

A Spreadsheet for Bayesian Posterior Compatibility Intervals and Magnitude-Based Decisions

Will G Hopkins

Sportscience 23, 5-7, 2019 (sportsci.org/2019/bayes.htm)

Institute for Health and Sport, Victoria University, Melbourne, Australia. [Email](#). Reviewer: Ross D Neville, School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland.

The usual compatibility (confidence) interval for an effect in a sample can be modified to a Bayesian posterior compatibility (credibility) interval by combining the value of the effect and its interval with a prior belief in the effect expressed as its own value and interval. The spreadsheet accompanying this article provides such analyses for four kinds of effect: differences in means and other t-distributed estimates; percent or factor effects for such means derived from analyses of log-transformed dependent variables; ratios of risks, odds, hazards, and counts derived from generalized linear models; and Pearson correlation coefficients. Inclusion of a smallest important value for the effect allows the spreadsheet to provide a probabilistic magnitude-based decision about implementation of a clinically or practically relevant effect and about adequate precision for a non-clinical effect. The spreadsheet shows that realistic weakly informative priors applied to compatibility intervals from typically small samples produce posterior intervals that are practically the same as the original intervals. The minimally informative prior implicit in the magnitude-based decision method therefore provides acceptable Bayesian probabilistic estimates of the true magnitude of effects. Weakly informative priors should nevertheless be used to shrink unrealistically large compatibility limits arising from very small sample sizes and to reduce bias in effect magnitudes from generalized linear models with sparse data. Use of more-informative priors is problematic, owing to the difficulty of quantifying a belief and to bias in the belief. KEYWORDS: bias, clinical decisions, confidence, inference, probability, sample.

[Reprint pdf](#) · [Reprint docx](#) · [Spreadsheet](#) (the *Bayesian analyses* tab)

Update July 2022. The [spreadsheet](#) now has panels for estimating a prior for the effect itself (what I now call a Greenland prior) that combines with the data to give the posterior provided by a full Bayesian analysis (where every parameter in the statistical model has its own prior). The Greenland prior is estimated using the Solver add-in in Excel, which you can install via File/Options/Add-ins/Manage Excel Add-ins Go.../select Solver Add-in and click OK. The Solver is then available in Data at far right. The aim of this update is to give more legitimacy to the Greenland prior: a full Bayesian posterior always boils down to a simple comprehensible Greenland prior for the effect itself.

A Bayesian analysis of a sample combines the sample data with a prior belief about the magnitude of the effect to produce a posterior probabilistic assessment about the true value, where *true* refers to the value you would expect to obtain with a very large sample. In a full Bayesian

analysis, the prior belief applies to all the parameters in the analytic model providing the effect, including covariates used to adjust the effect and magnitude thresholds used to derive the probabilistic assessment. However, it is possible to perform a Bayesian analysis for an adjusted effect simply by specifying a prior for that effect alone and by assuming that the thresholds have no uncertainty. The prior is expressed as a point value (the most likely value, in the belief of the researcher) with a compatibility interval (formerly *confidence* interval) reflecting the researcher's uncertainty in the belief. The sample data are represented by a point estimate and its compatibility interval provided by the usual frequentist analysis with a general or generalized linear model. A Bayesian posterior credibility or compatibility interval is calculated by "information-weighting" the prior and point estimates, using the inverse of their error variances (Greenland, 2006). Probabilistic statements and decisions about the true magnitude of the effect can then be derived using the magnitude-based

decision method. The sensitivity of the probabilities to uncertainty in the smallest important magnitude of the effect can be investigated by repeating the analysis with different reasonable values of the smallest important magnitude.

A spreadsheet for performing such Bayesian analyses accompanies this article. It contains panels for analyzing four kinds of effect: differences or changes in means and other t-distributed effects; percent or factor effects for such means derived from analyses of log-transformed dependent variables; ratios of risks, odds, hazards, or counts derived from generalized linear models; and Pearson correlation coefficients.

This article and the creation of the spreadsheet were motivated partly by the need to demonstrate to researchers and journal editors that weakly informative priors make no practical difference to the compatibility interval with the usual small sample sizes, and therefore that a Bayesian interpretation of the usual compatibility interval underlying the magnitude-based decision method is justified. For this reason, the example shown in the spreadsheet for each kind of effect statistic has a weakly informative prior: a zero or null point value with 90% compatibility limits consistent with borderline extremely large values of the effect. The observed effect in each example shows a value and compatibility limits that you could get with a small sample size: approximately one-tenth of that estimated for magnitude-based decisions (and one-thirtieth that of null-hypothesis testing with the usual Type-I and Type-II error rates), using [a spreadsheet for sample-size estimation](#) (Hopkins, 2006a). You will notice that the prior causes "shrinkage" of the point estimate towards the null prior, but that the shrinkage is negligible.

Some may argue that I have opted for a prior that is unrealistically weak, deliberately to make no practical difference to the posterior and thereby to vindicate the flat-prior Bayesian interpretation of MBD. But if you allow for effects to have extremely large magnitudes, then a prior with limits on the threshold for extremely large effectively implies that you cannot have extremely large effects (where *cannot* means a 5% chance, or very unlikely, that the true effect is extremely large of either sign). Arguably, then, I should set the compatibility limits for the weak prior somewhat *larger* than the threshold for extremely large, not smaller. *Extremely large* is a standardized difference in means >4.0 , a hazard

or count ratio >10 , and a correlation >0.9 (Hopkins et al., 2009; Hopkins, 2010). Such effect magnitudes do occur from time to time.

When he promoted the approach to Bayesian analysis presented here, Greenland noted that it is more properly called *semi*-Bayesian, in that it does not introduce explicit priors for all the free parameters in the model (Greenland, 2006). He stated that "semi-Bayes analyses are equivalent to Bayesian analyses in which those parameters are given non-informative priors... Results fall short of the accuracy that could be achieved if realistic priors were used." Given the challenges of quantifying realistic priors for every parameter in a model, it is possible that a single prior for an adjusted effect could sometimes give a more accurate posterior.

But should you use any informative prior, weak or otherwise? Weakly informative priors are easily specified and can be useful to shrink unrealistically large compatibility limits arising from unavoidably *very* small sample sizes and to reduce bias in effects from generalized linear models with unavoidably sparse data (Greenland et al., 2016). If you suspect that you have a very small sample size or sparse data, check whether a weakly informative prior (extremely large \pm or \times/\div compatibility limits) results in noticeable shrinkage ($>10\%$) of either of the compatibility limits of the effect; if it does, use it.

Establishing more-informative priors is much more challenging. I have yet to see a convincing explanation or example of how to turn clinical or practical experience of an effect into numbers representing the most likely value of the effect and its uncertainty. A process of consultation and consensus with researchers or practitioners could provide a prior, but it is unreasonable to expect it to be centered on the true value. Hence a prior belief is biased, and given that the prior causes shrinkage of the effect towards itself, the resulting posterior must also be biased. Of course, the original effect is itself inevitably biased by violation of assumptions about sampling and the analytic model, but applying an informative prior to shrink the estimate will not necessarily reduce this bias. I therefore have difficulty recommending use of belief-based informative priors. A possible solution is a prior provided by a meta-analysis of studies of an effect, but such a prior would be unbiased only if it could be derived for the particular study setting of your data. Hidden effect modifiers, whose values differ

from setting to setting, guarantee that a meta-analysis cannot provide an unbiased prior for your setting. A meta-analysis is definitely worth doing, but *after* your study, not *before*.

The spreadsheet was devised by modifying the spreadsheet for combining one or more effects in [an existing workbook](#) (Hopkins, 2006b), and it is now available in that workbook (on the *Bayes* tab). The number of effects was reduced to two, labeled as *prior* and *observed* effects, and combined using as weights the inverse of the squares of standard errors of the effects, which are derived from the compatibility limits you supply for the two effects. For correlations, the sample size can be specified instead of compatibility limits, and for the weakly informative prior shown in the spreadsheet, the sample size is set to the smallest number permissible (4) with use of the Fisher z transformation underlying that analysis.

Acknowledgments: I thank Sander Greenland and Alan Batterham for helpful suggestions.

References

- Greenland S (2006). Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International Journal of Epidemiology* 35, 765-775
- Greenland S, Mansournia MA, Altman DG (2016). Sparse data bias: a problem hiding in plain sight. *BMJ* 352, i1981
- Hopkins WG (2006a). Estimating sample size for magnitude-based inferences. *Sportscience* 10, 63-70
- Hopkins WG (2006b). A spreadsheet for combining outcomes from several subject groups. *Sportscience* 10, 51-53
- Hopkins WG, Marshall SW, Batterham AM, Hanin J (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise* 41, 3-12
- Hopkins WG (2010). Linear models and effect magnitudes for research, clinical and practical applications. *Sportscience* 14, 49-58

Published June 2019

[©2019](#)