# The Use of Hot Deck Imputation to Compare Performance of Further Education Colleges

Kay I. Penny[1], Mohammed Zegham Ashraf[2] and John C. Duffy[2]

[1]School of Accounting, Economics and Statistics, Napier University, Edinburgh, UK
[2]Scottish Funding Council, Donaldson House, Edinburgh, UK

There are problems associated with the use of existing educational performance indicators when comparing further education colleges in Scotland. Several approaches have been suggested to avoid or resolve problems in comparing performance indicators. This paper reports the results of a pilot study using hot-deck multiple imputation as a method of comparing performance on one particular course in six colleges. Two different modelling scenarios are examined and compared. Results from the two scenarios are consistent, showing that colleges which perform well using one modelling approach also perform well in the other.

*Keywords:* hot-decking, multiple imputation, educational performance indicators

## 1. Introduction

The Scottish Further and Higher Education Funding Council (SFC) is the body that distributes funding for teaching and learning, research and other activities in Scotland's universities and colleges. There are 43 colleges and 20 higher education institutions currently funded by the SFC, and data are collected on students attending these institutions. One of the key functions is to establish targets and indicators of performance, and to monitor student progress.

A Performance Indicator (PI) is a summary statistical measure on an institution or "system" that is intended to be related to the quality of its functioning. These measures can be designed to measure different aspects of the system and can be classified into three main categories: input, process and outcome [1]. The percentage of staff with a teaching qualification may be used as an input indicator, the number of staff contact hours per student is an example of a process indicator, and student pass rate is an outcome indicator. PIs on all Scottish Further Education colleges are currently produced by the SFC. They were primarily developed to meet the needs of the Council and the further education sector. They are also expected to be of interest to a wider audience e. g. prospective students, local enterprise companies and schools. The PIs cover areas such as the volume of further and higher education being delivered, the quality of provision reported by the Inspectorate, student satisfaction, student retention rates, student achievement rates, and staff qualifications and staff indicators.

However, given that institutions, whose performance is being measured, are competing for funds and market share, there is an incentive for them to create and publicise their own indicators of performance, in areas in which they perform well, or they may selectively report the rankings of external bodies [2]. Therefore there is the possibility of multiple and contradictory PIs being created, and there is a rationale for implementation of a single system of performance monitoring across the sector.

League tables i. e. rankings of institutions are commonly produced and used to compare institutions. However, problems can occur when these rankings are unadjusted. For example, in a study [3] of cardiac surgeons in New York, whose individual unadjusted patient death rates were published regularly, there was a tendency

to avoid taking on high risk cases, which led to a subsequent increase in mortality of cardiac patients.

In order to reduce such dysfunctional effects, value-added PIs are often advocated. For example, when looking at the performance of schools, these value-added measures are usually considered in terms of progress made, or change in performance, between successive cohorts of students. However, it is argued [4] that a value-added score is an unreliable indicator of school performance, and factors out with the school are important. Factors such as gender, economic deprivation, ethnicity and prior attainment have been found to be important in explaining differences in educational achievements [5-9]. Hence factors outside the control of an institution can be associated with student outcome, e. g. entry qualifications and the socio-demographic breakdown of the student group.

The aim of this study is to implement and examine a method of comparing the performances of further education colleges aimed at overcoming some of the problems associated with using PIs and value-added PIs for comparative purposes.

## 2. Methods

The dataset used in this study is a subset of the data collected annually from all the further education colleges in Scotland. The data include information on students who attended one particular college course (for one year) in each of the academic years from 2002–03 through 2004–05. This comparative study includes six colleges located in the same geographical area in Scotland, at which this particular course is offered. A total of 1 560 students are included in this study. The Further Education Statistics dataset contains 107 variables in the student record, of which only a subset are used in this analysis. The variables that were considered as important in this study are those associated with educational attainment, including gender, age, deprivation, ethnicity, mode of study (i. e. full-time or part-time study), and the number of Scottish Certificate of Education / General Certificate of Education (SCE / GCE) passes. SCE / GCE grades A-C are considered passes whereas grades D-E are fails. The outcome

variable in the study is categorised as achieving the course qualification aim.

The data analysis follows the approach suggested by Longford and Rubin [10], which avoids the use of value-added measures or direct adjustment of the data. In this approach, each student is associated with a potential outcome for each of the six colleges that he/she could have attended, although only one outcome is observed i. e. for the college that he/she actually attended. Thus the data have the form, $(W, \mathbf{Z}, \mathbf{Y})$, where $W$ indicates the college attended; $\mathbf{Z}$ represents the set of background variables, e. g. the student characteristics; and $\mathbf{Y}$ represents the set of potential outcomes. Comparisons may be made between the six colleges, $\mathbf{C}_1 \ldots \mathbf{C}_6$ in terms of summary statistics e. g. the proportion of successful students. In order for a summary statistics to be calculated, the dataset $(W, \mathbf{Z}, \mathbf{Y})$ must be completed. Only $W$ and $\mathbf{Z}$ are completely observed, but for $\mathbf{Y}$, only the observed outcome in the college the student attended is known, i. e. $\mathbf{Y}_W$ is observed. The outcome $\mathbf{Y}$ is missing for all 5 colleges that the student could potentially have attended but did not attend.

The missing values of $\mathbf{Y}$ are imputed using the method of multiple imputation [11], where each missing value is replaced with $M \geq 2$ imputed values, resulting in $M$ completed datasets. The $M$ complete datasets are then combined to form one inference that reflects the uncertainty due to "missingness" under that model. Since different imputations are created for a particular missing value for each of the different data subsets, the between-imputation variability can also be estimated. In this study, $M = 5$ imputations are considered, although the analysis is repeated with $M = 10$ for comparative purposes.

Hot-deck imputation [11] is used to create $M$ imputations for each missing observation. The method involves imputing the unobserved outcomes for a particular student, $\mathbf{Y}_W$ by individual outcomes, $\mathbf{Y}$, drawn from students with observed outcomes who are "similar" to the student in question. This is done by random selection of an observed student outcome drawn from the subset of students who have similar background variables, $\mathbf{Z}$, that are associated with $\mathbf{Y}$. In this study the background variables, $\mathbf{Z}$, were selected using logistic regression modelling with student outcome as the dependent variable. Students with unobserved outcomes

will then have their missing outcomes replaced with observed outcomes selected at random, with replacement, from the donor pool of students i. e. students who are similar with respect to the background variables, **Z**. If there are less than five observed outcomes in the donor pool, then the subset is collapsed by excluding the least important background variable from the matching, and this process is repeated until the donor pool contains at least five outcomes.

The six colleges are compared using the overall success rate for each college as the outcome, **Y**. The colleges are compared under two scenarios, using hot-decking to multiply impute the missing observations. In Scenario 1, the success rate is estimated as if the students attending a particular college had come from any of the other colleges in the sector, the sector being defined as all other colleges in Scotland offering the same course. In Scenario 2, the success rate is estimated for each college as if all students had attended each of the six colleges.

Once the M imputations have been carried out for each unobserved outcome, this gives M completed datasets. As outlined by Rubin [12], for each of the *M* completed datasets, a success rate, $\hat{Q}$, is calculated. An overall estimate of the success rate for each college is:

$$\bar{Q} = \frac{1}{M} \sum_{i=1}^{M} \hat{Q}_i \qquad (1)$$

The between-imputation variance, *B*, is given as:

$$B = \frac{1}{M-1} \sum_{i=1}^{M} \left( \hat{Q}_i - \bar{Q} \right)^2 \qquad (2)$$

Since M is finite, the total variance, *T*, is corrected by the term, $M + 1/M$:

$$T = \frac{M+1}{M} B \qquad (3)$$

Confidence intervals are calculated for $\bar{Q}$ using the Student's *t*-approximation:

$$\bar{Q} \pm t_{df} \sqrt{T} \qquad (4)$$

where $df = M - 1$.

Data analysis was carried out using SAS [13].

## 3. Results

The observed success rates in each of the six colleges are given in Table 1. Success rates vary between 62.6% and 87.4%, and the mean success rate for all six colleges is 78.1%.

The results of the logistic regression modelling show that number of SCE/GCEs (A-C), the number of SCE/GCEs (D-E), other qualifications, age, gender, mode of study, ethnicity and deprivation are all independently associated with student outcome. The variables, number of SCE/GCEs (D-E), other qualifications, and ethnicity were excluded from the division of the population into donor subsets since the majority of students (84%, 88% and 97% respectively) all fell into one category.

|  | Success Rate (%) |
|---|---|
| College 1 | 77.5 |
| College 2 | 87.4 |
| College 3 | 83.7 |
| College 4 | 67.0 |
| College 5 | 69.9 |
| College 6 | 62.6 |
| **Mean for 6 colleges** | **78.1** |
| **Mean for Sector** | **78.5** |

*Table 1*. Success rates in each college.

Logistic regression modelling was carried out for a second time discarding the three variables above, in order to determine the order of importance of the remaining variables for the subdivisions in both scenarios. The two types of modelling results are given in Sections 3.1 and 3.2 for Scenario 1 and Scenario 2 respectively.

### 3.1. Scenario 1 Results

The number of SCE/GCEs (A-C) was not significant in the logistic regression modelling, hence age, mode of study, gender and deprivation (in order of importance) were used to create the donor subgroups for Scenario 1.

Results of the imputations for Scenario 1 are shown in Table 2. The success rate for each college based on $M = 5$ imputations is given, as

| $\hat{Q}$ | College | | | | | |
|---|---|---|---|---|---|---|
| **for** | **1** | **2** | **3** | **4** | **5** | **6** |
| **M=5** | | | | | | |
| **1** | 0.80 | 0.82 | 0.79 | 0.78 | 0.79 | 0.74 |
| **2** | 0.80 | 0.77 | 0.79 | 0.77 | 0.82 | 0.74 |
| **3** | 0.81 | 0.77 | 0.76 | 0.78 | 0.78 | 0.70 |
| **4** | 0.75 | 0.81 | 0.72 | 0.80 | 0.63 | 0.73 |
| **5** | 0.79 | 0.79 | 0.79 | 0.76 | 0.81 | 0.78 |
| $\bar{Q}$ | **0.79** | **0.79** | **0.77** | **0.78** | **0.79** | **0.74** |
| **95% Confidence Limits:** | | | | | | |
| **upp** | 0.86 | 0.87 | 0.87 | 0.83 | 0.86 | 0.82 |
| **low** | 0.72 | 0.72 | 0.67 | 0.73 | 0.72 | 0.65 |

*Table 2.* Imputation results for Scenario 1.

well as the overall success rate and confidence intervals calculated using equations (1) and (4). The actual observed success rates for each of the six colleges, along with the overall success rates and 95% confidence intervals based on the multiple imputations, are presented in Figure 1.

In Figure 1, the imputed overall college success rates tend towards the sector average (78.5%). The imputed college success rates are greater than the observed rates for colleges 1, 4, 5 and 6 with increases of 1.7%, 10.7%, 9.3% and 11.0% respectively. The differences for colleges 4, 5 and 6 are statistically significant, providing evidence that these three colleges are performing

worse than expected. For college 2, the observed success rate of 87.4% lies just out with the confidence interval for the imputed success rate of 79.2%, providing evidence that this college may be performing better than it would be expected.

To investigate whether $M = 5$ imputations are sufficient, the process was repeated again by creating another set of 5 imputations. This second set of results shows three colleges whose performance is significantly different from what was expected; colleges 2 and 3 are now performing significantly better, and college 4 is performing significantly worse than expected. Colleges 5 and 6 which were previously performing significantly worse are still performing below their imputed success rates, but the differences are not statistically significant. The results from the two sets of imputation analyses appear to give inconsistent results.

The imputation process was then repeated two more times, using $M = 10$ imputations to explore whether increasing $M$ gives more consistent results. The results from both sets of $M = 10$ imputations give very consistent results. In both cases there is statistical evidence that colleges 2 and 3 are performing better than expected according to the imputation results, whereas colleges 4, 5 and 6 are performing worse. Results from one set of $M = 10$ imputations are presented in Figure 2.
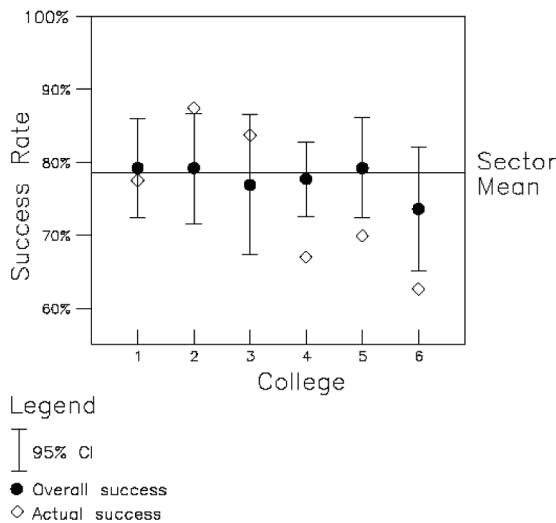


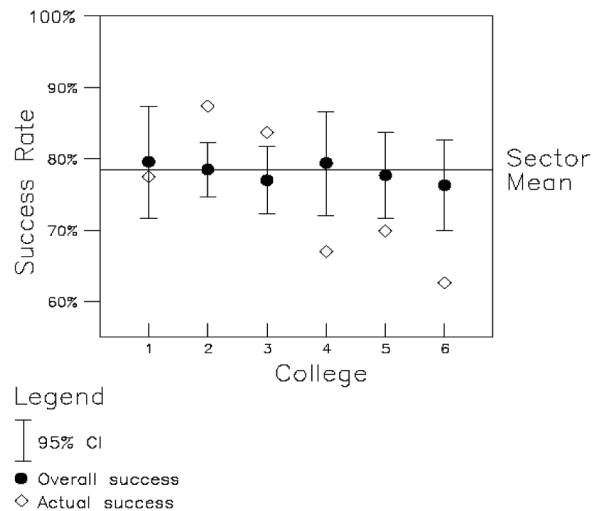*Figure 1.* Scenario 1: Success rates and 95% confidence intervals – first set of 5 imputations.



*Figure 2.* Scenario 1: Success rates and 95% confidence intervals – first set of 10 imputations.

## 3.2. Scenario 2 Results

Scenario 2 allows success rates to be estimated as if each individual college had the entire student population of all six colleges. This scenario allows comparison between colleges, and not that of assessing one college's performance against the sector.

In Scenario 2, using age, mode of study, gender and deprivation to create the donor subgroups resulted in many subgroups with no donors or small counts of donors. Hence the variable, gender, was discarded, and age, mode of study

| $\hat{Q}$ for $M=5$ | College | | | | | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** |
| **1** | 0.77 | 0.87 | 0.84 | 0.66 | 0.71 | 0.58 |
| **2** | 0.77 | 0.86 | 0.86 | 0.67 | 0.74 | 0.57 |
| **3** | 0.76 | 0.86 | 0.85 | 0.66 | 0.70 | 0.56 |
| **4** | 0.77 | 0.88 | 0.86 | 0.70 | 0.70 | 0.59 |
| **5** | 0.77 | 0.87 | 0.85 | 0.69 | 0.71 | 0.59 |
| **Q** | **0.77** | **0.87** | **0.85** | **0.68** | **0.71** | **0.58** |
| **95% Confidence Limits:** | | | | | | |
| **upp** | 0.78 | 0.89 | 0.88 | 0.72 | 0.76 | 0.62 |
| **low** | 0.75 | 0.85 | 0.83 | 0.63 | 0.66 | 0.55 |

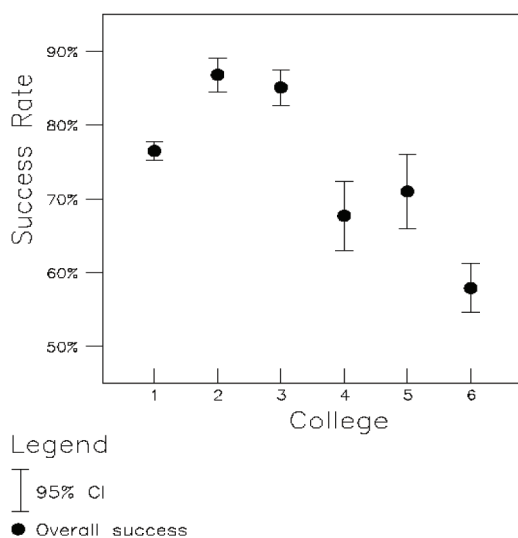*Table 3.* Imputation results for Scenario 2.



*Figure 3.* Scenario 1: Scenario 2: Success rates and 95% confidence intervals – 5 imputations.

and deprivation were used to create the donor subgroups for Scenario 2.

Imputed success rates and 95% confidence intervals for each college using $M = 5$ imputations are given in Table 3, and the results are shown graphically in Figure 3.

Colleges 2 and 3 with imputed overall success rates of 86.8% and 85.1% are clearly performing better than the other colleges, whereas college 6 with an imputed success rate of 57.9% is doing worse. College 1 is also performing better than college 4.

The analysis is repeated for Scenario 2, using $M = 10$ imputations. The results are consistent with $M = 5$.

## 4. Conclusions

This study demonstrates that multiple imputation techniques can be employed in the comparison of performance indicators for Scottish Further Education Colleges. The method requires unknown or unobserved values to be imputed, and permits investigation under two scenarios. In Scenario 1, the success rate is estimated as if the students attending a particular college were attending any of the other colleges in the sector. This allows the actual college performance to be assessed, as it allows comparison between observed success rate at the particular college and the estimated success rate if the students were attending one of the other five colleges. In Scenario 2, the method explores what might have happened if the entire student population of all six colleges had hypothetically attended each of the individual six colleges, thereby allowing a comparison of performance between the six colleges. The results show that College 2 appeared to perform well in both scenarios, whereas colleges 4, 5 and 6 appear to be underperforming for this particular course.

This technique also allows estimates of precision to be calculated for the estimated success rates. The results from $M = 10$ imputations were more stable and more precise than for $M = 5$, although 5 imputations appeared to be sufficient in this study for Scenario 2. By repeating the multiple imputations, results were found to be consistent, providing assurance that the modelling process was stable.

One difficulty in using this method is in determining the donor subgroups for the hot-deck multiple imputations. Having many variables in the donor subgroup division ensures that imputations are selected from a similar subgroup based on many important socio-demographic factors. However, having too many variables in the subgroup division leads to donor subgroups containing very small or zero counts of students, which in turn causes problems in selecting the imputations. Further work is needed to investigate alternative methods for selecting the multiple imputations, and model-based imputations are an attractive possibility for future investigation.

## References

[1] H. GOLDSTEIN, D. J. SPIEGELHALTER, League Tables and their Limitations: Statistical Issues in Comparisons of Institutional Performance. *Journal of the Royal Statistical Society*, Series A, **159** (3) (1996), 385–443.

[2] H. MORRISON, S. MAGENNIS, L. CAREY, Performance Indicators and League Tables: A Call for Standards. *Higher Education Quarterly*, **49** (2) (1995), 128–45.

[3] D. DRANOVE, D. KESSLER, M. MCCLELLAN, M. SATTERTHWAITE, Is More Information Better? The Effects of "Report Cards" on Health Care Providers. *Journal of Political Economy*, **111** (3) (2003), 555–88.

[4] J. TAYLOR, A. N. NGUYEN, An Analysis of the Value Added by Secondary Schools in England: Is the Value Added Indicator of Any Value? *Oxford Bulletin of Economics and Statistics* **68** (2) (2006), 203–24.

[5] D. GILLBORN, H. S. MIRZA, Educational Inequality. Mapping Race, Class and Gender. A Synthesis of Research Evidence. *Office for Standards in Education* (2000), http://www.ofsted.gov.uk [11/01/2007].

[6] D. T. HERBERT, C. J. THOMAS, School Performance, League Tables and Social Geography. *Applied Geography*, **18** (1998), 199–233.

[7] R. MCNABB, S. PAL, P. SLOANE, Gender Differences in Educational Attainment: The Case of University Students in England and Wales. *Economica*, bf 69 (275) (2002), 481–503.

[8] D. V. RAINEY, O. MUROVA, Factors Influencing Education Achievement. *Applied Economics*, **36** (21) (2004), 2397–404.

[9] J. SCOTT, Family, Gender and Educational Attainment in Britain: A Longitudinal Study. *Journal of Comparative Family Studies*, **35** (4) (2004), 565–90.

[10] N. T. LONGFORD, D. B. RUBIN, Performance Assessment and League Tables. Comparing Like with Like. Department d'Economica i Empresa, Universitat Pompeu Fabra, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain (2005), Working Paper No. 994.

[11] R. J. A. LITTLE, D. B. RUBIN, *Statistical Analysis with Missing Data*. John Wiley & Sons, New Jersey, 2002.

[12] D. B. RUBIN, *Multiple Imputation for Nonresponse in Surveys*. John Wiley, New York, 1987.

[13] SAS INSTITUTE INC., SAS/STAT User's Guide. Cary, *NC: SAS Institute Inc.*, fourth ed. (1989).

*Contact addresses:*
Kay I. Penny
School of Accounting,
Economics and Statistics
Napier University
Craiglockhart Campus
Edinburgh, EH14 1DJ
e-mail: k.penny@napier.ac.uk

Mohammed Zegham Ashraf
Scottish Funding Council
Donaldson House
97 Haymarket Terrace
Edinburgh, EH12 5HD
e-mail: zashraf@sfc.ac.uk

John C. Duffy
Scottish Funding Council
Donaldson House
97 Haymarket Terrace
Edinburgh, EH12 5HD
e-mail: jduffy@sfc.ac.uk

KAY PENNY studied at the University of Aberdeen, Scotland at both undergraduate and postgraduate levels, and was awarded a PhD in 1995. Since 1999, she has held the post of lecturer at Napier University in Edinburgh. Previously she held research posts at both The University of Aberdeen (1995 – 1997) and Edinburgh University (1997 – 1999). Her current research interests include missing data imputation methods, as well as data mining and medical statistics.

MOHAMMED ZEGHAM ASHRAF currently holds the post of a Statistics Officer, Scottish Funding Council, and was awarded an MSc in Applied Statistics at Napier University, Edinburgh in 2006.

JOHN DUFFY currently holds the post of the deputy director, Corporate Policy and Services, Scottish Funding Council (since 2003). Previous posts include Head of Statistics, Department of Primary Care, University of Birmingham (2001 – 2003); research manager, Chief Scientist Office, Scottish Office Department of Health (1996 – 1999); professorial research fellow, Department of Psychiatry, University of Edinburgh (1990 – 1996); and Senior lecturer, Department of Statistics, University of Edinburgh (1970 – 2001).