

ONS analogies for selected NHS datasets

“The government are very keen on amassing statistics. They collect them, add them, raise them to the nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damn pleases.”

Josiah Stamp’s adage applies much less to NHS data than to ONS surveys. In general, both because of what it is and because of how it is generated, NHS data is akin to the most sensitive data held by ONS – with data handling akin to the processes ONS uses for its least sensitive data.

One of the ONS Five Safes is “safe people”, through which ONS is able to reject access to people it is not assured will follow the rules, or where there are insufficient governance sanctions available. In many cases, NHS Digital is *obliged* to supply data to other public bodies – which may make their own decisions on assurance for their own suppliers, and where reliability of governance sanctions may be non-existent.

Patient-level data (HES, GP data – 100% administrative lifetime records)

- ONS Equivalent in Coverage: 100% Census data
- ONS Equivalent in Linkage: Longitudinal Study (100%)
- ONS Equivalent in Sensitivity: Economic / Business data

Differences:

Patient data is an administrative dataset without patient involvement in the answers; the ONS analogies have respondent involvement and so can.

Even if the National Data Opt-out was fully applied to HES or GP data, the remaining data would still have coverage higher than the census, i.e. around 95%. (Noting the census is a self-reported survey, so those with particular concerns about particular details can answer how they wish.)

Any mistake would undermine confidence in ONS statistical data handling in a way which would affect both HM Treasury and the Bank of England – as well as the wider ability to measure and forecast the economy, with consequences for all of Government. It is in professionals’ interests to ensure that risks are minimised.

NHS patient data has none of these protections, and is disseminated at high risk for analysis outside the control of the public body. HES / GP data are datasets with the detail and SDC of the 100% census dataset, disseminated as if they were an ONS EUL file.

When NHS bodies talk about “Hospital Episode Statistics, what they mean is not statistics in an ONS published sense, but rather the raw data that is an administrative patient-level census of all data.

Disease registries, e.g. the Cancer Registry (100% administrative lifetime records for those with a particular condition)

- ONS Equivalent in Coverage: 100% Census data
- ONS Equivalent in Linkage: ONS Longitudinal Study (100%)

Differences:

The Cancer Registry is the database of every person who has had cancer since the mid-1990s, with maximum detail of their cancer care, along with details of any other hospital care, and (sometimes) GP records. Equivalents exist for other diseases to a greater or lesser detailed extent.

Data release processes are equivalent to those of GP and Hospital data.

Prescribing data (transactional data; often disclosive about conditions)

- ONS Equivalent: 100% Census Origin / Destination (flow) statistics (i.e. the census data on home / work postcodes and methods of travel between them)

Differences:

[In some cases](#), this data is only available from a commercial entity for a fee with restrictions on use (pay more, fewer restrictions?)

For data that is available, the same style of unique considerations are required of the prescribing protections as are required of some O/D statistics – e.g. workers in GL510EX – *without* the protection of respondents being able to take any steps to protect themselves. **All** of the measures must be taken in the dataset itself; none are.

PROMS (surveys of a cohort of patients)

- ONS Equivalent: Surveys (sometimes longitudinal)

Differences:

Patient Reported Outcome Measures are how patients feel about the treatment they had, and (sometimes) whether they'd recommend it to others; a net promoter score. Sometimes PROMS are one-offs, sometimes longitudinal, sometimes they are run by the care provider, sometimes by charities – but they are all surveys, with all the advantages and disadvantages those can bring.

PROMS should be scaled up in the same way that ONS has scaled up and integrated some of its smaller surveys over time.

The views of ONS on the Covid Infection Study, and on other COVID datasets, would likely provide value to the NHS – especially as the disease registers rationalise into NHSD / X / E, as well as adding overall value to the NHS ecosystem.