Gibbison, B. (2017). Fifty percent of anaesthetists are worse than average at understanding statistics and risk. *Anaesthesia*, *71*(1), 118-121. https://doi.org/10.1111/anae.13690

**EDITORIAL**

## Fifty percent of anaesthetists are worse than average at understanding statistics and risk.

*B. Gibbison*
*Consultant in Cardiac Anaesthesia and Intensive Care*
*Honorary Senior Lecturer*
*Bristol Heart Institute*
*University Hospitals Bristol NHS Foundation Trust / University of Bristol*
*Bristol, UK*

Address for correspondence:
Dr Ben Gibbison
Department of Anaesthesia
Level 7 Bristol Royal Infirmary
University Hospitals Bristol NHS Foundation Trust
Upper Maudlin St
Bristol
UK
BS2 8HW
Email: mdbjjg@bristol.ac.uk

The UK referendum on whether to leave or remain in the European Union has brought statistics, risk and uncertainty back into the parlance of the general public. They have also learnt that statistics without context can be misleading, tolerance of an acceptable risk is opinion based and that both financial markets and individuals struggle to deal with uncertainty.

Statistics, risk and uncertainty form the basis of medicine – but there are questions over how much we really understand them, not just the nuts and bolts of the equations, but what the numbers *really* mean in context. How we can use these numbers to make better decisions for our patients and finally, how we can communicate these abstract concepts derived from (usually) populations into something meaningful for individual patients?

We may not like it, but uncertainty governs our world. The behaviour of even the smallest sub-atomic particle is subject to "Heisenberg's uncertainty principle". That is that one can only really know either where the particle is *or* its momentum, but not both. This can be put more simply as "everything that can happen does happen"[1] - it's just that some things are so unlikely to occur that we can assume they won't in any meaningful time-frame. Time-frames are incredibly important when dealing with probabilities.

Statistics and probability are the two disciplines that underpin the science of uncertainty and provide the tools to handle uncertainty in a structured manner. Statistics is the collection, analysis, presentation and interpretation of data. Probability is the chance that an event will occur. There is a perception that statistics and probability are "maths" and therefore difficult to understand. But mathematics is just a structured way of dealing with logical processes. It allows us to communicate 'quite a lot', 'a lot' and 'more likely than not' with precision[2]. Numeracy is clearly important, but only a broad knowledge of mathematics and an understanding of some simple principles is required to use statistics and probabilities to inform our practice. Studies show that doctors as a group are not good at this. Sixty-nine percent of American primary care physicians presented with evidence for two screening tests failed to recommend the correct test when questioned [3]. In 1995 the UK Committee on the Safety of Medicines issued a document stating that oral contraceptive pills containing *desogesterel* or *gestodene* were twice as likely to cause venous thrombo-embolism than other progestogen containing pills resulting in a large rise in both conceptions and abortions[4]. This was a large *relative* risk that translated to an *absolute* risk change from 1 in 7, 000 to 2 in 7, 000 women – a small number.

Misunderstanding of statistics is seen within the scientific manuscripts themselves. Journals are peppered with the phrase that something shows a "trend" towards statistical significance. A p-value of 0.06 does **not** show "a trend" towards statistical significance. A p-value describes the probability that the data conforms to the null hypothesis. Because we use p-values as a binary outcome – "significant" or "not significant", it allows us to remove any uncertainty from the question we have asked. It should not. This is an example of the psychological phenomenon of “attribute substitution” where people try and answer a complex question by answering a similar, but much easier question. A difficult question might be “which fluid is most likely to improve this patient’s long term outcome, given the risks and benefits of each in this circumstance and in this patient” and the easy question is: “is there a statistically significant difference between saline and albumin in the SAFE study[5]”?

Indeed, the reliance on the p-value in biological science has recently been criticised by the American Statistical Association[6] and has been postulated as one reason for the large number of studies that cannot be reproduced[7]. None of the statistical errors above are a result of poor arithmetic; rather they are a failure to understand the definitions and principles behind them. In an age where huge amounts of data can be dealt with in short periods of time by algorithms, being able to work out p-values is not important – but understanding where the numbers come from and where the errors might lie is. “Anaesthesia” has made an attempt to do this with its “Statistically Speaking”[2,8] series, which is a clear, simple and accessible approach to understanding the statistics behind all of our practices.

“Risk” is one of the most used terms in anaesthesia. Most people use it to mean the probability of a perceived negative outcome – although this is not necessarily the case. Risk is simply the potential of both losing *or* gaining something of value and so there are benefits to risk, for example the trial of a new medicine. Risk is created when there is uncertainty over an event. There are two classes of uncertainty (aleatory and epistemic) and because risk is a type of uncertainty, this feeds into risk. Aleatory uncertainty (randomness) is that that comes from a random process; will we flip heads or tails on a coin. No matter what we do, this sort of uncertainty cannot be reduced. The other type of uncertainty is epistemic uncertainty (ambiguity). This comes from a lack of information about the process or the system in question. We can reduce the epistemic uncertainty by knowing more about the system. This is the sort of uncertainty that we reduce by improving our sample size or composition, or by improving our measuring instruments. We should bear this in mind when using risk in clinical practice – could we know more about the risk to enable us to quantify it better, or is there just a random factor that we have no control over? Randomness is not as intuitive a concept as we might expect and can co-exist with ambiguity within the same system. The interaction between the two has been discussed within the pages of *Anaesthesia* this year[9].

So how do we use numerical statistics and probabilities that describe and are derived from populations to answer questions about individuals? My interest in this came from work looking at *ultradian* rhythms of the steroid hormone cortisol after major surgery and critical illness. There is a diurnal rhythm of cortisol and adrenocorticotrophic hormone (ACTH) production in the body that is highest in the early morning and lowest in the late afternoon. However, frequent measurements of cortisol and ACTH show that both hormones are pulsatile with pulse durations of about an hour[10]. The peaks of the circadian rhythm are formed by frequent, large amplitude pulses and the troughs by smaller pulses or no pulses at all. This pulsatility persists during and after major surgery[11], but with an altered pattern. To see these pulses, individual hormone profiles must be examined. Aggregated population data means that the peaks of some people are cancelled out by the troughs of others. Therefore, there is a situation whereby the aggregated mean, mode and median not only fail to describe the outliers in the population, they describe *no one* in the population. However, most of the decisions we make for individuals are based on point averages derived from a population. This coupled with the assumption that the models we use are linear, rather than biphasic, exponential, or a combination of many[2], means that the decisions we make are not as good as they could be. Should I give this patient blood? Is giving two units of blood twice as effective as one unit of blood? Should I just wait until the haemoglobin drops below 70g/L or is there a better way? This does not fly against protocol driven care – protocols help teams work more effectively and provide guidance, but as senior clinicians, it is incumbent on us to make better *individualised* decisions. We have to *know* the evidence and use the statistics, with the associated risk and probability to do so.

When working out how to use the aggregated data for the individual, what we are essentially doing is 'forecasting' – how much increase in cardiac output will this inotrope bring without side effects? What is the benefit of giving this patient fluid and which one should I give? This is a skill that is used throughout business, science and politics. The US intelligence services were so worried by how poor their 'expert' intelligence analysts were at forecasting that there were no weapons of mass destruction in Iraq, they worked with psychologists at the University of Pennsylvania to try and improve[12]. They used forecasting 'tournaments' where participants tried to predict the probability of rare political events. The participants were not experts, but some of them persistently outperformed the experts. When their approach to forecasting was analysed, 'super-forecasters'[13] used many similar features that can frame how to use risks and probabilities well in clinical practice:

- Use all of the evidence and weight it accordingly.
- Break seemingly intractable problems into more manageable sub-problems.
- Be probabilistic about thinking – very few things are 'certain' or 'impossible'. This also means that individual predictions will be wrong sometimes. Correctly predicting a 75% chance of death on the intensive care unit means that 25% of similar patients will survive – be wrong for the right reasons.
- Be granular about the risk – know the difference between a 60/40 risk and a 55/45 risk. Both of these are 'more likely than not'.
- When making predictions, start at the base-rate and narrow down for the individual circumstance.
- Update predictions regularly and don't be afraid to change your mind as the information and evidence changes.

- Make forecasts within a time-frame. It is easy to predict that something will happen at some point – “everything that can happen does happen”. Most risk data are allied with a time-frame; lifetime risk of cancer, in-hospital risk of mortality, improvement in cardiac output within 24 hours. Doing this allows monitoring predictions and performance feedback.
- Look for the errors behind mistakes in prediction, but be wary of hindsight bias.

The points above that are used to guide our thinking as clinicians can also help communicate these probabilities, risks and uncertainties to patients. The same is true in reverse. The persisting theme is that when numbers are presented in a meaningful context with uncertainties, they aid rather than hinder communication[14]. Chances of single events should be presented as simple frequencies with time-frames and an explicit denominator; “70 out of 100 patients with pneumonia as severe as your husband will die in hospital”, rather than “70% of patients like your husband will die”. Keeping the same denominator throughout a conversation is also very helpful. Denominators and base-rates are the key to understanding changes in risk; we should try to present an incremental absolute risk after showing a base-rate; “…about 30 in 100 patients will be sick after an operation. Giving ondansetron to 100 patients before surgery means that 7 fewer people will be sick whilst 93 will not benefit (23 will still vomit and 70 will still not vomit)”. Conveying the uncertainties inherent in probabilities are difficult and the differences between aleatory and epistemic uncertainty harder, with little evidence to guide us. Psychological studies show in healthcare that which we know from world outside; uncertainty leads to avoidance of decision making, pessimistic risk perceptions and worry related to the outcome of the final choice[15,16]. Doctors are at as great a risk of this as their patients.

Statistics, probabilities and risks are abstract human constructs on the world around us. They aim to put some objectivity into an uncertain world. However, there will always be some subjectivity to their interpretation and this is particularly difficult when comparing risks across different time frames. The first stage in trading off risks and probabilities is understanding the statistical principles that underpin them and thinking of the world in a more probabilistic way. Applying binary, linear models makes our decisions easier, but not necessarily better.

**References**


1. Cox B, Forshaw J. The Quantum Universe: Everything that can happen does happen. Penguin. London 2012

2. Poole D, Carlisle JB. Mirror, mirror on the wall…predictions in anaesthesia and critical care. *Anaesthesia* 2016;71:1104–9.

3. Wegwarth O, Schwartz LM, Woloshin S, Gaissmaier W, Gigerenzer G. Do physicians

understand cancer screening statistics? A national survey of primary care physicians in the United States. *Annals of Internal Medicine* 2012;156:340–9.

4. Furedi A. The public health implications of the 1995 'pill scare'. *Human Reproduction Update* 1999;5:621–6.

5. Finfer S, Bellomo R, Boyce N, French J, Myburgh J, Norton R. A comparison of albumin and saline for fluid resuscitation in the intensive care unit. *New England Journal of Medicine* 2004;350:2247–56.

6. Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician.* 2016;70:129–33.

7. Ioannidis JPA. Why Most Published Research Findings Are False. *PLoS Medicine*; 2005; 2: e128

8. Choi SW, Tran DHD. Heart attacks prevent chocolate. Or, the differences between correlation and causation. *Anaesthesia* 2016;71:586–9.

9. Pandit JJ. Deaths by horsekick in the Prussian army - and other "Never Events" in large organisations. *Anaesthesia*. 2015;71:7–11.

10. Walker JJ, Terry JR, Lightman SL. Origin of ultradian pulsatility in the hypothalamic-pituitary-adrenal axis. *Proceedings of the Royal Society B*; 2010;277:1627–33.

11. Gibbison B, Spiga F, Walker JJ, et al. Dynamic pituitary-adrenal interactions in response to cardiac surgery. *Critical Care Medicine* 2015;43:791–800.

12. Dhami MK, Mandel DR, Mellers BA, Tetlock PE. Improving Intelligence Analysis With Decision Science. *Perspectives on Psychological Science*. 2015;10:753–7.

13. Mellers B, Stone E, Murray T, et al. Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions. *Perspectives on Psychological Science*. 2015;10:267–81.

14. Trevena LJ, Zikmund-Fisher BJ, Edwards A, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. *BMC Medical Informatics and Decision Making* 2013;13:S7–15.

15. Han PKJ, Moser RP, Klein WMP. Perceived ambiguity about cancer prevention recommendations: associations with cancer-related perceptions and behaviours in a US population survey. *Health Expectations* 2007;10:321–36.

16. Han PKJ, Kobrin SC, Klein WMP, Davis WW, Stefanek M, Taplin SH. Perceived ambiguity about screening mammography recommendations: association with future mammography uptake and perceptions. *Cancer Epidemiology, Biomarkers and Prevention* 2007;16:458–66.

**Competing Interests**

None