Government Statistical Service

Statistics for policy professionals

Things that you need to know

First Edition
Version 1.0
January 2017

Contacts

Good Practice Team
Government Statistical Service
1 Drummond Gate
London SW1V 2QQ

Twitter: @UKGSS
Email us: goodpracticeteam@statistics.gov.uk

There are more resources on this topic on the GSS website: http://bit.ly/goodpracticeresources

Acknowledgements

The Good Practice Team would like to thank all of the statistical and policy colleagues in government who have helped us to design and deliver our sessions on "Ten things you really need to know about statistics" and provided advice, critique and suggestions for improving the content and materials.

We welcome feedback and suggestions on how to improve this guidance. Please email us at goodpracticeteam@statistics.gov.uk.

Contents

About this guidance

The case for a policy can usually be strengthened by including statistics that show context and help to demonstrate the predicted impact and cost. These might come from published information or you might ask your analyst colleagues for some new evidence.

The Good Practice Team (GPT) runs a seminar called **"Ten things you really need to know about statistics"** for policy colleagues across central government.

The aim of the seminar is to help colleagues to work effectively with analysts in their departments. In particular, we wanted to introduce some key ideas and concepts to help you to ask the right questions at the right time and to understand, interpret, use and challenge statistical information - expecting and acting upon caveats about its strengths and limitations. We talk about the need for accurate presentation and the need to check the plausibility of numbers in the context of the policy area.

We deliberately did not include an introduction to the theory and methods of statistics in the seminar, because these are better covered elsewhere. We did introduce local representatives from the Government Statistical Service and recommended a close and ongoing relationship early and throughout the policy development cycle.

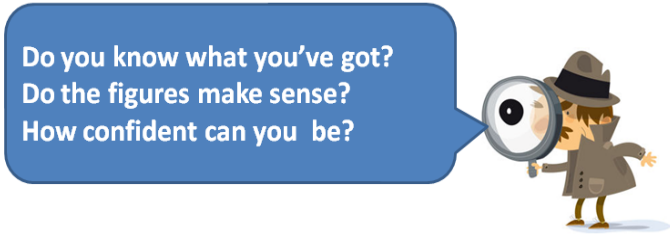
This guidance brings together the main points from the seminar and from further consultation with colleagues. Our aim is that it will help to reinforce the ideas introduced in the seminar and stand as a companion volume and supporting resource.

If you would like us to run a "Ten things …" session in your department, please get in touch with the Good Practice Team.

"Objective1: Inform decision making

We will provide a firm evidence base for sound decision, supporting the formulation of effective government policies and the management of public service delivery …"

Better Statistics, Better Decisions. Strategy for UK statistics, 2015 to 2020



Does it look right?

Take a critical look at the numbers

Before you commit your arguments for wider consideration, we suggest you take a critical look at the numbers. Do they look right? This is not to cast doubt on the competence of analysts, although mistakes can happen. Rather, are you sure that you have a clear, common understanding of the question that the figures address?

To illustrate, we picked a front page article from the Daily Telegraph: "Cannabis causing a quarter of psychosis", which asserted that 60,000 people in Britain are living with mental illness because of the drug. Normally we might take a passing interest in such a story, but picture yourself as the public servant responsible for a policy intervention to prevent this happening or to address the consequences. That is when healthy scepticism may help.

"The finding suggests that about 60,000 people in Britain are currently living with conditions involving hallucinations and paranoid episodes brought on by abuse of high-potency cannabis"

Daily Telegraph, 16 Feb 2015

The Telegraph article drew on findings from a Lancet Journal of Psychiatry research paper. The authors found that 24% of psychosis cases in their study group could be attributed to high strength cannabis, once other factors were accounted for. The results were based on a sample of 410 adults aged 18-65 in South London, where the prevalence of

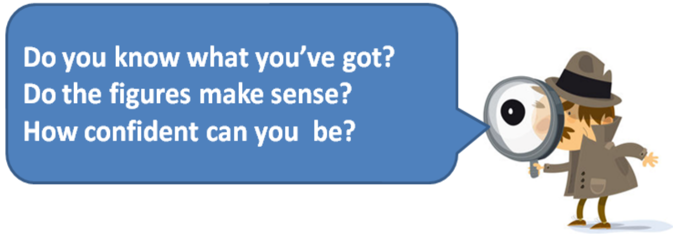

THE LANCET

Articles

Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: a case-control study

psychosis and the use of cannabis are both high compared to the rest of the country.

The research paper did not mention the "60,000 people in Britain" discussed in the Telegraph at all.



Does it look right?

If a policy intervention was planned on the basis of these figures, this would be a good time to check back with the original source to see whether the numbers are based on any important assumptions and to consider whether they are reasonable.

We recommend that you take out the trusty "back of the envelope" and try out some simple sums.

- 0.4% of adults have psychosis
- There are 64.1 million people in the country
- Back of envelope:

  0.4% of 64,100,000 = 256,000 people with psychosis

  (24% of psychosis is attributable to skunk)

  24% of 256,000 = **62,000**
- OK?

We don't know how the Telegraph actually got to 60,000, because the article doesn't explain this. But we can use information that is readily to hand to see whether 60,000 looks broadly sensible.

To start with, we know that there were about 64.1 million people in the UK at mid-year 2013. We can get this figure from government population estimates.

We also know from the 2007 Adult Psychiatric Morbidity Survey, another government statistical source, that 0.4% of adults suffer from psychosis.

Bringing those figures together, we could estimate that about 0.4% of 64.1 million – that's 256,000 - people have psychosis.

From the Lancet article, we think about 24% of psychosis cases are attributable to high strength cannabis, so taking 24% of 256,000 we get to **62,000 people**.

Do you know what you've got?
Do the figures make sense?
How confident can you be?

Does it look right?

Using the 24% statistic in this way also helps to reinforce that this is 24% of psychosis cases, not that, for example, 24% of cannabis smokers will go on to suffer from psychosis.

What assumptions have we made?

Although we've used only simple calculations and readily available information, this does seem to land us in same ball park as the Telegraph. However, by making this calculation, we've made some very important assumptions. Do they make sense?

Firstly, we've applied the 0.4% figure from the Adult Psychiatric Morbidity survey to the entire UK population, and not just to adults. By doing that, we're assuming that the psychosis rate is identical for adults and children. We do not have any evidence to support this assumption. It would have been better to exclude children from the population total when multiplying by 0.4%, but if we did that we would, quite reasonably, arrive at rather less than 62,000.

Secondly, we've applied the 24% to all psychosis sufferers. By doing this, we've generalised the Lancet findings from a specific population in South London, where we know there are some local issues, to the UK, and assumed that the 24% applies universally.

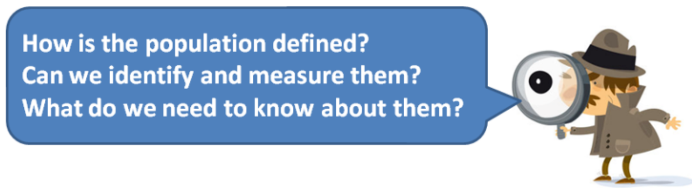
Thirdly, our data sources are from different time points. By bringing them together we've assumed this has no problematic effect. This is potentially less concerning than the other two assumptions (because we expect that the indicators we are measuring here vary quite slowly) but it is noteworthy anyway.

What are their effects?

When these assumptions are taken together, they include a lot of people who should probably be excluded from the total. As a result, we may be overstating things by quite a margin.

If this were a real policy scenario, this would be a good time to contact your analyst and talk through the concepts and assumptions made so you can build a common understanding of what reasonably can be said about this evidence and whether the findings are robust enough to warrant a major intervention.

What, exactly, are we measuring?



Government policies are designed to have a beneficial effect on, for example, “troubled families”, “historic buildings and monuments”, “cyber security breaches”, and “migration”. Such policies typically come with an intended or predicted impact that will need to be evaluated.

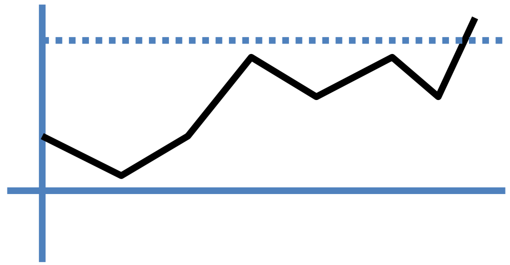
For an objective evaluation, **the group we want to measure must be clearly defined in advance** to provide accurate estimates and to avoid confusion or possible accusations of selectivity in reporting to paint a favourable picture of policy impact.

As well as a transparent definition, **we also need to know that we can, in practice, identify and measure the target group**. Is the group identifiable from information available in administrative systems? Do we need to ask people whether they fit the criteria? If so, what questions should we ask and would they be able to answer them accurately?

Given that we can identify the things we need to count, **we also need to consider what we are going to measure about them**. Again, this may be registered in an administrative system but if not we may have to ask questions to gather the information.

Sometimes, what we want to measure is not available, or not readily translated into questions, or we do not have the resources available to collect what we need. **Here, we may decide to use a proxy measure instead**. For example, the number of unemployment benefit claimants might be a good proxy for the number of unemployed people.

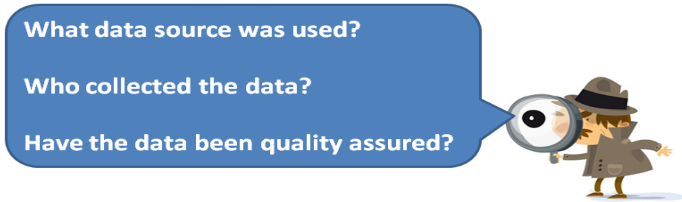
Once we can locate and count our target group and collect the right measures about them, **we need to think about the statistics we will derive from the data**.



These might be summary measures like an average, or more complex statistics showing the relationship between different quantities.

Finally we need to think in advance about **what levels of change in the numbers will be associated with success**, and whether the data we can collect will be accurate enough to clearly indicate whether that level has been met.

Where do the data come from?



We need data to understand how policies are working, and to assist in the policy-making process.

Data may be collected for statistical purposes in a **census** or a **survey**, or as a by-product of an administrative process. New sources of data might also be available, like mobile telephone traces, online transactions or social media posts.

In practice, data sources are always flawed to some extent. These flaws may be reflected in the quality of the resulting statistics and we will need to understand how when we take decisions using those statistics. **Understanding data collection processes helps to reveal the strengths and limitations of the data and the quality of statistics based on them.**

The frequency of data collection is also important. Data might be available annually, monthly, weekly, daily, or, for digital sources like social media posts or search engine traffic, every few seconds.

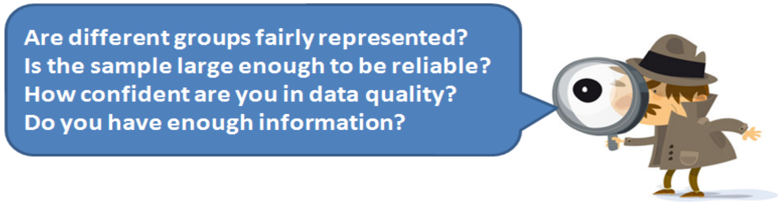
While timely data provide an up to date picture, very frequent measures tend to vary a lot, and discriminating between random noise and real change can be a challenge. We often need to smooth or summarise such data to extract patterns and trends.

When data are collected purposefully in a well-designed census or survey, policy questions can be directly answered from the questions asked. Those responsible for collecting the data should be able to explain clearly the steps they have taken to ensure that different groups are represented fairly and that the data they collect on each subject is as close to the truth as possible. The cost of collecting the data should also be considered.

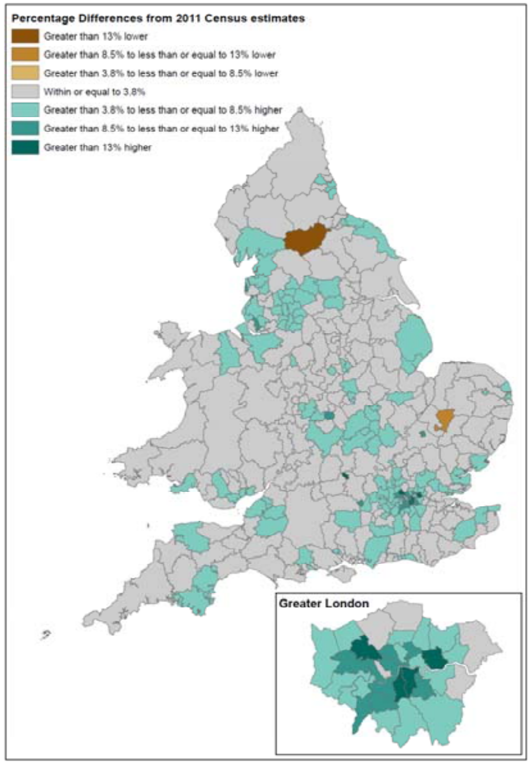
Survey analysts will aim to specify an adequate sample to produce sufficiently reliable estimates, but it is important to understand the level of uncertainty around the numbers that result. You should also explore the interests of those commissioning the survey - could the study be designed to reach particular conclusions?

Administrative data sources are attractive as they are by-products of existing processes, and therefore cost much less than censuses or surveys. Nevertheless, consider whether they are likely to fairly represent your population of interest. Often, the information collected in an administrative source may not match precisely the measure that you require. At the same time, the process of collecting and maintaining the data may introduce biases or errors into the data that will carry through into your analysis unless you adjust for them.

Where do the data come from?



ONS has explored how the NHS Patient Register compares with the Census for counting the population. The map below shows differences between the two. The Patient Register undercounts the population in some rural areas, and over counts in large cities. Why is this?



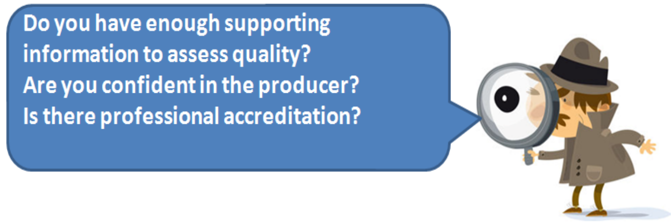
The differences are driven by how the register is administered and who is on it. It undercounts newborns (because of the time allowed to register them). It includes some people who have died or left the country until data maintenance removes them. Young adults and recent migrants, who move around a lot, are often registered in the wrong place. This drives the over count in large cities. Older people and families are well represented, because they interact a lot with the health system. Military personnel are not registered (they have their own health service). This explains the undercount in rural areas with large military bases. None of this is surprising when you understand how the register works, but it has profound implications when we use it to count people for statistical purposes.

To help you understand the quality of statistics from administrative data, think about the process that created them. Is there a clear account of the way the data are collected, by whom, and the rules used? **Are the data quality assured or audited?** Consider the motivations of those inputting the data. Might they favour recording or scoring cases to improve the picture of their own performance?

New data sources like social media posts or mobile phone traces have issues of bias and coverage just like surveys and administrative data. **Think about what they measure and its relevance to your policy question.** Who might be more or less likely to appear in the data? Could collection or processing affect this?

Consider the use of mobile telephone data to understand commuting patterns. Do the data include everyone, or only contract customers? Do numbers vary geographically across mobile operators? Do users opt out of sharing their information? All of these factors would affect the quality and reliability of such apparently comprehensive data.

How are the statistics calculated?



In practice, the statistical operations that create the numbers you use from raw data are usually closely entwined with data collection processes and knowledge about likely strengths and limitations.

We have separated out the earlier section on data collection from subsequent processing in this discussion, acknowledging that, as policy professionals, you usually have direct access to those producing the statistics and rely on them for reassurance about data collection.

You will need to make judgements about the credibility of those producing statistics. This may come from a formal commissioning process, taking into account professional standing and reputation. Consider whether the derivation and reporting of statistics could, like data collection, be influenced by any motivation to paint a favourable picture of the producer's own performance or that of their organisation.

You may have more confidence where independence and professional integrity are reinforced by the application of, for example, the **Code of Practice for Official Statistics**. Such codes also include principles about the ethical collection of data and protection of the confidentiality of individuals included in the data.

When presented with statistics, you should expect to see supporting information about their quality and the methods used in their derivation. This should including statements about the quality of the data used and explanations of steps takes to mitigate flaws in the data when deriving the statistics, including any assumptions made to justify the approach. This should be done with a critical perspective that focuses on the use of the statistics.

You might see statements about the extent to which definitions used in the statistics are harmonised with those used in other statistics, which helps to improve comparability. **This explanation of quality should help you understand the strengths and limitations of the statistics for your specific use.**

Where statistical processes or data collection processes are changed, this can risk creating a spurious change in the statistics that does not reflect true change. Such changes do inevitably happen: a new contract may be awarded or the administration of a policy may change.

Expect to see assurances about the continuity of statistical processes and data collection, or a description of steps taken to mitigate that change and information about the likely impact on the statistics.

The small print matters!
Are the measures about the same thing?
Do they include different metrics?

Making fair comparisons

We are sometimes faced with several statistics about the same topic. While this is helpful in building understanding of the policy issue, there are times when the numbers seem to paint conflicting or different pictures.

Consider this example. The Department of Culture, Media and Sport (DCMS) Creative Industries Economic Estimates for 2013 estimated the value (approximate Gross Value Added, or aGVA) of the Advertising and Marketing sector of the British economy at £10.2 billion, employing 167,000 people.

A report from the Advertising Association (AA) and Deloitte Consulting, also dating from 2013, told a rather different story. Their analysis reported that £16 billion was spent on advertising, leading to a £100 billion contribution to GDP and the employment of 550,000 people.

These numbers are very different – the AA figure is almost ten times larger than the DCMS one. What is going on here?

The answer lies in the way that the two different figures were calculated. The DCMS measure uses figures from the ONS Annual Business Survey to estimate the economic contribution of advertising and marketing. It focuses in on that one industry.

When you delve into the AA report, you find that the scope of their measure is much wider. As well as the direct contribution of advertisers themselves, they have also tried to estimate the total economic impact of advertising across all sectors. This might include all sorts of things, from the money that advertising executives spend on catering or fashion through to their investment in office cleaning services or printing. It's a much wider measure, and, unsurprisingly, the resulting figure is much, much bigger. This doesn't mean it's wrong, just that the scope is very different, and it provides an answer to a different question.

When comparing two statistics that apparently measure the same thing, take care that you understand how they are measured and whether comparing them directly is sensible. In this case, it isn't.

The small print matters!
Are the measures about the same thing?
Do they include different metrics?

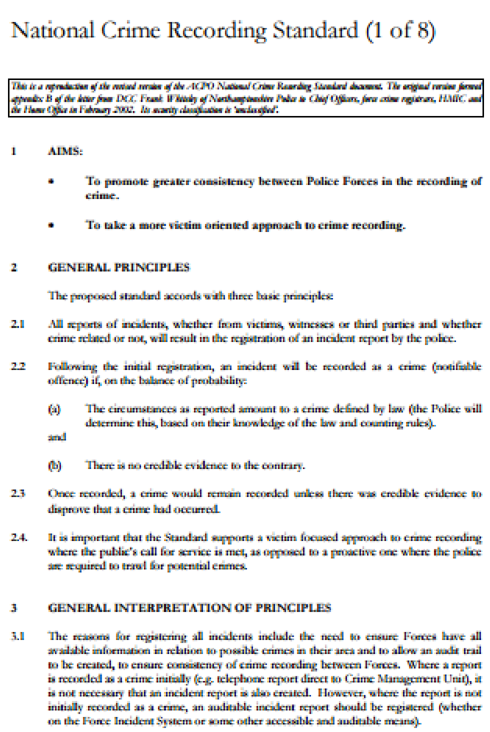
Making fair comparisons

The issue of fair comparisons also arises when we compare statistics about the same thing from different time points.

Comparisons through time can be invalidated when the measuring system changes during the time period of interest. Such changes can happen for lots of good reasons. The methods used to collect, collate or analyse statistics might improve, for example. Statistics might be updated as new information becomes available, improving accuracy. More rarely, the definitions that set out the population or topic that we are trying to measure can be revised.

In 2010, the UK Statistics Authority took issue with the comparison of police recorded crime statistics from the late 1990s with those from 2008/9.

The problem here was that the National Crime Recording Standard was introduced in 2002/3, resulting in marked change to the way that some offences were logged. This led to an apparent spike in the number of offences, particularly some types of violent crime, and a consistent rise in the count, but the change was due to the introduction of new processes and



National Crime Recording Standard (1 of 8)

This is a reproduction of the revised version of the ACPO National Crime Recording Standard document. The original version formed appendix B of the letter from DCC Frank Whiteley of Northamptonshire Police to Chief Officers, force crime registrars, HMIC and the Home Office in February 2002. Its security classification is 'unclassified'.

1 **AIMS:**

- **To promote greater consistency between Police Forces in the recording of crime.**
- **To take a more victim oriented approach to crime recording.**

2 **GENERAL PRINCIPLES**

The proposed standard accords with three basic principles:

2.1 All reports of incidents, whether from victims, witnesses or third parties and whether crime related or not, will result in the registration of an incident report by the police.

2.2 Following the initial registration, an incident will be recorded as a crime (notifiable offence) if, on the balance of probability:

(a) The circumstances as reported amount to a crime defined by law (the Police will determine this, based on their knowledge of the law and counting rules).

and

(b) There is no credible evidence to the contrary.

2.3 Once recorded, a crime would remain recorded unless there was credible evidence to disprove that a crime had occurred.

2.4. It is important that the Standard supports a victim focused approach to crime recording where the public's call for service is met, as opposed to a proactive one where the police are required to trawl for potential crimes.

3 **GENERAL INTERPRETATION OF PRINCIPLES**

3.1 The reasons for registering all incidents include the need to ensure Forces have all available information in relation to possible crimes in their area and to allow an audit trail to be created, to ensure consistency of crime recording between Forces. Where a report is recorded as a crime initially (e.g. telephone report direct to Crime Management Unit), it is not necessary that an incident report is also created. However, where the report is not initially recorded as a crime, an auditable incident report should be registered (whether on the Force Incident System or some other accessible and auditable means).

not a reflection of a real trend on the ground. Interestingly, the Crime Survey, which provides an alternative measure of crime levels, showed no such increase. In this situation, having an alternative source as reliable as the Crime Survey provides a helpful sense check that the change in recorded crime may not be what it seems, and exploring why the two measures differ would point towards the change in recording standards.

Some commentators compared the old figures with the new in the context of political debate about trends in rising crime without drawing attention to the change in recording standards, and the UK Statistics Authority ruled that this was misleading.

Analysis should ensure that official statistics caveat changes like this clearly and advise you on how to interpret the numbers accordingly. Take care to ensure that comparisons are valid.

Establishing causality is difficult to do
Random control trials can really help
Is the focus too narrow, and just looking at problem cases?

What caused what?

Being able to state confidently that something happened as a result of something else is incredibly useful, and hugely relevant in the context of policy evaluation.

In practice, determining causality is quite challenging. **There is a big difference between correlation (an association between factors) and causation (the fact that A causes B).** Establishing causation requires that we demonstrate:

1. That A is correlated with B.
2. That A happened before B.
3. That all other plausible causes of B have been ruled out.

The well-known and (nowadays) accepted causal relationship between smoking and lung cancer, established over years of careful scientific study, meets these criteria, but achieving this level of certainty is usually very challenging.

Consider the case of the "Scared Straight" Programme in the United States in the 1970s. This was an attempt to deter juveniles deemed to be at risk of criminal behaviour by bringing them into contact with prison inmates. The policy rationale was this:

Visit Prison

↓

Witness prison life and hate it

↓

Realise might end up in prison

↓

Delinquent behaviour averted!

Early studies showed success rates as high as 94 per cent, with widespread media attention and a TV documentary providing anecdotal evidence of success.

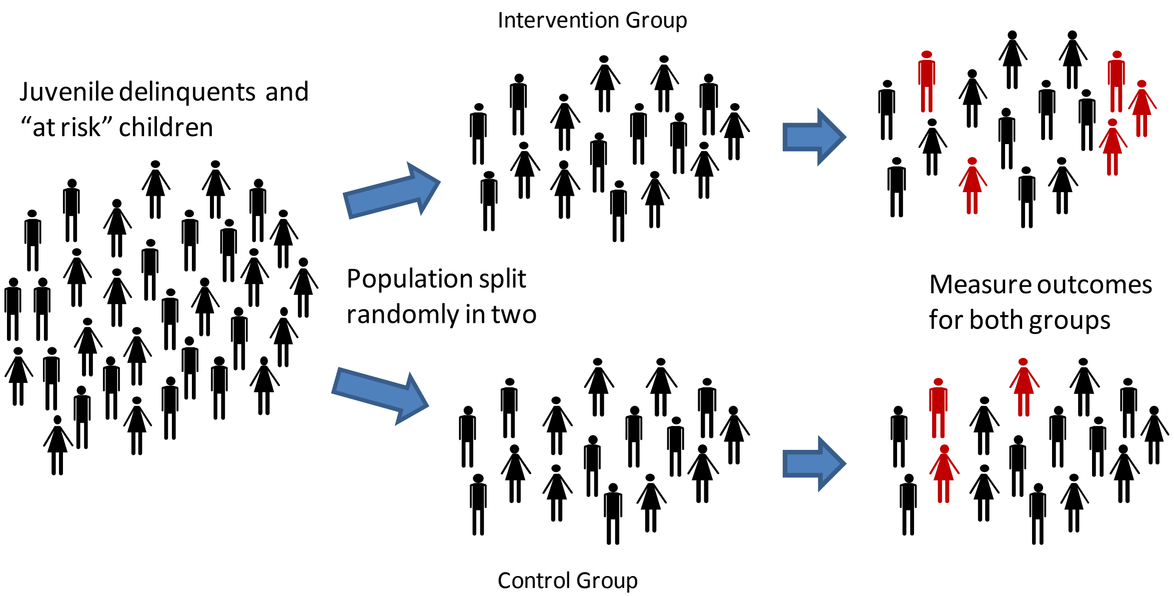
The problem was that the studies looked only at the group who went through the programme, and ignored similar individuals who did not. This meant that it was unreasonable to conclude definitively that the programme itself was causing the positive effect.

In 1982, the study was repeated as a randomised control trial. Randomised control trials introduce a randomly assigned control group, which enables you to compare the effectiveness of an intervention with what would have happened if you changed nothing.[1]

[1] Haynes, L., Service, O., Goldacre, B. and Torgerson, D., 2012, "Test, Learn, Adapt: Developing public policy with randomised controlled trials", Cabinet Office Behavioural Insights Team.

What caused what?

Establishing causality is difficult to do
Random control trials can really help
Is the focus too narrow, and just looking at problem cases?

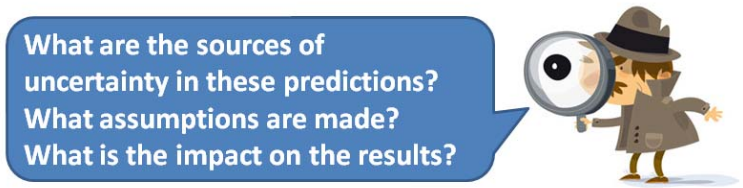


For the "Scared Straight" randomised control trial, individuals at risk were split into two groups at random. Half of them went through the programme, and the others experienced no intervention. At the end of the process, rates of criminal behaviour for the two groups were compared.

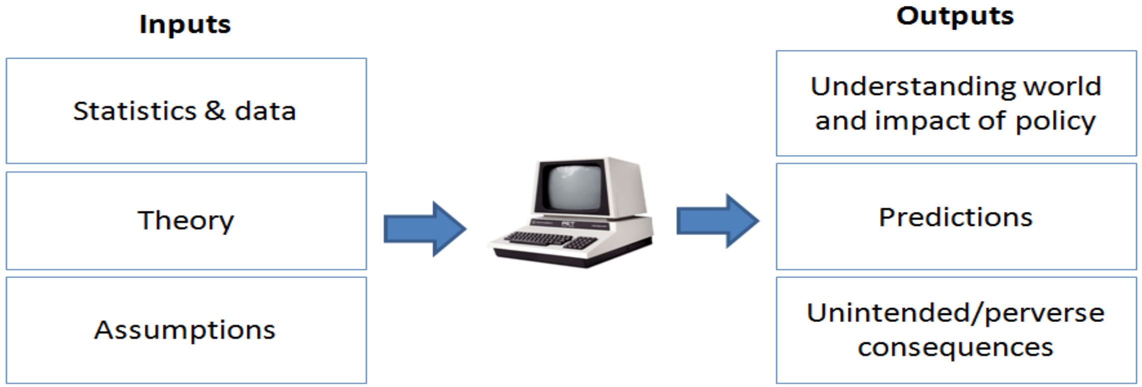
In fact, the random control trial results told a very different story. "Scared Straight" had no significant positive effect on behaviour when compared to similar individuals who did not participate in the programme and the cost was more than thirty times higher than the benefit.

The programme actually had the opposite effect to that intended, with some participants more likely to be arrested! These results are consistent with an alternative process:

It was only the use of the random control trial method that revealed the limited contribution that this policy intervention actually made.



Understand the assumptions



The numbers that you use to develop and evaluate policy originate in data collection and statistical processes. Such evidence may contain estimates or forecasts from a statistical model.

Models are useful. They help us to simplify and understand patterns, predict how something might work and identify potential unintended consequences of a policy decision. **Models construct a simplified view of the world to build insight.** They work by bringing together information, statistical theory and assumptions.

For an informed view of model results, **it is essential to understand what impact the assumptions have on the outcomes, how the techniques used translate into outputs, and the level of uncertainty around those outputs.**

The West Coast Mainline rail franchise bid is an infamous example of where failure to communicate assumptions caused huge difficulty. A new financial model was used by DfT to assess bids by rival operators to operate the franchise. Fundamental mistakes in the communication of model assumptions led to inconsistency in rival bids and incorrect conclusions being drawn about their relative merits. As a result, the government was forced to cancel the franchise competition and instigate a re-tendering process at a cost in excess of £50 million.

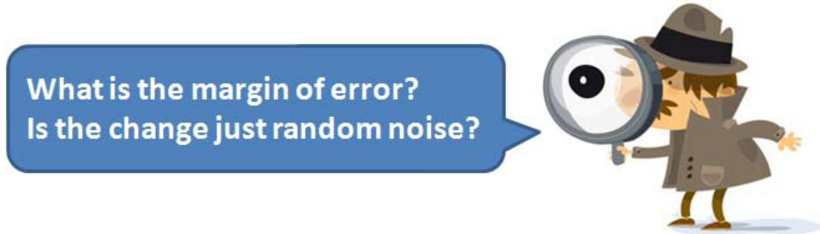
What can we learn from this in practice? It is vital to understand key limitations of model outputs and the assumptions that have been made, so that the evidence they provide can be treated appropriately. **This understanding of the impact of limitations and caveats must be passed on to those making decisions based on the model.**

The AQUA book[2] and guidance by the National Audit Office[3] emphasise this.

[2] https://www.gov.uk/government/publications/the-aqua-book-guidance-on-producing-quality-analysis-for-government

[3] https://www.nao.org.uk/report/framework-to-review-models/

How sure are you about this?



Statistical data, whether from surveys, administrative sources or elsewhere, are never perfect. When presented with some statistics, think about how and why they were collected and what this might indicate about the accuracy and reliability of the numbers in front of you.

NEWS POLITICS

Home World UK England N. Ireland Scotland Wales Business Politics Health Education Sci/Env

17 August 2011 Last updated at 12:31

'Worrying' jobless rise needs urgent action - Labour

Labour have said ministers need to take urgent action to reverse the "very worrying" rise in unemployment.

The number of people out of work rose by 38,000 to 2.49 million in the three months to June, official figures show.

The opposition said the situation was "very serious" and ministers must consider work guarantees for under-24s.

Ministers said the figures were "disappointing" but jobs were being created and "intensive" help being given to those looking for work.

Following the quarterly rise, unemployment levels are now at their highest since February 2010.

The number of 16 to 24-year-olds out of work rose by

jobcentreplus

Unemployment levels are at their highest for 18 months

Consider this example from BBC News[4], discussed by Ben Goldacre in his "Bad Science" blog[5].

It refers to statistics from the ONS Labour Force Survey (LFS). The survey reported an estimated rise in unemployment of 38,000 for April – June 2011, and the article cites opposition MPs requesting urgent action to address this.

38,000 sounds like a lot. However, this is a survey estimate and represents a change in the unemployment rate of 0.1 percentage points. It's an estimate because the survey doesn't ask everybody about their unemployment status, but takes a representative sample of people and uses that to estimate the level of unemployment for the entire population.

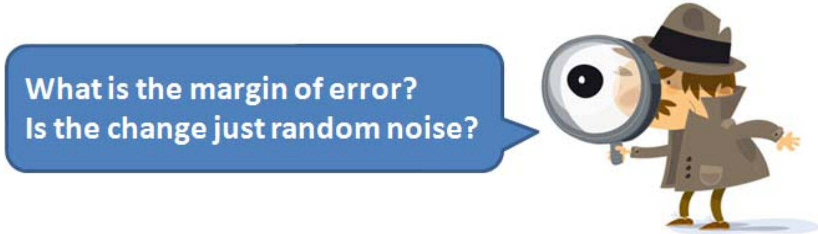
Alongside the 38,000, ONS provides a measure of "sampling variability" called a "95% confidence interval". The confidence interval defines a range around the survey estimate, within which the real number is likely to fall. The 95% tells us what we mean by 'likely'. It means that if we were to draw 100 LFS samples instead of 1 (we would never do this in practice, it would cost far too much!), we would expect that for 95 out of those 100 samples, the survey estimate of unemployment would be within the range set by the confidence interval.

In this quarter, the confidence interval is reported as ± 87,000, so the possible range for the change in unemployment runs from -49,000 (here unemployment would have fallen) to +125,000. The range includes 0 – no change at all.

Given all this, what can we say about the change in unemployment? Our best estimate *is* that it increased by 38,000. But the change is not statistically significant and the estimate is bouncing around "no change". In other words, there is little evidence that drastic new action is needed.

[4] http://www.bbc.co.uk/news/uk-politics-14558369

[5] https://www.theguardian.com/commentisfree/2011/aug/19/bad-science-unemployment-statistical-noise



How sure are you about this?

This example illustrates the need to be mindful of the level of uncertainty in survey data before getting too excited about an apparent change. **But confidence intervals can also reassure us that change is genuine.** The Lancet study of cannabis and psychosis that we began with also discusses uncertainty. Recall that 24% of psychosis cases in the study were attributed to skunk cannabis use. This measure also has a 95% confidence interval around it – this time giving a range of ± 6 per cent, from 18 to 30 per cent. The key difference here is that the range does not include 0 – the confidence interval supports the finding that there is a positive association between skunk cannabis use and psychosis, while the range around the 24% is sufficiently small and gives reassurance that the level of use estimated is reasonable.

For a well-designed survey estimate it should always be possible to obtain helpful measures of variability like these confidence intervals. The statistical properties of surveys are well understood, and we can use them to calculate the level of uncertainty in the numbers. Analysts can help to provide and explain such context.

However, there is a wider issue here. **It can be less straightforward to get reliable information about uncertainty and quality for non-survey sources** such as administrative data or "big data" because often little is known or reported about their statistical properties.

Administrative data produced by government should be easiest to deal with, because data collection and processing are usually well documented and issues of completeness, coverage and reliability are transparent. Your analyst colleagues should be able to reassure you here.

Third party data and tools can be less straightforward. It is often difficult to obtain information about how they work, and in some cases methods and inputs are commercially sensitive and not disclosed.

Comparison against similar indicators with known properties can help, but these are not always available. In such cases, **take a balanced view of the evidence that accounts for what you know about quality**.

Look for long term patterns and trends when measuring impacts
Things return to their normal level
Consider population size

Putting things in context

Statistical evidence is more than just the latest numbers. To consider and interpret such evidence **we must think about the wider context**, as this shapes how we interpret what the numbers tell us.

The first point here is that things tend to return to normal over time. Unusual or extreme cases typically become less so. Consider the example of a moderate football team which surprisingly tops the league. Unless there has been a fundamental change in the underlying strength of that club, we would expect that team to finish lower down the league in subsequent years.

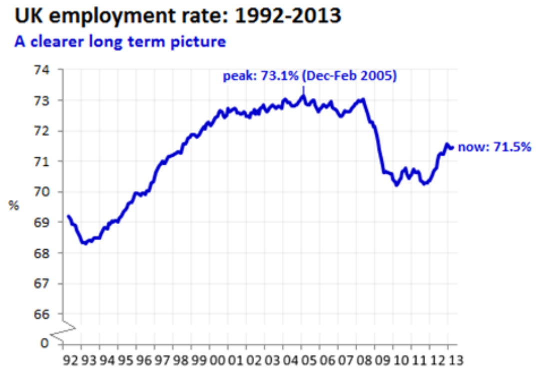
Changes in team composition, managerial policy, player injuries, the relative performance of other teams and a thousand other factors that come together under the umbrella of random variation mean that, in the long run, we'd expect performance to become less exceptional and the team to lose the top spot.

In statistics, we refer to this effect as "regression to the mean". It is very relevant to policy making because **if we focus intervention on the extremes – the very good or very bad cases – and they change, this can be due to the underlying tendency to return to normal** as a consequence of random variation, rather than a specific effect of the intervention. Taking a longer view, looking at broad trends (including the unexceptional cases) and whether other data echo the change you observe can help here.

The second point is that record numbers are not always remarkable. Consider this graph of the employment count:

Source: ONS Labour Market Statistics, June 2013

The graph shows that there were a record number of people employed in the UK in 2013, but that if we look at the longer term there had been record numbers every year (except during the financial crisis of 2008 and its aftermath). The fact that there were record numbers is not surprising at all because the population was increasing throughout this period.



Source: ONS Labour Market Statistics, June 2013

If we look instead at the employment rate (taking account of population growth) the picture is more nuanced. By 2013, the employment rate had not yet regained 2005 levels.

Putting things in context

The third point is that the way we report context can dramatically affect the impact of numbers. “*The tiger that isn’t*”[6] describes the case of an apparent link between mobile phones and cancer.

In 2005, the National Radiological Protection Board advised that children should not use mobile telephones because of the risk of radiation exposure leading to the development of acoustic neuroma, a type of tumour. Research published in the British Journal of Cancer[7] indicated that after ten years of mobile phone use, the risk of developing neuroma doubled.

Context is absolutely critical here. The baseline risk of developing acoustic neuroma without using a mobile phone at all was reported as 0.001%, or 1 in 100,000. The results of the study showed that after ten years, this risk doubled to 0.002%, or 2 in 100,000.

To put this in perspective, Blastland and Dilnot asked the study authors if they would prevent their own children from using mobile phones as a result of the increased risk. They indicated that they preferred to know where the children were and how to contact them, as the risks around not being able to do so were much more important than that of developing cancer.

We can report this risk information in a very different way. A doubling of risk from 0.001% to 0.002% also represents a **100% increase in the level of risk**. This sounds very serious, although actually the wider context suggests that the increase is probably not a cause for concern.

As the evidence accumulated, further studies failed to corroborate the original, and the doubling of risk reported here was found to be a statistical fluke.

[6] Blastland, M. and Dilnot. A., 2007, The tiger that isn’t: seeing through a world of numbers, Profile Books.

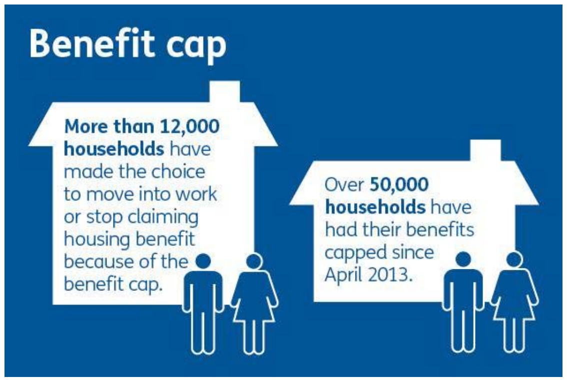
[7] Schoemaker, M.J. et al., 2005, Mobile phone use and risk of acoustic neuroma: results of the Interphne case control study in five North European countries, British Journal of Cancer, Vol. 93, No. 7, p. 842-848.

Presentation really matters

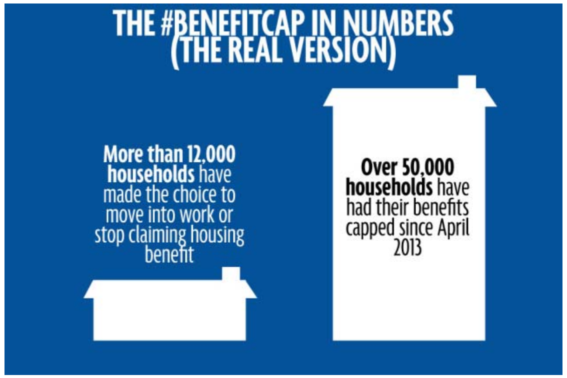
Are the statistics conveyed accurately?
Could the message be clearer?
Are visual effects distorting the image and making things harder for the reader?

How you present numbers has profound consequences. Poorly chosen presentation methods can obscure, mislead and confuse, reducing the impact of your findings and, most importantly, undermining your message. In contrast, **clear and effective presentation can really enhance the quality of evidence.**

In 2014, the DWP press office published this infographic alongside a press release[8]. It was also repeated in a Tweet:



Aside from the claim of causality (we have already seen how challenging it can be to confirm a causal link) the fact that the house on the left (representing 12,000 households) was much larger than the house on the right (representing 50,000) was picked up in a matter of hours by the press (in this case ampp3d in the Daily Mirror)[9]:



Ampp3d's response was widely shared, to the extent that the original messages were entirely lost in the subsequent discussion of inappropriate use of graphics and difficulty of attributing causality in this situation. Worse, the tone of the debate was very much that this was a deliberate attempt to mislead on the part of the government.

This example shows how important it is to **get the graphics right and make sure that the messages you intend to convey are covered appropriately**. Sign off from analyst colleagues before publication is a useful way to minimise the risk of a graphic that could be construed as misleading from appearing.

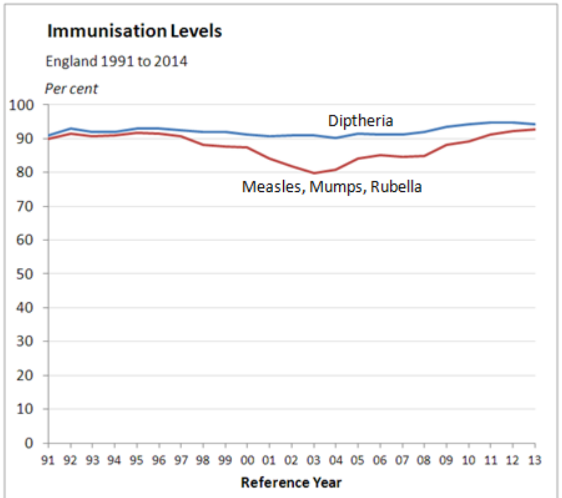
[8] https://www.gov.uk/government/news/benefit-cap-thousands-move-into-work

[9] http://www.mirror.co.uk/news/ampp3d/hey-dwp-just-fixed-your-4584836

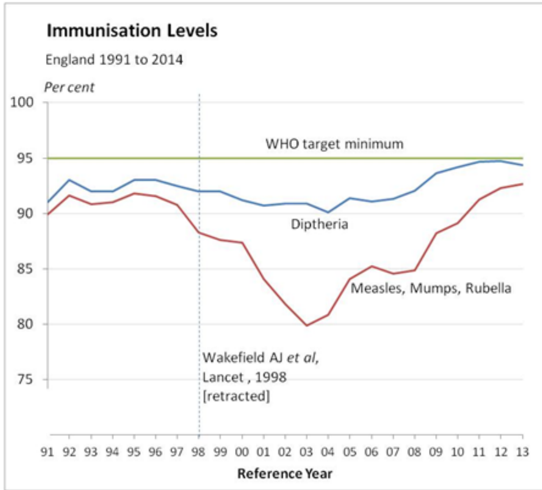
Presentation really matters

Are the statistics conveyed accurately?
Could the message be clearer?
Are visual effects distorting the image and making things harder for the reader?

Good statistical graphics, based on sound statistics, help you to tell the stories in the numbers by bringing out the messages that really matter clearly and coherently.



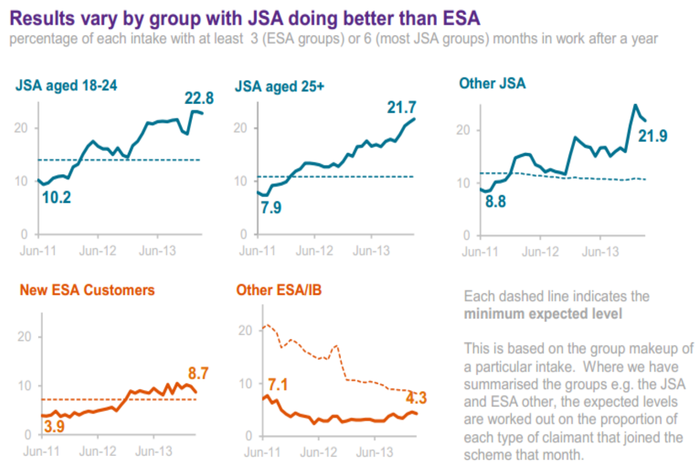
This line graph shows immunisation rates for diphtheria and MMR in England since 1991. There was a discernible dip in the rate for MMR after 1998, but the level seems to be recovering. All in all, the picture is quite reassuring.



Consider this alternative view of the statistics. We've zoomed in on the detail and annotated the graph to provide context.

The picture presented here is more worrying. The fall in the MMR rate associated with the 1998 article (subsequently retracted) by Wakefield and others linking MMR to autism is clear. Both diphtheria and MMR immunisation rates are below the WHO target minimum, and both appear to be levelling off. The WHO target may not be achieved if the trends we see here continue.

Statistical graphics in government are increasingly self contained, and tell the important stories in the data at a glance.



This example from DWP is a case in point. It uses well designed, annotated graphs and an active title to bring out the key messages clearly and simply.