Technical notes on our planned approach to statistical uncertainty and publication thresholds

Purpose

1. The Office for Students (OfS) has issued a consultation about the publication of the National Student Survey (NSS) results.[1] This document has been published as supporting information alongside the consultation to assist stakeholders in understanding the statistical methods we have proposed to use in the presentation and contextualisation of the NSS results. We anticipate that some readers of the consultation, particularly those with in-depth knowledge of statistical methods, will find this document useful for understanding the practical effects of implementing these changes.

2. This document describes the implementation we would expect based on the general principles proposed in the consultation document. This means descriptions in this document may be impacted by the outcomes of the consultation.

Communicating statistical uncertainty in the NSS

3. The approach currently used for the UK-wide NSS data published on the OfS website is as follows:

    a. For the provider-level data, we calculate a single set of confidence intervals.

    b. For the sector-level data, we calculate a range of confidence intervals between 50 per cent and 99 per cent and allow users to choose their confidence level.

    c. For both sets of data we produce confidence intervals using the Wilson method, for both the agreement rate and the difference from benchmark.

    d. For both sets of data, we account for multiple comparisons using a Bonferroni correction, and flag differences from benchmark which are significant at the respective confidence level.

4. In implementing the general policy proposed in the consultation document we would continue calculating and presenting statistical uncertainty for the UK-wide NSS data published on the OfS website, but would make some changes to our current approach.

5. The new approach we would expect to implement draws on the research and expert advice used to inform the consultation 'Constructing student outcomes and experience indicators for

---

use in OfS regulation'.[2] As such, it is fully consistent with the approach developed through that consultation and implemented on OfS dashboards.[3] For further detailed information on the approach see 'Supporting information about constructing student outcome and experience indicators for use in OfS regulation – Description of statistical measures'.[4] The rest of this section summarises the changes we would make when calculating and displaying statistical uncertainty.

Calculating statistical uncertainty

6. We would change our calculations as follows:

    a. Use the Jeffreys interval[5] to create confidence intervals. This is because the Jeffreys interval method has been shown to perform well both when the sample size is small and when the observed outcome is close to 0 per cent or 100 per cent.[6] This behaviour is important because NSS statistics often have one or both of these challenges.

    b. Calculate the proportion of the uncertainty distribution that falls materially above and below the benchmark (that is, more than 2.5 percentage points above, or more than 2.5 percentage points below, respectively).

    c. No longer account for multiple corrections using the Bonferroni correction. This is a correction which has been made to published NSS confidence intervals to reduce the risk, due to the amount of data which can be compared, of too many differences appearing significant. We know that there are a wide variety of users of the published NSS data, with a range of different needs, so any correction to the confidence intervals would at best be approximate and could have a disproportionate effect for some users. Further, the approach to displaying statistical uncertainty, outlined below, means that users would be able to see the distribution of statistical uncertainty, rather than relying on a single confidence interval or significance test.

Presenting statistical uncertainty

7. We would also change the way we display statistical uncertainty:

    a. No longer show a single confidence interval; instead we would show a range of upper and lower confidence limits. This range of confidence intervals would better show the uncertainty around the observed value, and allow users to make their own judgement using the extra information available.

---

[2] Available at www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/outcome-and-experience-data/.

[3] For instance, see www.officeforstudents.org.uk/data-and-analysis/tef-data-dashboard/data-dashboard/.

[4] Available at www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/outcome-and-experience-data/.

[5] Jeffreys, Harold (1946). An invariant form for the prior probability in estimation problems. Proc. Royal Society, London. A186453–461. http://doi.org/10.1098/rspa.1946.0056.

[6] Brown et al (2001). Interval estimation for a binomial proportion Statistical Science. Vol. 16, No. 2, pages 101-133. http://dx.doi.org/10.1214/ss/1009213286.

b. Display the proportion of the uncertainty distribution falling materially above and below the benchmark (that is, more than 2.5 percentage points above, or more than 2.5 percentage points below, respectively).

c. No longer flag whether figures are 'significantly' different from their benchmark. The range of confidence intervals displayed would help users better understand the uncertainty around the observed difference from benchmark and avoid the use of a 'cliff-edge' figure which could be overly simplistic.

Suppressions

8. The consultation document outlines our proposed approach to publication thresholds. There is one specific aspect which is explained in more detail here, which relates to the suppression of data in order to protect respondent confidentiality.

9. In general, our proposed approach to numeric thresholds (which for 2023 we expect would mean continuing to use a threshold of 10 students) protects the confidentiality of respondents. However, there are unusual cases where this is not the case. In particular, this is true of a case in which all the students in the relevant population responded to the survey and all of them gave the same response to a particular question. In this case, anyone who knows which students are in the population would know how each student responded to the question.

10. To avoid disclosing information about individuals in these circumstances we would introduce a new 'data protection' publication threshold: results are published only if the response rate is less than 100 per cent or the responses are not unanimous. Results would also be suppressed in instances where the response rate is 100 per cent and only one student responds differently. This would allow us to indicate whether a suppression is made for responses which are overwhelmingly positive or overwhelmingly negative.

11. Note that for this purpose we are not distinguishing between different types of positive response, or different types of negative response. For instance, if out of 10 respondents, five responded 'Well' and five 'Very well', that would count as 10 respondents giving positive responses, and could lead to suppression if there was a response rate of 100 per cent.

12. For cases which need suppressing due to the new data protection threshold we would suppress values relating to that 'row', for example that combination of provider, mode, level and subject. Specifically we would:

a. Suppress all values for that category except for those relating to the number of students in the population and the number of responses,

b. Show an indicator for why the data is suppressed and whether the value was high or low:

    i. [DPH] for measures close to 100 per cent ('data protection suppression, high value'), or

    ii. [DPL] for measures close to 0 per cent ('data protection suppression, low value').

13. Note that we would also expect to apply this new approach to the provider dissemination portal.

Statistical uncertainty and suppressions on Discover Uni

14. Uncertainty measures, such as confidence intervals or significance flags, are not currently included on Discover Uni and we would expect this to remain the case. User testing has shown various approaches to communicating uncertainty to be ineffective for the users of Discover Uni, who tend to be prospective students. Instead we currently offer guidance, in various formats, about how best to interpret the Discover Uni metrics, and we would continue to develop these. We would also continue to ensure that relevant information, such as population size and response rate, are prominently displayed alongside the headline Discover Uni metrics.

15. We are expecting to continue to apply the numeric threshold of 10 students to the data presented on Discover Uni, and not to apply any further suppression for the sake of respondent confidentiality. The reason for this is that the risk of identification of individuals is very low, and further suppression would not substantially reduce the chance of individuals being identified but would make the data more difficult to interpret and understand for prospective students. This is particularly the case for the Discover Uni website because of the structure – each course as displayed on the website uses data from a number of different historic courses.