

Annex A: Administrative data - a toolkit for non-statisticians in statistical producer bodies

The purpose of this toolkit is to provide a framework to guide the judgments of non-statisticians when confronted with statistics derived from administrative data.

Administrative data are the lifeblood of any organisation. They record what has been done, by whom, to whom, when and where. They range from customer records to payment details, from records of activity like the number of crimes recorded by police to records of geography and the environment like the mapping of flooding incidence by environmental regulators.

Administrative data are frequently, and increasingly, used as the basis for official statistics. This toolkit seeks to support policy makers when they encounter administrative data. Policy makers in government bodies will come across administrative data in different ways. These include when their organisation is responsible for the production of official statistics, and our questions have this primary use in mind. But policy makers may also be presented with administrative data when confronting a new policy problem, and they want to see what their organisation already knows about the problem; and when they are evaluating the success of a policy as it is implemented. The questions below will be a useful guide in those situations too.

While most of these questions are relevant to the other main source of official statistics – surveys – they are particularly pertinent to statistics based on administrative data because they are a relatively less well recognised feature of the statistical system. We are therefore seeking to help policy makers become more intelligent customers through this toolkit.

The toolkit takes the form of 10 key questions. While they work as a sequence of questions, there is no need to work through in methodical order; any of these questions can be asked on its own or in combination with any of the others. The key point is for the policy maker to be willing to interrogate the strength of the data and obtain assurance on the reliability of that data.

The 10 questions

1. Where do the data come from?

This is a simple, straightforward question. Your statisticians should be able to give you a clear answer, explaining the nature of the data, who produces them and why. This question is a good starting point – and if the answer sounds vague or unconvincing, this can be the jumping off point for further questions as you look to interrogate the data more rigorously.

2. Is there a consistent time series?

A good way of understanding the statistics is to see how they behave over time. A time series can demonstrate to you how the underlying data have been affected by contextual factors – which you may know better as a policymaker than the statisticians who work on the production of the official statistics.

3. If there's a limited time series, how do you caveat the statistics you've got, and what warnings do you give about the conclusions that can be drawn?

This is crucial. Your organisation is likely to be criticised – not least by the UK Statistics Authority – if you disseminate statistics which argue for a clear pattern or response to a policy initiative when you don't have a time series to enable you to make these claims with confidence. It's always worth considering the extent to which the time series might be subject to cyclical patterns, such as economic growth.

4. What is the story behind the pattern revealed by the statistics?

Understanding the story behind the patterns is important – though this is more an internal check: to satisfy yourself that the patterns are plausible and fit your own experience. Be cautious about going public with this story until you have explored other explanations for the patterns in the data.

5. Have you changed measurement – or data suppliers – and might this be a plausible explanation for the pattern you see in the statistics?

Sometimes, what seems to be a good story – how a particular initiative has produced a clear result – is in fact the product of changes in the way data are measured, collected or categorised. Data are particularly prone to change where the supplier itself has changed (e.g. from one contractor to another; or from a shift in organisational responsibilities; or a change programme within the supplier).

6. Is there any sense in which these data are obviously subject to confirmation bias – saying what those who commission it want to hear?

This is a difficult question for you to ask, because it could be that there is a tension between the role of statisticians as independent data collectors and your responsibilities for the successful delivery of policy. Nevertheless, it is an essential question to ask, even if the answer does not make for comfortable reading.

7. What level of assurance do you have over the administrative data?

The UK Statistics Authority has recently emphasised the importance of statisticians obtaining assurance over the underlying administrative data that

feeds into official statistics. So your statisticians ought to have a clear answer to the question about how they know the underlying data are reliable. If not you should ask them to get it.

8. How important are the data to the supplier – to meeting their own KPIs/success factors?

If the data supplier uses the underlying administrative data as the basis for a) its performance against a key performance indicator and b) for its returns to your statisticians, there is a potential for criticism of your statistics – because people may believe that the data supplier is incentivised to record the data in ways which make it look good. In this context, assurance is even more important.

9. What do you know about the processes by which the data have been compiled?

Like many of the other questions here, this is a remarkably simple question that can guide you as to how far you should rely on the statistics based on the data. If your teams don't seem to understand or trust the process, it's probably right to be concerned about the reliability of the official statistics.

10. How independent are the statisticians from the suppliers of the data?

The *Code of Practice for Official Statistics*¹⁴ (the *Code*) is the key document designed to ensure high quality statistical practice in government bodies. One of its key principles is the need for statisticians to act independently of both data suppliers and political decision-makers. So you should always ask your statisticians how they have complied with the independence requirements of the *Code*.

¹⁴ <http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html>