Nurse Education in Practice 85 (2025) 104377

Contents lists available at ScienceDirect

Nurse Education in Practice

journal homepage: www.elsevier.com/locate/issn/14715953

A model for predicting student nurse attrition during pre-registration training: A retrospective observations study using routinely collected administrative data

Elizabeth Crisp [1,*], Robert Cook, Sarahjane Jones [2]

University of Staffordshire, School of Health, Policing and Sciences, Staffordshire, UK

ARTICLE INFO

Keywords:
Student nurse retention
Student nurse education
Nursing education

ABSTRACT

Aim: To explore historical student data to identify patterns predictive of attrition risk among nursing students, and hence train a predictive model of an individuals' risk of leaving the course.

Background: The World Health Organization point to an international shortage of trained nurses, which poses a risk for patient safety and care worldwide. The risk is compounded where the workforce is also aging creating additional pressures on the delivery of quality care. To stabilize the workforce, a healthy supply of newly trained registered nurses is necessary; however undergraduate nursing has one of the highest rates of student attrition (approx. 24 %).

Methods: This study follows a knowledge discovery in databases (KDD) methodology performing an observational analysis of routinely collected student data. The data (1840 students, taken from the pre-existing university business intelligence systems) was modelled for three end points; 'attrition in 1st year', 'attrition in 2nd year', and 'failure to complete'. Analysis was performed via step-wise binomial regression.

Results: Several attrition factors have been identified by the model (e.g. students who return from periods of intermittence, are Male and/or non-mature have an increased likelihood to leave).

Conclusion: To our knowledge this is the first study to examine the role of study intermittence on student attrition, or to be built on the pre-existing university business intelligence (BI) systems. The use of pre-existing university BI systems as reported here can serve as the grounding for an individual, tailored approach to retention strategy rather than an approach built on demographic assessment alone.

1. Introduction

Nursing is the largest occupational group in the health sector accounting for approximately 59 % of health professions and is experiencing a global shortage (World Health Organisation, 2020). The United Kingdom (UK) faces the additional, common, problem of an aging nursing workforce which threatens the stability of nursing stock, as does the challenge in retaining graduate nurses. During 2020 the UK experienced an increase in enrolments for preregistration nursing courses, this was approximately an increase of 20 % compared with 2019 enrolments (World Health Organisation, 2023). Whilst investment in acceleration of nursing education is a plausible approach to address workforce shortages, an equal commitment to improving the retention of nursing students would also improve the supply of graduate nurses.

The limiting factor to the supply of registered nurses in the UK is its rightful status as a registered profession, which creates the requirement for nurses to be members of the Nursing and Midwifery Council (Professional Standards Authority for Health and Social Care, 2024). To obtain registration, a new nurse must possess appropriate training, often delivered via an accredited undergraduate degree. This process hence creates a bottle neck to increasing level of new nurses; a three year lead time during which student attrition is common and far above the average for other disciplines (approx. 24 % of students nurses do not complete the course) (Health Foundation). Increasing course capacity might lead to an increase in new nursing staff, but not if any increase in student numbers comes at the cost of student experience and quality of

* Corresponding author.
E-mail address: Elizabeth.Crisp@staffs.ac.uk (E. Crisp).
[1] ORCID: 0009-0002-3256-9464.
[2] ORCID: 0000-0003-4729-4029.

https://doi.org/10.1016/j.nepr.2025.104377
Received 24 February 2025; Received in revised form 15 April 2025; Accepted 17 April 2025
Available online 25 April 2025

teaching that could easily drive-up attrition. Hence, the most promising route for increasing our cohort of new nurses will be to take steps to retain students who may otherwise have left.

As student nurse retention is a significant concern within the UK, Higher Education Institutions (HEIs) are adapting their processes, and exploring how different support strategies encourage students to complete their studies (Edge and Gladstone, 2022). Cameron et al. (2011) sought to identify student characteristics and strategies that determine retention and attrition for nursing and midwifery preregistration programmes. Through their qualitative content analysis, two broad themes emerged: programme and personal. In terms of the programme apprehensions to retention, support from personal tutors was invaluable to cope with academic demands of their studies. Similar findings were reported by Bowden (2008) and Colalillo (2007) detailing how many students often feel overwhelmed by the demands of academia and how implementing academic mentoring sessions was likely to promote retention. In addition, students from racial and ethnic minority groups who were provided with specialist academic support had increased chances of completing their studies and registering as professionals (Shavelson et al., 2018; Bekhradnia, 2004; UCAS, 2024).

Current studies have explored student nurse retention and attrition predominantly focus upon qualitative exploration of students' perceptions or evaluations of co-designed interventions. As modern HEI business practices have been driven by national and internal performance indicators (Shavelson et al., 2018), the wealth of data they collect has increased, creating a space for easy retrospective analysis via a knowledge discovery in databases (KDD) approach. Moreover, the literature is dating and there is a need for a contemporary understanding of the current challenges students experience.

This study aimed to analyse historical student data to identify patterns predictive of attrition risk among nursing students, and hence train a predictive model of an individuals' risk of leaving the course. It explored the steps needed to identify, extract and analyse existing retrospective operational data from a large higher education institute, with the expectation of disseminating the model both as key inferential patterns and as a predictive algorithm.

2. Methodology

Prior to undertaking analysis, a series of collaborator engagement meetings were held. These sessions brought together the research team with students, lecturers, senior administrative staff, and the university's business intelligence (BI) analysts. During these sessions collaborators were encouraged to voice any perceived pressures (first hand or otherwise) on student retention. Between meetings, the research and BI teams undertook a discovery phase to identify data sources already routinely collected for oversight reporting that would either map directly to, or be expected to closely proxy, the experiences and concerns voiced. These variables then formed the predictive feature space for quantitative analysis.

During collaborator engagement consultation two sets of concerns were noted; one, that the pressures faced by students would be heterogeneous across years of study, and that data richness (breadth of variables) increases the longer a student is at the university. Hence, to allow different variable spaces to be defined for different scenarios, three outcomes of analytical interest were defined:

1. Likelihood for a student to leave a nursing course within a year of enrolling at the university *(Y1 Model) [with the least available information]*
2. Likelihood for a student to leave a nursing course within two years of enrolling at the university, given they did not leave within a year of enrolling *(Y2 Model) [including Y1 variables]*
3. Likelihood a student fails to complete a nursing course, given they did not leave within two year of enrolling *(Y3* Failure to Complete *Model) [including Y1 and Y2 variables]*

A 4th end point was defined as a secondary outcome after initial analyses due to the weaker predictive performance of the 'Y1 Model' (see Fig. 2 and Table 2):

1A. Likelihood for a student to leave a nursing course within a year of enrolling at the university, given they did not leave within 6 months of enrolling *(Y1A Model) [including results from the 1st half of Y1]*

Three data sets were extracted from the University data warehouse/SITS systems:

1. Student demographics & exit status
2. Course intermission data
3. Module outcome data

Module outcome data was divided based on timing since an individual's enrolment date (0–6 months *('Y1A: Early Results')*, 0–1 years *('Y1 Results')*, and 1–2 years *('Y2 Results')* since enrolment) and limited to the highest module score where re-attempts were undertaken (NB: re-attempt marks were capped at 40 % unless extenuating circumstances were granted). For each student at each results window, an average module score (percentages) weighted by module credits (CATS; 'Credit Accumulation and Transfer Scheme' (Bekhradnia, 2004)) was calculated and then converted to a module classification (see SI-1 for details). Where a student had no completed modules within a given results window, the module classification was labelled 'NA' (Not Available) and treated as its own category, hence each result window has the data domain '1st', '2:1', '2:2', '3rd', 'Fail' and 'NA'. Intermission data was reflected in the model based on timing since enrolment date (0–1 years *('Y1 Intermission')*, and 1–2 years *('Y2 Intermission')*). Each period was reflected as a binary variable, valued one if there was any intermission in this period and zero otherwise.

Student data were re-categorised to minimise groups with low representation (see SI-2, 3 and 4 for the mappings related to 'Ethnicity', 'Nationality', and 'Highest Qualification on Entry', respectively). In addition the student's 'Age' and 'Home IMD' were converted to ordinal groups based on a combination of equal size and heuristic knowledge. 'Home IMD' was converted to three approximately equal sized groups ('1–3', '4–7', and '8–10') and 'Age' was initially converted to equal size groups (18–21, 22–28, 29 +) before moving the 21 year olds to the middle group to match the definition of 'mature' students (UCAS, 2024). Table 1 summarises the variables as prepared for the analysis.

Analysis was performed via a Kaplan-Meier estimator (Bland and Altman, 1998) to visualise the sample survival curve, followed by bi-directional stepwise binary logistic regression. Stepwise analysis was performed for each outcome, using an offset-only model as both initial, and lower bound to model scope, and the main effects model as the upper bound to model scope. The analysis used the 'step' function implemented in the R 'stats' package (R Core Team,.) which judges proposed models based on Akaike Information Criteria (AIC), for a full explanation see the related 'stepAIC' function in Venables and Ripley (2013). Each data set was randomly divided into training and validation data sets in a 3:1 ratio, with the intentions of learning the regression models on the training data set and then judging generalisable model quality via the validation data set (Browne, 2000). Model quality was measured using both training and validation data sets via the ROC-AUC method (Hamel, 2009).

The data split and subsequent analysis were repeated from multiple initial random seeds with the aim of selecting the optimal model, judged by a consistent training-validation AUC-ROC (<0.01 difference) followed by highest AUC-ROC score. Model parameters are summarised as odds ratios of coefficients and 95 % confidence intervals (CIs; calculated with assumption of asymptotic normality).

3. Results

Fig. 1 presents the estimated survival curve (Kaplan-Meier method) for students across the first 3 years of their enrolment. Student attrition

Table 1
Analysis variable summaries broken down by year of study.

Variable	Value	Y1 Summary	Y2 Summary	Y3 Summary
Age	[18,21)	31 %	30 %	29 %
Age	[21,29)	34 %	34 %	34 %
Age	[29,59]	35 %	36 %	37 %
Care Leaver		0.41 %	0.56 %	0.51 %
Carer		0.083 %	0.19 %	0.2 %
Course	BSc Adult Nursing	69 %	69 %	69 %
Course	BSc (Hons) Nursing Practice (Child)	9.8 %	8.8 %	10 %
Course	BSc Mental Health Nursing	21 %	22 %	21 %
Disability		21 %	21 %	20 %
Enrolment Year	2013	9.9 %	9.5 %	9.8 %
Enrolment Year	2014	13 %	14 %	14 %
Enrolment Year	2015	14 %	14 %	15 %
Enrolment Year	2016	14 %	14 %	15 %
Enrolment Year	2017	16 %	16 %	15 %
Enrolment Year	2018	14 %	14 %	15 %
Enrolment Year	2019	19 %	19 %	16 %
Ethnicity	Black - African/ African British	16 %	16 %	16 %
Ethnicity	Other	10 %	11 %	10 %
Ethnicity	White	74 %	73 %	74 %
Gender	Female	88 %	89 %	89 %
Gender	Male	12 %	11 %	11 %
Gender	Other	0.083 %	0.093 %	0.1 %
Home IMD	[1, 4)	38 %	38 %	38 %
Home IMD	[4, 8)	40 %	40 %	41 %
Home IMD	[8,10]	22 %	22 %	21 %
Intermittence	During Y1		5.5 %	3.4 %
Intermittence	During Y2			7.5 %
Nationality	Other	10 %	10 %	11 %
Nationality	UK National	85 %	85 %	85 %
Nationality	Zimbabwean	4.9 %	5.1 %	4.3 %
Qualification on Entry	1st Degree	12 %	13 %	12 %
Qualification on Entry	A/AS level	12 %	12 %	12 %
Qualification on Entry	BTEC National Dip	19 %	19 %	18 %
Qualification on Entry	HE access course [QAA recognized]	18 %	18 %	19 %
Qualification on Entry	Level 3 [ex. BTEC Diploma]	22 %	21 %	22 %
Qualification on Entry	Level 4–5	7.5 %	7 %	6.6 %
Qualification on Entry	Other	9.8 %	9.7 %	11 %
Y1 Result*	1st		14 %	15 %
Y1 Result*	2:1		22 %	23 %
Y1 Result*	2:2		28 %	28 %
Y1 Result*	3rd		14 %	15 %
Y1 Result*	Fail		1.8 %	0.72 %
Y1 Result*	NA		19 %	18 %
Y2 Result*	1st			16 %
Y2 Result*	2:1			27 %
Y2 Result*	2:2			28 %
Y2 Result*	3rd			21 %
Y2 Result*	Fail			3.8 %
Y2 Result*	NA			4.2 %

* 'Y1 Result' and 'Y2 Result' refer to the achieved mark within the first, and second year of study relative to the students start date and not relative to Year 1/ Year 2 modules.

appeared to be strongest at annual transitions (end of one, two and three years) with a milder acceleration in loss in the period 3–5 months after enrolment. Outside these periods loss of students was relatively steady with a slight reduction is the rate of loss in the 3rd year. The relative losses of students were 11.4 % within the 1st year of enrolment, and 8.7 % within the second year of enrolment (of those that did not withdraw within one year), with 6.4 % of students who remained enrolled for two years failing to complete.

Fig. 2 presents the ROC curves for each analysis and demonstrates the improvement in model quality as the data set expands. Notably, the Y1 model shows rather poor performance (AUC = 0.6 and poor sensitivity/specificity levels) but improves on expansion to the Y1A model (inclusion of results known after 5 months of study). The Y1A model uses data generated in the initial 6 months of study, and hence excludes student withdrawals that occurred in the period of increased loss (3–5 months post enrolment. Hence, it is unclear if improved predictive power of the Y1A model over the Y1 model is due to increased data available or the removal of 'hard to predict' cases, and it is likely both factors play a role.

Table 3 presents the non-zero regression coefficients for each scenario (tree plots of model coefficients are included in SI-5 to aid interpretation). For the Y1 model ('Likelihood for a student to leave a nursing course within a year of enrolling at the university'), student gender and age are the most significant factors. Students who identified as 'male' had an elevated likelihood compared to students who identified as either 'female' or 'other', and students in the age range 18 (inclusive) to 21 (exclusive) had an elevated likelihood of leaving compared to the older students (21+, i.e. 'mature' students). The stepwise selection method identified two further features for consideration (reduced likelihood to leave for students coming from 'A/AS levels' and increased likelihood for students who enrolled in 2013) but neither has significant evidence within this study.

The Y1A model ('Likelihood for a student to leave a nursing course within a year of enrolling at the university, given they did not leave within 6 months of enrolling') identified 'highest qualification on entry', 'enrolment year', 'care leaver' status, and 'early results in year 1' as significant factors for student loss. There was significant evidence for a reduction in likelihood for students whose highest qualification was 'A/AS levels' (as compared to any other qualification) and for students who enrolled in the period 2014–2016 (as opposed to 2013/2017–2019). There was significant evidence for an increased likelihood for students who were averaging a 'Fail' or who had no results in the system ('Y1E Results: NA'), and for individuals who had been in care before entering the university. The model identified 2 further variables for consideration though neither has significant evidence in this study; an increased likelihood to leave for 'UK National' students, and student's whose highest qualification on entry was at 'Level 3' (excluding the BTEC Diploma).

The Y2 model ('Likelihood for a student to leave a nursing course within two years of enrolling at the university, given they did not leave within a year of enrolling') identified 'gender', 'student home IMD', 'qualification on entry', 'enrolment year', 'Y1 results' and 'intermittence' as significant factors for student loss. There was significant evidence for an increased 'likelihood to leave' for students who identified as 'Male' (as opposed to any other gender), who's highest qualification on entry was of 'Level 4–5' (e.g. foundation degrees), enrolled in 2019, were averaging a '2:2' or less (or had no available module results) across the modules taken in their 1st year of enrolment, or had any period of intermittence in their first year. There was significant evidence for a reduced likelihood to leave for student from areas of high social deprivation (*IMD 1–3*) or had enrolled in 2013 (as opposed to any other year in the study).

The Y3 Failure to Complete (FTC) model ('Likelihood a student fails to complete a nursing course, given they did not leave within two year of enrolling') identified 'student nationality' 'enrolment year', 'results' and 'intermittence' as significant factors for a student's failure to complete.

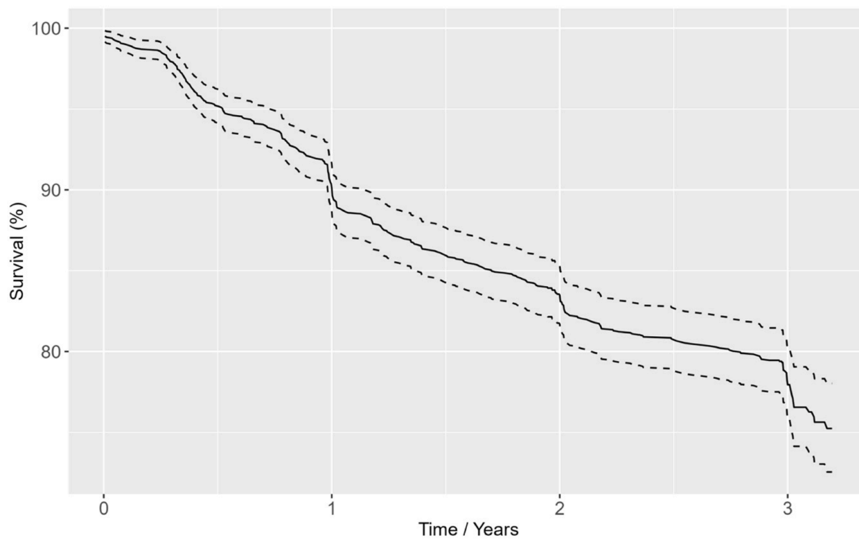

Fig. 1. Kaplan-Meier survival curve for the first 3 years from enrolment (doted lines represent the 95 % confidence interval). The curve shows step changes in survival at the one, two, and three year periods, with a smooth period of increased loss in the region 3–5 months post enrolment.

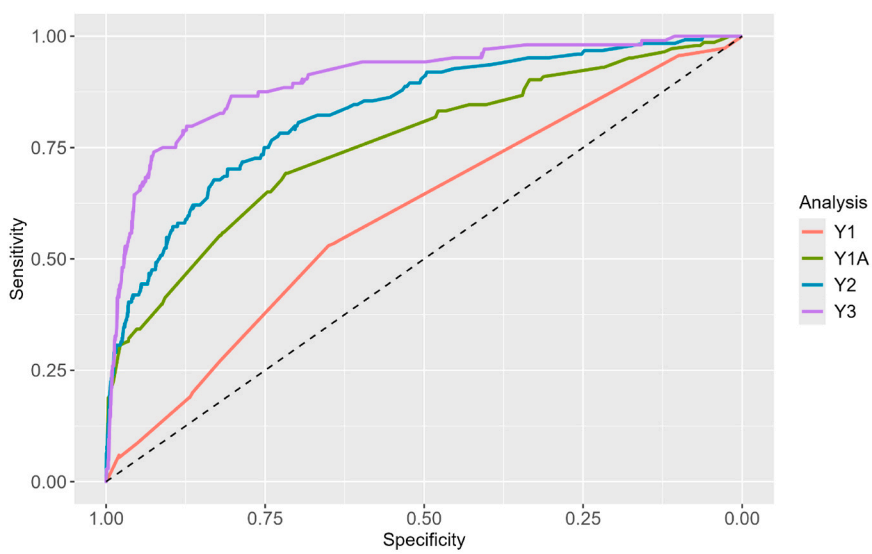

Fig. 2. Receiver Operator Characteristic (ROC) curves for each analysis. The associated area under the curve (AUC) scores for each cure are reported in Table 2, alongside an example sensitivity-specificity pair.

Table 2
Model quality summary metrics. The reported sensitivities and specificities were selected to minimize the difference between them, the applied threshold can be varied depending on the use case.

Model	AUC	Sensitivity	Specificity
Y1	0.60	0.54	0.63
Y1A	0.76	0.69	0.71
Y2	0.83	0.75	0.75
Y3	0.90	0.83	0.82

There was significant evidence for an increased ‘likelihood not to complete’ for students who enrolled in 2019 (as opposed to 2013–2018), had averaged a 3rd or less (or had no available module results) across the modules taken in their 2nd year of enrolment, or had undergone a period of intermittence within either the 1st or 2nd year of their enrolment. There was significant evidence for a reduced ‘likelihood not to complete’ for student’s who identified as ‘Zimbabwean’ (as opposed to any other nationality), had enrolled in 2017 (as opposed to 2013–2016, or 2018), had averaged a 1st/2:2 across modules taken in their 1st year of study or a 2:1 across modules taken in their 2nd year of study.

4. Discussion

Retention of nursing students in undergraduate programmes is not a novel area to research; the mixture of demand from the healthcare setting for new nurses to employ and the financial benefit to universities to retain a student body with historically high attrition rates has led to several existing studies. Of the features highlighted by this analysis, the demographic components resound with previous literature, and the role of gender, nationality, and pre-enrolment qualification have been illustrated in previous quantitative studies (Wray et al., 2012; Mulholland et al., 2008; Pryjmachuk et al., 2009). However, as we see in the ‘Y1 model’, demographic factors alone do not give strong predictive power; the pre-enrolment factors play a role in prediction but the role of post-enrolment factors (in the form of results and intermittence of study) are far stronger predictors of outcome.

Table 3
Summary of regression model coefficients.

Variable	Coefficient (O.R.)	O.R. 95 % CI	P-value	Interpretation†
Y1				
(Intercept)	0.104	[0.0807, 0.134]	< 0.001	***
Gender: Male	2.01	[1.25, 3.23]	0.004	**
Age: [18,21)	1.48	[1, 2.19]	0.048	*
Qualification on Entry: A/AS level	0.589	[0.31, 1.12]	0.105	-
Enrolment Year: 2013	1.49	[0.879, 2.54]	0.138	-
Y1A (Early Results)				
(Intercept)	0.0459	[0.023, 0.0918]	< 0.001	***
Nationality: UK National	1.71	[0.862, 3.38]	0.125	-
Qualification on Entry: A/AS level	0.242	[0.0735, 0.799]	0.02	*
Qualification on Entry: Level 3 [ex. BTEC Diploma]	1.49	[0.915, 2.43]	0.109	-
Enrolment Year: 2014	0.352	[0.166, 0.746]	0.006	*
Enrolment Year: 2015	0.246	[0.107, 0.565]	0.001	**
Enrolment Year: 2016	0.414	[0.203, 0.846]	0.016	*
Y1E Result: Fail	17.1	[9.34, 31.4]	< 0.001	***
Y1E Result: NA	2.16	[1.3, 3.59]	0.003	**
Care Leaver: Any	8.39	[1.01, 69.4]	0.048	*
Y2				
(Intercept)	0.0174	[0.00853, 0.0357]	< 0.001	***
Gender: Male	2.15	[1.1, 4.22]	0.026	*
Home IMD: [1, 4)	0.537	[0.296, 0.976]	0.041	*
Home IMD: [8,10]	1.64	[0.891, 3]	0.113	-
Qualification on Entry: Level 4–5	2.72	[1.26, 5.89]	0.011	*
Qualification on Entry: Other	2.33	[1.05, 5.17]	0.038	*
Course: BSc Mental Health Nursing	1.69	[0.978, 2.91]	0.06	-
Enrolment Year: 2013	0.22	[0.0668, 0.724]	0.013	*
Enrolment Year: 2014	0.465	[0.189, 1.15]	0.096	-
Enrolment Year: 2019	3.61	[1.97, 6.62]	< 0.001	* **
Y1 Result: 2:2	1.84	[0.9, 3.75]	0.095	-
Y1 Result: 3rd	3.39	[1.52, 7.57]	0.003	**
Y1 Result: Fail	120	[31.1, 463]	< 0.001	***
Y1 Result: NA	8.27	[3.64, 18.8]	< 0.001	***
Intermittence: During Y1	3.66	[1.71, 7.87]	0.001	**
Y3 FTC (Failure to complete)				
(Intercept)	0.0203	[0.00839, 0.049]	< 0.001	***
Gender: Male	2.03	[0.818, 5.03]	0.127	-
Nationality: Zimbabwean	0.0427	[0.0023, 0.795]	0.035	*
Course: BSc Adult Nursing	1.87	[0.93, 3.78]	0.079	-
Enrolment Year: 2017	0.366	[0.134, 0.994]	0.049	*
Enrolment Year: 2019	2.93	[1.38, 6.2]	0.005	**
Y1 Result: 1st	0.229	[0.0565, 0.927]	0.039	*
Y1 Result: 2:2	0.57	[0.281, 1.15]	0.119	-
Y2 Result: 2:1	0.385	[0.123, 1.21]	0.102	-
Y2 Result: 3rd	2.04	[0.958, 4.36]	0.064	-
Y2 Result: Fail	38	[14,103]	< 0.001	***
Y2 Result: NA	10.3	[3.87, 27.2]	< 0.001	***
Intermittence: During Y1	4.94	[1.56, 15.6]	0.007	*
Intermittence: During Y2	3.06	[1.31, 7.17]	0.01	*
Disability: Any	1.67	[0.864, 3.24]	0.127	-

†Referring to '-': 'p-value ≥ 0.05', *: 'p-value < 0.05', **: 'p-value < 0.005', ***: 'p-value < 0.0005'.

Qualitative studies have addressed person centric factors, with recurrent themes of perceived lack of physical and practical resources and perceptions of '**not being suited to be a nurse**' leading to disengagement. These themes can often be a precursor to poor academic performance, and disengagement with studies (Canzan et al., 2022). Students also discussed the challenges of organising and managing personal and family commitments, often citing poor course organisation and lack of pastoral support as reasons for taking a break in studies or leaving their programme (Mazzotta et al., 2024). Disengagement from studies is a prominent theme throughout the current evidence base, this highlights a missed opportunity to identify these students and implement tailored and timely support.

Qualitative methods utilised historically to address student retention offer deep insight into the needs and motivations of a student body and deserve deeper investigation. While lessons derived from previous qualitative studies are key, each study represents a static snapshot of a historical student cohort. The challenge this study has addressed is to understand to what extent we can perform a low resource evaluation of recent nursing cohorts to aid intervention with an active cohort, either as overarching lessons or as identification of at-risk individuals. The modern operational imperative to capture information about students, customers and service users offers a wealth of data which, when leveraged with the modern disciplines of statistical learning, pattern recognition and predictive analytics, presents an easy to implement analysis. By leveraging existing business intelligence warehouses to provide the wealth of data for modelling data collection costs are entirely removed, and instead we can focus purely on creating insight. The minimisation of resource cost of data creation offers a unique pathway for an easy return on investment so long as the inferences drawn from said data are reliable and valid.

By constructing a model atop a university's pre-existing business intelligence infrastructure, the predictive improvements offered by individual student performance become accessible and hence allows for an individual tailored approach to retention, rather than the current broad, demographic assessment. As predictive power improves, and with it the ability to identify more precise retention risk groups, retention interventions can be better focussed and hence poses a greater chance to succeed. To illustrate, consider a 100 person cohort with a 5 % attrition rate, the same resource cost can be better used if we can intervene with the 5–10 most at risk people rather than trying to deliver a broader intervention to the population. While it may seem unequal to not offer the benefit to those with a low prediction risk of attrition, it may be more equitable, and essential to optimizing nursing numbers to provide resources to our at risk groups.

5. Strengths & limitations

The initial collaborator engagement meetings identified key variables, some of which were directly mapped to the existing operational data. Some variables identified in the collaborator discussions were excluded as the data was not readily available. It is natural in an open discussion with collaborators to generate multiple hypotheses and requests that go beyond the scope of the project; however, they do identify places where practices may be improved. Notably, methods to adequately capture information on student placement for inclusion in the regression techniques.

Specific findings are limited to University of Staffordshire. However routinely collected data held on other courses or by other institutions could be explored and might provide intelligence and insight that could help retain students on courses. The proliferation of business intelligence systems across the modern HEI setting has created the pre-existing infrastructure to allow the transference of our approach across the sector.

6. Conclusion

To our knowledge this is the first study to examine the role of study intermittence on student attrition, or to be built on the pre-existing university business intelligence (BI) systems. The model has identified groups and individuals who are at risk of leaving their studies, allowing for more timely and targeted support to these individuals. The timely prediction is both unique to this model and critical in the implementation of early support interventions. In this case this has been led by a dedicated support lecturer whose role is solely to work with cohorts and individuals at risk, to provide pastoral support, build resilience and work with course teams to ensure that pastoral and academic support is seamless. The project team and wider collaborator network continue to network and engage in discussions as to how systemic changes to student support could be implemented and supported by the model developed in this study.

CRediT authorship contribution statement

Sarahjane Jones: Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Robert Cook:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elizabeth Crisp:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Patient or Public Contribution

Not applicable

Ethical consideration

Ethical approval was granted by University of Staffordshire on March 16th 2023.

Impact

It is anticipated that a tailored approach will allow a more adaptive student retention strategies, a better student experience and, due to the proliferation of BI systems across the sector, be simple to replicate in other environments.

Statistical analysis

There is a statistician on the author team (R.M. Cook).

Funding

This research was funded by Health Education England.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Elizabeth Crisp reports financial support was provided by NHS Health Education England. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.nepr.2025.104377.

References

Bekhradnia, B., 2004. Credit Accumulation and Transfer, and the Bologna Process: An Overview. Higher Education Policy Institute, Oxford.
Bland, J.M., Altman, D.G., 1998. Statistics notes: survival probabilities (the Kaplan-Meier method). BMJ. https://doi.org/10.1136/bmj.317.7172.1572.
Bowden, J., 2008. Why do nursing students who consider leaving stay on their courses? Nurse Res. https://doi.org/10.7748/nr2008.04.15.3.45.c6456.
Browne, M.W., 2000. Cross-validation methods. J. Math. Psychol. https://doi.org/10.1006/jmps.1999.1279.
Cameron, J., Roxburgh, M., Taylor, J., et al., 2011. An integrative literature review of student retention in programmes of nursing and midwifery education: why do students stay? J. Clin. Nurs. https://doi.org/10.1111/j.1365-2702.2010.03336.x.
Canzan, F., Saiani, L., Mezzalira, E., et al., 2022. Why do nursing students leave bachelor program? Findings from a qualitative descriptive study. BMC Nurs. https://doi.org/10.1186/s12912-022-00851-z.
Colalillo, G., 2007. Mentoring as a retention strategy in a diverse, multicultural, urban associate degree nursing program. Teach. Learn. Nurs. https://doi.org/10.1016/j.teln.2007.01.005.
Edge, D., Gladstone, N., 2022. Exploring support strategies for improving nursing student retention. Nurs. Stand. https://doi.org/10.7748/ns.2022.e11914.
Hamel, L., 2009. Model Assessment with ROC Curves, p.1316.
Health Foundation. Nursing Students Are Still Dropping Out in Worrying Numbers.
Mazzotta, R., Durante, A., Bressan, V., et al., 2024. Perceptions of nursing staff and students regarding attrition: a qualitative study. Int. J. Nurs. Educ. Scholarsh. https://doi.org/10.1515/ijnes-2023-0081.
Mulholland, J., Anionwu, E.N., Atkins, R., et al., 2008. Diversity, attrition and transition into nursing. J. Adv. Nurs. https://doi.org/10.1111/j.1365-2648.2008.04758.x.
Professional Standards Authority for Health and Social Care. The Nursing Midwifery Council, ⟨https://www.professionalstandards.org.uk/what-we-do/our-work-with-regulators/find-a-regulator/nursing-midwifery-council⟩ (Accessed 6th December 2024).
Pryjmachuk, S., Easton, K., Littlewood, A., 2009. Nurse education: factors associated with attrition. J. Adv. Nurs. https://doi.org/10.1111/j.1365-2648.2008.04852.x.
R Core Team. R: A Language and Environment for Statistical Computing.
Shavelson, Richard J., Zlatkin-Troitschanskaia, Olga, Mariño, Julián P., 2018. Performance indicators of learning in higher education institutions: an overview of the field. In: Anonymous Research Handbook on Quality, Performance and Accountability in Higher Education. Edward Elgar Publishing, United Kingdom, p. 249.
UCAS, 2024. Mature Undergraduate Student. https://www.ucas.com/applying/applying-university/mature-undergraduate-students (Accessed 6th December 2024).
Venables, W.N., Ripley, B.D., 2013. Modern Applied Statistics with S-Plus. Springer Science & Business Media.
World Health Organisation, 2020. State of the world's nursing: investing in education, jobs and leadership⟨https://iris.who.int/bitstream/handle/10665/331677/9789240003279-eng.pdf?sequence= 1⟩(Accessed 1 Aug 2024).
World Health Organisation, 2023, The International Year of the Nurse and the Midwife: taking stock of outcomes and commitments. ⟨https://iris.who.int/bitstream/handle/10665/375009/9789240081925-eng.pdf?sequence= 1⟩(Accessed 1st August 2024)..
Wray, J., Barrett, D., Aspland, J., et al., 2012. Staying the course: Factors influencing pre-registration nursing student progression into Year 2—a retrospective cohort study. Int. J. Nurs. Stud. https://doi.org/10.1016/j.ijnurstu.2012.06.006.