (Modified Poisson model), or had 1.221 times the risk of being retained (Log-Binomial). In comparison to the Modified Poisson estimate, the estimates from the Logistic and Log-Binomial models were exaggerated by factors of 3.211 and 0.895 respectively. These exaggerations were also present, though smaller in magnitude, in the models for completion (Table 7.4).

The following interpretations will use the Log-Binomial as an example (Tables 7.3 and 7.4). Students from the **Maths subset** who held an Advanced Higher Mathematics qualification were 22.1% [95% CI: 10.4%, 35.1%] more likely to progress and 44.3% [95% CI: 17.4%, 77.5%] more likely to complete their degree compared to their peers with only Higher Mathematics (Table 3). For each additional point (or grade) increase over the mean *Prior Attainment Points*, a student was 0.7% [95% CI: 0.2%, 1.1%] more likely to progress at the end of first year, and 3.7% [95% CI: 2.6%, 4.8%] more likely complete their degree within four years. Students from *Academic Cohort* 2015/16 were 28.5% [44.2% 8.4%] less likely to complete their degree compared to their peers from 2012/13. The remaining cohorts did not significantly differ from the 2012/13 cohort. Given that the sizes of these cohorts within the **Maths subset** are relatively small (around 100 students each year), some volatility is expected here. *Academic Cohort* may have some linear effect on retention, as the risk-ratios appear to decrease with each successive cohort. There does not appear to be any association between the academic outcomes and *Sex* and *SIMD Quintile*.

Table 7.3: Comparison of estimated odds and risk-ratios on retention outcome for students from Mathematics and Statistics (n = 754).

Variables	Regression Models (Retention Outcome)		
	Logistic	**Mod. Poisson**	**Log-Binomial**
(Intercept)	5.267 (1.785,15.540) [**]	0.682 (0.581,0.799) [***]	0.743 (0.650,0.850) [***]
Best Maths - Adv. Higher (vs Higher)	4.383 (2.622,7.325) [***]	1.365 (1.215,1.534) [***]	1.221 (1.104,1.351) [***]
Prior Attainment Points	1.172 (1.094,1.255) [***]	1.016 (1.009,1.023) [***]	1.007 (1.002,1.011) [**]
2013/14 Cohort (vs 2012/13)	1.256 (0.428,3.687)	0.992 (0.917,1.072)	0.981 (0.928,1.038)
2014/15 Cohort (vs 2012/13)	0.574 (0.220,1.498)	0.940 (0.857,1.030)	0.976 (0.906,1.052)
2015/16 Cohort (vs 2012/13)	0.364 (0.146,0.907) [*]	0.889 (0.798,0.991) [*]	0.933 (0.847,1.028)
2016/17 Cohort (vs 2012/13)	0.348 (0.140,0.864) [*]	0.879 (0.792,0.977) [*]	0.935 (0.856,1.021)
2017/18 Cohort (vs 2012/13)	0.317 (0.134,0.750) [**]	0.874 (0.799,0.956) [**]	0.935 (0.865,1.011)
2018/19 Cohort (vs 2012/13)	0.371 (0.151,0.913) [*]	0.888 (0.809,0.975) [*]	0.949 (0.874,1.031)
Female (vs Male)	0.943 (0.588,1.512)	0.987 (0.933,1.045)	0.996 (0.954,1.040)
SIMD Quintile 2 (vs 1)	1.037 (0.454,2.370)	1.038 (0.918,1.175)	1.023 (0.923,1.134)
SIMD Quintile 3 (vs 1)	0.926 (0.401,2.138)	1.052 (0.935,1.183)	1.024 (0.928,1.130)
SIMD Quintile 4 (vs 1)	0.946 (0.396,2.261)	1.039 (0.926,1.166)	1.021 (0.923,1.128)
SIMD Quintile 5 (vs 1)	1.481 (0.667,3.287)	1.097 (0.979,1.230)	1.038 (0.942,1.143)
Ethnic-minority (vs White)	2.618 (0.852,8.038)	1.109 (1.016,1.210) [*]	1.025 (0.976,1.077)

Wald's Test P-values: * < 0.05, **< 0.01, ***< 0.001.

Table 7.4: Comparison of estimated odds and risk-ratios on completion outcome for students from Mathematics and Statistics (n = 754).

Variables	**Regression Models (Completion Outcome)**		
	Logistic	**Mod. Poisson**	**Log-Binomial**
(Intercept)	1.201 (0.558,2.585)	0.424 (0.318,0.565) [***]	0.457 (0.350,0.596) [***]
Best Maths - Adv. Higher (vs Higher)	2.187 (1.441,3.319) [***]	1.575 (1.249,1.985) [***]	1.443 (1.174,1.775) [***]
Prior Attainment Points	1.189 (1.135,1.245) [***]	1.058 (1.044,1.073) [***]	1.037 (1.026,1.048) [***]
2013/14 Cohort (vs 2012/13)	0.681 (0.364,1.274)	0.867 (0.714,1.052)	0.933 (0.802,1.085)
2014/15 Cohort (vs 2012/13)	0.425 (0.225,0.803) [**]	0.733 (0.591,0.909) [**]	0.853 (0.707,1.028)
2015/16 Cohort (vs 2012/13)	0.304 (0.162,0.570) [***]	0.612 (0.464,0.807) [***]	0.715 (0.558,0.916) [**]
2016/17 Cohort (vs 2012/13)	0.542 (0.291,1.009)	0.805 (0.661,0.981) [*]	0.922 (0.785,1.082)
2017/18 Cohort (vs 2012/13)	0.660 (0.371,1.172)	0.876 (0.736,1.043)	0.964 (0.834,1.115)
2018/19 Cohort (vs 2012/13)	0.733 (0.398,1.350)	0.897 (0.749,1.076)	0.990 (0.846,1.158)
Female (vs Male)	1.135 (0.812,1.587)	1.051 (0.938,1.178)	1.025 (0.923,1.139)
SIMD Quintile 2 (vs 1)	0.987 (0.517,1.885)	1.012 (0.809,1.267)	0.978 (0.794,1.205)
SIMD Quintile 3 (vs 1)	1.049 (0.557,1.975)	1.073 (0.866,1.329)	0.986 (0.805,1.207)
SIMD Quintile 4 (vs 1)	1.067 (0.564,2.019)	1.055 (0.858,1.297)	0.973 (0.808,1.173)
SIMD Quintile 5 (vs 1)	1.075 (0.588,1.963)	1.066 (0.871,1.304)	1.015 (0.840,1.226)
Ethnic-minority (vs White)	1.061 (0.553,2.036)	1.028 (0.844,1.254)	1.065 (0.894,1.269)

Wald's Test P-values: * < 0.05, **< 0.01, ***< 0.001.

The deviance residuals from each model fit to the **Maths subset** (Figure 7.3) did not indicate any unduly influential observations. The expected value of the deviance residuals were also roughly centred around zero.

The *Prior Attainment Points* variable appeared to be linear with respect to the logit of the retention/completion outcomes in the Logistic models (Figure 7.4). The slope of the lines were steeper for the completion models than the retention models. This indicates that *Prior Attainment Points* are given more weight for the completion than the retention outcome. In the Modified Poisson and Log-Binomial models, the *Prior Attainment Points* variable appears linear with respect to the log of the completion outcome, but not sufficiently linear with respect to the log of the retention outcome. There appears to be two clusters of points for each of the Modified Poisson and Log-Binomial models. This is most noticeable in the models for retention (Figure 7.4 - (II) and (III)), where the relationship looks "S" shaped. Further inspection revealed that these two groups of observations corresponded to Higher and Advanced Higher Mathematics students. For the Modified Poisson models, there were several observations which had predicted probabilities of greater than 1 (i.e. where $\log(P) > 0$ in Figure 7.4 - (II) and (V)). This was not wholly unexpected given that these models were predicting mean values and not probabilities (see Section 6.5), and the prevalence of each outcome was high. The observations which had predicted mean value that exceeded 1 were all students who held Advanced Higher Mathematics qualifications and had higher than the mean *Prior Attainment Points*.

Taken together, the deviance residuals and linearity plots indicate that the models adequately fit the **Maths subset**, with some issues in the linearity of the *Prior Attainment Points* variable.

Figure 7.3: Comparing deviance residuals between retention/completion models applied to Mathematics and Statistics students (n = 754).

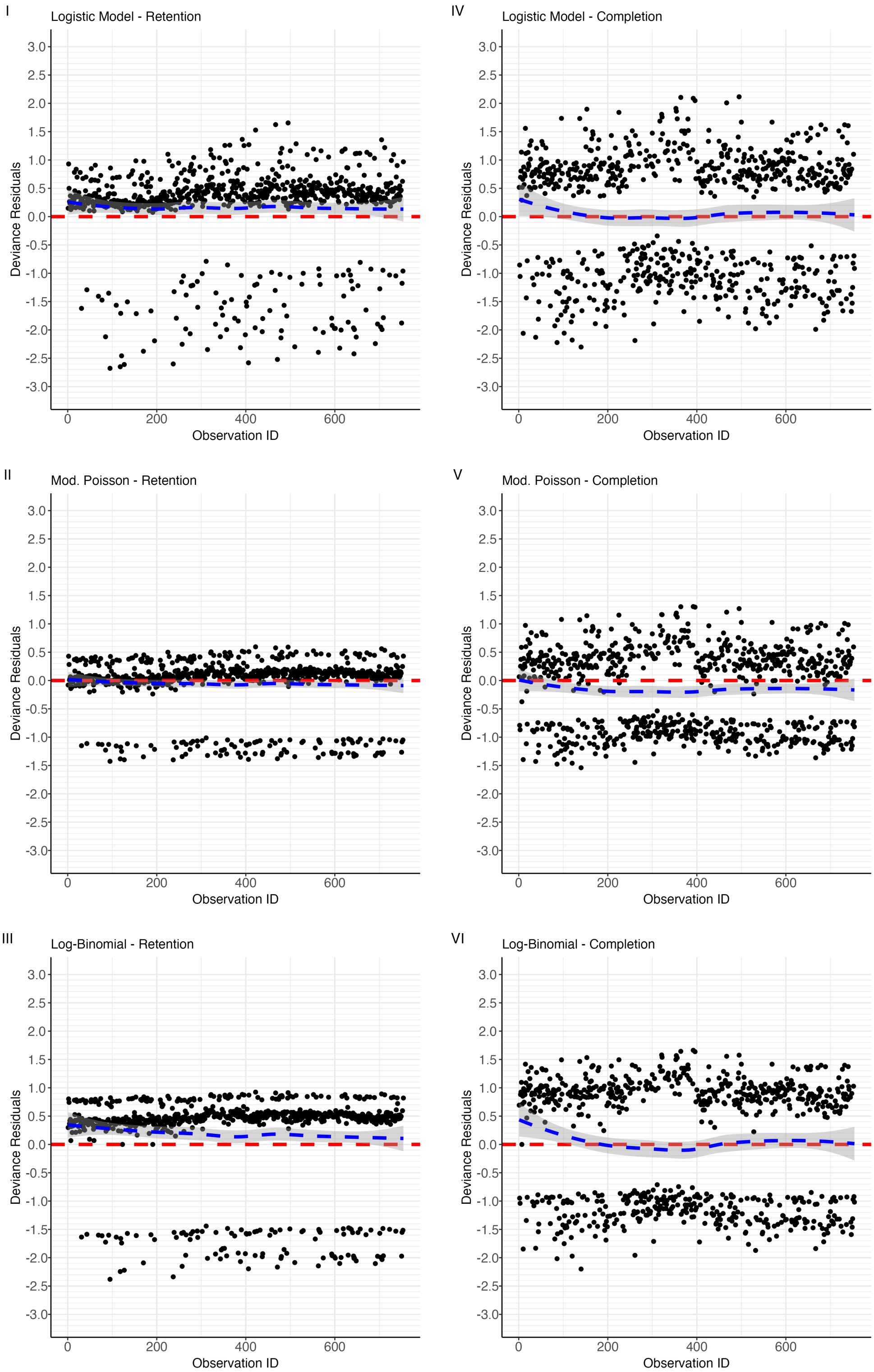


Figure 7.4: Comparing log/logit linearity of Prior Attainment Points variable in retention/completion models applied to Mathematics and Statistics students (n = 754).

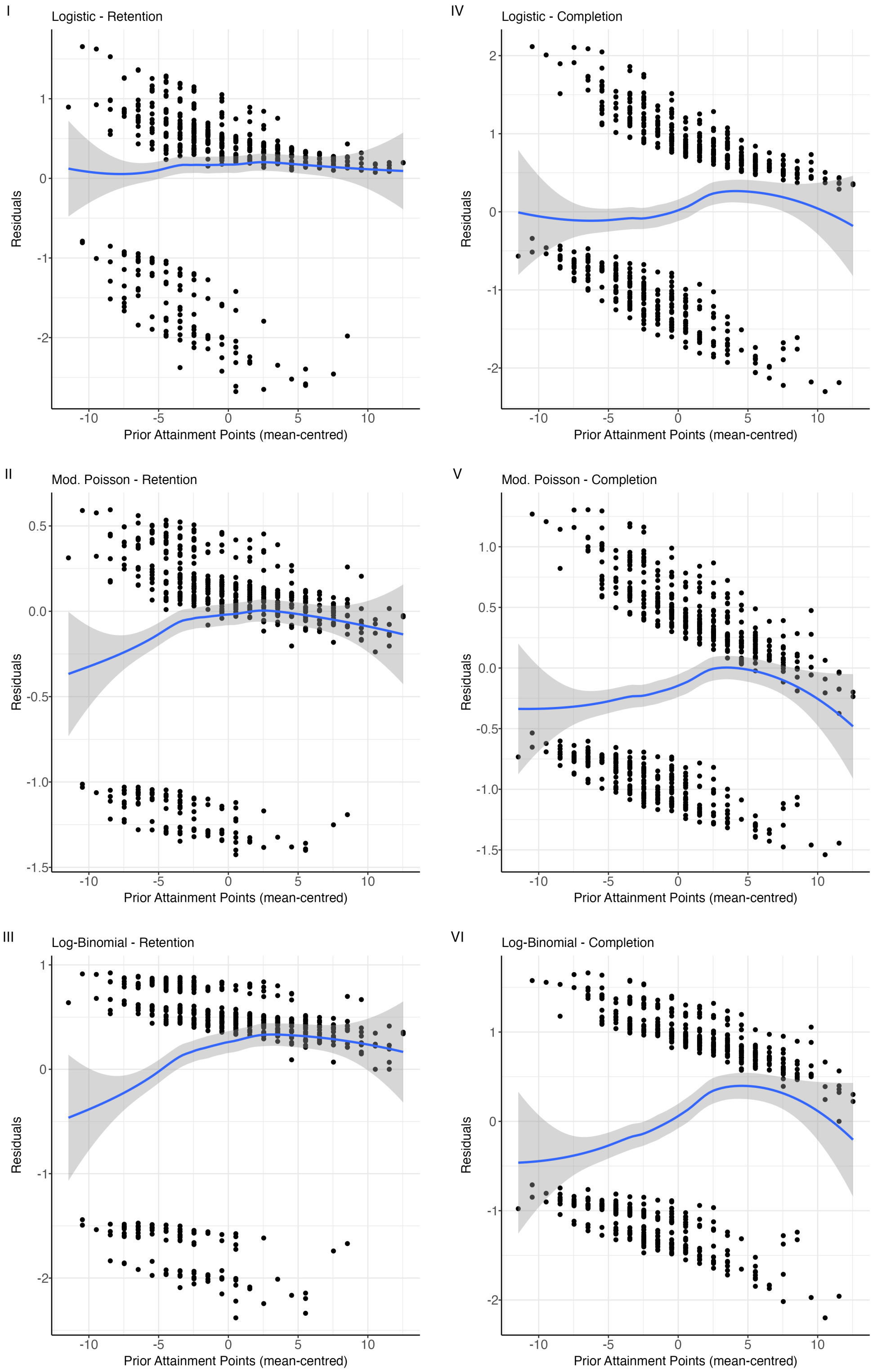
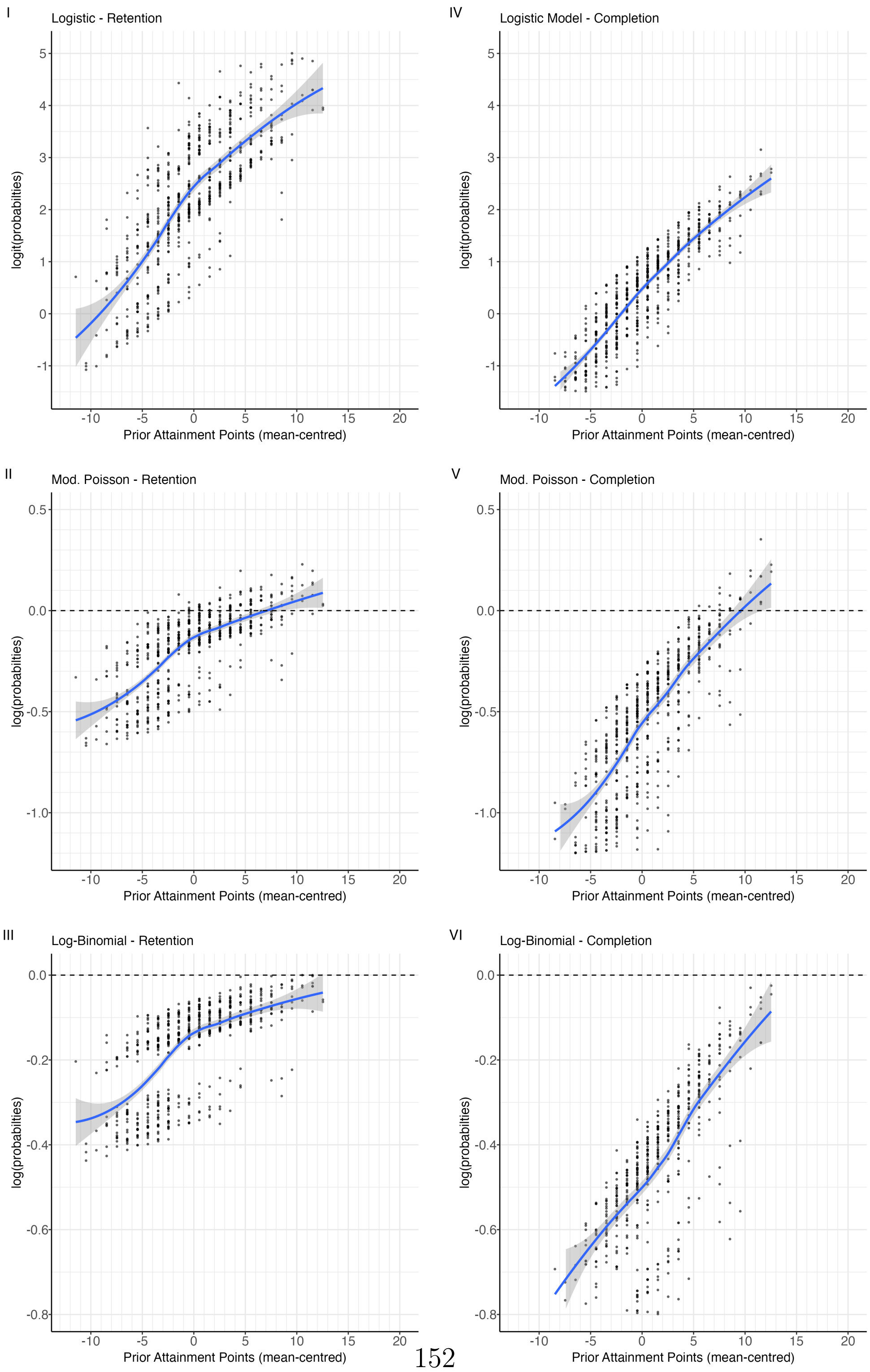


Figure 7.5: Predicted Probabilities of retention/completion versus Prior Attainment Points variable for each model applied to the Mathematics and Statistics students (n = 754).



7.6.2 Science and Engineering (SciEng) Subset

As seen in the **Maths subset**, the Logistic, Modified Poisson, and Log-Binomial models fit to the **SciEng subset** each had different estimated risk/odds ratios when compared to one another (Tables 7.5 and 7.6). However, the Logistic model estimates were not as exaggerated when compared to the other two models, and the Modified Poisson and Log-Binomial estimates were closer to one another.

The following interpretations will use the Modified Poisson as an example (Tables 7.5 and 7.6). Looking first at the retention outcome, for each additional point (or grade) increase over the mean *Prior Attainment Points*, students in the **SciEng subset** were 1.0% [95% CI: 0.8%, 1.2%] more likely to be retained at the end of first year. There was no association between retention and *Ethnicity*, nor *Sex*, nor *SIMD Quintile.* There was no association between retention and the *Academic Cohort* a student belonged to, except perhaps a small difference in the retention rates of those from *Academic Cohorts* 2012/13 and 2013/14. Similarly for the completion outcome, for each additional point (or grade) increase over the mean *Prior Attainment Points*, students from other Science and Engineering departments were and 2.9% [95% CI: 2.6%, 3.3%] more likely complete their degree within four years. Students from SIMD Quintile 5 were 14.5% [95% CI: 7.2%, 22.4%] more likely to complete their degree than students from SIMD Quintile 1. Females were also 4.0% [95% CI: 0.7%, 7.4%] more likely to complete their degree than males. There was no association between completion and *Ethnicity* nor *Academic Cohort.*

Table 7.5: Comparison of estimated odds and risk-ratios on retention outcome for students from Science and Engineering (n = 6,061).

Variables	**Regression Models (Retention Outcome)**		
	Logistic	**Mod. Poisson**	**Log-Binomial**
(Intercept)	8.630 (5.929,12.561) [***]	0.876 (0.839,0.915) [***]	0.888 (0.854,0.923) [***]
Best Maths - Adv. Higher (vs Higher)	1.182 (0.836,1.672)	1.022 (0.991,1.054)	1.011 (0.981,1.043)
Prog. Recommended AH Maths (vs not)	0.837 (0.684,1.025)	0.976 (0.950,1.003)	0.990 (0.966,1.015)
Prior Att. Points	1.125 (1.099,1.152) [***]	1.010 (1.008,1.012) [***]	1.004 (1.003,1.005) [***]
2013/14 Cohort (vs 2012/13)	1.421 (1.021,1.976) [*]	1.030 (1.001,1.061) [*]	1.014 (0.985,1.043)
2014/15 Cohort (vs 2012/13)	1.002 (0.732,1.372)	1.002 (0.972,1.033)	1.000 (0.972,1.029)
2015/16 Cohort (vs 2012/13)	0.868 (0.633,1.189)	0.986 (0.955,1.017)	0.995 (0.966,1.025)
2016/17 Cohort (vs 2012/13)	0.802 (0.588,1.094)	0.978 (0.946,1.010)	0.991 (0.961,1.022)
2017/18 Cohort (vs 2012/13)	0.982 (0.706,1.366)	0.997 (0.965,1.029)	1.000 (0.969,1.031)
2018/19 Cohort (vs 2012/13)	0.778 (0.567,1.068)	0.972 (0.940,1.005)	0.990 (0.960,1.022)
Female (vs Male)	1.073 (0.885,1.302)	1.008 (0.991,1.027)	1.002 (0.985,1.019)
SIMD Quintile 2 (vs 1)	0.968 (0.701,1.339)	0.999 (0.958,1.041)	1.001 (0.965,1.039)
SIMD Quintile 3 (vs 1)	1.060 (0.767,1.464)	1.013 (0.976,1.053)	1.007 (0.972,1.042)
SIMD Quintile 4 (vs 1)	1.295 (0.941,1.782)	1.032 (0.996,1.070)	1.016 (0.983,1.050)
SIMD Quintile 5 (vs 1)	1.170 (0.872,1.571)	1.024 (0.989,1.060)	1.013 (0.982,1.046)
Ethnic-minority (vs White)	1.275 (0.918,1.769)	1.019 (0.990,1.048)	1.009 (0.981,1.038)
Interaction: Adv. Higher & Rec.	1.753 (1.149,2.675) [**]	1.038 (1.000,1.078)	1.017 (0.980,1.056)

Wald's Test P-values: * < 0.05, **< 0.01, ***< 0.001.

Table 7.6: Comparison of estimated odds and risk-ratios on completion outcome for students from Science and Engineering (n = 6,061).

Variables	**Regression Models (Completion Outcome)**		
	Logistic	**Mod. Poisson**	**Log-Binomial**
(Intercept)	2.243 (1.731,2.906) [***]	0.663 (0.612,0.718) [***]	0.666 (0.618,0.718) [***]
Best Maths - Adv. Higher (vs Higher)	1.237 (0.968,1.581)	1.040 (0.988,1.095)	1.057 (1.006,1.110) [*]
Prog. Recommended AH Maths (vs not)	0.629 (0.541,0.732) [***]	0.861 (0.819,0.905) [***]	0.900 (0.859,0.944) [***]
Prior Att. Points	1.130 (1.112,1.148) [***]	1.029 (1.026,1.033) [***]	1.014 (1.012,1.017) [***]
2013/14 Cohort (vs 2012/13)	0.967 (0.785,1.191)	0.993 (0.935,1.054)	0.998 (0.943,1.056)
2014/15 Cohort (vs 2012/13)	1.077 (0.871,1.332)	1.025 (0.968,1.085)	0.986 (0.932,1.042)
2015/16 Cohort (vs 2012/13)	0.972 (0.783,1.206)	0.995 (0.939,1.053)	1.011 (0.958,1.068)
2016/17 Cohort (vs 2012/13)	1.040 (0.836,1.294)	1.011 (0.954,1.071)	1.019 (0.965,1.077)
2017/18 Cohort (vs 2012/13)	1.280 (1.018,1.610) [*]	1.057 (0.999,1.119)	1.051 (0.996,1.109)
2018/19 Cohort (vs 2012/13)	1.194 (0.950,1.501)	1.038 (0.980,1.099)	1.035 (0.983,1.090)
Female (vs Male)	1.182 (1.034,1.350) [*]	1.040 (1.007,1.074) [*]	1.009 (0.979,1.040)
SIMD Quintile 2 (vs 1)	1.272 (1.004,1.611) [*]	1.084 (1.005,1.170) [*]	1.075 (1.001,1.155) [*]
SIMD Quintile 3 (vs 1)	1.261 (1.004,1.584) [*]	1.084 (1.008,1.166) [*]	1.049 (0.978,1.124)
SIMD Quintile 4 (vs 1)	1.315 (1.056,1.638) [*]	1.100 (1.026,1.180) [**]	1.095 (1.026,1.169) [**]
SIMD Quintile 5 (vs 1)	1.570 (1.276,1.933) [***]	1.145 (1.072,1.224) [***]	1.129 (1.060,1.201) [***]
Ethnic-minority (vs White)	1.114 (0.894,1.388)	1.021 (0.968,1.078)	1.023 (0.974,1.074)
Interaction: Adv. Higher & Rec.	1.194 (0.897,1.590)	1.083 (1.014,1.157) [*]	1.066 (1.000,1.136) [*]

Wald's Test P-values: * < 0.05, **< 0.01, ***< 0.001.

The interpretation of the *Best Mathematics Qualification* and *AH Maths Recommended* variables are more complicated here due to the presence of the interaction term between the two. There appears to be an association between the interaction term and the retention outcome in the Logistic and Modified Poisson models, but not the Log-Binomial. Whereas an association between the interaction term and the completion outcome appears to be present in the Modified Poisson and Log-Binomial models, but not the Logistic. Across all models point estimates for *Best Mathematics Qualification* are greater than 1, indicating that if Advanced Higher Mathematics did have an effect on the academic outcomes of Science and Engineering students, it would be a positive effect. This contrasts with the *AH Maths Recommended* estimates which are less than 1, indicating a negative effect. Likelihood-ratio tests were conducted to test whether the inclusion of the interaction term between significantly improved upon a model fit which did not contain the interaction term (Table 7.7). For the retention outcome, the interaction term significantly improved the fit of the Logistic and Log-Binomial models, but not the Modified Poisson model. For the completion outcome, the interaction term significantly improved the fit of the Log-Binomial model, but not the Logistic nor Modified Poisson models.

Table 7.7: Likelihood-ratio tests which compared models with/without interaction term fit to the **SciEng subset**.

Outcome	**Model**	**Test Stat**	**LRT P-value**
Retention	Logistic	6.52	0.011
	Mod. Poisson	0.36	0.546
	Log-Binomial	2.63	0.105
Completion	Logistic	1.46	0.226
	Mod. Poisson	1.38	0.240
	Log-Binomial	13.63	<0.001

Visualising the interaction groups across both outcomes shows that Advanced Higher Mathematics students had higher retention and completion rates when compared to Higher Mathematics students (Figure 7.9); this was regardless of whether or not Advanced Higher was recommended. The retention rates for Advanced Higher Mathematics students increased when the qualification was recommended (Figure 7.9A). In contrast, there was a decrease in completion rates within Higher Mathematics students when the Advanced Higher was recommended (Figure 7.9B).

The deviance residuals from each model fit to the **SciEng subset** (Figure 7.6) indicated a few influential observations that corresponded to students with *Advanced Higher Mathematics* and very high *Prior Attainment Points* that did not complete their degree within four years. The expected value of the deviance residuals were roughly centred around zero.

The *Prior Attainment Points* variable appeared to be linear with respect to the transformation of the retention/completion outcomes across all models (Figure 7.7). Again, there were some observations in the Modified Poisson models with predicted probabilities of greater than 1 (Figure 7.7 - (II) and (V)). Taken together, the deviance residuals and linearity plots indicate that the models adequately fit the **SciEng subset**, with fewer issues in the linearity of the *Prior Attainment Points* variable compared to the **Maths subset**.

Figure 7.6: Comparing deviance residuals between retention/completion models applied to Science and Engineering students (n = 6,061).

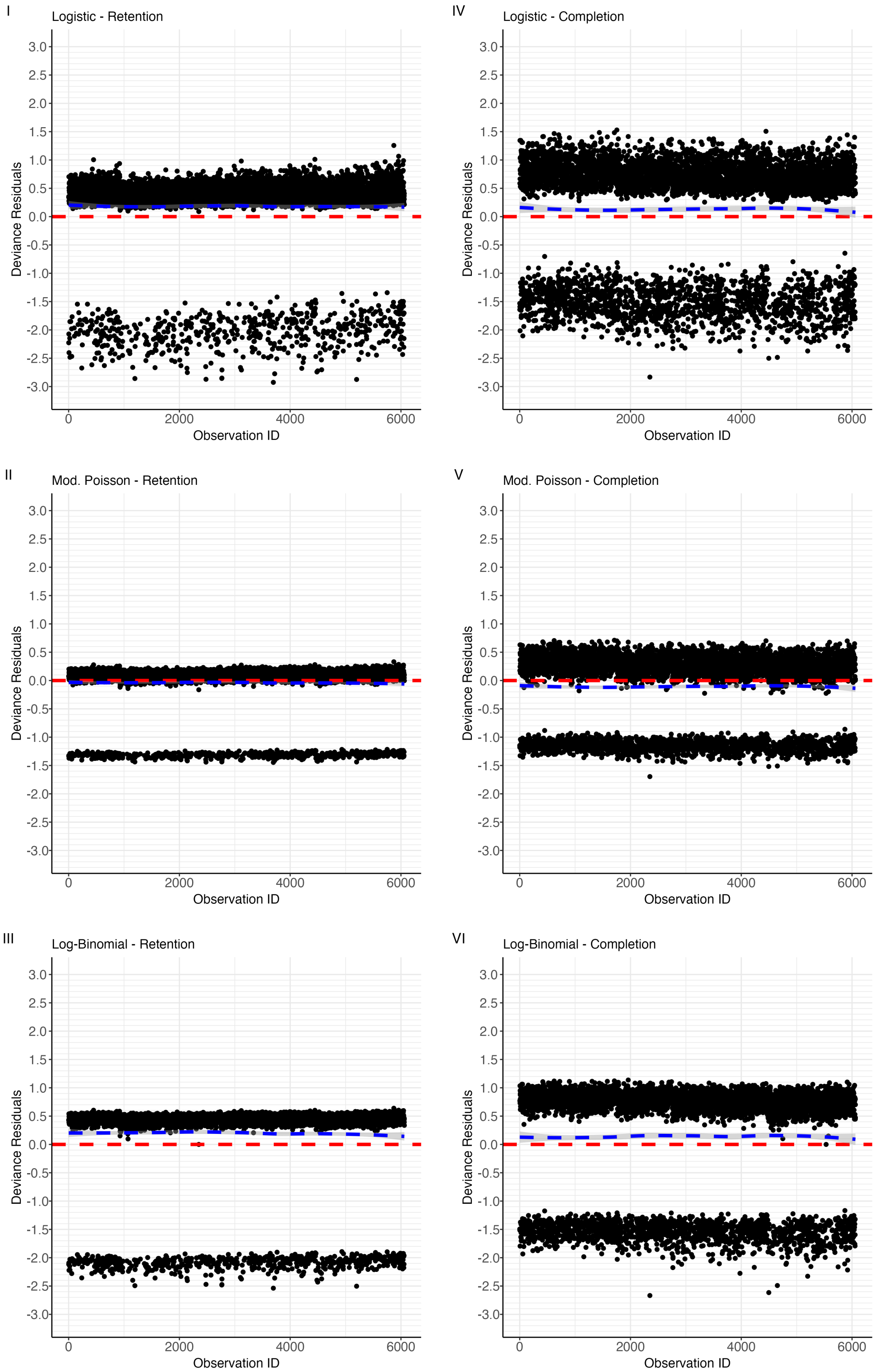


Figure 7.7: Comparing log/logit linearity of Prior Attainment Points variable in retention/completion models applied to Science and Engineering students (n = 6,061).

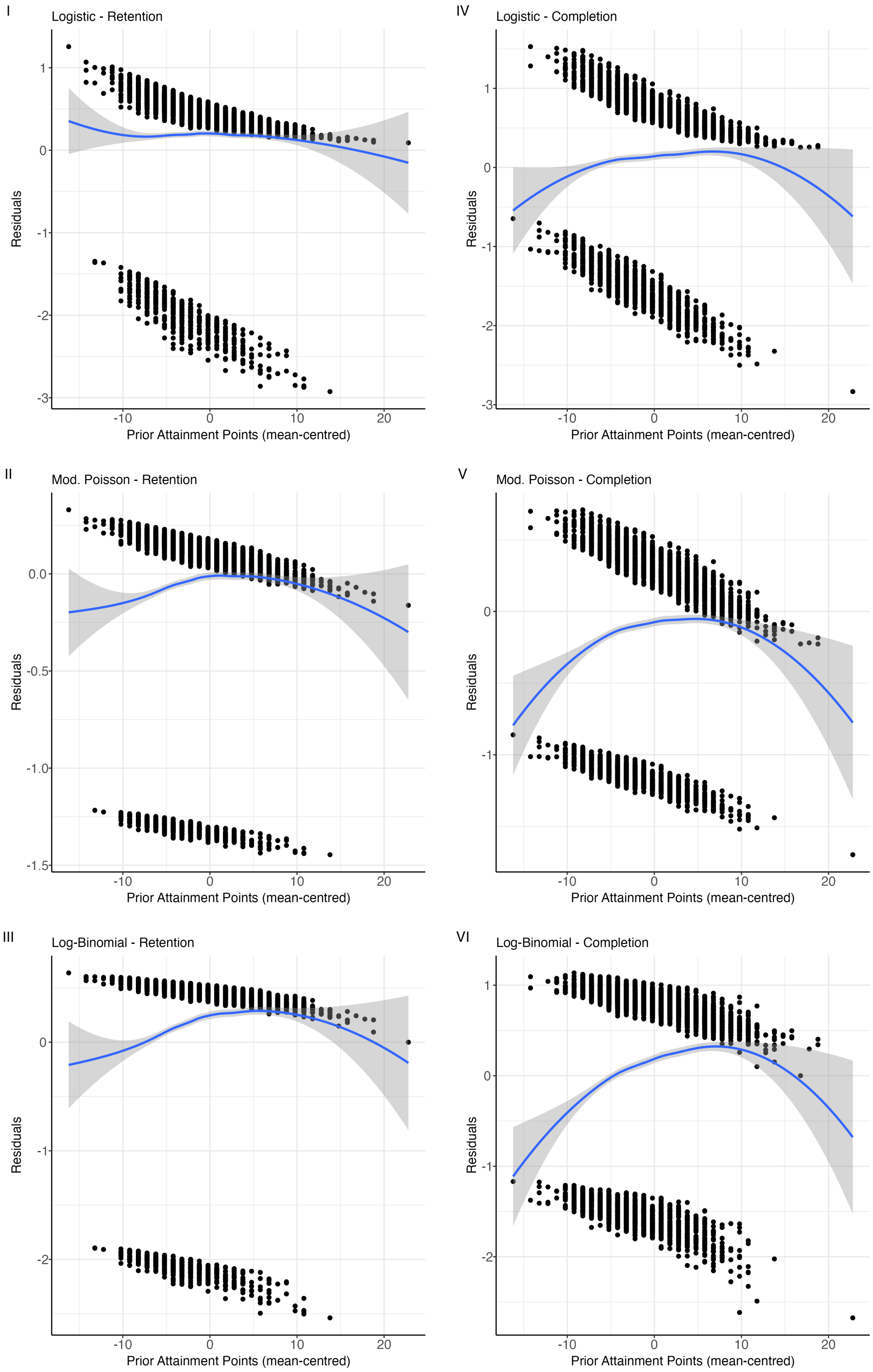


Figure 7.8: Predicted Probabilities of retention/completion versus Prior Attainment Points variable for each model applied to the Science and Engineering students (n = 6.061).

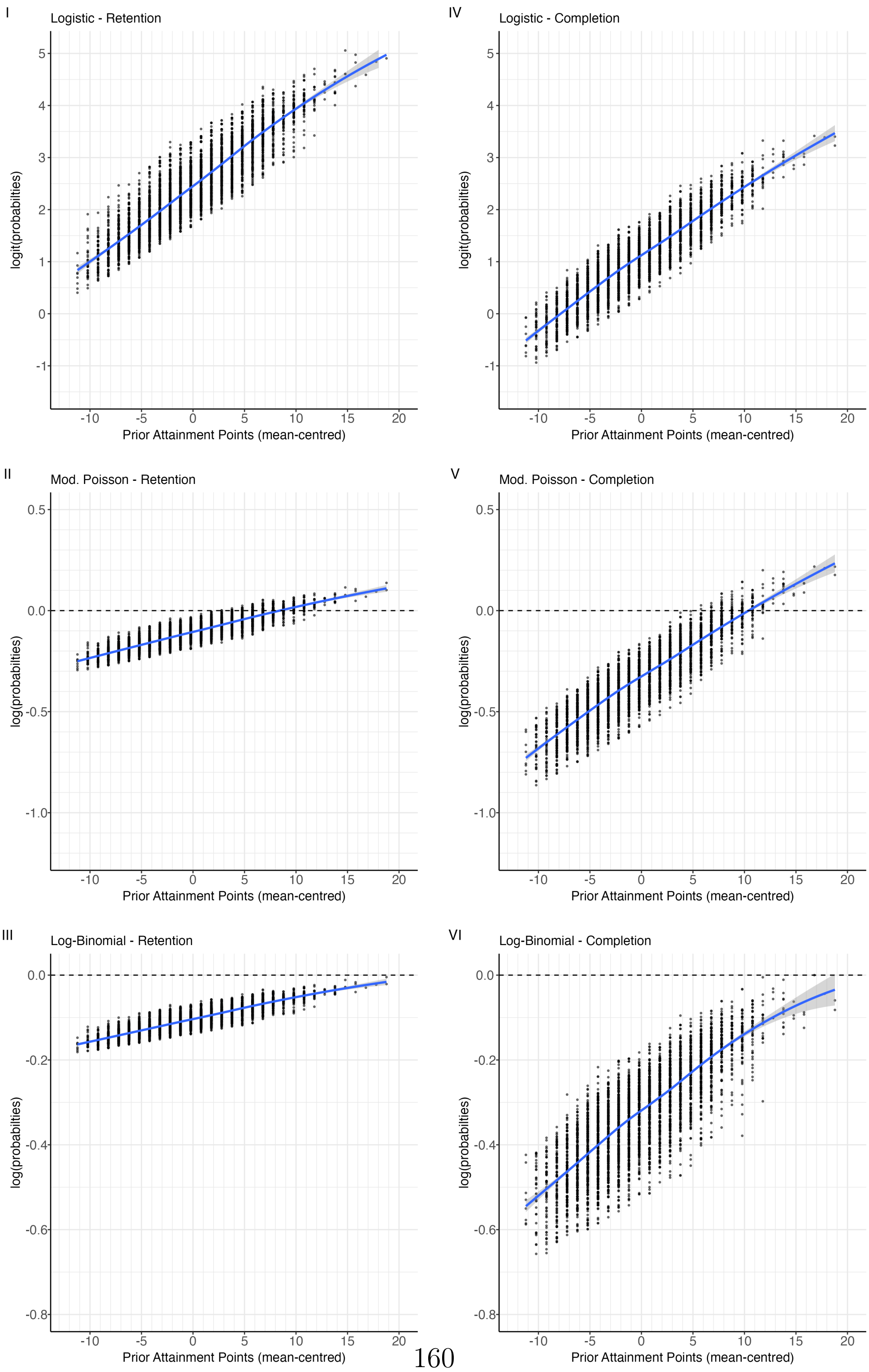
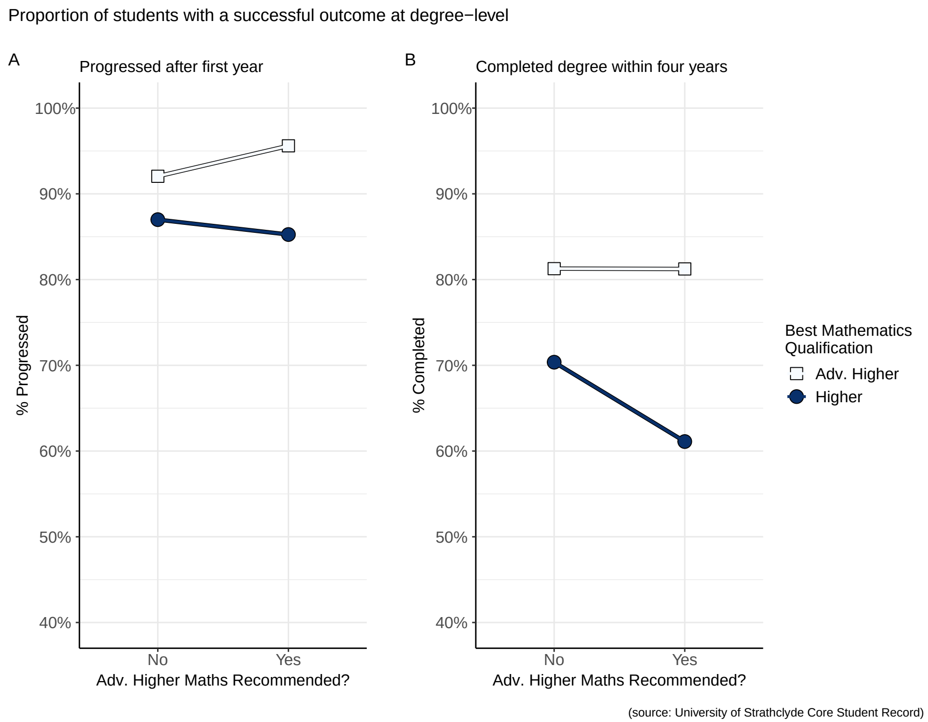


Figure 7.9: The proportion of successful retentions/completions amongst students in Science and Engineering programmes (not including Mathematics and Statistics), grouped by their *Best Mathematics Qualification* and whether or not their programme recommended Advanced Higher Mathematics.



7.7 Discussion

There was a significant difference between the proportion of students with an Advanced Higher Mathematics qualification across SIMD Quintiles, where 40.5% of students from the SIMD Quintile 1 had Advanced Higher Mathematics compared to 52.3% in SIMD Quintile 5. This should be interpreted carefully; it is not known why this difference exists. This result should not be interpreted as evidence for differences between SIMD Quintiles in who has access to Advanced Highers, since it does not take into account students who were unsuccessful in applying to the University, nor qualified applicants who had decided to accept an offer elsewhere. Further examination on the provision of Advanced Higher Mathematics across Scottish schools and the University's distribution of offers made across SIMD Quintiles would be necessary to explain this result. This result could also suggest that SIMD Quintile may be confounded with the relationship between Advanced Higher Mathematics and a successful outcome at university.

Advanced Higher Mathematics had a strong and positive association with the successful retention and completion rates of students enrolled on Mathematics and Statistics degrees. The effect of Advanced Higher Mathematics was more complicated to interpret for students on other Science and Engineering degrees. However, the results suggest that there may be an equivalent positive association, conditional upon whether or not Advanced Higher Mathematics was recommended by a student's chosen degree programme. Taken together, these findings indicate that Advanced Higher Mathematics is likely to be beneficial for all students studying Mathematics and Statistics or other Science and Engineering programmes at the University. For the latter group of students, it remains unclear whether or not these associations could be attributed to other factors, such as sex, socio-economic status, or prior attainment in other subjects. For example, this analysis did not account for the effects of other relevant Advanced Highers in science subjects, e.g. Physics, Chemistry, Biology, etc. In addition, *Prior Attainment Points* are an imperfect measure of the academic potential to succeed though are simple to implement and interpret. Further investigation is required into how best to measure and compare similar attainment profiles across Highers and Advanced Highers. These findings may highlight a potential problem within the secondary stages of Scottish education. If Advanced Higher Mathematics is not accessible by all learners across Scotland, as is believed (rightly or wrongly) by stakeholders (Section 7.1), then this could be evidence of an unfair system. If the accessibility of Advanced Higher is affected by the SIMD Quintile a learner is from, or the SIMD Quintile a school is located, then this undermines Widening Access objectives.

The *SIMD Quintile* a student came from seems to have little-to-no influence on Mathematics and Statistics students retention and completion outcomes once controlling for the effects of others factors. In fact, the only predictors which appeared to have any influence were *Prior Attainment Points*, whether or not a student held an Advanced Higher Mathematics qualification, and the *Academic Cohort* a student belonged to. This is in contrast to previous studies which found that success at university could not be predicted based on prior attainment alone [69, 76–79]. Yet, there was a strong influence of *SIMD Quintile* within the rest of Science and Engineering students, even after controlling for the same factors. This may suggest that factors which are predictive of success may be programme-specific. Then again, it should also be emphasised that attainment and socio-economic background are confounded with one another, meaning that the influence of socio-economic background still remains even if one were to only focus on prior attainment. Future analyses could perhaps tease out these associations in more detail within the **Maths** and **SciEng** subsets.

Care should be taken with the interpretation of the *Prior Attainment Points* variable in the Modified Poisson and Log-Binomial models; it was not linearly associated with the outcome when fit to data on the students from the Department of Mathematics and Statistics. This issue was also present in the Modified Poisson and Log-Binomial models applied to students from the rest of Science and Engineering but were less severe. However, given that the Logistic model had no such issues, and it agreed with the Modified Poisson and Log-Binomial in terms of the direction of the effect of *Prior Attainment Points*, this adds confidence that the variable is positively associated with the outcomes of retention and completion. The linearity issues could be addressed by categorising *Prior Attainment Points* in future models (for example, by using the *Prior Attainment Quintile* variable - see Section 4.4.1) or by fitting a quadratic or cubic term for *Prior Attainment Points*.

A small number of students ($n = 52$) who did not have a Mathematics qualification at Higher or Advanced Higher were removed. These were students who either had an A-level or no Mathematics qualification whatsoever, and were all students from the Faculty of Engineering (i.e. within the **SciEng subset**). These students also had the lowest retention and completion rates when compared to Higher and Advanced Higher Mathematics students. A sensitivity analysis was conducted which found that model estimates in Tables 7.5 and 7.6 did not change when these students were included. Removal of these students therefore did not have an impact on the interpretation of the model results.

A small number of students whose Ethnicities were unknown or missing were removed (see Section 3.5). Sensitivity analyses found that model estimates remained unchanged when these observations were included, their removal therefore has no impact on the interpretation of the model results.

The odds-ratios derived from the Logistic model were poor estimators of risk-ratios, making these models more difficult to interpret for a lay audience. The model estimates between the Modified Poisson and Log-Binomial models differed, likely because the prevalence of outcomes were so common (and very close to 1 in the case of retention) combined with the strong association of *Prior Attainment Points* variable. This is apparent when examining the linearity plots (Figures 7.4 and 7.7) where observations with high *Prior Attainment Points* had predicted probabilities greater than 1 in the Modified Poisson model, and predicted probabilities close to 1 in the Log-Binomial model. This may have caused convergence issues in the Log-Binomial model, which took 25 Fisher-iterations to converge, whereas the Logistic and Modified Poisson models took between 4-6 Fisher iterations. All of the previous issues may be related to the prevalence of a successful retentions and completions both being so common ($> 10\%$). They could be addressed by using an outcome variable that is rarer. For example, models could instead estimate the effect of covariates on dropping out of university rather than achieving a successful outcome.

There is no clear best-fitting model given that each of the three models (Logistic, Modified Poisson Log-Binomial) had some issues when fit to the data. If a preferred model were to be selected for an outcome variable that is common ($> 10\%$), it would be the Log-Binomial model. If the Log-Binomial model struggled to converge however, a Modified Poisson might be preferred given that it still derives risk-ratios. Logistic regression models are still valid, but care would have to be taken when interpreting the estimates to a lay audience.

7.7.1 Limitations and Future Research

The analyses conducted here assumes that any passing grade at Advanced Higher is equivalent to, or better than, any grade awarded at Higher. However, some stakeholders believe that an A at Higher is equivalent to, or better than C or D grades at Advanced Higher. For example, UCAS tariff points score A grades at Higher as 33 points, while D grades at Advanced Higher as 32 points. Future analyses could examine the specific grades students achieved and their association with academic outcomes at university.

Chen et al. [115] suggest that model misspecification occurs when important explanatory variables are omitted, when non-linear terms or critical interactions are not accounted for, or when there is measurement bias. It has been acknowledged (Section 4.7) that the **School-leavers dataset** does not contain any information on students' attainment at university-level, which is assumed to be a critical explanatory variable. There could be some underlying interactions that remain unaccounted for. It is assumed that the data collected for this analysis is correct thus measurement bias is not considered a factor here.

Many of the issues encountered in this chapter: exaggerated Logistic regression estimates, convergence of the Log-Binomial, differences between Log-Binomial and Modified Poisson estimates; can be ascribed to the retention and completion outcomes being so common (prevalence greater than 10%). An alternative approach is examined in Chapter 9 which uses time-to-event regression models. These models examine the rate of dropout over time, where the prevalence of dropout in any given academic session is at most 6% (Section 9.3), rare enough to avoid the issues encountered in this chapter. Thus, regression models may not be the most suitable method for measuring the academic outcomes of students at higher education, given that the most popular outcomes in the literature [76, 78] all have prevalences that are not rare (10%). However, their ease of interpretability should prove sufficient for stakeholders in Widening Access, if results on the boundary of the probability space $(0, 1)$ are interpreted with caution.

7.7.2 Conclusion

The results shown here suggest that, at the very least, having Advanced Higher Mathematics is associated with a positive outcome for students studying Mathematics and/or Statistics at degree-level. Future research should investigate whether this link is present across other Advanced Higher subjects and degree programmes and other higher education institutions. Further research should also establish whether there is a link between a student's socio-economic status and access to Advanced Higher. If access to Advanced Higher Mathematics (and other subjects) is not equal for all learners across Scotland, as is (rightly or wrongly) assumed by stakeholders, then this could be evidence of an unfair system. If students from more socio-economically deprived areas are disproportionately affected, then this could jeopardise Scotland's Widening Access ambitions.

Chapter 8

Comparing the Outcomes of Standard and Contextual Offer Students

The analysis from this chapter was published as a journal article in *Higher Education Quarterly* in January 2025 [145]. The article will also form part of an invited contribution to a special issue publication of *Higher Education Quarterly* due to be published later in 2025. The special issue will examine the topic of contextual offers and their impact at six UK institutions (including the University of Strathclyde). This chapter contains some additional detail that was not included in the original journal article, such as the effects of each *Faculty* in each model and the residuals and predicted probabilities plots (Figures 8.1 and 8.2, respectively).

8.1 Motivation and Research Questions

As established in Chapter 2, much of the emphasis on measuring Scotland's progress for Widening Access has been focused on admitting disadvantaged students into higher education; less emphasis has been placed on the performance of those students once they are on-programme. There is, however, some indication that this is changing after the new Commissioner for Fair Access recommended that equal weight be given to disadvantaged students' academic outcomes [7]. It was argued that the retention rates of target groups had seen little progress since 2016, and a commitment was made to investigate why this was the case. Indeed, there has been concern over a lack of support structures at institutions for students with lesser attainment due to their disadvantaged backgrounds and who may be struggling [97]. Having targets on the academic outcomes of disadvantaged students could prove to be sufficiently stretching for higher education institutions that have already achieved the Widening Access 2030 target of equal representation for students from the 20% most deprived areas.

The problem with measuring the academic outcomes of disadvantaged students is in deciding what an appropriate success rate is for such students. Ultimately, what is deemed appropriate is subjective. For example, it may seem appropriate to simply compare the success rates of contextual offer students compared to their peers. However, such a comparison may be unfair given that these students, by definition, have lower prior attainment from secondary education. Furthermore, the disadvantage that made these students eligible for a contextual offer does not disappear once they attend university. A more appropriate measure of their success may be to look at their chances of success, once controlling for other factors. For example, Boliver et al. [52] argued that an 80%+ probability of being retained at the end of first year, and a 65%+ probability of achieving a "good pass" at honours (first-class or upper second-class) could constitute "high-bars" of success for any student at highly-selective institutions. Similarly, one of the University of Strathclyde's key-performance indicators is to achieve between 90-95% retention from Year 1 to Year 2 for all students going towards 2030 [8].

8.2 Aims of the Chapter

This chapter aims to address the growing interest in post-admission academic outcomes for contextual offer students by investigating their retention and completion rates in the **School-leavers dataset**. This was achieved through two analysis aims. The first aim was to use risk-ratios to determine contextual offer students' chances of success compared to standard offer students, whilst controlling for other significant factors. This gives an assessment of how well contextual offer students are faring in higher education compared to their peers. The second aim was to determine the average predicted probability of success for contextual offer students, whilst controlling for other significant factors. These predicted probabilities are then compared to the progression (80%+) benchmark provided by Boliver et al. [52] as well as the University's 2030 Key Performance Indicator of 90-95% retention [8]. The predicted probability of success for standard offer students was also derived for comparison. Taken together, both the risks ratios and predicted probabilities give a clearer picture of the academic outcomes of contextual offer students in higher education.

8.3 Data

This analysis made use of the **Contextual Offer subset** (see Section 3.9 for more details) which contained $7,534$ students from *Academic Cohorts* 2015/16 – 2018/19 only. The two academic outcomes of interest were the retention and completion outcomes (see Section 4.2 for outcome definitions). A reminder that contextual offer students were defined as those from SIMD Quintiles 1 and 2 who achieved below the standard entry requirements for the degree programme they applied for, and were defined as standard offer students otherwise (see Section 4.5 for more details on the *Offer Received* variable). There were 1,049 students classified as contextual offer (13.9%) in the **Contextual Offer subset** (Table 8.1). In Section 5.1 it was established that the Faculty of HaSS also had the largest proportion of SIMD Quintile 1 and 2 students of all the four faculties. It was also shown that a small number of students from the Faculties of HaSS, Business and Science had missing entry requirements data. The removal of these data from the **Contextual Offer subset** means that model estimates for Faculty may contain a negligible amount of bias. Were these observations to be accounted for however, we do not expect this to substantially change the model estimates.

8.4 Methods

Given there was no clear best-fitting model to the data from Chapter 7, it was decided to use Modified Poisson regression due to its derivation of risk-ratios and ease of fit compared to the Log-Binomial. Logistic regression was ruled out due

Table 8.1: Count/proportion of students' SIMD Quintile versus whether ot not they met the standard entry requirements at the point of application. Contextual Offer students (highlighted) are those who attained below the standard entry requirements and were from SIMD Quintiles 1 and 2. Proportions rounded to 2 decimal places.

SIMD Quintile	**Met Standard Entry Requirements?**		**Sum (Prop.)**
	Equal to or Above	**Below**	
1	346 (0.05)	471 (0.06)	**817 (0.11)**
2	538 (0.07)	578 (0.08)	**1116 (0.15)**
3	813 (0.11)	396 (0.05)	**1209 (0.16)**
4	1200 (0.16)	398 (0.05)	**1598 (0.21)**
5	2158 (0.28)	636 (0.08)	**2794 (0.37)**
Sum (Prop.)	**5055 (0.67)**	**2479 (0.33)**	**7534 (1.00)**

to its sensitivity to changes in the model specification (see Section 6.4.3) and its derivation of odds-ratios, which are frequently misinterpreted as risk-ratios (see Section 6.1).

Two Modified Poisson regression models were fit to the data and denoted Models 1a and 1b, where "a" represented the model for the retention outcome, and "b" for the completion outcome. Models 1a and 1b were used to determine the adjusted risk-ratio of retention/completion for contextual versus standard offer students, whilst controlling for *Academic Cohort*, *Faculty*, *Sex* and *Ethnicity* which are common control variables in the literature [76, 78].

This was followed by two additional Modified Poisson regression models, denoted Models 2a and 2b, which had the same model specification as before but replaced *Offer Received* with *Best 5 Highers Appl. Points* and a new variable – *SIMD Group*. *SIMD Group* was a binary stratification of *SIMD Quintile* which grouped together SIMD Quintiles 1-2 and SIMD Quintiles 3-5. Models

2a and 2b were used to determine whether or not the adjusted risk-ratio of retention/completion for *Best 5 Highers Appl. Points* and *SIMD Group* were comparable to the estimates of *Offer Received* in Models 1a and 1b.

The average-adjusted probabilities of success for standard and contextual offer students were calculated from Models 1a and 1b. This was found by predicting the probability of retention/completion for each student in the **Contextual Offer subset**, once controlling for other variables in the model fits, and then taking the mean within each group. The average adjusted probabilities for students in each *SIMD Group* were also calculated from Models 2a and 2b. The 95% confidence intervals for each of these were constructed using the robust standard errors derived from the Modified Poisson regression fits.

For each fit, the reference groups for the categorical explanatory variables were: *Academic Cohort* – "2015/16", *Ethnicity* – "White", *Faculty* – "Engineering", *Offer Received* – "Standard Offer", *Sex* – "Male", *SIMD Group* – "Quintiles 3-5". *Best 5 Highers Appl. Points* was mean-centred in the model fits, where in the **Contextual Offer subset** the mean was 19.26 points (the equivalent of AAAAA-AC at Higher). The p-values from Wald's tests on each coefficient were derived using $\alpha = 0.05$ as the significance level.

8.4.1 Software Used for Analysis

All analyses were conducted using the statistical software `R (version 4.3.1)` [140]. Poisson regression models were fit using the `glm()` function from the stats package [140]. Robust variances for the Modified Poisson Regression model were derived using the `sandwich (3.1-0)` package [118]. Average adjusted probabilities were derived using the `avg_predictions()` function within the `marginaleffects (0.23.0)` package [146]. Additional packages for general data cleaning and visualisations were used from the `tidyverse (2.0.0)` [142].

8.5 Results - Models 1a and 1b Estimates

The derived risk-ratios from Models 1a (Table 8.2) showed the adjusted risk-ratios for *Offer Received* and other control variables, with respect to the retention outcome. Contextual offer students were 8.3% [95% CI: 5.6%, 10.9%] less likely to be retained compared to standard offer students. Students from the Faculty of Science were 4.9% [95% CI: 2.6%, 7.1%] less likely to be retained compared to students from the Faculty of Engineering. Students from the faculties of Business and HaSS did not appear to be more or less likely to be retained compared to students from the Faculty of Engineering. Students from *Academic Cohort* 2018/19 were 2.4% [95% CI: 0.2%, 4.5%] less likely to be retained compared to students from the 2015/16 cohort. Males and females did not appear to have different rates of retention compared to one another, and similarly between whites and ethnic-minorities.

The derived risk-ratios from Models 1b (Table 8.3) showed similar adjusted risk-ratios but with respect to the completion outcome. Contextual offer students were 18.6% [95% CI: 14.5%, 22.5%] less likely to complete their degree compared to standard offer students. Students from the Faculty of Business were 5.4% [95% CI: 1.9%, 9.0%] more likely to complete their degree compared to students from the Faculty of Engineering. In contrast, students from the faculties of HaSS

Table 8.2: Modified Poisson Model 1a - Comparing standard and contextual offer students on progression.

Variables	P-values	Coefficients (Robust S.E.)	Adjusted Risk-Ratio (95% C.I.)
(Intercept)		-0.082 (0.010)	0.921 (0.904,0.938)
Contextual Offer (vs Std.)	<0.001	-0.086 (0.015)	0.917 (0.891,0.944)
Business (vs Engineering)	0.207	0.013 (0.011)	1.013 (0.993,1.035)
HaSS (vs Engineering)	0.202	-0.014 (0.011)	0.986 (0.965,1.008)
Science (vs Engineering)	<0.001	-0.050 (0.012)	0.951 (0.929,0.974)
2016/17 Cohort (vs 2015/16)	0.419	-0.009 (0.011)	0.991 (0.971,1.012)
2017/18 Cohort (vs 2015/16)	0.856	-0.002 (0.011)	0.998 (0.977,1.019)
2018/19 Cohort (vs 2015/16)	0.036	-0.024 (0.011)	0.976 (0.955,0.998)
Female (vs Male)	0.432	0.007 (0.009)	1.007 (0.990,1.024)
Ethnic-minority (vs White)	0.543	0.009 (0.015)	1.009 (0.980,1.040)

and Science were 4.8% [95% CI: 1.3%, 8.2%] and 14.8% [95% CI: 11.3%, 18.3%] less likely to complete their degree compared to students from the Faculty of Engineering. Females were 6.0% [95% CI: 3.1%, 9.0%] more likely to complete their degree compared to males. There did not appear to be different rates of completion between each of the *Academic Cohorts* when compared to the 2015/16 cohort. Similarly, there did not appear to be any differences in completion rates between whites and ethnic-minorities.

The deviance residuals from Models 1a and 1b (plots (I) and (II) in Figure 8.1, respectively) appear to be centred around the mean value with no influential observations observed. Hence, this suggests that both models adequately fit the data.

Table 8.3: Modified Poisson Model 1b - Comparing standard and contextual offer students on completion.

Variables	P-values	Coefficients (Robust S.E.)	Adjusted Risk-Ratio (95% C.I.)
(Intercept)		-0.269 (0.017)	0.764 (0.739,0.789)
Contextual Offer (vs Std.)	<0.001	-0.206 (0.025)	0.814 (0.775,0.855)
Business (vs Engineering)	0.002	0.053 (0.017)	1.054 (1.019,1.090)
HaSS (vs Engineering)	0.008	-0.049 (0.019)	0.952 (0.918,0.987)
Science (vs Engineering)	<0.001	-0.161 (0.021)	0.852 (0.817,0.887)
2016/17 Cohort (vs 2015/16)	0.347	0.017 (0.018)	1.017 (0.981,1.055)
2017/18 Cohort (vs 2015/16)	0.078	0.032 (0.018)	1.033 (0.996,1.070)
2018/19 Cohort (vs 2015/16)	0.164	0.026 (0.019)	1.026 (0.989,1.065)
Female (vs Male)	<0.001	0.058 (0.014)	1.060 (1.031,1.090)
Ethnic-minority (vs White)	0.410	0.021 (0.026)	1.021 (0.971,1.074)

8.6 Results - Models 2a and 2b Estimates

The derived risk-ratios from Models 2a (Table 8.4) showed the adjusted risk-ratios for *Best 5 Highers Appl. Points* and *SIMD Group*, as well as other control variables, with respect to the retention outcome. For each additional point increase over the mean *Best 5 Highers Appl. Points*, a student was 0.9% [95% CI: 0.6%, 1.2%] more likely to be retained at the end of first year (Table 8.4). This meant that for each additional A grade a student achieved at Higher in S5, they were 2.7% more likely to be retained. Students from SIMD Quintiles 1 and 2 were 5.2% [95% CI: 3.1%, 7.1%] less likely to be retained compared to students from SIMD Quintiles 3-5. Students from the Faculty of Science were 3.0% [95% CI: 0.5%, 5.3%] less likely to be retained compared to students from the Faculty of Engineering. Students from the faculties of Business and HaSS did not appear to be more or less likely to be retained compared to students from the Faculty of Engineering. Students from *Academic Cohort* 2018/19 were 2.8% [95% CI: 0.6%,

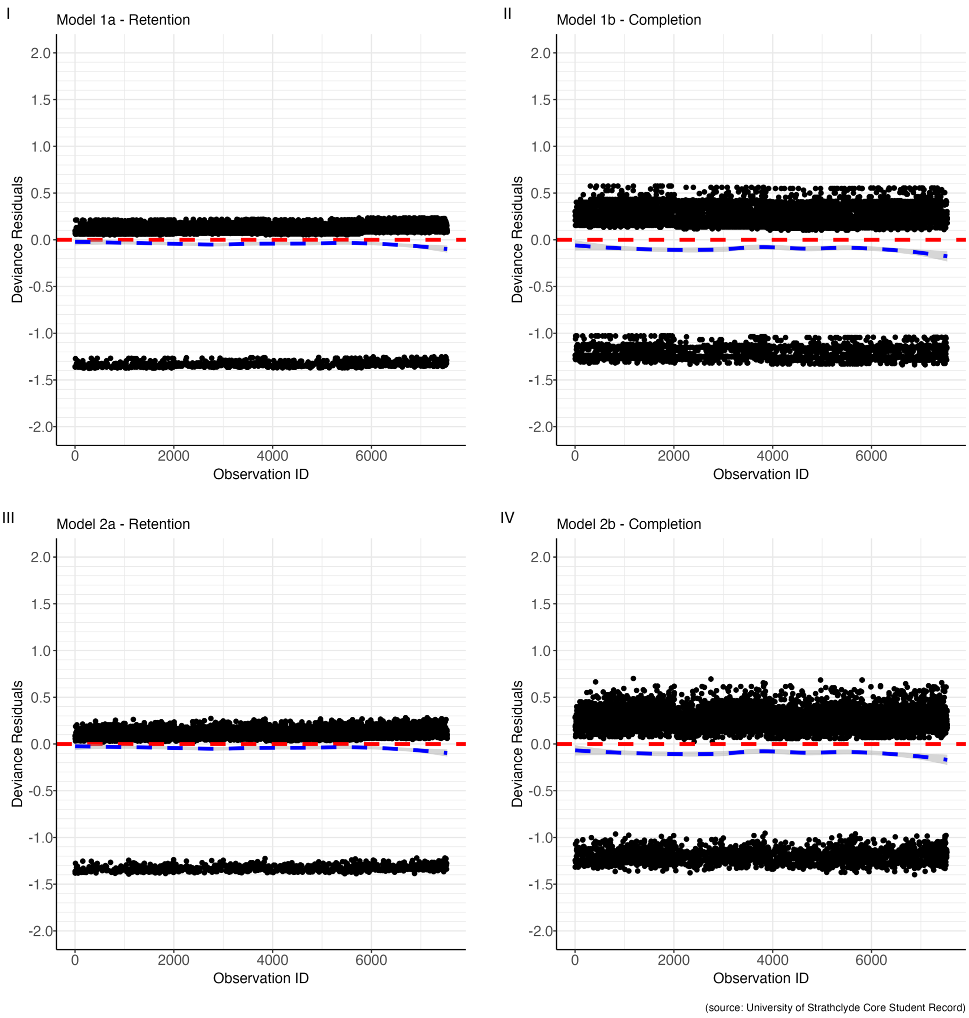


Figure 8.1: Deviance residuals for Modified Poisson Models.

5.0%] less likely to be retained compared to students from the 2015/16 cohort. Males and females did not appear to have different rates of retention compared to one another, and similarly between whites and ethnic-minorities.

The derived risk-ratios from Models 2b (Table 8.5) showed the similar adjusted risk-ratios but with respect to the completion outcome. For each additional point increase over the mean *Best 5 Highers Appl. Points*, a student was 2.6% [95% CI: 2.0%, 3.2%] more likely to complete their degree within four years (Table

8.5). This meant that for each additional A grade a student achieved at Higher in S5, they were 7.8% more likely to complete their degree. Students from SIMD Quintiles 1 and 2 were 9.5% [95% CI: 6.3%, 12.6%] less likely to complete their degree compared to students from SIMD Quintiles 3-5. Students from the Faculty of Business were 6.5% [95% CI: 3.0%, 10.2%] more likely to complete their degree compared to students from the Faculty of Engineering. In contrast, students from the Faculty of Science were 10.0% [95% CI: 6.0%, 13.8%] less likely to complete their degree compared to students from the Faculty of Engineering. Students from the Faculty of HaSS did not appear to be more or less likely to complete their degree compared to the Faculty of Engineering. Females were 6.2% [95% CI: 3.2%, 9.2%] more likely to complete their degree compared to males. There did not appear to be different rates of complete between each of the *Academic Cohorts* when compared to the 2015/16 cohort. Similarly, there did not appear to be any differences in completion rates between whites and ethnic-minorities.

Similar to Models 1a and 1b, the deviance residuals from Models 2a and 2b appear to be centred around the mean value with no influential observations observed (plots (III) and (IV) in Figure 8.1, respectively). The deviance residuals appeared to linearly decrease with each increase in the *Best 5 Highers Appl. Points* variable, suggesting that the assumptions of log-linearity for Models 2a and 2b were satisfied here (Figure 8.2). There were also no predicted probabilities greater than 1, which would have indicated inadequate fit (Figure 8.3).

Table 8.4: Modified Poisson Model 2a - Measuring effect of prior attainment and SIMD Quintile on progression.

Variables	**P-values**	**Coefficients (Robust S.E.)**	**Adjusted Risk-Ratio (95% C.I.)**
(Intercept)		-0.088 (0.010)	0.916 (0.898,0.934)
Best 5 Highers at Application (Points)	<0.001	0.009 (0.002)	1.009 (1.006,1.012)
SIMD 20-40 (vs SIMD 60-100)	<0.001	-0.053 (0.011)	0.948 (0.929,0.969)
Business (vs Engineering)	0.121	0.016 (0.011)	1.017 (0.996,1.038)
HaSS (vs Engineering)	0.726	-0.004 (0.012)	0.996 (0.973,1.019)
Science (vs Engineering)	0.018	-0.030 (0.013)	0.970 (0.947,0.995)
2016/17 Cohort (vs 2015/16)	0.269	-0.012 (0.011)	0.988 (0.968,1.009)
2017/18 Cohort (vs 2015/16)	0.630	-0.005 (0.011)	0.995 (0.974,1.016)
2018/19 Cohort (vs 2015/16)	0.012	-0.029 (0.011)	0.972 (0.950,0.994)
Female (vs Male)	0.343	0.008 (0.009)	1.008 (0.991,1.025)
Ethnic-minority (vs White)	0.190	0.020 (0.015)	1.020 (0.990,1.051)

Table 8.5: Modified Poisson Model 2b - Measuring effect of prior attainment and SIMD Quintile on completion.

Variables	**P-values**	**Coefficients (Robust S.E.)**	**Adjusted Risk-Ratio (95% C.I.)**
(Intercept)		-0.292 (0.017)	0.747 (0.722,0.773)
Best 5 Highers at Application (Points)	<0.001	0.026 (0.003)	1.026 (1.020,1.032)
SIMD 20-40 (vs SIMD 60-100)	<0.001	-0.100 (0.018)	0.905 (0.874,0.937)
Business (vs Engineering)	<0.001	0.063 (0.017)	1.065 (1.030,1.102)
HaSS (vs Engineering)	0.350	-0.018 (0.019)	0.982 (0.946,1.020)
Science (vs Engineering)	<0.001	-0.105 (0.022)	0.900 (0.862,0.940)
2016/17 Cohort (vs 2015/16)	0.680	0.008 (0.018)	1.008 (0.972,1.044)
2017/18 Cohort (vs 2015/16)	0.245	0.021 (0.018)	1.021 (0.986,1.059)
2018/19 Cohort (vs 2015/16)	0.572	0.011 (0.019)	1.011 (0.974,1.049)
Female (vs Male)	<0.001	0.060 (0.014)	1.062 (1.032,1.092)
Ethnic-minority (vs White)	0.069	0.047 (0.026)	1.048 (0.996,1.102)

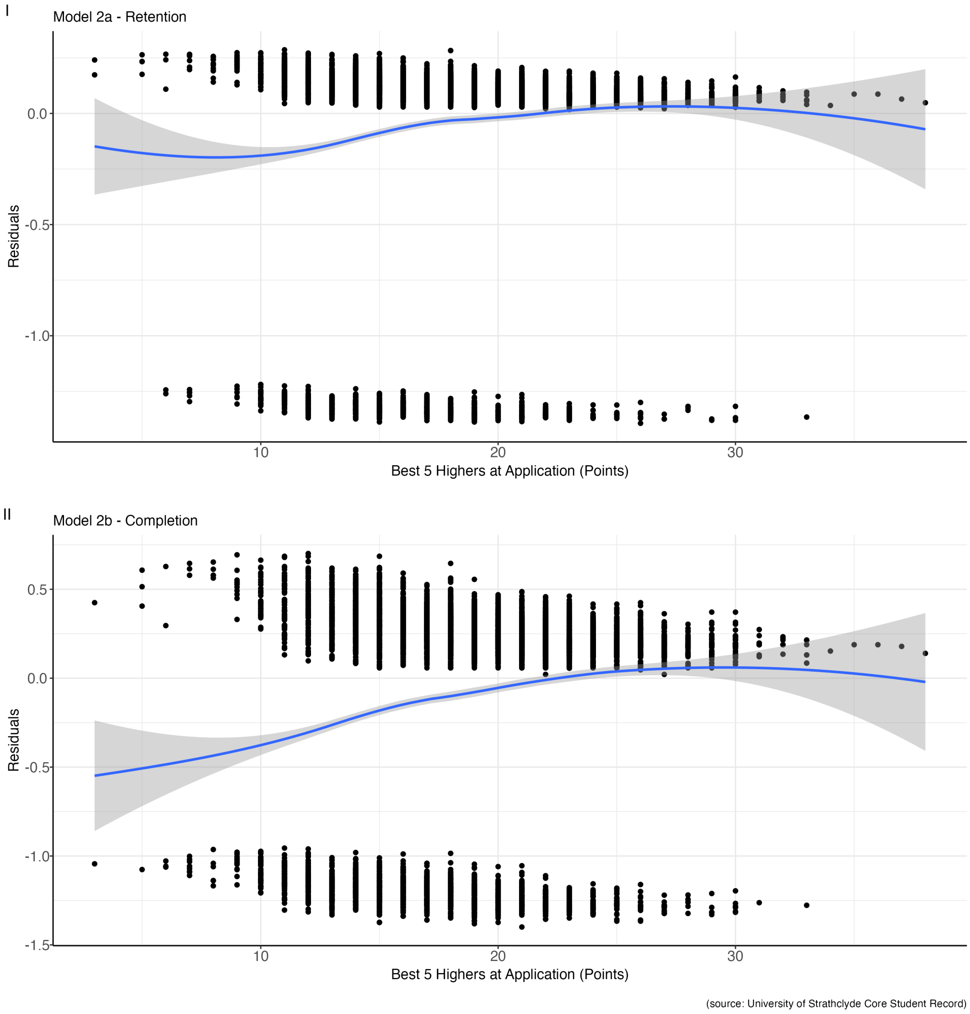


Figure 8.2: Deviance residuals versus Best 5 Highers Appl. Points for Models 2a and 2b.

8.7 Results - Average Predicted Probabilities

Using each Modified Poisson regression fit, the average adjusted probability of retention/completion for standard and contextual offer students (Models 1a and 1b), as well as for each SIMD Quintile (Models 2a and 2b), were calculated.

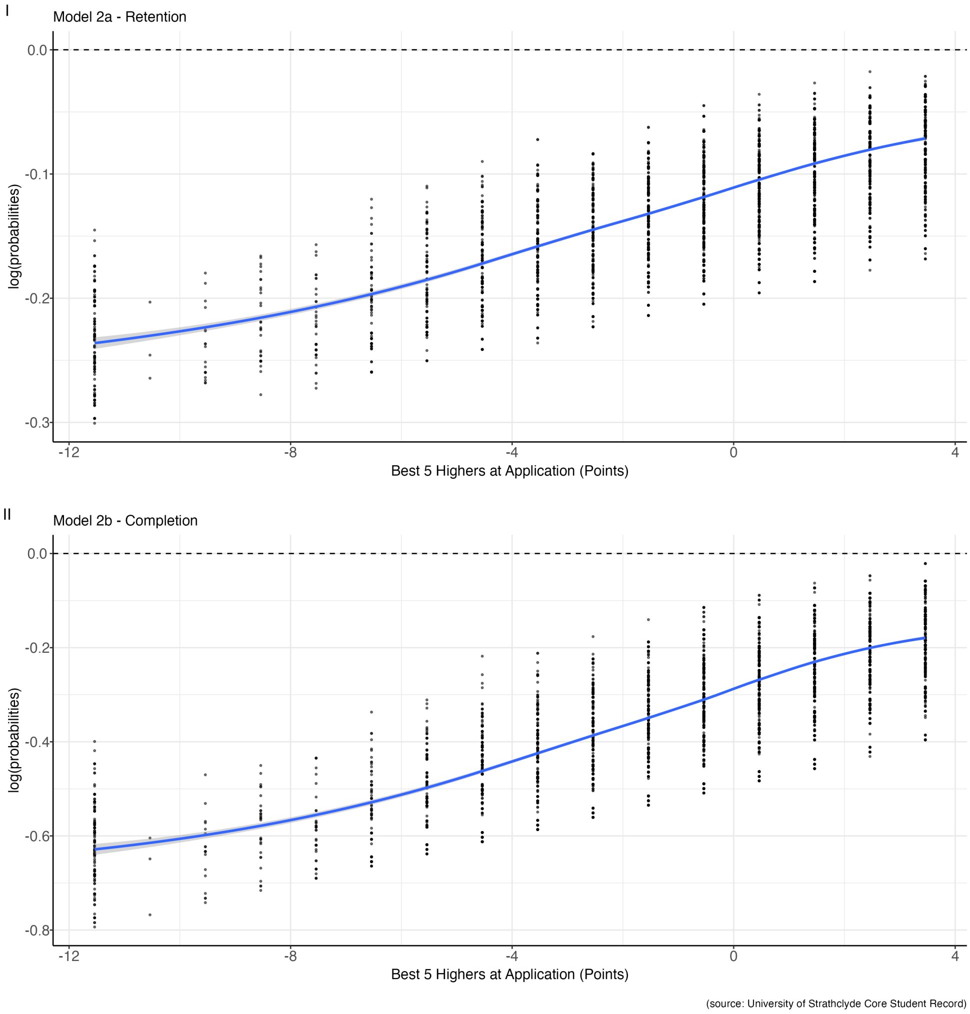


Figure 8.3: Predicted probabilities of retention/completion versus Best 5 Highers Appl. Points for Models 2a and 2b.

Models 1a and 1b showed that once controlling for *Academic Cohort*, *Sex* and *Ethnicity*; contextual offer students had an 82.7% [95% CI: 80.4%, 84.9%] chance of retention and a 62.2% [95% CI: 59.3%, 65.2%] chance of completion. Meanwhile, standard offer students had an 90.6% [95% CI: 89.9%, 91.3%] chance of retention and a 77.4% [95% CI: 76.4%, 78.4%] chance of completion. This is equivalent to a gap between standard and contextual offer students of 7.9 percentage-points for retention and 15.2 percentage-points for completion.

Similarly, Models 2a and 2b showed that, once controlling for *Academic Cohort*, *Sex*, *Ethnicity* and *Best 5 Highers Appl. Points*, those from SIMD Quintiles 1-2 had an 84.8% [95% CI: 83.2%, 86.4%] chance of retention and a 67.3% [95% CI: 65.2%, 69.3%] chance of completion. This contrasted with those from SIMD Quintiles 3-5 who had a 91.1% [95% CI: 90.4%, 91.9%] chance of retention and a 78.1% [95% CI: 77.0%, 79.2%] chance of completion. This constituted a gap between SIMD Quintiles 1 and 5 of 6.3 percentage-points for retention and 10.8. percentage-points for completion.

8.8 Discussion

The aims of this analysis were twofold: (i) to determine contextual offer students' controlled chances of success compared to standard offer students, and (ii) to determine the average adjusted probability of success for contextual offer students. Based on the modelled definitions of success, deprivation, prior attainment and the offers received, contextual offer students were less likely to be successful at university compared to their standard offer peers. Contextual offer students' predicted chances of retention exceeded the "high-bar" of 80%+ argued for by Boliver et al. [52]. These chances are, however, lower than the University of Strathclyde's benchmark of 90-95% retention for all students. Most contextual offer students were predicted to successfully complete their degree (around 62.2%), though there exists no benchmark for comparison.

The models also identified a gap in the retention and completion rates of those from SIMD Quintiles 1-2 versus Quintiles 3-5, even when they have the same levels of prior attainment at Higher in S5. These findings highlight that while the University has commendably achieved its Widening Access target on the equal representation of entrants from lower SIMD Quintiles [68], achieving outcome equality for these students is still in progress. Models also suggested that prior attainment from S5 was more strongly associated with completion than retention. This appears to be reflected in the gap between the adjusted predicted probabilities of standard and contextual offer students, which is higher for the completion outcome than the retention outcome.

8.8.1 Implications on Widening Access

These results could be used to argue either for or against Widening Access interventions. To expect contextual offer students to achieve at a level similar to their standard offer peers at university is perhaps unrealistic. This is because, by definition, contextual offer students are very likely to have lower prior attainment when they commence their degree and come from areas with higher levels of deprivation (as defined by SIMD), both of which are negatively associated with a successful outcome in our models. Furthermore, the disadvantage that makes a student eligible for a contextual offer may not completely disappear once the student starts university; it may persist until they graduate and join the workforce. For example, the experiences of Widening Access students at an elite Scottish university suggest that they are more likely to take on full-time or part-time jobs to support themselves and/or family members, do not always have the same levels of support at home, and can struggle with a sense of belonging at university and amongst their peers [147]. It should also be noted that without a contextual offer, many if not all of these students would not have been admitted to the University. Given such circumstances, and the fact that, at this university, around 217 contextual offer students were retained each academic session and around 163 contextual offer students per academic session completed their degree, their achievements are perhaps understated. Furthermore, without a contextual offer these students may have instead enrolled in colleges, taken apprenticeships or non-professional occupations. By entering University, this means that more spaces in these areas can be given to young people who would not have met either the contextual nor standard offers for university.

The policy of contextual offers and Widening Access targets have had a positive impact on students who may not have otherwise had the opportunity to attend higher education. Yet if the aim of Widening Access is to progress towards equality of outcome [7], then universities may need to take a more active role in supporting students admitted via these policies. Setting targets for the proportion of Widening Access students who successfully complete a degree programme, in addition to those currently set on access to higher education and retention for all 1st year to 2nd year students, could help with this. Such targets may cultivate more trust from the public that Widening Access policies are providing measurable impact.

8.8.2 Limitations and Future Analyses

Rather than using UCAS tariff points to measure prior attainment, a single-point based system was adopted, based on attainment from Higher in S5 (see Sections 4.4 and 4.5 for more details). This was done to improve the interpretability of model fits and because the analysis only considered Scottish school-leavers. Those not from SIMD Quintiles 1 or 2 but satisfying the other criteria for a contextual offer will not be correctly classified as standard offer students in this analysis. If such students have higher/lower rates of retention or completion, then they will bias the results such that the gap between standard and contextual offer students may appear larger/smaller than the true gap. Without access to data on the other eligibility criteria, the true rate of correct classification is unknown. However, anecdotal evidence from the University's Widening Access Team suggests that the majority of students who receive a contextual offer come from SIMD Quintiles 1 or 2, and that there is significant overlap between those from SIMD Quintiles 1 and 2 and the other eligibility criteria. Thus, there is confidence that most students will be correctly classified as either standard or contextual offer. The analysis presented here could be improved with direct access to applicant data. This would remove the need for the proxy indicator.

The outcomes modelled by the fitted regression models were based on defined time periods; retention was measured at the end of one year and completion at the end of four years. Some students may have had periods of suspension or may have repeated one or more years and thus results may underestimate final success rates. If these behaviours are more prevalent for contextual offer students, then the observed gaps in outcomes may be less pronounced. The effect sizes of some control variables also differed between the models for retention when compared to the models for completion. While some overlap is expected since those who failed to be retained were also counted as failing to complete their degree, it may also be the case that some factors are more associated with one outcome versus the other. These discussion points could be addressed in future analyses by fitting time-to-event models that instead track if and when a student drops out of university [148].

Students from *Academic Cohorts* 2016/17 - 2018/19 were affected at some point in their registration by the COVID-19 pandemic which began in March 2020. Any potential effects from the pandemic should be controlled for in the model fits via the *Academic Cohort* variable. It would be of interest to compare the academic outcomes of students pre- and post-pandemic, for example using the *No Detriment Retention* or *No Detriment Completion* variables.

Data on the university-level attainment of students was not available for this study. Individual module attainment is a key factor in the overall academic success rates. It would also distinguish between those who do not progress/complete for personal versus academic reasons. Comparing attainment between contextual and standard offer students could thus provide further insight into the reasons for the gaps between the two groups of students.

A small number of students whose Ethnicities were unknown or missing were removed (see Section 3.5). Sensitivity analyses found that model estimates remained unchanged when these observations were included, their removal therefore has no impact on the interpretation of the model results.

The results presented here are for a single Scottish university and are not necessarily representative of other universities across Scotland or the rest of the UK. The analysis does not consider students registered with Widening Access specific programmes, only standard or contextual offer students on traditional degree programmes. The data are also cross-sectional and so the associations highlighted should not be interpreted as causal. Future analyses could endeavour to combine results across universities to give a more holistic view of the success rates of contextual offers students across the country. Of particular interest would be an examination of the four other Glasgow institutions who have already achieved current Widening Access targets [68], though in-practice gaining access to this data would be a difficult endeavour.

8.9 Conclusion

The results presented in this chapter add to the evidence on the associations between area-level deprivation and prior attainment from secondary education with higher education outcomes. The work also estimates, for one university, the size of the attainment gap between standard and contextual offer students. The analysis is, to the best of our knowledge, the first Scottish study to do so using data from the post-Commission on Widening Access [3] period. Widening Access policies have resulted in commendable progress on admission rates for disadvantaged students into higher education. Going forward, it is important that these policies should focus not only on admissions but also on provision of support and targets focused on the academic outcomes of students admitted via these policies. This may cultivate more trust from the public that Widening Access policies are providing measurable impact.

Chapter 9

Modelling Student Drop-outs Over Time

The regression models applied in Chapters 7 and 8 had various issues when fit to their respective subsets of the **School-leavers dataset**. Each of these issues were related to the outcomes variables being so common ($> 10\%$). These models also did not account for students who may have achieved a successful outcome over longer than expected periods of time. For example, completing a Bachelor's with Honours degree programme in five years rather than the typical four years. This chapter aims to address these limitations by applying survival models to estimate the risk of dropping out of higher education, an outcome which should be sufficiently rare (around 9% - see Section 5.3). This chapter will compare the fits of traditional regression and survival models and make a determination on the preferred technique for measuring academic outcomes in higher education.

9.1 Examples of Survival Models in the Literature

The application of survival models to estimate the risk of dropping out of university over time is also referred to as the “student drop-out problem”. There is widespread and international interest in the student drop-out problem, for example there are studies from Europe [148–151], North America [152–156], Latin America [157, 158], Asia [159], and Oceania [160]. These studies tend to be concerned with determining whether drop-outs over time are more prevalent within some groups of students than others [149, 152, 159] or predicting if and when students drop-out [151, 154, 155, 158].

In 2004, Arulampalam, Naylor, and Smith [148] was the first known UK study that applied survival analysis methods to the student drop-out problem. They were interested in how the drop-out of medical school students could affect the supply of future NHS staff. The study considered a large number of variables including prior attainment (in A-levels, Highers and other qualifications), high-school attended, sex, age, nationality, parental social class and whether or not the father was a doctor. They found that variables related to academic “preparedness”, such as prior attainment in A-levels/Highers and subject choice, had the largest impact on drop-out rates. Having a father who was a doctor also had a significantly positive effect on survival time, however, there was “little” evidence that the parent’s social class background or the school attended had an impact on student drop-out rates. They cautioned that Widening Access students (on medical programmes) could have low probabilities of success at university should they not be adequately supported while at university. They also warned that the type of school a student attended (state, grammar, independent, etc.) may not be an appropriate measure for deciding whether or not to award a reduced offer of entry to students. This research was, in many ways, ahead of its time given the recent and similar findings from Boliver and Gorard (2017-2022) [51–55] on appropriate contextual indicators.

The majority of the literature highlighted previously used discrete time-to-event survival methods [148–150, 153, 156, 159, 160], with some examining continuous time-to-event survival methods [151, 152, 155, 158]. A review of the literature on student drop-outs conducted by Kim et al. [156] in 2018, found that the most popular survival method was the Cox-Proportional Hazards (CPH) model, which assumes that follow-up time is continuous. This is despite student drop-out problems typically involving discrete follow-up time intervals, either every academic year or every semester. Researchers [103, 156] have made strong recommendations that studies involving educational data should appropriately model follow-up time as discrete, rather than continuous. When follow-up time is misspecified in this way, it can lead to biased estimates [156]. This is particularly true when the hazard at any given time point is high, the sample size is large, or the proportion of censored observations is low [156]. Additionally, CPH models can run into issues with tied survival times [103, 156], which is common when the follow-up time occurs in discrete intervals, though there are methods for addressing this [100, p. 77].

9.2 Aim of the Chapter

The aim of this chapter is to determine whether or not discrete and/or continuous survival methods are more appropriate fits to the data when compared to traditional regression methods. The association between students' prior attainment/socio-economic background and their risk of dropping out, will also be examined.

9.3 Data

This chapter used the **Drop-outs subset** (derived in Section 3.9) in both a person-period and person-level format. This subset only considered students from the 7 academic cohorts 2012/13-2018/19 such that all students had the chance to complete their Bachelor's with Honours degree (typically four years in duration). The subset contained $13,319$ school-leavers, roughly 9% of which eventually dropped out (Table 9.1).

Both the person-level and person-period formats were required to support the fitting of discrete and continuous models (see Section 3.6). The outcome of interest was the time until drop out occurred (*Ever Dropped Out*/*Drop-out Status*). A reminder that observations were censored at the relevant point in time if they had either: (1) completed their Bachelor's with Honours degree, (2) continued into their 6th registration year (*Reg. Year Count* > 5), or (3) had not dropped out by the end of the observation period (academic session 2021/22). See Section 6.9 for more details. An example of this censoring setup is visualised and explained in Figure 6.2. No time-dependent covariates were considered for this analysis. Therefore every time-dependent variable was fixed at the value that was true in the student's first instance. The probability of drop-out peaked in the first academic session (6%) and subsequently decreased over time (Table 9.1). This is in agreement with studies conducted at other international higher education institutions [148, 152, 159]. Of all the students to have ever dropped out (1165), 70% did so in their first academic session (Table 9.1).

Table 9.1: Life table of the school-leavers person-period dataset. *Anyone registered longer than 5 academic sessions were censored and added to the 5th academic session count for censored observations.

Academic Sessions	**Total Students**	**Censored Count**	**Dropped-out**	
			Count	**Prop.**
1	13319	12501	818	0.0614
2	12432	12213	219	0.0176
3	11875	11803	72	0.0061
4	11443	11402	41	0.0036
5 or more*	3980	3965	15	0.0038

9.4 Methods

Similar to Chapter 7, Logistic, Modified Poisson and Log-Binomial regression models were fit to the **Drop-outs subset**, though using *Ever Dropped Out* as the outcome variable and *Sex*, *Ethnicity*, *Faculty*, *SIMD Quintile* and *Prior Attainment Quintile* as explanatory variables. The regression model estimates were interpreted and goodness-of-fit assessed by examining deviance residuals for influential observations. There was no need to assess the linearity assumption here since all explanatory variables were categorical.

For the survival analysis, the data were first explored by finding estimates of the survival function, $\hat{S}(t)$, and the baseline hazard function, $\hat{h_0}(t)$, whilst controlling for a single explanatory variable. The survival functions were estimated using Kaplan-Meier curves (see Section 6.11 for details) which controlled for: *Sex*, *Ethnicity*, *Faculty*, *SIMD Quintile*, *Prior Attainment Quintile*. To create these plots, the explanatory variables had to be categorical, hence why *Prior Attainment Quintile* was used for this analysis instead of *Prior Attainment Points*. The baseline hazard functions were estimated by fitting several Logit Discrete Time-to-Event (DTE) models (see Section 6.12 for details) which treated time until drop-out, T, as a categorical explanatory variable. These estimates controlled for the same explanatory variables as before. The proportional-hazards assumption for each explanatory variable was informally assessed by examining the complementary log-log (cloglog) transformation of the estimated baseline hazard functions. Survival and hazard estimates were also created for the *Offer Received*, *Urban/Rural Status*, *Disability Status* explanatory variables, though these were not used in the multivariable regression or survival models.

Discrete survival models were applied to the **Drop-outs Person-period subset** while continuous models were applied to the **Drop-outs Person-level subset**. It was of interest to determine whether or not the continuous survival models produced sensible fits and estimates even when the time-to-event variable, T, was clearly discrete (drop-outs occurred every semester - see Section 6.9.4 for more details). Other studies have also applied continuous models when the follow-up time was clearly discrete [151, 152, 155, 158].

Discrete survival models require an explicit form for the baseline hazard function, $h_0(t)$, to be provided. To explore viable estimates of the baseline hazard function, several intercept-only Logit DTE models were fit to the data which assumed different polynomial relationships between the risk of drop-out over time, namely: linear, quadratic, and cubic. These were compared to an intercept-only model which assumed time until drop-out was a categorical variable, denoted the "General Logit DTE" model. From this comparison an appropriate approximation of the General function was determined.

Two multivariable Logit DTEs were then fit to the data: one which used the most appropriate approximation of the baseline hazard function (whether linear, quadratic or cubic), and another General Logit DTE. Each model controlled for the same effects as those seen in the regression models. Two continuous survival models were also fit to the data: a Cox-Proportional Hazards (Std. CPH) model and a Parametric Weibull (PW) model. The Efron approximation was used to deal with tied survival times. A follow-up Stratified Cox-Proportional Hazards (Strat. CPH) was fit to the data which stratified by the variables which violated the proportional hazards assumption; the estimated effects before and after stratification were compared to detect any changes. Finally, all regression and survival models were compared on their estimated effects and goodness-of-fit to the data. A final determination was made on the appropriateness of both modelling frameworks.

9.4.1 Software Used for Analysis

All analyses were conducted using the statistical software `R (version 4.3.1)` [140] and regression models were fit using the `glm()` function [140]. Robust variances for the Modified Poisson Regression model were derived using the `sandwich (3.1-0)` package [118]. Survival methods were applied using the following `R` packages: `survival (3.5-5)` [120, 121] and `survminer (0.4.9)` [161] for the fitting of survival models and plotting survival curves. Additional packages used for general cleaning and visualisations included: `tidyverse (2.0.0)` [142], `ggfortify (0.4.16)` [162], `patchwork (1.1.3)` [143], and `xtable (1.8-4)` [144].

9.5 Results - Regression Modelling

All regression models have similar odds/risk-ratio estimates for each of the covariates in the model (Table 9.2). This is likely because the the outcome variable - drop-out status - has low prevalence in the dataset (roughly 9% - Table 9.2). Since this is considered sufficiently rare (< 10%), the odds-ratios from the Logistic models could be interpreted as estimates of risk-ratios. However, the Logistic estimates still exaggerated the Modified Poisson estimates by at most 11% (ignoring the intercept term - see Table 9.3). According to the Modified Poisson fit, students from Prior Attainment Quintile 5 were 56.7% [95% CI: 44.2%, 76.4%] less likely to drop-out than students from Prior Attainment Quintile 3, while students from Prior Attainment Quintile 1 were 116.0% [95% CI: 83.8%, 153.8%] more likely. Students from SIMD Quintile 5 were 38.1% [95% CI: 26.5%, 47.8%] less likely to drop-out compared to their peers from SIMD Quintile 1. Students from the Faculty of Business were 28.8% [95% CI: 23.0%, 41.7%] less likely to drop-out compared to students from the Faculty of Engineering. Females were 17.1% [95% CI: 6.5%, 26.4%] less likely to drop-out compared to males, and students from ethnic-minority groups were 30.7% [95% CI: 10.7%, 46.2%] less likely to drop-out compared to white students. The deviance residuals from each regression model indicate that there are no influential observations (Figure 9.1). The linearity assumption is not assessed here since there are no continuous covariates.

Table 9.2: Comparison of the exponentiated regression model estimates (and 95% CIs) for the drop-out outcome.

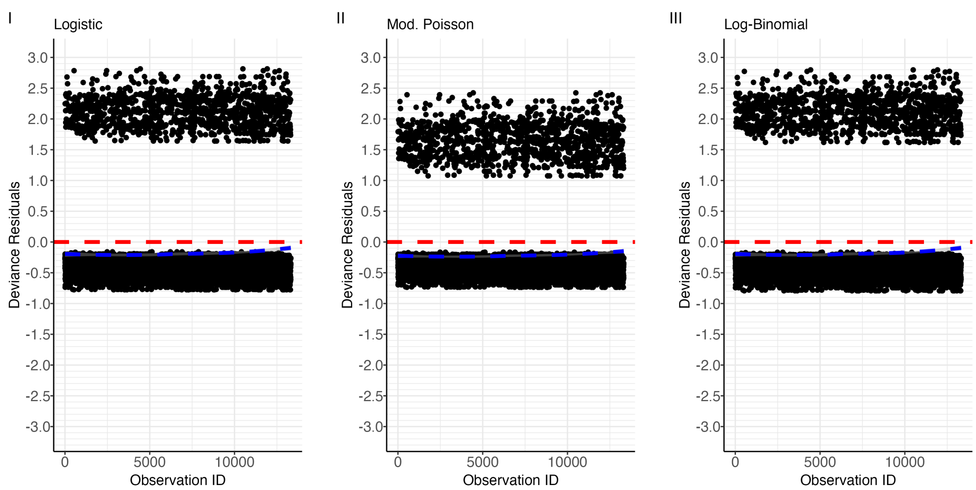
Variables	**Regression Models**		
	Logistic	**Mod. Poisson**	**Log-Binomial**
(Intercept)	0.146 (0.114,0.188) [***]	0.125 (0.101,0.156) [***]	0.125 (0.100,0.155) [***]
Prior Att. Quintile 1 (vs 3)	2.402 (2.003,2.879) [***]	2.160 (1.838,2.538) [***]	2.163 (1.839,2.545) [***]
Prior Att. Quintile 2 (vs 3)	1.411 (1.162,1.712) [***]	1.368 (1.147,1.632) [***]	1.372 (1.150,1.637) [***]
Prior Att. Quintile 4 (vs 3)	0.844 (0.679,1.048)	0.854 (0.698,1.045)	0.855 (0.699,1.046)
Prior Att. Quintile 5 (vs 3)	0.413 (0.317,0.539) [***]	0.433 (0.336,0.558) [***]	0.433 (0.336,0.558) [***]
SIMD Quintile 2 (vs 1)	0.758 (0.610,0.941) [*]	0.791 (0.657,0.951) [*]	0.799 (0.665,0.960) [*]
SIMD Quintile 3 (vs 1)	0.734 (0.591,0.910) [**]	0.769 (0.639,0.926) [**]	0.774 (0.644,0.931) [**]
SIMD Quintile 4 (vs 1)	0.627 (0.507,0.775) [***]	0.669 (0.556,0.804) [***]	0.673 (0.560,0.808) [***]
SIMD Quintile 5 (vs 1)	0.576 (0.474,0.701) [***]	0.619 (0.522,0.735) [***]	0.628 (0.531,0.744) [***]
Business (vs Engineering)	0.689 (0.553,0.858) [***]	0.712 (0.583,0.870) [***]	0.713 (0.585,0.871) [***]
HaSS (vs Engineering)	0.982 (0.823,1.172)	0.983 (0.840,1.151)	0.975 (0.835,1.138)
Science (vs Engineering	0.883 (0.741,1.051)	0.895 (0.769,1.042)	0.894 (0.768,1.041)
Female (vs Male)	0.809 (0.707,0.925) [**]	0.829 (0.736,0.935) [**]	0.825 (0.733,0.929) [**]
Ethnic-Minority (vs White)	0.661 (0.500,0.874) [**]	0.693 (0.538,0.893) [**]	0.693 (0.539,0.890) [**]

Wald's Test P-values: * < 0.05, **< 0.01, ***< 0.001.

Table 9.3: A comparison of how much each of the three regression models exaggerated one another in terms of their exponentiated point-estimates. LR = Logistic Regression, MP = Modified Poisson, LB = Log-Binomial.

Variables	**LR/MP**	**LB/MP**	**LR/LB**
Prior Att. Quintile 1 (vs 3)	1.11	1.00	1.11
Prior Att. Quintile 2 (vs 3)	1.03	1.00	1.03
Prior Att. Quintile 4 (vs 3)	0.99	1.00	0.99
Prior Att. Quintile 5 (vs 3)	0.95	1.00	0.95
SIMD Quintile 2 (vs 1)	0.96	1.01	0.95
SIMD Quintile 3 (vs 1)	0.95	1.01	0.95
SIMD Quintile 4 (vs 1)	0.94	1.01	0.93
SIMD Quintile 5 (vs 1)	0.93	1.01	0.92
Business (vs Engineering	0.97	1.00	0.97
HaSS (vs Engineering)	1.00	0.99	1.01
Science (vs Engineering	0.99	1.00	0.99
Female (vs Male)	0.98	1.00	0.98
Ethnic-Minority (vs White)	0.95	1.00	0.95

Figure 9.1: Deviance residuals for each of the regression models fit to the drop-out outcome.



9.6 Results - Estimating the Survival and Hazard Function

The Kaplan-Meier Estimates of the survival function (Figures 9.2 and 9.3) indicated that there were lower drop-out rates over time amongst: females versus males, ethnic-minorities versus whites, and standard offer versus contextual offer students. SIMD Quintile 1 students were significantly more likely to drop-out compared to every other quintile, and the drop-out rate across *SIMD Quintiles* followed an ordinal pattern with drop-outs in $1 > 2 > 3 > 4 > 5$ (Figure 9.2). Notably, SIMD Quintile 1 students appeared to have a much higher drop-out rate in 2nd year compared to other quintiles. There was a clear association between *Prior Attainment Quintile* and probability of survival (Figure 9.2); those in the lower attainment groups had much higher drop-out rates than those from the higher attainment groups. The only groups of students where the prevalence of drop-out exceeded 10% was for students from Prior Attainment Quintile 1, and students from SIMD Quintile 1, who dropped out before the end of their first academic session. Looking at *Faculty*, Business had lower drop-out rates compared to the other faculties, followed by Engineering, whilst Science and HaSS had overlapping drop-out rates (Figure 9.2). There were no significant differences in drop-out rates over time detected between the levels for *Disability Status* nor *Urban/Rural Status* (Figure 9.3). Log-rank tests agreed with these results, finding significant chi-square test statistics for each of the explanatory variable fits with the exception of *Disability Status* ($\chi^2 = 0.5$ on 1 degree of freedom, $p = 0.5$) and *Urban/Rural Status* ($\chi^2 = 0.16$ on 2 degrees of freedom, $p = 0.5$).

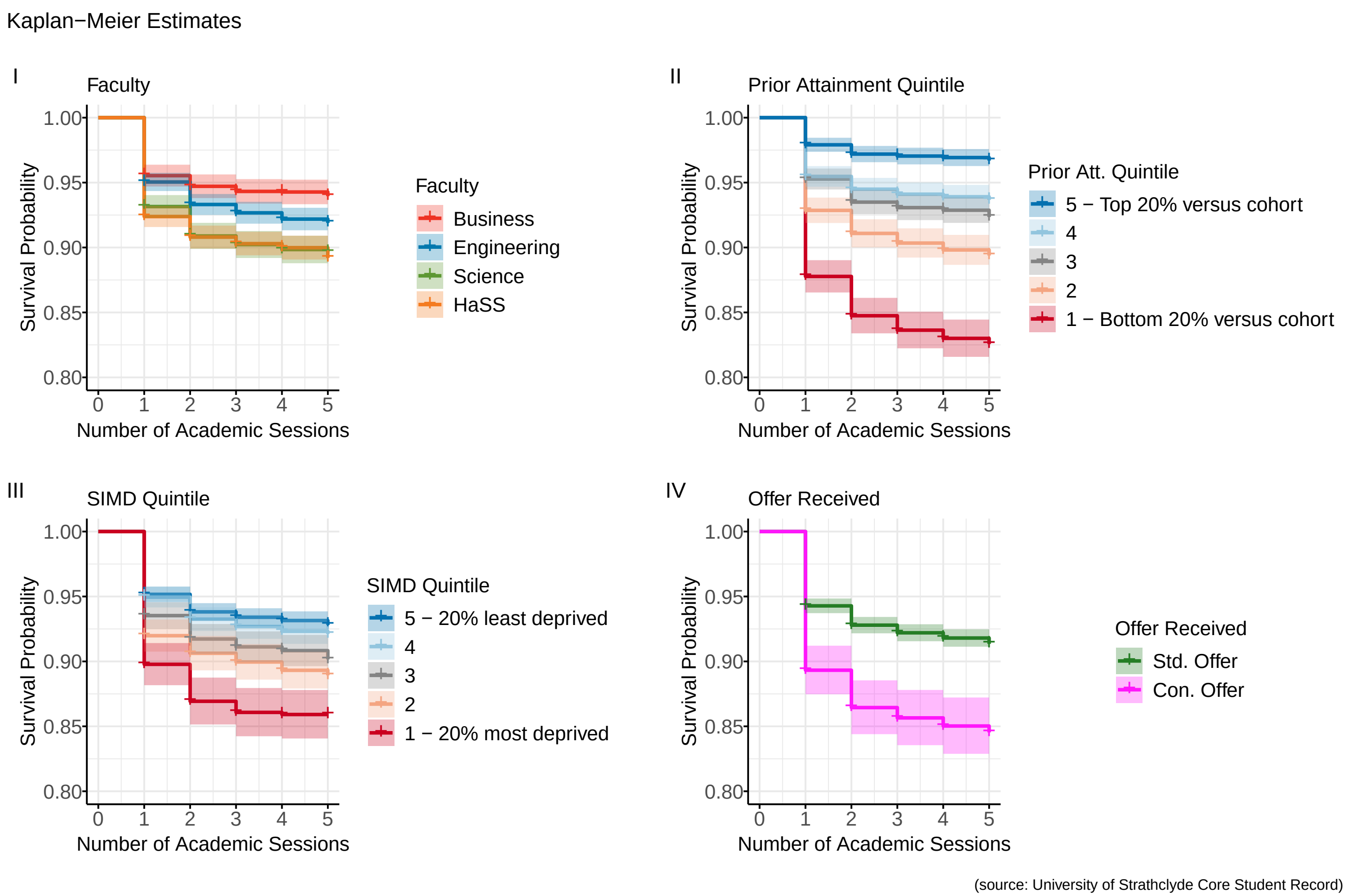
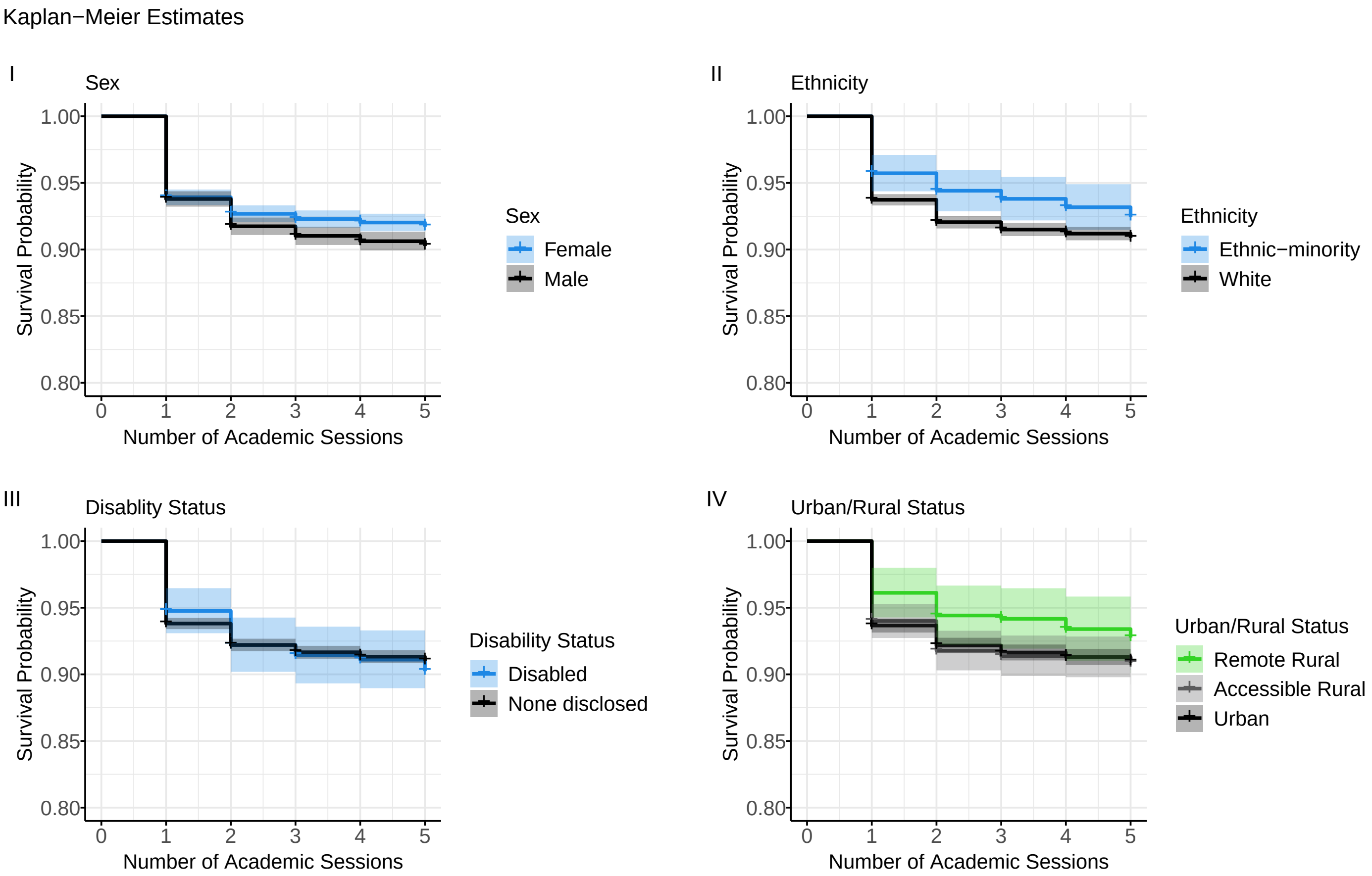


Figure 9.2: Kaplan-Meier Estimates (including 95% C.I.s) of drop-outs against explanatory variables (1 of 2). The legend of each sub-plot (I-IV) is ordered according to highest survival probability after one academic session.



(source: University of Strathclyde Core Student Record)

Figure 9.3: Kaplan-Meier Estimates (including 95% C.I.s) of drop-outs against explanatory variables (2 of 2). The legend of each sub-plot (I-IV) is ordered according to highest survival probability after one academic session.

The Logit estimates of the baseline hazard functions, $\hat{h_0}(t)$, indicated that drop-out rates peaked in the first academic session and levelled off between the fourth and fifth academic sessions (Figure 9.4). The exception to these were the plots for the Faculty of Business and SIMD Quintile 1, though this was because the number of observations in these categories was low (< 5). Plots of the complementary log-logistic (cloglog) transformation of the hazard function (see Section 6.13 for details) can assess whether or not variables satisfied the proportional-hazards assumption (Figure E.4). These plots were more or less identical to the logit plots (Figure 9.4), which was expected given that both approximate the one another when the hazard of experiencing drop-out at any given time point is low [103, p. 422]. From these plots it was not obvious whether any variable violated the proportional hazards (or proportional odds) assumption, meaning that schoenfeld tests would be required (Section 9.7.1). These plots also suggest that assuming the baseline hazard function follows a Weibull distribution is appropriate here given that they are always decreasing over time [163].

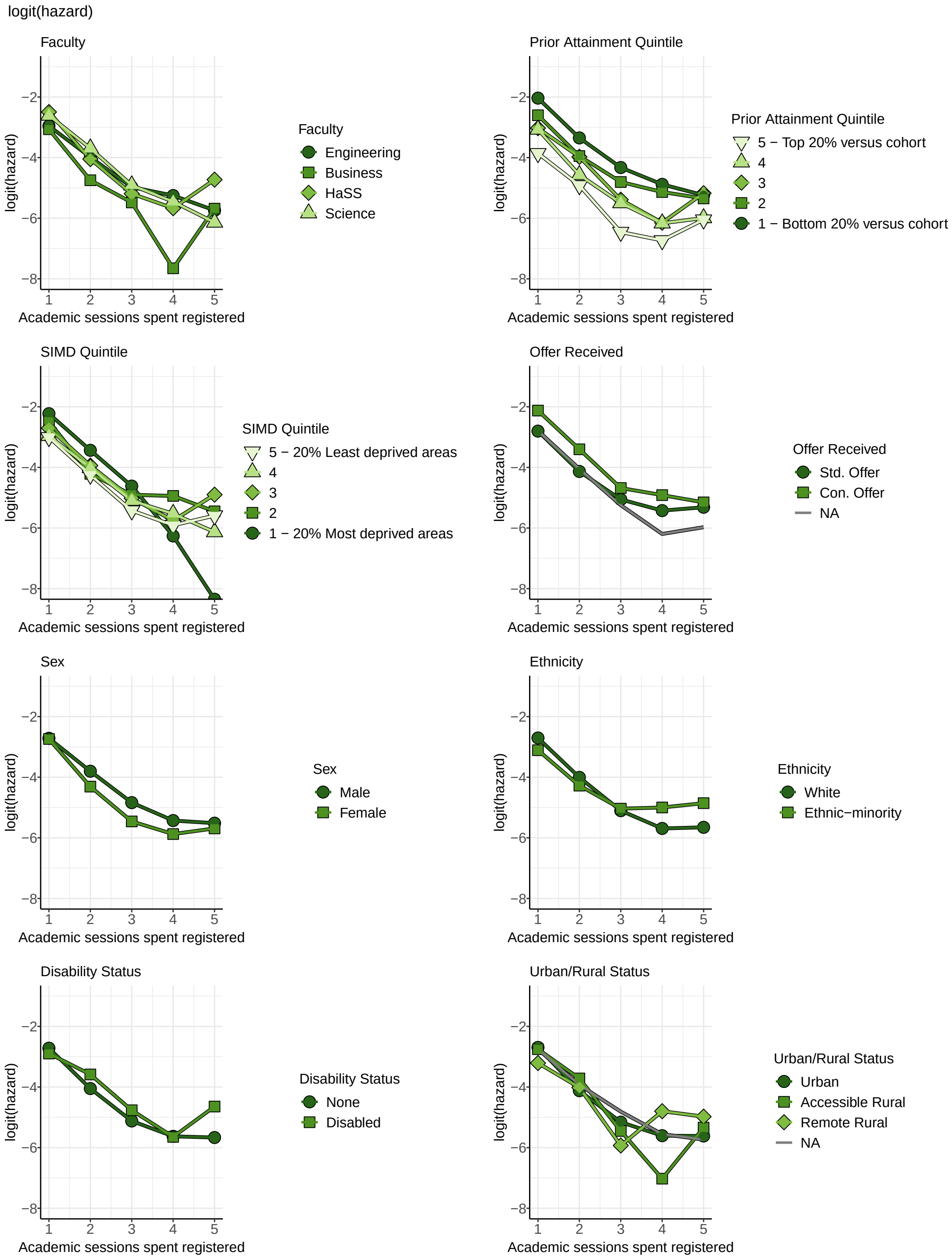


Figure 9.4: Comparison of logit(hazard) calculated descriptively at each time period, split by levels of each explanatory variable.

9.6.1 Results - Approximating the Baseline Hazard Function

The intercept-only General Logit DTE was well-approximated by the Quadratic and Cubic Logit DTE models (Figure 9.5). Comparing deviance residuals between each of the polynomial models (Table 9.4), the difference between the general, quadratic, and cubic models was negligible. The Quadratic Logit DTE was therefore selected as the preferred model to estimate the baseline hazard function given that it was the simplest of the three. This was also reflected in the AIC for the quadratic function which was the lowest of all models (Table 9.4). Thus, a multivariable General Logit DTE and a multivariable Quadratic Logit DTE model were fit to the **Drop-outs Person-level subset**, each controlling for the effects of *Sex*, *Ethnicity*, *Faculty*, *SIMD Quintile* and *Prior Attainment Quintile*.

Table 9.4: Comparison of Deviance and AIC values between baseline discrete time-to-event models which assume different shapes for the effect of time (academic sessions spent registered).

Model	Number of Covariates	Deviance	AIC	Log-Likelihood
Linear	2	10009.26	10013.26	-5004.63
Quadratic	3	9973.66	9979.66	-4986.83
Cubic	4	9972.32	9980.32	-4986.16
General	5	9972.23	9982.23	-4986.12

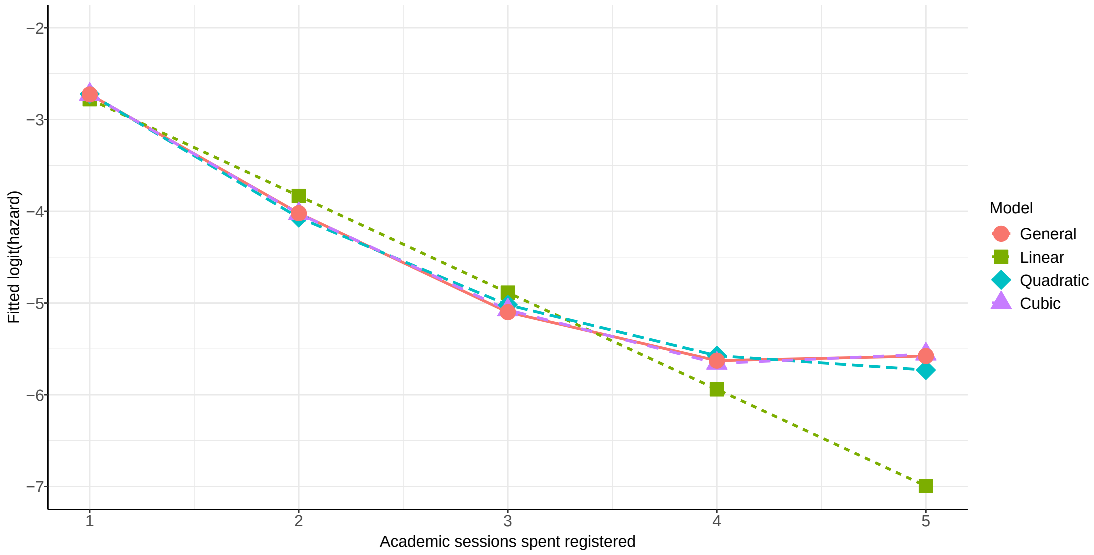


Figure 9.5: Comparison of logit(hazard) models with time as the only covariate. Each model either treated time as a categorical variable (general) or a linear, quadratic, or cubic continuous function.

9.7 Results - Discrete Time-to-Event Survival Models

The baseline hazard function, $\hat{h}_0(t)$, for the General Logit DTE are described by a combination of the intercept and *Sessions survived* covariates (Table 9.5). The intercept term represents the logit hazard (-2.337) of drop-out in the first academic session. The *Sessions survived* covariates suggest that with each increase in the time period, the odds of dropping out decreased over time before levelling off after four registration sessions.

In the Quadratic Logit DTE the baseline hazard function, $\hat{h}_0(t)$, is represented by the intercept and the linear and quadratic covariate for *Academic Sessions* (Table 9.6). Again, the intercept term represents the logit hazard (-2.329) of drop-out in the first academic session. The linear coefficient for academic sessions represents the instantaneous change in logit hazard (-1.547) of drop-out in the first academic session. The sign of the quadratic coefficient for academic session is positive (0.204), which indicates that the logit hazard function is convex. In other words, the the odds of dropping out decreased over time until it reached a stationary point, in agreement with the General Logit DTE. The quadratic function for time had a stationary point at session $t = 1 - \frac{1}{2}(-1.547/0.204) = 4.792$. This is in agreement with the shape of the logit hazard derived for the models with academic sessions as the only predictors(s) (Figure 9.5), which also had a stationary point between sessions 4 and 5.

Both the General and Quadratic Logit DTE models had near identical estimated effects for each covariate (Tables 9.5 and 9.6). Using the model estimates from the Quadratic Logit DTE, students from Prior Attainment Quintile 5 had 58.3% [95% CI: 46.0%, 68.1%] lower odds to drop-out than students from Prior Attainment Quintile 3, while students from Prior Attainment Quintile 1 had 135.8% [95% CI: 97.9%, 181.8%] higher odds. Students from SIMD Quintile 5 had 41.4% [95% CI: 29.1%, 51.4%] lower odds to drop-out compared to their peers from SIMD Quintile 1. Students from the Faculty of Business had 29.2% [95% CI: 12.5%, 43.1%] lower odds to drop-out compared to students from the Faculty of Engineering. Females had 18.7% [95% CI: 7.3%, 28.6%] lower odds to drop-out compared to males, and students from ethnic-minority groups had 33.5% [95% CI: 13.4%, 48.8%] lower odds to drop-out compared to white students. The deviance, AIC, and log-likelihood values for both the General and Quadratic Logit DTE models were very similar (Table 9.7). The deviance residuals from each regression model indicate that there are no unduly influential observations (Figure 9.6).

Table 9.5: Model estimates for the General Logit DTE Model.

Variables	**P-values**	**Coefficients (S.E.)**	**Odds-Ratio (95% C.I.)**
(Intercept)		-2.337 (0.124)	0.097 (0.075,0.123)
Sessions survived: 2 (vs 1)	<0.001	-3.624 (0.137)	0.027 (0.020,0.035)
Sessions survived: 3 (vs 1)	<0.001	-4.690 (0.168)	0.009 (0.007,0.013)
Sessions survived: 4 (vs 1)	<0.001	-5.206 (0.197)	0.005 (0.004,0.008)
Sessions survived: 5 (vs 1)	<0.001	-5.089 (0.283)	0.006 (0.003,0.010)
Prior Att. Quintile 1 (vs 3)	<0.001	0.858 (0.090)	2.358 (1.979,2.818)
Prior Att. Quintile 2 (vs 3)	<0.001	0.340 (0.097)	1.406 (1.164,1.701)
Prior Att. Quintile 4 (vs 3)	0.124	-0.167 (0.109)	0.846 (0.683,1.046)
Prior Att. Quintile 5 (vs 3)	<0.001	-0.876 (0.134)	0.416 (0.319,0.539)
SIMD Quintile 2 (vs 1)	0.013	-0.264 (0.106)	0.768 (0.623,0.946)
SIMD Quintile 3 (vs 1)	0.005	-0.301 (0.106)	0.740 (0.602,0.912)
SIMD Quintile 4 (vs 1)	<0.001	-0.453 (0.104)	0.636 (0.518,0.781)
SIMD Quintile 5 (vs 1)	<0.001	-0.534 (0.097)	0.586 (0.486,0.709)
Business (vs Engineering)	0.002	-0.342 (0.110)	0.710 (0.571,0.878)
HaSS (vs Engineering)	0.985	0.002 (0.088)	1.002 (0.844,1.190)
Science (vs Engineering)	0.209	-0.109 (0.086)	0.897 (0.757,1.063)
Female (vs Male)	0.002	-0.206 (0.067)	0.814 (0.714,0.927)
Ethnic-minority (vs White)	0.003	-0.407 (0.139)	0.666 (0.502,0.866)

Table 9.6: Model estimates for the Quadratic Logit DTE Model.

Variables	**P-values**	**Coefficients (S.E.)**	**Odds-Ratio (95% C.I.)**
(Intercept)		-2.329 (0.124)	0.097 (0.076,0.124)
Academic Sessions	<0.001	-1.547 (0.092)	0.213 (0.178,0.255)
(Academic Sessions)^2	<0.001	0.204 (0.032)	1.226 (1.151,1.303)
Prior Att. Quintile 1 (vs 3)	<0.001	0.858 (0.090)	2.358 (1.979,2.818)
Prior Att. Quintile 2 (vs 3)	<0.001	0.340 (0.097)	1.405 (1.164,1.700)
Prior Att. Quintile 4 (vs 3)	0.124	-0.167 (0.109)	0.846 (0.683,1.046)
Prior Att. Quintile 5 (vs 3)	<0.001	-0.876 (0.134)	0.417 (0.319,0.540)
SIMD Quintile 2 (vs 1)	0.013	-0.264 (0.106)	0.768 (0.623,0.946)
SIMD Quintile 3 (vs 1)	0.005	-0.300 (0.106)	0.741 (0.602,0.912)
SIMD Quintile 4 (vs 1)	<0.001	-0.453 (0.104)	0.636 (0.518,0.781)
SIMD Quintile 5 (vs 1)	<0.001	-0.534 (0.097)	0.586 (0.486,0.709)
Business (vs Engineering)	0.002	-0.346 (0.110)	0.708 (0.569,0.875)
HaSS (vs Engineering)	0.989	-0.001 (0.088)	0.999 (0.842,1.187)
Science (vs Engineering)	0.202	-0.110 (0.086)	0.896 (0.756,1.061)
Female (vs Male)	0.002	-0.206 (0.067)	0.813 (0.714,0.927)
Ethnic-minority (vs White)	0.003	-0.407 (0.139)	0.665 (0.502,0.866)

Table 9.7: Comparison of Deviance and AIC values between Quadratic and General Logit DTE models.

Model	**Number of Covariates**	**Deviance**	**AIC**	**Log-Likelihood**
Quadratic	16	9547.21	9579.21	-4773.60
General	18	9545.57	9581.57	-4772.78

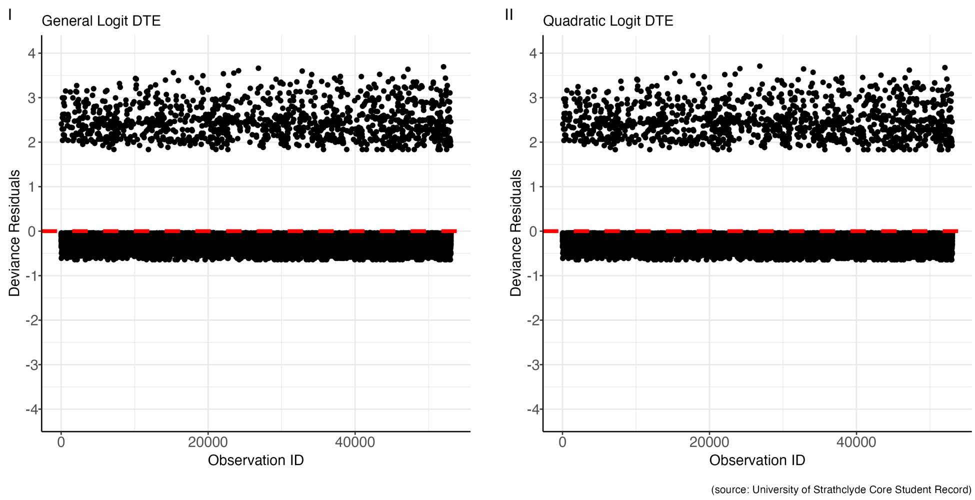


Figure 9.6: Deviance residuals for each of the fitted discrete survival models.

9.7.1 Results - Continuous Time-to-Event Survival Models

The Standard CPH and Parametric Weibull models had very similar estimates to one another (Tables 9.8, E.2 and 9.9, respectively). Using the Standard CPH as an example (Table 9.8), students from Prior Attainment Quintile 5 were 57.8% [95% CI: 45.3%, 67.4%] less likely to drop-out than students from Prior Attainment Quintile 3, while students from Prior Attainment Quintile 1 were 129.5% [95% CI: 93.3%, 172.4%] more likely. Students from SIMD Quintile 5 were 40.0% [95% CI: 28.1%, 50.0%] less likely to drop-out compared to their peers from SIMD Quintile 1. Students from the Faculty of Business were 28.2% [95% CI: 11.4%, 41.8%] less likely to drop-out compared to students from the Faculty of Engineering. Females were 18.0% [95% CI: 7.0%, 27.7%] less likely to drop-out compared to males, and students from ethnic-minority groups were 32.8% [95% CI: 12.4%, 48.4%] less likely to drop-out compared to white students. The Parametric Weibull fit assumed that the baseline hazard function followed a Weibull distribution with scale parameter $\hat{\sigma} = 1.123$, and mean $\hat{\mu}_0 = 3.311$. The point estimates from the Parametric Weibull (Table 9.9) were slightly larger in magnitude than the CPH models, though the standard errors were very similar.

Schoenfeld tests on the Standard CPH indicated that the proportional-hazards assumption was violated for the *Faculty* and *Sex* explanatory variables. The model estimates from these variables may therefore be biased. Thus, a follow-up "Stratified CPH" was fit to the data which stratified by these two variables. Schoenfeld tests conducted on the Stratified CPH model indicated that no variables violated the proportional-hazards assumption (Table E.1). However, the remaining covariates in the model had nearly identical estimates to the Standard

CPH model (Table E.2). Trivially, the Stratified CPH could not provide estimates for *Faculty* nor *Sex*. The deviance residuals from each of the continuous survival models do not indicate any influential observations (Figure 9.7).

Table 9.8: Model estimates for the Standard CPH Model.

Variables	**P-values**	**Coefficients (S.E.)**	**Hazard Ratio (95% C.I.)**
Prior Att. Quintile 1 (vs 3)	<0.001	0.831 (0.087)	2.295 (1.933,2.724)
Prior Att. Quintile 2 (vs 3)	<0.001	0.334 (0.094)	1.397 (1.161,1.680)
Prior Att. Quintile 4 (vs 3)	0.128	-0.162 (0.107)	0.850 (0.690,1.048)
Prior Att. Quintile 5 (vs 3)	<0.001	-0.863 (0.132)	0.422 (0.326,0.547)
SIMD Quintile 2 (vs 1)	0.014	-0.250 (0.102)	0.779 (0.638,0.951)
SIMD Quintile 3 (vs 1)	0.005	-0.287 (0.102)	0.750 (0.615,0.916)
SIMD Quintile 4 (vs 1)	<0.001	-0.435 (0.100)	0.647 (0.531,0.788)
SIMD Quintile 5 (vs 1)	<0.001	-0.511 (0.093)	0.600 (0.500,0.719)
Business (vs Engineering)	0.002	-0.331 (0.107)	0.718 (0.582,0.886)
HaSS (vs Engineering)	0.987	0.001 (0.085)	1.001 (0.848,1.182)
Science (vs Engineering)	0.206	-0.106 (0.084)	0.900 (0.764,1.060)
Female (vs Male)	0.002	-0.199 (0.065)	0.820 (0.723,0.930)
Ethnic-minority (vs White)	0.003	-0.397 (0.135)	0.672 (0.516,0.876)

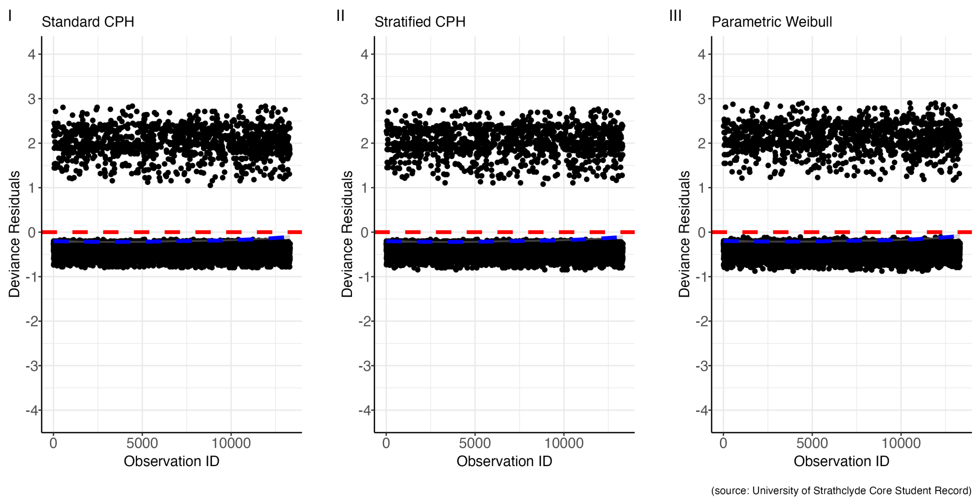


Figure 9.7: Deviance residuals for each of the fitted continuous survival models.

Table 9.9: Model estimates for the Parametric Weibull Model.

Variables	P-values	Coefficients (S.E.)	Hazard Ratio (95% C.I.)
(Intercept)		-3.311 (0.132)	0.036 (0.028,0.047)
Prior Att. Quintile 1 (vs 3)	<0.001	0.858 (0.090)	2.358 (1.976,2.815)
Prior Att. Quintile 2 (vs 3)	<0.001	0.345 (0.095)	1.412 (1.173,1.700)
Prior Att. Quintile 4 (vs 3)	0.110	-0.170 (0.107)	0.843 (0.684,1.040)
Prior Att. Quintile 5 (vs 3)	<0.001	-0.890 (0.134)	0.411 (0.316,0.535)
SIMD Quintile 2 (vs 1)	0.013	-0.255 (0.102)	0.775 (0.634,0.947)
SIMD Quintile 3 (vs 1)	0.003	-0.299 (0.102)	0.742 (0.607,0.906)
SIMD Quintile 4 (vs 1)	<0.001	-0.448 (0.101)	0.639 (0.524,0.779)
SIMD Quintile 5 (vs 1)	<0.001	-0.529 (0.094)	0.589 (0.490,0.708)
Entered Faculty Bus. (vs Eng.)	0.012	-0.270 (0.107)	0.764 (0.619,0.943)
Entered Faculty HaSS (vs Eng.)	0.453	0.064 (0.085)	1.066 (0.903,1.258)
Entered Faculty Sci. (vs Eng.)	0.444	-0.064 (0.084)	0.938 (0.796,1.105)
Female (vs Male)	0.001	-0.207 (0.065)	0.813 (0.716,0.923)
Ethnic-minority (vs White)	0.003	-0.406 (0.136)	0.666 (0.511,0.869)

9.7.2 Results - Comparing Model Fits

The estimated effect of each covariate (and their 95% confidence intervals) were very similar across all regression and survival models applied to the data (Figure 9.8). The models which derived odds ratios (Logistic, General Logit DTE, Quadratic Logit DTE) were near identical in their estimates.

As mentioned previously (Section 9.5), the regression model estimates were similar to one another regardless of whether they derived risk or odds ratios (Table 9.3). It can be seen here that this was also the case for the survival models (Table 9.12). In fact, the difference between the continuous Std. CPH and Quadratic Logit DTE survival estimates were smaller than the difference between

Table 9.10: Schoenfeld test statistics for Standard CPH Model. Calculated using Kaplan-Meier transformed time.

Variables	Chi-Square	D.o.f	P-values
Prior Attainment Quintile	2.94	4	0.568
SIMD Quintile	3.55	4	0.470
Faculty	15.95	3	<0.001
Sex	13.90	1	<0.001
Ethnicity	3.20	1	0.073
Global	30.71	13	<0.001

the Logistic and Modified Poisson/Log-Binomial regression models. This could be because the prevalence of the drop-out outcome was even smaller within each academic session (roughly 6% in first year for example, as opposed to 9% overall - Table 9.2). This means that the survival and regression models which derived odds-ratios could be reasonably interpreted as approximations of risk-ratios.

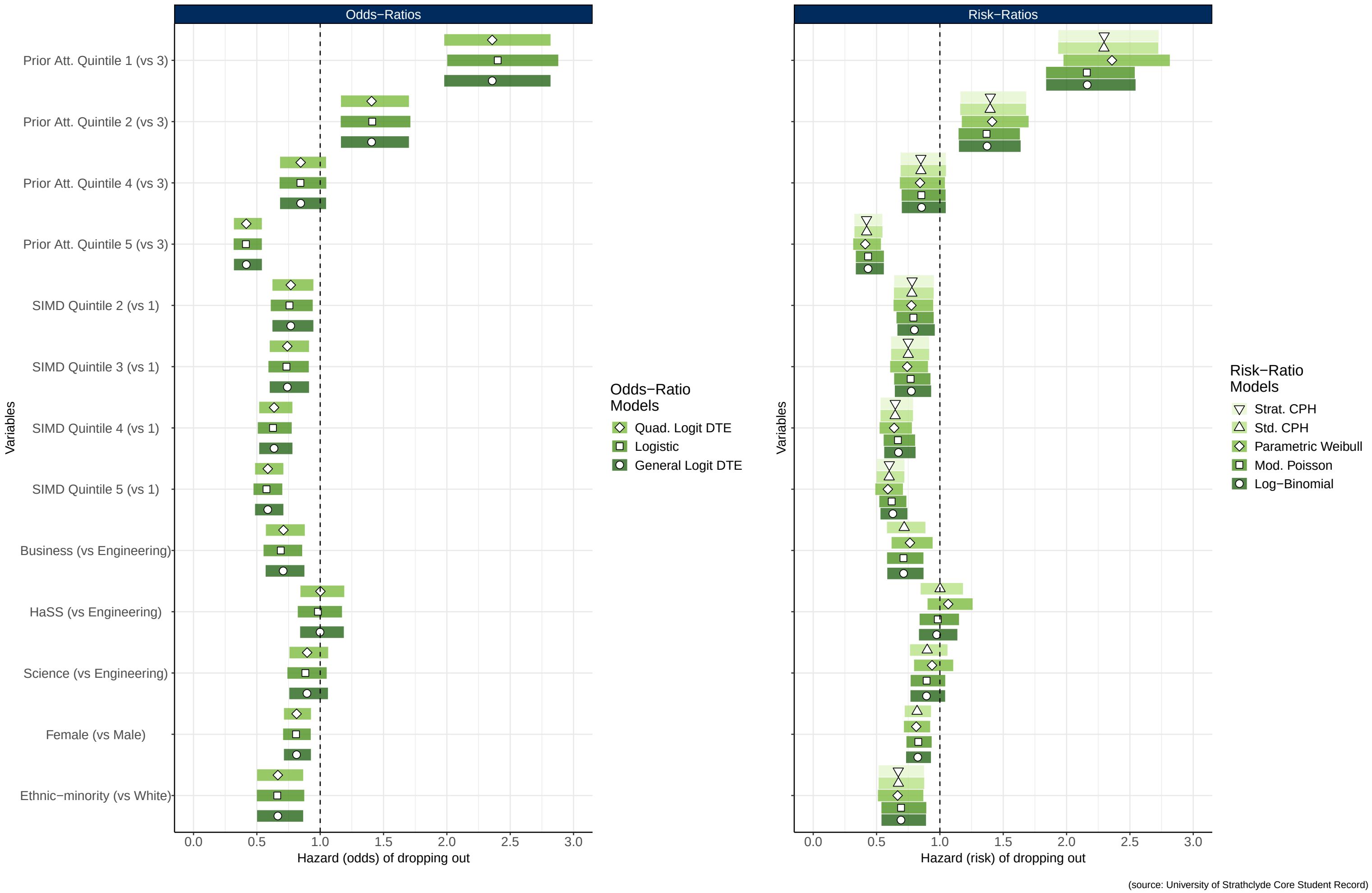


Figure 9.8: Comparison of each survival model's exponentiated estimates.

Table 9.11: Comparison of exponentiated estimates across all regression and survival models. LR = Logistic, QDTE = Quadratic Logit DTE, GDTE = General Logit DTE, MP = Modified Poisson, LB = Log-Binomial, Std. CPH = Standard Cox-Proportional Hazards, PW = Parametric Weibull.

Variables	**LR**	**QDTE**	**GDTE**	**MP**	**LB**	**Std. CPH**	**PW**
(Intercept)	0.146 [***]	0.097 [***]	0.097 [***]	0.125 [***]	0.125 [***]		0.036 [***]
Sessions survived: 2 (vs 1)		0.027 [***]					
Sessions survived: 3 (vs 1)		0.009 [***]					
Sessions survived: 4 (vs 1)		0.005 [***]					
Sessions survived: 5 (vs 1)		0.006 [***]					
Academic Sessions			0.213 [***]				
(Academic Sessions)^2			1.226 [***]				
Prior Att. Quintile 1 (vs 3)	2.402 [***]	2.358 [***]	2.358 [***]	2.160 [***]	2.163 [***]	2.295 [***]	2.358 [***]
Prior Att. Quintile 2 (vs 3)	1.411 [***]	1.406 [***]	1.405 [***]	1.368 [***]	1.372 [***]	1.397 [***]	1.412 [***]
Prior Att. Quintile 4 (vs 3)	0.844	0.846	0.846	0.854	0.855	0.850	0.843
Prior Att. Quintile 5 (vs 3)	0.413 [***]	0.416 [***]	0.417 [***]	0.433 [***]	0.433 [***]	0.422 [***]	0.411 [***]
SIMD Quintile 2 (vs 1)	0.758 [*]	0.768 [*]	0.768 [*]	0.791 [*]	0.799 [*]	0.779 [*]	0.775 [*]
SIMD Quintile 3 (vs 1)	0.734 [**]	0.740 [**]	0.741 [**]	0.769 [**]	0.774 [**]	0.750 [**]	0.742 [**]
SIMD Quintile 4 (vs 1)	0.627 [***]	0.636 [***]	0.636 [***]	0.669 [***]	0.673 [***]	0.647 [***]	0.639 [***]
SIMD Quintile 5 (vs 1)	0.576 [***]	0.586 [***]	0.586 [***]	0.619 [***]	0.628 [***]	0.600 [***]	0.589 [***]
Business (vs Engineering)	0.689 [***]	0.710 [**]	0.708 [**]	0.712 [***]	0.713 [***]	0.718 [**]	0.764 [*]
HaSS (vs Engineering)	0.982	1.002	0.999	0.983	0.975	1.001	1.066
Science (vs Engineering)	0.883	0.897	0.896	0.895	0.894	0.900	0.938
Female (vs Male)	0.809 [**]	0.814 [**]	0.813 [**]	0.829 [**]	0.825 [**]	0.820 [**]	0.813 [***]
Ethnic-minority (vs White)	0.661 [**]	0.666 [**]	0.665 [**]	0.693 [**]	0.693 [**]	0.672 [**]	0.666 [**]

Wald's Test P-values: * < 0.05, **< 0.01, ***< 0.001.

Table 9.12: Comparison of Standard CPH risk-ratios and Quadratic Logit DTE odds-ratios.

Variables	Std. CPH	Quad. Logit	Std. CPH/ Quad. Logit
Prior Att. Quintile 1 (vs 3)	2.30	2.36	0.97
Prior Att. Quintile 2 (vs 3)	1.40	1.41	0.99
Prior Att. Quintile 4 (vs 3)	0.85	0.85	1.01
Prior Att. Quintile 5 (vs 3)	0.42	0.42	1.01
SIMD Quintile 2 (vs 1)	0.78	0.77	1.01
SIMD Quintile 3 (vs 1)	0.75	0.74	1.01
SIMD Quintile 4 (vs 1)	0.65	0.64	1.02
SIMD Quintile 5 (vs 1)	0.60	0.59	1.02
Business (vs Engineering)	0.72	0.71	1.01
HaSS (vs Engineering)	1.00	1.00	1.00
Science (vs Engineering)	0.90	0.90	1.00
Female (vs Male)	0.82	0.81	1.01
Ethnic-minority (vs White)	0.67	0.67	1.01

9.8 Discussion

The aim of this chapter was to determine whether or not discrete and/or continuous survival methods are more appropriate fits to the data when compared to traditional regression methods. All methods derived similar covariate estimates compared to one another. This is despite the warnings that biased estimates can occur when follow-up time is misspecified as a continuous random variable [156]. The small differences between estimates may have been due to the hazard of drop-out in any given academic session being low ($< 6\%$) and that the proportion of censored observations was high [156]. Despite the similarity between the traditional regression, continuous survival and discrete survival model estimates, it is recommended that discrete survival models be applied to the data. The survival models are preferred over regression models since the latter are more sensitive to issues (explained in Section 7.7.1) when the prevalence of the outcome is not rare ($> 10\%$). In survival models, the prevalence of drop-out is smaller within each academic session (e.g. roughly 6% in first year - Table 9.2) than the prevalence of drop-out over all academic sessions in regression models (roughly 9% overall - Table 9.2). However, given that the estimated effects were very similar across regression and survival models, it appears unlikely that incorrect conclusions would be drawn from the regression models. Of the two discrete survival models examined in this chapter, the Quadratic Logit DTE is preferred over the General Logit DTE since both had very similar deviance, AIC, and log-likelihood values and the former was the simpler of the two.

An advantage to the survival approach in this chapter is that the risk, or hazard, of dropping out each academic session is perhaps more intuitive for a lay audience than regression approach in previous Chapters (8 and 7) which estimated the "risk" of achieving a successful outcome. For example, the hazard of drop-out over time peaked in the first academic session and decreased thereafter, reaching a stationary point between the fourth and fifth academic session. In the Scottish context, this could be because most students become censored after the fourth and fifth academic session due to completing their Bachelor's with Honours degree, or progressing onto the final stage of their Integrated Masters degree.

No examples could be found of survival analysis techniques being applied to the student drop-out problem in Scotland. This means that this analysis is likely the first of its kind as it relates to Widening Access in Scotland. The associations observed between socio-economic/demographic background and dropping out of university are very similar to those already observed in previous chapters. Specifically, factors associated with drop-out were lower prior attainment, coming from more deprived areas, being male, being white, and there were differences in the drop-out rates between faculties. This adds to the evidence that there are factors beyond academic "potential" which affect students' outcomes in higher education. These findings differ from those found by Arulampalam et al. [148], though it should be noted that both the population of interest and explanatory variables used in the model fits were considerably different. Trivially, neither the results from this chapter, nor those from Arulampalam et al. [148], should be taken as representative of all higher education institutions in the UK or globally. It is likely that individual institutions, or even individual degree programmes, will have unique relationships between students' backgrounds/prior attainment and their drop-out rates. Thus, to understand the impact of Widening Access policies across Scotland, it would be necessary to collect and analyse data from each individual institution. This is perhaps unrealistic given that the student data required covers protected characteristics (*Sex*, *Ethnicity*, etc.) and that institutions may be unwilling to share registration records with external researchers.

9.8.1 Limitations and Future Analyses

A small number of students whose Ethnicities were unknown or missing were removed (see Section 3.5). Sensitivity analyses found that model estimates remained unchanged when these observations were included, their removal therefore has no impact on the interpretation of the model results.

Bias has been introduced to model fits via informative censoring since all students were censored after their 5th academic session (Sections 6.9.3 and 6.9.5). However, the number of students this affects, and thus the level of bias, should be very small. Another potential source of informative censoring is from students who remain in voluntary or academic suspension before eventually dropping out. Since the **Drop-outs subset** is derived from registration records, such students should eventually be recorded as drop-outs. However, their survival time will appear to be longer than the time they spent meaningfully engaged with the degree programme. It is unlikely that the exact moment a student decided to drop-out of university can be identified for all students since some will inevitably not respond to follow-up. Students in some form of suspension could be examined descriptively to infer why they have been registered for unusually long periods of time.

Where other studies [115, 151, 154, 158, 159] had follow-up times each semester, this analysis only followed-up each academic session. Measuring each semester has the advantage of identifying more precisely when a student drops out, perhaps even narrowing down the potential reasons for their exit. For example, it could be investigated whether an unbalanced workload over semesters lead to more drop-outs within one semester over another. These insights could be used to understand when pressure is highest and when students may need more support. Smaller units of time could also improve estimates derived from the continuous-time models such as the CPH and Weibull. Models could also be extended to consider a competing risks framework that would allow the outcome of completion to be modelled alongside drop-out. This approach has already been taken by Arulampalam et al. [148] and Cvetkovski et al. [160]. This would be particularly useful given that the prevalence of completion is $> 70\%$ across the whole school-leavers dataset (see Chapter 5).

The model fits may be improved with additional explanatory variables. This analysis only examined one socio-economic variable – *SIMD Quintile*. There are likely many more factors that influence a student's decision to drop-out, some of which have already been highlighted in the literature (Section 2.3). Besides a student's contextual background, two more key variables are worth investigation. The first is, trivially, the reason for student drop-out. This information is hard to obtain, but would be of interest to Widening Access efforts, particularly if the target groups are leaving for reasons related to their socio-economic background. This information could be captured via exit-interviews or surveys, but would depend on the response rates of these students. The second key variable, is the student's university-level attainment achieved each academic session.

This is an example of a time-dependent covariate, which would require models to be extended to consider more complicated relationships between the outcome and explanatory variables. There are examples of time-dependent covariates in the literature, for example Respondek et al. [150] modelled the interactive effect between university-level attainment and student's perceived control over their studies, in relation to student drop-out. Cvetkovski et al. [160] looked at the effects of many time-dependent covariates, such as employment status, mental health status, and whether students were living at home or not. These studies serve as examples for what extensions could be applied to the **Drop-outs subset** (and also the **School-leavers dataset** in future analyses. In particular, university-level attainment would be of interest because a student's probability of survival could update each session to more accurately reflect their current academic performance. Other examples of time-dependent covariates include *Faculty*, *Department*, *Programme Title*, *Repeated Stage* and *Break* (see Section 4.1). University-attainment could also be used to infer a student's reason for drop-out, if their academic performance suggests they did not have the necessary credits required to progress to the next stage just prior to dropping out. It is highly recommended that future analyses of the data aim to include the effect of time-dependent covariates.

An increasingly popular topic is the prediction of student drop-outs [151, 153–155, 158]. Many studies are comparing the performance of survival methods to traditional regression and machine learning methods [154, 155, 158] since the latter can also predict when a student will drop out. There is evidence to suggest that survival methods could be more useful than traditional machine learning algorithms for predicting drop-outs at the early stages of a students journey [154, 155]. The prediction of student drop-outs using survival analysis methods would be a natural next step for the **School-leavers dataset**. The findings in this analysis forms the foundations for which variables are worth consideration in a predictive model. Applying such techniques across Scottish universities would be of interest for Widening Access efforts considering its current trajectory towards improving the positive outcomes of target students [7] (see also Section 2.5.2).

9.9 Conclusion

The regression and survival models considered in this chapter support the notion that a range of factors are associated with the rate of student drop-outs from the University of Strathclyde. These factors include prior attainment from secondary education, sex, ethnicity, area-level deprivation, and the faculty a student is registered with. While discrete survival models were the more natural fit to the student drop-out data due to satisfying more assumptions, the estimates from continuous survival models and traditional regression models were very similar. This may not be the case when applied to other datasets from other institutions. While it is recommended that discrete survival models be used, researchers should use their best judgment when selecting an appropriate model for student drop-out data; consideration should be given to how the model is interpreted to its target audience.

It is recommended that the models be extended to consider the time-dependent covariates which were not available in the current edition of the **School-leavers dataset**, such as module performance or credit-weighted averages from year-to-year. Future analyses could also consider the competing risk of completion of a degree programme. Students' reasons for drop-out should be prioritised as information to be added to the data, if possible. It is also recommended that the prediction of student drop-outs be investigated in future analyses, as it could allow the University to more effectively intervene and support students who are at risk of drop-out. If possible, Widening Access efforts should focus on collecting and analysing data from across all of Scotland's 18 higher education institutions, although access to these data may prove to be a challenge.

Chapter 10

Conclusions and Future Work

10.1 Aims of the Thesis

The aims of the thesis were as follows:

1. To explore the demographic, socio-economic and prior attainment information of school-leavers at the University of Strathclyde.
2. To determine whether or not contextual offer students are achieving similar levels of academic success as their standard offer peers.
3. To measure how much academic success/failure is affected by a student's:
   (a) Socio-economic background,
   (b) Prior attainment from secondary education,
   (c) Demographics,
   (d) Choice of degree programme.

4. To determine what the most appropriate method is for modelling the effects on academic outcomes.

10.1.1 Aim 1 - Exploration of School-Leavers

The first aim was addressed across Chapters 2, 3, 4 and 5. Chapter 2 summarised the relevant background information needed to understand the Scottish education system, the disadvantage present in the system, the arguments for and against Widening Access policies, and the progress that has been made towards Widening Access targets up to 2024. This was necessary to properly interpret any findings from the analysis of student data from the University of Strathclyde. Chapter 3 detailed how data from the University and the Scottish Government were gathered, joined, and cleaned to derive the **School-leavers dataset** giving the population of interest. Chapter 4 then defined the relevant outcome and explanatory variables in the dataset which were visualised and explored in Chapter 5. This chapter also measured the associations between variables. For example, it was found that there appeared to be associations between *Faculty* and the variables: *Sex*, *Ethnicity*, and *SIMD Quintile*. Males and ethnic-minorities were more represented in Science and Engineering while females and whites were more represented in Business and HaSS. Those from lower *SIMD Quintiles* were more represented in HaSS and Science. *Prior Attainment Points* appeared to be higher within the Faculties of Business and Engineering, those from higher *SIMD Quintiles*. Chapter 5 also showed the prevalence of retention after first year (90%), completion of a degree programme within four years (74%), and drop-out (9%) in the **School-leavers dataset**. Additional variables that were not used in the analyses, such as those which identify students who changed or repeated a stage of their degree programme, or had taken a break at some point in their registration, were briefly explored in Appendices C and D. From the context provided in these Chapters (2, 3, 4 and 5), research questions were able to be posed and answered in the subsequent analyses Chapters (7, 8 and 9).

10.1.2 Aim 2 - Contextual Offer Students

The second aim was addressed in Chapter 8. It was found that standard offer students were significantly more likely to be retained at the end of first year and significantly more likely to complete their Bachelor's with Honours degree within four years, compared to contextual offer students. For example, it was found that contextual offer students had an 82.7% [95% CI: 80.4%, 84.9%] chance of retention and a 62.2% [95% CI: 59.3%, 65.2%] chance of completion, once controlling for the effects of *Academic Cohort*, *Sex* and *Ethnicity*. These chances for contextual offer students were lower than the University of Strathclyde's 2030 benchmark of 90-95% retention for all students [8]. Most contextual offer students were predicted to successfully complete their degree (around 62.2%), though there exists no benchmark for comparison. It is perhaps unrealistic however, to expect contextual offer students to achieve at a level similar to their standard offer peers at university. This is because, by definition, contextual offer students are very likely to have lower prior attainment when they commence their degree and come from areas with higher levels of deprivation (as defined by SIMD), both of which are negatively associated with a successful outcome in our models. It should also be emphasised that the majority of contextual offer students managed to complete their Bachelor's with Honours degree within four years. It may also be the case that these students are more likely to repeat a stage of their degree programme, meaning that this could be an under-estimate of the true completion rate for contextual offer students.

10.1.3 Aim 3 - Relationships With Success/Failure

Chapters (7, 8 and 9) each addressed the third aim using different model specifications.

Increased levels of prior attainment from secondary education always appeared to have a highly significant and positive effect on school-leavers' chances of a successful academic outcome at the University. In Chapter 7, it was found that having Advanced Higher Mathematics was positively associated with successfully being retained and completing a degree programme in Mathematics and/or Statistics. The results also suggested that Advanced Higher Mathematics may have a similarly positive effect for students in other Science and Engineering programmes, conditional upon whether or not Advanced Higher Mathematics was recommended by the student's chosen degree programme. A student's *Prior Attainment Points* had a significant and positive effect on retention and completion in all of the models it appeared in across both Chapters 7 and 8, though there were some conflicting evidence over whether or not this effect was linear. In Chapter 9, students who had the lowest levels of prior attainment compared to their peers (Prior Attainment Quintile 1) were more than twice as likely to drop out of the University at any time period compared to their peers with average levels of attainment (Prior Attainment Quintile 3 - Table 9.6), who themselves were more than twice as likely to drop out compared to their peers with the highest levels of prior attainment (Prior attainment Quintile 5 - Table 9.6).

Each chapter suggested that school-leavers from SIMD Quintile 1 had significantly lower chances of a successful outcome at the University compared to their peers from SIMD Quintiles 2-5, even when controlling for their prior attainment. For example, Chapter 8 found that school-leavers from SIMD Quintiles 1 and 2 were 5.2% [95% CI: 3.1%, 7.1%] less likely to be retained and 9.5% [95% CI: 6.3%, 12.6%] less likely to complete their degree compared to students from SIMD Quintiles 3-5, once controlling for the effects of *Prior Attainment Points*, *Academic Cohort*, *Sex*, *Ethnicity* and *Faculty*. Chapter 7 identified a similarly significant gap between the completion rates of school-leavers from *SIMD Quintiles* 1 and 5, specifically within the Faculties of Science and Engineering. Finally, in Chapter 9 students from lower SIMD Quintiles had consistently higher drop-out rates than those from higher quintiles, both before and after controlling for the effects of other significant covariates. For example, a Cox-Proportional Hazards model found that school-leavers from SIMD Quintile 5 had 40.0% [95% CI: 28.1%, 50.0%] reduced hazard of to drop-out compared to their peers from SIMD Quintile 1.

Each of the analysis chapters (7, 8 and 9) found that *Sex*, *Ethnicity*, and *Faculty* significantly affected retention, completion and drop-out of university. However, the significance (but not direction) of these effects sometimes changed across model specifications. When examining school-leavers across the whole University, it appeared that females were more likely to achieve degree completion compared to males when controlling for other significant effects. For example, Chapter

7 found that sex had little-to-no influence on retention nor completion rates of students within the Faculties of Science and Engineering. Yet, in Chapter 8 , while there was no influence of sex on students' retention outcomes across the whole University, females were 6.2% [95% CI: 3.2%, 9.2%] more likely to complete their degree compared to males across the entire University, once controlling for the effects of student's *Best 5 Highers at Application*, their *SIMD Quintile, Faculty, Academic Cohort*, and *Ethnicity.*

At times, ethnic-minorities appeared to be more likely to achieve successful academic outcomes compared to whites. In Chapters 7 the significance of *Ethnicity* changed depending on the model specification, while in Chapter 9, ethnic-minority students were found to be significantly less likely to drop out compared to their white peers. These results suggest that perhaps there is some confounder that is moderating the effect of *Ethnicity* on the chances of achieving a successful/unsuccessful academic outcome, which has not been accounted for in the models fit to the data. For example, it may be that perhaps there is too much variability between the academic outcomes of different ethnicity groups within the "ethnic-minority" category. Future analyses should perhaps examine these groups separately, with care taken in the which variables to control for given the relatively low number of ethnic-minority to the number of white students in the school-leaver population.

In Chapter 7 there was evidence that some *Academic Cohorts* were perhaps more likely to be retained compared to others. However, given that the sizes of the cohorts considered in these models were relatively small (around 100 students each year) some volatility is expected here. *Academic Cohort* may be a significant if the cohorts affected by the COVID-19 pandemic were examined (2019/20 - 2021/22). The *Prior Attainment Points* of students increased with each *Academic Cohort* suggesting grade inflation amongst applicants.

10.1.4 Aim 4 - Identifying Appropriate Models

The fourth aim was address by Chapter 7, which compared different regression models, and Chapter 9, which compared regression and survival models, when applied to subsets of the **School-leavers dataset**. The results from these Chapters suggested that both regression and survival models were adequate fits to the data, although the regression models had some issues.

The first issue was that the regression models required the achievement of academic outcomes to be defined within a specified time frame for every student. For example, being retained at the end of first year and completing a Bachelor's with Honours degree within four years. This meant that students who took longer to achieve these outcomes were recorded as failures. If certain groups of students are more likely to repeat a stage of their degree programme than others, for example SIMD Quintile 1 or contextual offer students, then the results of the regression models will be biased against these groups.

The second issue was that the regression models had problems related to the retention/completion outcomes being so common ($> 10\%$). The Logistic regression derived odds-ratio estimates, which are harder to interpret and are frequently misinterpreted as risk-ratios by both a researcher and lay audience. This is problematic given that they exaggerated the risk-ratio estimates from the Modified Poisson and Log-Binomial models because the retention and completion outcomes were so common. The Log-Binomial models required starting values to be supplied to aid with convergence, though each Log-Binomial model applied to the **School-leavers dataset** was successful. The Modified Poisson models sometimes predicted probabilities greater than 1, again likely due to the retention and completion outcomes being so common paired with some very strong associations like *Prior Attainment Points* and *SIMD Quintile.* Despite both the Log-Binomial and Modified Poisson models both deriving risk-ratios, model estimates between each appeared to differ most when the outcome variable was more common.

Chapter 9 aimed to address these limitations by examining a different outcome variable and using survival methods. The outcome examined was time until a student dropped out of university which could be defined as a continuous or discrete random variable. Both were attempted, despite the follow-up time being naturally discrete (student drop-outs were recorded every academic session). When the discrete and continuous survival methods were compared to their equivalent regression method (Logistic, Modified Poisson, Log-Binomial), only small differences in the estimates were observed and all models adequately fit the data. This was likely because the prevalence of drop-out depended on each academic session, and the highest prevalence was 6% for first-year students. This also meant the odds-ratios could be interpreted as estimates of risk-ratios, since prevalence is less than 10%. Discrete survival methods are recommended for future analyses given that time until drop-out is a naturally discrete random variable and that survival methods can account for time-dependent covariates, unlike the regression methods. However, if regression methods were to be applied instead, it seems unlikely that incorrect conclusions would be drawn.

10.2 Future Research

It is hoped that the work presented here will allow future researchers to reproduce the **School-leavers dataset** and the results contained in this thesis, such that they can be expanded into future analyses. Key pieces of information were not available in the **School-leavers dataset**, though this is likely to change with future editions. For example, perhaps one of the most important predictors of success at University that was not included in the **School-leavers dataset** is students' academic performance, or "university-level" attainment. This information could take the form of average marks per registration session, final marks in each module that make up a degree programme, or marks across weekly assignments. It is highly likely that the models considered in this thesis would have better fit the data with access to any of these information. The survival models in particular, could have incorporated these new variables as time-dependent covariates. Such models could be used to update the risk of a student dropping out for each *Academic Session* they spend registered at the University. Examples of these models have already been tested on data from institutions in Germany [150] and Australia [160]. Time-dependent covariates would also likely be vital for any prediction model that aims to identify students at risk of dropping out. There is evidence to suggest that survival methods could be more useful than traditional machine learning algorithms for predicting dropouts at the early stages of a students journey [154, 155]. Machine learning models were attempted on the **School-leavers dataset** but did not provide satisfying predictions. If access to university-level attainment is secured, then it is recommended that Machine Learning methods also be investigated in future analyses.

This thesis also did not investigate the effects of other time-dependent covariates such as changing degree programme, repeating a stage of a degree programme, or taking a break in studies. It is of interest to measure the effects of these variables on a student's chances of success/failure at University. Attendance is another example of a time-dependent covariate, though it may prove to be difficult to access accurate historic attendance records for every student in the University.

With regards to socio-economic background, this thesis only considered one indicator: *SIMD Quintile*. Sections 2.3 and A.2 list a range of indicators that could be considered for future analyses. Of particular interest would be those used by the University of Strathclyde in its contextualised admissions criteria: attendance of a low-progression school, care-experience and caring responsibilities [56].

This thesis only considered Scottish school-leavers; it did not consider mature students, RUK or international students, those who entered via college, nor those who were admitted through a Widening Access initiative. The latter group of students is of particular interest for Widening Access efforts. At the University of Strathclyde for example, there are the Engineering Academy and Natural Sciences programmes which are targetted specifically at SIMD Quintile 1 and other eligible secondary-school students who do not meet the minimum entry requirements. Other examples of initiatives include the Top-up and Summer School programmes. Knowing the academic outcomes of these students would help identify whether or not these programmes are successful in developing students who can succeed in higher education.

Another area of research that has not been addressed here is whether or not minimum entry requirements are set at an appropriate threshold. In other words, are they too high or too low? This was discussed by Boliver et al. in 2017 [52] who argued for drastically lower minimum entry requirement thresholds for Scottish higher education institutions. Yet, since their introduction in 2019, most institutions have not changed these from around one to two grades less than the standard entry requirements.

Finally, these results are not representative of the higher education sector in Scotland, the UK nor internationally. It is recommended that future research prioritise similar analyses of data from other institutions to compare trends and relationships with those derived in this thesis. In particular, it is strongly emphasised that data from Scotland's 18 higher education institutions be compared to one another, if at all possible. In light of the Commissioner for Fair Access's recommendation that equal weight be given to Widening Access student's academic outcomes at University [7], cross-institutional analyses would be integral towards measuring progress on this front.

Appendix A

Widening Access Topics

A.1 Impact of COVID-19 Pandemic

Academic session 2019/20 saw the cancellation of secondary-level examinations across Scotland and the other UK nations [31, 32]. Instead, students were awarded results based on "teacher-assessed grades", where teachers would rank their students and give estimates of which grade band (A, B, C, D) they believed their students had achieved [164]. These estimates were then checked and approved by the SQA before being awarded to students on results day [164]. It was found that grades under teachers' estimates had inflated compared to previous years, and were subsequently revised down using the SQA's "alternative certification model" to ensure "fairness to all learners" and to maintain the "integrity and credibility of the qualifications system" [33]. This was ultimately reversed after backlash from the public and the original teacher estimates were re-instated [34, 35] . Critics pointed out that students who had attended schools with historically low attainment rates or came from more socio-economically deprived areas were disproportionately penalised compared to their more affluent peers [35]. For example, those from SIMD Quintile 1 areas had their grades revised down by 15.2 percentage-points compared to just 6.9 percentage-points for those from Quintile 5 [33, p. 69]. Ofqual, the awarding body for GSCE and A-level examinations similarly reversed their initial revised-down grades for students after protests [165]. The practice of teacher-assessed grades in Scotland was repeated for academic session 2020/21, but subsequently returned to traditional in-person examinations from session 2021/22 onwards [166–168]

The final qualifications awarded in Scotland during the pandemic experienced significant grade-inflation. The rate at which students attained an A to C grade at Higher increased from 74.9% in 2018/19 to 89.3% in 2019/20, or 14.4 percentage-points [36]. In 2020/21 this was 87.3%, while in 2021/22 – when students had returned to in-person exams – this was 80.3%, lower than the pandemic levels but still 3 percentage-points higher than pre-pandemic levels [36]. Similar levels of inflation were also observed within Advanced Highers. In response to the increased number of qualified applicants to higher education, the Scottish Government increased the number of funded places for Scottish domiciled students [37].

At the University of Strathclyde, the end of the second semester of academic session 2019/20 saw the suspension of all face-to-face teaching and some campus-based examinations [38, 39]. Disruption varied from programme-to-programme, though generally speaking, students only sat examinations that were absolutely necessary. For example, examinations that required accreditation to be given by awarding accrediting bodies or where there was insufficient evidence of a student's academic performance [38, 39, 41, 42].

Replacement tests took the form of "open-book" remote exams, where students were allowed access to approved "books, notes, and reference material" [40–42]. The "no-detriment" policy was also introduced, which modified the University's compensation policy to allow students who had achieved an average mark of 40% (lowered from 45%) but had failed one or more classes (raised from only one class) to "Pass by compensation" if they had achieved at least 30-39% in that class [47–49]. It also gave students extra favour in areas such as personal circumstances claims, flexibility in extension requests and discretionary credits (under certain circumstances allowing modules to be discounted if they bring down the student's average grade) [49].

Restrictions and policies (including no-detriment [49]) extended to academic sessions 2020/21 and some of 2021/22, where a "blended learning" was introduced that allowed for some in-person classes to go ahead [43–46]. However, in reality most students attended classes online due to Scottish Government restrictions, which were not completely lifted until the 21st March 2022 [169]. The return to traditional teaching and examinations did not occur until academic session 2022/23.

It was unknown at the beginning of the pandemic whether it would unfairly impact students from socio-economically deprived backgrounds. There were concerns over "digital poverty", the ability to support students face-to-face and of course the inability to run traditional outreach events [50].

A.2 Contextual Indicators

There are three broad categories of contextual indicators: Individual-level, school-level, and area-level.

Individual-level – Factors that identify a specific individual's contextual background. Examples include being in receipt of free-school meals or being a refugee/asylum-seeker. Due to the intensely personal nature of these indicators, they are often either the hardest to collect data on, or the hardest to verify. In some cases they are the best forms of identifying truly disadvantaged individuals given that they speak to an individual's experience rather than relying on aggregated data.

School-level – Factors that identify the contextual background of the school an individual attended. Examples include schools which have a large percentage of students in receipt of free-school meals, or those from low-income households. Other examples include attendance of a state/non-fee-paying school or attendance of a school with a low-progression to higher education rate. These indicators are popular since there are less barriers to access the necessary data. They also anonymise the nature of individual's deprived circumstances, and thus face reduced data protection concerns.

Area-level – Factors that identify the contextual background of the area an individual comes from. These include the Scottish Index of Multiple Deprivation (SIMD), The participation of local areas (POLAR) [170], tracking under-representation by area (TUNDRA) [171], and Acorn [172]. SIMD, POLAR and TUNDRA are government-made tools while Acorn was created and is maintained by CACI Inc. at the cost of a subscription. These area-level indicators are popular due to their ability to connect to postcode data – the most readily available data on students.

A.2.1 Participation of Local Areas (POLAR)

Participation of local areas (POLAR) is the area-level measure predominantly used by English higher education institutions and the Office for Students to identify how many young people within a particular area participate in higher education [170]. Similarly to SIMD, POLAR ranks areas by their participation rate which are then split into five groups or Quintiles. The interpretation for POLAR is similar to SIMD, where POLAR Quintile 1 represents areas that rank in the bottom 20% for progression to higher education. The most recent edition – POLAR4 – was released in September 2020 and will be the final release for POLAR [170]. POLAR is the measure used by the Office for Students for their own Widening Access targets to be achieved by 2039 [173].

As an area-level indicator POLAR also suffers from criticisms of bias, for example its high false-positive and false-negative rates in socio-economically diverse areas such as London [51, 53]. POLAR as a measure is also not as effective in Scotland given its high rates of participation in higher education [68]. As a result this indicator is used by the University of Strathclyde but only for students that register from the rest of the United Kingdom, while SIMD is used for Scottish-domiciled students [56].

A.2.2 Free-School Meals (FSM)

Free-School Meals (FSM) are available to the children (aged 5 to 16) of those on governmental support schemes such as Universal Credit, Jobseeker's Allowance, and Child Tax Credits [174–176]. FSM is considered a highly valid and highly reliable indicator of an individual student's disadvantage [51, 52]. Data from 2007-9 showed only 13.3% of pupils on FSM attained five or more SQA Highers by the end of sixth year, compared to 47.6% for those not on FSM [52]. In England, FSM is available to all students within state-funded schools in the academic stages of reception to year 3 due to the Children and Families Act 2014 [175]. In 2021, The Scottish Government committed to universal FSM for all pupils from P1 to P5, and to pupils in P6 and P7 by 2024, though the plan to extend this to secondary school students has been postponed due to budgetary constraints [176, 177]. Were this to be rolled out to all secondary school students in Scotland, it could render obsolete the use of FSM as a contextual indicator. In 2021, UCAS rolled out its Modernised Contextual Data Service that aimed to give institutions better access to applicants FSM status, however this was only for applicants within schools in England. In Scotland, data on free-school meals is held by each local authority with no existing centralised service that can provide access to this data. The University of Strathclyde does not consider FSM data in its basket of indicators [56].

A.2.3 Other Popular Indicators

Other examples of popular indicators include being a refugee/asylum-seeker, being the first person in the family to attend higher education (known as "first-generation"), indicators related to parents' backgrounds (occupation, education, etc.), attending a state versus fee-paying school, and household income. None of these are considered within the University of Strathclyde's basket of indicators.

Refugee/asylum-seeker status is considered to be a valid indicator that could be verified on a case-by-case basis [51]. First-generation and school status were cited in some studies aiming to measure inequalities within education [74, 81, 83, 84]. First-generation is considered by Boliver et al. [51] to be an invalid and unreliable indicator for disadvantage, while school status was considered an invalid indicator with the caveat that attending a fee-paying school could be used as a disqualifier for Widening Access support. Indicators related to parent's background also suffers from validity and reliability issues [51]. Household income, while a very valid and reliable indicator in the opinion of Boliver et al. [51] would require action from the government to set in place the necessary infrastructure to verify and distribute data. They point out that this could be a contentious policy to approve given the sensitivity of the data [51].

A.2.4 Summary of indicators used by Scottish institutions

In 2017, Boliver et al. [52] provided a table of the indicators that were in use across each of Universities Scotland's member institutions [22]. Given that several years have passed and more research on the suitability of various indicators has been published, it is of interest to see whether each of Scotland's institutions have changed or kept the same approach to contextualised admissions as they had in 2017. Here we provide a table of all indicators used within contextualised admissions in 2023 and discuss any changes (Tables A.1 and A.2). This list is representative of the admissions policies and website guidance pages that could be online through each institution's website [56, 178–197].

An indicator was considered "in use" if it was explicitly stated as one of the indicators a university used in their contextual admission policies, and "not in use" otherwise. Some indicators were referenced within documents in sections related to Widening Access but were not explicitly referenced as used for contextualised admissions - these were not counted for the purposes of Tables A.1 and A.2. Some indicators referenced by institutions were grouped together, such as "Priority School" which referred to any indicator that targeted certain schools such as the Schools for Higher Education Programme (SHEP), FOCUS West, etc. In total, there were 39 unique indicators identified in to be in use: 10 area-level, 1 school-level, and 22 individual-level indicators (Tables A.1 and A.2). The indicator for "Participation in a WA programme" was considered its own category of indicator due to the variety of eligibility criteria between different Widening Access programmes. This indicator also did not include any programmes encompassed in the definition of other indicators such as "Priority School" or "Mature/SWAP/Access students". Every institution used SIMD Quintile 1 and care experience as indicators of disadvantage since they were recommended by the Commission on Widening Access [3]. The most popular indicators outwith these two were (in decreasing order of popularity): Caring responsibilities, Priority School, Estranged students, SIMD Quintile 2, Refugee/Asylum-seeker, and

Participation in WA/outreach programme were used by more than half of institutions. Free-school meals – despite strong recommendations for its use by Boliver et al. [51] – was only used by 5 institutions, likely highlighting the issues around access to this information.

The median count of indicators used by institutions was 7.5. Institutions which used more than this were (in increasing order): Edinburgh Napier (8), Glasgow School of Art (9), Edinburgh (9), Abertay (10), Heriot-Watt (11), Scotland's Rural College (13), St. Andrews (14), Dundee (14), and Aberdeen (16). Five of these institutions come from outside Scotland's major metropolitan areas of Edinburgh and Glasgow. Institutions which used less than the median count of indicators were (in decreasing order): Glasgow (7), Glasgow Caledonian (7), Stirling (7), Highlands and Islands (6), Queen Margaret (6), Royal Conservatoire of Scotland (6), Strathclyde (6), West of Scotland (6), Robert Gordon (5) – only two of which come from outside Edinburgh and Glasgow. Five institutions used separate indicators for rest of UK students: St. Andrews (5), Aberdeen (1), Edinburgh (1), Heriot-Watt (1), and Strathclyde (1). These indicators included ACORN, POLAR4, and each of the devolved nations' (England, Wales, Northern Ireland) Index of Multiple Deprivation (IMD) indicators.

Seven institutions used indicators in exceptional cases – where a student was told they "might" be eligible for a contextual offer, or they were only guaranteed an offer if they satisfied more than one of these indicators. This category of indicator was the broadest and included SIMD Quintile 2, Caring responsibilities, being an Estranged student, and others. The institutions that used these indicators were Dundee (8), Heriot-Watt (8), Glasgow Caledonian (3), Queen Margaret (3), Abertay (1), Edinburgh (1), and Glasgow (1).

Table A.1: Indicators in use across Universities Scotland's 18 members (1 of 2). Does not include the Open University. Y = Yes, N = Not in use, E = Used in exceptional cases, RUK = only used for Rest of UK applicants.

Indicator	Aberdeen	Abertay	Dundee	Edinburgh	Edinburgh Napier	Glasgow	Glasgow Caledonian	Glasgow School of Art	Heriot-Watt	Highlands and Islands	Queen Margaret	Robert Gordon	Royal Conserv. of Scotland	Scotland's Rural College	St. Andrews	Stirling	Strathclyde	West of Scotland
Outreach Indicator																		
Participation in WA/outreach programme		Y	E	Y		E	E			Y				Y		Y	Y	
Area-level Indicators																		
ACORN				Y												RUK		
English IMD																RUK		
Gaelic-speaking area															Y			
Low-participation area		Y																
N.I. IMD																RUK		
Remote & Rural	Y							Y										
SIMD Quintile 1	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y
SIMD Quintile 2			E	E		Y	E	Y	E				Y	Y	Y	Y	Y	Y
Welsh IMD																RUK		
POLAR4			RUK					RUK							RUK		RUK	
School-level Indicators																		
Priority School (low-prog/high FSM/SHEP/FOCUS West...)	Y		Y	Y		E	Y	E	Y		Y	Y	Y	Y	Y	Y	Y	
Individual-level Indicators																		
*Under-represented groups														Y				
Armed forces (applicant)										E			Y					
Armed forces (parent)													Y					
Care-experience	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y

Table A.2: Indicators in use across Universities Scotland's 18 members (2 of 2). Does not include the Open University. Y = Yes, N = Not in use, E = Used in exceptional cases, RUK = only used for Rest of UK applicants.

Indicator	Aberdeen	Abertay	Dundee	Edinburgh	Edinburgh Napier	Glasgow	Glasgow Caledonian	Glasgow School of Art	Heriot-Watt	Highlands and Islands	Queen Margaret	Robert Gordon	Royal Conserv. of Scotland	Scotland's Rural College	St. Andrews	Stirling	Strathclyde	West of Scotland
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Caring responsibilities	Y		Y	Y	Y	Y	E	Y		E	Y		Y	Y	Y	Y	Y	Y
College applicant			Y															
Custodial sentence (parent)																		
Disability (mental)	E																	E
Disability (physical)	E																	E
Disruption to education	E																	E
Education Maintenance Allowance							E											
Estranged students	Y	Y	Y	Y		Y	E			E			Y	Y			Y	Y
Ethnic minority						Y												
Few/no formal qualifications																		
First generation to university								Y										
Free-school meals	Y						E						Y					Y
Mature/SWAP/Access applicant	E	Y	Y				E					Y	Y		Y		Y	E
Recognition of prior learning																		
Refugee/Asylum-seeker	Y		Y	Y	Y	Y	E						Y	Y			Y	Y
Transgender	E																	E
Travelling community													Y					
Vulnerable group	E																	E

Appendix B

Data Protection Impact Assessment

To gain access to the **School-leavers dataset**, a Data Protection Impact Assessment had to be written and agreed upon between the investigators and data owners (Strategy & Policy, University of Strathclyde). This appendix includes the details of the latest DPIA agreement between all parties.

PROFESSIONAL SERVICES STRATEGY & POLICY

University of Strathclyde Glasgow

Data Protection Impact Assessment Form (DPIA)

Before completing the form you should read the relevant guidance on our Sharepoint pages on how to conduct a DPIA. If you are unsure if you should complete a full DPIA you should complete the Screening Questions.

For assistance in completing this form please follow the step by step guidance in Completing the DPIA Template – Guidance for Staff.

Complete this form in as much detail as you can. You should then submit it to the Information Governance Unit (IGU). If you have any questions whilst you are completing the form please contact the IGU.

You can find more information about DPIAs in the guidance from the Information Commissioner's Office.

Please note: the term 'project' has been used throughout for simplicity. This should be understood to mean any project/policy/system etc. which involves the processing of personal data.

Governance and Contacts

Project title/brief description	The title of the PhD project is "The relationship between widening access and success at university: a data-driven statistical investigation" A large focus of this project is how to meet the recommendations made by the Commission in its 'Final Report of the Commission on Widening Access - Blueprint for Fairness'. This includes setting access thresholds which applicants from the most deprived backgrounds should be assessed.
Name and job title of person responsible for project (project manager)	Dr Louise Kelly Senior Lecturer and Associate Dean (Recruitment and Admissions) Department of Mathematics and Statistics, University of Strathclyde
Name and job title of team/departmental contact (if different to above)	Nathan Burns PhD Student

	Department of Mathematics and Statistics, University of Strathclyde
Faculty/School/Dept/Professional Service Area	Department of Mathematics and Statistics, University of Strathclyde

Version Control and History

Version	Date	Author	Reason for change
1.0	12/11/2019	Kayleigh Kerr	First Draft
2.0	13/10/2020	Nathan Burns	Department personnel change, Section B.5 storage retention end date changed,
2.1	04/11/2020	Nathan Burns	Amendments to sections based on comments of Fiona Wilbraham. Sections are as follows: A.1, A.2, A.3, B.1.1, B.1.2, B.3, B.5, B.6.1, B.6.2, B.6.4, B.6.6, C.1.2, C.1.3, C.1.4, C.1.7, E, F
2.2	09/11/2020	Nathan Burns	Amendments to sections based on change of how data will be stored, updated, secured, and accessed (Due to inability to access a campus office computer). Sections are as follows: B.4.1, B.5.2, B.6.1, B.6.2, B.6.3
2.3	14/03/2024	Nathan Burns	Updates to sections A.2, B.3.1, 3.4.1, B.5.1; which clarify the language surrounding the source and formatting of the data, as well as the descriptions of relevant data fields. Fixed minor grammatical errors. Updated the expected end date of the PhD research.
3.0	14/03/2024	Nathan Burns	Additional data fields requested relating to student marks derived from Core Student Record. Amendments to sections: A.2, B.3.1.

Section A – Description of the project/processing and consultation

A.1 Describe the project/reason you will be processing personal data. **What is the purpose and why are you doing it?** **How will you collect/access the personal data? Include any data flow diagrams/process maps.**
Scottish Government have set out recommendations and guidelines to achieve fair access to university. In line with this Strathclyde have introduced contextual admissions to widen access for those currently underrepresented in higher education. This requires information about the relationship between these widening access (WA) covariates and success at university to make informed decisions on the minimum entry requirements. I will be assisting Strathclyde in investigating this relationship as the main aim of my PhD project, and leading from this establish a statistical model that can be used to set and test access thresholds. The funding for the PhD is the Research Excellence Award (the award letter has been attached at the end of this document). Gaining access to this data is necessary for the study since data-driven evidence is essential in supporting the setting of access thresholds and understanding the relationship between factors which contribute to WA status and successful completion of a degree programme. The data is collected by the university and I will be processing this data on behalf of the university. This data is a derivation of the Core Student Record, therefore the custodian of the data is Gianna Devin. Once given access to this data it will only be accessed through my personal computer within the university.

A.2 Describe the type of personal data you will be processing. Explicitly state the category of individual to whom it relates e.g. staff/students/applicants. **State in detail the type of data e.g. name/address/ethnicity/disability etc.** **If special category data or criminal conviction data is involved state this where indicated. Special category data relates to: race; ethnic origin; politics; religion; trade union membership; genetics; biometrics (where used for ID purposes); health; sex life; or sexual orientation.**
Category of individual, e.g. staff/student/applicants/research subjects: Applicants to Strathclyde University The data being requested is Strathclyde's Core Student Record formatted to HESA standards on potential, current and former students. This extends back 10 years. Access to data held on former students is necessary to allow for adequate analysis of any changes in the success at university since the introduction of contextual admissions. Access to data held on potential students will allow analysis to understand the differences in those who apply and enrol and those who do not enrol. Data held on current students will include those who have received contextual admissions and so is crucial for the study. **Personal data:** • Unique Student identifiers including registration numbers: This will be used as an identity variable in the analysis rather than the individual's name. • Date of birth and Gender: These are crucial variables in the analysis as similar analyses have identified these as significant variables in the success at university. • Postcode and Term-time postcode: These would facilitate the derivation of a student's Scottish Index of Multiple Derivation and Urban-Rural status. These are crucial variables in

DPIA Template v3.0- Updated May 2019

the analysis as similar analyses have identified these as significant variables in the success at university.

- Student marks in university examinations: Crucial variables in understanding a student's chances of success at university.

Special category data: N/A

Criminal conviction data: N/A

A.3 Stakeholders and consultees **Who is involved in the project and who may be affected by it – both internal and external stakeholders.** **Who has been consulted during the design/review of this process? Describe the consultation process and feedback.**
Dr Louise Kelly, Dr David Young (Mathematics and Statistics); Strategy and Policy; Al Blackshaw (Widening Access Team); Dr Andrea Sherriff (University of Glasgow Dental School)

Section B – Compliance with Data Protection Legislation

B.1 Principle 1: Lawfulness, Fairness and Transparency

B.1.1 What is your lawful basis for processing personal data? **- if you are processing special category data you must identify an additional lawful basis. Include this in the box below** **- if processing criminal conviction data you must choose a relevant lawful basis for 'personal data' and also identify the relevant section from the DPA 2018 (consult IGU in advance if required). Include this in the box below.**
Personal data: Public task.
Special category data i.e. data relating to: race; ethnic origin; politics; religion; trade union membership; genetics; biometrics (where used for ID purposes); health; sex life; or sexual orientation: N/A
Criminal conviction data: N/A

B.1.2 What information is provided to data subjects to ensure that they are aware of this processing? When and how are they provided with this information?

Include links to anywhere you provide information on how you will use personal data, e.g. privacy notices/handbooks/contracts etc. Or attach a copy.
Strathclyde publish privacy notices to inform students and applicants on how their data will be used. Research purposes is listed within this. https://www.strath.ac.uk/whystrathclyde/universitygovernance/accesstoinformation/dataprotection/privacynotices/

B.2 Principle 2: Purpose Limitation

B.2.1 Are you using existing data for a new purpose, or will you use the data you collect in a way which is different to that described in A.1? If so, explain the different purposes here.
All aims described in A.1

B.3 Principle 3: Data Minimisation

B.3.1 You should have listed all the types of data you are collecting in A.2. Indicate why you need to collect this data. **Where you have to collect special category/criminal conviction data you must explicitly state why you require this data.**
• Unique student identifiers including registration number: This will be used as an identity variable in the analysis rather than the individual's name. • Date of birth and Gender: These are crucial variables in the analysis as similar analyses have identified these as significant variables in the success at university. • Postcode and Term-time postcode: These would facilitate the derivation of a student's Scottish Index of Multiple Derivation and Urban-Rural status. These are crucial variables in the analysis as similar analyses have identified these as significant variables in the success at university. • Student marks in university examinations: Crucial variables in understanding a student's chances of success at university. The study involves testing which personal characteristics affect success at university and so processing of some personal data is essential.

B.3.2 How will you review the collection/use/processing of personal data to determine whether it is still required?
All data will be included in the analysis and will be required throughout the study.

B.4 Principle 4: Accuracy

B.4.1 Describe how the accuracy of personal data will be monitored and maintained.
The dataset is the Core Student Record so will be accurate at the time of gaining access. The current year's data will be provisional and not final until the last quarter of the following year.

Updates or amendments that will be made will be passed on through a new release of the data every year, if required, until the end of the PhD.

B.5. Principle 5: Storage Limitation

B.5.1 Indicate the retention rule(s) which will govern how long this data is retained for. Include any relevant links/appendices.
The data will be held for the duration of the PhD, which is expected to end in July 2024. Afterwards the data will be held by the university in accordance with the data management plan.

B.5.2 How will the retention rules be implemented, for records in any formats, e.g. hard copy or electronic? Can any systems delete data according to the retention rules?
There will only be one copy of this dataset, which will be stored on a shared cloud drive in the university. This will be held in accordance with the data management plan for the project.

B.6 Principle 6: Security, Integrity and Confidentiality

B.6.1 Describe who will have access to the personal data?
Predominantly myself as the PhD student – Nathan Burns. Peter Black from Strategy and Policy, as part of the data handover, will have access and, by default, the computer security officers.

B.6.2 Describe what organisational controls will be in place to support the process and protect the personal data.
The dataset will be saved on a password protected shared cloud drive within the University. There will be no copies made on any other devices.

B.6.3 Describe what technical controls will be in place to support the process and protect the personal data.
The file itself will be password protected and stored on an i-drive, which is secured by the university.

B.6.4 Will the personal data be shared with any other organisation(s)?	
Yes	
No	Personal data will not be shared with any other organisation.

B.6.5 If yes to B.6.4 do you have a data sharing/processing agreement in place?	
Yes **(if yes, attach a copy)**	
No **(if no, why not?)**	

B.6.6 Will personal data be transferred from/to the UK? NB this includes storing data.	
Yes	
No	No data will not be transferred.

B.6.7 If yes to above, state where the data is transferred to/from?

Section C – Complying with Individual Rights

Data protection legislation provides data subjects with certain rights about their data.

The Information Governance Unit (IGU) will generally co-ordinate all requests from data subjects who wish to exercise their rights but it is important that <u>local</u> processes are designed to support these.

C.1 Not all rights apply in all circumstances but you should provide relevant information to allow the IGU to assess if rights could be upheld, <u>if required.</u> In some cases you may not be able to comply with rights, you should explain why not.

C.1.1 Right to be informed
How will you provide information to individuals about how their information will be used? **You should already have provided a full explanation in B.1.2. If so, state this below. If you will not provide information to individuals, state why not.**
See B.1.2.

C.1.2 Right to access personal data
The IGU centrally manages any requests for access to personal data (Subject Access Requests or SARs). **If required, are you able to identify and retrieve the personal data within 2 weeks?** **Identify the processes you have in place, which will enable you to comply with a SAR.**
The names of the individuals will not be available so it will not be possible to comply with this. However, as this data is collected and owned by Strathclyde, I do not believe that I, as a data processor, would be responsible for any Subject Access Requests.

C.1.3 Right to rectification
Are you able to correct personal information, if it is incorrect/incomplete? Can you do this to meet an internal timescale of 2 weeks? **Identify how you would correct/complete inaccurate information.**
The dataset being used for the project is the Core Student Record from Strathclyde. These data should be accurate and up to date.

C.1.4 Right to erasure
Could you erase personal data? Describe how this would be done. **If not, please state why not (this may be for technical or operational reasons).**
It will not be possible to identify individuals, so it would be impossible to comply with this.

C.1.5 Right to restrict processing
Do you have appropriate methods in place to restrict the processing of personal data on your systems, e.g. temporarily moving the data to another system; making the data unavailable to users; or temporarily removing published data from a website?
As mentioned above, I will not be able to identify individuals. Restricting processing would require restricting the processing on the entire dataset. If this were for an extended period this would severely affect the project.

C.1.6 Right to data portability
Could you supply information, that an individual has provided to the University, in a structured, commonly-used and machine readable format? **Could you provide data in this format to meet an internal timescale of 2 weeks?**
This is not applicable here. The data controller would comply with this right if necessary.

C.1.7 Right to object to processing
If an individual objects to the use of their personal data, particularly in relation to direct marketing, are you able to cease processing their data? **Describe how you could comply with an objection to the use of data.**
Without a way to identify individuals this would not be possible.

C.1.8 Rights related to automated decision making, including profiling
If any automated decision-making or profiling exists, i.e. where no human intervention is involved in the decision-making process, do you: - **give individuals information about the processing;** - **have simple ways for them to request human intervention or challenge a decision; and** - **carry out regular checks to make sure that your systems are working as intended?** **Describe how you do the above. If no automated decision-making or profiling exists, please state this.**
No automated decision making or profiling exists. The data will be used to look at the relationship between the characteristics of individuals and not used to make any decisions.

Section D – Direct Marketing

D.1 Will your project involve direct marketing by electronic means, e.g. phone/fax/email/SMS?	
Yes	
No	No direct marketing involved in the project

D.2 If yes to D.1 You must ensure that any direct marketing complies with the Privacy and Electronic Communications Regulations (PECRs). State how you will ensure the processing will comply with PECRs.

PROFESSIONAL SERVICES STRATEGY & POLICY

University of Strathclyde Glasgow

Section E - Identifying Privacy Risks and Actions

Any issues that have identified through the project and/or DPIA process should be listed here. These may have been identified through the consultation process, by a stakeholder or when completing the DPIA. For example, you may have identified that you are sharing data but you have no data sharing agreement in place.

You can add in as many additional rows as required.

Risk no.	Privacy risk/issue	What outcome is required to mitigate the risk?	Action required to meet outcome	Action assigned to

Section F - Mitigating Actions and Risk Assessment

This section records the mitigating actions taken and associated risk assessment in relation to privacy risks. This section is intended to indicate if the mitigating actions have been completed and records any residual risk and whether it is acceptable.

Ref no	Action required to meet outcome	Specific actions taken. Outcome met?	Date Completed	Residual Risk Rating (High, Medium, Low)	Risk Owner	Risk Accepted Yes/No	Comments

If any of the risks **cannot** be accepted then you must provide comments in the table above as to next steps.

Section G – Record DPO advice and sign offs

You must submit your DPIA to the IGU for comment, along with any supporting documentation. The IGU will respond with comments/actions.

Assessment - IGU

IGU Assessment Summary:	
Date Completed:	
DPO Opinion/Advice:	

DPIA Template v3.0- Updated May 2019

Date Completed	

Assessment - Department

DPO Opinion/Advice Accepted? If not, why not? NB this section must be completed by a senior member of staff in the relevant area.	
Date Completed:	
New Review Date:	

Appendix C

Additional Datasets Information

C.1 The Prospectus Dataset

The **Prospectus dataset** was developed by the Widening Access Team (University of Strathclyde). It was created through manual recording on an excel spreadsheet of the entry requirements shown on each prospectus release from 2015/16 - 2023/24. A list of the variables contained in the data is shown in Table C.1. It is possible there are some manual data entry errors as this data was not taken from a centrally maintained database or independently verified by administrative staff. There were missing data that varied in size between *Academic Sessions*. A summary of the missing data contained in each *Entry Path* per academic session is shown in Table C.2. Only the "Std. Entry" column was used in the **School-leavers dataset** since it had the most complete data for each *Academic Session*.

Table C.1: Summary of the Undergraduate Prospectus data gathered from 2015/16 - 2022/23. There are n = 6,331 rows each representing a unique entry path to a given degree programme and academic year.

	Variables	**Role**	**Unique Values**	**Description**
1	UCAS Code	Primary Key/ID	266	UCAS degree programme code
2	Academic Year	Variable	9	Academic year of entry upon acceptance
3	UCAS Code Description	Variable	264	Degree programme title
4	Level	Variable	18	Bachelor's (with Honours) or Master's in Engineering, Arts, etc.
5	Entry Path	Variable	8	Standard/Minimum entry, 1st/2nd sitting
6	Attainment	Variable	2	Whether Higher or Advanced Higher
7	Grade	Variable	33	Grade Profile asked for (e.g. AAAA)
8	Attainment Points	Variable	17	Grade Profile converted into "Simple Points"
9	Entry Tariff Total	Variable	19	Total "simple points" asked for by programme for stipulated year of entry
10	Num Grades	Variable	6	Total number of grades asked for by programme for stipulated year of entry

Table C.2: Count of degree programmes by year, and the corresponding number (and proportion) of missing entry requirements data.

Entry Path	Academic Session	Programme Count	Missing Data
Std Entry	2015-16	214	1 (0.5%)
	2016-17	235	1 (0.4%)
	2017-18	234	1 (0.4%)
	2018-19	228	2 (0.9%)
	2019-20	230	2 (0.9%)
	2020-21	230	2 (0.9%)
	2021-22	235	5 (2.1%)
	2022-23	227	4 (1.8%)
	2023-24	227	4 (1.8%)
Std Entry 2nd Sitting	2015-16	214	78 (36.4%)
	2016-17	235	78 (33.2%)
	2017-18	234	81 (34.6%)
	2018-19	228	79 (34.6%)
	2019-20	230	80 (34.8%)
	2020-21	230	78 (33.9%)
	2021-22	235	134 (57%)
	2022-23	227	82 (36.1%)
	2023-24	227	83 (36.6%)
Min Entry	2015-16	214	214 (100%)
	2016-17	235	235 (100%)
	2017-18	234	222 (94.9%)
	2018-19	228	228 (100%)
	2019-20	230	230 (100%)
	2020-21	230	3 (1.3%)
	2021-22	235	5 (2.1%)
	2022-23	227	4 (1.8%)
	2023-24	227	5 (2.2%)
Min Entry 2nd Sitting	2015-16	214	214 (100%)
	2016-17	235	235 (100%)
	2017-18	234	234 (100%)
	2018-19	228	228 (100%)
	2019-20	230	230 (100%)
	2020-21	230	78 (33.9%)
	2021-22	235	145 (61.7%)
	2022-23	227	141 (62.1%)
	2023-24	227	141 (62.1%)

C.2 The Demographic and Socio-economic Variables

Table C.3: Binary Groupings for all ethnicities in the School-leavers dataset.

Ethnicity Description (UCAS)	**Ethnicity (School-leavers dataset)**
Arab	Ethnic-minority
Asian - Bangladeshi	Ethnic-minority
Asian - Chinese	Ethnic-minority
Asian - Indian	Ethnic-minority
Asian - Other	Ethnic-minority
Asian - Pakistani	Ethnic-minority
Black - African	Ethnic-minority
Black - Caribbean	Ethnic-minority
Black - Other	Ethnic-minority
Gypsy, Traveller or Irish Traveller	Ethnic-minority
Other	Ethnic-minority
Other Mixed	Ethnic-minority
White and Asian	Ethnic-minority
White/Black African	Ethnic-minority
White/Black Caribbean	Ethnic-minority
White	White
White - British	White
White - Irish	White
White - Scottish	White
White - other background	White
Information Refused	NA
Not given (Dom=Home)	NA

Table C.4: Binary Groupings for all disability statuses in the School-leavers dataset.

Disability Description (UCAS)	**Disability Status**
A condition or impairment not listed	Disabled
A hearing impairment (e.g. deafness or partial hearing)	Disabled
A learning difference (e.g. dyslexia, dyspraxia, or AD(H)D)	Disabled
A long-term illness or health condition which may involve pain or cause fatigue, loss of concentration or breathing difficulties - including any effects from taking associated medication.	Disabled
A mental health condition, challenge or disorder (e.g. anxiety or depression)	Disabled
A physical impairment or challenges with mobility (e.g. climbing stairs or uneven surfaces), or dexterity (e.g. using a keyboard or laboratory equipment)	Disabled
A social, behavioural or communication impairment (e.g. an autistic spectrum condition or Tourette's Syndrome)	Disabled
A visual impairment uncorrected by glasses (e.g. blindness or partial sight)	Disabled
Two or more impairments or conditions	Disabled
None	None

C.3 The Attainment Data

Table C.5: List of all attainment codes removed from the Attainment on Entry data table.

Attainment Code	Description
ACCESS	ACCESS
ACCESS8	Access to HE Diploma (Open Awards)
BA	Bachelor of Arts
BA-HONS	Bachelor of Arts with Honours
BACC-HONS	Bachelor of Accounting with Honours
BENG	Bachelor of Engineering
BENG-HONS	Bachelor of Engineering with Honours
BSC	Bachelor of Science
BSC-HONS	Bachelor of Science with Honours
BTECDIP90	Pearson (BTEC) Diploma 90 Credit
BTECSUBDIP	Pearson (BTEC) Subsidiary Dip (was National Award 60+)
CERT-HE	Certificate of Higher Education
DIP-HE	Higher Education Diploma
DIP-PROF	Professional Diploma
DIP-UG	Undergraduate Diploma
EXT_PRO	EXTENDED PROJECT
FOUNDATION	Foundation course at HE
HNC	HNC
HND	HND
MA	Master of Arts
MA-ARCH-I	Master of Architecture

MA-I	Master of Arts
MSC-I	Master of Science
MUS-T6	Music Theory Level 6
MUS-T7	Music Theory Level 7
NONUKQUAL	Non UK Qualification - level unknown
NVQ	National Vocational Qualification
OPENUNIUN	Open University Unit
PG-CERT	Postgraduate Certificate
PG-DIP	Postgraduate Diploma
PRE-U-CERT	Practice Certificate
PROF-QUALS	Professional Qualifications
SWAP	Scottish Wider Access Programme

C.4 Prior Attainment Points/Quintile

UCAS employs a point system called "UCAS Tariff Points" which assigns Highers as: A - 33 points, B - 27 points, and C - 21 points (gap of 6 points between each); and Advanced Highers as: A - 56 points, B - 48 points, and C - 40 point (gap of 8 points). However, it was noticed that larger point gaps between the discrete grades led to "gap inflation" - where students with many "lower-quality grades" would appear to have equivalent point totals to students with "higher-quality" grades, when it could be argued that these were not equivalent. This discrepancy was most apparent when comparing students' points to the entry requirements of the relevant degree programmes (see Section 4.5 for more details on how entry requirement data were gathered and cleaned). For example, given entry requirements at AABB, it would be debatable whether an attainment of ABBCC would be deemed enough to warrant an offer of entry with one interpretation being that it would be on the boundary of consideration. If we were to compare their equivalent UCAS tariff point totals, these would be 120 and 132, respectively, meaning that the attainment would be comfortably over the threshold. This is an example of bias due to both gap inflation and comparing an unequal number of grades (4 vs 5). To counteract this problem, it was decided to introduce a "simple" tariff system that assigned points accordingly: A - 3 points, B - 2 points, C - 1 point[1], which seemed to mitigate these errors for the most part (using the above example, AABB would be 10 points and ABBCC would be 9 points, now under the threshold). Any Higher or Advanced Highers at a grade of D or less were ignored. The interpretation of simple points were also easier than UCAS tariff points. A single point increase can be interpreted as an increase in one grade (B to A) or the attainment of an extra C grade (AA to AAC).

C.4.1 Prior Attainment Quintile Tie-breakers

Tie-breaker rules were used to rank students within the same *Academic Cohort* by their prior attainment from secondary education. Having thirteen tie-breaker rules was sufficient to rank all students into their correct *Prior Attainment Quintile*. Defining *Prior Attainment Quintile* in this way meant that there is a near-equal distribution of students within each quintile across all *Academic Cohorts* (Figure C.1).

[1]Applied to both Highers and Advanced Highers.

Table C.6: List of 13 tie-breaker rules used to rank, within each Academic Cohort, students' prior attainment at secondary school. Listed in order of decreasing priority.

Priority	Variable	Arrange by
1	*All Highers Reg. Points*	Maximum Value
2	*All Adv. Highers Points*	Maximum Value
3	*Count of A Grades at Higher*	Maximum Value
4	*Count of B Grades at Higher*	Maximum Value
5	*Count of C Grades at Higher*	Maximum Value
6	*Count of D Grades at Higher*	Maximum Value
7	*Count of F Grades at Higher*	Minimum Value
8	*Count of A Grades at Adv. Higher*	Maximum Value
9	*Count of B Grades at Adv. Higher*	Maximum Value
10	*Count of C Grades at Adv. Higher*	Maximum Value
11	*Count of D Grades at Adv. Higher*	Maximum Value
12	*Count of F Grades at Adv. Higher*	Minimum Value
13	*All Highers Appl. Points*	Maximum Value

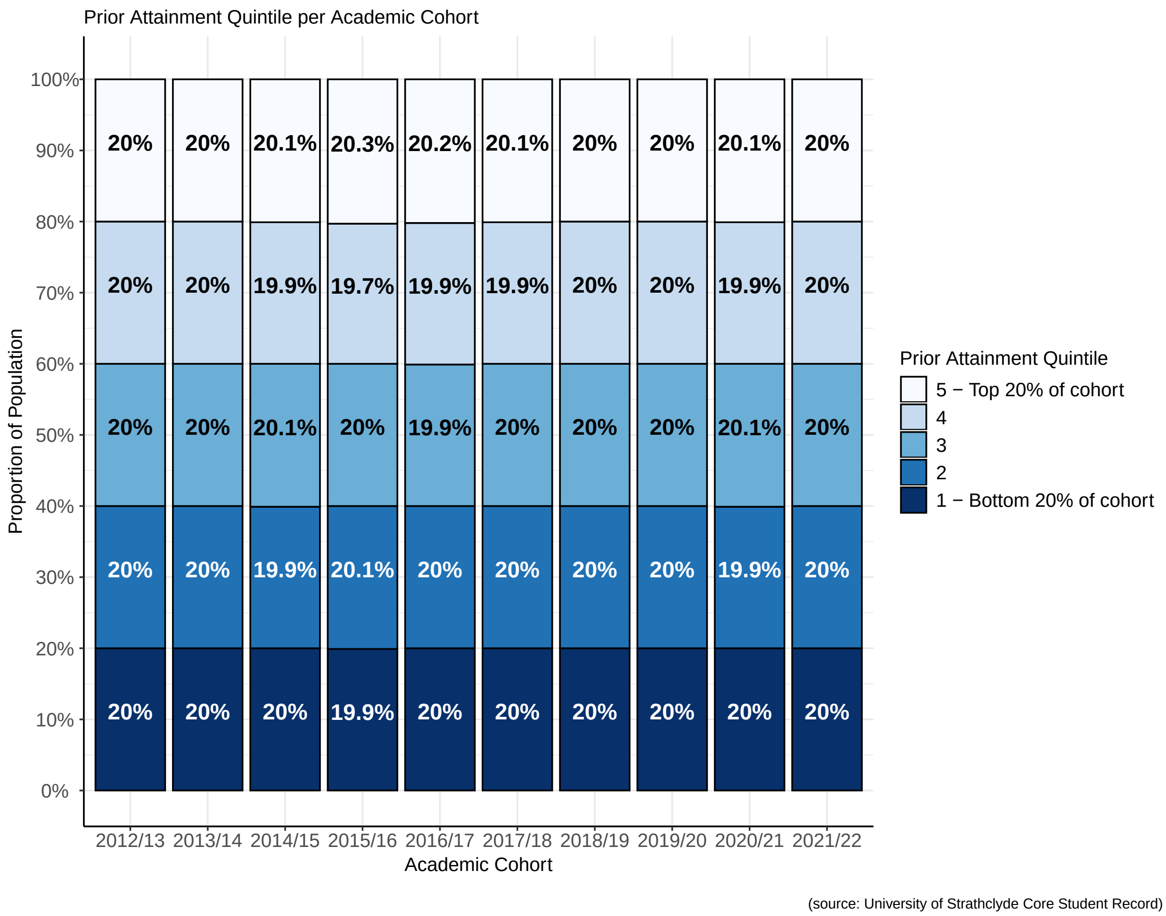


Figure C.1: Proportion of school leavers from each Prior Attainment Quintile within each Academic Cohort.

C.5 Faculties, Departments, and Degree Programmes

The Faculty of Engineering has 8 unique departments (Table C.7), while HaSS has 7 departments, and Science has 6 departments. The Faculty of Business has no departments since all degree programmes are managed centrally by the faculty.

Table C.7: List of departments per Faculty within the University of Strathclyde.

Faculty	**Department**
Business	-
Engineering	Architecture Biomedical Engineering Chemical and Process Engineering Civil and Environmental Engineering Design, Manufacturing and Engineering Management Electronic and Electrical Engineering Mechanical and Aerospace Engineering Naval Architecture, Ocean and Marine Engineering
HaSS	Education Government & Public Policy Humanities Law Psychological Sciences & Health Social Work & Social Policy Centre for Lifelong Learning
Science	Computer and Information Sciences Electronic and Electrical Engineering Mathematics and Statistics Physics Pure and Applied Chemistry Strathclyde Institute of Pharmacy and Biomedical Sciences

In the **School-leavers dataset**, students could be registered to one of 306 unique degree programmes. The Faculty of HaSS was assigned 167 of these programmes, which is the most out of the four faculties. This was followed by Business with 69 programmes, Science with 38 and then Engineering with 32. Note that these counts are not inclusive of all programmes run by the University in this time period. For example, it does not include the MPharm Pharmacy, MPhys Physics with Advanced Research, Natural Sciences, or Engineering Academy degree programmes since these were removed (see Section 3.5). A more complete list of the degree programmes offered by the University of Strathclyde for full-time undergraduates between 2012/13 and 2021/22 is provided in Tables C.8 to C.15.

Degree programmes at the University of Strathclyde are delivered by a single department, or jointly between two or more departments. These departments may or may not belong to the same Faculty. For example, "MEng Chemical and Process Engineering" is delivered primarily by the "Department of Chemical and Process Engineering" (Faculty of Engineering) but also by the "Department of Pure and Applied Chemistry" (Faculty of Science). In the **School-leavers dataset** however, each student can only be assigned to a single department. This meant that, for example, any student registered on the "MEng Chemical and Process Engineering" programme was assigned to the "Department of Chemical Engineering". Assigning students to a single department presented a particular problem for those in the Faculty of HaSS. This faculty offers a flexible BA (with Honours) programme which allows students to combine several subjects into a single degree. For this reason, most of the students on these programmes are assigned to a department named the "Faculty of HaSS" in the **School-leavers dataset** rather than a single department. Similarly, all programmes within Business are run centrally, hence students are assigned to a department named the "Faculty of Business". In contrast, all students within the Faculties of Science and Engineering are assigned to one of the named departments in Table C.7.

Table C.8: List of all unique degree programmes offered by Strathclyde Business School (2012/13 - 2021/22) [1 of 2].

Programme Title
Accounting
Accounting and Business Enterprise
Accounting and Business Law
Accounting and Economics
Accounting and Finance
Accounting and Human Resource Management
Accounting and Management
Accounting and Management Science
Accounting and Marketing
Accounting and Mathematics and Statistics
Business
Business Administration
Business Analysis and Technology and Business Enterprise
Business Analysis and Technology and Business Law
Business Analysis and Technology and Economics
Business Analysis and Technology and Finance
Business Analysis and Technology and Hospitality and Tourism Management
Business Analysis and Technology and Human Resource Management
Business Analysis and Technology and Management
Business Analysis and Technology and Marketing
Business Analysis and Technology and Mathematics and Statistics
Business Enterprise and Business Law
Business Enterprise and Business Technology
Business Enterprise and Economics
Business Enterprise and Finance
Business Enterprise and Hospitality and Tourism
Business Enterprise and Hospitality and Tourism Management
Business Enterprise and Human Resource Management
Business Enterprise and Management
Business Enterprise and Management Science
Business Enterprise and Marketing
Business Law and Economics
Business Law and Finance
Business Law and Hospitality and Tourism
Business Law and Human Resource Management
Business Law and Management
Business Law and Marketing
⋮

Table C.9: List of all unique degree programmes offered by Strathclyde Business School (2012/13 - 2021/22) [2 of 2].

Programme Title
⋮
Business - Business Technology and Management Business - Business Technology and Marketing
Business Technology and Management Business Technology and Marketing
Economics Economics and Finance Economics and Hospitality and Tourism Economics and Human Resource Management Economics and Management Economics and Marketing Economics and Mathematics and Statistics Economics and Psychology Economics and Psychology - SBS
Finance and Hospitality and Tourism Management Finance and Human Resource Management Finance and Management Finance and Marketing Finance and Mathematics and Statistics
Hospitality and Tourism Management and Human Resource Management Hospitality and Tourism Management and Management Hospitality and Tourism Management and Marketing Hospitality and Tourism and Management Hospitality and Tourism and Marketing
Human Resource Management and Management Human Resource Management and Marketing Human Resource Management and Psychology Human Resource Management and Psychology - SBS
International Business International Business and Modern Languages International Business with Modern Languages International Business with a Modern Language
Management Science and Marketing Management and Management Science Management and Marketing Marketing and Psychology Marketing and Psychology - SBS
Total: 71 programmes

Table C.10: List of all unique degree programmes offered by Faculty of Humanities and Social Sciences (2012/13 - 2021/22) [1 of 3].

Programme Title
Economics and French
Economics and History
Economics and Journalism and Creative Writing
Economics and Journalism, Media and Communication
Economics and Law
Economics and Politics
Economics and Politics and International Relations
Economics and Psychology
Economics and Social Policy
Economics and Spanish
Education and English
Education and English and Creative Writing
Education and French
Education and History
Education and Human Resource Management
Education and Italian
Education and Journalism and Creative Writing
Education and Journalism, Media and Communication
Education and Law
Education and Politics
Education and Politics and International Relations
Education and Psychology
Education and Social Policy
Education and Spanish
Education and Sport
English and Creative Writing and French
English and Creative Writing and History
English and Creative Writing and Journalism, Media and Communication
English and Creative Writing and Law
English and Creative Writing and Politics and International Relations
English and Creative Writing and Psychology
English and Creative Writing and Social Policy
English and Creative Writing and Spanish
English and French
English and History
English and Human Resource Management
English and Journalism and Creative Writing
English and Journalism, Media and Communication
English and Law
English and Politics
English and Politics and International Relations
English and Psychology
English and Social Policy
English and Spanish
⋮

Table C.11: List of all unique degree programmes offered by Faculty of Humanities and Social Sciences (2012/13 - 2021/22) [2 of 3].

Programme Title
⋮
French and History
French and Hospitality and Tourism
French and Hospitality and Tourism Management
French and Human Resource Management
French and Italian
French and Journalism and Creative Writing
French and Journalism, Media and Communication
French and Law
French and Marketing
French and Politics
French and Politics and International Relations
French and Psychology
French and Social Policy
French and Spanish
History and Human Resource Management
History and Italian
History and Journalism and Creative Writing
History and Journalism, Media and Communication
History and Law
History and Politics
History and Politics and International Relations
History and Psychology
History and Social Policy
History and Spanish
Hospitality and Tourism Management and Spanish
Hospitality and Tourism and Italian
Hospitality and Tourism and Spanish
Human Resource Management and Journalism and Creative Writing
Human Resource Management and Law
Human Resource Management and Politics
Human Resource Management and Politics and International Relations
Human Resource Management and Psychology
Human Resource Management and Social Policy
Human Resource Management and Spanish
Humanities and Social Sciences
Italian and Journalism and Creative Writing
Italian and Marketing
Italian and Politics and International Relations
Italian and Psychology
Italian and Social Policy
Italian and Spanish
⋮

Table C.12: List of all unique degree programmes offered by Faculty of Humanities and Social Sciences (2012/13 - 2021/22) [3 of 3].

Programme Title
⋮
Journalism and Creative Writing and Law
Journalism and Creative Writing and Politics
Journalism and Creative Writing and Politics and International Relations
Journalism and Creative Writing and Social Policy
Journalism and Creative Writing and Spanish
Journalism, Media and Communication and Law
Journalism, Media and Communication and Politics and International Relations
Journalism, Media and Communication and Psychology
Journalism, Media and Communication and Social Policy
Journalism, Media and Communication and Spanish
Law
Law (Clinical)
Law and Politics
Law and Politics and International Relations
Law and Psychology
Law and Social Policy
Law and Spanish
Law with a Modern Language
Law with a Modern Language - French and Law
Law with a Modern Language - Italian and Law
Law with a Modern Language - Law and Spanish
Marketing and Spanish
Mathematics and Psychology
Philosophy, Politics and Economics
Politics and International Relations and Psychology
Politics and International Relations and Social Policy
Politics and International Relations and Spanish
Politics and Psychology
Politics and Social Policy
Politics and Spanish
Primary Education
Psychology
Psychology and Counselling
Psychology and Social Policy
Psychology and Spanish
Psychology and Sport
Scots and English Law
Scots and English Law (Clinical)
Social Policy and Spanish
Social Work
Speech and Language Pathology
Sport and Physical Activity
Total: 127 programmes

Table C.13: List of all unique degree programmes offered by Faculty of Engineering (2012/13 - 2021/22).

Department	**Programme Title**
Architecture	Architectural Studies
Biomedical Engineering	Biomedical Engineering Prosthetics and Orthotics
Chemical and Process Engineering	Chemical Engineering
Civil and Environmental Engineering	Civil Engineering Civil and Environmental Engineering Structural and Architectural Engineering
Design, Manufacturing and Engineering Management	Manufacturing Engineering with Management Product Design Engineering Product Design and Innovation Production Engineering and Management Sports Design Engineering Sports Engineering
Electronic and Electrical Engineering	Computer and Electronic Systems Computer and Electronic Systems with International Study Electrical Energy Systems Electrical and Mechanical Engineering Electrical and Mechanical Engineering with International Study Electronic and Digital Systems Electronic and Electrical Engineering Electronic and Electrical Engineering with Business Studies Electronic and Electrical Engineering with International Study
Mechanical and Aerospace Engineering	Aero-Mechanical Engineering Mechanical Engineering Mechanical Engineering with Aeronautics Mechanical Engineering with Financial Management Mechanical Engineering with International Study Mechanical Engineering with Materials Engineering
Naval Architecture, Ocean and Marine Engineering	Naval Architecture and Marine Engineering Naval Architecture with High Performance Marine Vehicles Naval Architecture with Ocean Engineering Naval Architecture with Small Craft Engineering
None - Faculty of Engineering	Engineering Academy
Total	**33 programmes**

Table C.14: List of all unique degree programmes offered by Faculty of Science (2012/13 - 2021/22) [1 of 2].

Department	**Programme Title**
Computer and Information Sciences	Business Information Systems Computer Science Computer Science with Law Software Engineering Computer and Electronic Systems[1] Computer and Electronic Systems with International Study[1]
Mathematics and Statistics	Data Analytics Mathematics Mathematics and Computer Science Mathematics and Physics Mathematics with Teaching Mathematics, Statistics and Accounting Mathematics, Statistics and Business Analysis Mathematics, Statistics and Economics Mathematics, Statistics and Finance Mathematics, Statistics and Management Science
Physics	Physics Physics with Advanced Research Physics with Teaching
Pure and Applied Chemistry	Applied Chemistry and Chemical Engineering[2] Chemistry Chemistry with Drug Discovery Chemistry with Forensic Chemistry Chemistry with Teaching Forensic and Analytical Chemistry
⋮	

[1] Delivered jointly with the Department of Electronic and Electrical Engineering.

[2] Delivered jointly with the Department of Chemical and Process Engineering.

Table C.15: List of all unique degree programmes offered by Faculty of Science (2012/13 - 2021/22) [2 of 2].

Department	**Programme Title**
Strathclyde Institute of Pharmacy and Biomedical Sciences	Biochemistry Biochemistry and Immunology Biochemistry and Microbiology Biochemistry and Pharmacology Biomedical Science Biomolecular Sciences Forensic Biology Immunology Immunology and Microbiology Immunology and Pharmacology Microbiology Microbiology and Pharmacology Pharmaceutical Sciences Pharmacology Pharmacy[3]
None - Faculty of Science	Natural Sciences[4]
Total	**41 programmes**

[3] A 4 stage Integrated Masters programme (as opposed to the traditional 5 stage). Students are expected to enter directly into 2nd stage of the programme.

[4] Widening Access programme.

C.6 Changes in Registration Status

Students may request to change degree programme at any time. If approved, the **School-leavers dataset** will indicate the relevant instance in which the student began the new programme. For example, if in their first academic session, a student switches from "BSc Electrical & Mechanical Engineering" in semester 1 to "BSc Mathematics & Statistics" in semester 2, then the **School-leavers dataset** will show that student was registered with the *Programme Title* "Mathematics & Statistics" in their first academic session. In this respect, the **School-leavers dataset** likely under-estimates the true number of students who change degree programme.

However, it is also known that some degree programmes are functionally identical to one another in the first few stages, before diverging into distinct programmes. For example, "BSc Mechanical Engineering" and "MSc Mechanical Engineering" are distinct programmes in the **School-leavers dataset**, but are functionally identical from stages 1 to 4 until the 5th stage (the Integrated Masters stage). Thus, students may be recorded as having changed programme despite simply continuing their studies. Additionally, changes in programmes may have practically occurred much earlier than they were recorded in the **School-leavers dataset**. For example, "BSc Mathematics & Statistics" and "BSc Mathematics" are functionally identical from stages 1 to 3, but the registration status in the CSR likely only changes in the 4th stage if the student did/did not pick a sufficient number of credits in Statistics modules from stages 3 to 4. Finally, students from the faculties of Business and HaSS likely over-inflate the true number of changes that occurred. This is because of the large number of combinations for joint honours degrees in each that depend on the student successfully attaining a certain number of credits throughout their degree. Taken together, the *Changed Programme* variable likely over-estimates the true number of changes.

Around 33% of students were recorded in the **School-leavers dataset** as having changed degree programme at some point during their registration with the University (Table C.16). Similarly, around 20% of students were recorded as changing department, while only 2% changed faculty. The difference in the proportions seen here between faculties/departments/programmes can be explained by the way that the data is recorded in the **School-leavers dataset**. Thus, *Changed Department* and *Changed Faculty* may prove to be more useful proxy indicators than *Changed Programme*, for measuring whether or not a student truly changed discipline.

Table C.16: Summary of variables examined in the school-leavers dataset (2 of 2).

Variables	Levels	Count	Proportion
Break	No	17951	0.94
	Yes	1037	0.06
Changed Department	No	15277	0.81
	Yes	3711	0.20
Changed Faculty	No	18654	0.98
	Yes	334	0.02
Changed Programme	No	12721	0.67
	Yes	6267	0.33
Repeated Stage	No	18396	0.97
	Yes	592	0.03

C.6.1 Taking a Break from Studies

Students may take a break in their studies through suspension of their degree programme, returning at any time to complete their studies. There is no official limit to the length of time a student may suspend their studies. Suspension may be requested for a variety of reasons not excluded to: financial, health, or personal reasons, or even to take up work experience/placement. In the **School-leavers dataset** 1037 (5.5%) students took a break at some point during their studies. The drop-out rate was slightly smaller within the group of students who took a break (8.4%) compared to the group of students who did not take a break (8.8%). Ignoring the length of time it takes to complete a Bachelor's with Honours degree, around 59.9% of students who took a break would go on to complete their degree versus 81.7% of students who did not take a break that completed their degree. This difference will be influenced by students who return from a break that are still actively studying towards completing their degree.

C.6.2 Repeating a Stage

Students who do not attain the necessary credits to progress to the next stage of a degree programme may have to repeat a stage. Typically, students may only repeat one stage of a degree programme before taking compulsory withdrawal. In the **School-leavers dataset** 592 (3.1%) students repeated at least one stage of their degree programme. The drop-out rate was substantially larger within the group of students who repeated a stage (20.1%) compared to the group of students who did not repeat a stage (8.45%). Ignoring the length of time it takes to complete a Bachelor's with Honours degree, around 31.4% of students who repeated a stage would go on to complete their degree within four years versus 82.3% of students who did not repeat a stage that completed their degree. This suggests that those who repeat a stage of degree programme are may be more likely to have an unsuccessful outcome at university, although it should be noted that some students will still be actively studying towards completing their degree.

Table C.17: Cross tabulation of explanatory variables versus whether or not a student repeated a stage of their degree programme (1 of 2). Proportions rounded to 2 decimal places.

Variables	Levels	Repeated Stage		Did Not Repeat	
		Count	**Prop.**	**Count**	**Prop.**
Academic Cohort	2012/13	85	0.05	1742	0.95
	2013/14	79	0.04	1787	0.96
	2014/15	81	0.04	1872	0.96
	2015/16	76	0.04	1909	0.96
	2016/17	59	0.03	1913	0.97
	2017/18	67	0.04	1826	0.96
	2018/19	53	0.03	1770	0.97
	2019/20	46	0.03	1802	0.97
	2020/21	33	0.02	1780	0.98
	2021/22	13	0.01	1995	0.99
Age at Entry	17 or under	296	0.03	9101	0.97
	18	296	0.03	9295	0.97
Disability Status	Disabled	48	0.04	1038	0.96
	None	544	0.03	17358	0.97
Ethnicity	Ethnic-minority	59	0.04	1327	0.96
	White	533	0.03	17069	0.97
Faculty	Business	35	0.01	3276	0.99
	Engineering	154	0.03	5182	0.97
	HaSS	149	0.03	5746	0.97
	Science	254	0.06	4192	0.94
Local to Glasgow	Glasgow-based	316	0.03	8888	0.97
	Outside Glasgow	276	0.03	9508	0.97
SIMD Quintile	1	97	0.04	2097	0.96
	2	132	0.05	2665	0.95
	3	83	0.03	2955	0.97
	4	116	0.03	3957	0.97
	5	164	0.02	6722	0.98
Sex	Female	230	0.02	9240	0.98
	Male	362	0.04	9156	0.96
Urban/Rural Status	Accessible Rural	43	0.02	1853	0.98
	Remote Rural	14	0.02	561	0.98
	Urban	374	0.03	11288	0.97
	Unknown	161	0.03	4694	0.97
Overall	**-**	**592**	**0.03**	**18396**	**0.97**

Table C.18: Cross tabulation of explanatory variables versus whether or not a student repeated a stage of their degree programme (2 of 2). Proportions rounded to 2 decimal places.

Variables	**Levels**	**Repeated Stage**		**Did Not Repeat**	
		Count	**Prop.**	**Count**	**Prop.**
Break	No	529	0.03	17422	0.97
	Yes	63	0.06	974	0.94
Changed Department	No	545	0.04	14732	0.96
	Yes	47	0.01	3664	0.99
Changed Faculty	No	581	0.03	18073	0.97
	Yes	11	0.03	323	0.97
Changed Programme	No	454	0.04	12267	0.96
	Yes	138	0.02	6129	0.98
Offer Received	Con. Offer	116	0.06	1937	0.94
	Std. Offer	226	0.02	10725	0.98
	Unknown	250	0.04	5734	0.96
Prior Att. Quintile	1	232	0.06	3569	0.94
	2	155	0.04	3636	0.96
	3	101	0.03	3700	0.97
	4	72	0.02	3710	0.98
	5	32	0.01	3781	0.99
Overall	**-**	**592**	**0.03**	**18396**	**0.97**

C.7 Advanced Higher Recommendations

The summaries in this section refer to the **Advanced Higher subset** (see Section 3.9) and not the whole **School-leavers dataset**. Programmes in the Faculties of Science and Engineering recommended a range of Advanced Highers in various Science subjects (Table C.19). The only departments which did not recommend Advanced Highers for any of their programmes were the three Faculty of Engineering departments: "Architecture", "Civil & Environmental Engineering", and "Design, Manufacturing & Engineering Management". The remaining departments recommended an Advanced Higher for at least one of their programmes. The most popular Advanced Higher to recommend was Advanced Higher Mathematics. For example, departments in the Faculty of Engineering tended to recommend Advanced Highers in at least Mathematics and Physics, while departments in the Faculty of Science tended to recommend Advanced Higher Mathematics and at least one other relevant science subject. The Faculty of Engineering departments were more explicit in their recommendation of Advanced Highers for entry into stage 1 of their programmes. In contrast, the Faculty of Science departments tended to encourage Advanced Highers and only explicitly recommended them for entry into stage 2 of their programmes.

The language surrounding the recommendation of Advanced Highers varied from programme-to-programme and over time. An illustrative example are the "MEng Chemical and Process Engineering" and "BSc Honours Mathematics" programmes in the 2014/15 and 2021/22 prospectuses. In 2014/15, the "MEng Chemical and Process Engineering" explicitly recommended Advanced Higher Mathematics for stage 1 entrants while "BSc Honours Mathematics" did not explicitly recommend it, though did encourage it. In contrast, the 2021/22 prospectus showed the inverse: "MEng Chemical and Process Engineering" did not explicitly recommend Advanced Higher Mathematics while "BSc Honours Mathematics" did. When constructing binary indicators (see Section 4.6), it was decided that any programme which had ever recommended or encouraged an Advanced Higher would be recorded as having "recommended" it.

Some departments did not consistently recommend Advanced Highers for all of their programmes. For example, the department of "Biomedical Engineering" recommended Advanced Highers in Mathematics and Physics for students applying to its "MEng Biomedical Engineering" programme, but not its BEng Honours equivalent. It also recommended Advanced Highers in Mathematics, Physics and Biology for its "BSc Prosthetics & Orthotics" programme. From 2013/14 - 2015/16, the department of "Chemical & Process Engineering" recommended Advanced Highers in Mathematics, Physics and Chemistry; it did not mention any recommendations for Advanced Highers from 2016/17 onwards.

The "MPharm Pharmacy" and "MPhys Physics with Advanced Research" programmes recommended Advanced Highers, though students from these programmes were removed from the **School-leavers dataset** and its relevant subsets (see Section 3.7 for more details).

Table C.19: Summary of the Advanced Higher subjects that University of Strathclyde STEM departments have recommended for some or all of their degree programmes (according to prospectus data from 2013/14 - 2023/24).

Faculty	**Department**	**Count of Programmes**	**Advanced Higher Subjects Recommended?**
Engineering	Architecture	1	None
	Biomedical Engineering[1]	3	Mathematics, Physics, Biology[2]
	Chemical & Process Engineering[3]	2	Mathematics, Physics, Chemistry
	Civil & Environmental Engineering	4	None
	Design, Manufacturing & Engineering Management	10	None
	Electronic & Electrical Engineering	12	Mathematics, Physics
	Mechanical & Aerospace Engineering	9	Mathematics, Physics
	Naval Architecture, Ocean & Marine Engineering	7	Mathematics, Physics
Science	Strathclyde Institute of Pharmacy & Biomedical Sciences[4]	14	Mathematics, Physics, Biology, Chemistry
	Computer & Information Sciences	5	Mathematics, Computing Science
	Mathematics & Statistics	12	Mathematics
	Physics	4	Mathematics, Physics
	Pure and Applied Chemistry	9	Mathematics, Chemistry

[1] Does not recommend Advanced Highers for its BEng Honours Biomedical Engineering programme.

[2] Only recommends Advanced Higher Biology for its Prosthetics & Orthotics programme.

[3] Did not explicitly recommend any Advanced Highers from 2016/17 onwards.

[4] The MPharm Pharmacy programme expects most students to enter directly into stage 2. It is expected that students have Advanced Highers in Biology and Chemistry; if not possible then may accept Mathematics and Physics as alternatives.

Appendix D

Exploratory Visualisation

This appendix contains some explorations considered too large for inclusion in the main body of in Chapter 5, as well as some additional explorations.

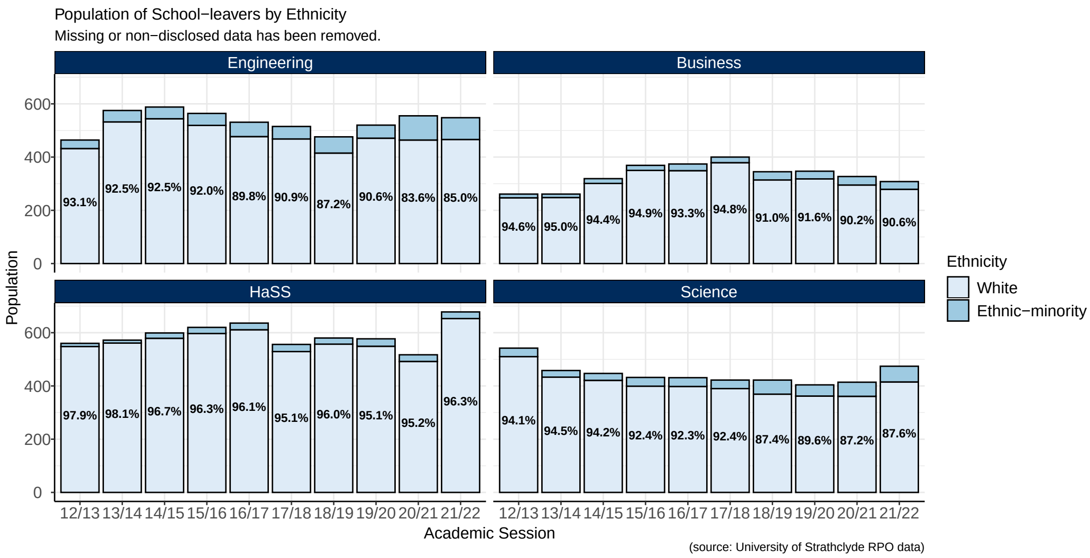


Figure D.1: Count/proportion of White/Ethnic-minority school-leavers who entered each faculty from 2012/13 to 2021/22.

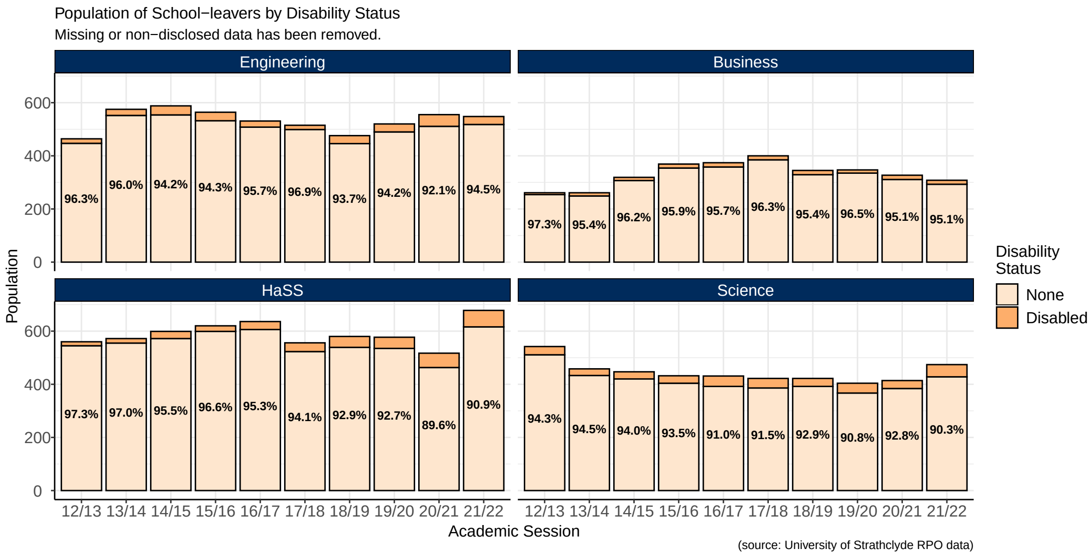


Figure D.2: Count/proportion of school-leavers with disabilities/no disclosed disabilities who entered each faculty from 2012/13 to 2021/22.

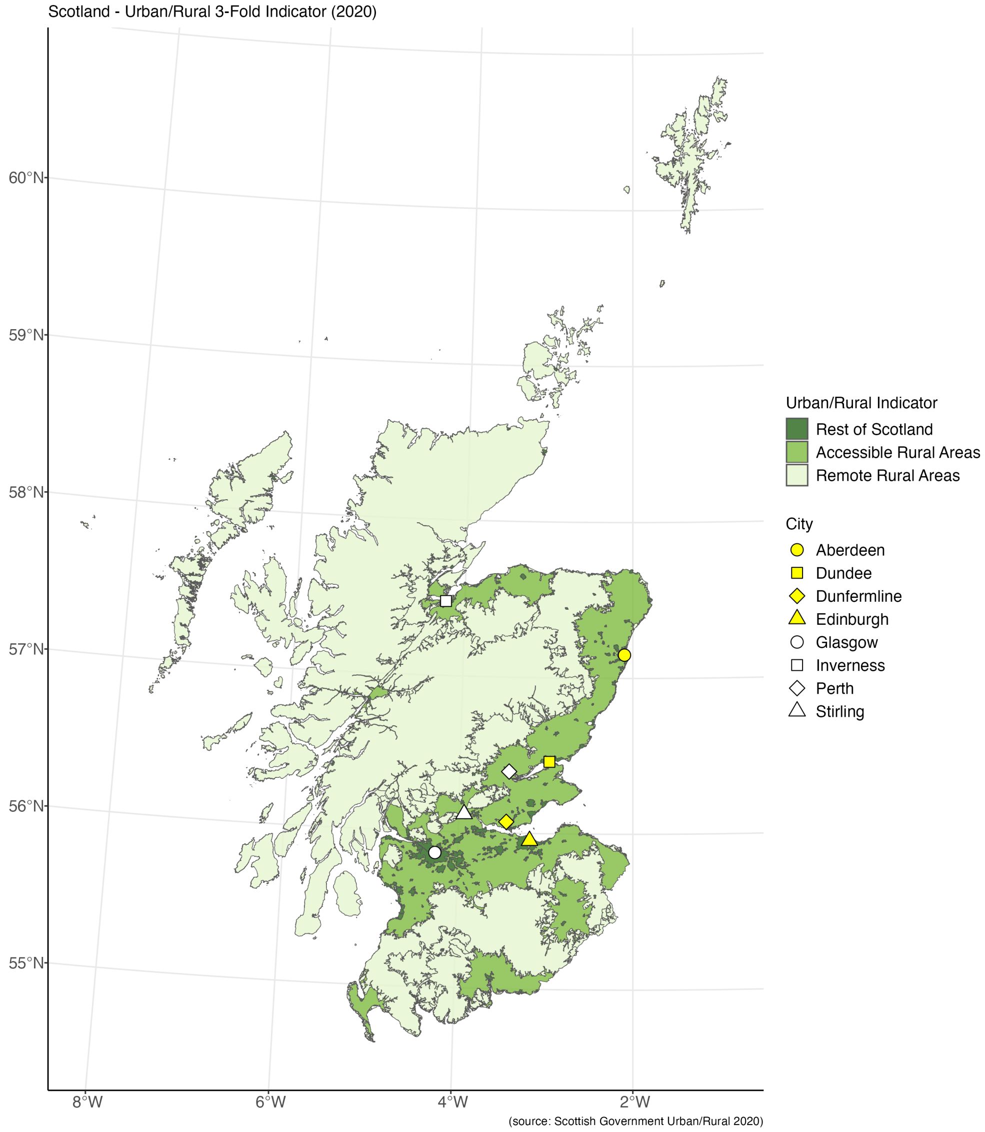


Figure D.3: Urban/Rural 3-Fold Indicator (2020) across Scotland. Scotland's 8 major cities are indicated with dots. Plot created using "shapefiles" provided by the Scottish Government [2].

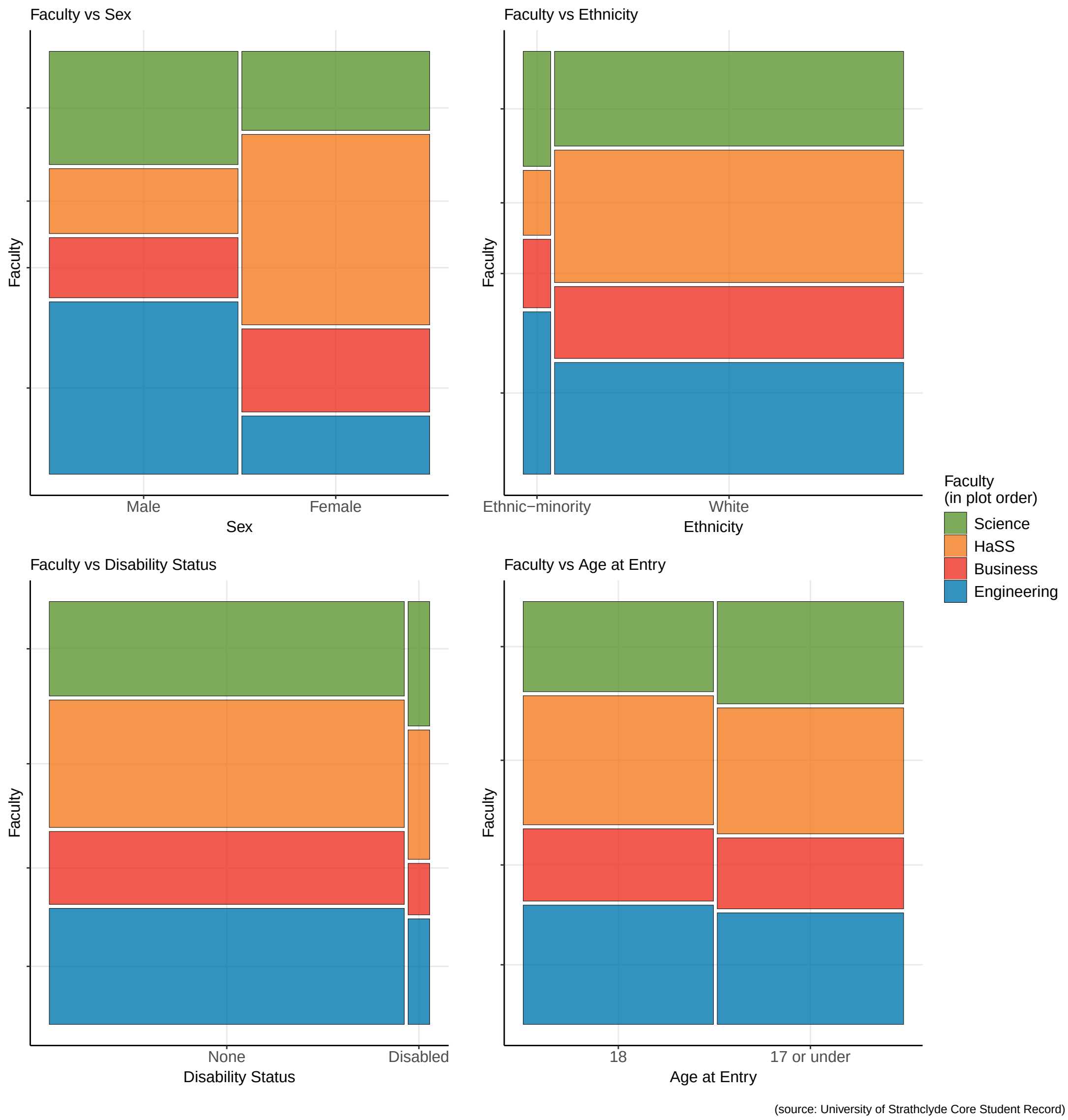


Figure D.4: Mosaic plots showing proportion of school-leavers within each strata: Faculty versus explanatory variables (1 of 2).

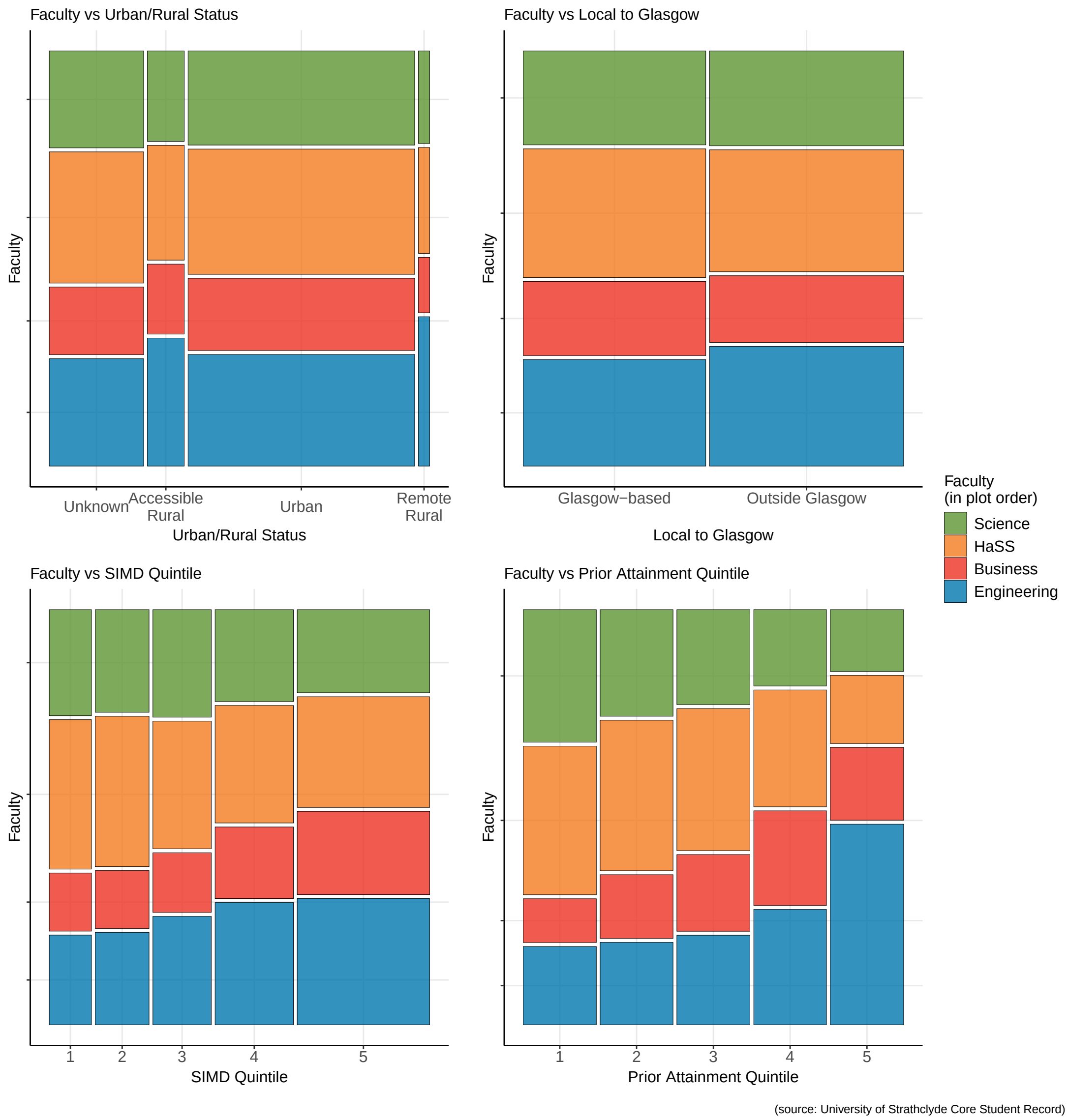


Figure D.5: Mosaic plots showing proportion of school-leavers within each strata: Faculty versus explanatory variables (2 of 2).

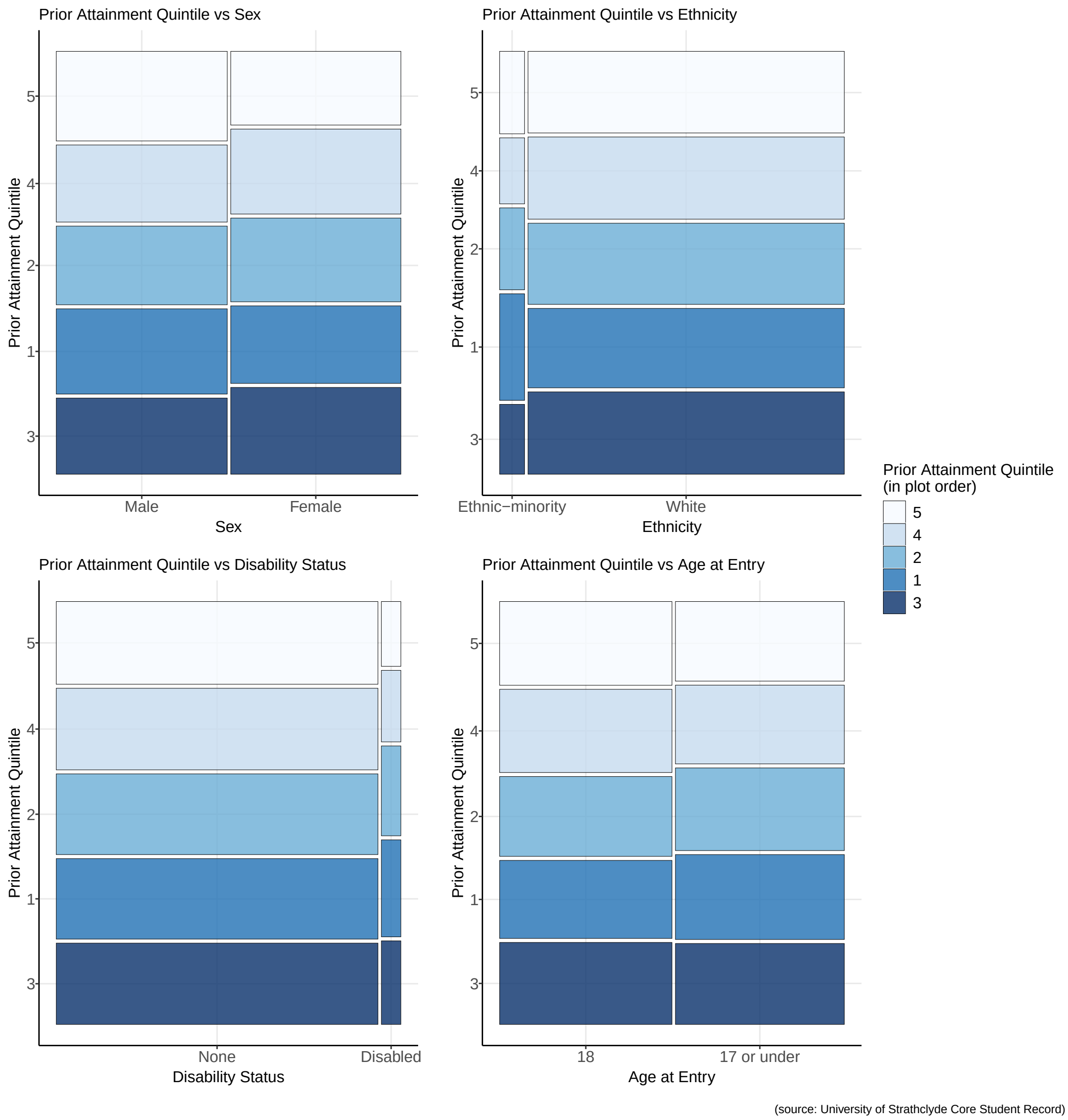
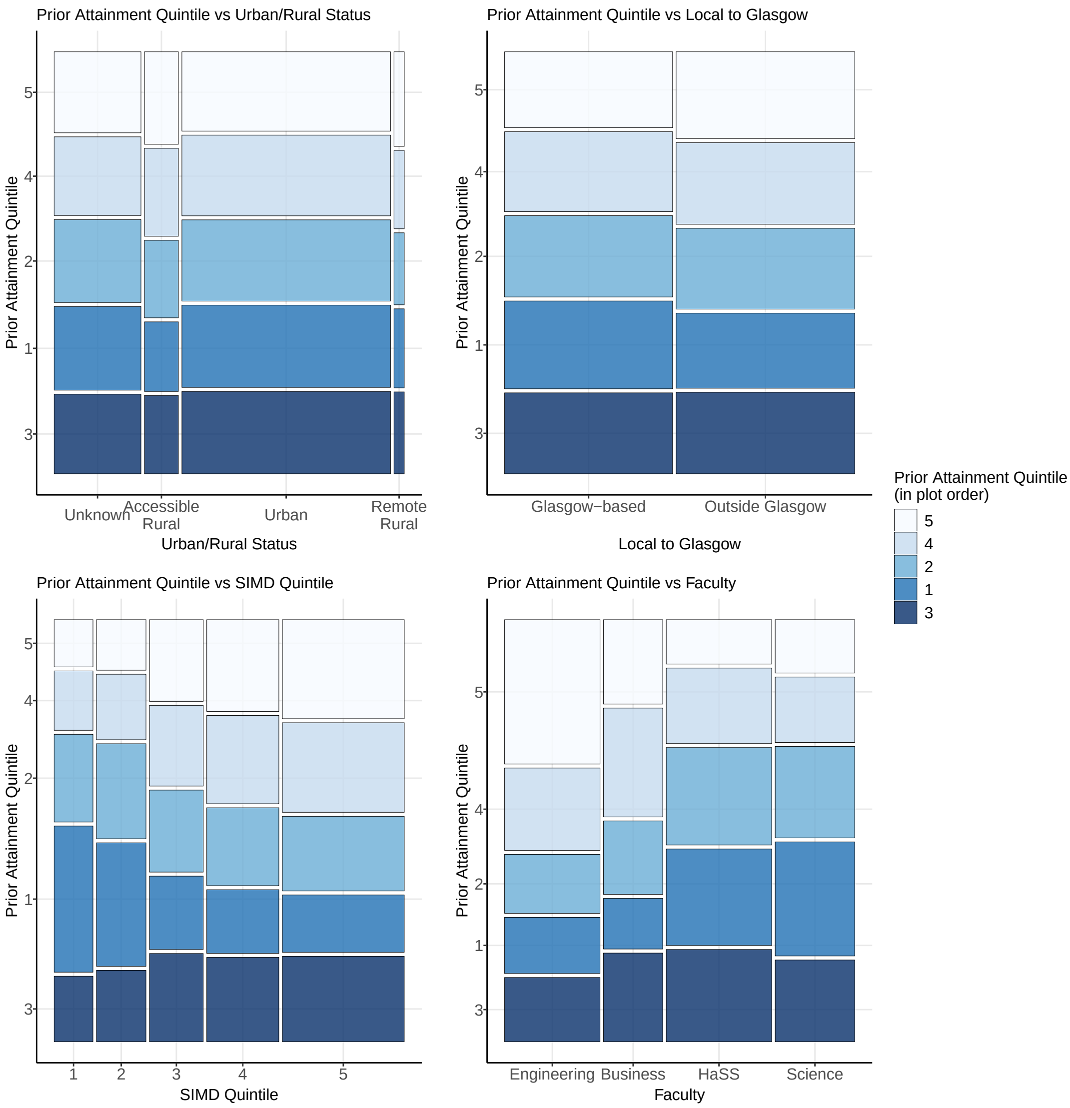


Figure D.6: Mosaic plots showing proportion of school-leavers within each strata: Prior Attainment Quintile versus explanatory variables (1 of 2).



(source: University of Strathclyde Core Student Record)

Figure D.7: Mosaic plots showing proportion of school-leavers within each strata: Prior Attainment Quintile versus explanatory variables (2 of 2).

Table D.1: Cross tabulation of explanatory variables versus retention outcome (1 of 2). Proportions rounded to 2 decimal places.

Variables	**Levels**	**Retained**		**Not Retained**	
		Count	**Prop.**	**Count**	**Prop.**
Academic Cohort	2012/13	1635	0.90	192	0.10
	2013/14	1703	0.91	163	0.09
	2014/15	1744	0.89	209	0.11
	2015/16	1795	0.90	190	0.10
	2016/17	1765	0.90	207	0.10
	2017/18	1705	0.90	188	0.10
	2018/19	1595	0.88	228	0.12
	2019/20	1693	0.92	155	0.08
	2020/21	1641	0.91	172	0.10
	2021/22	1720	0.86	288	0.14
Age at Entry	17 or under	8410	0.90	987	0.10
	18	8586	0.90	1005	0.10
Disability Status	Disabled	954	0.88	132	0.12
	None	16042	0.90	1860	0.10
Ethnicity	Ethnic-minority	1253	0.90	133	0.10
	White	15743	0.89	1859	0.11
Faculty	Business	3029	0.92	282	0.09
	Engineering	4850	0.91	486	0.09
	HaSS	5233	0.89	662	0.11
	Science	3884	0.87	562	0.13
Local to Glasgow	Glasgow-based	8174	0.89	1030	0.11
	Outside Glasgow	8822	0.90	962	0.10
SIMD Quintile	1	1867	0.85	327	0.15
	2	2418	0.86	379	0.14
	3	2713	0.89	325	0.11
	4	3686	0.91	387	0.10
	5	6312	0.92	574	0.08
Sex	Female	8530	0.90	940	0.10
	Male	8466	0.89	1052	0.11
Urban/Rural Status	Accessible Rural	1698	0.90	198	0.10
	Remote Rural	534	0.93	41	0.07
	Urban	10436	0.90	1226	0.10
	Unknown	4328	0.89	527	0.11
Overall	**-**	**16996**	**0.90**	**1992**	**0.10**

Table D.2: Cross tabulation of explanatory variables versus retention outcome (2 of 2). Proportions rounded to 2 decimal places.

Variables	**Levels**	**Retained**		**Not Retained**	
		Count	**Prop.**	**Count**	**Prop.**
Break	No	16198	0.90	1753	0.10
	Yes	798	0.77	239	0.23
Changed Department	No	13597	0.89	1680	0.11
	Yes	3399	0.92	312	0.08
Changed Faculty	No	16840	0.90	1814	0.10
	Yes	156	0.47	178	0.53
Changed Programme	No	11105	0.87	1616	0.13
	Yes	5891	0.94	376	0.06
Repeated Stage	No	16551	0.90	1845	0.10
	Yes	445	0.75	147	0.25
Offer Received	Con. Offer	1706	0.83	347	0.17
	Std. Offer	9912	0.91	1039	0.10
	Unknown	5378	0.90	606	0.10
Prior Att. Quintile	1	3061	0.81	740	0.20
	2	3331	0.88	460	0.12
	3	3449	0.91	352	0.09
	4	3506	0.93	276	0.07
	5	3649	0.96	164	0.04
Overall	**-**	**16996**	**0.90**	**1992**	**0.10**

Table D.3: Cross tabulation of explanatory variables versus completion outcome (1 of 2). Proportions rounded to 2 decimal places.

Variables	**Levels**	**Completed**		**Not Completed**	
		Count	**Prop.**	**Count**	**Prop.**
Academic Cohort	2012/13	1304	0.71	523	0.29
	2013/14	1342	0.72	524	0.28
	2014/15	1442	0.74	511	0.26
	2015/16	1473	0.74	512	0.26
	2016/17	1485	0.75	487	0.25
	2017/18	1446	0.76	447	0.24
	2018/19	1370	0.75	453	0.25
Age at Entry	17 or under	4894	0.74	1706	0.26
	18	4968	0.74	1751	0.26
Disability Status	Disabled	449	0.67	219	0.33
	None	9413	0.74	3238	0.26
Ethnicity	Ethnic-minority	626	0.74	216	0.26
	White	9236	0.74	3241	0.26
Faculty	Business	1924	0.83	405	0.17
	Engineering	2849	0.77	864	0.23
	HaSS	3047	0.74	1076	0.26
	Science	2042	0.65	1112	0.35
Local to Glasgow	Glasgow-based	4700	0.72	1803	0.28
	Outside Glasgow	5162	0.76	1654	0.24
SIMD Quintile	1	874	0.65	466	0.35
	2	1291	0.69	567	0.30
	3	1614	0.72	611	0.28
	4	2176	0.75	735	0.25
	5	3907	0.78	1078	0.22
Sex	Female	5035	0.77	1516	0.23
	Male	4827	0.71	1941	0.29
Urban/Rural Status	Accessible Rural	987	0.75	330	0.25
	Remote Rural	314	0.76	98	0.24
	Urban	6106	0.74	2161	0.26
	Unknown	2455	0.74	868	0.26
Overall	**-**	**9862**	**0.74**	**3457**	**0.26**

Table D.4: Cross tabulation of explanatory variables versus completion outcome (2 of 2). Proportions rounded to 2 decimal places.

Variables	**Levels**	**Completed**		**Not Completed**	
		Count	**Prop.**	**Count**	**Prop.**
Break	No	9792	0.78	2702	0.22
	Yes	70	0.09	755	0.92
Changed Department	No	6858	0.71	2836	0.29
	Yes	3004	0.83	621	0.17
Changed Faculty	No	9802	0.75	3228	0.25
	Yes	60	0.21	229	0.79
Changed Programme	No	5418	0.70	2371	0.30
	Yes	4444	0.80	1086	0.20
Repeated Stage	No	9855	0.77	2964	0.23
	Yes	7	0.01	493	0.99
Offer Received	Con. Offer	653	0.62	396	0.38
	Std. Offer	5022	0.77	1463	0.23
	Unknown	4187	0.72	1598	0.28
Prior Att. Quintile	1	1533	0.57	1133	0.42
	2	1835	0.69	826	0.31
	3	2009	0.75	655	0.25
	4	2143	0.81	508	0.19
	5	2342	0.88	335	0.12
Overall	**-**	**9862**	**0.74**	**3457**	**0.26**

Table D.5: Cross tabulation of explanatory variables versus dropout outcome (1 of 2). Proportions rounded to 2 decimal places.

Variables	**Levels**	**Dropped Out**		**Censored**	
		Count	**Prop.**	**Count**	**Prop.**
Academic Cohort	2012/13	158	0.09	1669	0.91
	2013/14	132	0.07	1734	0.93
	2014/15	166	0.09	1787	0.92
	2015/16	155	0.08	1830	0.92
	2016/17	177	0.09	1795	0.91
	2017/18	174	0.09	1719	0.91
	2018/19	203	0.11	1620	0.89
Age at Entry	17 or under	559	0.09	6041	0.92
	18	606	0.09	6113	0.91
Disability Status	Disabled	61	0.09	607	0.91
	None	1104	0.09	11547	0.91
Ethnicity	Ethnic-minority	59	0.07	783	0.93
	White	1106	0.09	11371	0.91
Faculty	Business	134	0.06	2195	0.94
	Engineering	295	0.08	3418	0.92
	HaSS	416	0.10	3707	0.90
	Science	320	0.10	2834	0.90
Local to Glasgow	Glasgow-based	593	0.09	5910	0.91
	Outside Glasgow	572	0.08	6244	0.92
SIMD Quintile	1	188	0.14	1152	0.86
	2	199	0.11	1659	0.89
	3	208	0.09	2017	0.91
	4	224	0.08	2687	0.92
	5	346	0.07	4639	0.93
Sex	Female	525	0.08	6026	0.92
	Male	640	0.10	6128	0.91
Urban/Rural Status	Accessible Rural	116	0.09	1201	0.91
	Remote Rural	28	0.07	384	0.93
	Urban	724	0.09	7543	0.91
	Unknown	297	0.09	3026	0.91
Overall	**-**	**12154**	**0.09**	**1165**	**0.91**

Table D.6: Cross tabulation of explanatory variables versus dropout outcome (2 of 2). Proportions rounded to 2 decimal places.

Variables	Levels	Dropped Out		Censored	
		Count	Prop.	Count	Prop.
Break	No	1102	0.09	11392	0.91
	Yes	63	0.08	762	0.92
Changed Department	No	1115	0.12	8579	0.89
	Yes	50	0.01	3575	0.99
Changed Faculty	No	1136	0.09	11894	0.91
	Yes	29	0.10	260	0.90
Changed Programme	1	1061	0.14	6728	0.86
	Yes	104	0.02	5426	0.98
Repeated Stage	No	1071	0.08	11748	0.92
	Yes	94	0.19	406	0.81
Offer Received	Con. Offer	157	0.15	892	0.85
	Std. Offer	538	0.08	5947	0.92
	Unknown	470	0.08	5315	0.92
Prior Att. Quintile	1	452	0.17	2214	0.83
	2	272	0.10	2389	0.90
	3	193	0.07	2471	0.93
	4	163	0.06	2488	0.94
	5	85	0.03	2592	0.97
Overall	**-**	**12154**	**0.09**	**1165**	**0.91**

Appendix E

Survival Analysis

E.1 Exploratory Analysis

Unlike for other faculties, those from Engineering were more likely to stay registered for at least 5 academic sessions rather than 4 (Figure E.1). This is most likely due to the number of Integrated Masters degrees being higher within Engineering than the other faculties. This could also explain why more Males were registered for at least 5 academic sessions than Females (Figure E.2) given that Sex was associated with Faculty (see Figures 5.7 and D.4). Of those who were registered for at least 6 academic sessions or longer, 73.1% were school-leavers who had taken academic suspension or had repeated a year at some point on their journey (results not shown due to small sample sizes). The number of direct-entry-to-second-stage (direct to "2nd year") students were small, however only 2.0% of direct-entry students dropped out compared to 9.4% of standard first stage entry ("1st year") students (results not shown due to small sample sizes).

The *cloglog* plots (Figure E.4) were near-identical to the *logit* plots (Figure 9.4) in Section 9.6 because the hazard of dropout in any given academic session was low ($< 10\%$).

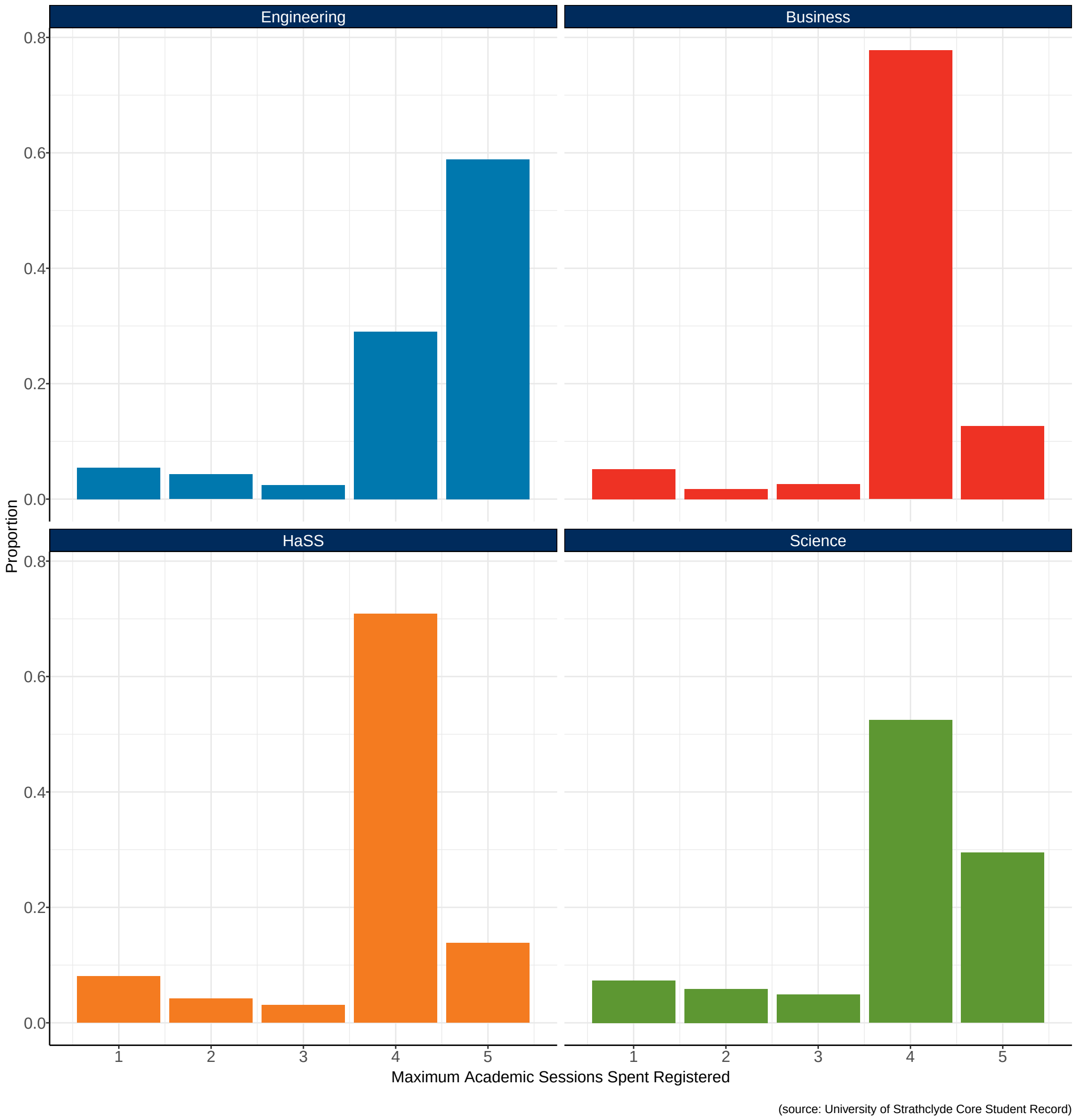


Figure E.1: Proportion of students by the maximum number of sessions they spent registered at the University. Grouped by Faculty.

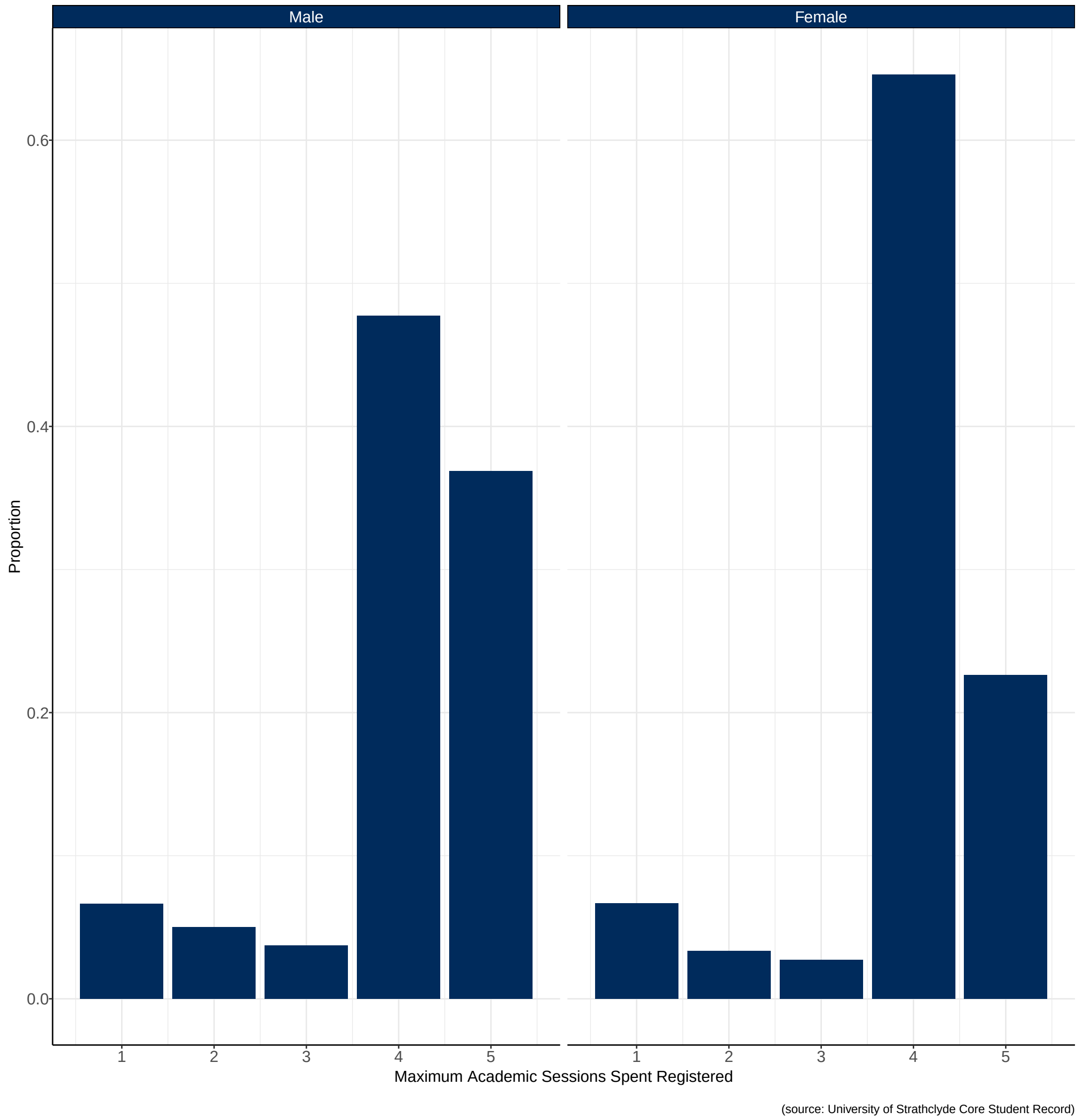


Figure E.2: Proportion of students by the maximum number of sessions they spent registered at the University. Grouped by Sex.

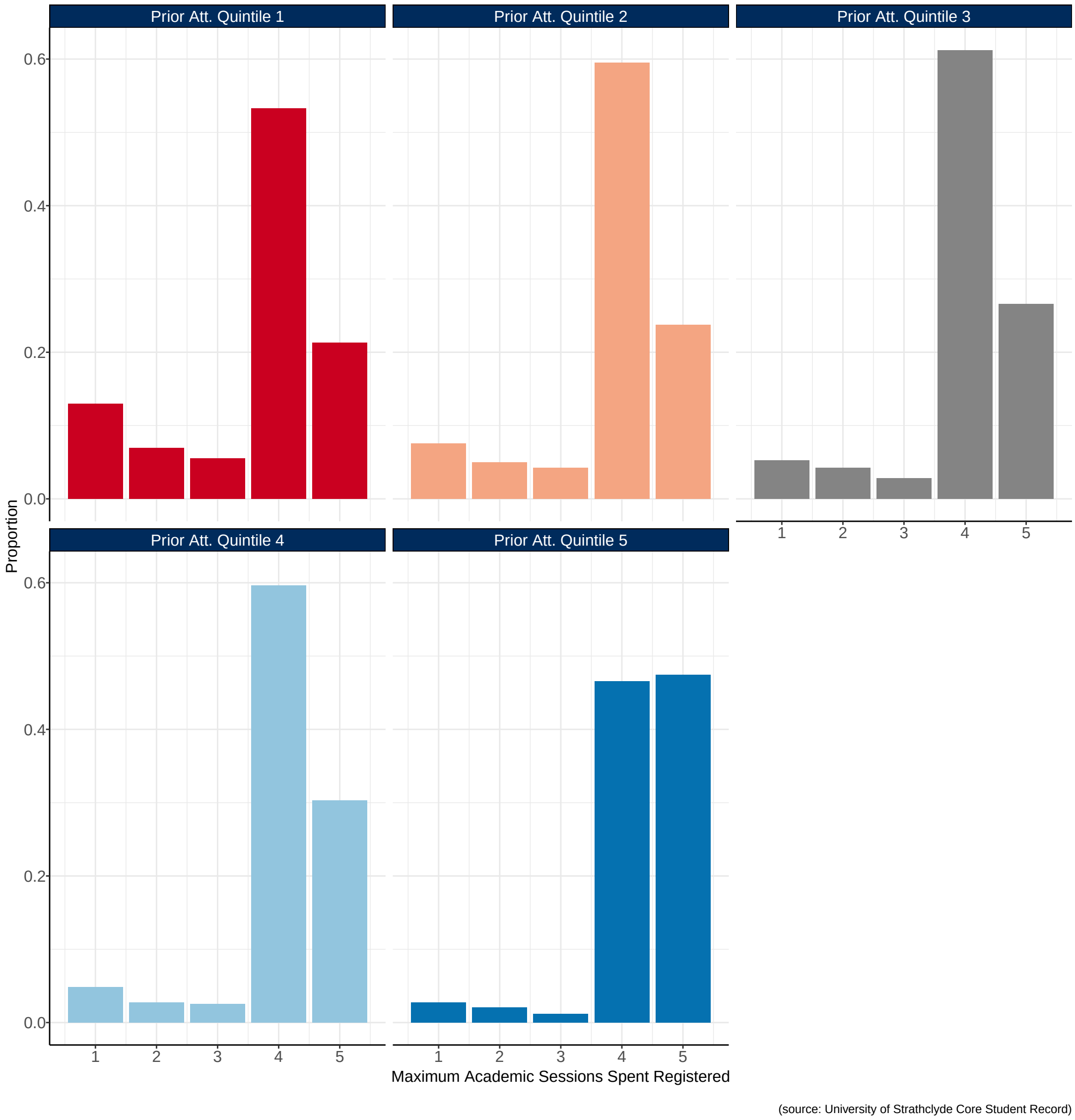


Figure E.3: Proportion of students by the maximum number of sessions they spent registered at the University. Grouped by Prior Attainment Quintile.

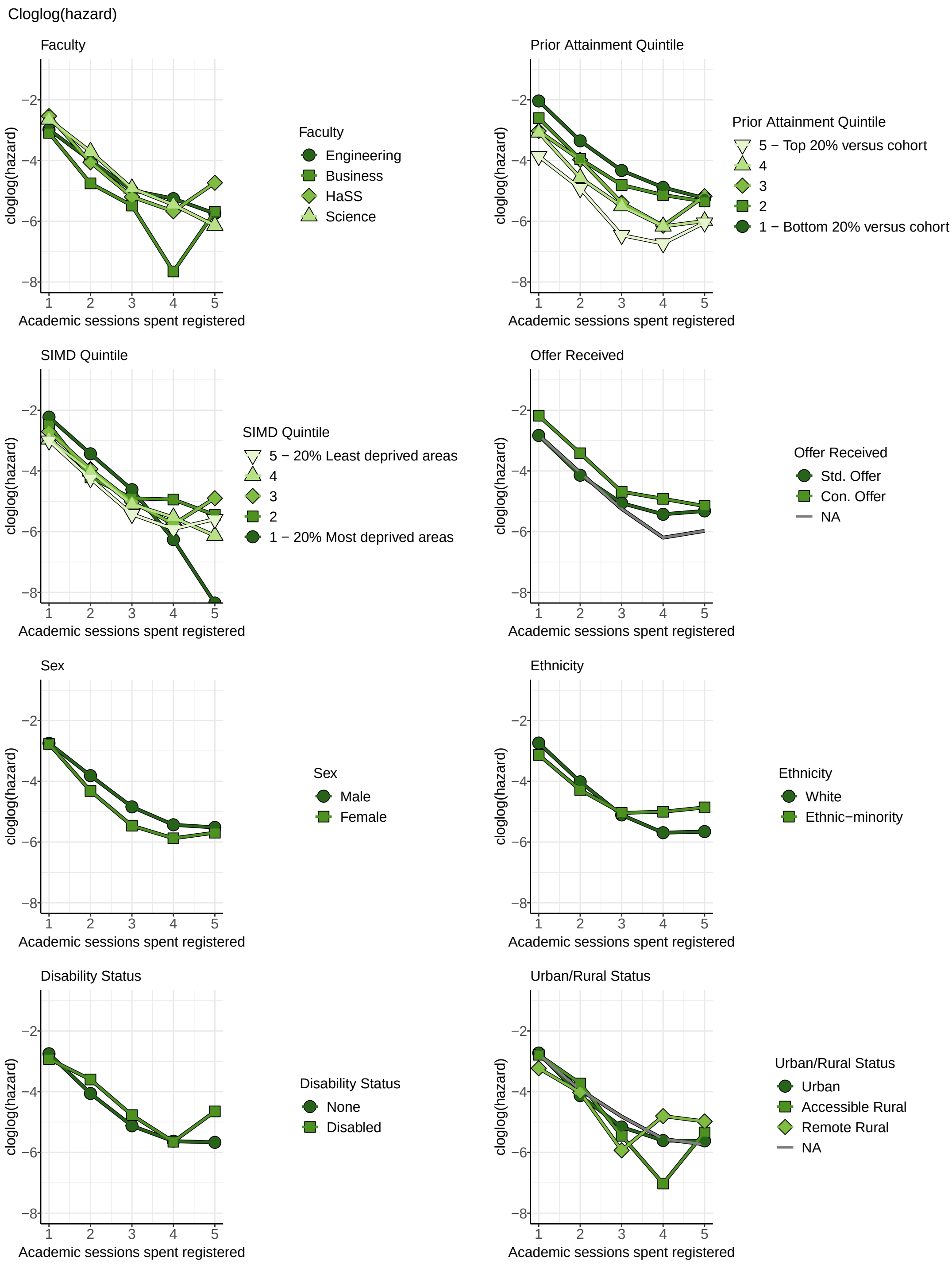


Figure E.4: Comparison of cloglog(hazard) calculated descriptively at each time period, split by levels of each explanatory variable.

E.2 Kaplan-Meier Fits

Additional Kaplan-Meier estimates were derived for Academic Cohort (Figure E.5). Higher rates of dropouts were observed amongst those who entered in 2016/17, 2017/18 and 2018/19 – the cohorts affected by the COVID-19 pandemic assuming the students followed the typical length of a degree programme (4 years). 2018/19 in particular had a large number of dropouts after first year. The cohort with the fewest dropouts was 2013/14.

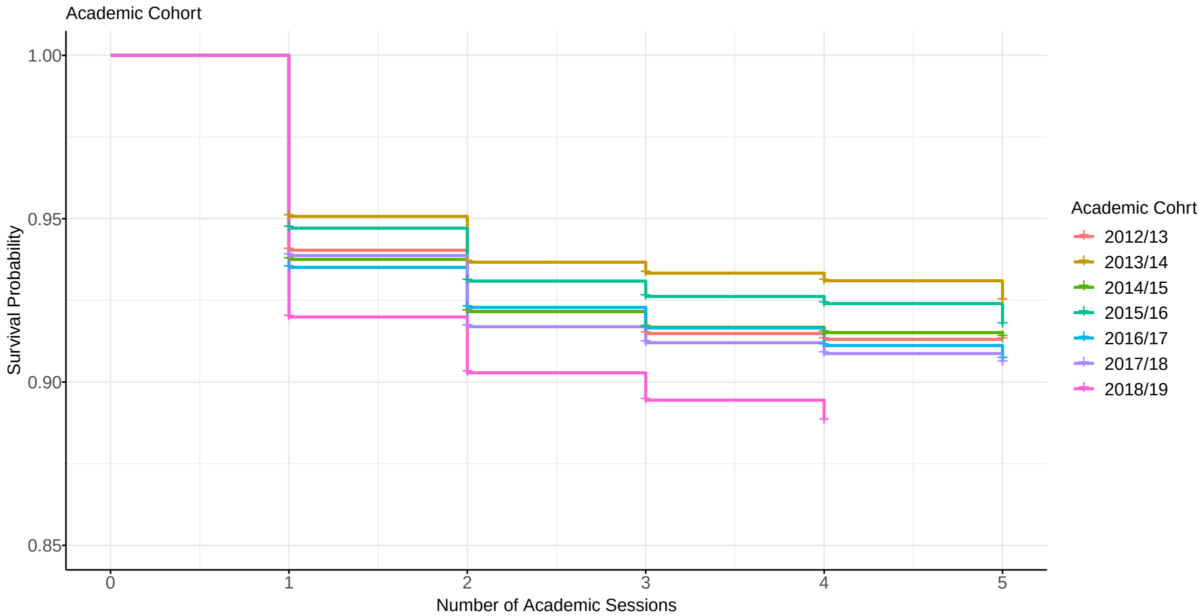


Figure E.5: Kaplan-Meier estimates of dropouts against Academic Cohort.

E.3 Continuous Time-to-Event Multivariable Models

None of the explanatory variables in the Stratified CPH violated the proportional-hazards assumption (Table E.1).

Table E.1: Schoenfeld test statistics for Stratified CPH. Calculated using Kaplan-Meier transformed time.

Variables	Chi-Square	D.o.f	P-values
Prior Attainment Quintile	3.23	4	0.52
SIMD Quintile	2.79	4	0.59
Ethnicity	2.22	1	0.14
Global	8.59	9	0.48

Table E.2: Model estimates for the Stratified CPH Model.

Variables	P-values	Coefficients (S.E.)	Hazard Ratio (95% C.I.)
Prior Att. Quintile 1 (vs 3)	<0.001	0.832 (0.088)	2.297 (1.935,2.727)
Prior Att. Quintile 2 (vs 3)	<0.001	0.335 (0.094)	1.398 (1.162,1.682)
Prior Att. Quintile 4 (vs 3)	0.126	-0.163 (0.107)	0.850 (0.689,1.047)
Prior Att. Quintile 5 (vs 3)	<0.001	-0.865 (0.132)	0.421 (0.325,0.546)
SIMD Quintile 2 (vs 1)	0.015	-0.248 (0.102)	0.780 (0.639,0.953)
SIMD Quintile 3 (vs 1)	0.005	-0.288 (0.102)	0.750 (0.614,0.915)
SIMD Quintile 4 (vs 1)	<0.001	-0.434 (0.100)	0.648 (0.532,0.789)
SIMD Quintile 5 (vs 1)	<0.001	-0.510 (0.093)	0.601 (0.501,0.721)
Ethnic-minority (vs White)	0.003	-0.398 (0.135)	0.672 (0.515,0.875)

References

[1] Scottish Government, "Scottish Index of Multiple Deprivation 2020," https://spatialdata.gov.scot/geonetwork/srv/eng/catalog.search#/metadata/02866b0b-66e5-46ab-9b1c-d433dc3c2fae, Jan. 2020, coordinate reference system - OSGB 1936 / British National Grid (EPSG:27700) Available under the terms of the Open Government Licence. The following attribution statement must be used to acknowledge the source of the information: Copyright Scottish Government, contains Ordnance Survey data © Crown copyright and database right (insert year). Dataset is complete for Scotland. Care should be taken when using this dataset with lookups to other postcode based geographies. Some postcode unit boundaries will have changed since data zones were created therefore exact match of the boundaries are unlikely. [Online]. Available: https://spatialdata.gov.scot/geonetwork/srv/eng/catalog.search#/metadata/02866b0b-66e5-46ab-9b1c-d433dc3c2fae

[2] ——, "Urban Rural 2016 Classifier | Supporting documents," Mar. 2018, iSBN: 978-1-78851-620-4 (web only) Any enquiries regarding this publication should be sent to us at The Scottish Government St Andrew's

House Edinburgh EH1 3DG. [Online]. Available: http://www.gov.scot/publications/scottish-government-urban-rural-classification-2016/

[3] Commission on Widening Access, "A Blueprint for Fairness | The Final Report of the Comission on Widening Access," The Scottish Government, Edinburgh, Tech. Rep., Mar. 2016, iSBN: 978-1-78652-094-4 ISBN: 978-1-78652-095-1 (EPUB) ISBN: 978-1-78652-096-8 (MOBI) Any enquiries regarding this publication should be sent to us at The Scottish Government St Andrew's House Edinburgh EH1 3DG. [Online]. Available: https://www.gov.scot/binaries/content/documents/govscot/publications/progress-report/2016/03/blueprint-fairness-final-report-commission-widening-access/documents/00496619-pdf/00496619-pdf/govscot%3Adocument/00496619.pdf

[4] Universities Scotland, "Futures Not Backgrounds," Aug. 2016. [Online]. Available: http://www.universities-scotland.ac.uk/wp-content/uploads/2016/09/10537-âĂć-Futures-Not-Backgrounds-web.pdf

[5] Scottish Government, "A Plan for Scotland: The Government's Programme for Scotland 2016 -17," Sep. 2016, any enquiries regarding this publication should be sent to us at The Scottish Government St Andrew's House Edinburgh EH1 3DG ISBN: 978-1-78652-439-3. [Online]. Available: https://www.gov.scot/binaries/content/documents/govscot/publications/strategy-plan/2016/09/plan-scotland-scottish-governments-programme-scotland-2016-17/documents/00505210-pdf/00505210-pdf/govscot%3Adocument/00505210.pdf

[6] Commissioner for Fair Access, “Laying the Foundations for Fair Access: Annual Report 2017,” The Scottish Government, Edinburgh, Tech. Rep., Dec. 2017, iSBN: 978-1-78851-512-2. [Online]. Available: https://www.gov.scot/publications/laying-foundations-fair-access-annual-report-2017-commissioner-fair-access/

[7] ——, “Renewing the Alliance for Fair Access | Annual Report for 2024,” The Scottish Government, Edinburgh, Tech. Rep., Jan. 2024, any enquiries regarding this publication should be sent to us at The Scottish Government St Andrew’s House Edinburgh EH1 3DG ISBN: 978-1-83521-906-5 (web only). [Online]. Available: https://www.fairaccess.scot/wp-content/uploads/2024/01/Sixth-Renewing-Alliance-for-Fair-Access-annual-report-January-2024.pdf

[8] University of Strathclyde, “Strathclyde 2030,” 2024. [Online]. Available: https://www.strath.ac.uk/media/1newwebsite/strategicplan2030/Strathclyde-2030-30012024.pdf

[9] Scottish Government, “Chapter 2: School Leaver Cohort,” Feb. 2023. [Online]. Available: http://www.gov.scot/publications/summary-statistics-attainment-initial-leaver-destinations-no-5-2023-edition/pages/4/

[10] Scottish Credit and Qualifications Framework, “Interactive Framework,” Nov. 2023. [Online]. Available: https://scqf.org.uk/about-the-framework/interactive-framework/

[11] UK Government, “What qualification levels mean,” Nov. 2023. [Online]. Available: https://www.gov.uk/what-different-qualification-levels-mean/list-of-qualification-levels

[12] Scottish Qualifications Authority, “What we do,” publisher: SQA. [Online]. Available: https://www.sqa.org.uk/sqa/100149.html

[13] National Careers Service, “A levels | National Careers Service,” Nov. 2023. [Online]. Available: https://nationalcareers.service.gov.uk/explore-your-education-and-training-choices/a-levels

[14] Scottish Qualifications Authority, “Introducing Advanced Highers,” Sep. 2009. [Online]. Available: https://www.sqa.org.uk/files_ccc/NQIntroducingAdvancedHigherLeaflet.pdf

[15] UCAS, “Advanced Higher - new | UCAS Qualification Information Profiles,” 2023. [Online]. Available: https://qips.ucas.com/qip/advanced-higher-new

[16] University of Strathclyde, “How to choose your subjects at school | University of Strathclyde.” [Online]. Available: https://www.strath.ac.uk/studywithus/undergraduate/beforeyouapply/howtochooseyoursubjectsatschool/

[17] University of Oxford, “Response to Scottish Qualifications Reform | University of Oxford.” [Online]. Available: https://www.ox.ac.uk/admissions/undergraduate/courses/entrance-requirements/response-scottish-qualifications-reform

[18] UCAS, “Don’t miss your deadline!” 2023. [Online]. Available: https://www.ucas.com/ucas/undergraduate/getting-started/when-apply

[19] ——, “Not holding any offers? You might be able to apply for more with Extra!” 2023. [Online]. Available: https://www.ucas.com/ucas/undergraduate/apply-and-track/track-your-application/extra-choices

[20] ——, “Types of undergraduate offers,” 2024. [Online]. Available: https://www.ucas.com/applying/after-you-apply/types-undergraduate-offers

[21] ——, “Admissions tests,” 2023. [Online]. Available: https://www.ucas.com/undergraduate/applying-university/admissions-tests

[22] Universities Scotland, “Who We Represent,” Oct. 2023. [Online]. Available: https://www.universities-scotland.ac.uk/about-us/who-we-represent/

[23] Russell Group, “Russell Group | About,” Jan. 2024. [Online]. Available: https://russellgroup.ac.uk/about/

[24] Higher Education Statistics Agency, “Higher Education Student Statistics: UK, 2022/23 - Where students come from and go to study | HESA,” Aug. 2024. [Online]. Available: https://www.hesa.ac.uk/news/08-08-2024/sb269-higher-education-student-statistics/location

[25] Times Higher Education, “Best universities in the UK 2024 - University Rankings,” Feb. 2024. [Online]. Available: https://www.timeshighereducation.com/student/best-universities/best-universities-uk

[26] Scottish Funding Council, "Funding for universities." [Online]. Available: https://www.sfc.ac.uk/funding/university-funding/university-funding.aspx

[27] ——, "Outcome Agreements." [Online]. Available: https://www.sfc.ac.uk/funding/outcome-agreements/outcome-agreements.aspx

[28] Student Awards Agency Scotland, "Undergrad Funding - SAAS - Higher Education - Student Loan - Bursary," Nov. 2023. [Online]. Available: https://www.saas.gov.uk/full-time/funding-information-undergraduate

[29] Audit Scotland, "Audit of higher education in Scottish universities," Jul. 2016, audit Scotland, 4th Floor, 102 West Port, Edinburgh EH3 9DN T: 0131 625 1500 E: info@audit-scotland.gov.uk ISBN 978 1 911494 00 3. [Online]. Available: https://www.audit-scotland.gov.uk/uploads/docs/report/2016/nr_160707_higher_education.pdf

[30] ——, "Finances of Scottish universities," Sep. 2019, audit Scotland, 4th Floor, 102 West Port, Edinburgh EH3 9DN T: 0131 625 1500 E: info@audit-scotland.gov.uk ISBN 978 1 913287 08 5 AGS/2019/6. [Online]. Available: https://www.audit-scotland.gov.uk/uploads/docs/report/2019/nr_190919_finances_universities.pdf

[31] BBC News, "Coronavirus: Scottish exams cancelled for first time in history," *BBC News*, Mar. 2020. [Online]. Available: https://www.bbc.com/news/uk-scotland-51958208

[32] H. Richardson, "Coronavirus: Teachers to estimate grades after exams cancelled," *BBC News*, Mar. 2020. [Online]. Available: https://www.bbc.com/news/education-51980831

[33] Scottish Qualifications Authority, "2020 Alternative Certification Model: Equality Impact Assessment," Aug. 2020. [Online]. Available: https://www.sqa.org.uk/sqa/files_ccc/2020-sqa-alternative-certification-model-equality-impact-assessment.pdf

[34] BBC News, "Nicola Sturgeon 'sorry' over Scottish exam results," *BBC News*, Aug. 2020. [Online]. Available: https://www.bbc.com/news/uk-scotland-scotland-politics-53719477

[35] ——, "Scottish school pupils have results upgraded," *BBC News*, Aug. 2020. [Online]. Available: https://www.bbc.com/news/uk-scotland-53740588

[36] S. Price, "Official Statistics Attainment Statistics - December 2022," Dec. 2022. [Online]. Available: https://www.sqa.org.uk/sqa/files_ccc/attainment-statistics-december-2022-summary.pdf

[37] Commissioner for Fair Access, "Discussion Paper: Retention, Outcomes and Destinations," Jan. 2018, iSBN: 978-1-78851-567-2 (web only) Any enquiries regarding this publication should be sent to us at The Scottish Government St Andrew's House Edinburgh EH1 3DG. [Online]. Available: https://www.gov.scot/publications/commissioner-fair-access-discussion-paper-retention-outcomes-destinations/

[38] J. McDonald, "Update 15 March | University of Strathclyde," Mar. 2020. [Online]. Available: https://www.strath.ac.uk/coronavirus/students/studentupdates/update15march/

[39] ——, "Update 18 March | University of Strathclyde," Mar. 2020. [Online]. Available: https://www.strath.ac.uk/coronavirus/students/studentupdates/update18march/

[40] Assessment and Feedback Working Group, "Policy on invigilation of examinations," 2019. [Online]. Available: https://www.strath.ac.uk/media/ps/cs/gmap/academicaffairs/policies/Policy_on_Invigilation_of_Examinations.pdf

[41] J. McDonald, "Update 26 March | University of Strathclyde," Mar. 2020. [Online]. Available: https://www.strath.ac.uk/coronavirus/students/studentupdates/update26march/

[42] ——, "Update 30 March | University of Strathclyde," Mar. 2020. [Online]. Available: https://www.strath.ac.uk/coronavirus/students/studentupdates/update30march/

[43] ——, "Update 19 February 2021 | University of Strathclyde," Feb. 2021. [Online]. Available: https://www.strath.ac.uk/coronavirus/students/studentupdates/update19february2021/

[44] ——, "Update 21 September | University of Strathclyde," Sep. 2021. [Online]. Available: https://www.strath.ac.uk/coronavirus/students/studentupdates/update21september/

[45] ——, “Update 9 October | University of Strathclyde,” Oct. 2021. [Online]. Available: https://www.strath.ac.uk/coronavirus/students/studentupdates/update9october/

[46] ——, “Update 12 April 2021 | University of Strathclyde,” Apr. 2021. [Online]. Available: https://www.strath.ac.uk/coronavirus/students/studentupdates/update12april2021/

[47] Assessment and Feedback Working Group, “Policy on Compensation Scheme,” 2019. [Online]. Available: https://www.strath.ac.uk/media/ps/cs/gmap/academicaffairs/policies/Policy_on_Compensation_Scheme.pdf

[48] University of Strathclyde, “Equality Mainstreaming Report 2021 | University of Strathclyde,” University of Strathclyde, Equality Mainstreaming Report, 2021. [Online]. Available: https://www.strath.ac.uk/media/ps/sees/equality/Equality_Mainstreaming_Report_2021.pdf

[49] K.-M. Burns, “No Detriment Policy,” Jan. 2021. [Online]. Available: https://www.strathunion.com/news/article/strathunion/no-detriment-policy/

[50] Commissioner for Fair Access, “The Impact of COVID-19 on Fair Access to Higher Education,” Dec. 2020, any enquiries regarding this publication should be sent to us at The Scottish Government St Andrew’s House Edinburgh EH1 3DG ISBN: 978-1-80004-501-9 (web only). [Online]. Available: https://www.gov.scot/publications/impact-covid-19-fair-access-higher-education/

[51] V. Boliver, S. Gorard, and N. Siddiqui, "Who counts as socioeconomically disadvantaged for the purposes of widening access to higher education?" *British Journal of Sociology of Education*, vol. 43, no. 3, pp. 349–374, Apr. 2022, publisher: Routledge _eprint: https://doi.org/10.1080/01425692.2021.2017852. [Online]. Available: https://doi.org/10.1080/01425692.2021.2017852

[52] V. Boliver, S. Gorard, M. Powell, and T. Moreira, "Mapping and evaluating the use of contextual data in undergraduate admissions in Scotland." Scottish Funding Council, Edinburgh, Tech. Rep., Oct. 2017. [Online]. Available: https://dro.dur.ac.uk/23459/

[53] ——, "The use of access thresholds to widen participation at Scottish universities," *Scottish Affairs*, vol. 29, no. 1, Jan. 2020, publisher: Edinburgh University Press. [Online]. Available: https://durham-repository.worktribe.com/output/1277721

[54] S. Gorard, N. Siddiqui, and V. Boliver, "An Analysis of School-Based Contextual Indicators for Possible Use in Widening Participation," *Higher Education Studies*, vol. 7, no. 2, May 2017, publisher: Canadian Center of Science and Education (CCSE). [Online]. Available: https://durham-repository.worktribe.com/output/1357788

[55] S. Gorard, V. Boliver, N. Siddiqui, and P. Banerjee, "Which are the most suitable contextual indicators for use in widening participation to HE?" *Research Papers in Education*, vol. 34, no. 1, pp. 99–129, 2019.

[56] University of Strathclyde, "Widening access | University of Strathclyde," Jan. 2025. [Online]. Available: https://www.strath.ac.uk/studywithus/wideningaccess/

[57] Scottish Government, "Scottish Index of Multiple Deprivation 2020: introduction," Jan. 2020. [Online]. Available: http://www.gov.scot/publications/scottish-index-multiple-deprivation-2020/

[58] ——, "Scottish Index of Multiple Deprivation 2020: technical notes," Jan. 2020. [Online]. Available: http://www.gov.scot/publications/scottish-index-multiple-deprivation-2020/

[59] ——, "Scottish Index of Multiple Deprivation 2020 - gov.scot," 2020. [Online]. Available: https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/

[60] University of Strathclyde, "University of Strathclyde Outcome Agreement 2021-22," May 2022. [Online]. Available: https://www.sfc.ac.uk/funding/outcome-agreements/2021-22/university-strathclyde-oa2021-22.aspx

[61] Scottish Government, "Scottish Index of Multiple Deprivation 2016 - gov.scot," 2016. [Online]. Available: https://webarchive.nrscotland.gov.uk/20190708060827/https://www2.gov.scot/Topics/Statistics/SIMD

[62] ——, "Scottish Index of Multiple Deprivation (Historical) - 2004-2012." [Online]. Available: https://statistics.gov.scot/data/scottish-index-of-multiple-deprivation-historical-i

[63] ——, “Scottish Index of Multiple Deprivation 2020v2 local and national share calculator,” Jan. 2020. [Online]. Available: http://www.gov.scot/publications/scottish-index-of-multiple-deprivation-2020v2-local-and-national-share-calculator-2/

[64] University of Strathclyde, “Admissions Policy,” 2024. [Online]. Available: https://www.strath.ac.uk/media/ps/cs/gmap/academicaffairs/policies/Admissions_Policy_July_2024_(1).pdf

[65] UCAS, “Students with caring responsibilities,” 2024. [Online]. Available: https://www.ucas.com/undergraduate/applying-university/individual-needs/students-caring-responsibilities

[66] A. Blackshaw, “Contextual Data in Admissions Policy – Proposed changes to the low progression schools list,” May 2021, contact a.blackshaw@strath.ac.uk to request access to this document.

[67] A. Learmonth, “Revealed: The Scottish unis where fewer than 10 of new students are working class,” *The Herald*, Jul. 2023. [Online]. Available: https://www.heraldscotland.com/news/23678207.three-scottish-universities-fail-meet-working-class-student-target/

[68] Commissioner for Fair Access, “Maintaining the Momentum Towards Fair Access: annual report 2022,” May 2022, iSBN: 978-1-80435-482-7 (web only) Any enquiries regarding this publication should be sent to us at The Scottish Government St Andrew’s House Edinburgh EH1 3DG. [Online]. Available: http://www.gov.scot/publications/maintaining-momentum-towards-fair-access-annual-report-2022/

[69] L. Croxford, G. Docherty, R. Gaukroger, and K. Hood, "Widening Participation at the University of Edinburgh: Contextual Admissions, Retention, and Degree Outcomes," *Scottish Affairs*, vol. 23, no. 2, pp. 192–216, May 2014. [Online]. Available: https://www.euppublishing.com/doi/epdf/10.3366/scot.2014.0017?src=getftr

[70] Scottish Government, "Achievement of Curriculum for Excellence (CfE) Levels 2021/22," Scottish Government, Tech. Rep., Dec. 2022.

[71] L. Croxford, "Change Over Time in the Context, Outcomes and Inequalities of Secondary Schooling in Scotland, 1985-2005," The Scottish Government, Edinburgh, Tech. Rep., May 2009.

[72] J. McVitie, "Policy and New Products | Research Report 3 | National Qualifications: a short history," Oct. 2008. [Online]. Available: https://www.sqa.org.uk/files_ccc/PNP_ResearchReport3_NationalQualificationsAShortHistory.pdf

[73] Scottish Government, "Fair Access: Analysis of school leaver attainment by SIMD quintile," May 2018. [Online]. Available: http://www.gov.scot/publications/fair-access-analysis-of-school-leaver-attainment-by-simd-quintile/

[74] V. Boliver, "How fair is access to more prestigious UK universities?" *The British Journal of Sociology*, vol. 64, no. 2, pp. 344–364, 2013, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-4446.12021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-4446.12021

[75] Commissioner for Fair Access, “Building on Progress Towards Fair Access: Annual Report 2019,” The Scottish Government, Tech. Rep., Jun. 2019. [Online]. Available: https://www.gov.scot/binaries/content/documents/govscot/publications/corporate-report/2019/06/commissioner-fair-access-annual-report-2019-building-progress-towards-fair-access/documents/building-progress-towards-fair-access-annual-report-2019/building-progress-towards-fair-access-annual-report-2019/govscot%3Adocument/building-progress-towards-fair-access-annual-report-2019.pdf

[76] L. Lasselle, J. McDougall-Bagnall, and I. Smith, “School grades, school context and university degree performance: evidence from an elite Scottish institution,” *Oxford Review of Education*, vol. 40, no. 3, pp. 293–314, May 2014, publisher: Routledge _eprint: https://doi.org/10.1080/03054985.2014.900485. [Online]. Available: https://doi.org/10.1080/03054985.2014.900485

[77] C. Crawford, “The link between secondary school characteristics and university participation and outcomes,” Institute for Fiscal Studies and University of Warwick, Londond, Tech. Rep., Jun. 2014, iSBN: 978-1-78105-356-0. [Online]. Available: https://assets.publishing.service.gov.uk/media/5a7e2fc2ed915d74e33f09cc/RR353_-_The_link_between_secondary_school_characteristics_and_university_participation_and_outcomes_FINAL.pdf

[78] A. Hoare and R. Johnston, “Widening participation through admissions policy – a British case study of school and university performance,” *Studies in Higher Education*, vol. 36, no. 1, pp. 21–41, Feb. 2011, publisher: Routledge _eprint: https://doi.org/10.1080/03075070903414297. [Online]. Available: https://doi.org/10.1080/03075070903414297

[79] A. Cameron, Y. Wharton, and J. B. Scally, “An investigation into the comparative learning gain and ‘value added’ for students from widening participation and non-widening participation groups: a case study from sports degrees,” *Higher Education Pedagogies*, vol. 3, no. 1, pp. 83–102, Jan. 2018, publisher: Routledge _eprint: https://doi.org/10.1080/23752696.2018.1435297. [Online]. Available: https://doi.org/10.1080/23752696.2018.1435297

[80] Universities Scotland, “84% of Scottish graduates say going to university got them the job they wanted,” Jul. 2023. [Online]. Available: https://www.universities-scotland.ac.uk/graduate-success/

[81] T. Byrom and N. Lightfoot, “Interrupted trajectories: the impact of academic failure on the social mobility of working-class students,” *British Journal of Sociology of Education*, vol. 34, no. 5-6, pp. 812–828, Nov. 2013, publisher: Routledge _eprint: https://doi.org/10.1080/01425692.2013.816042. [Online]. Available: https://doi.org/10.1080/01425692.2013.816042

[82] B. Waltmann, L. Van Der Erve, L. Dearden, and J. Britton, “The impact of undergraduate degrees on lifetime earnings,” The IFS, Tech. Rep., Feb. 2020. [Online]. Available: https://www.ifs.org.uk/publications/14729

[83] L. Macmillan and A. Vignoles, “Mapping the occupational destinations of new graduates,” Social Mobility and Child Poverty Commission, Research report, Oct. 2013, iSBN: 978-1-78105-276-1 Any enquiries regarding this publication should be sent to us at Social Mobility and Child Poverty Commission, Sanctuary Buildings, 20 Great Smith Street, London, SW1P 3BT. Email: contact@smcpcommission.gov.uk.

[84] A. Vignoles and C. Crawford, “Heterogeneity in graduate earnings by socio-economic background,” IFS, Tech. Rep., Oct. 2014. [Online]. Available:
https://www.ifs.org.uk/uploads/publications/wps/WP201430.pdf

[85] Higher Education Funding Council for England, “Differences in employment outcomes: equality and diversity characteristics,” Sep. 2015, enquiries to Alison Brunt, tel 0117 931 7166, email a.brunt@hefce.ac.uk, or Quantitative Analysis for Policy Team, email qapt@hefce.ac.uk Reference 2015/23. [Online]. Available:
http://dera.ioe.ac.uk/24435/1/HEFCE2015_23.pdf

[86] C. Belfield, J. Britton, F. Buscha, M. Dickson, A. Vignoles, I. Walker, and Y. Zhu, “The relative labour market returns to different degrees,” Institute for Fiscal Studies, Research Report, Jun. 2018, reference: DFE-RR787 ISBN: 978-1-78105-912-8 Any enquiries regarding this publication should be sent to us at: chris_b@ifs.org.uk,

jack_b@ifs.org.uk, laura_v@ifs.org.uk or www.education.gov.uk/contactus. [Online]. Available: https://assets.publishing.service.gov.uk/media/5b1906e540f0b634b73dbeb6/The_relative_labour_market-returns_to_different_degrees.pdf

[87] S. Schwarz, "Fair admissions to higher education: recommendations for good practice," Department for Education and Skills Publications, Nottingham, Tech. Rep., 2004. [Online]. Available: https://dera.ioe.ac.uk/id/eprint/5284/1/finalreport.pdf

[88] The National Committee of Inquiry into Higher Education, "The Dearing Report | Higher Education in the learning society," 1997. [Online]. Available: https://education-uk.org/documents/dearing1997/dearing1997.html

[89] Scottish Parliament, "Meeting of the Parliament," Nov. 2014, archive Location: UK Last Modified: 24/01/2014 15:58:03 Publisher: The Scottish Parliament. [Online]. Available: https://archive2021.parliament.scot/parliamentarybusiness/report.aspx?r=9653&mode=html#iob_87656

[90] Commission on Widening Access, "Commission on Widening Access Interim Report: November 2015," The Scottish Government, Tech. Rep., Nov. 2015, iSBN: 978-1-78544-813-3 ISBN: 978-1-78544-814-0 (EPUB) ISBN: 978-1-78544-815-7 (MOBI) Any enquiries regarding this publication should be sent to us at The Scottish Government St Andrew's House Edinburgh EH1 3DG. [Online]. Available: https://dera.ioe.ac.uk/id/eprint/24745/1/00489004.pdf

[91] Commissioner for Fair Access, "Re-Committing to Fair Access: A Plan for Recovery Annual Report 2021," The Scottish Government, Tech. Rep., Jun. 2021. [Online]. Available: https://www.gov.scot/publications/re-committing-fair-access-plan-recovery-annual-report-2021/

[92] V. Boliver, M. Powell, and T. Moreira, "Organisational Identity as a Barrier to Widening Access in Scottish Universities," *Social Sciences*, vol. 7, no. 9, p. 151, Sep. 2018, number: 9 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2076-0760/7/9/151

[93] Universities Scotland, "Working to Widen Access," Nov. 2017. [Online]. Available: https://www.universities-scotland.ac.uk/wp-content/uploads/2019/02/Working-to-Widen-Access.pdf

[94] A. Mountford-Zimdars and J. Moore, "Identifying merit and potential beyond grades: opportunities and challenges in using contextual data in undergraduate admissions at nine highly selective English universities," *Oxford Review of Education*, vol. 46, no. 6, pp. 752–769, Nov. 2020, publisher: Routledge _eprint: https://doi.org/10.1080/03054985.2020.1785413. [Online]. Available: https://doi.org/10.1080/03054985.2020.1785413

[95] V. Boliver, "Misplaced optimism: how higher education reproduces rather than reduces social inequality," *British Journal of Sociology of Education*, vol. 38, no. 3, Mar. 2017, publisher: Taylor and Francis Group. [Online]. Available: https://durham-repository.worktribe.com/output/1362703

[96] Higher Education Statistics Agency, "Who's studying in HE? | HESA." [Online]. Available: https://www.hesa.ac.uk/data-and-analysis/students/whos-in-he

[97] V. Boliver and M. Powell, "Competing conceptions of fair admission and their implications for supporting students to fulfil their potential at university," *Perspectives: Policy and Practice in Higher Education*, vol. 27, no. 1, pp. 8–15, Jan. 2023, publisher: Routledge _eprint: https://doi.org/10.1080/13603108.2022.2063429. [Online]. Available: https://doi.org/10.1080/13603108.2022.2063429

[98] Scottish Government, "Scottish Government Urban Rural Classification 2020," May 2022, iSBN: 978-1-80435-481-0 (web only). [Online]. Available: https://www.gov.scot/binaries/content/documents/govscot/publications/advice-and-guidance/2022/05/scottish-government-urban-rural-classification-2020/documents/scottish-government-urban-rural-classification-2020/scottish-government-urban-rural-classification-2020/govscot%3Adocument/scottish-government-urban-rural-classification-2020.pdf

[99] A. J. Dobson and A. G. Barnett, *An Introduction to Generalized Linear Models*, 4th ed. New York: Chapman and Hall/CRC, Apr. 2018.

[100] D. F. Moore, *Applied Survival Analysis Using R*, ser. Use R! Cham: Springer International Publishing, 2016. [Online]. Available: http://link.springer.com/10.1007/978-3-319-31245-3

[101] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R / Second Edition*, 2nd ed., ser. Springer Texts in Statistics. New York: Springer, 2021, corrected 2023-06-01 ISSN: 1431-875X ISBN: 9781071614174 ISBN: 1071614177 EISSN: 2197-4136 EISBN: 9781071614181 EISBN: 1071614185 OCLC: 1262436825. [Online]. Available: https://www.statlearning.com

[102] E. Vittinghoff, D. V. Glidden, S. C. Shiboski, and C. E. McCulloch, *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*, ser. Statistics for Biology and Health. Boston, MA: Springer US, 2012. [Online]. Available: https://link.springer.com/10.1007/978-1-4614-1353-0

[103] J. D. Singer and J. B. Willett, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* Oxford University Press, May 2003. [Online]. Available: https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

[104] G. Tutz and M. Schmid, *Modeling Discrete Time-to-Event Data*, ser. Springer Series in Statistics. Cham: Springer International Publishing, 2016. [Online]. Available: http://link.springer.com/10.1007/978-3-319-28158-2

[105] T. M. Therneau, P. M. Grambsch, and T. R. Fleming, "Martingale-Based Residuals for Survival Models," *Biometrika*, vol. 77, no. 1, pp. 147–160, 1990, publisher: [Oxford University Press, Biometrika Trust]. [Online]. Available: https://www.jstor.org/stable/2336057

[106] D. G. Altman, J. J. Deeks, and D. L. Sackett, "Odds ratios should be avoided when events are common," *BMJ : British Medical Journal*, vol. 317, no. 7168, p. 1318, Nov. 1998. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1114216/

[107] B. Chu, M. Liu, E. C. Leas, B. M. Althouse, and J. W. Ayers, "Effect size reporting among prominent health journals: a case study of odds ratios," *BMJ Evidence-Based Medicine*, vol. 26, no. 4, pp. 184–184, Aug. 2021, publisher: Royal Society of Medicine Section: Original research. [Online]. Available: https://ebm.bmj.com/content/26/4/184

[108] H. T. O. Davies, I. K. Crombie, and M. Tavakoli, "When can odds ratios mislead?" *BMJ*, vol. 316, no. 7136, pp. 989–991, Mar. 1998, publisher: British Medical Journal Publishing Group Section: Information in practice. [Online]. Available: https://www.bmj.com/content/316/7136/989

[109] B. A. F. Martinez, V. B. Leotti, G. d. S. E. Silva, L. N. Nunes, G. Machado, and L. G. Corbellini, "Odds Ratio or Prevalence Ratio? An Overview of Reported Statistical Methods and Appropriateness of Interpretations in Cross-sectional Studies with Dichotomous Outcomes in Veterinary Medicine," *Frontiers in Veterinary Science*, vol. 4, p. 193, 2017.

[110] E. C. Norton, B. E. Dowd, and M. L. Maciejewski, "Odds Ratios—Current Best Practice and Use," *JAMA*, vol. 320, no. 1, p. 84, Jul. 2018. [Online]. Available: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2018.6971

[111] G. Tajeu, B. Sen, D. B. Allison, and N. Menachemi, “Misuse of Odds Ratios in Obesity Literature: An Empirical Analysis of Published Studies,” *Obesity (Silver Spring, Md.)*, vol. 20, no. 8, pp. 1726–1731, Aug. 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3399983/

[112] E. C. Norton and B. E. Dowd, “Log Odds and the Interpretation of Logit Models,” *Health Services Research*, vol. 53, no. 2, pp. 859–878, 2018, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-6773.12712. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-6773.12712

[113] L. C. Kleinman and E. C. Norton, “What’s the Risk? A Simple Approach for Estimating Adjusted Risk Measures from Nonlinear Models Including Logistic Regression,” *Health Services Research*, vol. 44, no. 1, pp. 288–302, 2009, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-6773.2008.00900.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-6773.2008.00900.x

[114] T. Williamson, M. Eliasziw, and G. H. Fick, “Log-binomial models: exploring failed convergence,” *Emerging Themes in Epidemiology*, vol. 10, no. 1, p. 14, Dec. 2013. [Online]. Available: https://doi.org/10.1186/1742-7622-10-14

[115] W. Chen, L. Qian, J. Shi, and M. Franklin, “Comparing performance between log-binomial and robust Poisson regression models for estimating risk ratios under model misspecification,” *BMC Medical Research Methodology*, vol. 18, no. 1, p. 63, Jun. 2018. [Online]. Available: https://doi.org/10.1186/s12874-018-0519-5

[116] C. Zocchetti, D. Consonni, and P. A. Bertazzi, “Estimation of Prevalence Rate Ratios from Cross-Sectional Data,” *International Journal of Epidemiology*, vol. 24, no. 5, pp. 1064–1065, 1995. [Online]. Available: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/24.5.1064

[117] G. Zou, “A Modified Poisson Regression Approach to Prospective Studies with Binary Data,” *American Journal of Epidemiology*, vol. 159, no. 7, pp. 702–706, Apr. 2004. [Online]. Available: https://doi.org/10.1093/aje/kwh090

[118] A. Zeileis, “Object-Oriented Computation of Sandwich Estimators,” *Journal of Statistical Software*, vol. 16, no. 9, pp. 1–16, 2006. [Online]. Available: <https://doi.org/10.18637/jss.v016.i09>

[119] S. P. Jenkins, “Survival Analysis,” 2004. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9bb46b98492c0d8e33ffbddab4a0f99d84f3f0c0

[120] T. M. Therneau, “A Package for Survival Analysis in R,” 2023. [Online]. Available: https://CRAN.R-project.org/package=survival

[121] ——, “A package for survival analysis in R,” Feb. 2024. [Online]. Available: https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf

[122] D. R. Cox, "Regression Models and Life-Tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1972.tb00899.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1972.tb00899.x

[123] B. Efron, "The Efficiency of Cox's Likelihood Function for Censored Data," *Journal of the American Statistical Association*, vol. 72, no. 359, pp. 557–565, 1977, publisher: [American Statistical Association, Taylor & Francis, Ltd.]. [Online]. Available: https://www.jstor.org/stable/2286217

[124] J. Johnson and D. G. Hayward, "Scottish Highers and Advanced Highers," Jul. 2008. [Online]. Available: https://www.ucas.com/sites/default/files/ScottishQs.pdf

[125] Scottish Qualifications Authority, "A to Z of National Qualification (NQ) subjects," publisher: SQA. [Online]. Available: //www.sqa.org.uk/sqa/CCC_FirstPage.jsp

[126] University of Edinburgh, "Undergraduate Study 2024 Entry | SQA Highers and Advanced Highers," Nov. 2023. [Online]. Available: https://www.ed.ac.uk/studying/undergraduate/entry-requirements/scottish-qualifications/highers

[127] University of Glasgow, "University of Glasgow - Undergraduate study - Information about entry requirements - 2024 Admissions guidance," 2023. [Online]. Available: https://www.gla.ac.uk/undergraduate/entryrequirements/guidance/

[128] University of Strathclyde, "SUBJECT GUIDANCE FOR THE SENIOR PHASE," Feb. 2024. [Online]. Available: https://www.strath.ac.uk/media/1newwebsite/documents/Senior_Phase_Guide_2024.pdf

[129] Scottish Government, "Review of the Impact of Regional Improvement Collaboratives on Pupils and Practitioners," Jan. 2024, any enquiries regarding this publication should be sent to us at The Scottish Government St Andrew's House Edinburgh EH1 3DG ISBN: 978-1-83521-824-2 (web only). [Online]. Available: https://www.gov.scot/binaries/content/documents/govscot/publications/corporate-report/2024/01/review-impact-regional-improvement-collaboratives-pupils-practitioners/documents/review-impact-regional-improvement-collaboratives-pupils-practitioners/review-impact-regional-improvement-collaboratives-pupils-practitioners/govscot%3Adocument/review-impact-regional-improvement-collaboratives-pupils-practitioners.pdf

[130] University of Dundee, "External partners for widening access | University of Dundee, UK." [Online]. Available: https://www.dundee.ac.uk/widening-access/external-partners

[131] K. MacFarlane, "An Evaluation of the Advanced Higher Hub: The Learner Perspective," Jan. 2018. [Online]. Available: https://archive2021.parliament.scot/S5_Education/Inquiries/20190311GCU_AdvancedHigherHub.pdf

[132] B. Borland, “More than one in 10 schools no longer offer Advanced Higher England and Maths,” Mar. 2023, section: Scottish News. [Online]. Available: https://www.scottishdailyexpress.co.uk/news/scottish-news/more-one-10-scots-schools-29536024

[133] The Herald, “Warning over lack of access to Advanced Highers for poor pupils,” May 2018. [Online]. Available: https://www.heraldscotland.com/news/16218177.warning-lack-access-advanced-highers-poor-pupils/

[134] S. Phipps, “‘Postcode lottery’ of Scotland’s Advanced Highers,” May 2018, section: Examinations & Qualifications. [Online]. Available: https://www.sec-ed.co.uk/content/news/postcode-lottery-of-scotland-s-advanced-highers/

[135] The Herald, “Award-winning Scots higher education programme to be scrapped amid cuts,” Jan. 2024. [Online]. Available: https://www.heraldscotland.com/news/24057933.glasgow-advanced-higher-hub-close-amid-government-council-cuts/

[136] L. Mclaughlin, “A sign of decline: Shutting down Scotland’s Advanced Higher Hub is a backwards step,” Jan. 2024. [Online]. Available: https://www.heraldscotland.com/opinion/24069211.glasgow-advanced-higher-hub-closure-backward-step/

[137] J. Wands, “’Glasgow’s Advanced Higher Hub got me to Oxford, its closure saddens me’,” Jan. 2024. [Online]. Available: https://www.heraldscotland.com/opinion/24069222.closure-glasgow-caledonians-advanced-higher-hub-saddens/

[138] J. McEnaney, “Calls for U-turn over closure of "essential" Advanced Highers Hub,” Jan. 2024. [Online]. Available: https://www.heraldscotland.com/news/24069943.calls-u-turn-closure-essential-education-hub/

[139] C. Cannon, “Council defends Advanced Highers Hub closure plan,” Jan. 2024. [Online]. Available: https://www.heraldscotland.com/opinion/24075819.glasgows-new-model-advanced-highers-will-increase-opportunities/

[140] R Core Team, “R: A Language and Environment for Statistical Computing,” Vienna, Austria, Jun. 2023. [Online]. Available: https://www.R-project.org/

[141] D. Lüdecke, “sjPlot: Data Visualization for Statistics in Social Science,” 2024, r package version 2.8.16. [Online]. Available: https://CRAN.R-project.org/package=sjPlot

[142] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani, “Welcome to the tidyverse,” *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, 2019. [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.01686

[143] T. L. Pedersen, “patchwork: The Composer of Plots,” 2024, r package version 1.2.0. [Online]. Available: https://CRAN.R-project.org/package=patchwork

[144] D. B. Dahl, D. Scott, C. Roosen, A. Magnusson, and J. Bates, "xtable: Export Tables to LaTeX or HTML," 2019, r package version 1.8-4. [Online]. Available: https://CRAN.R-project.org/package=xtable

[145] N. P. Burns, D. Young, A. Sherriff, P. Black, A. Blackshaw, and L. Kelly, "Tracking the Success of Contextual Offer Students at One Scottish Higher Education Institution," *Higher Education Quarterly*, vol. 79, no. 1, p. e70011, 2025, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/hequ.70011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/hequ.70011

[146] V. Arel-Bundock, N. Greifer, and A. Heiss, "How to Interpret Statistical Models Using marginaleffects in R and Python," *Journal of Statistical Software*, vol. 111, no. 9, pp. 1–32, 2024. [Online]. Available: https://marginaleffects.com

[147] K. L. Friend, "The creation of social networks: social capital and the experiences of widening participation students at three elite institutions in the US, England, and Scotland," *Pedagogy, Culture and Society*, vol. 29, no. 3, pp. 359–377, May 2021, number: 3 Publisher: Routledge. [Online]. Available: https://doi.org/10.1080/14681366.2020.1735496

[148] W. Arulampalam, R. A. Naylor, and J. P. Smith, "A Hazard Model of the Probability of Medical School Drop-out in the UK," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, vol. 167, no. 1, pp. 157–178, 2004, publisher: [Wiley, Royal Statistical Society]. [Online]. Available: https://www.jstor.org/stable/3559804

[149] C. Aina, “Parental background and university dropout in Italy,” *Higher Education*, vol. 65, no. 4, pp. 437–456, Apr. 2013. [Online]. Available: https://doi.org/10.1007/s10734-012-9554-z

[150] L. Respondek, T. Seufert, J. M. Hamm, and U. E. Nett, “Linking changes in perceived academic control to university dropout and university grades: A longitudinal approach.” *Journal of Educational Psychology*, vol. 112, no. 5, pp. 987–1002, Jul. 2020. [Online]. Available: https://doi.apa.org/doi/10.1037/edu0000388

[151] J. Berens, K. Schneider, S. Görtz, S. Oster, and J. Burghoff, “Early Detection of Students at Risk – Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods,” Rochester, NY, 2018. [Online]. Available: https://papers.ssrn.com/abstract=3275433

[152] P. A. Murtaugh, L. D. Burns, and J. Schuster, “Predicting the Retention of University Students,” *Research in Higher Education*, vol. 40, no. 3, pp. 355–371, Jun. 1999. [Online]. Available: https://doi.org/10.1023/A:1018755201899

[153] G. Nicholls, H. Wolfe, M. Besterfield-Sacre, and L. Shuman, “Predicting Post Secondary Educational Outcomes With Survival Analysis,” in *Predicting Post-Secondary Educational Outcomes with Survival Analysis Survival Analysis*, Jun. 2009, pp. 14.966.1–14.966.16, iSSN: 2153-5965. [Online]. Available: https://strategy.asee.org/predicting-post-secondary-educational-outcomes-with-survival-analysis

[154] S. Ameri, M. J. Fard, R. B. Chinnam, and C. K. Reddy, "Survival Analysis based Framework for Early Prediction of Student Dropouts," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management.* Indianapolis Indiana USA: ACM, Oct. 2016, pp. 903–912. [Online]. Available: https://dl.acm.org/doi/10.1145/2983323.2983351

[155] Y. Chen, A. Johri, and H. Rangwala, "Running out of STEM: a comparative study across STEM majors of college students at-risk of dropping out early," in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, ser. LAK '18. New York, NY, USA: Association for Computing Machinery, Mar. 2018, pp. 270–279. [Online]. Available: https://dl.acm.org/doi/10.1145/3170358.3170410

[156] S. Kim, M. Chang, and J. Park, "Survival analysis for Hispanic ELL students' access to postsecondary schools: discrete model or Cox regression?" *International Journal of Research & Method in Education*, vol. 41, no. 5, pp. 514–535, Oct. 2018, publisher: Routledge _eprint: https://doi.org/10.1080/1743727X.2017.1307956. [Online]. Available: https://doi.org/10.1080/1743727X.2017.1307956

[157] S. Barragán, L. González, and G. Calderón, "Modelling Student Dropout Risk Using Survival Analysis and Analytic Hierarchy Process for an Undergraduate Accounting Program," *Interchange*, vol. 53, no. 3, pp. 407–427, Dec. 2022. [Online]. Available: https://doi.org/10.1007/s10780-022-09463-7

[158] D. A. Gutierrez-Pachas, G. Garcia-Zanabria, E. Cuadros-Vargas, G. Camara-Chavez, and E. Gomez-Nieto, "Supporting Decision-Making Process on Higher Education Dropout by Analyzing Academic, Socioeconomic, and Equity Factors through Machine Learning and Survival Analysis Methods in the Latin American Context," *Education Sciences*, vol. 13, no. 2, p. 154, Feb. 2023, number: 2 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2227-7102/13/2/154

[159] E. M. Sosu and P. Pheunpha, "Trajectory of University Dropout: Investigating the Cumulative Effect of Academic Vulnerability and Proximity to Family Support," *Frontiers in Education*, vol. 4, Feb. 2019, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/articles/10.3389/feduc.2019.00006

[160] S. Cvetkovski, A. F. Jorm, and A. J. Mackinnon, "Student psychological distress and degree dropout or completion: a discrete-time, competing risks survival analysis," *Higher Education Research & Development*, vol. 37, no. 3, pp. 484–498, Apr. 2018, publisher: Routledge _eprint: https://doi.org/10.1080/07294360.2017.1404557. [Online]. Available: https://doi.org/10.1080/07294360.2017.1404557

[161] A. Kassambara, M. Kosinski, and P. Biecek, "survminer: Drawing Survival Curves using 'ggplot2'," 2021, r package version 0.4.9. [Online]. Available: https://CRAN.R-project.org/package=survminer

[162] Y. Tang, M. Horikoshi, and W. Li, "ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages." *The R Journal*, vol. 8, no. 2, pp. 474–485, 2016. [Online]. Available: https://doi.org/10.32614/RJ-2016-060

[163] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, "Survival Analysis Part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit," *British Journal of Cancer*, vol. 89, no. 4, pp. 605–611, Aug. 2003, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/6601120

[164] BBC News, "Scottish exam grades will be based on teacher estimates," *BBC News*, Apr. 2020. [Online]. Available: https://www.bbc.com/news/uk-scotland-52359692

[165] ——, "A-levels and GCSEs: U-turn as teacher estimates to be used for exam results," *BBC News*, Aug. 2020. [Online]. Available: https://www.bbc.com/news/uk-53810655

[166] ——, "SQA chief defends 'assessments' replacing Nat 5 and Higher exams," *BBC News*, Aug. 2021. [Online]. Available: https://www.bbc.com/news/uk-scotland-58089687

[167] ——, "Scottish exam pass rate drops from pandemic high," *BBC News*, Aug. 2022. [Online]. Available: https://www.bbc.com/news/uk-scotland-scotland-politics-62463451

[168] Scottish Qualifications Authority, "Assessment of National Qualifications in 2021-22," publisher: SQA. [Online]. Available: //www.sqa.org.uk/sqa/CCC_FirstPage.jsp

[169] Scottish Government, "Update on COVID regulations," Mar. 2022. [Online]. Available: http://www.gov.scot/news/update-on-covid-regulations/

[170] Office for Students, "About POLAR and Adult HE - Office for Students," Sep. 2022, archive Location: Worldwide Publisher: Office for Students. [Online]. Available: https://www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-polar-and-adult-he/

[171] ——, "About the TUNDRA area-based measures data - Office for Students," Sep. 2020, archive Location: Worldwide Publisher: Office for Students. [Online]. Available: https://www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-tundra/

[172] CACI Inc., "About Us," Oct. 2023. [Online]. Available: https://acorn.caci.co.uk/about-us/

[173] Office for Students, "Regulatory notice 1 - Access and participation plan guidance," Feb. 2019. [Online]. Available: https://www.officeforstudents.org.uk/media/0bcce522-df4b-4517-a4fd-101c2468444a/regulatory-notice-1-access-and-participation-plan-guidance.pdf

[174] UK Government, "Apply for free school meals," Oct. 2023. [Online]. Available: https://www.gov.uk/apply-free-school-meals

[175] ——, "Free school meals," Feb. 2023. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1133262/Free_school_meals.pdf

[176] Scottish Government, "Free school meals," Jun. 2021. [Online]. Available: http://www.gov.scot/news/free-school-meals-1/

[177] BBC News, "Free school meals for all Scottish primary pupils will go ahead, says deputy FM," *BBC News*, May 2023. [Online]. Available: https://www.bbc.com/news/uk-scotland-scotland-politics-65479990

[178] Abertay University, "Recruitment and Admissions Policy," Jan. 2023. [Online]. Available: https://www.abertay.ac.uk/about/the-university/recruitment-and-admissions-policy/

[179] Edinburgh Napier University, "Minimum Offers," 2024. [Online]. Available: https://www.napier.ac.uk/study-with-us/widening-participation/minimum-offers/

[180] Glasgow Caledonian University, "Contextualised Admissions Policy," Oct. 2022. [Online]. Available: https://www.gcu.ac.uk/__data/assets/word_doc/0020/37028/Contextualised-Admissions-Policy.docx#:~:text=The%20University%20currently%20gives%20consideration,Program%20for%20Schools%20(LEAPs).

[181] Glasgow School of Art, "Eligibility," 2024. [Online]. Available: https://www.gsa.ac.uk/about-gsa/key-information/our-structure/widening-participation/eligibility/

[182] Heriot-Watt University, "Fair Access Admissions Policy V2.0," 2022. [Online]. Available: https://www.hw.ac.uk/uk/services/docs/learning-teaching/policies/fairaccessadmissionspolicy.pdf

[183] ——, "Widening participation," 2024. [Online]. Available: https://www.hw.ac.uk/study/widening-participation.htm

[184] Queen Margaret University, "Admissions Policy | Queen Margaret University," 2020.

[185] Robert Gordon University, "Entry Requirements Explained | Apply | RGU," 2023. [Online]. Available: https://www.rgu.ac.uk/study/apply/entry-requirements-explained

[186] Royal Conservatoire of Scotland, "Quality Assurance Handbook," 2018. [Online]. Available: https://www.rcs.ac.uk/wp-content/uploads/2018/11/Admissions-Policy.pdf

[187] Scotland's Rural College, "Admissions Policy | Scotland's Rural College," 2022. [Online]. Available: https://www.sruc.ac.uk/media/dbohg04y/sruc-admissions-policy.pdf

[188] University of Aberdeen, "Widening access criteria | Study Here | The University of Aberdeen," 2024. [Online]. Available: https://www.abdn.ac.uk/study/undergraduate/widening-access-criteria--2848.php

[189] University of Dundee, "Contextual admissions policy | University of Dundee, UK," 2024. [Online]. Available: https://www.dundee.ac.uk/corporate-information/contextual-admissions-policy

[190] University of Edinburgh, "Widening access offers criteria," Oct. 2023. [Online]. Available: https://www.ed.ac.uk/studying/undergraduate/access-edinburgh/widening-access-offers/criteria

[191] University of Glasgow, "University of Glasgow - Study - Widening Participation - Adjusted Offers and Contextual Admissions," 2024. [Online]. Available: https://www.gla.ac.uk/study/wp/adjustedoffers/

[192] University of Highlands and Islands, "Admissions Policy (Higher Education)," 2018. [Online]. Available: https://www.uhi.ac.uk/en/t4-media/one-web/university/about-uhi/governance/policies-and-regulations/policies/Admissions-Policy-(Higher-Education).pdf

[193] ——, "Considering university? - Admission to courses," 2024. [Online]. Available: https://www.uhi.ac.uk/en/studying-at-uhi/first-steps/admissions/

[194] University of St Andrews, "Contextual admissions - Subjects - University of St Andrews," Oct. 2023. [Online]. Available: https://www.st-andrews.ac.uk/subjects/entry/contextual-admissions/

[195] University of Stirling, "2022-23 Admissions Policy | Study," 2022. [Online]. Available: https://www.stir.ac.uk/study/important-information-for-applicants/admissions-policy/2022-23-admissions-policy/

[196] University of Strathclyde, "Undergraduate Prospectus 2023 Entry | University of Strathclyde," 2023. [Online]. Available: https://www.strath.ac.uk/media/publications/prospectus/ug/2023/UG_2023_Prospectus.pdf

[197] University of the West of Scotland, “Undergraduate Entry Requirements,” Feb. 2017. [Online]. Available: https://www.uws.ac.uk/study/undergraduate/admissions-application/undergraduate-entry-requirements/

[198] Scottish Qualifications Authority, “Attainment Statistics,” Jan. 2023, publisher: SQA. [Online]. Available: https://www.sqa.org.uk/sqa/103928.html