Risk of Disclosure – Physical Health Checks for people with Severe Mental Illness (SMI)

Coverage

This paper assesses confidentiality and data disclosure issues of the commissioner-based "Physical Health Checks for people with Severe Mental Illness" collection.

Background

1. Statisticians 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 T6[1]: "Organisations should look after people's information securely and manage data in ways that are consistent with relevant legislation and serve the public good". 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 General Data Protection Regulation and the Data Protection Act, which place obligations on organisations to protect personal information and the Freedom of Information Act, which creates a public right of access to information. Statisticians also need to act in accordance with the common law duty of confidentiality.

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[2] 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 patient 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] https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/02/Code-of-Practice-for-Statistics.pdf
[2] https://gss.civilservice.gov.uk/guidances/methodology/statistical-disclosure-control/#tables-produced-from-administrative-sources

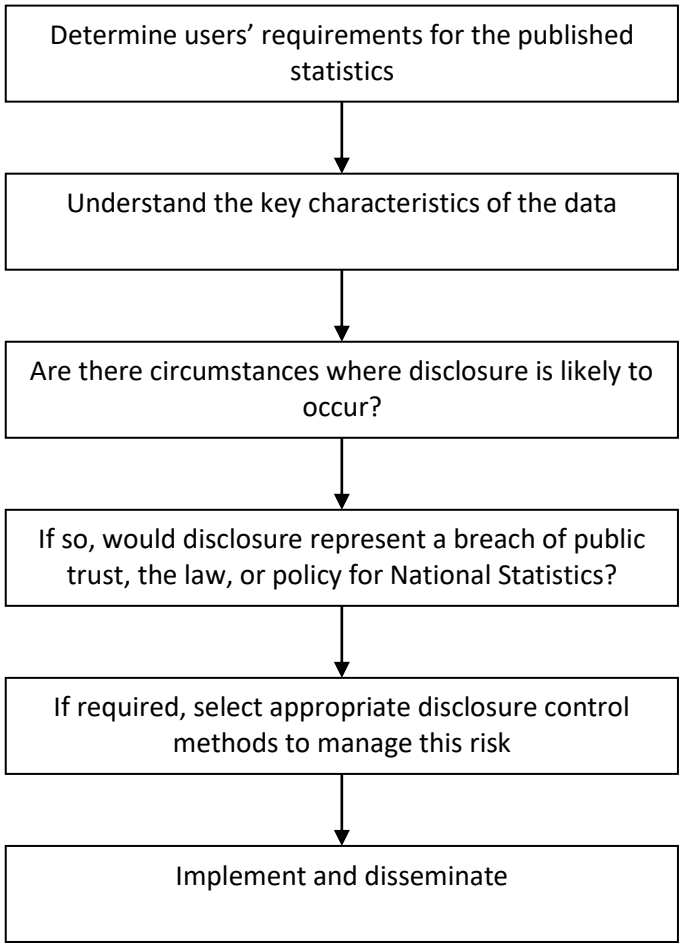

Figure 1: Main steps for ensuring access to non-disclosive statistics

Step 1 – Determining users' requirements

5. The requirements for this data were set out in **"Physical Health Checks for people with Severe Mental Illness[3]".** This material includes the standard which the collection was created to monitor, that at least 50% of patients with severe mental illness should receive a comprehensive physical health check in the primary care setting each year.

6. A collection was established using the Strategic Data Collection System (SDCS) to collect this data on a quarterly basis from all English commissioners, drawing from data collected from GPs and other sources.

7. The very first data collection is not planned to be published, but will be shared within the NHS so that information about the relative completeness of the data collection and variation in physical health checks can be reported. Data quality

[3] https://www.england.nhs.uk/mental-health/resources/smi/

allowing, the data will be published as soon as possible and is planned to be published in February 2019. When published the collection will allow members of the public and those working within the system to have access to up-to-date information. This value to users underlies the case for publishing data subject to any confidentiality constraints in a timely way.

8. **Physical Health Checks for people with Severe Mental Illness** data will be published to give patients and commissioners an insight into the performance of their local CCG, and allows them to compare against all other CCGs in England. The subjects covered in this data include:

    - **The count of people on the SMI register:** how many people are on the SMI register at the end of the reporting period

    - **The count of people to have had each, and all of, the specified health checks:** how many people on the SMI register had each of the health checks in the rolling 12-month period to the end of the reporting period. And, how many had all of the health checks in the period.

        The health checks for the first data collection are:
        1. Alcohol use
        2. Blood glucose check
        3. Blood lipid check
        4. Blood pressure check
        5. BMI / Weight check
        6. Smoking status

9. There is converse public interest in ensuring 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

10. This is an aggregated data source. The data is submitted by CCGs based on patient level information that is taken from an administrative data source within General Practices and other providers of primary care services in the area. While CCGs may have access to practice level data to collate it; the actual data supplied is aggregated to CCG before being supplied.

11. There is a process of data cleaning and validation within the collection system. The template used to collect the data prevents some data input errors, e.g. the numerators must be less than or equal to the denominator. Once received there are further checks to see whether the denominator is consistent with the SMI register already published as part of QOF[4].

12. As above, a version of the denominator is already in the public domain, albeit published annually while this data is collected quarterly. The numerators represent the number of people who have had a physical health check in the period; there is no information about the finding of the health check, just that the check was performed in the 12 months to the end of the period.

[4] https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement-prevalence-and-exceptions-data

13. The QOF mental health register is approximately 550,000 in the 2017/18 publication; ranging from a minimum of over 600 to over 14,000 per CCG. It is expected that data completeness issues will mean that the reported denominator in this data collection will be somewhat lower than this, and numerators will be lower again as not every patient will have had the required health check.

Step 3 – Evidence of risk of disclosure

14. Publication of any data may increase risks of disclosure of information relating to an individual patient. It is important to note that these data do not include any personal identifiers, so it is not possible to identify patients directly from the published data. Instead the categories of disclosure risk (situations in which disclosure might arise) are as follows:
    - Self-identification risk: When a patient recalls their circumstances during the time-period of the data collection and can recognise, from the context, which data refers to them. This would only likely cause distress within smaller counts.
    - Motivated intruder risk: Where there are reasons for a third party to seek further information about cases of a patient, for example where a 'celebrity' case arises or where cases in an organisation happen with a newsworthy frequency or pattern. This type of risk can be broken down further into two types:
        a. Identity disclosure: Where a third party is able to determine who the data relates to using the data itself and other information available to that third party.
        b. Attribute disclosure: Where a third party is able to infer additional information about an individual.

    It can be concluded that there is no risk of identity disclosure, as the possible population size of the collection is large and the collection does not contain any personal identifiers. Instead, this document focuses on the motivated intruder risk regarding attribute disclosure.

<u>Self Identification risk</u>

15. There may be circumstances where a patient can self-identify. Current published tables can contain small numbers. This is not in itself a reason for suppressing data. An appropriate test is defined by the Data Protection Act 1998, which requires the matter to be considered (although it does not directly require all self-identification to be avoided). There is a need to confirm that the published data would not cause, or be likely to cause, unwarranted and substantial damage or distress.

16. It is conceivable that a patient may identify themselves within an aggregate count. This requires recognition of primary care experiences during the time-period.

17. It is considered highly unlikely that distress would be caused by self-identification unless some sort of negative emotion is evoked from recalling the event. However, as the information does not include the findings of the health checks, it is felt this is unlikely to cause unwarranted and substantial damage or distress.

18. The broad conclusion is that the consequences of self-identification are highly unlikely to cause substantial damage or distress to the individual patient. There is therefore no need to suppress any small numbers to avoid self-identification.

Motivated intruder risk

19. The risks of being identified by a third party are similar to those arising from self identification, except in the following aspects:
    - The third party may not have access to information that the individual is aware of (regarding themselves), so in some areas risk is reduced.
    - However, it may be a breach of confidentiality if a third party can deduce anything about the individual.
    - We need to consider carefully the extent to which a third party might become a motivated intruder, with an incentive to explore the data and deduce information about the individual.

20. The published data does not contain any personal identifiers. If someone had access to GP data then identification could be possible, but these data sources are subject to their own security and rules concerning confidentiality. The data considered here cannot be linked to another data set in a way that would increase the risk of identification. The additional risk that publishing small numbers allows a motivated intruder to deduce information about an individual is next considered.

21. The incentive, and consequently the risk, may be higher when celebrities are known to have attended primary care during the period. There may also be scenarios where someone would seek information about a friend or relative.

22. Because the published data reports on the number of SMI registered patients in receipt of specified health checks, because there are multiple SMI registered patients in each CCG, and because the data does not indicate the findings of these checks or other information about the individual, it is not possible to determine additional information about an individual from the data published.

Step 4 – Would disclosure represent a breach of public trust, the law, or policy for National Statistics?

23. GSS protocols on confidentiality state that disclosure control methods should be judged sufficient when, taking account of information likely to be available to third parties, it would take a disproportionate amount of time, effort or expertise for an intruder to identify a statistical unit to others, or to reveal information about that person that is not already in the public domain.

24. Where patients can identify themselves in the data, there is a risk that the patient could view this as disclosive. As discussed above, such self-identification is highly unlikely to result in substantial distress.

25. In this collection there is no additional data from which an individual can be identified. If a third party was able to access other data sources, such as GP data, to further identify a patient, these secondary sources would have to be fully disclosive

in their own right in order for an individual to be identified. As discussed above, GP systems have their own security protocols.

26. Due to the aggregate nature and content of the data collection, the data does not allow further information about an individual to be determined. i.e. attribute disclosure risk is not considered to be possible.

27. Disclosure would not therefore represent a breach of public trust, the law or official statistics policy.

Conclusion

28. The risk of disclosure and/or harm or distress to data subjects is minimal. It is not possible to identify patients or infer some other fact about the patient's condition or treatment using this data set alone, nor in conjunction with information likely to be available to third parties.

29. It is possible that some patients will be able to identify themselves, but there have been no instances of public disquiet about this and risk of harm from self-identification is very low.

30. For these reasons, the publication of this information about physical health checks for people with SMI involves minimal risk.

31. People with severe mental illness have a life expectancy 15-20 years lower than the general population and monitoring whether they are offered health checks and interventions is included in the Five Year Forward View for Mental Health[5]. Our conclusion is therefore that the publication of such information does not represent a breach of public trust, the law or official statistics policy; and is strongly in the public interest.

[5] https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf