

Data Disclosure Control Policy

Introduction

1. This policy applies, irrespective of length of service or duration of contract to:
 - employees of HEFCW
 - temporary or contract staff engaged by HEFCW, including where their services are contracted via a third party
 - consultants employed by HEFCW for a specific project
 - employees of organisations carrying out work on behalf of HEFCW
 - secondees
 - recipients of data from HEFCW via a data sharing agreement

Statement of Policy

2. The aim of this policy is to inform all parties referred to in paragraph 1 above of their responsibilities with respect to data disclosure controls adopted by HEFCW. Data disclosure controls are methods applied to aggregated data in order to prevent the possibility of revealing information about individuals. This policy applies to all data about individuals held and used in HEFCW.
3. Examples of individual data held and used in HEFCW include:
 - HESA student and staff data held and used by the Statistics and Funding Team
 - HESA student and staff unrounded data downloaded by Heidi plus Gold users
 - HEFCW staff data
 - Go Wales data
4. In light of changing availability and use of data in the public domain, and as a requirement of our data sharing agreement with the Higher Education Statistics Agency (HESA) regarding our use of HESA data, HEFCW is required to apply suitable data disclosure control methods when providing aggregate data analyses externally.
5. Examples of when it may be possible to deduce information about an individual from aggregated data include:
 - data where the population is small and categories give rise to small cell values (generally considered to be fewer than 5)
 - when two sets of data which contain large populations which are similar but different are taken one away from the other to produce small cell values (known as data differencing)
6. Although aggregate data is not considered to be personal data and is not covered by the requirements of the Data Protection Act, if it is possible to deduce information about an individual from the aggregated data, then the data deduced

Data Disclosure Control Policy

is covered by the requirements of the Data Protection Act. Care must be taken with aggregate data to reduce the risk of individual information being released, and particularly in the case of sensitive personal data (e.g. ethnicity, religious belief, physical and mental health).

Responsibilities and reporting

7. All staff and individuals outlined in paragraph 1 are responsible for ensuring that suitable data disclosure controls are applied to aggregate data relating to individuals in any data that they share externally. Data is considered to be shared externally if it is shared with anyone outside of those listed in paragraph 1.
8. The Statistics and Funding Team are responsible for ensuring that suitable data disclosure controls have been applied where appropriate when providing aggregate data internally or externally, for providing advice on data disclosure controls, and for maintaining the Data Disclosure Control Policy and Procedure.

Supporting documentation

9. The Data Disclosure Control policy is supported by the following:
 - Data Disclosure Control Procedure
 - Data Protection Policy
 - Data Protection Act Guidance
 - Data Sharing Guidance

| Version | Date | Comment |
|---------|------------|---|
| 0.1 | 13/10/2016 | First draft |
| 0.2 | 31/10/2016 | Second draft |
| 0.3 | 25/11/2016 | Amendments post Funding Group |
| 1.0 | 06/12/2016 | Approved by Management Board |
| 1.1 | 06/01/2017 | Amendments post impact assessment screening |

Data Disclosure Control Procedure

Introduction

1. This procedure details how and when HEFCW's data disclosure control methodology must be applied to aggregated data about individuals.
2. HEFCW has adopted a data disclosure control methodology based on the HESA Services standard rounding methodology. The methodology is outlined in Annex A and examples of applying the methodology can be found in Annex C. Advice on how to apply the methodology using SAS is in Annex D. SAS is software used by the Statistics and Funding Team for data manipulation and analysis. As part of the procedure, HEFCW reserves the right to use its discretion regarding when to apply this methodology and when to apply suitable alternative data disclosure controls. When this might occur is explained in more detail below.
3. This data disclosure control procedure applies to all data about individuals held or used in HEFCW.
4. Examples of individual data held and used in HEFCW include:
 - HESA student and staff data held and used by the Statistics and Funding Team
 - HESA student and staff unrounded data downloaded by Heidi plus Gold users
 - HEFCW staff data
 - GO Wales data
5. The procedure is effective from December 2016.
6. The data disclosure control methodology will be applied in all cases relating to aggregated data about individuals, unless a reason not to can be identified.

Data previously published

7. Aggregated data already created, shared or published will not be updated with the new methodology unless part of a new analysis.
8. All data in new analyses which include historic data will be subject to the data disclosure control procedure.

Sharing data externally

9. When sharing aggregated data about individuals, data disclosure controls will always be applied, other than where there is a specific reason not to, as described in paragraphs 12 and 16.
10. When data which has been subjected to data disclosure controls is displayed on our website, intranet or extranet, used in a spreadsheet or in a table in a document, or a single figure is quoted in a document, a footnote must be inserted

Data Disclosure Control Procedure

stating that HEFCW's data disclosure control procedure (or the alternative data disclosure control used) has been applied. A standard footnote, to be included when HEFCW's data disclosure control procedure is applied, can be found in Annex E.

11. When sharing with an institution its own data, the data shall not be subjected to any data disclosure controls.
12. If, after discussion, it is agreed that there is a need to share the data in its original form with an external requester, then the requester will be required to enter into a data sharing agreement with HEFCW. In this case, the reason not to apply the data disclosure control methodology to the data supplied to the external requester should be recorded.

Sharing data internally

13. In many cases aggregated data about individuals that is shared internally or externally originates from the Statistics and Funding Team. If a member of staff (i.e. internal request) requests data about individuals from the Statistics and Funding Team then the Team will either
 - routinely provide the data with the data disclosure control methodology applied
 - provide the data in its original form with no data disclosure control methodology applied, if no external use will be made of the data and a preference has been expressed for unsuppressed data
 - provide the data in its original form with no data disclosure control methodology applied, if timing is an issue and the data is needed as soon as possible for internal use. Data with the data disclosure control methodology applied will be made available subsequently as appropriate.
14. If data is provided with no data disclosure control methodology applied, then the data will be clearly labelled as 'Non-disclosable' e.g. in a watermark, header or footer. The staff member making the request will be told not to disclose the data externally and to seek assistance from the Team prior to any sharing of data externally. This instruction will also be provided in the table of data as a footnote. The staff member should also ensure that the data is stored securely and that appropriate measures are adopted to protect the data from unauthorised access.
15. Other teams in HEFCW also share data about individuals both internally and externally, and must apply this data disclosure control procedure to their data. Whilst discretion may be used when sharing data internally, care must be taken when sharing data externally, and all staff providing data about individuals are responsible for ensuring that the data disclosure control methodology has been applied appropriately before any data about individuals is shared externally. If in

Data Disclosure Control Procedure

doubt about how or when to apply the data disclosure control methodology, staff can seek advice from the Statistics and Funding Team.

Using alternative data disclosure control methodologies

16. The data disclosure control methodology stated in this procedure should be used in the first instance. If it is not used then the reasons for this must be documented, alternative data disclosure controls used where necessary and the risk of data disclosure assessed. Examples of when a decision may be made not to apply the HEFCW data disclosure control methodology include:

- Data used in HEFCW's funding allocations is published in the interest of transparency and enables readers to understand how the allocations have been calculated. Rounding this data would not be in the interests of transparency.
- Corporate Strategy targets 2013/14 to 2016/17 and the data used to monitor progress towards them were not subjected to data disclosure controls. This data disclosure control procedure has been introduced whilst these targets are still being monitored. It would be difficult and confusing to start rounding data part way through this monitoring.

17. Alternative data disclosure controls include:

- presenting small cell values less than 5 as * and other cells as + to prevent recalculation of the *s,
- using alternative aggregations or
- excluding certain groups of data from the overall analysis.

18. The Statistics and Funding Team will maintain a record of when decisions have been taken to use either an alternative, or no data disclosure control procedure (refer to paragraphs 12 and 16). All staff who make a decision to either use an alternative, or no data disclosure control procedure must notify a member of the Statistics and Funding Team.

Publication of data

19. When data is published by HEFCW (including by all those listed in paragraph 1 of the Data Disclosure Control Policy), or data provided by HEFCW is published by external requesters, the data disclosure control methodology must have been applied unless otherwise exempt as described in paragraph 16. In particular, if data that has not had the data disclosure control methodology applied has been supplied to an external requester as part of a data sharing agreement, any aggregations of the supplied data published by that external requester must have the HEFCW data disclosure control methodology applied.

Data Disclosure Control Procedure

Data sharing agreements

20. Any new data sharing agreements with external organisations should refer to this procedure and existing data sharing agreements will be updated to refer to this procedure. Where appointment of external consultants or contractors includes the provision of data to them by HEFCW, the procurement process will include commitment by the consultants or contractors to adhere to our data disclosure control methodology.

Data this procedure applies to

21. This data disclosure control procedure relates to aggregated data about individuals, it does not relate to other types of data such as financial data.

Further advice

22. The Statistics and Funding Team can provide advice on how and when the data disclosure control methodology should be applied.

Annexes

23. Annexes to the Data Disclosure Control Procedure are:

- Annex A HEFCW data disclosure control methodology
- Annex B FTE and FPE explained
- Annex C Examples of applying the HEFCW data disclosure control methodology
- Annex D SAS code for applying the HEFCW data disclosure control methodology
- Annex E Standard data disclosure control footnote

Supporting documentation

24. The Data Disclosure Control Procedure is supported by the following:

- Data Disclosure Control Policy
- Data Protection Policy
- Data Protection Act Guidance
- Data Sharing Guidance

| Version | Date | Comment |
|---------|------------|---|
| 0.1 | 13/10/2016 | First draft |
| 0.2 | 31/10/2016 | Second draft |
| 0.3 | 25/11/2016 | Amendments post Funding Group |
| 1.0 | 06/12/2016 | Approved by Management Board |
| 1.1 | 06/01/2017 | Amendments post impact assessment screening |

HEFCW data disclosure control methodology

1. HEFCW's Data Disclosure Control Procedure uses the methodology described below. This is based on the [HESA Services standard rounding methodology](#) in order to be consistent with both HESA and other organisations using this methodology.
2. The methodology involves rounding all numbers to the nearest multiple of 5 and suppressing percentages and averages based on small populations.
3. HEFCW data disclosure control methodology:
 - a) All numbers are rounded to the nearest multiple of 5
 - b) Any number lower than 2.5 is rounded to 0
 - c) Halves are always rounded upwards (e.g. 2.5 is rounded to 5)
 - d) Percentages based on fewer than 22.5 individuals are suppressed
 - e) Averages based on 7 or fewer individuals are suppressed
 - f) The above requirements apply to headcounts, counts of enrolments, registrations, full person equivalents (FPEs) and full-time equivalents (FTEs)
 - g) Financial data and data that does not relate to individuals, such as numbers of days delivered, numbers of spin-outs and number of contracts, are not rounded
4. Total figures are also subject to this rounding methodology after calculation; so the sum of numbers in each row or column may not match the total shown. Suppressed values are normally represented as '.' in published tables.
5. FTEs and FPEs are explained in Annex B.
6. Examples of applying the HEFCW data disclosure control methodology are in Annex C.

What is FTE?

1. For staff, full-time equivalent compares an individual's workload to a standard full-time, full-year workload. For students, full-time, full year students would normally be returned as 1.0 and part-time students returned as a proportion of an equivalent full-time course (further guidance is provided in the link below).
2. A full-time student is 1.0 FTE for a year of study. A member of staff is 1.0 FTE working full-time for a year. A student on a part-time course that is 60% of a full-time course would be 0.6 FTE. A member of staff working 2.5 days a week for the year would be 0.5 FTE.
3. Because FTE calculations look at a typical full-year workload, a member of staff working full-time for six months of the year would be 0.5 FTE. A member of staff working 2.5 days a week for six months of the year would be 0.25 FTE.

What is FPE?

4. Full-person equivalent (FPE) looks at how much of the (whole) person's time is engaged in a particular activity (and is not related to FTE in anyway). For students the activity is in each subject area, for staff the activity is in each cost centre.
5. A student example:

Student A is studying a joint course with equal amounts of Mathematics and English. They are represented as 0.5 FPE in each subject.

Student B is studying Mathematics (50%) with Physics (25%) and French (25%). They are represented as 0.5 FPE in Mathematics, and 0.25 FPE in Physics and in French.

Student C is only studying French. They are represented as 1 FPE in French.

Student B is are studying full-time but student A is studying part-time. Student A's enrolment is still counted as 1 and split equally between subjects giving 0.5 FPE in Mathematics and 0.5 FPE in English.

6. For more information on FTE and FPE refer to the HESA website:

Staff

<https://www.hesa.ac.uk/support/definitions/staff#staff-full-time-equivalent>

<https://www.hesa.ac.uk/support/definitions/staff#staff-full-person-equivalent>

https://www.hesa.ac.uk/collection/c15025/fte_vs_fpe/

Students

<https://www.hesa.ac.uk/collection/c14051/fte/>

<https://www.hesa.ac.uk/support/definitions/students#count-fpe-fte>

Applying the HEFCW data disclosure control methodology

Examples

1) Original data (which will not be published)

| Age | Student enrolments | Percentages |
|--------------|--------------------|-------------|
| 20 and under | 18 | 40.9% |
| 21 to 24 | 12 | 27.3% |
| 25 to 29 | 7 | 15.9% |
| 30 and over | 6 | 13.6% |
| Unknown | 1 | 2.3% |
| Total | 44 | 100.0% |

2) Original data (which will not be published)

| Age | Student enrolments ^c | | Percentages | |
|--------------|---------------------------------|-----------|-------------|-----------|
| | Subject A | Subject B | Subject A | Subject B |
| 20 and under | 9 | 9 | 40.0% | 41.9% |
| 21 to 24 | 6 | 6 | 26.7% | 27.9% |
| 25 to 29 | 3.5 | 3.5 | 15.6% | 16.3% |
| 30 and over | 3.5 | 2.5 | 15.6% | 11.6% |
| Unknown | 0.5 | 0.5 | 2.2% | 2.3% |
| Total | 22.5 | 21.5 | 100.0% | 100.0% |

1) Data to be published (with data disclosure control methodology applied)

| Age | Student enrolments | Percentages ^b |
|--------------|--------------------|--------------------------|
| 20 and under | 20 | 40.9% |
| 21 to 24 | 10 | 27.3% |
| 25 to 29 | 5 | 15.9% |
| 30 and over | 5 | 13.6% |
| Unknown | 0 | 2.3% |
| Total | 45 ^a | 100.0% |

2) Data to be published (with data disclosure control methodology applied)

| Age | Student enrolments | | Percentages ^b | |
|--------------|--------------------|-----------------|--------------------------|------------------------|
| | Subject A | Subject B | Subject A | Subject B ^d |
| 20 and under | 10 | 10 | 40.0% | . |
| 21 to 24 | 5 | 5 | 26.7% | . |
| 25 to 29 | 5 | 5 | 15.6% | . |
| 30 and over | 5 | 5 | 15.6% | . |
| Unknown | 0 | 0 | 2.2% | . |
| Total | 25 | 20 ^a | 100.0% | . |

3) Original data (which will not be published)

| Age | NETFEE | | | | Average NETFEE £ |
|--------------|--------|-------|-------|-------|------------------|
| | £7550 | £8000 | £8500 | £9000 | |
| 20 and under | 1 | 0 | 1 | 16 | 8,892 |
| 21 to 24 | 3 | 0 | 0 | 9 | 8,638 |
| 25 to 29 | 0 | 1 | 0 | 6 | 8,857 |
| 30 and over | 0 | 1 | 2 | 3 | 8,667 |
| Unknown | 0 | 0 | 1 | 0 | 8,500 |
| Total | 4 | 2 | 4 | 34 | 8,777 |

3) Data to be published (with data disclosure control methodology applied)

| Age | Average NETFEE ^e £ |
|--------------|-------------------------------|
| 20 and under | 8,892 |
| 21 to 24 | 8,638 |
| 25 to 29 | . |
| 30 and over | . |
| Unknown | . |
| Total | 8,777 |

Notes

All data is illustrative only

- a. Example of where the original total is rounded and does not equal the sum of the individual rounded data
- b. Percentages are calculated on original data not rounded data
- c. Example of FPEs - all students are doing 50% subject A and 50% subject B except for one aged 30 and over who is doing 100% subject A
- d. Percentages based on fewer than 22.5 individuals are suppressed
- e. Averages based on 7 or fewer individuals are suppressed

SAS code for applying the HEFCW data disclosure control methodology

1. This annex provides advice on how to apply the HEFCW data disclosure control methodology using SAS. This advice is based on producing a final dataset containing only the data to be shared. Variables which have been used to calculate the final dataset but which are not to be shared are removed from the final dataset.
2. In practice, the coder may wish to retain these variables for checking purposes. Therefore, it may be good practice to output two datasets: the final dataset for sharing, as described below and, a final dataset retaining all variables used for checking purposes only and not to be shared.

Options statement

3. An options statement is used to set missing data to '.'. This is used to ensure that the suppressed data are shown correctly when the data are exported from SAS.

```
OPTIONS missing=.;
```

Rounding and percentages

4. The data in dataset 'example2' is individualised student data which has been aggregated into the summary statistics shown in example 2 of Annex C; the number of FPEs in subject A are contained in a variable named subja and the number of FPEs in subject B are contained in a variable named subjb. Overall total FPEs in subject A are contained in the variable subja_tot, and overall total FPEs in subject B are contained in the variable subjb_tot.

Dataset example2

| agegrp | subja | subjb | subja_tot | subjb_tot |
|--------------|-------|-------|-----------|-----------|
| 20 and under | 9 | 9 | 22.5 | 21.5 |
| 21 to 24 | 6 | 6 | 22.5 | 21.5 |
| 25 to 29 | 3.5 | 3.5 | 22.5 | 21.5 |
| 30 and over | 3.5 | 2.5 | 22.5 | 21.5 |
| Unknown | 0.5 | 0.5 | 22.5 | 21.5 |
| Total | 22.5 | 21.5 | 22.5 | 21.5 |

5. The macro 'round1' inputs a variable, rounds it to the nearest five and creates a new variable with the same name but with an a appended to contain the rounded number, thus for the variable subja with a value of 9, the new variable subjaa will contain 10.

```
** rounding macro **;  
%macro round1(var);  
&var.a=round(&var,5);  
%mend;
```

6. A new dataset 'example2a' is created from the dataset 'example2'. In this dataset, percentages are calculated on unrounded data and are contained in the variables psubja and psubjb. The macro is employed to create rounded data which is contained in the variables subjaa and subjba. If any of the percentages have been calculated based on a population of less than 22.5 individuals they are set to '.'. Variables that are not to be contained in the final data are dropped, these are the original unrounded data subja and subjb and the totals subja_tot and subjb_tot which are no longer needed.

```

data example2a(drop=subja subjb subja_tot subjb_tot);
  set example2;

  ** calculate percentages on unrounded data **;
  psubja=subja/subja_tot;
  psubjb=subjb/subjb_tot;

  ** apply HEFCW data disclosure control methodology **;
  * numbers rounded *;
  %round1(subja);
  %round1(subjb);

  * percentages based on lt 22.5 are suppressed *;
  if subja_tot < 22.5 then psubja=.;
  if subjb_tot < 22.5 then psubjb=.;

run;

```

Data example2a

| agegrp | subjaa | subjbb | psubja | psubjb |
|----------------|--------|--------|--------|--------|
| 1 20 and under | 10 | 10 | 40.0% | . |
| 2 21 to 24 | 5 | 5 | 26.7% | . |
| 3 25 to 29 | 5 | 5 | 15.6% | . |
| 4 30 and over | 5 | 5 | 15.6% | . |
| 5 Unknown | 0 | 0 | 2.2% | . |
| 6 Total | 25 | 20 | 100.0% | . |

Averages

7. The data in dataset 'example3' is individualised student data which has been aggregated into summary statistics similar to those shown in example 3 of Annex C; the number of student enrolments are contained in a variable named num and the average net fee is contained in a variable named ave.

Dataset example3

| agegrp | num | ave |
|--------------|-----|-------|
| 20 and under | 18 | 8,892 |
| 21 to 24 | 12 | 8,638 |
| 25 to 29 | 7 | 8,857 |
| 30 and over | 6 | 8,667 |
| Unknown | 1 | 8,500 |
| Total | 44 | 8,777 |

8. A new dataset 'example3a' is created from the dataset 'example3'. In this dataset, if any of the averages have been calculated based on a population of 7 or fewer individuals they are set to '.'. Variables that are not to be contained in the final data are dropped, these are the number of enrolments num which are no longer needed.

```
data example3a(drop=num);  
  set example3;  
  
  ** apply HEFCW data disclosure control methodology **;  
  ** averages based on 7 or fewer are suppressed **;  
  if num le 7 then ave=.;  
  
run;
```

Dataset example3a

| agegrp | ave |
|--------------|-------|
| 20 and under | 8,892 |
| 21 to 24 | 8,638 |
| 25 to 29 | . |
| 30 and over | . |
| Unknown | . |
| Total | 8,777 |

Standard data disclosure control footnote

1. Paragraph 10 of the Data Disclosure Control Procedure states that when data have been subjected to data disclosure controls, a footnote must be inserted stating that HEFCW's data disclosure control procedure (or the alternative data disclosure control used) has been applied.
2. A standard footnote is provided below, which is to be included with data where HEFCW's data disclosure control procedure has been applied. The relevant points should be selected, depending on the analysis, as not all points will be applicable in all cases.

HEFCW's standard data disclosure control procedure has been applied. This means that:

- All numbers are rounded up or down to the nearest multiple of 5. Any number lower than 2.5 is rounded to 0. Halves are rounded upwards (e.g. 2.5 is rounded to 5).
 - Percentages are calculated on unrounded data. Percentages calculated on populations which contain fewer than 22.5 individuals are suppressed and represented as ".".
 - Averages based on 7 or fewer individuals are suppressed and represented as ".".
3. The above footnote should be amended as appropriate when alternative disclosure control methods are used.