ANALYSIS OF ATTRITION AND RETENTION RATES USING THE GENERALIZED LINEAR MODEL

Temesgen Zewotir [1]
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, 4001, South Africa.
e-mail: *zewotir@ukzn.ac.za*
and
Delia North
School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, 4001, South Africa.

Key words: Dampening effect, dropout, link function, odds ratio, over-dispersion, progression, relative risk.

Summary: Improving the progression rates of students and reducing the numbers of students dropping out from institutions of higher education are critical to get maximum return for financial subsidies received or private fees paid, as well as being key components for producing skilled workers within the developing economy. Institutions of higher education in South Africa are accordingly grappling with finding a delicate balance between access, equity, redress and quality. A study of attrition is a sensitive, yet essential issue for university planning offices. An appropriate modelling approach is essential for identifying factors that contribute to attrition. This study presents two models for attrition, with slow and fast drifts of attrition, as students progress from year to year, with constant and varying dampening effects. The fast dampening model has the property of relative risk, whilst the slow drifting model has the property of odds ratio. The effect of faculty, gender, race and entry batch year, on the progression and attrition rates was examined in the study. The results of the analysis show that the first year attrition rate of White students is higher than that of the other race groups, whilst the retention and graduation rates of White students, is much greater than that of the Black and Indian students, from second year onwards. Throughout the three-year study period, the attrition rates of female students was found to be consistently lower than the corresponding attrition rates of male students.

1. Introduction

Despite the fact that a higher educational qualification is regarded as being extremely important for breaking the cycle of poverty, the opportunity of entering an institution of higher education is

[1] Corresponding author.

AMS: 62J12, 62J02, 62P99

only a reality to a small fraction of school leavers. Among this small group, not all are able to successfully complete their university education. In fact, the proportion of students who discontinue their studies on account of academic dismissal or voluntary dropout, referred to as attrition, forms a fairly substantial proportion of students that enter higher education institutions. This remains a major problem for these institutions to grapple with. Though most studies on attrition focus on the first-year level (Singell and Waddell, 2010; Zewotir, North and Murray, 2011, and references therein), there is growing concern about attrition patterns beyond the first year of study (Strauss and Volkwein, 2004; Glynn, Sauer and Miller, 2005; Hoffman and Lowitzki, 2005; Pocock, 2012).

Attrition is a concern for any type of educational or certification programme, since costs are incurred with respect to time, resources and tuition for students and institutions (Bennett, 2003; Schneider and Yin, 2011). The magnitude of attrition rates, or its counterpart, the progression rates of students, that occurs at higher education institutions, around the country, provide an important indicator of the overall operational efficiency of the learning environment. Student success is enhanced by a positive learning environment which is able to meet the basic psychological needs of students, promote intrinsic motivation and optimize learning outcomes such as knowledge transfer, meta-cognition, and engagement (Copeland and Levesque-Bristol, 2010). It is not only what happens to a student after arriving at a university that determines their progression or attrition, we thus need to include an appropriate set of student specific covariates in the attrition/progression rates model (Baumgart and Johnstone, 1977; Sagy, 2000; Danaher, Bowser and Somasundaram, 2008). The basic elements that tend to define student attrition in a programme include characteristics of the students, as well as the characteristics of the programme (Tinto, 1975; Zewotir and North, 2007).

The student characteristics include socio-demographic attributes (race, gender, etc.) as well as student's coping and studying skills (Tinto, 1975; Hirschy, Bremer and Castellano, 2011). The characteristics of a programme that can limit or facilitate the development and integration of individuals within the programme include the resources, facilities, structural and organizational arrangements, as well as the members of the programme (Yukeselturk and Inan, 2006; Zewotir and North, 2007). Several programme-related reasons that have been cited for students discontinuing with a course of study, include course schedule and pacing, insufficient feedback, quality of learning materials, lack of interaction between students and instructors, inexperienced instructors, lack of social integration and lack of student support. Students' attrition may further have implications for mismatch that could occur between the academic and social environments of the institution, and the expectations of the student (North and Zewotir, 2006; Pocock, 2012).

In any given study of attrition rates, one should expect to see a dampening effect of attrition occur as each entry batch of students progress from one year to the next, since the probability of leaving the programme generally diminishes as the student progresses through the system from year to year. Such a dampening phenomenon of attrition can be attributed to the fact that as each student cohort group progresses from one year to the next, the students generally become more experienced, having completed more subjects and further become more motivated, as their date of final qualification gets closer than what it was the previous year. The identification of factors that help to determine these dampening rates of attrition therefore forms an important component in understanding how a student cohort group progresses from one year to the next through an educational system.

Further key features of student throughput that university planners need pay attention to are retention, progression and graduation rates, where graduation rate is defined as the percentage of an entering batch that graduates within three years with a baccalaureate degree, whilst the retention rate is a year-by-year successful completion rate, including graduation rate.

Besides the model discussed being useful for identifying the effect of mentioned factors on attrition/retention of students at the institution the model is further useful for budget planners and administrators to predict the retention, progression and graduation rates of students, which would assist in determining the number of new students to admit into a programme.

The purpose of this paper is to develop an attrition model, with a dampening effect which is suitable to identify the effect of related factors and covariates on the attrition of students at an institution of higher education. The application of the model was demonstrated by assessing the effect of race, gender, faculty and entry batch, on the attrition and retention of students entering the various three-year, non-professional degree programmes in Faculties of Science and Humanities, at the University of KwaZulu-Natal (UKZN), between 2006 and 2008.

2. Data description

The number of students who enrolled and dropped out over the study period, along with the socio-demographic characteristics and classifications as supplied by the UKZN Data Management and Information (DMI) office database, formed the data for this study. The cohorts of entry year batches of students from 2006 to 2008, were tracked over a 3-year period (the desired length of the degree programme), by tracing the number of first, second and third year students enrolled in undergraduate programmes, at the beginning of each year, in the Faculties of Science and Humanities. Likewise, the number of first, second and third year students who had dropped out of programmes either on a voluntary basis, or due to academic exclusions, by the end of the year, were recorded as attritions.

For each entry year batch of students (i.e. those starting in 2006 to 2008), year-on-year registrations and dropouts were classified by faculty, race and gender. The 2005 entry year cohort was deliberately excluded from this study as this cohort of students began their studies during the time of greatest change in the University's history, namely, the physical merger between the University of Natal and the University of Durban-Westville. This change in the system and the subsequent necessary re-orientation and dealing with "pipeline students", which continued throughout 2005, may accordingly have an undue influence on a study of attrition (Pocock, 2012).

Race was categorized into three levels: Black, White and Indian. Gender was categorized as being male or female. Faculty was categorized into Science and Humanities. Of all the faculties at UKZN, these two faculties were chosen as the faculties of focus, due to the similarly structured three-year-degree programmes in these faculties. Other faculties at UKZN, that are not part of this study, have many professional certification and degree structures (e.g., Nursing in the Faculty of Health Science, Accounting in the Faculty of Management Studies), as well as mainline professional degree structures, with four or more year programmes (e.g., the Faculties of Medicine and Engineering).

Table 1 displays the attrition rates that were recorded by the different cohorts (batches) groups and Faculty, over the three-year study period. It is noted that as the students progress from year to year, the attrition rates decrease, for both faculties and for each of the entering cohort groups. Also note from Table 1 that at each entry batch year, attrition rates in the Science Faculty are consistently higher than those in the Humanities Faculty at first and second years.

Table 1: Observed attrition rates by Entry year batch, Year and Faculty.

Faculty	Entry year batch	Year		
		1st	2nd	3rd
Humanities	2006	0.158	0.0861	0.0643
	2007	0.1394	0.0735	0.061
	2008	0.1361	0.0692	0.0561
Science	2006	0.1729	0.099	0.0761
	2007	0.1691	0.1139	0.056
	2008	0.1483	0.0892	0.0415

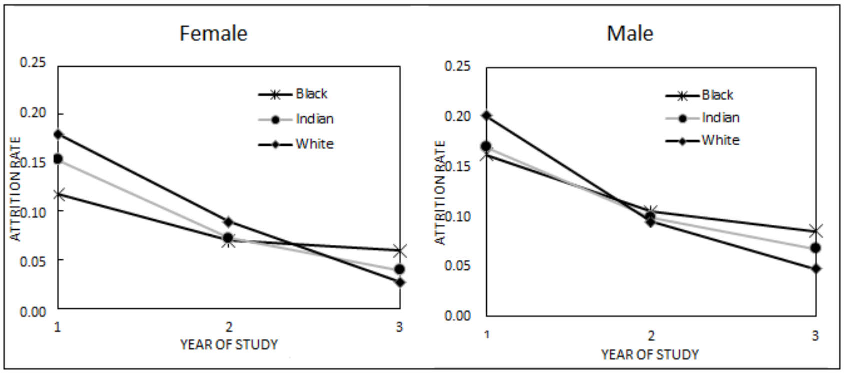

Figure 1: Observed attrition rates by gender and race.

Figure 1 depicts that the first year attrition rates are higher for White students than for other races, irrespective of gender. For White students, the attrition rate at the final (third) year is lower than the corresponding attrition rates for any of the other races. This is further true both for males and females. However the attrition rates of females is consistently lower than that of males.

3. The model

Suppose that each student is required to successfully complete K years of study before being allowed to graduate from a given programme. Some students may decide to leave a particular programme, whereas others, with an otherwise equivalent academic background, may choose to re-enter the programme. It will be the purpose of this study to follow the progression of a given cohort of students, noting that the data requirements for such a study will need to focus on the total number of students that are enrolled for each year (i.e. the new student intake, plus those that were carried over from the previous year) and the number of attritions that occurred at the end of each year. The notations presented in Table 2 will be used as the basis for our student progression data modelling.

In terms of notation introduced in Table 2, it clearly follows that $b_i = N_i - a_i$ and $p_i = \frac{a_i}{N_i}$ for all $i \in \{1, 2, \ldots, K\}$. The model is developed based on the assumption that as we progress from one

Table 2: A layout and notation for the attrition data.

Period (year)	1	2	...	K
Number of students enrolled	N_1	N_2	...	N_K
Number of attritions (students)	a_1	a_2	...	a_K
Number of successful completions (students)	b_1	b_2	...	b_K
Attrition rate	p_1	p_2	...	p_K

semester/year to the next there is a dampening effect on the rates of attrition. That is, one would expect to see

$$E(p_1) \geq E(p_2) \geq \ldots \geq E(p_K).$$

Guttman and Olkin (1989) have chosen to model the process outlined in Table 2, by using a parameter $\pi, (0 \leq \pi \leq 1)$ to denote an initial probability of attrition for the first year of study, for a typical student. They also use a parameter value $\rho, (0 \leq \rho \leq 1)$, to represent a dampening effect for each future period of study. It follows from their study that the probability for attrition during the second semester of study is given by $\pi\rho$, the probability for attrition during the third semester is correspondingly given by $\pi\rho^2$, and so on. The model of Guttman and Olkin (1989), for a given cohort, thus gives an expected probability of the attrition rate year i, as

$$E(p_i) = \pi\rho^{i-1}, \quad i = 1, 2, ..., K. \tag{1}$$

Estimating the non-linear parameters, initial attrition (π) and the dampening effect (ρ), subject to the additional constraint that both π and ρ should lie between 0 and 1, are statistically intractable. Moreover, the model (Equation (1)) is not suitable for the inclusion of explanatory factors such as gender, faculty and race into the model structure. This paper resolves these shortcomings by using a flexible generalized linear model approach that allows the inclusion of the explanatory factors, as a covariate vector, **x**.

A plot of the attrition rate against year of study, suggests that an exponential model of the form

$$E[p_i(\mathbf{x})] = e^{\beta_0(\mathbf{x}) + (i-1) \times \beta_1(\mathbf{x})}, \quad i = 1, 2, \ldots, K$$

should be fitted to the data where, $\exp(\beta_0(\mathbf{x}))$ denotes an initial attrition rate and $\exp(\beta_1(\mathbf{x}))$ denotes the relative dampening of attrition as the students with covariate set **x** progress from year to year after the first year. If $\beta_1(\mathbf{x})$ is close to 0, then $\exp[\beta_1(\mathbf{x})]$ is close to 1 and the covariate vector **x** will accordingly have a negligible dampening on the initial attrition rate as students progress from year to year. Should $\beta_1(\mathbf{x})$ take on a negative value, then this reflects that students associated with covariate set **x**, have a slower dampening of attrition rate, beyond the first year. In situations where $\beta_1(\mathbf{x})$ is positive, then the attrition rate diminishes faster for students with covariate set **x**, as they progress from year to year. The alternative interpretation of $\exp[\beta_1(\mathbf{x})]$ is considering it as the "relative risk" (RR) of attrition as the students progress from year to year.

If one assumes that the observations are independently generated by a binomial distribution, then one can derive maximum likelihood estimators for $\beta_0(\mathbf{x})$ and $\beta_1(\mathbf{x})$, using a computer-aided, iterative method (Demidovich and Maron, 1987). Since the binomial distribution is a member of the Exponential family of distributions (Agresti, 2002; McCullagh and Nelder, 1989), the dampening effect, given in the exponential form, can further be derived using a generalized linear model, with the following log link function:

$$\log\left(E[p_i(\mathbf{x})]\right) = \beta_0(\mathbf{x}) + (i-1)\times\beta_1(\mathbf{x}), \quad i = 1,2,\ldots,K. \tag{2}$$

From Equation (2) the initial attrition rate is $\exp(\beta_0(\mathbf{x}))$, and equivalently $[1-\exp(\beta_0(\mathbf{x})]$ is the probability that a student, with a covariate profile $\mathbf{x}$, will progress from first year to second year. Likewise the dampening effect of attrition as students progress from year to year is $\exp(\beta_1(\mathbf{x}))$.

A hypothesis of interest that one may want to test involves determining whether the dampening effect and/or initial attrition rate differs from one subgroup (for example faculty, gender or race) to another. In statistical terms, $H_0 : \beta_0(\mathbf{x}) = \beta_0$ for all $\mathbf{x}$ and/or $H_0 : \beta_1(\mathbf{x}) = \beta_1$ for all $\mathbf{x}$. These hypotheses can be tested using the contrast statements that are available in a typical generalized linear model analysis package. If the attrition rate is assumed to drift more slowly downward from an initial value, then a generalized linear model with the following logistic link function may be more appropriate:

$$\log\left(\frac{E[p_i(\mathbf{x})]}{1-E[p_i(\mathbf{x})]}\right) = \beta_0(\mathbf{x}) + (i-1)\times\beta_1(\mathbf{x}), \quad i = 1,2,\ldots,K \tag{3}$$

which implies that

$$E[p_i(\mathbf{x})] = \frac{e^{\beta_0(\mathbf{x})+(i-1)\times\beta_1(\mathbf{x})}}{1+e^{\beta_0(\mathbf{x})+(i-1)\times\beta_1(\mathbf{x})}}, \quad i = 1,2,\ldots,K.$$

When compared with model Equation (2), model Equation (3) will have a slower dampening effect with an initial rate of attrition being given by $\frac{e^{\beta_0(\mathbf{x})}}{1+e^{\beta_0(\mathbf{x})}}$. The generalized linear model, Equation (3), is commonly referred to as logistic regression. Logistic regression analysis is commonly available in standard statistical packages. The interpretation of $\exp[\beta_1(\mathbf{x})]$ now needs to be given in terms of an odds ratio attrition, as the students with profile vector $\mathbf{x}$, progress from year to year.

The models given in Equations (2) and (3), make use of a constant covariate dependent dampening effect $\beta_1(\mathbf{x})$, for each of the K years that are required to complete a given degree. A further model that has an attrition rate effect that decreases by a factor $\beta_1(\mathbf{x})$, during year 2 and then continues to decrease by an additional factor $\beta_2(\mathbf{x})$, during year 3 can be created by introducing the following two variables into the model structure

$$t_{1i} = \begin{cases} 0, & \text{if } i = 1 \\ 1, & \text{if } i \neq 1 \end{cases}$$

and

$$t_{2i} = \begin{cases} 0, & \text{if } i = 1,2 \\ i-2, & \text{otherwise.} \end{cases}$$

The model then becomes

$$\eta\left(E[p_i(\mathbf{x})]\right) = \beta_0(\mathbf{x}) + t_{1i}\beta_1(\mathbf{x}) + t_{2i}\beta_2(\mathbf{x}), \qquad i = 1,2,\ldots,K \tag{4}$$

where $\eta(\cdot)$ denotes an appropriately chosen link function, either a log or logit function. If $\beta_1(\mathbf{x}) = \beta_2(\mathbf{x})$, then a constant dampening effect for attrition over the years, similar to the situation modelled in Equations (2) or (3), will apply.

In the application of the GLM with binomial distribution and link functions either log or logit, the deviance and Pearson Chi-square divided by the degrees of freedom must be used to detect over-dispersion or under-dispersion in the binomial distribution. Values greater than 1 indicate over-dispersion, that is, the true variance of the response variable is greater than what it should be under the given model. If this happens, the resulting estimates are consistent, but estimates of the variance are not, since the over-dispersion or under-dispersion can result in spuriously small or large standard errors of the estimates (Barron, 1992). This inconsistent variance estimation invalidates any hypothesis testing. The most widely implemented approach to remedy this is the use of quasi-likelihood, which overcomes the problem of over and under-dispersion as discussed. This adjustment further provides valid inference, guarding against drawing of incorrect conclusions, discussed (Allison, 1999). The interpretation of the parameter estimates in the quasi-likelihood approach, remains as discussed above.

4. Results

A generalised linear model (GLM) analysis was carried out to investigate the effect of gender, race, cohort batch and faculty on both students initial attrition rate and the dampening of the attrition rate as students' progress from year to year. All the analyses were performed using SAS procedures (GENMOD procedure). The Pearson Chi-square and deviance divided by the degrees of freedom were used to determine for over/under-dispersion in the fitted model. The results indicated evidence of over-dispersion (Table 3). This evidence of over-dispersion indicates an inadequate fit of the ordinary GLM model. The model was refitted by adjusting for over-dispersion, using the quasi-likelihood approach. In the ordinary GLM model, there was no allowance for over-dispersion, whilst in the quasi-likelihood approach, adjustment for over-dispersion was employed. For adjusted models the values for Pearson Chi-square and deviance were sufficiently close to 1 for both links (Table 3).

Table 3: Assessment of over-dispersion in the fit.

Approach		Log Link			Logit Link		
	Criterion	DF	Value	Value/DF	DF	Value	Value/DF
Ordinary GLM	Deviance	94	207.1192	2.2034	94	206.2797	2.1945
	Pearson Chi-Square		203.0989	2.1606		202.8875	2.1584
Quasi-Likelihood	Deviance	94	94.0000	1.0000	94	94.0000	1.0000
	Pearson Chi-Square		92.1754	0.9806		92.4542	0.9836

The results from the quasi-likelihood estimates are presented in Table 4. From the parameter estimates presented in Table 4, the initial attrition rate estimates, i.e. the $\exp(estimate)$ column, of model Equations (2) and (3) are identical up to the thousandth decimal point. The standard errors of the log link estimate however are smaller than their corresponding logit link estimates. With regard to the dampening effect, parameter estimates for both models produced the same result. The overall first year attrition rate was 18.97%, with an overall attrition dampening multiplier of 0.4588, for every year the students' progress after the first year. That is, as the level of study increases by a year, from first to second year, or from second to third year, the attrition rate decreases by 54.12%. As the dampening effect gets closer to 1, the attrition rate remains constant throughout the study period.

The interpretation of $\exp(estimate)$ is that the outcome for a given category is given by $\exp(estimate)$ times that of the reference category. The complementary interpretation is the sign of the *estimates*: a positive coefficient shows the increase of the outcome of interest relative to its reference category, whilst a negative coefficient shows the decrease of the outcome of interest relative to its reference category.

At 5% level of significance, the initial attrition rate is not significantly different for the two faculties, or for the three entry batches. However the initial (i.e., first year) attrition rate is significantly different for the different racial groups, and also for the different gender groups. In particular, the attrition of first year students that were of the Black race group is $28.10\%[(1-0.7190)\times 100\%]$ lower than the corresponding attrition rate for first year students of the White race group. Likewise, the attrition rate of first year Indian students is 16.77% lower than the first year attrition rate of the White students. The difference between attrition rates of first year students that are of Black and Indian race is also significant (p = 0.0341). After first year, however, the dampening rate of attrition for Black and Indian students is slower than that of the White students. In other words, once students progress from first year, the attrition rate of White students diminishes faster than the corresponding figure for Black and Indian students. The attrition rate of Indian students diminishes at a higher rate than that of Black students, after the first year of study.

With regard to gender, the attrition rate of first year female students is significantly lower than that of the male students, but that their progression rate is identical after the first year of study.

The basic assumption in Table 4 was that of a constant dampening effect throughout the study period. However, the attrition rate seems to dampen at a higher rate from first year to second year, than from a second year to third year of study. This suggests that a piecewise model, such as the one given in Equation (4), might provide a more appropriate fit. Accordingly, Table 5 presents the results of an analysis using the piecewise approach given in model Equation (4). The overall first year attrition rate logit link estimate is slightly higher than the estimate obtained from the log link. Otherwise, the test results are identical for both links. At second year level, the overall attrition dampening multiplier from the first year is 0.4431 for the log link and 0.3925 for the logit link. The results further show that none of the factors have a significant effect on the dampening of second year attrition rates. All the categories were found to have an identical dampening multiplier.

The overall attrition dampening multiplier from the second year to the third year level is 0.4655 for the log link and 0.4447 for the logit link. It is important to compare the overall dampening effect at second and third year levels. The t-test supports identical dampening at second year and third year levels (p = 0.4423). This test favours the goodness-of-fit of the constant dampening effect model fit presented in Table 4. Moreover, the deviance difference between the model fit results in Tables 4 and 5 is 20.4116. When this deviance difference divided by the difference in the number of parameters in the two models is compared with the hypothetical chi-square distribution with 7 degrees of freedom, it is found to be insignificant (p = 0.8926). Consequently, the constant dampening effect goodness-of-fit is better than the piecewise model fitness.

Table 4: Over-dispersion adjusted estimates of initial attrition and dampening effect.

Initial attrition rate							
		Log link			Logit link		
Parameter	Level	Estimate	St Err	Exp(estimate)	Estimate	St Err	Exp(estimate)
Overall		-1.6625*	0.0966	0.1897*	-1.4511*	0.1156	0.1898*#
Entry batch (ref=2008)							
Year	2006	0.1047	0.0736	1.1104	0.1258	0.0863	1.1104
Year	2007	0.0701	0.0729	1.0726	0.0808	0.0852	1.0726
Gender	Female	-0.1944*	0.0618	0.8233*	-0.2307*	0.0723	0.8233*
Race (ref=White)							
	Black	-0.3299*	0.0824	0.7190*	-0.4012*	0.1000	0.7190*
	Indian	-0.1836*	0.0902	0.8323*	0.2286*	0.1096	0.8323*
Faculty	Science	0.1026	0.0620	1.1080	0.1312	0.0727	1.1080
Dampening Effect							
Overall		-0.7792*	0.1142	0.4588*	-0.8731*	0.1263	0.4588*
Entry batch (ref=2008)							
Year	2006	0.1094	0.0794	1.1156	0.1102	0.0879	1.1156
Year	2007	0.0575	0.0800	1.0592	0.0580	0.0882	1.0592
Gender	Female	-0.1211	0.0667	0.8859	-0.1191	0.0737	0.8859
Race (ref=White)							
	Black	0.4469*	0.0997	1.5635*	0.5114*	0.1109	1.5635*
	Indian	0.2171*	0.1092	1.2425*	0.2546*	0.1214	1.2425*
Faculty	Science	-0.0585	0.0673	0.9432	-0.0817	0.0747	0.9432
$*$ significant at 5% level, and # is obtained using $\frac{\exp(estimate)}{1+\exp(estimate)}$							

Table 5: Quasi-likelihood piecewise attrition and dampening effect estimate.

		Log link			Logit link		
		First year attrition rate					
		Estimate	St Err	Exp(estimate)	Estimate	St Err	Exp(estimate)
Overall		-1.6546*	0.0983	0.1912*	-1.4385*	0.1182	0.1918*#
Entry batch (ref=2008)							
Year	2006	0.1166	0.0748	1.1237	0.1409	0.0883	1.1513
Year	2007	0.0648	0.0744	1.0669	0.0744	0.0872	1.0772
Gender	Female	-0.2048*	0.0629	0.8148*	-0.2447*	0.0741	0.7829*
Race (ref=White)							
	Black	-0.2965*	0.0842	0.7434*	-0.3610*	0.1024	0.6970*
	Indian	-0.1595	0.0921	0.8526	-0.1973	0.1123	0.8209
Faculty	Science	0.0663	0.0631	1.0685	0.0851	0.0745	1.0888
Second year dampening effect							
Overall		-0.8139*	0.1868	0.4431*	-0.9353*	0.2101	0.3925*
Entry batch (ref=2008)							
Year	2006	0.0413	0.1393	1.0422	0.0342	0.1558	1.0348
Year	2007	0.0839	0.1376	1.0875	0.0874	0.1535	1.0913
Gender	Female	-0.0630	0.1167	0.9389	-0.0498	0.1303	0.9514
Race (ref=White)							
	Black	0.2602	0.1601	1.2972	0.3158	0.1815	1.3714
	Indian	0.0722	0.1750	1.0749	0.0982	0.1982	1.1032
Faculty	Science	0.1474	0.1166	1.1588	0.1495	0.1306	1.1613
Third year dampening effect							
Overall		-0.7646*	0.2822	0.4655*	-0.8103*	0.2987	0.4447*
Entry batch (ref=2008)							
Year	2006	0.1980	0.1869	1.2190	0.2037	0.2012	1.2259
Year	2007	0.0165	0.1893	1.0166	0.0141	0.2029	1.0142
Gender	Female	-0.2001	0.1569	0.8186	-0.2101	0.1688	0.8105
Race (ref=White)							
	Black	0.7333*	0.2496	2.0819*	0.7856*	0.2647	2.1937*
	Indian	0.4581	0.2718	1.5811	0.4894	0.2879	1.6313
Faculty	Science	-0.3529*	0.1600	0.7026*	-0.3952*	0.1725	0.6735*
$*$ significant at 5% level, and # is obtained using $\frac{\exp(estimate)}{1+\exp(estimate)}$							

5. Conclusion

The models adopted in this study are flexible enough to include a downward drift in the attrition rate as a student progresses from one year to the next. The log link function allows one to fit an attrition rate that can drift down more rapidly when compared with a model that uses a logit link function. The variables, t_{1i} and t_{2i}, introduced in the models, allow one to model an attrition dampening effect after the second year of study. It should be noted, however, that this break point can easily be moved to any particular year by appropriately recoding t_1 and t_2.

For the UKZN data analysed, the constant dampening effect throughout the study period was found to be the best fit. The log and logit links dampening effect produced identical results. In order to understand why, it is important to realise that the magnitudes of the odds ratio is approximately equal to the relative risk, if the probability of the outcome of interest is close to zero, for both groups (Agresti, 2002; Lachin, 2011).

The exclusion of interaction terms was tested by first including all the main effects into the model and then evaluating whether any interaction term, one-at-a-time, needed to be included in the model. None of the interaction effects were found to be significant. The final model for an analysis of the data obtained from the University of KwaZulu-Natal, showed that the rate of attrition in the first year of study does not differ significantly between the Science and Humanities faculties. Black students' attrition rate is lower at first year than other races, but this group experiences a higher rate of attrition as the level of study progresses. On the contrary, White students experience a relatively higher first year attrition rate, which highly diminishes as the level of study advances. Throughout the three year study period, female students' retention rate is consistently better than that of males. The insignificance of faculty and entry batch year on effect of the first year attrition rate, or the attrition dampening after first year, might be a reflection of the effect of resources, facilities, structural/organizational arrangements, etc. on the development and integration of incoming stability in the programmes of the university.

This article aimed to make a contribution to the study of attrition/retention by simplifying the complexity in the modelling. The list of factors that used in the study is by no means exhaustive, but were restricted by the availability of information from the university database. The proposed model capability and novelty is, however, well demonstrated by using the available data.

The fact that the model demonstrated that students in the study of different races had a vastly different initial attrition rate, with inverse dampening effects, alerts one to the need for investigating whether this effect is due to voluntary drop-out or academic exclusion. The future direction of this research is therefore to extend this study to the competing risk approach, with three possible outcomes: voluntary drop-out, academic exclusion and progression.

6. Acknowledgement

The anonymous reviewers and the associate editor are gratefully acknowledged for the constructive comments which have helped the authors to improve the paper. The authors are also grateful to Prof Michael Murray for reading and commenting on the first draft. This work was supported by the teaching-learning competitive grant from UKZN. The authors are grateful for this support.

References

AGRESTI, A. (2002). *Categorical Data Analysis*. Second edition. John Wiley and Sons: New York.

ALLISON, P. D. (1999). *Logistic Regression Using the SAS System: Theory and Application*. SAS Institute Inc.: Cary NC.

BARRON, D. N. (1992). The analysis of count data: over-dispersion and autocorrelation. *Sociological Methodology*, **22**, 179–220.

BAUMGART, N. L. AND JOHNSTONE, J. N. (1977). Attrition at an Australian university: A case study. *Journal of Higher Education*, **48**, 553–570.

BENNETT, R. (2003). Determinants of undergraduate student drop out rates in a university business studies department. *Journal of Further and Higher Education*, **27** (2), 123–141.

COPELAND, K. J. AND LEVESQUE-BRISTOL, C. (2010). The retention dilemma: Effectively reaching the first-year university student. *Journal of College Student Retention*, **12** (4), 485–515.

DANAHER, P. A., BOWSER, D., AND SOMASUNDARAM, J. (2008). The student departure puzzle: Do some faculties and programs have answers? *Higher Education Research and Development*, **27** (3), 271–280.

DEMIDOVICH, B. P. AND MARON, I. A. (1987). *Computational Mathematics*. Fourth edition. Mir Publisher: Moscow.

GLYNN, J. G., SAUER, P. L., AND MILLER, T. E. (2005). Configural invariance of a model of student attrition. *Journal of College Student Retention*, **7** (3), 263–281.

GUTTMAN, I. AND OLKIN, I. (1989). Retention and attrition models. *Journal of Educational Statistics*, **14** (1), 1–20.

HIRSCHY, A. S., BREMER, C. D., AND CASTELLANO, M. (2011). Career and technical education (CTE) student success in community colleges. *Community College Review*, **39** (3), 296–318.

HOFFMAN, J. L. AND LOWITZKI, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. *Review of Higher Education*, **28** (4), 455–474. Available at `http://dx.doi.org/10.4102/sajs.v108i3/4.634`.

LACHIN, J. M. (2011). *Biostatistics Methods: The Assessment of Relative Risks*. John Wiley and Sons: New York.

MCCULLAGH, P. AND NELDER, J. A. (1989). *Generalized Linear Models*. Second edition. Chapman and Hall: New York.

NORTH, D. AND ZEWOTIR, T. (2006). Teaching statistics to social science students: Making it valuable. *South African Journal of Higher Education*, **20** (4), 503–514.

POCOCK, J. (2012). Leaving rates and reasons for leaving in an Engineering faculty in South Africa: A case study. *South African Journal of Science*, **108** (3/4), 8 pages. Available at `http://dx.doi.org/10.4102/sajs.v108i3/4.634`.

SAGY, S. (2000). Factors influencing early dropout: The case of Russian immigrant students attending an Israeli university. *Journal of Applied Behavioral Science*, **36** (3), 362–375.

SCHNEIDER, M. AND YIN, L. (2011). The Hidden Costs of Community Colleges. American Institute for Research: Washington, DC. Available at `http://www.air.org/sites/default/files/downloads/report/Oct2011_0.pdf`.

SINGELL, L. D. AND WADDELL, G. R. (2010). Modeling retention at a large public university: Can at-risk students be identified early enough to treat? *Research in Higher Education*, **51** (6), 546–572.

STRAUSS, L. C. AND VOLKWEIN, J. F. (2004). Predictors of student commitment at two-year and four-year institutions. *Journal of Higher Education*, **75** (2), 203–227.

TINTO, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, **45** (1), 89–125.

YUKESELTURK, E. AND INAN, F. (2006). Examining the factors affecting student dropout. *The Turkish Online Journal of Distance Education*, **7**. Available at `http://tojde.anadolu.edu.tr/tojde23/pdf/article_6.pdf`.

ZEWOTIR, T. AND NORTH, D. (2007). Focus on the statistical education of prospective engineers in South Africa. *Pythagoras*, **65**, 18–23.

ZEWOTIR, T., NORTH, D., AND MURRAY, M. (2011). Student success in entry level modules at the University of KwaZulu-Natal. *South African Journal of Higher Education*, **25** (6), 1233–1244.

Manuscript received, 2014-01-10, revised, 2015-03-25, accepted, 2015-04-27.