Nurse Education Today (2008) **28**, 469–475

ELSEVIER

Nurse Education Today

intl.elsevierhealth.com/journals/nedt

Predicting who will drop out of nursing courses: A machine learning exercise

Laurence G. Moseley *, Donna M. Mead

HESAS, Glyntaff Campus, University of Glamorgan, Portypridd CF37 1DL, UK

Accepted 31 July 2007

KEYWORDS
Student drop outs;
Machine learning;
Data mining;
Prediction

Summary
Introduction: The concepts of causation and prediction are different, and have different implications for practice. This distinction is applied here to studies of the problem of student attrition (although it is more widely applicable).
Background: Studies of attrition from nursing courses have tended to concentrate on causation, trying, largely unsuccessfully, to elicit what causes drop out. However, the problem may more fruitfully be cast in terms of predicting who is likely to drop out.
Methods: One powerful method for attempting to make predictions is rule induction. This paper reports the use of the Answer Tree package from SPSS for that purpose.
Data: The main data set consisted of 3978 records on 528 nursing students, split into a training set and a test set. The source was standard university student records.
Results: The method obtained 84% sensitivity, 70% specificity, and 94% accuracy on previously unseen cases.
Discussion: The method requires large amounts of high quality data. When such data are available, rule induction offers a way to reduce attrition. It would be desirable to compare its results with those of predictions made by tutors using more informal conventional methods.

Background

In recent years, the number of students entering higher education institutions to study nursing has shown a substantial increase. Selected examples from data provided by an ad hoc query by HESA (personal communication) show that the number of students studying nursing (HESA code B7) had risen (see Table 1).

The picture can be grasped more simply by presenting the whole series of HESA data graphically, as in Fig. 1.

These figures show a remarkable growth in nurse training, with annual growth rates in the early years ranging from 3.1% to 10.2%, falling back to 1.5% in the most recent year — but still rising nonetheless.

* Corresponding author. Tel.: +44 1594 826 354.
E-mail address: LGMoseley@btinternet.com (L.G. Moseley).

0260-6917/$ - see front matter
doi:10.1016/j.nedt.2007.07.012

Table 1 Numbers of nurses in training 1998/9 to 2005/6

1998/99	138,365
2002/03	176,500
2005/06	194,335

The cumulative effect has been dramatic: a 40% rise over 7 years in the number of nurses being trained in higher education institutions.

One might have expected such an increase to place a strain on nursing education institutions, and on those students going through the system. A possible outcome of that strain would have been an increase in the numbers who fail to complete their courses — a phenomenon often referred to as attrition.

Attrition from nursing courses in the UK is important. It matters to the students and their families, to the higher education institutions which train them, to the funding bodies, to the future employers, especially the NHS, and to patients who because of attrition will not receive as much skilled care as would have otherwise been obtainable. These undesirable outcomes give reasons for wishing to reduce the level of attrition. What is that level?

It has at times proved difficult to find reliable data on the matter. In the UK the relevant Minister refused a parliamentary debate on the issue in 2006, but the RCN (Royal College of Nursing) instituted a query under the Freedom of Information Act, and extracted an estimate in 2006. The reported overall figure was that almost 25% of nursing students withdrew before qualifying (Nursing Standard, 2006, p. 5). The cost of attrition was £57 million per annum. Although there are substantial variations between institutions in such rates (from 3% to 65%), it is clear that the problem is sufficiently serious to merit research and remediation.

Selection of method

There are at least two possible models which one could adopt in trying to reduce attrition.

Causative

Historically, research has concentrated on trying to elicit the causes of premature withdrawal (Comptroller and Auditor General, 2001; Comptroller and Auditor General for Wales, 2001; Deary et al., 2003; Glossop, 2001, 2002; Johnes and McNabb, 2004; Last and Fullbrook, 2003, Wharrad et al., 2003). Although it is rarely discussed overtly, presumably the logic behind this approach is that if the causes were known, they could be removed or alleviated, and the attrition would be reduced.

With that approach, the major problem is one of attributing causation. For that to be demonstrated, a necessary, if not sufficient, condition is that students who drop out exhibit the posited causal factors, while students who do not drop out, do not exhibit them. In attrition research that design has rarely been used. Unfortunately, most studies have looked at students who have withdrawn, but not at more successful ones. In other words, there have been no control groups.

One can overcome this problem by studying not only those who had dropped out but also those who had thought of dropping out, but had eventually stayed the course. In my own institution this was done. The reasons given for withdrawing, or

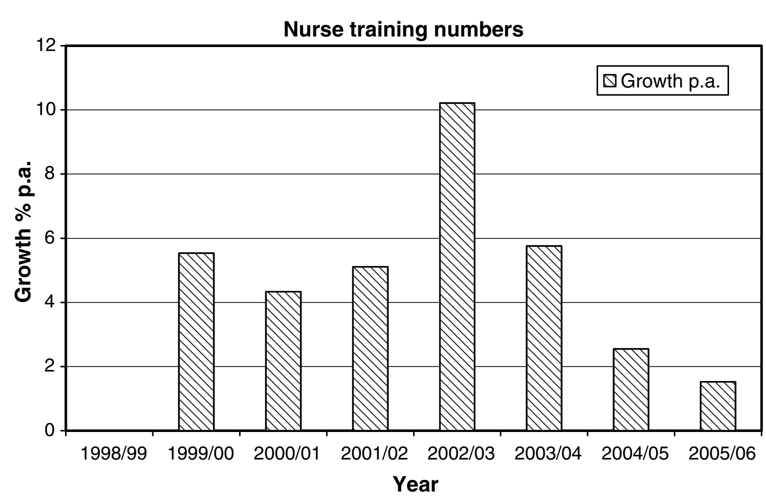

Figure 1 Annual growth in number of nurses in training source: (HESA 2007) — Code 57.

for thinking of withdrawing (even when that thought had not been acted upon) were the ones which had been mentioned in the literature – financial, health, and family problems (Glossop, 2001, 2002). Financial, health and family problems characterised both groups. They were factors which did not distinguish withdrawing students from others.

That is not to say that these practical problems are unimportant. They may not be ultimate causes, but they may sometimes be proximate causes. It appears that academically strong students may weather financial or other problems, whilst for weaker students such problems can prove the last straw.

Predictive

Many tutors have commonsense ideas about how to help students who are at risk – financial support, the services of a student counsellor, study skills sessions, more liberal re-sits, etc. Many of these commonsense solutions will not be aimed at all the students in a cohort. They are usually targeted at only some students – those whom the staff think may be in difficulty. In such cases, it may be believed that the solutions are known; but the problem is to identify the students to whom assistance should be offered, and to do so in advance of the signs becoming obvious. This means the problem is not one of causation, but one of prediction. How though would one make predictions? This question has been examined (Alexander and Brophy, 1997; Hayes, 2005), although rarely in the UK.

There was a broad choice between using an informal or a formal method of trying to make the predictions. The choice between the two has been an active research area for 50 years. The weaknesses of unaided human information process have been voluminously documented (Kahneman et al., 1982; Baron, 2000; Gilovich et al., 2002; Myers, 2004; Tetlock, 2005). Direct comparisons of the accuracy of clinical/intuitive (informal) and algorithmic/statistical (formal) ways of making decisions have appeared every decade (Meehl, 1954; Sawyer, 1966; Dawes and Corrigan, 1974; Dawes, 1979; Dawes et al., 1989; Marchese, 1992; Grove and Meehl, 1996; Moseley, 1999). This culminated in a meta-analysis of 136 studies (Grove et al., 2000). In only 8 of those 136 studies did clinical/intuitive approaches outperform the algorithmic ones.

A proposal was therefore prepared to undertake research on attrition using an algorithmic method and it duly received funding. Thanks are due to the Welsh Assembly Government for funding the study.

Data used

Two sorts of data were gathered. There were those items which were time-invariant – those known on entry: age, gender, entry qualifications, branch of nursing. There were other time-varying items, largely to do with the student's performance, which changed over time – grades awarded each semester for various modules, and gross and net attendance records. These were chosen largely because they were routinely available on the university's database system, not because they were expected to be predictive. Data were gathered on 528 students over a 5-year entry period. In total the project used 3978 individual records. Nearly 4000 records is a large data set by nursing research standards, but it is tiny compared with similar exercises in industry, commerce, and the natural sciences.

Methods: data analysis

The first step was the obvious one of undertaking univariate analyses of the relationship of the time-invariant data items to the propensity to withdraw early (e.g. were gender, age-group, or branch of nursing on their own related to attrition?). However, the problem was clearly likely not to be a univariate one, but rather a multivariate one – trying to see the effect of all our variables taken together, e.g. what was the effect of gender when all other factors had been taken into account. A conventional statistical procedure (e.g. logistic regression) could have been chosen. However, such procedures, powerful though they are, produce results which are in the form of equations, coefficients, residuals, etc. – none of which fit in with the usual mental sets of many nurse educators or nurses in practice. Something more widely interpretable was needed.

A method which comes from the field of machine learning or data mining (Jones, 2001) was therefore chosen. It is called rule induction. It takes a set of data (called the training set) about a group of students and tries to induce rules which explain the relationships between the independent variables (age, gender, marks, etc.) and the dependent one (probability of withdrawing prematurely). These are then expressed in the form of IF–THEN rules. For example ''IF the student is well qualified (A-level or better) AND is on the learning disability branch AND is female AND has failed more than 23% of their first year modules THEN they have a high (42%) chance of withdrawing prematurely''.

Such rules are produced automatically by the system; no human intervention is needed. It was thought that expressing relationships as rules would be more comprehensible to the target readership than a logistic regression equation would be. Although that belief appears to be true in medicine, psychology, and the law (Gigerenzer, 2002), it is not known whether it holds in nursing as well.

Software used

The data were exported from the university database system into Excel, and some transformations undertaken as a preparatory step. For example, each grade in each module for each student was automatically characterised, according to reproducible criteria, as a Clear Pass, Borderline Pass, or Clear Fail, a step which would probably not have been taken had a parametric technique been used. The data were then read into SPSS using a syntax file rather than the menu system. That permitted the data reading to be undertaken automatically with no possibility of subsequent typing errors.

There are many data mining packages available (Haughton et al., 2003). The rule induction package which was used in this project was called Answer Tree, and comes from the SPSS stable. It normally reads its data from an SPSS Sav file, and can export the induced rules in SPSS or SQL syntax. This means that once the rules have been learned, they can be written to a new SPSS syntax file and that file can be run on a new, previously unseen, data set, i.e. it could easily become part of routine practice. If one had good reason so to do, one could of course modify by hand the automatically induced rules.

Training sets and test sets

The notion of running the rules against a previously unseen data set is vital to the testing of any induced rule set (or any other representation of statistical relationships, or of any other posited relationships, for that matter). The software is easy to use, and it is a simple matter to generate rules. However, as in all studies, from the most quantitative to the most qualitative, there is the ever-present danger of over-fitting. Any human or machine learning method can fit a data set. In this case, that means that it can generate a set of rules which describe a data set in an accurate yet parsimonious way.

However, the danger is that the rules will describe that data set too well. In particular, just as people do, an algorithmic system may wrongly take into account all sorts of irrelevant relationships which lurk in the data. At the ultimate, it will produce one rule to describe each and every student, and therefore describes the data set as perfectly as is possible given the variables included. That, unfortunately, means that it may well be influenced by serendipitous and random fluctuations which apply to the individuals on whom it has learned the rules, but which may very well not be present for any other potential students to whom one wishes to apply the rules in future.

To be able to describe past experience, however parsimoniously, is no guarantee of the ability to predict the future. The application of that fundamental principle is unusual in nursing research, in which the generation of theory is often restricted effectively to describing, parsimoniously or otherwise, a set of data, with no attempt to make predictions of as yet unseen events. However, the distinction is central to the epistemology of the successful natural sciences. It was applied in the study described here.

For this reason, one uses two sets of data. The first is called the training set, and is the one on which the system generates its rules. Most nursing studies take that step. However, to be sure that they generalise well enough for later use, one must try out those generated rules against a set of previously unseen data called the test set — a process which is less common in the nursing literature. In the current case, that test data set consisted of students who were not in the training set, and could therefore not have been be over-fitted. To obtain the two sets the overall data file was split into two using random numbers.

All the results given below are based on performance of the system on the test set. That means that it reflects the likely performance in real-world usage, since all the cases to which the rules were applied were previously unseen ones. There was no hindsight bias of the type which is so common with unaided human judgement (Jay and Cynthia, 1991; Guilbault et al., 2004).

Rule induction in more detail

Rule induction is one of the methods used in machine learning. There are many ways of doing it. The one which was used in this project is called CHAID (Chi-Square Automatic Interaction Detector).

CHAID works by building a decision tree. It starts with the whole training set, finds the best single predictor and divides the set into two child nodes.

It then repeats the process with each of those child nodes. It continues building the tree and recursively descending until it reaches some pre-determined limit (a stopping rule) or runs out of variables. The process of generating rules is simple. The program merely traces each path from the root (top) of the tree to each of the leaf nodes (bottom of a branch). In the current study, the final induced rule set contained 17 rules, none of which had been articulated by tutors before the study was undertaken.

Results: how successful was the method?

Clearly, the system would be successful if it made correct predictions. However, there are at least three ways of defining correct predictions. They are:

Overall accuracy

This represents the percentage of predictions which turned out to be true. In the current study that figure was 94%. That sounds impressive. However, the problem with this measure is that if the data set is unbalanced this is too easy a criterion to achieve. In the current case, the overall attrition rate was only 16%, leaving the other 84% to complete successfully. One could therefore obtain an apparently impressive 84% overall accuracy by the use of the simple rule ''Predict that NO students will withdraw prematurely''. Impressive though its performance would appear to be, such a rule would be useless in practice, since it would fail to predict even a single one of the eventual drop outs. One has therefore to consider the other two conventional measures. They answer two different questions. Which question one wishes to ask is a policy matter, not a research one.

Sensitivity

This gives an answer to the question ''What proportion of those who eventually dropped out did the system correctly identify in advance?'' Maximising that quantity should enable one to concentrate resources in the proper place and thus reduce attrition.

Specificity

This gives an answer to the question ''What proportion of those whom the system predicted would drop out eventually did so?'' Maximising this quantity should help to avoid wrongly labelling students as being at risk when they were not in fact so.

These two quantities are inherently in tension. The higher the proportion of the eventual drop outs one identifies in advance, the more of those so identified will not in fact eventually withdraw. Conversely, the better one is at avoiding wrongly labelling students as being at-risk, the less likely one is to identify all those who really are at-risk. One has to do one's best to maximise both sensitivity and specificity.

How well did the system perform?

The factors which were investigated were:

- Sensitivity
- Specificity
- Overall accuracy
- Did using the student's performance on the course add anything to making predictions on the time-invariant data available at their time of entry?

Thus, with the test set of previously unseen data, and using only a small list of routinely recorded variables available before the semester in which they withdrew, the system was able to identify, in advance, 84% of those who later withdrew prematurely. Of those whom it identified as at risk, 70% eventually withdrew prematurely. Together they meant that the predictions were 94% accurate (Table 2).

Do data emerging during the course add to the accuracy of prediction?

Presumably the earlier one makes an accurate prediction, the more useful it will be. One therefore has to ask whether predictions made as the course proceeds (and marks, and attendance data become available) are better than those made on the basis of the simple time-invariant data available at the point of the student's entry to their course. Perhaps the most important comparison is between

Table 2 Success with unseen cases using three measures

	% Success
Sensitivity	84
Specificity	70
Accuracy	94

Table 3 Measures of % success in predicting withdrawal at two points in time

	At start of course	During course
Sensitivity	31	84
Specificity	44	70

the degree to which one might have been able to predict withdrawal from data available at the start of the course using time-invariant variables alone, and the degree to which one could do so as grades and attendance data became available, but more than a semester before the decision to withdraw was made. Table 3 draws that comparison.

Clearly, as one had expected, having data on student performance during the course adds considerably to the effectiveness of prediction, increasing both sensitivity and specificity. It is a matter of some concern that even before the course began, the rules induced in this study would have identified 31% of those who eventually withdrew. There is an implication that if the use of such algorithmic methods to assist in the decision to admit might on its own substantially reduce attrition, by helping to minimise the admission of students who had little chance of succeeding.

It is difficult to assess how good in the more general scheme of things, the sensitivity and specificity figures reported here are. The only comparable study (Boudreaux, 2004) obtained 74% specificity, but in a different culture, and with more data. The obvious way of testing the method would be to compare it with predictions made by experienced tutors who know their students. Unfortunately, there appear to be no reports of such formal predictions being made, still less of how accurate they are. It is to be hoped that over time they will become a routine part of practice.

Discussion

The method appears to work fairly well, but it does need good data. The early stages of the current project found gaps and inaccuracies in our central records system. Of course, since then attempts have been made to improve the quality of the recorded data. For this application, it appears that one should not initially discard missing data — in the early univariate analyses the existence of missing data items turned out to be predictive in many cases, especially in the cases of students who were at risk.

Given space constraints, this account has of necessity been brief. There is much more that one can do even with routinely available software. Apart from changing the maximum depth of the tree permitted, one can change the sensitivity of the chi-squares used, and one can define costs. One can, for example, specify that wrongly labelling a student as at risk is *x*% more or less serious than failing to identify a student who really is at risk. Making such judgements is again a policy, not a research, matter. Until consensus on the formal setting of such costs has been achieved, or even considered, they cannot be taken into account. Clearly they should be.

This study started by looking at methods developed in a cognate discipline — computer science. Nurse researchers could also look for solutions from other disciplines, especially if those other disciplines have a track record of leading to successful technologies. One discipline which should be a close relative of nursing is cognitive psychology. A recent review of 125 meta-analyses of psychological test validity concluded that the validity of such tests was comparable to that of diagnostic tests in medicine (Meyer et al., 2001). One would expect there to be benefit to nursing of borrowing ideas and methods from that discipline as well.

Conclusion

Rule induction, even when implemented in a simple and widely available package, holds out some promise as a routine solution to part of the attrition problem. If the research reported here was repeated on a wider scale, it would be of considerable help in attempts to reduce attrition from nursing courses. More generally, an algorithmic and statistical approach, far from losing depth and richness, actually increases both. However it requires good quality data, and the provision of such data should be a consideration for most courses.

There are many branches of nursing research in which the idea of machine learning could lead to a paradigm shift which would probably benefit patients. Above all, one can obtain many insights by moving from a causative model to a predictive one. After all, an expert has been anonymously defined as ''Someone who can make money by betting on the likely occurrence of future events''.

References

Alexander, J.E., Brophy, G.H., 1997. A five-year study of graduates' performance on NCLEX-RN. Journal of Nursing Education 36 (9), 443—445.

Auditor General for Wales, 2001. Educating and Training the future health professional workforce for England. Presented to the National Assembly, 1 March 2001.

Baron, J., 2000. Thinking and Deciding, third ed. Cambridge, UP.
Boudreaux, R.W., 2004. Predicting the success of vocational nursing students. Ed D Dissertation University of Houston.
Comptroller and Auditor General, 2001. Educating and Training the future health professional workforce for England. Report by the Comptroller and Auditor General HC 277, 1 March 2001.
Dawes, R.M., 1979. The robust beauty of improper linear models in decision-making. American Psychologist 34, 571—582.
Dawes, R.M., Corrigan, B., 1974. Linear models in decision-making. Psychological Bulletin 81, 95—106.
Dawes, R.M., Faust, D., Meehl, P.E., 1989. Clinical versus actuarial prediction. Science 243, 1668—1674.
Deary, I.J., Watson, R., Hogston, R., 2003. A longitudinal cohort study of burnout and attrition in nursing students. Journal of Advanced Nursing 43 (1), 71—81.
Gigerenzer, G., 2002. Reckoning with Risk. Penguin Books, UK.
Gilovich, T., Griffin, D., Kahneman, D., 2002. Cambridge University Press, New York.
Glossop, C., 2001. Student nurse attrition from pre-registration courses: investigating methodological issues. Nurse Education Today 21 (3), 170—180.
Glossop, C., 2002. Student nurse attrition: use of an exit-interview procedure to determine student's leaving reasons. Nurse Education Today 22 (5), 375—396.
Grove, W.E., Meehl, P.E., 1996. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical—statistical controversy. Psychology, Public Policy and Law 2, 293—323.
Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, C.E., Nelson, C., 2000. Clinical versus Mechanical prediction: a meta-analysis. Psychological Assessment 12, 19—30.
Guilbault, R.L., Bryant, F.B., Howard, J., Brockway, E., Posavac, J., 2004. A meta-analysis of research on hindsight bias. Basic and Applied Social Psychology 26 (2—3), 103—117.
Haughton, D., Deichmann, J., Eshghi, A., Sayek, S., Teebagy, N., Topi, H., 2003. A review of software packages for data mining. American Statistician 57 (4), 290—309.
Hayes, C., 2005. Identification of cognitive and non-cognitive predictive variables related to attrition in baccalaureate nursing education programs in Mississippi. Ed D dissertation. Delta State University.
Jay, J.C.-S., Cynthia, F.W., 1991. The hindsight bias: a meta-analysis. Organizational Behavior and Human Decision Processes 48 (1), 147—168.
Johnes, G., McNabb, R., 2004. Never give up on the good times: student attrition in the UK. Oxford Bulletin of Economics and Statistics 66 (1), 23—47.
Jones, J.K., 2001. The role of data mining technology in the identification of signals of possible adverse drug reactions: value and limitations. Current Therapeutic Research 62 (9), 664—672.
Kahneman, D., Slovic, P., Tversky, A., 1982. Judgement Under Uncertainty: Heuristics and Biases. Cambridge University Press, Part IV.
Last, L., Fullbrook, P., 2003. Why do student nurses leave? Suggestions from a Delphi study. Nurse Education Today 23 (6), 449, pp. 39—58.
Marchese, M.C., 1992. Clinical versus actuarial prediction: a review of the literature. Perceptual and Motor Skills 75 (2), 583—594.
Meehl, Paul E., 1954. Clinical Versus Statistical Prediction. University of Minnesota Press, Minneapolis.
Meyer, G.J., Finn, S.E., Eyde, L.D., Kay, G.G., Moreland, K.L., Dies, R.R., Eisman, E.J., Kubiszyn, T.W., Reed, G.M., 2001. Psychological testing and psychological assessment — A review of evidence and issues. American Psychologist 56 (2), 128—165.
Moseley, L.G., 1999. Is Thinking Just Too Difficult? Inaugural Lecture. University of Glamorgan.
Myers, D.G., 2004. Intuition: its powers and perils. Yale UP, New Haven.
Nursing Standard, 20, 23, Feb 15 2006, p 5.
Sawyer, J., 1966. Measurement and prediction clinical and statistical. Psychological Bulletin 66 (3), 178—200.
Tetlock, P., 2005. Expert Political Judgment: How Good Is It? How Can We Know? Princeton UP, Cited in New Yorker, December 5.
Wharrad, H.J., Chapple, M., Price, N., 2003. Predictors of academic success on a Bachelor of Nursing course. Nurse Education Today 23 (4), 246—254.