

Can everyone stop using the 'F' word?

A STATISTICIANS' DIALOGUE

[Enter two statisticians, deep in conversation]

X: 'Sticking to frequentist principles of data analysis can only get you so far. You frequentists—'

Y: 'Hold on — can you please stop using the "F" word?'

X: 'The "F" word?'

Y: "'Frequentist.'"

X: 'Aren't you a frequentist?'

Y: 'No.'

X: 'So you're a Bayesian then.'

Y: 'Certainly not!'

X: 'But I thought that statisticians who weren't Bayesians were called frequentists.'

Y: 'That's like suggesting that people who aren't Jewish are called Zoroastrians. Frequentism means something very particular. The frequentist principle is that what works well in the long run is also a good rule in any unique instance. The frequentist approach to hypothesis testing, for example, is to follow a procedure that leads to few errors in the long run. Jerzy Neyman and Egon Pearson, who invented modern hypothesis testing, didn't even think it was sensible to try to evaluate uncertainty, such as the uncertainty over competing hypotheses in any unique instance of testing, and suggested that the best we could do was take comfort in the guaranteed long-run performance of their method. There's a serious problem with this logic, though: we can't be sure that the Neyman–Pearson method will have good performance in the long run — we can be fairly confident but not certain.

Now, if we're happy to endorse the method in the long run based on something less than certainty, then why can't we decide which hypothesis to endorse in a unique instance of testing based on evidence that is less than certain? Why do we even need a long-run justification for doing this? Most people are quite comfortable with judging each case on its own merits; I don't know any medical

statistician who is a true frequentist.'

X: 'We're agreed, then: I believe all medical statisticians are Bayesians, even if they don't know it.'

Y: 'Interesting theory. What's your definition of a Bayesian?'

X: 'A Bayesian is someone who understands that prior belief and new evidence from research combine to form a posterior belief.'

Y: 'But that's simply the definition of a statistician! A Bayesian is something more specific. A Bayesian holds that the degree to which we believe in anything can be given a numerical value that follows very particular mathematical laws: the laws of probability. This allows a Bayesian to prove the value of a posterior belief using something called Bayes' theorem.'

X: 'OK — I can see that's a very rigid, mathematical theory of mind, but might it not be exactly how we do think?'

Y: 'In some situations undoubtedly, but there are difficulties, such as how to represent ignorance with probabilities. Suppose I've gone into hiding somewhere in France, but you've no idea where, or even the basis for my choice of hiding place. There are 96 *départements* in Metropolitan France, and 322 *arrondissements*. Does this mean that the probability I'm in any particular *département* is 1/96? That the probability I'm in any particular *arrondissement* is 1/322? Or is every square metre of France equally probable? These things can't all be true because different *départements* contain different numbers of *arrondissements* and different areas. So there are some kinds of belief that just don't fit with this probabilistic way of looking at things.'

X: 'So if you're not a Bayesian and you're not a frequentist, what are you?'

Y: 'I prefer the term non-Bayesian.'

X: 'Isn't that just political correctness gone mad? Wait — you said that all statisticians combine new evidence with prior beliefs, but where is this apparent in frequentist, I mean non-Bayesian methods? When do

you ever take account of prior beliefs?'

Y: 'Every time we specify a statistical model it's informed, implicitly, by our prior beliefs. Every time we set up a null hypothesis we acknowledge the principle of Occam's razor: that simple explanations are more convincing a priori than complicated ones.'

X: 'All right, but where is the well-developed non-Bayesian and non-frequentist approach to statistical inference? Surely what we have currently is a collection of methods — confidence interval estimation and hypothesis testing — founded on dodgy frequentist principles?'

Y: 'You're right. Perhaps that's what we should all be working on: the foundation of our subject. Practising statisticians have shifted the principles of hypothesis testing away from their frequentist roots — most medical statisticians interpret a *P*-value as a measure of strength of evidence, for example — but this isn't as coherently developed as Bayesian statistics. What if we could improve on this? What if, say, Neyman–Pearson hypothesis testing could be reinvented from first principles that were more general than Bayes, and without the logical flaws of frequentism? If everyone confines themselves to Bayesianism we might never find out. Aren't the possibilities richer if we don't limit ourselves to just one way of thinking?'

[Exeunt, in even more animated discussion]

Richard Hooper,

Reader in Medical Statistics, Centre for Primary Care & Public Health, Queen Mary University of London, London.

E-mail: r.l.hooper@qmul.ac.uk

DOI: <https://doi.org/10.3399/bjgp18X695201>

Further reading

Austin PC, Brunner LJ, Hux JE. Bayeswatch: an overview of Bayesian statistics. *J Eval Clin Pract* 2002; **8(2)**: 277–286.

Flam FD. The odds, continually updated. *New York Times* 2014; **29 Sep**: <http://www.nytimes.com/2014/09/30/science/the-odds-continually-updated.html> (accessed 6 Feb 2018).

Freedman L. Bayesian statistical methods. *BMJ* 1996; **313(7057)**: 569.

Lilford RJ, Braunholtz D. For debate: the statistical basis of public policy: a paradigm shift is overdue. *BMJ* 1996; **313(7057)**: 603.