 doi:10.1111/add.13053

Using Bayesian analysis for hypothesis testing in addiction science

Bayesian statistical analysis provides a useful alternative to classical statistical methods for testing hypotheses. Tools are now available to apply this approach to most hypothesis tests in the field of addiction. It will often be informative to supplement classical hypothesis testing with calculation of Bayes factors, especially to distinguish 'no effect' from 'unsure about presence of an effect'.

INTRODUCTION

A revolution is sweeping through the world of statistical analysis. Although the seeds of the revolution were sown 250 years ago, it is only in the past few years that the tools have become sufficiently accessible to turn theory into reality [1]. The revolution stems from Bayes' theorem, applied to testing hypotheses about populations on the basis of sample data [2].

This editorial explains why this revolution is gathering pace and how to become part of it. A key message is that researchers could usefully supplement, and perhaps replace, classical frequentist hypothesis testing with a Bayesian approach [2].

WHAT IS BAYESIAN ANALYSIS?

There are two aspects to Bayesian analyses. One is the use of Bayes Factors to assess how far a set of data should change one's degree of belief in one hypothesis versus another. The other is how to combine this with prior degrees of belief in a given hypothesis or hypotheses. Use of 'Bayes Factors' and 'priors' are separate; it is possible to use the former without the latter. The former is largely uncontroversial, while concerns are still raised about the extent and under what circumstances the latter should be used. This editorial focuses primarily on the former.

Bayesian analysis uses the axioms of probability theory to do what most researchers already act as though they are doing when they use statistics to test hypotheses. It provides a method for using data to assess how strongly they should believe a particular hypothesis, or how strongly they should believe it relative to another hypothesis.

The Bayes Factor (BF) is a ratio showing the degree to which a set of data (even just one data point), changes the probability that a given hypothesis (H1) is true versus an alternative hypothesis (H0). If the BF is 1, then the data support (or fail to support) both hypotheses equally and are thus completely uninformative on the matter. If the BF is 3 then, given these data, H1 is three times more likely to be correct than H0. If the BF is 1/3 then H0 is three times more likely to be correct than H1.

It so happens that a BF of 3 is found typically when in classical frequentist statistics the *P*-value would be 0.05 on the same data, and so 3 and 1/3 are used commonly as thresholds for statements that there is substantial evidence for one or other hypothesis being correct [3]. For different decisions one may wish to use different thresholds.

Suppose one has undertaken a study of 300 people (150 in each group) comparing gradual and abrupt methods for smoking cessation and finds an odds ratio favouring abrupt cessation of 1.56 [95% confidence interval (CI) = 0.77–3.11, $P = 0.22$]. Thus, using the classical hypothesis testing approach, we would say there was no statistically significant difference and and research often conclude that there is 'no effect'. With a Bayesian approach we can go further. We could specify in advance two hypotheses: H0 as the hypothesis that the odds ratio in the population is 1 (i.e. there is no difference), and H1 as the hypothesis that there is a clinically useful and plausible advantage to abrupt cessation of something between 1.1 and 3.0. We can calculate a BF to tell us which of these is most likely from the data and by how much. In fact, it turns out to be 1.5, which means that H1 is 50% more likely than H0. The appropriate conclusion, then, is that the data favour the hypothesis of a clinically important difference relative to none at all, but not strongly. This is more informative than simply concluding that there is no statistically significant difference and asserting that there is 'no effect'.

OTHER REASONS FOR USING BAYES FACTORS

Apart from being more informative than the conventional approach to hypothesis testing, there are a number of other advantages to using Bayes Factors.

Use of Bayes Factors can make more efficient use of data

Using classical statistics to test hypotheses promotes data-gathering to take place in pre-specified samples of a size that should normally be determined in advance using power calculations. Once a study has been conducted and a particular result obtained, it is not generally acceptable to add further cases unless pre-planned corrections are made; and it is not acceptable to look at the data part-way through a study without increasing the threshold for statistical significance. Thus, if one has undertaken a study of, for instance, 1000 people and found a *P*-value relating to an association between two variables of 0.11, one has to report 'no significant association'. Bayesian analysis allows adding cases to the study until it is clear that there is an association, or that there is not, and it is acceptable to review the data as one goes along to assess whether there is a need to continue [4].

Use of Bayes Factors promotes clarity of hypotheses to be tested

Using classical statistics, the experimental hypothesis is typically not tested—only the null hypothesis. The experimental hypothesis is thus typically only that there is a non-zero difference or association, yet in practice one is typically interested in something more specific. In a randomized trial, for example, one typically undertakes the trial to test the hypothesis that an intervention is better to a degree that is clinically useful than some comparator such as usual care. Use of Bayes Factors forces the user to be more precise about what the experimental hypothesis is so that it can be tested.

WHAT ABOUT ACCUMULATION OF EVIDENCE?

With classical statistics, *P*-values are calculated on the basis that nothing was known previously about the hypotheses being tested. If there are three studies previously showing an effect of an intervention, the statistical analysis for the fourth study takes no account of that except in determination of the required sample size. The results may be combined in a meta-analysis, but this is performed in only a fraction of cases. Bayesian analysis can use the available information from past studies to generate a starting point or 'prior' strength of belief from which the current study can build [5]. Use of priors is contentious, because of concern that it is open to subjectivity and bias. This can be addressed in part by the use of what are known as 'objective priors' where the values used are clearly defined and arise directly from previous studies that have tested the hypothesis in question.

WHY TEST HYPOTHESES AT ALL AND WHAT HAS THIS GOT TO DO WITH *ADDICTION*?

It has been suggested that hypothesis testing should be replaced with effect size estimation, with confidence intervals [6]. However, where decisions have to be made as to whether there is sufficient evidence for an effect or association for a particular policy or clinical intervention to be warranted, hypothesis testing remains useful. This kind of question arises frequently in studies in the field of addiction. In fact, Bayesian analysis applies equally to estimating effect sizes as it does hypothesis testing.

WHAT NEXT?

Many readers will already be familiar with Bayesian analyses. Those readers are encouraged to include such analyses in their papers. Those who are unfamiliar are encouraged to read introductory texts on the subject and use tools available (e.g. www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm) to add Bayes Factors to their analyses. Researchers who do not feel able to use Bayes Factors should, at the very least, express 'negative results' in terms of lack of clear evidence for an effect rather than concluding 'no effect'.

Addiction will publish a more extensive paper in the Methods and Techniques series explaining how to undertake Bayesian analyses and the difference it makes to the conclusions of studies.

Declaration of interests

R.W. undertakes research and consultancy for companies that manufacture smoking cessation medications. He is an unpaid adviser to the UK's National Centre for Smoking cessation and Training and is a trustee of the charity, QUIT.

Acknowledgement

R.W.'s salary is funded by Cancer Research UK.

Keywords Bayesian analysis, hypothesis testing, statistical analysis.

ROBERT WEST
Department of Epidemiology and Public Health, University College London, Gower Street, London WC1E 6BT, UK
E-mail: Robert.west@ucl.ac.uk

References

1. Dienes Z. *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. London: Palgrave MacMillan; 2008.
2. Berger J. The case for objective Bayesian analysis. *Bayesian Anal* 2006; **1**: 385–402.
3. Rouder J. N., Speckman P. L., Sun D., Morey R. D., Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bull Rev* 2009; **16**: 225–37.
4. Rouder J. N. Optional stopping: no problem for Bayesians. *Psychonomic Bull Rev* 2014; **21**: 301–8.
5. Hartung J., Knapp G., Sinha B. K. Bayesian meta-analysis. *Statistical Meta-Analysis with Applications*. New York: Wiley; 2008, pp. 155–70.
6. Gardner M. J., Altman D. G. Confidence intervals rather than *P* values: estimation rather than hypothesis testing. *BMJ* 1986; **292**: 746–50.

