

Top 20 things policy makers need to know about science

The following tips, originally addressed to politicians and first published in *Nature*, were compiled by William Sutherland, a zoologist, and David Spiegelhalter, a statistician – both from the University of Cambridge – and Mark Burgman, an ecologist at the University of Melbourne.

The trio argued the “immediate priority is to improve policy makers’ understanding of the imperfect nature of science” by suggesting 20 concepts that should be taught to government ministers and public servants.

1. Differences and chance cause variation

The real world varies unpredictably. Science is mostly about discovering what causes the patterns we see. Why is it hotter this decade than last? Why are there more birds in some areas than others? There are many explanations to such trends, so the main challenge of research is teasing apart the importance of the process of interest (for example, the effect of climate change on bird populations) from the innumerable other sources of variation.

2. No measurement is exact

Practically all measurements have some error. If the measurement process were repeated, one might record a different result. In some cases, the measurement error might be large compared with real differences. Thus, if you are told that the economy grew by 0.13% last month, there is a moderate chance that it may actually have shrunk.

3. Bias is rife

Experimental design or measuring devices may produce atypical results in a given direction. For example, determining voting behaviour by asking people on the street, at home or through the internet will sample different proportions of the population, and all may give different results. Because studies that report “statistically significant” results are more likely to be written up and published, the scientific literature tends to give an exaggerated picture of the magnitude of problems or the effectiveness of solutions.

4. Bigger is usually better for sample size

The average taken from a large number of observations will usually be more informative than the average taken from a smaller number of observations. That is, as we accumulate evidence, our knowledge improves.

5. Correlation does not imply causation

It is tempting to assume that one pattern causes another. However, the correlation might be coincidental, or it might be a result of both patterns being caused by a third factor – a “confounding” or “lurking” variable. For example, ecologists at one time believed that poisonous algae were killing fish in estuaries; it turned out that the algae grew where fish died. The algae did not cause the deaths.

6. Regression to the mean can mislead

Extreme patterns in data are likely to be, at least in part, anomalies attributable to chance or error. The next count is likely to be less extreme. For example, if speed cameras are placed where there has been a spate of accidents, any reduction in the accident rate cannot be attributed to the camera; a reduction would probably have happened anyway.

7. Extrapolating beyond the data is risky

Patterns found within a given range do not necessarily apply outside that range. Thus, it is very difficult to predict the response of ecological systems to climate change, when the rate of change is faster than has been experienced in the evolutionary history of existing species, and when the weather extremes may be entirely new.

8. Beware the base-rate fallacy

The ability of an imperfect test to identify a condition depends upon the likelihood of that condition occurring (the base rate). For example, a person might have a blood test that is “99% accurate” for a rare disease and test positive, yet they might be unlikely to have the disease.

9. Controls are important

A control group is dealt with in exactly the same way as the experimental group, except that the treatment is not applied. Without a control, it is difficult to determine whether a given treatment really had an effect. The

control helps researchers to be reasonably sure that there are no confounding variables affecting the results.

10. Randomisation avoids bias

Experiments should, wherever possible, allocate individuals or groups to interventions randomly. Comparing the educational achievement of children whose parents adopt a health program with that of children of parents who do not is likely to suffer from bias.

11. Seek replication, not pseudoreplication

Results consistent across many studies, replicated on independent populations, are more likely to be solid. The results of several such experiments may be combined in a systematic review or a meta-analysis to provide an overarching view of the topic with potentially much greater statistical power than any of the individual studies.

12. Scientists are human

Scientists have a vested interest in promoting their work, often for status and further research funding, although sometimes for direct financial gain. This can lead to selective reporting of results and occasionally, exaggeration. Peer review is not infallible: journal editors might favour positive findings and newsworthiness. Multiple, independent sources of evidence and replication are much more convincing.

13. Significance is significant

Expressed as P, statistical significance is a measure of how likely a result is to occur by chance. Thus $P = 0.01$ means there is a 1-in-100 probability that what looks like an effect of the treatment could have occurred randomly, and in truth there was no effect at all. Typically, scientists report results as significant when the P-value of the test is less than 0.05 (1 in 20).

14. Separate no effect from non-significance

The lack of a statistically significant result (say a P-value > 0.05) does not mean that there was no underlying effect: it means that no effect was detected. A small study may not have the power to detect a real difference. For example, tests of cotton and potato crops that were genetically modified to produce a toxin to protect them from damaging insects suggested that there were no adverse effects on beneficial insects such as pollinators. Yet

none of the experiments had large enough sample sizes to detect impacts on beneficial species had there been any.

15. Effect size matters

Small responses are less likely to be detected. A study with many replicates might result in a statistically significant result but have a small effect size (and so, perhaps, be unimportant). The importance of an effect size is a biological, physical or social question, and not a statistical one.

16. Data can be dredged or cherry picked

Advertisement

Evidence can be arranged to support one point of view. To interpret an apparent association between consumption of yoghurt during pregnancy and subsequent asthma in offspring, one would need to know whether the authors set out to test this sole hypothesis, or happened across this finding in a huge data set.

17. Extreme measurements may mislead

Any collation of measures (the effectiveness of a given school, say) will show variability owing to differences in innate ability (teacher competence), plus sampling (children might by chance be an atypical sample with complications), plus bias (the school might be in an area where people are unusually unhealthy), plus measurement error (outcomes might be measured in different ways for different schools). However, the resulting variation is typically interpreted only as differences in innate ability, ignoring the other sources.

18. Study relevance limits generalisations

The relevance of a study depends on how much the conditions under which it is done resemble the conditions of the issue under consideration. For example, there are limits to the generalisations that one can make from animal or laboratory experiments to humans.

19. Feelings influence risk perception

Broadly, risk can be thought of as the likelihood of an event occurring in some time frame, multiplied by the consequences should the event occur. People's risk perception is influenced disproportionately by many things, including the rarity of the event, how much control they believe they have, the adverseness of the outcomes, and whether the risk is voluntarily or not.

For example, people in the US underestimate the risks associated with having a handgun at home by 100-fold and overestimate the risks of living close to a nuclear reactor by 10-fold.

20. Dependencies change the risks

It is possible to calculate the consequences of individual events, such as an extreme tide, heavy rainfall and key workers being absent. However, if the events are interrelated, (for example a storm causes a high tide, or heavy rain prevents workers from accessing the site) then the probability of their co-occurrence is much higher than might be expected.