SENSE ABOUT SCIENCE AND STRAIGHT STATISTICS

MAKING SENSE OF STATISTICS

Publications

Other publications by Sense About Science
All are available as free downloads from www.senseaboutscience.org

Making Sense of Testing: why scans and health tests for well people aren't always a good idea

Making Sense of Screening: a guide to weighing up the benefits and harms of health screening programmes.

Making Sense of Radiation: a guide to radiation and its health effects

Making Sense of Chemical Stories: a briefing document for the lifestyle sector on misconceptions about chemicals

Making Sense of GM: what is the genetic modification of plants and why are scientists doing it?

Making Sense of Weather & Climate: an introduction to forecasts and predictions of weather events and climate change

"I've got nothing to lose by trying it": a guide to weighing up claims about cures and treatments

Peer Review and the Acceptance of New Scientific Ideas

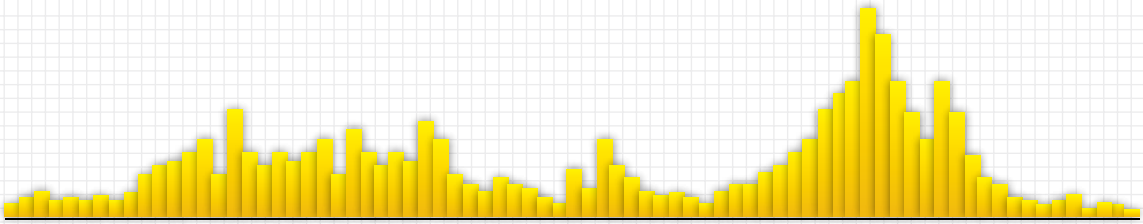
Science and Celebrities Review (2006, 2007, 2008)

Standing up for Science: a guide to the media for early career scientists

Standing up for Science II: the nuts and bolts

There Goes the Science Bit... a hunt for the evidence

"I don't know what to believe": a short guide to peer review



Introduction

good statistics, bad statistics

Statistics are used to measure and make sense of the world. They are produced by the Government, political parties, the civil service, the Bank of England, opinion polls, campaign groups, social research, scientific papers, newspapers and more. But when confronted with stories such as "Crime rate rising again", "Polls put Tories up to 7% ahead" "Child heart surgery halted at hospital after four deaths" or "Swine flu 'could kill up to 120m'" how can we work out whether to believe them and what they really mean?

Statistics can be hyped and sensationalised by the use of an extreme value to make a story more dramatic or by reporting a relative increase in risk without including the absolute change. Data may be analysed and presented in different ways to support contradictory arguments or to reach different conclusions, whether deliberately or by mistake.

But while statistics can be misrepresented, they can also unpick arguments. By knowing the right questions to ask we can discriminate between the proper use of statistics and their misuse. We asked statisticians, journalists and scientists to tell us how they make sense of statistics and what pitfalls to look out for. They gave us the following insights:

- Statistics borrow from mathematics an air of precision and certainty but also call on human judgment and so are subject to bias and imprecision
- Knowing what has been counted, and how, tells us whether a study can really answer the question it addresses
- Like words, numbers and statistics mean different things in different contexts
- Just because something is statistically significant it doesn't mean it is practically significant or of importance to society

This guide is not meant to be a lesson in statistics but a source of questions you can ask and pitfalls to avoid. Knowing something about statistics can help you test and debunk arguments and establish where the truth is likely to lie.

Nigel Hawkes and Leonor Sierra

Contributors

Rosemary A. Bailey is Professor of Statistics at Queen Mary, University of London. Previously, she spent ten years in agricultural research. She was a member of the Royal Statistical Society's working party on First-In-Man Studies.

Harriet Ball completed her PhD in psychiatric genetics at King's College London in 2009. She is part of Sense About Science's Voice of Young Science network (VOYS), investigating issues such as the marketing of "detox" products. She is now studying medicine in Oxford.

Michael Blastland is a writer and broadcaster. He was the creator of Radio 4's More or Less programme about numbers, the author of two critically acclaimed books one about understanding numbers in the news, co-written with Andrew Dilnot, and called *The Tiger That Isn't*.

Simon Briscoe is Statistics Editor at The Financial Times. He is on the editorial board of *Significance* and the executives of Straight Statistics and the Statistics Users Forum. He has written several books on the use of statistics.

Andrew Garratt is press and public affairs officer for the Royal Statistical Society. He combines his knowledge from a science based education with experience working in public relations and political affairs, having worked for MPs and in university public relations.

Nigel Hawkes is the Director of Straight Statistics. As a journalist, he worked for *The Observer* as Science Correspondent and for *The Times* as Science Editor, then Health Editor.

Christina Pagel is a senior research fellow at the Clinical Operational Research Unit at University College London. Christina works on diverse projects including risk models for congenital heart surgery, the reorganisation of mental health care in the UK and modelling the impact of interventions to improve maternal and child health in low-income countries.

Oliver Ratmann is a postdoctoral research associate at the Department of Biology, Duke University, USA. Oliver develops new statistical methods to better understand the emergence and spread of human influenza viruses. He is on the editorial board of *Significance* and a committee member of the Young Statisticians section of the Royal Statistical Society.

Leonor Sierra is Science and Policy Manager at Sense About Science. She joined Sense About Science following a PhD in Physics at the University of Cambridge.

David Spiegelhalter is a statistician who specialises in trying to put numbers on risks. He is Winton Professor of the Public Understanding of Risk at the University of Cambridge, as well as Senior Scientist in the MRC Biostatistics Unit.

Hazel Thornton is an Independent Advocate for Quality in Research and Healthcare, an Honorary Visiting Fellow, Department of Health Sciences, University of Leicester. She campaigns for better communication of `numbers` to combat the dangers of `statistical illiteracy`.

Shaun Treweek is a senior lecturer at the University of Dundee. He is interested in how to design clinical trials that are more relevant to those who will use the results, such as doctors, nurses, patients and managers. He is also part of two international projects that aim to make it easier for health research to find its way into health policy in low and middle-income countries.

Sense About Science (www.senseaboutscience.org) is a UK charity that equips people to make sense of science and evidence.

Straight Statistics (www.straightstatistics.org) is a campaign to improve the understanding and use of statistics by government, politicians, companies, advertisers and the mass media.

Sense About Science and Straight Statistics are grateful to all the contributors and to those who have read through all or part of the document or helped with specific points, including Richard Alldritt, Nina Alphey, James Bedford, Sheila Bird, Claire Coleman, Joanna Ellis, Tom Foxley, Lucy Goodchild, Simon Lacoste-Julien, Catherine de Lange, Richard Laux, Nick Ross, Tom Sheldon, Jessica Strangward and Rose Wu.

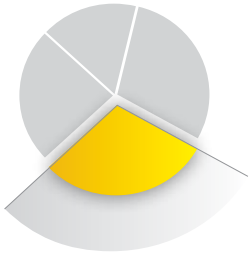
This booklet is published by Sense About Science and Straight Statistics, which have final responsibility for the content.

Contents

p.06 - p.07

Section 1: If a statistic is the answer, what was the question?

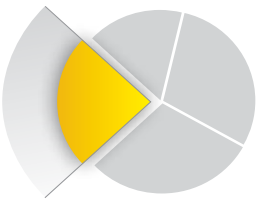
Statistics are the product of conscious choices: what to count, how to count it and how to express the results. To understand them we need to consider what choices were made when the study was designed, ask how big the sample was, how it was chosen and, in making projections or forecasts, what assumptions were used.



p.08 - p.10

Section 2: Common pitfalls

There is more than one type of average; each one can give a different answer and we need to find out why a particular one was used. To make a story more dramatic, people regularly use the most extreme number from a range of possible values, that is, an outlier – a possible but not very likely value.



p.11 - p.12

Section 3: How sure are we?

Statisticians check if a result is consistent with chance or if it is 'statistically significant'. Even if a result is statistically significant it doesn't mean it is practically significant or of importance to society. Confidence intervals give the scale of potential uncertainties in counting, measuring or observing data. Just because there has been a run of events – deaths at a hospital, accidents on a road, draws in football matches – it doesn't necessarily mean that something beyond chance has caused it.

p.13 - p.14

Section 4: Percentages and risk; knowing the absolute and relative changes

To understand the importance of any increase or decrease we need to know both the absolute and relative change. To know if a change in risk matters to an individual we need to know what the risk was to begin with.

1. If a statistic is the answer, what was the question?

Statistics is the science of making use of numerical data. Statistics are used to forecast (the weather, the economy), to analyse past events (has the crime rate gone up or down?) and to make decisions (should a new drug be available on NHS?).

Statistics are the product of conscious choices: what to count, and how to count it. The results borrow from mathematics an air of precision and certainty. But choosing what to count, and how, comes down to human judgement about the best way to get the answer to the question, whether it is by designing and carrying out an experiment, a survey, a poll, a clinical trial, an observational study or a census.

"How big is the gap between the earnings of men and women? According to the Office for National Statistics (ONS), it is 12.8%. But the Government Equalities Office (GEO) says it is 23%. And the Equality and Human Rights Commission (EHRC) says it's 17.1%[1].

The differences in these figures arise from the different methods used to produce them: the ONS includes only full-time employees, excluding overtime and part-time workers. The GEO includes part-time workers because it says more women than men work part-time and it is wrong to exclude them. The EHRC figure uses the ONS data but compares the mean salaries not the median. It justifies this by saying than men are over-represented at one extreme of the earnings range, and women at the other.

Three figures – all of them right – but asking what is being compared and how it was calculated tells us why there is a difference." **Nigel Hawkes**

It can be too difficult or not practical to make a complete count – it is impossible to know the exact number of illegal immigrants, for example, and we can only estimate the number. The only time when the whole population is asked for information is during the national census, which takes place every ten years and provides figures from a national to neighbourhood level. And even during the census not everyone will be counted.

[1] UK Statistics Authority publication Gender Pay Gap (11 June 2009)

The only time every household within the population is asked for information is during the national census, which takes place every ten years and provides figures from a national to neighbourhood level for the Government, local authorities, business and others. And even during the census not everyone will be counted.

We can usefully make statements about a particular group or population by using a sample. But the selection of the sample has to be carefully considered. If we wanted to find out what time Londoners wake up, for example, and carried out a survey at underground stations at 7am the results would not adequately answer the question. When surveys are reported statisticians look at what has been counted, and how, before assessing whether the conclusions drawn are fair and valid.

"*Literary Digest* carried out a survey before the 1936 US Presidential Election. It mailed out millions of ballot papers and got two million back; a huge sample, most of which backed the Republican candidate Alf Landon. But the addresses to which they had been sent came from a directory of car owners and from the telephone directory: a biased sample, since in 1936 only the better-off owned cars or had telephones. Franklin D Roosevelt, the Democrat, won the election in a landslide."

There is no magic number that tells you how large a sample should be to give valid results. It depends on what information you are trying to extract. For a political poll 1,000 people are normally considered enough. A trial of a drug however might need many thousands of participants to detect small benefits and identify rare side effects. Deciding how large a sample must be is a central part of statistical science.

To predict future events such as changes in the climate or the world population statistical models are used. Using models to forecast possible future scenarios means making assumptions using the evidence currently available. These scenarios can then be disputed because the same model run with different assumptions can give different results.

There will always be uncertainty in projections. For example, the projections for the number of deaths from swine flu depended on both the mortality rate and the number of people infected. It turned out that both were over-estimated and the number of deaths was much lower than projected.

Once a projection or forecast is produced it may lead to an intervention, which can change the course of events. Reality can therefore turn out differently from the forecast. In this case, it does not mean that the forecast was 'wrong'.

Figures aren't always generated from models or surveys; administrative data can also be used. We might use the number of people that claim benefits for unemployment, for example, to calculate how many people are unemployed. In this case the figure would not be accurate, not because of flaws in designing and conducting a study, but because we counted how many people claimed benefits not how many were unemployed.

Whatever the statistic we must also consider whether it is realistic given what else we know. When it was claimed that in the ten most deprived areas in the UK 54% of teenage girls were likely to fall pregnant before the age of 18, it didn't take long for people to realise this could not be true – it would mean over half of teenage girls from these areas being pregnant. The real figure was 5.4%.

The decisions that go into collecting data (how big was the sample, how it was chosen, how the questions in a survey are phrased, how the experiment was designed, was there a control group and the assumptions that form the basis of a model) affect the results. This information can help us see when a study may have been designed spuriously to achieve a particular answer, can't tell us the definitive result, or it is technically flawed.

2. Common pitfalls

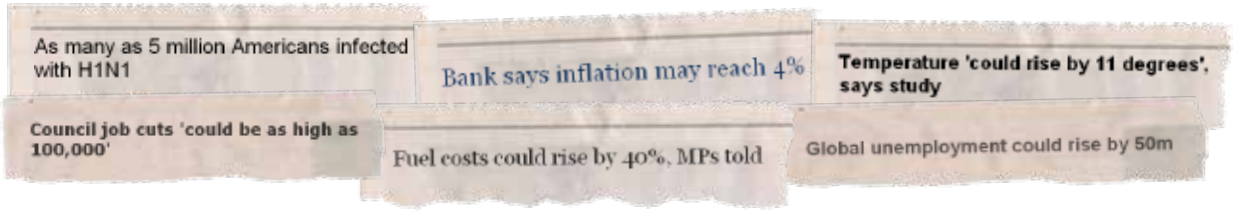
The way statistical data is summarised or presented can lead to wrong conclusions being drawn even if the statistics are correct.

'Average' in news reports is often used to refer to the mean (that is, the sum of the listed values divided by the number of values in the list). But the mean doesn't always give us the most useful information: we might need the median (the middle point) or the mode (the most common value). If we wanted to know the typical salary earned in the UK the most representative value would be the median salary. This is because a few thousand people earning millions of pounds a year will affect the mean more than the multitudes of people earning tens of thousands of pounds a year.

For example, if you had a room with ten teachers all earning between £20,000 - £30,000, with a mean salary of £24,900 and a median salary of £25,000 and then someone who earns a million pounds walked into the room, the mean would increase to £114,000 but the median would hardly change. By using the median this distortion is reduced, providing a more representative average salary.

"You might want to know the average number of children in each household in a particular place. You work out that the mean is 2.3 by adding up all the children in the area and dividing by the number of households. But this number wouldn't tell you that the most common household setup has no children (that is, the mode is 0). If you ordered the households according to how many children they had and the household in the middle had just one child,the median would be 1." **Harriet Ball**

Averages are useful ways of encapsulating data, but when averages are usedconsider what sort of average it is and whether it is representative of what we are trying to find out.



Most projections are given as a range, because we can't know precisely what is going to happen in the future. When the emphasis is on the most extreme value – "as many as 5 million Americans infected with H1N1", "Council job cuts 'could be as high as 100,000'" – it sounds alarming until you consider the likelihood (the probability) of what's being reported actually happening.

"'Could be as high as', 'may reach' ... phrases such as these hint at mischief at work, depending as they do on the most extreme possibilities rather than the most likely. Headlines like these might more accurately read: 'could be, but most likely won't'.

"Temperature could rise by 11 degrees C"[2]. This was one result from a model of climate sensitivity to rising CO_2 levels. But the model was run 2,000 times and this outcome was generated only once. The most common result was that temperature would rise by 3°C . Whilst 11°C is possible it's not the most likely result, but it was still widely reported. When reading such stories it is important to consider is it likely, given what else we know?" **Michael Blastland**

Big numbers can be difficult to get to grips with because the majority of us do not use millions and billions in our daily lives. To make them meaningful we need to divide them by the number of items they relate to.

[2] www.scidev.net/en/news/temperature-could-rise-by-11-degrees-says-study.html

"A million pounds is a fortune to you and me, but peanuts in the national accounts. The British Government recently announced £250 million over five years to 'kick start a green revolution in transport' in the UK. That works out at less than £1 per person per year.

If a new figure is announced for public spending divide it by the population, and then divide by the number of weeks in the year. The result tells you how much the figures represents, per week, for every individual." **Michael Blastland**

The Taxpayers' Alliance hit the headlines in June 2009 with its claim that civil servants' taxi bills amounted to £8 million a year, a seemingly over the top amount. But there are hundreds of thousands of civil servants. If we take just the top civil servants, roughly 20,000 people, and share between them the £8 million spent, their annual bill would be £400 each, or £8 a week. That's one short taxi ride a week at London prices.

The cost of the NHS is £110 billion a year but divide that by the population of the UK – approximately 60 million – and it works out at just under £2,000 a head - men, women and children. This is a more manageable number to consider.

"All numbers, big and small, need to be put in context to be meaningful: a day that is 1°C warmer than the previous one is hardly different but a 1°C rise in the average temperature of the Earth would distort the whole ecosystem. A disease with a 0.1% mortality rate doesn't sound too worrying until you consider that it could still cause the deaths of thousands of people every year if it is a common disease."**Oliver Ratmann**

3. How sure are we?

When evaluating data statisticians have to decide whether the results they are seeing are 'statistically significant'.

What does 'statistically significant' mean?

"To be honest, it's a tricky idea. It can tell us if the difference between a drug and a placebo or between the life expectancies of two groups of people, for example, could be just down to chance.

If a result is statistically significant we can be fairly confident that something real is going on. It means that a difference as large as the one observed is unlikely to to have occurred by chance alone.

Statisticians use standard levels of 'unlikely'. Commonly they use significant at the 5% level (sometimes written as p=0.05). In this case a difference is said to be 'significant' because it has a less than 1 in 20 probability of occurring if all that is going on is chance." **David Spiegelhalter**

Knowing whether something passes this test does not conclusively tell us whether or not it's important or practically significant, that the study is a good study, or that one thing causes another – it's simply a yardstick used by statisticians to evaluate possible conclusions.

Another way statisticians express how sure they are about an inference from data is to include 'confidence intervals', which give the range within which they are 90 or 95 or 99 % confident the true answer lies. They give a scale of potential uncertainties in counting, measuring or observing data, with a lower and an upper limit.

Confidence intervals are used in opinion polls. A sample of 1,000 representative voters will give a result for each party with confidence intervals stretching from three percentage points below the survey result to three percentage points above it. If a poll says Party X has the support of 32% of the voters, this could be as low as 29% or as high as 35 %. In this case the range of six percentage points from 29 to 35 is the confidence interval.

Difficulties arise when one poll is compared with another, and the confidence intervals are ignored. A two percentage point 'fall' in a political party's popularity may be headlined, but this change lies within the confidence interval and so we can't be sure that there has been any real change at all.

All other things being equal the larger the sample, the smaller the margins of error (or confidence interval). But we cannot always have large samples; there might be only a few people with the disease being studied; it is also costly and time consuming to run large trials. Statistics allows us to use a sample to draw wider conclusions, but we need to be aware of any limitations and check that any interpretation of the results has acknowledged these.

"In 2007 an article reported that the risk of heart attack or stroke was nine times higher for some individuals being treated for osteoarthritis when taking ibuprofen rather than the drug Lumiracoxib.

What wasn't mentioned was the confidence interval, which ranged from 1.13 to 72 times greater risk of stroke or heart attack. This is a very large range because it is based on a small number of individuals who had an adverse reaction. A result on its own is half the picture; for the full picture you need to know about the level of uncertainty."
Shaun Treweek

Statisticians also lookout for 'regression to the mean' – that is the tendency following an extreme measurement for further measurements to be closer to the average. A footballer scoring hat-tricks in a couple of matches is likely to return to scoring his usual number of goals over time.

Regression to the mean

Regression to the mean can lead to false conclusions: speed cameras are often located in places where there have been traffic accidents and so a run of accidents makes it more likely that a camera will be placed there. Any subsequent reduction in the accident rate is credited to the effect of the camera.

But part of the improvement could actually be regression to the mean. A run of accidents on any stretch of road may simply be the result of chance. Over time, accidents along this stretch will tend to become closer to the average. And since the trigger for placing the camera will have been a high level of accidents, the most likely course of events is for these to decline – whether the camera is there or not.

4. Percentages and risk; knowing the absolute and relative changes

"Bacon butties give you colon cancer", "passive smoking causes dementia", "knife killings at new high" – how should we react to such stories? We need to work out how many people are really affected and what our individual risks are.

Take the story about bacon sandwiches giving you colon cancer. It reported a 20% rise in the risk of colon cancer from eating red or processed meat. This may sound alarming, but what does it mean? It depends on how large your risk of getting colon cancer is to start with.

A person has around a 5% chance of getting colon cancer during their lifetime – the 'absolute' risk. If you eat a bacon sandwich every day you increase your risk of getting colon cancer by 20% – the 'relative' risk increase. So what it means is that your lifetime risk of getting colon cancer is now 6%, an increase of 1% (that is, 20% of 5% = 1%).

"Why do reports prefer to talk about relative percentage risks without mentioning the absolute risk? The suspicion must be that this allows the use of 'bigger numbers': 20% is big enough to be a scare, the absolute change 1% or even 1 person in every 100, is less disturbing." **Michael Blastland**

Changing the 'framing' of the risk gives us a different perspective on that risk. Knowing the relative and absolute changes tells us how big the risk is and how it might affect us.

"Recent newspaper articles reported that '9 in 10 people carry a gene that increases chance of high blood pressure'. The actual study had found a genetic variant in 1 in 10 people that reduced the risk of high blood pressure – this would not have received much press coverage, but by shifting the attention to the group who did not have this reduced risk the story became 'news'." **David Spiegelhalter**

Representing something as a proportion does not tell us what the absolute numbers are. A Whiskas advert told us that "8 out of 10 cats prefer Whiskas", but not how many cats had been asked... This was then changed to "8 out of 10 owners that expressed a preference said their cats preferred Whiskas". A more likely situation but we're still not told how many owners were asked and how many of these expressed an opinion.

Or, for example, the actual number of violent crimes might not have changed in a few years, but if the number of other types of crime has gone down then violent crimes as a proportion of the total numbers of crimes will increase.

"Before reacting to a percentage you have to think what it is really telling you and to do that you need to put it in context. Take, for example, the statistic that 99% of deaths in the first four weeks of life occur in developing countries. Although that sounds horrifying, around 90 per cent of all births take place in developing countries. And so the chances of a baby dying in its first four weeks are 'only' 11 times greater there – bad enough, in all conscience." **Christina Pagel**

Further information and useful reading

Organisations involved in the scrutiny of statistics:

The Royal Statistical Society (www.rss.org.uk) is a learned and professional society, founded in 1834, devoted to promoting the public understanding of statistics, and providing support to those who use statistics in their professional lives. It publishes an internationally renowned peer-reviewed journal and a magazine (*Significance*) aimed at a popular audience.

Straight Statistics (www.straightstatistics.org) is a campaign established by journalists and statisticians to improve the understanding and use of statistics by government, politicians, companies, advertisers and the mass media. It aims to restore public confidence in statistics by exposing bad practice and rewarding good.

The UK Statistics Authority (www.statisticsauthority.gov.uk) is an independent body operating at arm's length from the Government, to promote and safeguard the production and publication of official statistics that serve the public good. Its two main functions are the oversight of the Office for National Statistics (ONS) – responsible for the production of a wide range of economic and social statistics – and the independent scrutiny (monitoring and assessment) of all official statistics produced in the UK.

Books, websites and programmes about understanding statistics:

NHS Choices: Behind the Headlines http://www.nhs.uk/news/Pages/NewsIndex.aspx
This site is intended for the public and health professionals provides an evidence-based analysis of health stories that make the news.

Stats: take a quantitative leap. George Mason University http://stats.org/
This site provides resources for journalists and policy makers on the use and abuse of science and statistics in the media.

Understanding Uncertainty http://understandinguncertainty.org/
This site looks at how uncertainty plays out in the real world, and how a bit of understanding about probability can allow us to look critically at stories in the news.

The Tiger that Isn't: Seeing through a world of numbers. Blastland, M and Dilnot, A. London: Profile Books Ltd, 2008. A layman's guide to life in numbers using vivid and everyday news stories and ideas.

Panicology. Briscoe, S and Aldersley-Williams, H. London: Viking, 2008.
Two statisticians explain what's worth worrying about (and what's not) in the 21st century.

Statistics: A very short introduction. Hand, D. Oxford: Oxford University Press, 2008.
This Very Short Introduction sets the study of statistics in context, describing its history and impact, and how the field of statistics has become centrally important to how we understand our world.

How to lie with statistics. Huff, D. London: W.W.Norton & Company Ltd., 1993.
An introduction to statistics for the general reader, aimed to dispel our 'terror in numbers'.

More or Less (http://news.bbc.co.uk/1/hi/programmes/more_or_less/default.stm)
is a BBC Radio 4 series, devoted to the 'world of numbers.'

Contact a scientist

If you are...

a commentator, medical charity, journalist, employer, educator, information service, local authority, national body, parliamentarian, health care provider, professional association, community group, NGO, lifestyle writer
...or any other civic group in need of help or comment on a difficult or controversial area of science call Sense About Science on
020 7478 4380

For more copies or further information, contact us on
enquiries@senseaboutscience.org

Published in 2010 by

sense about science

Sense about Science
Registered Charity No. 1101114
25 Shaftesbury Avenue
London W1D 7EG
www.senseaboutscience.org

straight Statistics
The campaign for the honest use of statistics

Straight Statistics
Company No: 6822790
8/9 Well Court
London EC4 9DN
www.straightstatistics.org

This document is licensed under Creative Commons Attribution-Noncommercial-No Derivative Works 2.0 UK: England & Wales License. (cc)

In collaboration with the Royal Statistical Society

ROYAL STATISTICAL SOCIETY

design Peel me www.peel-me.com