



1. Home (<https://www.gov.uk/>)
 2. Coronavirus (COVID-19) (<https://www.gov.uk/coronavirus-taxon>)
 3. Testing for coronavirus (COVID-19) (<https://www.gov.uk/coronavirus-taxon/testing>)
 4. Office for National Statistics recommendation for publishing data at small geographies for NHS Test and Trace (<https://www.gov.uk/government/publications/office-for-national-statistics-recommendation-for-publishing-data-at-small-geographies-for-nhs-test-and-trace>)
- Department of Health & Social Care (<https://www.gov.uk/government/organisations/department-of-health-and-social-care>)

Guidance

Disclosure risk assessment for NHS Test and Trace: counts at small geographies

Published 22 October 2020

Contents

Background

Disclosure control

Publishing Test and Trace counts

Balance of risk and utility

Recommendation



© Crown copyright 2020

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3) (<https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3>) or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available at <https://www.gov.uk/government/publications/office-for-national-statistics-recommendation-for-publishing-data-at-small-geographies-for-nhs-test-and-trace/disclosure-risk-assessment-for-nhs-test-and-trace-counts-at-small-geographies>

Background

In monitoring the COVID-19 pandemic, the NHS Test and Trace system has given rise to a range of data publications containing testing data. This has largely been to support and enable the operational response to the pandemic, but also to act quickly to put relevant figures into the public domain.

The numbers of tests undertaken and the numbers of those resulting in people testing positive for COVID-19 have been published at geographic levels down to Middle Layer Super Output Area (MSOA). MSOAs are geographical sub-divisions of England and Wales. The minimum population in an MSOA is 5,000 people and the mean is around 7,200.

MSOAs can be broken down into smaller geographies called Lower Layer Super Output Areas (LSOAs). LSOAs have a minimum population of 1,000 and a mean around 1,700. These areas are both geographic hierarchies designed to improve the reporting of small area statistics in England and Wales.

The publication of MSOA data on the number of positive cases is via the public-facing COVID-19 dashboard. Data is presented weekly on a map and can be downloaded in a range of formats (<https://coronavirus.data.gov.uk/cases>).

This paper assesses the disclosure issues of publishing Test and Trace data at small geographies. It will consider both the risk and impact of disclosure, especially of low counts, alongside the practical use of providing these counts publicly.

Disclosure control

Disclosure control is not an exact science; it is about anticipating risks associated with different datasets that are produced. The most important thing is that there is a thorough discussion around disclosure risk on a study, and disclosure risk is discussed at every point where relevant. Decisions can be flexible and made on a case-by-case basis; disclosure control is not a 'one size fits all' process but relies on understanding the context of any dataset and its intended outputs.

Disclosure control balances safeguarding confidentiality and maximising data, both of which are vital to maintaining trust in statistics. Detail relating to individual statistical units should be protected but the released data must also be of high practical utility for users. Disclosure control is also important for convincing respondents that we take the protection of their details very seriously, which is a prerequisite for collecting high-quality information and obtaining good cooperation and participation.

Any outputs that have zero disclosure risk will be likely to be useless for researchers and policymakers. However, the risk must be managed in such a way as to make the likelihood that an individual could be identified negligible. The balance of risk and utility is contextual, predominantly relating to legal and ethical principles, and on the nature of the data and its sensitivity.

The Data Protection Act (DPA, 2018) and GDPR (2016) stipulate how personal information can be used by organisations, businesses or the government, who must make sure the information collected is used fairly, lawfully and transparently. Under GDPR, health data (under which Test and Trace statistics would fall) is classified as special category and so require additional protections.

Publishing Test and Trace counts

Bearing in mind the legal considerations on disclosure for publishing statistical outputs, it is therefore essential that an individual cannot be identified, either directly from the data or from the data in conjunction with other information, either in the public domain or held privately. The disclosure risk need not be absolutely zero but needs to be negligible using likely scenarios.

Currently, counts of positive tests are released at **MSOA** level. Counts of below 3 are suppressed. The disclosure risk here is low and the potential for any identification at this geography is predominantly via already existing knowledge among friends, acquaintances and work colleagues who are privy to it. A person receiving a positive test would reasonably report this (as a minimum) to other people in their household, other people in their 'bubble' and some work colleagues. One scenario is that one or more of those people might report that to others, but that could be the case for whatever count of positive cases in their area. It appears unlikely that other outside influences would attempt to discover the identities of cases within that area where the count was 3 or greater. This is not so much the statistic being the risk, but the communication of non-statistical information between individuals.

Where the count at **MSOA** is 1 or 2, the above risk is also present. However, there is an additional risk that for certain types of information, rarity or uniqueness may encourage others to seek out the individual. The risk of being identified (as a direct result of their uniqueness) stems from:

- interest within the community wanting to know who the case is
- a small number of people being party to the information as to which individual person has tested positive
- communication from one or more of the knowledgeable parties to the community – this is likely to be on a social media platform (note that this scenario is much less likely with a much larger count)

The threat or reality of this could cause harm or distress to the individual or may lead them to claim that the statistics are inadequate to protect them, and therefore others. The information at threat of being disclosed is the health status of an individual person within the Test and Trace data, and – if this scenario is realistic – would be a breach of the **DPA** and **GDPR**.

A similar argument is present for **LSOA** level data. Moreover, there is then greater likelihood of 'false positive identification' – whereby others may make ill-informed guesses based on partial information within a smaller community. **LSOA** is probably the highest level of geography where an individual in the community could well know everyone, or almost everyone, who is resident there. Sight of a person 'looking a bit unwell' or 'not having left their house for a while' might encourage incorrect claims that could spread rumours within the community. This might cause significant distress to the individual forced to deny that they had tested positive, and they may or may not be believed. This is not a direct disclosure risk per se but would be an unintended consequence of small counts at a low geography.

The consequences of any of the above scenarios are likely to be beyond causing harm or distress to an individual for the reality or perception of their personal data being unprotected. Any significant media coverage of such a case or cases could compromise the Test and Trace system through a wider population mistrust of statistics generally, mistrust of those processing Test and Trace data, and mistrust of the government and officialdom. This may then manifest in a reticence to take part and get tested or provide correct contact details in Test and Trace settings.

Balance of risk and utility

The aim of statistical disclosure control is to reach a fair trade-off between protecting confidentiality and providing outputs that are useful; the risk–utility balance. It is obviously pointless collecting information if it is going to be of no use to researchers, academics or policymakers. However, it is counterproductive to produce statistics so detailed that they allow (either by themselves or in combination with other available information) sensitive information on identifiable individuals to be derived.

Publishing information on a sensitive topic at a local geography of **MSOA** would normally require a threshold to be placed on counts, in accordance with disclosure guidance agreed across the Government Statistical Service (<https://gss.civilservice.gov.uk/policy-store/gssgsr-disclosure-control-guidance-for-tables-produced-from-administrative-sources/>). This threshold would typically be 3, meaning that no counts of 0, 1 or 2 could be published. In the case of an infectious disease, it is more possible that an **MSOA** count of 3, 4 or even higher could relate to individuals all from one household. If we were considering zero or even negligible risk, a threshold of 5 could be considered reasonable. However, the usefulness of the information is important here; the likelihood of only one household contributing to the total of 3 or 4 in an **MSOA** must be low in relation to the usefulness that that information might offer to data users. In that sense, a threshold of 3 does raise the disclosure risk slightly above where the threshold might be 5, but not significantly when measured against the data utility.

At **LSOA** level, the data is more localised and so a count of 3, 4 or higher is that much more likely (than in the **MSOA** case) to stem from one household. Hence, we are not supporting the case for **LSOA** counts to be released, since any threshold would have to be sufficiently high to avoid the 'all in one household' scenario, which might very adversely affect the utility of the data, while the corresponding above-threshold **MSOA** count would be available in any case.

The discussion of utility should be driven by how the information is likely to be used. Counts at **MSOA** are useful in gaining insights into the existence of local spikes. Even a count of 3 positive cases in an **MSOA** would normally translate into a rate of around 40 per 100,000 population in that area but would be unlikely to trigger action such as a local lockdown, unless replicated across an area of contiguous **MSOAs**. Important conclusions can be obtained from higher counts at **MSOA** level in terms of local spikes whereas counts of 0, 1 and 2 could be little more than noise, in some instances perhaps isolated 'false positives', and easy to misinterpret.

The use of **LSOA** counts appears to have minimal marginal gain over the use of those at **MSOA** in these respects. Even a count of 1 in an average-sized **LSOA** (population 1,700) translates into a rate of around 60 per 100,000 but it is of little use policy-wise above and beyond the larger counts in the parent **MSOA**.

The use of **MSOAs** does allow a more nuanced approach than data at higher geographies, such as local authority district. One can group contiguous **MSOAs** to create larger bespoke geographies that still effectively pinpoint hotspots of positive cases. The approach would allow these hotspots to be identified fairly locally, to focus scrutiny and policy on a part of a large city rather than the whole city, for example.

Recommendation

The lowering of the lowest geography to **LSOA** and/or the removal of the threshold of 3 would increase the risk of disclosure to above what **ONS** Statistical Disclosure Control (SDC) would consider acceptable. At present, the threshold of 3 does introduce a risk that is slightly higher than would be usual for sensitive health data but it is accepted that – in the case of Test and Trace – the importance of provision of information outweighs the additional protection offered by a threshold higher than 3.

My recommendation is that an **MSOA** geography with a threshold of 3 (no counts of 0, 1 or 2) is appropriate for this case.

Keith Spicer
Head of **ONS** Statistical Disclosure Control (SDC) Methodology
September 2020