Publication and Risk of Disclosure – Indicative Health Visitor (IHVC) workforce data

Coverage

This paper assesses confidentiality and data disclosure issues for the publication of monthly IHVC workforce data.

Background

1. The Government's statistical services have a professional duty to protect the confidentiality of individual level data obtained to produce statistics. The Code of Practice for Official Statistics sets this out in Principle 5: "Private information about individual persons (including bodies corporate) compiled in the production of official statistics is confidential and should be used for statistical purposes only". The Code of Practice also states arrangements for confidentiality protection should be sufficient to protect privacy but not so restrictive as to limit unduly the practical utility of statistics. The main legal instruments governing this balance are the Data Protection Act, which places obligations on organisations to protect personal information and the Freedom of Information Act, which creates a public right of access to information.

2. The design of a statistic should meet the obligation to protect against disclosure, but should then be optimised to include as much detail in the statistic as reasonably possible, to fully meet the needs of the users.

3. There is a need to assess whether this data is potentially disclosive.

Guidance from ONS – the structure of this assessment

4. Guidance from ONS[1] on confidentiality sets out guidelines for any assessment of disclosure risk. It stops short of setting out hard and fast rules, but is clear on the need to protect confidentiality while at the same time maximising public access to official data. This guidance summarises the six main steps for ensuring access to non-disclosive statistics as shown in Figure 1.

[1] GSS/GSR Disclosure Control policy for tables produced from administrative data sources (the document is available at the following link: http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-policy-for-tables/index.html)

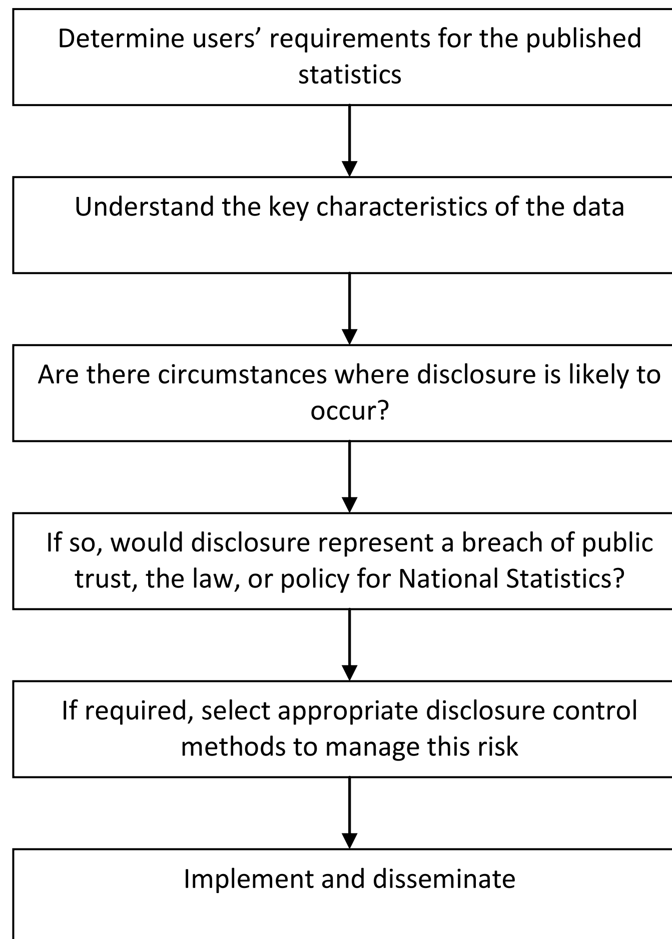

Figure 1: Main steps for ensuring access to non-disclosive statistics

Step 1 – Determining users' requirements

5. The requirements for this data are set out on the Health Visitor Workforce web page published on NHS England's Statistics page. This includes details of the purpose of the collection and highlights its intended use.

6. A collection was established using UNIFY2, to collect this data on a monthly basis from all English providers that are commissioned to provide a health visiting service.

7. The collection allows members of the public and those working within the system to have access to up-to-date information in a timely way.

8. The data are published to give the public an insight into the numbers of health visitors employed at a regional and national level and how this changes month on month.

9. The data fields covered in the IHVC are included in a separate 'Definitions' document.

10. There is a requirement to ensure that information about the experience of individuals is safeguarded in an appropriate way. A balance must be struck between measures to protect confidentiality and the public good arising from publication.

Step 2 – The characteristics of the data

11. This is an aggregated data source. The data are submitted by providers based mainly on workforce information taken from administrative data sources within the organisation.

12. There is a process of basic data cleaning and validation within the collection system over a short timeframe: The Area Teams check the data of each provider for workforce turnover (FTE and HC) being equal to the difference of starters and leavers. Also checked is that the number of student starters is in general agreement with what was forecast in the planning for 2014/15.

13. As the main purpose of the data collection is for management information. It is used by managers to take early action on managing the workforce, the most important factor in this data collection is the speed at which the data are collected, validated and published. The official health visitor workforce dataset (HV MDS) is published by the Health and Social Care Information Centre (HSCIC). http://www.hscic.gov.uk/healthvisitors. To retrieve the data go to the 'find data' web page at http://www.hscic.gov.uk/searchcatalogue and search for 'HV MDS' with the current year

14. Revisions policy. Once the collection has closed, about two weeks after the end of the reporting period, the data are prepared for publication and are published a week later. No revisions are made to the data once the collection has closed. The official health visitor workforce dataset (HV MDS) published by the HSCIC undergoes a more rigorous QA process. Provisional monthly data are published about two months later than the IHVC. The HSCIC also publish annual workforce data as part of their data catalogue at http://www.hscic.gov.uk/searchcatalogue, and search for 'NHS workforce statistics'.

15. Data Quality issues. In response to the need for more timely and detailed data, the Indicative Health Visitor Collection (IHVC) was developed and started collecting data in April 2014. This has since been collected on a monthly basis, alongside the HV MDS.

    The IHVC provides a useful early indication of overall health visitor workforce totals, however, data quality issues arise as a result of:

    - the IHVC being a new data collection with limited validation and quality assurance
    - manual data collection to supplement the data collected on the Electronic Staff Record (ESR) system

    Known data quality issues:

    - Manual recording of starters and leavers in-month, are in some cases inconsistent with the overall changes in the workforce when comparing aggregates between months.
    - Some data definitions having been developed over time as a result of better understanding of workforce dynamics. For example, changes in FTE by individual health visitors are no longer captured as starters or leavers to an

organisation. So, an increase in the number of hours worked by established health visitors is no longer included in the starters section, or likewise a decrease in the leavers section.

- The 'staff in administrative processing' is indicative of the number of new Health Visitors working in the system but not yet recorded on ESR. These HVs are not included in the HV MDS until they are officially on the system.
- The HV MDS includes all Health Visitors who are coded as such on the ESR, whether or not they are commissioned by the NHS Area Teams. The IHVC focuses on those providers where health visitors are directly commissioned by the Area Teams.
- In a small number of providers, recording of some student starters has been based on expectation rather than actual number of starters in a month, inflating the workforce figures for that month. This has been followed by recording of these starters again in the actual month they joined the workforce.

16. The majority of the aggregated data does not present a risk of disclosure, but depending on provider and area team size and general workforce numbers and movement within the workforce, a number of individual data items at provider and area team level return small numbers. Therefore, data are not being published at provider level and limited data are being published at area team level.

Step 3 – Evidence of risk of disclosure

17. Publication of any detailed data may increase risks of disclosure of information relating to an individual health visitor. It is important to note that these data do not include any personal identifiers, so it is not possible to identify health visitors directly from the published data. The categories of disclosure risk (situations in which disclosure might arise) are as follows:
    - Self-identification risk: When a health visitors recognises themselves, in the data.
    - Motivated intruder risk: Where there are reasons for a third party to seek further information about a health visitor, for example where it is possible to work out an individual's working hours or when they joined or left an organisation.

18. There may be circumstances where a health visitor can be identified if detailed data were published at a provider or area team level. Publishing tables containing small numbers is not in itself a reason for suppressing data. Under the Data Protection Act 1998, there is a need to confirm that the published data would not cause, or be likely to cause, unwarranted and substantial damage or distress.

Step 4 – Preventing disclosure

19. In order to comply with the data protection act and avoid identification of individuals or small groups of individuals, there will be:

    - no data published at provider level
    - limited publication of data at area team level. Overall workforce numbers at area team level will be published at an aggregated level. Starter and leaver

numbers will have suppression rules applied to individual data items where disclosure remains a possibility.

- publication at regional and national level. Suppression rules are applied to individual data items where disclosure remains a possibility.

Step 5 – Suppression methodology

20. Values less than 5 are suppressed to prevent disclosure. Where the suppressed value could be disclosed by calculating the difference of the subtotal of the remaining values from a total value, an additional value (5 or larger) is also suppressed. In the case of the sub categories of new starters if either was less than 5, both are suppressed.