

Impact of increasing the suppression threshold to 52.5

UKPITG 13/11

Issue

1. Impact assessment of increasing the suppression threshold from 22.5 to 52.5 in the UK Performance Indicators (UKPI) tables. Should the UKPIs be brought in line with HESA's standard and increase the suppression level, or should they remain in line with other standards, such as UNISTATS.

Recommendation

2. Suppression threshold in the UKPIs should remain at 22.5.

Discussion

3. Currently within the UKPIs, indicators and benchmarks are suppressed where the denominator is less than 22.5. What would be the impact on the UKPI tables if this suppression level was changed to 52.5?

4. Below summarises the impact of changing the suppression levels from 22.5 to 52.5 in the 2011/12 UKPI tables T1, T2, T3, T5 and E1a-E1d. Number of institutions indicates the additional number of institutions to be suppressed if this change were to be made. Where there is high impact, a percentage has been given to indicate the total proportion of rows of data which would be suppressed at the 52.5 level.

Table	Indicator	Number of institutions	Notes
T1a	All indicators	3	Minimal impact
T1b	All indicators	3	Minimal impact
T1c	State school	13	Moderate impact and no data remains for NI
T1c	NS-SEC	12	Moderate impact and no data remains for NI
T1c	Low participation	14	Moderate impact and no data remains for NI
T2a (mature FT UG)	First degree	7	Minimal impact
T2a (mature FT UG)	Other undergraduate	6	Minimal impact
T2b (PT UG)	Young	21	High impact and 1 institution remains in NI (61% total suppressed)
T2b (PT UG)	Mature / All	8	Moderate impact
T2c (mature FT)	Low participation	24	High impact and 1 institution remains in NI (50% total suppressed)

OUG)			suppressed)
T3a (FT FD)	Young	3	Minimal impact
T3a (FT FD)	Mature	11	Moderate impact
T3a (FT FD)	All	2	Minimal impact
T3b (young FT FD)	Low participation	6	Moderate impact
T3b (young FT FD)	Other	3	Minimal impact
T3c (mature FT FD)	Previous HE	17	Moderate impact
T3c (mature FT FD)	No previous HE	11	Moderate impact
T3d (FT OUG)	Young	13	Moderate impact and no data remains for NI
T3d (FT OUG)	Mature	22	High impact and no data remains for NI (48% total suppressed)
T3d (FT OUG)	All	19	Moderate impact and 1 institution remains in NI
T3e (PT FD)	30 & under	25	High impact and 1 institution remains in NI (69% total suppressed)
T3e (PT FD)	Over 30	31	High impact (63% total suppressed)
T3e (PT FD)	All	22	High impact (53% total suppressed)
T5 (projected)	All indicators	3	Minimal impact
E1a (FT FD)	All indicators	3	Minimal impact
E1b (PT FD)	All indicators	17	Moderate impact (52% total suppressed)
E1c (FT OUG)	All indicators	9	Moderate impact
E1d (PT OUG)	All indicators	8	Moderate impact

5. Data suppression in the UKPIs is intended to reduce the risk of identifying individuals from the published data. The standard threshold for percentages in HESA outputs is a population of 52 or fewer. The HESA standard threshold is designed to be applied to any novel extract with the same confidence of minimal risk. The possibly higher risk of a lower threshold for UKPIs is mitigated by their predictability – the same indicators are calculated each year and novel cross-tabulations are not possible with the published data.

6. Whether the population threshold is set at 22.5 or 52.5, or any value in between, it is possible in theory to deduce information that describes one individual. However this does not

mean that one could necessarily identify who that individual is and learn new information about them. The risk is described, with examples, in the paper Assessment of Data Protection risk in PIs (PITG 13/02), whose conclusions are summarised here.

7. It is possible for a case to arise in the UKPIs where one can deduce that a single individual at an institution meets a particular description. However these cases only arise if:

- There is in fact only one individual,
- The combination of rounded counts and the percentage calculation reveal only one possible combination of un-rounded raw figures,
- Those figures differ by one,
- It can also be deduced that there are 0 unknowns.

8. This combination of factors is not likely to occur often, and if it does one still cannot learn new information about an individual. It is also worth considering whether anyone would have the desire or motivation to make this deduction if such a combination did arise. The likelihood of this combination of factors occurring is marginally reduced by using a higher threshold population for suppression because smaller populations are more likely to have only one individual that meets a particular description. However the combination of factors can still occur with population of more than 52.

9. The likelihood of there being, in fact, only one individual who meets specific UKPI criteria within an institution's UKPI population varies according to the UKPI criteria. Very roughly the overall UK proportions of students in each UKPI minority group are as follows:

- | | |
|-----------------------|-----|
| • Independent schools | 12% |
| • NS-SEC 4-7 | 30% |
| • LPN (POLAR 2) | 10% |
| • DSA | 5% |
| • Non-continuation | 7% |
| • Unemployed | 10% |

10. Therefore it is fairly likely that a population of 23 will have exactly one DSA recipient or one non-continuer, but probably more than one of each other category. There may therefore be an argument for raising the percentage threshold for DSA and non-continuation UKPIs, but not for the others. However, as stated above the fact that one individual may be deduced from the counts does not in itself present a data protection problem as no new information can be deduced about that individual.

Recommendation: Retain the current suppression levels within the UKPIs as it allows more data to be displayed. Increasing the suppression levels to 52.5 would not offer any significant data protection benefits. The suppression level of 22.5 is already used within UNISTATS, so consistency with other HE statistic still remains.

Further information

For further information contact Suzie Dent (tel: 01242 211109; e-mail: suzie.dent@hesa.ac.uk)