Annex B

Adjusted sector benchmarks – technical notes and detailed information

1. This annex contains the technical details and assumptions made in producing the adjusted sector benchmarks for Tables T1 – T3 and T7. It also covers the location-adjusted benchmarks, and the calculations for the standard deviations. Details of the subject and entry qualifications breakdown used to obtain the benchmarks, and tables showing the numbers of students in each category and the proportion of students in each category with different characteristics, are given at the end of this annex.

Adjusted sector benchmarks

2. The adjusted sector benchmarks make allowance for various factors which affect the indicators. The main factors used are the subject mix of an institution, and the entry qualifications of its students. In addition, for the indicators which cover all age groups the benchmark also takes into account the proportions of young (under 21) and mature students entering the institution.

3. This year, the access indicators in Tables T1 and T2 are provided with a second benchmark, the location-adjusted benchmark. This uses a student's region of domicile as an additional factor. The benchmarks produced for the employment indicators use other factors, see Annex E.

Technical notes

4. The factors allow the population to be broken down into well-defined categories, which are used in the calculation of the adjusted sector benchmark. In addition, the 'sector population' needs to be defined, as it is not the same in all cases. Each indicator relates to a specific sub-set of the institution's students, for example, young full-time first degree students, or mature part-time undergraduates, and the adjusted sector benchmark is based on the equivalent sub-set of the sector population.

5. The sub-set of the population used will only contain students for whom information to calculate the indicator is available. The institution's profile is also based only on those of its students with that information available. So, for example, if the information about school type is available for only 80 per cent of an institution's students, the institutional profile used to obtain the benchmark for the indicator will be based on that 80 per cent.

6. The number of categories used in the calculation of the benchmarks will depend on which factors are included. As there are 13 subject groups and 21 entry qualification groups, the original adjusted sector benchmark for the access indicators is based on $13 \times 21 = 273$ categories. For the non-continuation indicator for all ages, where age is also taken into account, the number of categories will double to 546. And for the location-adjusted benchmark for the access indicators, where region is also a factor, there will be

$273 \times 12 = 3{,}276$ categories.

7. The calculations of the benchmarks use what we call the 'basic grid' of the population. This is the two-way classification of all students with known data for that indicator by institution and by the factor categories. For the original benchmark for the Social Class indicator, for example, the basic grid will be a classification all students with known Social Class by institution and by a category giving their subject of study and their entry qualifications.

8. Assume there are C categories, numbered from 1 to C, and U institutions, numbered from 1 to U. Let the number of students in institution j in category k be n_{jk}. Then the total number of students at institution j is

$n_{j.} = \sum_{k=1}^{C} n_{jk}$, the number from the sector in category k is $n_{.k} = \sum_{j=1}^{U} n_{jk}$, and the total number of students in the sector is $N = \sum_{k=1}^{C} \sum_{j=1}^{U} n_{jk}$.

9. Let $p_{.k}$ be the proportion of students in the sector from category k who have the characteristic of interest, for example, are from state schools, or have left HE after a year, and the equivalent proportion for institution j be p_{jk}. The proportion of students in institution j with the characteristic of interest can be found as

$$p_{j.} = \frac{1}{n_{j.}} \sum_{k=1}^{C} n_{jk} p_{jk}$$

This is the value of the indicator. If the proportion of students with the characteristic at the institution in each subject/entry qualification category was the same as in each category in the sector, then the overall proportion with the characteristic would be

$$\frac{1}{n_{j.}} \sum_{k=1}^{C} n_{jk} p_{.k}$$

This is what we have called the 'adjusted sector benchmark'.

10. Another way of interpreting this is to say it is the value that the sector average would have if the sector students were split across the C categories in the same proportions as at the institution.

Standard deviations

11. In general, small differences between an indicator and its benchmark are not important. However, it is not always obvious what constitutes a small difference. A standard deviation measures the amount by which one would expect a statistic to change, based solely on random sampling, and can therefore be used to say that a particular difference is significant or not. We have calculated the standard deviations of the differences between the indicators and their benchmarks, using a method developed by Professor David Draper and Mark Gittoes, formerly at the University of Bath. (Note that, because these are standard deviations of a statistic, they are more usually called standard errors.)

12. The mechanics of the calculations are explained below. More details of the statistical model used can be found in 'Statistical Analysis of Performance Indicators in UK Higher Education', Gittoes M. (2001), PhD thesis, School of Mathematical Sciences, University of Bath.

13. The calculations again use the basic grid of the population, see paragraph 8 above. The actual indicator at institution j, $p_{j.}$, is a weighted average of the form

$$p_{j.} = \frac{\sum_{k=1}^{C} n_{jk} p_{jk}}{\sum_{k=1}^{C} n_{jk}} = \frac{1}{n_{j.}} \sum_{k=1}^{C} n_{jk} p_{jk}.$$

14. The proportion of students in the sector in category k, $p_{.k}$, is

$$p_{.k} = \frac{\sum_{j=1}^{U} n_{jk} p_{jk}}{\sum_{j=1}^{U} n_{jk}} = \frac{1}{n_{.k}} \sum_{j=1}^{U} n_{jk} p_{jk}$$

and the benchmark for institution j, E_j, is

$$E_j = \frac{1}{n_{j.}} \sum_{k=1}^{C} n_{jk} p_{.k} = \frac{1}{n_{j.}} \sum_{k=1}^{C} \sum_{i=1}^{U} \frac{n_{jk} n_{ik} p_{ik}}{n_{.k}}$$

The difference between the indicator for institution j and its benchmark, $D_j = p_{j.} - E_j$, can then be written as a weighted sum of all $C \times U$ cells in the basic grid:

$$D_j = \sum_{k=1}^{C} \sum_{i=1}^{U} \lambda_{ijk} p_{ik}$$

where

$$\lambda_{ijk} = \frac{n_{jk}}{n_{j.}} \left(\delta_{ij} - \frac{n_{ik}}{n_{.k}} \right)$$

and δ_{ij} = 1 if i = j

0 if $i \neq j$

15. Assuming that the n_{jk} students at institution j in category k are like a random sample (with replacement) from the population of all such future students, the values p_{ik} and D_j can be estimated as $\hat{p}_{ik}$ and $\hat{D}_j$ respectively. The variance of $\hat{D}_j$ is given by

$$\mathrm{var}\left(\hat{D}_j\right) = \sum_{k=1}^{C} \sum_{i=1}^{U} \lambda_{ijk}^2 \, \mathrm{var}\left(\hat{p}_{ik}\right)$$

We then have to estimate the variance of $\hat{p}_{ik}$.

16. Draper and Gittoes show that a reasonable estimate of this variance is obtained by using a shrinkage estimation procedure. The value used here is

$$\mathrm{var}\left(\hat{p}_{jk}\right) = \hat{p}^*_{jk} \left(1 - \hat{p}^*_{jk}\right) \Big/ n_{jk}$$

where $\hat{p}^*_{jk} = 0.5\,\hat{p}_{..} + 0.5\,\hat{p}_{jk}$, and $\hat{p}_{..}$ is the estimated percentage with the characteristic of interest in the sector as a whole.

17. The square root of the estimated variance, which is the standard deviation of the indicator, can then be used to test whether the difference between the indicator and its benchmark is small or not. A difference that is less than twice the size of the standard deviation can certainly be said to be small, but we have been more conservative. In the tables, we have marked as 'large' those institutions where the difference is both greater than three times the standard deviation and greater than three percentage points. This is to draw attention to areas where the difference is large in both statistical and practical terms.

18. If an institution is marked in this way, it should be taken as an invitation to investigate possible causes for the differences that have been identified, whether they arise from an indicator that is better than the

benchmark (marked +), or worse than the benchmark (marked -). Where the difference is not marked, the indicator is either within the range that would be expected given random fluctuations, or is less than three percentage points away from the benchmark.

Projected outcomes

19. The adjusted sector benchmarks for the projected outcomes indicators in Table T5 are obtained by adjusting the transition matrix rather than the actual indicators. The standard deviations have therefore been obtained by assuming students have been selected at random from the outcome categories. These are simplifications, but appear to give realistic results in most cases. Further details of the methods used are available in Annex C.

Subject and entry qualifications breakdown for the sector

20. The assumption behind the adjusted sector calculations is that each category used contains students who form a homogeneous group. The groupings are kept under continuous review. Subjects studied have been combined into 13 groups, while entry qualifications have been combined into 21 groups. The actual categories are described below.

Subject categories

21. The subject categories have been obtained from the subject of qualification aim, HESA field SBJQA. Table B1 shows the codes used for the subject categories, and the subjects they represent. These differ from the categories normally used by HESA in that several have been merged, for example, Biological Sciences with Physical Sciences, and Mathematics with Computing. Two categories have been merged only if the characteristics of the students in them appear to be broadly similar.

22. Any combined studies programme whose subjects lie completely within one category have been included in that category. For example, a student on a joint degree programme studying Mathematics and Computing will be included in category E.

23. The third column of table B1 shows the HESA subject codes that the category covers. The HESA subject codes consist of a letter followed by a number, and if all codes beginning with a particular letter are in the category, only the letter has been given. Any invalid subject codes, or combinations not wholly within one category, have been allocated to the combined studies category.

Entry qualifications

24. The majority of students in the UK still enter higher education with either A-levels or Scottish Highers, and this is recognised in the groupings of entry qualifications used. As with subjects, the groupings have been chosen so that as far as possible the students within each group are relatively homogeneous.

25. Both A-level grades and Scottish Highers grades are converted to a number of points. In deciding on the category to which a student should be allocated, A-levels and Scottish Highers have been treated as equivalent as far as allocated points are concerned, so that a student with 20 A-level points is allocated to the same category as a student with 20 Scottish Higher points. If a student has both Scottish Highers and A-levels, the number of points used is the greater of the two. (The introduction by HESA of new fields to take the revised UCAS tariff for different qualifications into account will lead to a change of these categories from next year.)

26. A few students now enter higher education with a Baccalaureate. It has been decided to allocate these students to the same category as students with A-levels or Scottish Highers having up to four points. This will be reviewed when the new tariff is introduced.

27. Access and foundation courses have been given a separate category, as have BTEC and related qualifications.

28. Students' entry qualifications have been taken from the HESA database if possible, but if they are not available from HESA then the UCAS database has been used. The HESA fields QUALENT2, ALEVELS, HIGHERS, VOCQUALS, ALEVPTS, and HIGHPTS have been used to determine the category (or the equivalent fields from UCAS).

29. Table B2 shows the categories used, their descriptions, and the values of the QUALENT2 field used in the definition. In addition, if QUALENT2 has a value of 40 (A-levels or Scottish Highers or GNVQ level 3) then the values in the fields ALEVELS, HIGHERS and VOCQUALS have been used to determine whether the student has GNVQ only, or some combination of A-levels or Scottish Highers. Only students with QUALENT2 coded 40, ALEVELS and HIGHERS coded blank, 0, 98 or 99, and VOCQUALS coded 1, 2, 3 or 97 are included in the GNVQ category.

Regions

30. The regions used in the location-adjusted benchmark are the nine Government Office regions in England, plus the three other countries of the United Kingdom – Wales, Scotland and Northern Ireland. The percentages of young entrants to full-time first degrees from each region who are from state schools, Social Classes IIIm – V, and low participation neighbourhoods are shown in Table A4 in Annex A2.

Tables

31. Tables B3 and B4 show the actual numbers of full-time degree entrants to higher education across the sector in each of the defined categories in 2000-01; Table B3 shows the number of young entrants, and Table B4 the number of mature entrants. Tables B5 to B12 provide the proportions in each category with various characteristics, again for the whole sector, split into young and mature entrants. Table B5 shows the proportion of young entrants in 2000-01 who come from state schools in each category, and Table B6 shows the proportions from Social Classes IIIM – V. Tables B7 and B8 give the proportions of 2000-01 full-time entrants from low participation neighbourhoods, for young and mature respectively. Tables B9 and B10 show the proportion of 1999-2000 entrants in each category who did not continue in higher education after their first year, for young and mature entrants. Table B11 shows the percentage of full-time first degree students in receipt of the Disabled Students Allowance in each category in 2000-01.

Table B1 **Subject categories and their equivalent subject codes**

	Subject description	**HESA subject codes**
A	Medicine, Dentistry, and Veterinary science	A, D1
B	Subjects allied to medicine	B
C	Biological sciences, and Physical sciences	C, F
D	Agriculture and related subjects	D, except D1
E	Mathematical sciences and Computer science	G
F	Engineering and Technology	H, J
G	Architecture, Building, and Planning	K
H	Social studies and Law	L, M
I	Business & administrative studies and Librarianship & information science	N, P
J	Languages and Humanities	Q, R, T, V
K	Creative arts and Design	W
L	Education	X
M	Combined subjects	Y

Table B2 **Entry qualifications and codes used in their definition**

Abbreviation	Entry qualification	Qualent2 code
A pts NK	A-levels / AS levels / Scottish Highers, number of points unknown	39, 40
A pts 4	A-levels / AS levels / Scottish Highers with 1 to 4 points, or Baccalaureate	39, 40, 47
A pts 8	A-levels / AS levels / Scottish Highers with 5 to 8 points	39, 40
A pts 10	A-levels / AS levels / Scottish Highers with 9 to 10 points	39, 40
A pts 12	A-levels / AS levels / Scottish Highers with 11 to 12 points	39, 40
A pts 14	A-levels / AS levels / Scottish Highers with 13 or 14 points	39, 40
A pts 16	A-levels / AS levels / Scottish Highers with 15 or 16 points	39, 40
A pts 18	A-levels / AS levels / Scottish Highers with 17 or 18 points	39, 40
A pts 20	A-levels / AS levels / Scottish Highers with 19 or 20 points	39, 40
A pts 22	A-levels / AS levels / Scottish Highers with 21 or 22 points	39, 40
A pts 24	A-levels / AS levels / Scottish Highers with 23 or 24 points	39, 40
A pts 26	A-levels / AS levels / Scottish Highers with 25 or 26 points	39, 40
A pts 28	A-levels / AS levels / Scottish Highers with 27 or 28 points	39, 40
A pts 30	A-levels / AS levels / Scottish Highers with 29 or 30 points	39, 40
ACCFND	Foundation or Access course	29, 43, 48
GNVQ3	GNVQ or equivalent, level 3	40
BTEC/ONC	BTEC, ONC, SCOTVEC or equivalent	41
HE	Higher education qualification	1, 2, 3, 4, 5, 10, 11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, 27, 28, 30
NONE	No previous qualification	92, 93, 98
OTHERS	Other qualifications not given elsewhere	55, 56, 97
UNKNOWN	Unknown qualification	99