NHS England

Risk of Disclosure – Cancer Waiting Times (CWT) Official Statistics publication

Coverage

This paper assesses confidentiality and data disclosure issues of the commissioner-based and provider-based 'Cancer Waiting Times (CWT)' 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

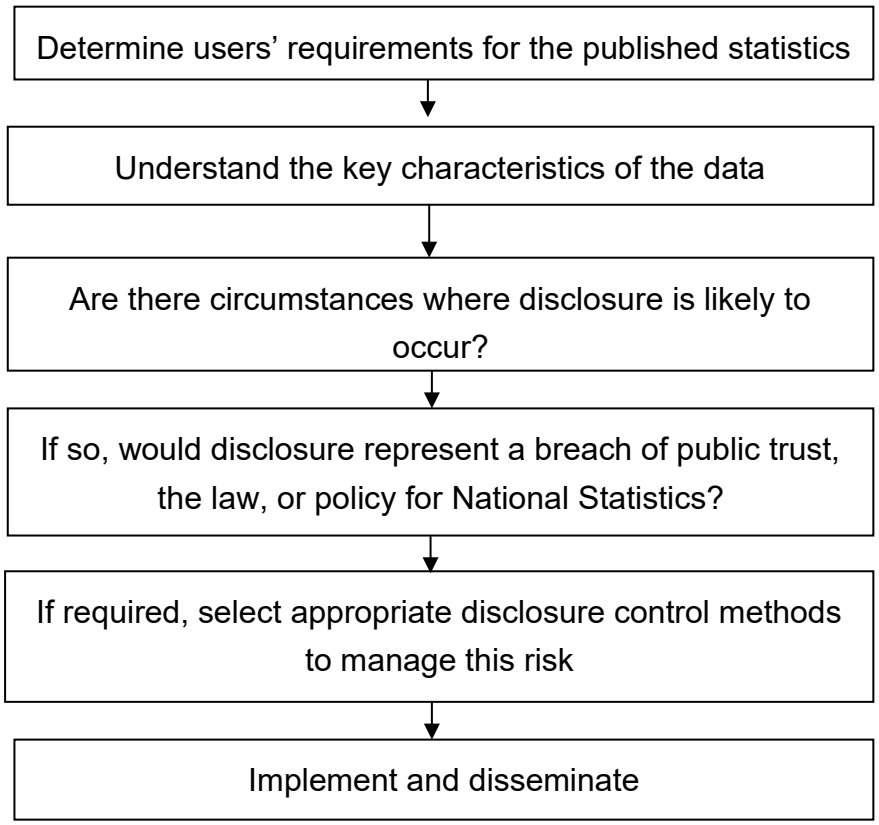
[1] https://code.statisticsauthority.gov.uk/

possible, to fully meet the needs of users. There is then a need to assess whether these data are potentially disclosive.

Guidance from Office for National Statistics (ONS) – the structure of this assessment

3. ONS[2] Guidance on confidentiality sets out guidelines for any assessment of disclosure risk. Whilst the guidance doesn't contain prescriptive rules, it is clear on the need to protect patient confidentiality while at the same time maximising public access to official data. It summarises the six main steps for ensuring access to non-disclosive statistics as shown in Figure 1.

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



[2] https://analysisfunction.civilservice.gov.uk/policy-store/gssgsr-disclosure-control-guidance-for-tables-produced-from-administrative-sources/

Step 1 – Determining users' requirements

4. The requirements for the CWT (Cancer Waiting Times) publication were set out following a Clinical Review of Cancer Standards in 2022[3]. The three current cancer standards specify that:

    - 75% of people should have cancer ruled out or receive a diagnosis within 28 days of referral (the Faster Diagnosis Standard).
    - 85% of people with cancer should begin treatment within 2 months (62 days) of referral.
    - 96% of people with cancer should begin treatment within 1 month (31 days) of deciding to treat the cancer.

5. Data from all English commissioners and providers is submitted monthly, via the Strategic Data Collection System (SDCS).

6. The data are published on the NHS England website[4] on the 2nd Thursday[5] of every month (otherwise known as 'SuperStats day'). Due to data flows and quality assurance processes, the data are published two months in arrears. For example, October data are published in December. When published, the collection allows 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 in a timely way subject to any confidentiality constraints.

7. CWT data are published to give patients and commissioners an insight into the performance of their local area and allows them to compare against all other areas in England. The data covered in the publication for each standard include:

    - Cancer performance over time.
    - Numbers of within standard and breaches, by tumour type; treatment modality; and geography.

8. Conversely, there is 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.

[3] B1320-clinically-led-review-of-nhs-cancer-standards-models-of-care-and-measurement_090322.pdf
[4] Statistics » Cancer Waiting Times
[5] Statistics » 12 months statistics calendar

Step 2 – The characteristics of the data

9. Data are submitted from trusts based on patient level data that is taken from local administrative systems. Once within NHS England, the data used for the CWT publication is then aggregated.

10. There is a multi-stage process for quality assuring and validating that data. This includes automated validation rules to check for quality and completeness; and providers also have the opportunity to review and check their submissions prior to analysis. As part of the production process, the most recent data are compared to the previous 6 months to see if there are any unexplained outliers, and data quality notes can be added to the publication.

11. The numerators represent the number of people who have met the cancer standard within the period. The denominators represent the number of people who are included within the scope of the standard in the period, irrespective of whether they have met the standard or not. Data are also included on tumour sites and provider/commissioner. However, there is no information about the health outcome e.g. the stage of any cancer found, or any details about the demographic characteristics of patients such as ethnicity or deprivation.

12. In terms of volumes, in 2024, between 250,000 and 290,000 people per month were either told they had a diagnosis of cancer or had cancer ruled out.

Step 3 – Evidence of risk of disclosure

13. 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 and are considered in turn below:

    - 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 refer to them.
    - 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:
    - 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.

- Attribute disclosure: Where a third party is able to infer additional information about an individual.

Self-identification risk

14. It is conceivable that a patient may identify themselves within an aggregate count. This would require recognition of trust or commissioner, access route, treatment modality, tumour type and whether their case did or did not breach the given standard during the time-period. However, 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.

15. This would only likely cause distress within smaller counts. Current published tables can contain small numbers. However, it is considered highly unlikely that substantial damage or distress would be caused by such self-identification, as the data do not include further details about the cancer such as stage at diagnosis or health outcomes. The conclusion is therefore that there is no need to suppress any small numbers to avoid self-identification.

Motivated intruder risk

16. A third party might have little motivation to explore the data and deduce information about an individual. The incentive, and consequently the risk, may be higher when celebrities are known to have attended care during the period. There may also be scenarios where someone would seek information about a friend or relative. However, the motivation is likely to be limited in the case of this data set given the very limited set of information included (e.g. it is likely a third party would be much less motivated to determine if treatment started within 62 days, than if the data set enabled them to determine stage of diagnosis or outcome).

17. Where a third party is motivated, they may not have access to information that the individual is aware of (regarding themselves) and so the risk is reduced.

18. It may be a breach of confidentiality if a third party (e.g. a health professional) discloses anything about the individual.

Identity disclosure

19. To assess the identity disclosure risk, we need to consider the size of the underlying population – that is, the number of people to whom the statistics might relate. The CWT

publication is both a provider and commissioner-based publication. Patient characteristics such as age, sex or ethnicity, are not included, although some tumour sites (e.g. breast) are dominated by patients of a particular sex. Providers do not have defined catchments so here the underlying population is all patients registered or resident in England, meaning it is not possible to identify individuals from the provider data. In contrast, commissioners do have defined areas, but the underlying populations[5] are large, with the smallest ICB sub-location having a population over 100,000, of whom over 60,000 are female. If someone had access to an individual trust's 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. It can therefore be concluded that there is no risk of identity disclosure.

Attribute disclosure

20. Attribution disclosure risk is greater where a row or column total is very small. In such circumstances, the risk of attribute disclosure is affected by the fact that separate breakdowns are published for tumour site and treatment modality. A specific nuance with the CWT data is that care and patient allocation for the 62-day standard can be shared across providers. This occurs when patients are transferred between different providers for different parts of their treatment, and the different providers take a certain amount of responsibility for the time a patient has waited. A count can therefore relate to a greater number of patients meaning it is not possible to infer how many patients fall into a particular cell, reducing attribute disclosure risk.

21. Given the combination of the above factors, the risk of attribute disclosure associated with the data as published is very small.

Conclusion

22. Government Statistical Service (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.

5 https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/clinicalcommissioninggroupmidyearpopulationestimates

23. 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.

24. 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 trust 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, trust systems have their own security protocols.

25. Due to the aggregate nature and content of the data collection, identity disclosure is not considered to be possible, and the risk of attribute disclosure is assessed to be very small.

26. For these reasons, the publication of this information about Cancer Waiting Times is assessed as offering sufficient disclosure control.