



## UvA-DARE (Digital Academic Repository)

### What Are the Odds? Modern Relevance and Bayes Factor Solutions for MacAlister's Problem From the 1881 Educational Times

Jamil, T.; Marsman, M.; Ly, A.; Morey, R.D.; Wagenmakers, E.-J.

**DOI**

[10.1177/0013164416667980](https://doi.org/10.1177/0013164416667980)

**Publication date**

2017

**Document Version**

Final published version

**Published in**

Educational and Psychological Measurement

**License**

Article 25fa Dutch Copyright Act

[Link to publication](#)

**Citation for published version (APA):**

Jamil, T., Marsman, M., Ly, A., Morey, R. D., & Wagenmakers, E.-J. (2017). What Are the Odds? Modern Relevance and Bayes Factor Solutions for MacAlister's Problem From the 1881 *Educational Times*. *Educational and Psychological Measurement*, 77(5), 819-830. <https://doi.org/10.1177/0013164416667980>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*



# What Are the Odds? Modern Relevance and Bayes Factor Solutions for MacAlister's Problem From the 1881 *Educational Times*

Educational and Psychological  
Measurement

2017, Vol. 77(5) 819–830

© The Author(s) 2016

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164416667980

journals.sagepub.com/home/epm



Tahira Jamil<sup>1</sup>, Maarten Marsman<sup>1</sup>, Alexander Ly<sup>1</sup>,  
Richard D. Morey<sup>2</sup>, and Eric-Jan Wagenmakers<sup>1</sup>

## Abstract

In 1881, Donald MacAlister posed a problem in the *Educational Times* that remains relevant today. The problem centers on the statistical evidence for the effectiveness of a treatment based on a comparison between two proportions. A brief historical sketch is followed by a discussion of two default Bayesian solutions, one based on a one-sided test between independent rates, and one on a one-sided test between dependent rates. We demonstrate the current-day relevance of MacAlister's original question with a modern-day example about the effectiveness of an educational program.

## Keywords

contingency tables, Bayes factor, evidence

In 1881, Donald MacAlister posed a famous problem in the *Educational Times*, a problem that represents one of the earliest instances concerning the comparison of two proportions in small samples:

Of 10 cases treated by Lister's method, 7 did well and 3 suffered from blood-poisoning; of 14 cases treated with ordinary dressings, 9 did well and 5 had blood-poisoning; what are the odds that the success of Lister's method was due to chance?

<sup>1</sup>University of Amsterdam, Amsterdam, The Netherlands

<sup>2</sup>Cardiff University, Cardiff, UK

## Corresponding Author:

Eric-Jan Wagenmakers, Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands.

Email: ej.wagenmakers@gmail.com

**Table 1.** MacAlister's 1881 Data Displayed as a Contingency Table.

Method	Outcome		Total
	Did well	Blood poisoning	
Lister	7	3	10
Traditional	9	5	14
Total	16	8	24

It is clear that the answer to this question is of considerable statistical relevance, far exceeding the specifics of the problem at hand. In modern-day educational research, one often wants to quantify the evidence for the effectiveness of a new program or instruction; if the new program seems to result in a beneficial outcome, the immediate question is identical that posed by MacAlister: “What are the odds that the success of the new method is due to chance”?

Before proceeding, a few remarks are in order. First and foremost, the traditional  $p$  value cannot be used to address MacAlister's question, as the  $p$  value is based on a single hypothesis (i.e., the null hypothesis) and therefore does not produce an odds. Moreover, for the MacAlister data the  $p$  value is not even close to being significant ( $p > .70$  for all standard classical methods); based on this  $p$  value, one might suspect that the evidence supports the null hypothesis. But to what degree? Second, MacAlister did not pose his question with mathematical exactness, and this requires that it has to be interpreted at least to some extent. The solutions offered in 1882 demonstrate how easy it is to misunderstand the problem (Dale, 1999: pp. 435–438; Winsor, 1948). Third, the problem as posed cannot be solved without involving the prior odds that Lister's method is effective. To appreciate the importance of the prior odds, consider the fact that Lister was a famous scientist who had advocated the use of antiseptic dressings to reduce the possibility of postsurgical infection, based on the theory that these infections were caused by germs (Lister, 1867/1967). The idea that antiseptic dressings fail to reduce the rate of postsurgical infection will strike the modern reader as absurd; consequently, the prior odds that the method's success is due to chance are extremely low. In MacAlister's example, we have the rare case that we know the answer—that Lister was correct—before we begin, so we can focus without distraction on the evidence provided by the data. These “data odds” can then be multiplied by the prior odds to obtain the posterior odds, as explained below. Fourth, the results may be presented in familiar form using a contingency table, as presented in Table 1.

The solution proposed by MacAlister was based on a procedure developed by Liebermeister (1877). Denote the probability of recovery by Lister's method and by the traditional method as  $\theta_l$  and  $\theta_t$ , respectively. It is clear that the interest partly concerns the probability of Lister's method outperforming the traditional method, that is,  $\mathcal{H}_+ : p(\theta_l > \theta_t | y)$ , where  $y$  denotes the observed data. But what should this

probability be compared to? MacAlister assumed independent uniform priors for  $\theta_l$  and  $\theta_t$  and computed  $p(\theta_l > \theta_t | y) = 0.59825$ . MacAlister compared this proportion to its complement,  $p(\theta_l < \theta_t | y)$ , and concluded, “We may wager nearly 3 to 2 that the difference in the results is not due to chance.” We may understand “due to chance” as “due to *mere* chance.” Note that MacAlister’s solution quantifies evidence in favor of the effectiveness of the treatment, despite the fact that the  $p$  value is not even close to being significant.

The hypothesis that the difference is due to mere chance, however, plays no role in MacAlister’s solution, as no prior mass is assigned to the invariance or general law that the treatments are equally effective: that is,  $\mathcal{H}_0 : \theta_l = \theta_t$  (Wrinch & Jeffreys, 1921). By failing to assign prior mass to mere chance (i.e., the hypothesis that the treatments are in fact equally effective), the question at hand cannot be answered. MacAlister’s odds of “nearly 3 to 2” address a different question, namely, “What are the odds that the success of Lister’s method is due to its superiority versus its inferiority over the traditional method?”

### Two Default Bayes Factor Solutions

To address MacAlister’s problem, we contrast two hypotheses. The first hypothesis represents the assertion that both treatments are equally effective, that is,  $\mathcal{H}_0 : \theta_l = \theta_t$ ; the second hypothesis represents the assertion that Lister’s treatment is superior to the standard treatment, that is,  $\mathcal{H}_+ : \theta_l > \theta_t$ .

We now wish to compute the evidence that the data provide for  $\mathcal{H}_+$  over  $\mathcal{H}_0$ . Recall that Bayes’ rule can be recast as follows:

$$\underbrace{\frac{p(\mathcal{H}_+ | y)}{p(\mathcal{H}_0 | y)}}_{\text{Posterior odds}} = \underbrace{\frac{p(\mathcal{H}_+)}{p(\mathcal{H}_0)}}_{\text{Prior odds}} \times \underbrace{\frac{p(y | \mathcal{H}_+)}{p(y | \mathcal{H}_0)}}_{\text{Bayes factor}}.$$

Thus, data  $y$  are used to update the prior odds to posterior odds. The assessment of prior odds is inherently subjective and depends on background information that informs one’s initial skepticism about the hypotheses under consideration. Indeed, in commenting on MacAlister’s solution to his own problem, Miss Elizabeth Blackwood stated—quite correctly, in our view,

I will merely remark that Dr. MacAlister would probably feel less satisfied as to the correctness of his result, if Lister were not the eminent man of science he is, but some superstitious old woman who, while really expert in dressing wounds, relied for protection against blood-poisoning mainly upon some mysterious charms and incantations.

MacAlister’s response made it clear that he did not consider prior odds to factor into the problem at all:

Miss Elizabeth Blackwood has perhaps not read my solution: there is no symbol in it representing Mr. Lister's science. For algebraical purposes I might substitute Mumbo Jumbo for Lister throughout, as I substituted the letter A, and no step of the reasoning on which alone the result depends would be altered.<sup>1</sup>

Here we adhere to the intention from MacAlister and focus on the Bayes factor  $BF_{+0}$ , that is, the change from prior to posterior model odds brought about by the data (Jeffreys, 1961).

The Bayes factor

$$BF_{+0} = \frac{p(y|\mathcal{H}_+)}{p(y|\mathcal{H}_0)},$$

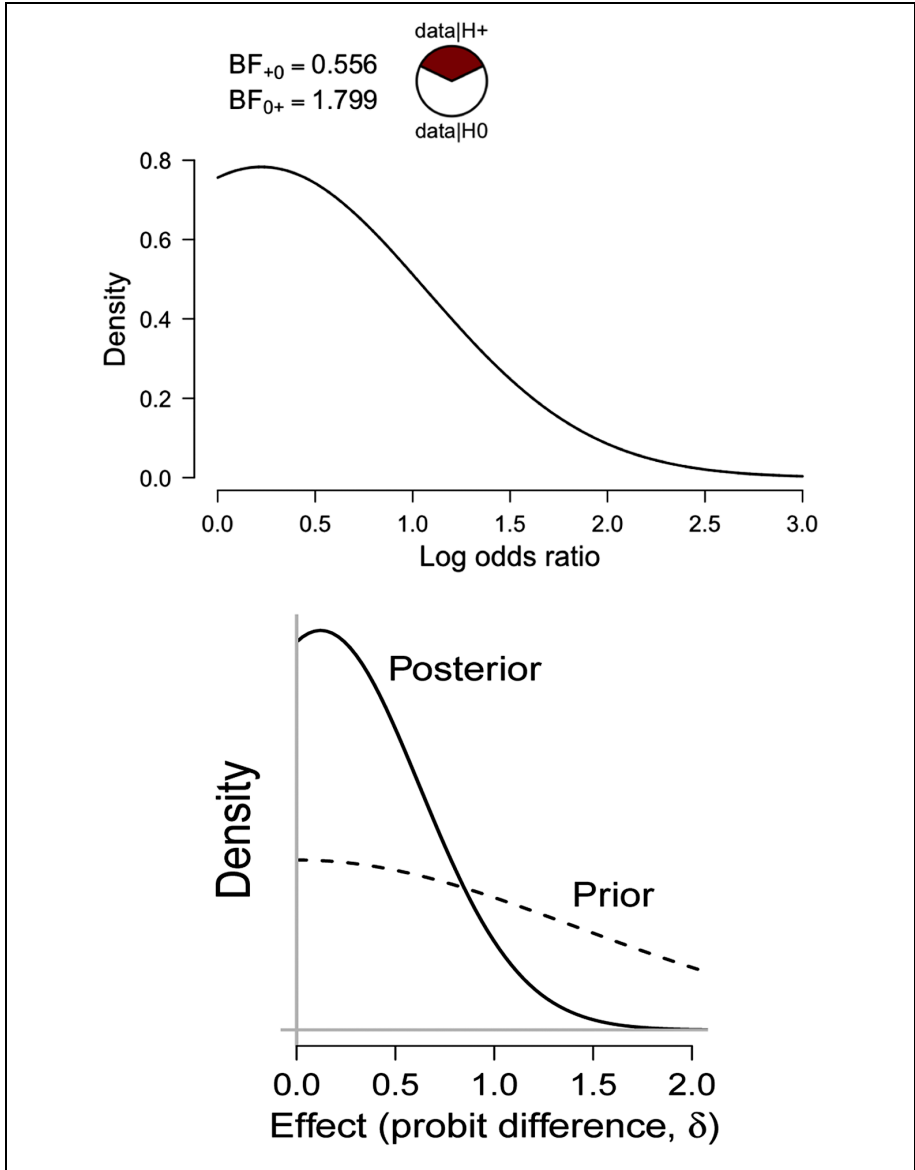
expresses the evidence in the data for the one-sided hypothesis  $\mathcal{H}_+ : \theta_l > \theta_t$ , asserting that Lister's treatment is superior to the standard treatment, against the point hypothesis  $\mathcal{H}_0 : \theta_l = \theta_t$ , asserting that both treatments are equally effective. In order to compute  $p(y|\mathcal{H}_+)$  and  $p(y|\mathcal{H}_0)$ , we need to assign prior distributions to the rate parameters  $\theta_l$  and  $\theta_t$ . This can be accomplished in many ways. Here we explore two default solutions: a model in which  $\theta_l$  and  $\theta_t$  are independent, and a model in which  $\theta_l$  and  $\theta_t$  are dependent. Both models yield a similar outcome.

### *Solution 1: Prior Independence of $\theta_l$ and $\theta_t$*

The default Bayes factor approach contrasts the single-rate model  $\mathcal{H}_0$  to the dual-rate model  $\mathcal{H}_1$ . The dual-rate model usually does not include information about the predicted direction of the effect. However, with any two-sided Bayes factor in hand a simple correction produces the desired one-sided version (see the appendix for details).

To obtain the default two-sided Bayes factor  $BF_{10}$  we assume that under the dual-rate model, each rate has an independent uniform prior distribution ranging from 0 to 1 (de Braganca Pereira & Stern, 1999; Gunel & Dickey, 1974; Jeffreys, 1935).<sup>2</sup> Based on this default prior specification, the one-sided Bayes factor can be computed easily in JASP ([jasp-stats.org](http://jasp-stats.org)), a free and open-source statistical software program with a graphical user interface familiar to users of SPSS. The same result is available for R users through the BayesFactor package (Morey & Rouder, 2015). The top panel of Figure 1 shows the JASP output.

As shown in the top panel of Figure 1, the  $BF_{0+} \approx 1.8$ , which means that the observed data are almost twice as likely under the single rate model  $\mathcal{H}_0$  than under the dual-rate model  $\mathcal{H}_+$ . The panel also features a probability wheel (i.e., a circle of area 1; Tversky, 1969) that visualizes the strength of the evidence; under equal prior odds, the white area equals the posterior probability for  $\mathcal{H}_0$  and the red area equals the posterior probability for  $\mathcal{H}_+$ . The strength of evidence can then be assessed as follows. Imagine the wheel is a dart board. You put on a blindfold and the board is attached to the wall in a random orientation. You then throw a dart and you are told



**Figure 1.** Two default one-sided Bayes factor analyses of the MacAlister data. Top panel: JASP output for the prior independent rate model, consisting of a posterior distribution for the log odds ratio and a visualization of the Bayes factor by means of a probability wheel. The corresponding .jasp file with data, analyses, and annotations is available at <https://osf.io/nvdqh/>. Bottom panel: Prior and posterior distributions for the difference parameter  $\delta$  under the prior dependent probit rate model. The Bayes factor in favor of  $\mathcal{H}_0$  is 2.3, which equals the ratio of posterior and prior ordinates at  $\delta=0$  (e.g., Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010).

it has hit the board. You remove the blindfold and observe that the dart has hit the red area instead of the white area. *How surprised are you?* This measure of imagined surprise, we suggest, conveys properly the degree of evidence that a particular Bayes factor imparts.

Consider again our Bayes factor  $BF_{0+} \approx 1.8$  for the MacAlister data. According to the classification scheme proposed by Jeffreys (1961, appendix B), this level of evidence is “not worth more than a bare mention.” Assuming that the single rate model and the dual-rate model are equally likely a priori, we can use MacAlister’s terminology and state that “we may wager nearly 2 to 1 that the difference in the results is due to mere chance.” Regardless of the inconclusive nature of the evidence in this particular instance, this result does answer MacAlister’s question.

### *Solution II: Prior Dependence of $\theta_l$ and $\theta_t$*

An alternative model specification views the two rates as dependent (e.g., Howard, 1998). Such a dependence is reasonable in many such problems; the probabilities of the two groups are typically similar. As clarified by Howard (1998)

[. . .] do English or Scots cattle have a higher proportion of cows infected with a certain virus? Suppose we were informed (before collecting any data) that the proportion of English cows infected was 0.8. With independent uniform priors we would now give  $H_1$  ( $p_1 > p_2$ ) a probability of 0.8 (because the chance that  $p_2 > 0.8$  is still 0.2). In very many cases this would not be appropriate. Often we will believe (for example) that if  $p_1$  is 80%,  $p_2$  will be near 80% as well and will be almost equally likely to be larger or smaller. (p. 363)

Thus, instead of thinking about the separate probabilities at which the two groups recover, it is convenient to instead frame the problem in terms of an overall recovery rate and the difference of the two groups from that overall rate (see also Kass & Vaidyanathan, 1992). This induces a reasonable dependency between the two groups.

The two parameters—the overall rate and the difference between the two groups—are best expressed on the probit scale, to avoid the common problem of compression of the probability scale at the extremes:

$$\Phi^{-1}(\theta_l) = \mu + \delta/2,$$

and

$$\Phi^{-1}(\theta_t) = \mu - \delta/2,$$

where  $\Phi^{-1}$  denotes the the probit transformation; that is, the inverse of the standard normal cumulative distribution function. Parameter  $\mu$  is the overall recovery rate on the probit scale, and  $\delta$  is the difference between the two groups and represents the effect of interest. Next,  $\mu$  and  $\delta$  are assigned normal priors. For demonstration, we assign a Normal(0, .707) prior distribution to  $\mu$  and  $\delta$  is assigned a folded (i.e., positive-only, to incorporate knowledge about the hypothesized direction of the effect)

**Table 2.** Number of Students Retained as a Function of Having Attended a Learning Strategies Course.

Group	Retained		Total
	Yes	No	
Course takers	328	23	351
Non-course takers	300	51	351
Total	628	74	702

Note. Data reported in Tuckman and Kennedy (2011).

normal prior with mean 0 and standard deviation  $\sigma$ . The test will not be very sensitive to the prior choice on  $\mu$ ; however, a reasonable prior on  $\delta$  is important, as it is the parameter of interest. For demonstration we choose  $\sigma = \sqrt{2}$  as a default value. We choose these settings because they yield the same marginal priors on  $\theta_1$  and  $\theta_2$  as under Solution I.

The Bayes factor of interest is based on a comparison between two models,  $\mathcal{H}_0 : \delta = 0$  versus  $\mathcal{H}_+ : \delta > 0$ . To obtain  $BF_{+0}$ , we use Gaussian quadrature.<sup>3</sup> Analyzing the MacAlister data, the bottom panel of Figure 1 shows the prior and posterior distributions for the difference  $\delta$  under  $\mathcal{H}_+$ . At the value of interest,  $\delta = 0$ , the posterior distribution is about 2.3 times as high as the prior distribution, and hence,  $BF_{0+} \approx 2.3$  (Dickey & Lientz, 1970; Wagenmakers et al., 2010). As for the default analysis using the independent priors, the Bayes factor indicates that the data are more likely to occur under  $\mathcal{H}_0$  than under  $\mathcal{H}_+$ , but the strength of this evidence is not impressive.

### A Modern Example From Education Research

To underscore the relevance of MacAlister’s problem for current-day research we turn to a study by Tuckman and Kennedy (2011) published in the *Journal of Experimental Education*. These authors investigated the effect of a learning strategies course on students’ academic performance as quantified by several dependent variables including retention rate, that is, the proportion of students that return to college the following year. The data showed that from a total of  $n_1 = 351$  first-year students who took the course, 93.4% returned to college the next year; from a total of  $n_2 = 351$  matched students who did not take the course, 85.5% returned. Table 2 shows the data in the form of a contingency table. For these data, MacAlister’s question is again relevant: *What are the odds that the success of the learning strategies course was due to mere chance?*

We address this question as we did before, by contrasting two hypotheses. The null hypothesis states that the course has no effect,  $\mathcal{H}_0 : \theta_1 = \theta_2$ . The alternative hypothesis has direction and states that the course increases the retention rate,  $\mathcal{H}_+ : \theta_1 > \theta_2$ .



As before, the change from prior to posterior odds for  $\mathcal{H}_0$  versus  $\mathcal{H}_+$  is expressed through the Bayes factor  $BF_{+0}$ .

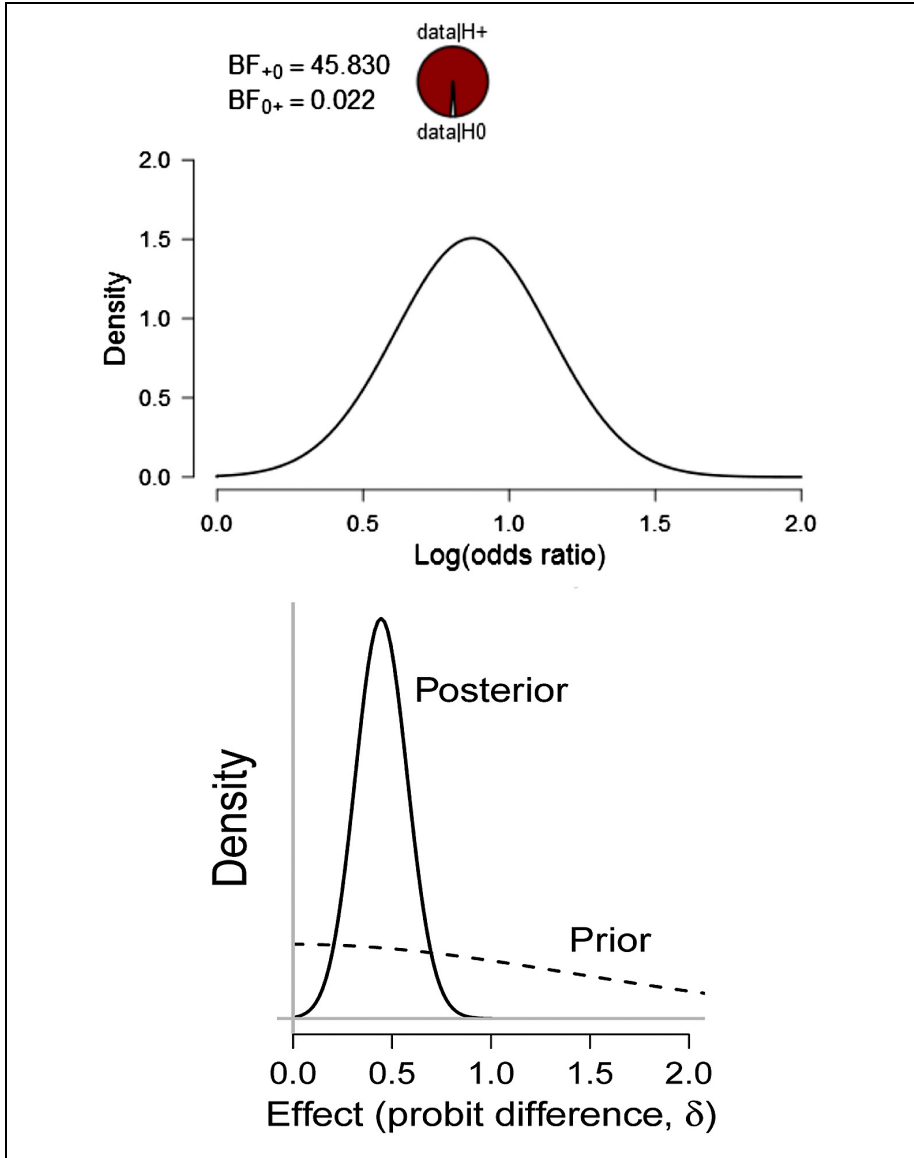
First, the results from the independent prior analysis (Gunel & Dickey, 1974; Jeffreys, 1935) are displayed in the top panel of Figure 2. The output shows that  $BF_{+0} = 45.83$ , meaning that the observed data are 45.83 times more likely to occur under  $\mathcal{H}_+$  than under  $\mathcal{H}_0$ . According to Jeffreys' classification scheme, this constitutes "very strong" evidence in favor of the effectiveness of the course on retention rate.

Second, the results from the dependent prior analysis with the probit model are displayed in the bottom panel of Figure 2. As before, the distributions are for the difference  $\delta$  under  $\mathcal{H}_+$ . At the value of interest,  $\delta = 0$ , the prior distribution is about .0142 times as high as the posterior distribution, and hence,  $BF_{+0} \approx 70$ . Even though the two methods give slightly different results, they agree that the data provide considerable support in favor of  $\mathcal{H}_+$ .

## Conclusion

We have outlined a Bayesian method to quantify the support that the data provide for the equality or inequality of two rates. This Bayesian method allows one to address the key problem posed by MacAlister in 1881: What are the odds that the success of a particular treatment is based on mere chance? In our solution to MacAlister's problem, we compared a single rate model  $\mathcal{H}_0$  against an order-restricted default dual-rate model  $\mathcal{H}_+$ , using two fundamentally different prior specifications, one dependent and one independent. As usual, it should be acknowledged that the default prior distributions can often be enriched and adjusted by incorporating substantive knowledge about the problem at hand. Moreover, in applied settings, one might extend the current framework and use model-averaging to obtain superior predictions (e.g., Hoeting, Madigan, Raftery, & Volinsky, 1999); in addition, one might specify utilities and combine these with the fundamental unknowns in order to make the best possible decision in a coherent manner (e.g., Lindley, 1985, 2006). Both prediction and decision making require the consideration of the prior odds for the competing hypotheses, an endeavor that is often inherently subjective.

Despite these reservations, we believe that in many situations the default prior specifications provide an appropriate reference analysis. For a range of standard statistical models, such reference analyses can be easily conducted using the R BayesFactor package (Morey & Rouder, 2015) or the free and open-source program JASP ([jasp-stats.org](http://jasp-stats.org)). Prominent advantages of the default Bayesian analysis include the possibility to monitor evidence as the data accumulate and the ability to discriminate evidence of absence from the absence of evidence. The problem posed by MacAlister in 1881 is still relevant today, and Bayesian methods such as the one outlined in this article constitute a solution that is theoretically elegant and practically relevant.



**Figure 2.** Two default one-sided Bayes factor analyses of the data from Tuckman and Kennedy (2011). Top panel: JASP output for the prior independent rate model, consisting of a posterior distribution for the log odds ratio and a visualization of the Bayes factor by means of a probability wheel. The corresponding .jasp file with data, analyses, and annotations is available at <https://osf.io/nvdqh/>. Bottom panel: Prior and posterior distributions for the difference parameter  $\delta$  under the prior dependent probit rate model. The Bayes factor in favor of  $\mathcal{H}_+$  is approximately 70, which equals the ratio of posterior and prior ordinates at  $\delta = 0$  (e.g., Wagenmakers et al., 2010).

## Appendix

### Obtaining the One-Sided Bayes Factor

The Bayes factor  $BF_{+0}$  can be easily obtained by decomposing it into two parts (Morey & Wagenmakers, 2014; Pericchi, Liu, & Torres, 2008):

$$\begin{aligned} BF_{+0} &= \frac{p(y|\mathcal{H}_+)}{p(y|\mathcal{H}_0)} \\ &= \frac{p(y|\mathcal{H}_+)}{p(y|\mathcal{H}_1)} \times \frac{p(y|\mathcal{H}_1)}{p(y|\mathcal{H}_0)} \\ &= BF_{+1} \times BF_{10}, \end{aligned} \tag{A1}$$

where  $BF_{+1}$  quantifies the evidence for the hypothesis  $\mathcal{H}_+$  that Lister's method is superior against the undirected hypothesis  $\mathcal{H}_1 : \theta_l \neq \theta_t$  that simply asserts that the two treatments have a different effect.

Equation (A1) demonstrates that to compute the one-sided Bayes factor  $BF_{+0}$ , we multiply the two-sided  $BF_{10}$  by the factor  $BF_{+1}$ . Klugkist, Laudy, and Hoijtink (2005) noted that  $BF_{+1}$  equals the ratio of posterior and prior mass under  $\mathcal{H}_1$  that is consistent with the restriction postulated by  $\mathcal{H}_+$ . That is,

$$BF_{+1} = \frac{p(\theta_l > \theta_t | y, \mathcal{H}_1)}{p(\theta_l > \theta_t | \mathcal{H}_1)}, \tag{A2}$$

which simplifies to  $BF_{+1} = 2 \times p(\theta_l > \theta_t | y, \mathcal{H}_1)$  whenever the prior distributions do not express any knowledge or preference for the ordering of  $\theta_l$  and  $\theta_t$ , meaning that  $p(\theta_l > \theta_t) = 1/2$ .

### Authors' Note

Supplemental materials and annotated JASP files are available on the Open Science Framework at <https://osf.io/nvdqh/>.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by an ERC grant (10.13039/501100000781 283876) from the European Research Council.

### Notes

1. In a later rejoinder, Miss Blackwood retorted as follows:

As to the right which Dr. MacAlister claims to substitute, if he chooses, Mumbo Jumbo for Lister in his solution, I am inclined to think he has already exercised that right. But, granting this mathematical license of substitution “for algebraical purposes” (a euphemism apparently for juggling purposes) is Lister to be arbitrarily valued at  $\frac{1}{2}$  merely because we don’t know what other value to assign to him? Poor Lister!

2. Another popular default prior distribution for rate parameters is the Beta(1/2, 1/2) prior, which is also known as Jeffreys’s prior (e.g., Zhu & Lu, 2004). However, Jeffreys proposed this prior specifically for estimation problems, whereas for testing problems Jeffreys consistently used the uniform Beta(1, 1) prior. Another option—less popular, but worthy of more attention—is to use nonlocal priors (Johnson & Rossell, 2010).
3. There are a number of other ways to obtain the Bayes factor, including importance sampling and Markov chain Monte Carlo sampling. An interactive application to compute and visualize the Bayes factor can be found at [richarddmorey.shinyapps.io/probitProportions](http://richarddmorey.shinyapps.io/probitProportions), and R code to compute the Bayes factor and plots can be downloaded at [gist.github.com/richarddmorey/4c7a408a45c3045ab949](https://gist.github.com/richarddmorey/4c7a408a45c3045ab949).

## References

- Dale, A. I. (1999). *A history of inverse probability: From Thomas Bayes to Karl Pearson* (2nd ed.). New York, NY: Springer.
- de Braganca Pereira, C. A., & Stern, J. M. (1999). Evidence and credibility: Full Bayesian significance test for precise hypotheses. *Entropy*, *1*, 99-110.
- Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *Annals of Mathematical Statistics*, *41*, 214-226.
- Gunel, E., & Dickey, J. (1974). Bayes factors for independence in contingency tables. *Biometrika*, *61*, 545-557.
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*, 382-417.
- Howard, J. V. (1998). The  $2 \times 2$  table: A discussion from a Bayesian viewpoint. *Statistical Science*, *13*, 351-367.
- Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society*, *31*, 203-222.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, England: Oxford University Press.
- Johnson, V. E., & Rossell, D. (2010). On the use of non-local prior densities in Bayesian hypothesis tests. *Journal of the Royal Statistical Society, Series B*, *72*, 143-170.
- Kass, R. E., & Vaidyanathan, S. K. (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. *Journal of the Royal Statistical Society, Series B*, *54*, 129-144.
- Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, *10*, 477-493.
- Liebermeister, C. (1877). Ueber Wahrscheinlichkeitsrechnung in Anwendung auf therapeutische Statistik [On probability theory as applied to therapeutic statistics]. In R. Volkmann (Ed.), *Sammlung Klinischer Vorträge No. 110 (Innere Medizin No. 39)* (pp. 935-962). Leipzig, Germany: Breitkopf & Härtel.
- Lindley, D. V. (1985). *Making decisions* (2nd ed.). London, England: Wiley.
- Lindley, D. V. (2006). *Understanding uncertainty*. Hoboken, NJ: Wiley.

- Lister, J. (1967). Antiseptic principle in the practice of surgery. *British Medical Journal*, 2, 9-12. (Original work published 1867)
- Morey, R. D., & Rouder, J. N. (2015). *Bayes Factor 0.9.11-1*. Comprehensive R Archive Network.
- Morey, R. D., & Wagenmakers, E.-J. (2014). Simple relation between Bayesian order-restricted and point-null hypothesis tests. *Statistics and Probability Letters*, 92, 121-124.
- Pericchi, L. R., Liu, G., & Torres, D. (2008). Objective Bayes factors for informative hypotheses: “Completing” the informative hypothesis and “splitting” the Bayes factor. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 131-154). New York, NY: Springer Verlag.
- Tuckman, B. W., & Kennedy, G. J. (2011). Teaching learning strategies to increase success of first-term college students. *Journal of Experimental Education*, 79, 478-504.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31-48.
- Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, 60, 158-189.
- Winsor, C. P. (1948). Probability and Listerism. *Human Biology*, 20, 161-169.
- Wrinch, D., & Jeffreys, H. (1921). On certain fundamental principles of scientific inquiry. *Philosophical Magazine*, 42, 369-390.
- Zhu, M., & Lu, A. Y. (2004). The counter-intuitive non-informative prior for the Bernoulli family. *Journal of Statistics Education*, 12, 1-10.