Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2006;**35**:765–775

Advance Access publication 30 January 2006
doi:10.1093/ije/dyi312

Bayesian perspectives for epidemiological research: I. Foundations and basic methods

Sander Greenland

Accepted 15 December 2005

One misconception (of many) about Bayesian analyses is that prior distributions introduce assumptions that are more questionable than assumptions made by frequentist methods; yet the assumptions in priors can be more reasonable than the assumptions implicit in standard frequentist models. Another misconception is that Bayesian methods are computationally difficult and require special software. But perfectly adequate Bayesian analyses can be carried out with common software for frequentist analysis. Under a wide range of priors, the accuracy of these approximations is just as good as the frequentist accuracy of the software—and more than adequate for the inaccurate observational studies found in health and social sciences. An easy way to do Bayesian analyses is via inverse-variance (information) weighted averaging of the prior with the frequentist estimate. A more general method expresses the prior distributions in the form of prior data or 'data equivalents', which are then entered in the analysis as a new data stratum. That form reveals the strength of the prior judgements being introduced and may lead to tempering of those judgements. It is argued that a criterion for scientific acceptability of a prior distribution is that it be expressible as prior data, so that the strength of prior assumptions can be gauged by how much data they represent.

Keywords Bayesian methods, biostatistics, odds ratio, relative risk, risk assessment

Introduction

Beginning with Laplace in the 18th century, Bayesian statistical methods were used freely alongside other methods until the 1920s. At that time, several influential statisticians (R. A. Fisher, J. Neyman, and E. Pearson) developed bodies of frequentist techniques intended to supplant entirely all others, based on notions of objective probability represented by relative frequencies in hypothetical infinite sequences of randomized experiments or random samplings. For over half a century these methods dominated advanced statistical research and became the sole body of methods taught to most students.

In the randomized trial and random-sample survey in which they were developed, these frequentist techniques appear to be highly effective tools. As the use of the methods spread from designed surveys and experiments to observational studies, however, an increasing number of statisticians questioned the objectivity and realism of hypothetical infinite sequences.[1–7] They argued that a subjective-Bayesian approach better represented situations in which the mechanisms generating study samples and exposure status were heavily non-random and poorly understood. In those settings, which typify most of epidemiology, the personal judgements of the investigators play an unavoidable and crucial role in making inferences and often over-ride technical considerations that dominate statistical analyses (as perhaps they should[8]).

In the wake of such arguments, Bayesian methods have become common in advanced training and research in statistics,[9–12] even in the randomized-trial literature for which frequentist methods were developed.[13,14] Elementary training appears to have lagged, however. The present paper argues that the Bayesian perspective should be included in elementary statistical training and illustrates how common frequentist methods can be used to generate Bayesian analyses. In particular, it shows how basic analyses can be conducted with a hand calculator or ordinary software packages. Thus, as far as computation is concerned, it is a small matter to extend current training to encompass Bayesian methods.

The first, philosophical section criticizes standard objections to Bayesian approaches, and delineates key parallels and differences between frequentist and Bayesian methods. It does not address distinctions within frequentist and Bayesian traditions. By 'frequentist methods' I will refer to the incoherent blend of Fisherian *P*-values and Neyman–Pearson hypothesis tests and confidence intervals found in research articles (see Goodman[15] for an accessible review of the profound divergence

Departments of Epidemiology and Statistics, University of California, Los Angeles, CA 90095-1772, USA. E-mail: lesdomes@ucla.edu

between Fisherian and Neyman–Pearsonian frequentism). I argue that observational researchers (not just statisticians) need training in subjective Bayesianism[1,2,16] to serve as a counterweight to the alleged objectivity of frequentist methods. For this purpose, neither so-called 'objective' Bayesian methods[17] nor 'pure likelihood' methods[18] will do, because they largely replicate the pretence of objectivity that render frequentist methods so misleading in observational settings.

An irony of the Bayesian renaissance is that some Bayesian statisticians have become a major obstacle to the dissemination of Bayesian approaches. They have done this by agonizing over representations of ignorance (rather than information) and by insisting on precision in prior specification and analytical computation far beyond anything required of frequentist methods or by the messy problems of observational data analysis. Their preferred methods—Markov-Chain Monte Carlo (MCMC)—require special software, have their own special problems, obscure important parallels between frequentist and Bayesian methods, and are unnecessary for the imprecise data and goals of everyday epidemiology (which is largely only semi-quantitative inference about an adjusted risk comparison). Subjective Bayesian methods are distinguished by their use of informative prior distributions; hence, their proper use requires a sound understanding of the meaning and limitations of those distributions, rather than a false sense of precision. In observational studies neither Bayesian nor other methods require precise computation, especially in light of the huge uncertainties about the processes generating observational data (represented by the likelihood function), as well as uncertainty about prior information.

After the philosophical section, the paper focuses on basic Bayesian methods that display prior distributions as prior estimates or prior data and employ the same approximate formulas used by frequentist methods.[5,19–24] Subsequent papers will illustrate extensions of these methods to non-normal priors, regression, and analysis of biases.[25] Even for those who prefer MCMC methods, the representation of prior distributions as prior data is helpful in understanding the strength of the prior judgements, and hence one should be able to supply this representation if the contextual acceptability of the prior is questioned.

Frequentism vs subjective Bayesianism

There are several objections that frequentists have raised against Bayesian methods. Some of these are legitimate but apply in parallel to frequentist methods (and indeed to all of statistics) in observational studies; they include the fact that the assumptions or models employed are at best subjective judgements. Others are propaganda; e.g. that adopting a Bayesian approach introduces arbitrariness that is not already present. In reality, the Bayesian approach makes explicit subjective and arbitrary elements shared by all statistical inferences. Because these elements are hidden by frequentist conventions, Bayesian methods are left open to criticisms, which make it appear that only they are using those elements.

Subjective probabilities should not be arbitrary

In subjective (personalist) Bayesian theory, my prior for a parameter is a probability distribution *P*(parameters) that shows how I would bet about parameters if I disregarded the data under analysis. This prior need not originate from evidence temporally prior to the study; rather, it represents information apart from the data being analysed. When the only parameter is a risk ratio (RR), the 50th percentile (median) of my prior, *P*(RR), is a number $\mathrm{RR}_{\mathrm{median}}$ for which I would give even odds that $\mathrm{RR} < \mathrm{RR}_{\mathrm{median}}$ vs $\mathrm{RR} > \mathrm{RR}_{\mathrm{median}}$, i.e. I would assign $P(\mathrm{RR} < \mathrm{RR}_{\mathrm{median}}) = P(\mathrm{RR} > \mathrm{RR}_{\mathrm{median}})$ if I disregarded the analysis data. Similarly, my 95% prior limits are a pair of numbers $\mathrm{RR}_{\mathrm{lower}}$ and $\mathrm{RR}_{\mathrm{upper}}$ such that I would give $95:5 = 19:1$ odds that the true risk ratio is between these numbers, i.e. $P(\mathrm{RR}_{\mathrm{lower}} < \mathrm{RR} < \mathrm{RR}_{\mathrm{upper}}) = 0.95$ if I disregarded the analysis data.

Prior limits may vary considerably across individuals; yours can be very different from mine. It does not, however, mean that the limits are arbitrary. When betting on a race with the goal of minimizing my losses, I would not regard it reasonable to bet everything on a randomly drawn contestant; rather, I would place different bets on different contestants, based on their previous performance (but taking account of differences in the past conditions from the present). Similarly, if I wanted a Bayesian analysis to seem reasonable or credible to others, my prior should reflect results from previous studies or reviews. This reflection should allow for possible biases and lack of generalizability among studies, so that my prior limits might be farther apart than frequentist meta-analytical confidence limits (even if the latter incorporated random effects).

The prior *P*(parameters) is one of two major inputs to a Bayesian analysis. The other input is a function *P*(data | parameters) that shows the probability the analyst would assign the observed data for any given set of parameter values (usually called the likelihood function). In subjective-Bayesian analysis this function is another set of bets: The model for *P*(data | parameters) summarizes how one would bet on the study outcome (the data) if one knew the parameters (e.g. the exposure-covariate specific risks). Any such model should meet the same credibility requirements as the prior. This requirement parallels the frequentist concern that the model should be able to approximate reality; in fact, any competent Bayesian has the same concern, albeit perhaps with more explicit doubts about whether that can be achieved with standard models.

The same need for credibility motivates authors to discuss other literature when writing their research reports. Credible authors pay attention to past literature in their analyses, e.g. by adjusting for known or suspected confounders, by not adjusting for factors affected by exposure, and by using a dose–response model that can capture previously observed patterns (e.g. the J-shaped relation of alcohol use to cardiovascular mortality). They may even vary their models to accommodate different views on what adjustments should be done. In a similar manner, Bayesian analyses need not be limited to using a single prior or likelihood function. Acceptability of an analysis is often enhanced by presenting results from different priors to reflect different opinions about the parameter, by presenting results using a prior that is broad enough to assign relatively high probability to each discussant's opinion (a 'consensus' prior), and by presenting results from different degrees of regression adjustment (which involves varying the likelihood function).

The posterior distribution

Upon seeing the outcome of a race on which I had bet, I would want to update my bets regarding the outcome of another race involving the same contestants. In this spirit, Bayesian analysis produces a model for the posterior distribution $P(\text{parameters} \mid \text{data})$, a probability distribution that shows how one should bet about the parameters after examining the analysis data.

As a minimal criterion of reasonable betting, suppose you would never want to place your bets in a manner that allows an opponent betting against you to guarantee your loss. This criterion implies that your bets should obey the laws of probability, including Bayes Theorem,

$$P(\text{parameters}|\text{data}) = P(\text{data}|\text{parameters})P(\text{parameters})/P(\text{data}),$$

where the portion $P(\text{data})$ is computed from the likelihood function and the prior [for an accessible review of these arguments see Ref. (26)]. The 50th percentile (median) of my posterior about a risk ratio RR is a number $\text{RR}_{\text{median}}$ for which $P(\text{RR} < \text{RR}_{\text{median}}|\text{data}) = P(\text{RR} > \text{RR}_{\text{median}}|\text{data})$, where '|data' indicates that this bet is formulated in light of the analysis data. Similarly, my 95% posterior limits are a pair of numbers RR_{lower} and RR_{upper} such that after analysing the data I would give 95 : 5 = 19 : 1 odds that the true relative risk is between these numbers, i.e. $P(\text{RR}_{\text{lower}} < \text{RR} < \text{RR}_{\text{upper}}|\text{data}) = 0.95$.

As with priors, posterior distributions may vary considerably across individuals, not only because they may use different priors $P(\text{parameters})$ but also because they may use different models for the data probabilities $P(\text{data} \mid \text{parameters})$. This variation is only to be expected given disagreement among observers about the implications of past study results and the present study's design. Bayesian analyses can help pinpoint sources of disagreement, especially in that they distinguish sources in the priors from sources in the data models.

Frequentist-Bayesian parallels

It is often said (incorrectly) that 'parameters are treated as fixed by the frequentist but as random by the Bayesian'. For frequentists and Bayesians alike, the value of a parameter may have been fixed from the start or may have been generated from a physically random mechanism. In either case, both suppose it has taken on some fixed value that we would like to know. The Bayesian uses formal probability models to express personal uncertainty about that value. The 'randomness' in these models represents personal uncertainty about the parameter's value; it is not a property of the parameter (although we should hope it accurately reflects properties of the mechanisms that produced the parameter).

A crucial parallel between frequentist and Bayesian methods is their dependence on the model chosen for the data probability $P(\text{data} \mid \text{parameters})$. Statistical results are as sensitive to this choice as they are to choice of priors. Yet, this choice is almost always a default built into statistical software, based on assumptions of random sampling or random treatment assignment (which are rarely credible in observational epidemiology), plus additivity assumptions. Worse, the data models are often selected by mechanical algorithms oblivious to background information, and as a result often conflict with contextual information. These problems afflict the majority of epidemiological analyses today in the form of models (such as the logistic, Poisson, and proportional-hazards models) that make interaction and dose–response assumptions, which are rarely if ever justified. These models are never known to be correct and in fact cannot hold exactly, especially when one considers possible study biases.[27,28]

Acceptance of results derived from these models (whether the results are frequentist or Bayesian) thus requires the doubtful assumption that existing violations have no important impact on results. The model for $P(\text{data} \mid \text{parameters})$ is thus a weak link in the chain of reasoning leading from data to inference, shared by both frequentist and Bayesian methods. In practice, the two approaches often use the same model for $P(\text{data} \mid \text{parameters})$, whence divergent outputs from the methods must arise elsewhere. A major source of divergence is the explicit prior $P(\text{parameters})$ used in Bayesian reasoning. The methods described below will show the mechanics of this divergence and provide a sense of when it will be important.

Empirical priors

The addition of the prior $P(\text{parameter})$ raises the point that the validity of the Bayesian answer will depend on the validity of the prior model as well as the validity of the data model. If the prior should not just be some arbitrary opinion, however, what should it be?

One answer arises from frequentist shrinkage-estimation methods (also known as Stein estimation, empirical-Bayes, penalized estimation, and random-coefficient or ridge regression) to improve repeated-sampling accuracy of estimates. These methods use numerical devices that translate directly into priors[4,5,29] and so leave unanswered the same question asked of subjective Bayesians: where should these devices come from? Empirical-Bayes and random-coefficient methods explicitly assume that the parameters as well as the data would vary randomly across repetitions, according to an actual frequency distribution $P(\text{parameters})$ that can be estimated from available data. As in Bayesian analyses, these methods compute posterior coefficient distributions using Bayes theorem. Given the randomness of the coefficients, however, the resulting posterior intervals are also frequentist confidence intervals, in the sense of containing the true (if varying) parameter values in the stated percentage of repetitions.[10]

Those who wish to extend Bayes-frequentist parallels into practice are thus led to the following empirical principle: When 'chances' or true frequency distributions exist and are known for the data or the parameter distribution (as in multilevel random sampling[30]), they should be used as the distributions in Bayesian analysis. This principle reflects the idea of placing odds on race contestants based on their past frequencies of winning, and corresponds to common notions of induction.[31] Such frequency-based priors are more accurately termed 'empirical' rather than 'subjective', although the decision to accept the empirical evidence remains a subjective judgement (and subject to error). Empirical priors are mandated in much of Bayesian philosophy, e.g. in the 'Principal Principle' of Lewis, 1981: When frequency probabilities exist and are known (as in games of chance and in quantum physics), use them as your personal probabilities. More generally, an often-obeyed (if implicit) inductive principle is that the prior should be found by fitting to available empirical frequencies (in the same manner as a frequentist would estimate hyperparameters in a hierarchical model by marginal or Type-II maximum-likelihood[5,20,32]). The fitted prior is thus no more arbitrary than (and may even be functionally identical to) a fitted second-stage frequentist model. With empirical priors, the resulting frequentist and Bayesian interval estimates may be numerically identical.

Frequentist-Bayesian divergences

Even when a frequentist and Bayesian arrive at the same interval estimate for a parameter, the interpretations remain quite different. Frequentist methods pretend that the models are laws of chance in the real world (indeed, much of the theoretical literature encourages this illusion by calling distributions 'laws'). In contrast, subjective-Bayesian methods interpret the models as nothing more than summaries of tentative personal bets about how the data and the parameters would appear, rather than as models of a real random mechanism. The prior model should be based on observed frequencies when those are available, but the resulting model for the posterior $P(\text{parameters}\,|\,\text{model})$ is a summary of personal bets after seeing the data, not a frequency distribution (although if the parameters are physically random it would also represent a personal estimate of their distribution).

It is important to recognize that the subjective-Bayesian interpretation is much less ambitious (and less confident) than the frequentist interpretation, insofar as it treats the models and the analysis results as systems of personal judgements, possibly poor ones, rather than as some sort of objective reality. Probabilities are nothing more than expressions of opinions, as in common phrasings like 'it will probably rain tomorrow'. Reasonable opinions are based heavily on frequencies in past experience, but they are never as precise as results from statistical computations.

Frequentist fantasy vs observational reality

For Bayesian methods, there seems no dispute that the results should be presented with reference to the priors as well as to the data models and the data. For example, a posterior interval should be presented as '*Given these priors, models, and data*, we would be 95% certain that the parameter is in this interval'.

A parallel directive should be applied to frequentist presentations. For example, 95% confidence intervals are usually presented as if they account for random error, without regard for what that random error is supposed to represent. For observational research, one of the many problems with frequentist (repeated-sampling) interpretations is that it is not clear what is 'random' when no random sampling or randomization has been done. Although 'random variation' may be present even when it has not been introduced by the investigator, in observational studies there is no basis for claiming it follows the distributions that frequentist methods assume, or any known distribution.[27] At best, those distributions refer only to thought experiments in which one asks, '*if* data were repeatedly produced by the *assumed* random-sampling process, the statistics would have their stated properties (e.g. 95% coverage) across those repetitions'. They do not refer to what happens under the distributions actually operating, for the latter are unknown. Thus, what they do say is extremely hypothetical, so much so that to understand them fully is to doubt their relevance for observational research.[4]

Frequentist results are hypothetical whenever one cannot be certain that the assumed data model holds, as when uncontrolled sources of bias (such as confounding, selection bias, and measurement error) are present. In light of such problems, claims that frequentist methods are 'objective' in an observational setting seem like propaganda or self-delusion.[4–6,26,28] At best, frequentist methods in epidemiology represent a dubious social convention that mandates treating observational data as if they arose from a dream or fantasy of a tightly designed and controlled randomized experiment on a random sample (that is, as if a thought experiment were reality). Like many entrenched conventions, they provoke defences that claim utility[12,33] without any comparative empirical evidence that the conventions serve observational research better than would alternatives. Other defences treat the frequentist thought experiments as if they were real—an example of what has been called the mind-projection fallacy.[34]

Were we to apply the same truth-in-packaging standard to frequentists as to Bayesians, a 'significant' frequentist result would be riddled with caveats like '*If* these data had been generated from a randomized trial with no drop-out or measurement error, these results would be very improbable were the null true; but because they were not so generated we can say little of their actual statistical significance'. Such brutal honesty is of course rare in observational epidemiology because emphasizing frequentist premises undermines the force of the presentation.

Summary

A criticism of Bayesian methods is that the priors must be arbitrary or subjective in a pernicious or special way. In observational studies, however, the prior need be no more arbitrary than the largely arbitrary data models routinely slapped on data, and prior models can often be given a scientific foundation as or more firm than that of frequentist data models. Like any analysis element, prior models should be scrutinized critically (and rejected as warranted), just as should frequentist models. When relevant and valid external frequency data are available, they should be used to build the prior model (which may lead to inclusion of those data as part of the likelihood function, so that the external and current data become pooled). When prior frequency data are absent or invalid, however, other sources of priors will enter, and must be judged critically.

Below I will show how normal log relative-risk priors can be translated into 'informationally equivalent' prior frequency data. Elsewhere I show how this translation extends to non-normal priors and to other parameters.[23,25] I will argue that this translation should become a standard method for evaluating whether a prior is contextually reasonable, even if the translation is not used to compute posteriors.

Simple approximate Bayesian methods

Exact Bayesian analysis proceeds by computing the posterior distribution via Bayes theorem, which requires $P(\text{data})$. The latter can be difficult to evaluate (usually requiring multiple integration over the parameters), which seems to have fostered the misimpression that practical Bayesian analyses are inherently more complex computationally than frequentist analyses. But this impression is based on an unfair comparison of exact Bayesian methods to approximate frequentist methods.

Frequentist teaching evolved during an era of limited computing and so focused on simple, large-sample approximate methods for categorical data. In contrast, the Bayesian resurgence occurred during the introduction of powerful personal computers and advanced Monte Carlo algorithms, hence much Bayesian teaching focuses on exact methods, often presented as if simple approximations are inadequate. But Bayesian approximations suitable for categorical data have a long history[19,20] and are as accurate as frequentist approximations—which are accurate enough for epidemiological studies. The approximations also provide insights into the meaning of both Bayesian and frequentist methods, and hence are the focus of the remainder of this paper.

In the examples below the outcome is very rare, so we may ignore distinctions among risk, rate, and odds ratios, which I will refer to generically as 'relative risks' (RR). Because a normal distribution has equal mode, median, and mean, we may also ignore distinctions among these measures of location when discussing a normal ln (RR). When we take the antilog $e^{\ln(\mathrm{RR})} = \mathrm{RR}$, however, we obtain a lognormal distribution, for which mode < median = geometric mean < arithmetic mean. Only the median transforms directly: median RR = $e^{\text{median } \ln(\mathrm{RR})}$.

Information-weighted averaging

Information (or precision) is defined here as the inverse of the variance.[9] Weighting by information shows how simple Bayesian methods parallel frequentist summary estimation based on inverse-variance weighting.[32,35] It assumes that both the prior model and the data model are adequately approximated by normal distributions. This comes down to assuming that the sample sizes (both actual and prior) are large enough for the approximation to be adequate. There is no hard-and-fast rule on what size is adequate, in part because of disagreement about how much inaccuracy is tolerable (which depends on context), but the same approximations in frequentist categorical statistics are arguably adequate down to cell sizes of 4 or 5.[35,36]

A single two-way table

Table 1 shows case–control data from Savitz *et al*. (1988), the first widely publicized study to report a positive association between residential magnetic fields and childhood leukaemia. Although previous studies had reported positive associations between household wiring and leukaemia, at the time strong field effects seemed unlikely and very strong effects seemed very unlikely. To start, suppose we model these a priori ideas by placing 2 : 1 odds on a relative risk (RR) between ½ and 2, and 95% probability on RR between ¼ and 4. These bets would follow from a normal prior for the log relative risk ln (RR) that satisfies

$$\exp(\text{prior mean} - 1.96 \cdot \text{prior standard deviation}) = \tfrac{1}{4},$$
$$\exp(\text{prior mean} + 1.96 \cdot \text{prior standard deviation}) = 4.$$

Solving this pair of equations, we get

$$\begin{aligned}\text{prior mean of } \ln(\mathrm{RR}) &= \text{average of the limits}\\ &= \left[\ln(\tfrac{1}{4}) + \ln(4)\right] \Big/ 2 = 0,\end{aligned}$$

$$\begin{aligned}&\text{prior standard deviation of } \ln(\mathrm{RR})\\ &\quad = (\text{width of interval in } \ln(\mathrm{RR}) \text{ units})/\\ &\qquad (\text{width of interval in standard deviation units})\\ &\quad = (\ln(4) - \ln(\tfrac{1}{4}))/(2 \cdot 1.96) = 0.707,\end{aligned}$$

$$\text{prior variance of } \ln(\mathrm{RR}) = 0.707^2 = 0.500 = \tfrac{1}{2}.$$

Thus, the normal prior distribution that would produce the stated bets has mean zero and variance ½.

Three of 36 cases and 5 of 198 controls had measured average fields >3 milliGauss (mG). These data yield the frequentist RR estimates:

$$\begin{aligned}&\text{estimated RR} = \text{sample odds ratio} = 3(193)/5(33) = 3.51,\\ &\text{estimated variance of log odds ratio}\\ &\quad = 1/3 + 1/33 + 1/5 + 1/193 = 0.569,\\ &95\% \text{ confidence limits} = \exp\left[\ln(3.51) \pm 1.96 \cdot 0.569^{\ 1/2}\right]\\ &\qquad = 0.80, 15.4.\end{aligned}$$

Assuming there is no prior information about the prevalence of exposure, an approximate posterior mean for ln (RR) is just the average of the prior mean ln(RR) of 0 and the data estimate ln(3.51), using the information weights of 1/(½) and 1/.569, respectively:

$$\begin{aligned}&\text{posterior mean for } \ln(\mathrm{RR}) = \text{expected } \ln(\mathrm{RR}) \text{ given data}\\ &\quad \approx \frac{\left[(0/(\tfrac{1}{2})) + (\ln(3.51)/0.569)\right]}{\left[(1/(\tfrac{1}{2})) + (1/0.569)\right]} = 0.587.\end{aligned}$$

The approximate posterior variance of ln(RR) is just the inverse of the total information:

$$\text{posterior variance for } \ln(\mathrm{RR}) \approx \frac{1}{\left[(1/(\tfrac{1}{2})) + (1/0.569)\right]} = 0.266.$$

Together this mean and variance produce

$$\begin{aligned}&\text{posterior median for RR} \approx \exp(0.587) = 1.80,\\ &95\% \text{ posterior limits for RR}\\ &\quad \approx \exp(0.587 \pm 1.96 \cdot 0.266^{1/2}) = 0.65, 4.94.\end{aligned}$$

The posterior RR of 1.80 is close to a simple geometric averaging of the prior RR (of 1) with the frequentist estimate (of 3.51), because the data information is 1/0.569 = 1.76 whereas the prior information is 1/(½) = 2, giving almost equal weight to the two. This reflects the fact that both the study (with only three exposed cases and five exposed controls) and the prior are weak. Note too that the posterior RR of 1.80 is much

Table 1 Case–control data on residential magnetic fields and childhood leukaemia (Savitz *et al*., 1988) and frequentist results

	***X* = 1**	***X* = 0**	
Cases	3	33	Table odds ratio = RR estimate = 3.51
Controls	5	193	95% Confidence limits = 0.80–15.4

ln(OR) = ln(RR) estimate = ln(3.51), estimated variance = 0.569.

closer to the frequentist odds ratios from other studies,[28] which average 1.7, even though the prior here is centred on the null.

Bayesian interpretation of frequentist results

The weighted-averaging formula shows that the frequentist results are what one gets from the Bayesian calculation when the prior information is made negligibly small relative to the data information. In this sense, frequentist results are just extreme Bayesian results, ones in which the prior information is zero, asserting that absolutely nothing is known about the RR outside of the study. Some promote such priors as 'letting the data speak for themselves'. In reality, the data say nothing by themselves: The frequentist results are computed using probability models that assume complete absence of bias and so filter the data through false assumptions.

A Bayesian analysis that uses these frequentist data models is subject to the same criticism. Even with no bias, however, assuming absence of prior information is empirically absurd. Prior information of zero implies that a relative risk of (say) 10^{100} is as plausible as a value of 1 or 2. Suppose the relative risk was truly 10^{100}; then every child exposed >3 mG would have contracted leukaemia, making exposure a sufficient cause. The resulting epidemic would have come to everyone's attention long before the above study was done because the leukaemia rate would have reached the prevalence of high exposure, or ~5/100 annually in the US, as opposed to the actual value of 4 per 100 000 annually; the same could be said of any relative risk >100. Thus there are ample background data to rule out such extreme relative risks.

So-called 'objective-Bayes' methods differ from frequentist methods only in that they make these unrealistic 'noninformative' priors explicit, and so produce posterior intervals that represent the inference of no one (except perhaps someone who understood nothing of the subject under study or even the meaning of the variable names). Genuine prior bets are more precise. Even exceptionally 'strong' relations in non-infectious-disease epidemiology (such as smoking and lung cancer) involve RR on the order of 10 or 1/10, and few non-infectious study relations are even that far from the null. This reflects the fact that, for a relation to be discovered by formal epidemiological study, its effects must be small enough to have gone undetected by clinical practice or by surveillance systems. There is almost always some surveillance (if only informal, through the health-care system) that implies limits on the effect size. If these limits are huge, frequentist results serve as a rough approximation to a Bayesian analysis that uses an empirically based prior for the RR; otherwise the frequentist results may be very misleading.

Adjustment

To adjust for measured confounders (without using explicit priors for their confounding effects), one need only set a prior for the adjusted RR and then combine the prior ln(RR) with the adjusted frequentist estimate by inverse-variance averaging. For example, a pooled analysis of 14 studies of magnetic fields (>3 mG vs less) and childhood leukaemia [Table 1 in Ref. (28)] produced a summary maximum-likelihood common odds-ratio estimate of 1.69 and 95% confidence limits of 1.28, 2.23; thus the log odds-ratio is ln (1.69) = 0.525 with variance estimate $(\ln\ (2.23/1.28)/3.92)^2 = 0.0201$. Combining this frequentist result with a normal (0,½) prior yields

$$\text{posterior mean for ln(RR)}$$
$$\approx \frac{\left[(0/(\tfrac{1}{2})) + (\ln(1.69)/0.0201)\right]}{\left[(1/(\tfrac{1}{2})) + (1/0.0201)\right]} = 0.504,$$

$$\text{posterior variance for ln(RR)} \approx \frac{1}{\left[(1/(\tfrac{1}{2}))+(1/0.0201)\right]} = 0.0193,$$

$$\text{posterior median for RR} \approx \exp(0.504) = 1.66,$$

$$95\%\ \text{posterior limits for RR} \approx \exp(0.504 \pm 1.96 \cdot 0.0193^{1/2})$$
$$= 1.26, 2.17,$$

This posterior barely differs from the frequentist results, reflecting that the data information is 1/0.0201 = 50, or 25 times the prior information of 1/(½) = 2. In other words, the data information dominates the prior information.

One can also make adjustments based on priors for confounding, which may include effects of unmeasured variables.[28,35–37]

Varying the prior

Many authors have expressed extreme scepticism over the existence of an actual magnetic-field effect, so much so that they have misinterpreted positive findings as null because they were not 'statistically significant' (e.g. UKCCS, 1999). The Bayesian framework allows this sort of prejudice to be displayed explicitly in the prior, rather than forcing it into misinterpretation of the data.[38] Let us suppose that the extreme scepticism about the effect[39] is expressed as a normal prior for ln(RR) with mean zero and 95% prior limits for RR of 0.91 and 1.1. The prior standard deviation is then [ln(1.1) – ln (0.91)]/3.92 = 0.0484. Averaging this prior with the frequentist summary of ln(1.69) yields 95% posterior RR limits of 0.97, 1.16. Here, the prior weight is $1/0.0484^2 = 427$, over eight times the data information of 50, and so the prior dominates the final result.

It can be instructive to examine how the results change as the prior changes.[14,28] Using a normal (0, v) prior, a simple approach examines the outputs as the variance v ranges over values that different researchers hold. For example, when examining a relative risk RR, prior variances of ⅛, ½, 2, 4 for ln(RR) correspond to 95% prior intervals for RR of (½,2), (¼,4), (1/16,16), (1/50,50). The frequentist results represent one (gullible) extreme based on two false assumptions: First, that the likelihood (data) model is correct (which is falsified by biases); and second, that nothing is known about any explicit parameter, corresponding to infinite v and hence no prior upper limit on RR (which is falsified by surveillance data). At the other extreme, assertions of sceptics often correspond to priors with $v < \tfrac{1}{8}$ and hence a 95% prior interval within (½, 2).

Bayes vs semi-Bayes

The above example analyses are semi-Bayes in that they do not introduce an explicit prior for all the free parameters in the problem. For example, they do not use a prior for the population exposure prevalence $P(X = 1)$ or for the relation of adjustment factors to exposure or the outcome. Semi-Bayes analyses are equivalent to Bayesian analyses in which those parameters are given non-informative priors, and correspond to frequentist mixed models (in which some but not all coefficients are

Table 2 General notation for 2 × 2 prior-data layout

	$X = 1$	$X = 0$	
Cases	A_1	A_0	Table RR = $\text{RR}_{\text{prior}} = (A_1/N_1)/(A_0/N_0)$ $= (A_1/A_0)/(N_1/N_0)$
Total	N_1	N_0	

random). As with frequentist analyses, the cost of using no prior for a parameter is that the results fall short of the accuracy that could be achieved if a realistic prior were used. The benefit is largely one of simplicity in not having to specify priors for many parameters. Good[5] provides a general discussion of cost-benefit tradeoffs of analysis complexity, under the heading of Type-II rationality. Good[5] and Greenland[40] also describe how multilevel (hierarchical) modelling subsumes frequentist, semi-Bayes, and Bayes methods, as well as shrinkage (empirical-Bayes) methods.

Prior data: frequentist interpretation of priors

Having expressed one's prior bets as intervals about the target parameter, it is valuable to ask what sort of data would have generated those bets as confidence intervals. In the above examples we could ask: What would constitute data 'equivalent' to the prior? That is, what experiment would convey the same information as the normal (0, ½) prior for ln(RR)? Answers to such Bayesian questions can be found by frequentist thought experiments[38 (Appendix 2)], which show how Bayesian methods parallel frequentist methods for pooled analysis of multiple studies.

Suppose we were given the results of a trial with N_1 children randomized to exposure ($X = 1$) and N_0 to no exposure (a trial infeasible and unethical in reality but, as yet, allowed in the mind), as in Table 2. With equal allocation, $N_1 = N_0 = N$, and the frequentist RR estimate would then equal the ratio of the number of treated cases A_1 to the number of untreated cases A_0:

$$\text{estimated RR} = (A_1/N)/(A_0/N) = A_1/A_0.$$

Given the rarity of leukaemia, N would be very large relative to A_1 and A_0. Hence $1/N \approx 0$, and

$$\text{estimated variance for ln(RR)}$$
$$= (1/A_1) + (1/A_0) - (1/N) - (1/N) \approx (1/A_1) + (1/A_0).$$

[see Ref. (41), Ch. 14]. To yield our prior for RR, these estimates would have to satisfy

$$\text{estimated RR} = A_1/A_0 = 1,$$

so $A_1 = A_0 = A$, and

$$\text{estimated variance of ln(RR) estimate}$$
$$\approx 1/A_1 + 1/A_0 = 1/A + 1/A = 2/A = {}^1\!/_2,$$

so $A_1 = A_0 = A = 4$. Thus, data roughly equivalent to our normal (0,½) prior would comprise four cases in each of the treated and the untreated groups in a very large randomized trial with equal allocation, yielding a prior estimate RR_{prior} of 1 and a ln(RR) variance of ½. The value of N would not matter provided it was large enough so that $1/N$ was negligible relative to $1/A$. Table 3 shows an example.

Table 3 Example of Bayesian analysis via frequentist methods: data approximating a lognormal prior, reflecting 2:1 certainty that RR is between ½ and 2, and 95% certainty that RR is between ¼ and 4, and result of combination with data from Table 1

	$X = 1$	$X = 0$	
Cases	4	4	Table RR = $\text{RR}_{\text{prior}} = 1$
Total	100 000	100 000	Approximate 95% prior limits = 0.25–4.00

$\ln(\text{RR}_{\text{prior}}) = 0$, approximate variance = ¼ + ¼ = ½.

Approximate posterior median and 95% limits from stratified analyses combining prior with Table 1.

From information (inverse-variance) weighting of RR estimates: 1.80, 95% limits 0.65–4.94.

From maximum-likelihood (ML) estimation: 1.76, 95% limits 0.59–5.23.

Expressing the prior as equivalent data leads to a general method for doing Bayesian and semi-Bayes analyses with frequentist software:

(i) Construct data equivalent to the prior, then
(ii) Add those prior data to the actual study data as a distinct (prior) stratum.

The resulting point estimate and C% confidence limits from the frequentist analysis of the augmented (actual + prior) data provide an approximate posterior median and C% posterior interval for the parameter.

In the example this method leads to a frequentist analysis of two strata: one stratum for the actual study data (Table 1) and one stratum for the prior-equivalent data (Table 3). Using information weighting (which assumes both the prior and the likelihood are approximately normal), these strata produce a point estimate of 1.80 and 95% limits of 0.65, 4.94, as above. A better approximation is supplied by using maximum likelihood (ML) to combine the strata, which yields here a point estimate of 1.76 and 95% limits of 0.59, 5.23; this approximation assumes only that the posterior distribution is approximately normal.

With other stratification factors in the analysis, the prior remains just an extra stratum, as above. For example, in the pooled analysis there were 14 strata, one for each study [Table 1 in Ref. (28)]. Adding the prior data used above with $A = 4$ and $N = 100\,000$ as if it were a 15th study, and applying ML, the approximate posterior median RR and 95% limits are 1.66 and 1.26, 2.17, the same as that from information weighting.

After translating the prior to equivalent data, one might see the size of the hypothetical study and decide that the original prior was overconfident, implying a prior trial larger than seemed justified. For a childhood-leukaemia incidence of $4/10^5$ years, 8 cases would require 200 000 child-years of follow-up, which is quite a bit larger than any real trial of childhood leukaemia. If one were not prepared to defend the amount of one's prior information as being this ample, one should make the trial smaller. In other settings one might decide that the prior trial should be larger.

Reverse-Bayes analysis

Several authors describe how to apply Bayes' theorem in reverse (inverse-Bayes analysis) by starting with hypothetical posterior results and asking what sort of prior would have led to those results, given the actual data and data models used.[5,42] One hypothetical posterior result of interest has the null as one of the 95% limits. In the above pooled analysis, this posterior leads to the question: how many prior cases per group (A) would be needed to make the lower end of the 95% posterior interval equal to 1?

Repeating the ordinary Bayes analysis with different A and N until the lower posterior limit equals 1, we find $A = 275$ prior leukaemia cases per group (550 total) forces the lower end of the 95% posterior interval to 1.00. This number is over twice the number of exposed cases seen in all epidemiological studies to date. At a rate of about 4 cases/10^5 person-years, a randomized trial capable of producing 2A = 550 leukaemia cases under the null would require roughly $550/(4/10^5) > 13$ million child-years of follow-up. The corresponding prior variance is $2/275 = 0.00727$, for a 95% prior interval of

$$\exp(0 \pm 1.96 \cdot 0.00727^{1/2}) = 0.85, 1.18.$$

While this is an extremely sceptical prior, it is not as sceptical as many of the opinions written about the relation.[41] Upon seeing this calculation, we might fairly ask of sceptics, 'do you actually have evidence for the null that is equivalent to such an impossibly large, perfect randomized trial?' Without such evidence, the calculation shows that any reasonable posterior scepticism about the association must arise from methodological shortcomings of the studies (which correspond to shortcomings of standard frequentist data models[28]).

Priors with non-null centre

Suppose we shift our prior estimate RR_{prior} for RR to 2, with 95% prior limits of ½ and 8. This corresponds to $\ln(RR_{prior}) = \ln(2)$ with a prior variance of ½. Combining this prior with the Savitz data by information weighting yields

posterior variance

$$\approx \frac{1}{\left[(1/(\tfrac{1}{2})) + (1/0.569)\right]} = 0.266 \text{ (as before)},$$

posterior ln(RR) median

$$\approx \frac{\left[(\ln(2)/(\tfrac{1}{2})) + (\ln(3.51)/0.569)\right]}{\left[(1/(\tfrac{1}{2})) + (1/0.569)\right]} = 0.956,$$

posterior RR median $\approx \exp(0.956) = 2.60$,

95% posterior RR limits

$$\approx \exp(0.956 \pm 1.96 \cdot 0.266^{1/2}) = 0.95, 7.15.$$

One can accomplish the same by augmenting the observed dataset with a stratum of prior data. To preserve approximate normality, we keep $A_1 = A_0$ (so $A_1/A_0 = 1$) and adjust the denominator quotient N_1/N_0 to obtain the desired $RR_{prior} = (A_1/A_0)/(N_1/N_0) = 1/(N_1/N_0) = N_0/N_1$. In the above example this means keeping $A_1 = A_0 = 4$, $N_1 = 100\,000$ but making $N_0 = 200\,000$, so that

$$RR_{prior} = (4/100\,000)/(4/200\,000) = 200\,000/100\,000 = 2,$$

while the approximate prior variance of ln(RR) remains ¼ + ¼ = ½. Thus, data equivalent to our upshifted prior would be the observation of four cases in each of the treated and the untreated groups in a randomized trial with a 1 : 2 allocation to $X = 1$ and $X = 0$.

Choosing the sizes of the prior denominators

The absolute size of N_1 and N_0 used will matter little, provided both $N_1 > 100 \cdot A_1$ and $N_0 > 100 \cdot A_0$. Thus, if we enlarge A_1 and A_0 we simply enlarge N_1 and N_0 proportionally to maintain disease rarity in the prior data. Although it may seem paradoxical, this rarity is simply a numerical device that can be used even with common diseases. This is because standard frequentist RR estimators do not combine baseline risks across strata. By placing the prior data in a separate stratum, the baseline risk in the prior data may take on any small value without affecting either the baseline risk estimates for the actual data or the posterior RR estimates. N_1 and N_0 are only used to move the prior estimate RR_{prior} to the desired value: When they are very large they cease to influence the prior variance and only their quotient N_1/N_0 matters.

For the thought experiment used to set N_1 and N_0, one envisions an experimental group that responds to treatment (X) with the relative risk one expects, but in which the baseline risk is so low that the distinctions among odds, risk, and rate ratios become unimportant. The estimator we apply to the total (augmented) data will determine what we are estimating: an odds ratio estimator will produce an odds-ratio estimate, a risk-ratio estimator will produce a risk-ratio estimate, and a rate-ratio estimator will produce a rate-ratio estimate. For rate-ratio analyses, N_1 and N_0 represent person-time rather than persons.

Non-normal priors

The addition of prior data shown above (with very large N_1, N_0) corresponds to using an F distribution with $2A_1$, $2A_0$ degrees of freedom as the RR prior.[19,25] With $A_1 = A_0 = A$, the above lognormal approximation to this prior appears adequate down to about $A = 4$, e.g. at $A = 4$, the ~95% RR interval of (¼,4) has 93.3% exact prior probability from an $F(8,8)$ distribution; at $A = 3$ the ~95% interval is (1/5,5) and has 92.8% exact probability from an $F(6,6)$. These are minor discrepancies compared with other sources of error, and the resulting discrepancies for the posterior percentiles are smaller still: as with the accuracy of ML, the accuracy of the posterior approximation depends on the total information across strata (prior + data). Nonetheless, if we want to introduce prior data that represent even less information or that represent non-normal ln(RR) priors, we can employ prior data with $A_1 \neq A_0$ to induce ln(RR)-skewness, and with A_1, $A_0 < 3$ to induce heavier tails than the normal. Generalizations beyond the F distribution are also available.[23,25]

Further extensions

Prior-data methods extend easily to multivariable modelling and to settings where some or all variables (including the outcome) have multiple levels. For example, one may add a prior stratum for each regression coefficient in a model; coefficients for continuous variables can be represented as trials comparing two levels of the variable (e.g. 800 vs 0 mcg/day folic-acid supplementation); and prior correlations can be induced using a hierarchical prior-data structure.[25,37]

Checking the prior

A standard recommendation is to check homogeneity of measures before summarizing them across strata. An analogous recommendation is to check the compatibility of the data and the prior,[43] which can be subsumed under the more general topic of Bayesian model checking.[11,13,44] For normal priors, one simple approximate check examines the *P*-value from the 'standardized' difference,

$$(\text{frequentist estimate} - \text{prior estimate})$$
$$/(\text{frequentist variance} + \text{prior variance})^{½},$$

which is the analogue of the frequentist two-stratum homogeneity statistic.[41] Like frequentist homogeneity tests, this check is neither sensitive nor specific, and assumes the counts are 'large' (>4). A small *P*-value does, however, indicate that the prior and the frequentist results are too incompatible to average by information weighting (in that chance alone would rarely produce a discrepancy as or more extreme).

For the pooled magnetic-field data with a normal (0,½) prior ($A_1 = A_0 = 4 << N_1 = N_0$), the check is $[\ln(1.69) - 0]/(0.0201 + ½)^{½} = 0.72$, $P = 0.47$. Thus, by this check the prior and the frequentist result appear compatible, largely because the prior is compatible with such a broad range of results. Despite this compatibility, their average may still be misleading (e.g. due to study biases). In contrast, with the sceptical normal $(0,0.0484^2)$ prior ($A_1 = A_0 = 427 << N_1 = N_0$), the check is $[\ln(1.69) - 0]/(0.0201 + 0.0484^2)^{½} = 3.50$, $P = 0.0005$, which indicates extreme incompatibility of the prior and the frequentist result and suggests that the average would be misleading because at least one is misleading (and perhaps both are).

The *P*-value for the exposure–disease (*X*–*Y*) association in the actual data equals the homogeneity *P*-value from comparing the frequentist result to a dogmatic normal (0,0) prior concentrated entirely at the null with zero variance (equivalent to overwhelming prior data, e.g. $A = 10^{100}$ and $N = 10^{100\,000}$). For the pooled-data example this *P*-value is 0.0002, corresponding to the usual interpretation that a small *P*-value is indicative of a conflict between the null and the data. The *P*-value comparing the prior mean to the frequentist estimate can be viewed as a generalization of the usual *P*-value to allow testing of 'fuzzy' (non-dogmatic) hypotheses, in which the true parameter is only specified up to a distribution rather than asserted to exactly equal a number.

Discussion

Data alone say nothing at all

It is sometimes recommended that the prior should be given up if it appears in conflict with the 'actual data'.[45] The conflict, however, is between the prior and the frequentist result from the data model; without a model for the data-generating mechanism, the data alone can conflict with nothing.[46]

If we truly believed the frequentist results were from perfect data from a randomized trial conducted on a random sample from the target population, we would no doubt afford them precedence over a prior composed from mere impressions of other evidence. Indeed, a key inferential property of randomized studies is that, if precise enough, they force agreement among those who believe that the randomization was carried out properly and not undermined by subsequent events. Observational studies stray from this ideal, however, and do so to an unknown extent. A conflict between a prior and a frequentist result can arise from an invalid data model (due to study biases or an incorrect analytical method), as opposed to an incorrect prior; thus, an apparent conflict only calls attention to a discrepancy in need of explanation. This situation is starkly illustrated by the magnetic-field controversy, in which many scientists still postulate that the frequentist results (rather than their sceptical prior) must be in error.

While (as often said) frequentist statistics much better reflect the data than do Bayesian statistics, those data should not be regarded as sacrosanct when making inferences beyond the observations. Even rough but contextually well-informed priors can provide information as or more reliable than current observations. To put it negatively: One may be justifiably afraid of the unreliability of subjective priors, but this fear does not license exclusive reliance on unreliable data. Frequentist statistics and their 'objective' Bayesian analogues indeed stay closer to the observations, but this closeness will harm inference when these observations are riddled with error and better external information is available. Conversely, for the goal of data description (as opposed to inference), neither Bayesian nor frequentist modelling results are an adequate substitute for tabular and graphical data summaries.

Data priors as a general diagnostic device

The data representation of priors is far more important and general than realized by most of the statistics community today. For teaching, data priors provide both a Bayesian interpretation of frequentist statistics (as Bayesian statistics with no prior data) and a frequentist interpretation of Bayesian statistics (as frequentist statistics based in part on external data). For analysis, data priors provide a critical perspective on a proposed prior and can lead to refinements to better match the level of prior information one wishes to assume. Other prior representations are also conceptually helpful; for example, the penalized-likelihood approach illustrates how Bayesian statistics can be viewed as frequentist statistics with 'fuzzy' constraints on parameters.[22,47] These interpretations show that entering and deleting variables in a regression are extremes along a continuum in which the variable may enter partially to the extent allowed by the constraint.[4,47,48]

I thus argue that translation between forms should become a routine contextual (as opposed to statistical) diagnostic for priors. In doing so, I emphasize that data representations are not limited to conjugate priors (priors that have the same functional form as the likelihood function). Any prior that can be viewed as a product of likelihood functions can be translated into data, namely the data arising from the statistical studies represented by the functions. These functions may be of varying forms, and those forms may differ from that of the actual-data likelihood (i.e. they may be non-conjugate). The translation clarifies the evidential claims that the prior is making. Conversely, to say a given prior could not be viewed as augmenting data means that one could not envision a series of studies that would lead to the prior (let alone point to actual studies that produce the prior). I would assert that such a prior is non-empirical in principle and hence scientifically meaningless (in the same sense that a theory empirically untestable in principle is scientifically meaningless).

The role of Markov-Chain Monte Carlo

Translation of the prior into various forms does not dictate the manner of posterior computation. One could for example translate for diagnostic purposes and then sample from the posterior distribution using an MCMC program such as WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml). Nonetheless, MCMC is subject to technical problems that are not always easily detected; indeed, the current WinBUGS website carries the warning 'MCMC is inherently less robust than analytical statistical methods'. MCMC also does not display the parallels between frequentist and Bayesian analyses. Hence I find it inadequate alone or as a starting point for Bayesian instruction, and in practice recommend analytical methods for starting and checking MCMC analyses.

Furthermore, I have yet to see MCMC make a scientifically meaningful difference in everyday epidemiological problems, even with small or sparse datasets.[22,23,47,49] This observation is unsurprising, given that data augmentation uses approximations that work well in most frequentist analyses and that approximation errors are typically far below the size of random errors and biases. The exceptions occur in studies so small or sparse as to have almost no information. Thus, upon considering all the poorly understood sources of bias that plague observational research, stringent levels of accuracy cannot be justified. For these reasons, I reject the primacy of MCMC over other methods, even though I find it a useful tool.

Connections to sensitivity analysis

There is a close and complementary connection between Bayesian methods and sensitivity analyses, in which parameters fixed at defaults by frequentist methods are varied to see their impact on statistics. Simple sensitivity analyses[50,51] will reveal unlimited sensitivities to certain variations and, hence, convey no information unless coupled with contextual information to determine what variations are meaningful.[52] That is none other than prior information, which can be formalized in a prior distribution for use in Bayesian or analogous prior-sampling methods for risk and decision analysis.[28,37,51] In this format, one can also conduct Bayesian sensitivity analysis by seeing how results vary as the prior is varied; in this process, data representations can help one judge the priors that are credible enough to examine.

Conclusions

Bayesian and related methods become worthwhile exactly when frequentist statistics become questionable and the priors matter, as when confronting sparse data, multiple comparisons, collinearity, or non-identification.[23,37,47,49,53,54] The simple examples above, nonetheless, illustrate the differences in interpretation between frequentist and Bayesian statistics. Frequentist interval estimates inevitably get interpreted as if they were Bayesian, without appreciating that the priors implicit in these interpretations are rarely if ever contextually plausible. Because this appreciation seems essential for proper interpretation of frequentist as well as Bayesian methods in observational settings,[4,23,26,28,37,40,41,47,55] the inclusion of Bayesian perspectives in teaching would be helpful even if frequentist results remained the norm for presentation.

In summary, Bayesian analysis can be performed easily by information weighting of prior estimates with frequentist estimates (as if doing a meta-analysis of prior studies and the current study). More generally, it can be done by representing prior information as hypothetical study data to be added to the analysis as new strata (as if doing a pooled analysis of prior studies and the current study); this approach provides a diagnostic for contextual strength and relevance of the prior. Both approaches allow one to produce Bayesian analyses from formulas and software for frequentist analyses, and both facilitate introduction of Bayesian ideas into introductory statistics training—where they should take their place alongside frequentist approaches.

Acknowledgements

The author is grateful to Stephen Cole, Michael Höfler, Katherine Hoggatt, Robert Weiss, and the referees for helpful comments.

References

[1] Lindley D. *Introduction to Probability and Statistics from a Bayesian Viewpoint*. Cambridge: Cambridge University Press, 1965.

[2] DeFinetti B. *The Theory of Probability, Vol. I*. New York: Wiley, 1974.

[3] Cornfield J. Recent methodological contributions to clinical trials. *Am J Epidemiol* 1976;**104:**408–24.

[4] Leamer E. *Specification Searches*. New York: Wiley, 1978.

[5] Good I. *Good Thinking*. Minneapolis: University of Minnesota Press, 1983.

[6] Berger JO, Berry DA. Statistical analysis and the illusion of objectivity. *Am Sci* 1998;**76:**159–65.

[7] Berk RA, Western B, Weiss RE. Statistical inference for apparent populations. *Sociol Methodol* 1995;**25:**421–58.

[8] Susser M. Judgment and causal inference. *Am J Epidemiol* 1977; **105:**1–15.

[9] Leonard T, Hsu JSJ. *Bayesian Methods*. Cambridge: Cambridge University Press, 1999.

[10] Carlin B, Louis TA. *Bayes and Empirical-Bayes Methods of Data Analysis*. 2nd edn. New York: Chapman and Hall, 2000.

[11] Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. 2nd edn. New York: Chapman and Hall/CRC, 2003.

[12] Efron B. Bayesians, frequentists, and scientists. *J Am Statist Assoc* 2005;**100:**1–5.

[13] Spielgelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Assessment*. London: John Wiley and Sons, 2004.

[14] Spielgelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials (with discussion). *J Royal Statist Soc* 1994; **156:**357–416.

[15] Goodman S. *P*-values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. *Am J Epidemiol* 1993;**137:**485–96.

[16] Goldstein M. Subjective Bayesian analysis: principles and practice. In: Proceedings of the workshop on Bayesian Analysis. 2006 (to appear) (preprint at http://www.stat.cmu.edu/bayesworkshop/2005/panel.html).

[17] Berger JO. The case for objective Bayesian analysis. In: Proceedings of the workshop on Bayesian Analysis. 2006 (to appear) (preprint at http://www.stat.cmu.edu/bayesworkshop/2005/panel.html).

[18] Royall R. *Statistical Inference: A Likelihood Paradigm*. New York: Chapman and Hall, 1997.

[19] Lindley D. The Bayesian analysis of contingency tables. *Annals of Mathematical Statistics* 1964;**35:**1622–43.

[20] Good I. *The Estimation of Probabilities*. Boston: MIT Press, 1965.

[21] Bedrick EJ, Christensen R, Johnson W. A new perspective on generalized linear models. *J Am Statist Assoc* 1996;**91:**1450–60.

[22] Greenland S. Putting background information about relative risks into conjugate priors. *Biometrics* 2001;**57:**663–70.

[23] Greenland S. Generalized conjugate priors for Bayesian analysis of risk and survival regressions. *Biometrics* 2003;**59:**92–99.

[24] Greenland S, Christensen R. Data augmentation for Bayesian and semi-Bayes analyses of conditional-logistic and proportional-hazards regression. *Stat Med* 2001;**20:**2421–28.

[25] Greenland S. Bayesian perspectives for epidemiologic research. II. Extensions to non-normal priors and regression, manuscript, 2006.

[26] Greenland S. Probability logic and probabilistic induction. *Epidemiology* 1998;**9:**322–32.

[27] Greenland S. Randomization, statistics, and causal inference. *Epidemiology* 1990;**1:**421–29.

[28] Greenland S. Multiple-bias modeling for analysis of observational data (with discussion). *J Royal Statist Soc* 2005;**168:**267–308.

[29] Titterington D. Common structure of smoothing techniques in statistics. *Int Stat Rev* 1985;**53:**141–70.

[30] Goldstein H. *Multilevel Statistical Models*. 3rd edn. New York: Oxford, 2003.

[31] Greenland S. Induction versus Popper: Substance versus semantics. *Int J Epidemiol* 1998b;**27:**543–48.

[32] Good I. Hierarchical Bayesian and empirical Bayesian methods (letter). *Am Stat* 1987;**41:**92.

[33] Zeger S. Statistical reasoning in epidemiology. *Am J Epidemiol* 1991;**134:**1062–66.

[34] Jaynes ET, Bretthorst GL. *Probability Theory: The Logic of Science*. New York: Cambridge University Press, 2003.

[35] Leamer E. False models and post-data model construction. *J Am Stat Assoc* 1974;**69:**122–31.

[36] Graham P. Bayesian inference for a generalized population attributable fraction. *Stat Med* 2000;**19:**937–56.

[37] Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc* 2003;**98:**47–54.

[38] Higgins JPT, Spiegelhalter. Being skeptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction. *Int J Epidemiol* 2000;**31:**96–104.

[39] Taubes G. Fields of fear. *Atlantic* 1994;**274:**94–100.

[40] Greenland S. Principles of multilevel modelling. *Int J Epidemiol* 2000b;**29:**158–67.

[41] Rothman KJ, Greenland S. Chapter 14, 15. *Modern Epidemiology*. 2nd edn. Philadelphia: Lippincott–Raven, 1998, pp. 197–99.

[42] Matthews R. Methods for assessing the credibility of clinical trial outcomes. *Drug Inf J* 2001;**35:**1469–78.

[43] Box GEP. Sampling and Bayes inference in scientific modeling and robustness. *J Roy Statist Soc A* 1980;**143:**383–430.

[44] Geweke J. Simulation methods for model criticism and robustness analysis. In Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). *Bayesian Statistics 6*. Oxford University Press: New York, 1998.

[45] Robins JM, Greenland S. The role of model selection in causal inference from nonexperimental data. *Am J Epidemiol* 1986;**123:** 392–402.

[46] Robins J. Data, design, and background knowledge in etiologic inference. *Epidemiology* 2001;**12:**313–20.

[47] Greenland S. When should epidemiologic regressions use random coefficients? *Biometrics* 2000a;**56:**915–21.

[48] Greenland S. Multilevel modeling and model averaging. *Scand J Work Environ Health* 1999;**25(Suppl. 4):**43–48.

[49] Greenland S. Methods for epidemiologic analyses of multiple exposures: a review and a comparative study of maximum-likelihood, preliminary testing, and empirical–Bayes regression. *Stat Med* 1993;**12:**717–36.

[50] Greenland S. Basic methods for sensitivity analysis of bias. *Int J Epidemiol* 1996;**25:**1107–16.

[51] Eddy DM, Hasselblad V, Schachter R. *Meta-Analysis by the Confidence Profile Method*. New York: Academic Press, 1992.

[52] Greenland S. Sensitivity analysis, Monte-Carlo risk analysis, and Bayesian uncertainty assessment. *Risk Anal* 2001b;**21:** 579–83.

[53] Greenland S, Finkle WD. A retrospective cohort study of implanted medical devices and selected chronic diseases in Medicare claims data. *Ann Epidemiol* 2000;**10:**205–13.

[54] Greenland S, Schwartzbaum J, Finkle WD. Problems from small samples and sparse data in conditional logistic regression analysis. *Am J Epidemiol* 2000;**151:**531–39.

[55] Rubin D. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics* 1999;**47:**1213–34.

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2006;**35:**775–777
doi:10.1093/ije/dyl055

Commentary: On Bayesian perspectives for epidemiological research

James R Carpenter

London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: james.carpenter@lshtm.ac.uk

Prevention is better and cheaper than cure, and the contribution of epidemiology to public health has been immense. Yet, when the number of practising epidemiologists is probably at an all time high, the discipline suffers angst.[1–3] Why is this, and what should be done?

Arguably, at a time when typical exposure effects are likely to be smaller, so the potential for being mistaken about exposure risks larger, the number of studies and (fuelled by the pressure to publish) the number of articles they generate is increasing. The result is more 'false-positive' results published, and an increasingly sceptical reception to all publications.[4]

Greenland lays into frequentist inference with relish and effectively tackles some common misunderstandings surrounding the Bayesian paradigm, arguing for its inclusion in epidemiological students' curricula as a matter of course.[5] He describes approximate procedures for using sceptical priors to shrink estimates towards the null.

Would epidemiology be much better off without frequentist statistics? Certainly it is true that the Neyman–Pearson hypothesis testing paradigm, which forces a choice between two alternatives, is not appropriate when we are in the process of gathering evidence on possible risks. Further we are all much more 'Bayesian' than we often admit, with prior beliefs motivating research questions, designs, and interpretations. You do not have to be a card-carrying Bayesian to vaguely expect to see a horse, and glimpse a donkey, then convince yourself you have seen a mule[6]—witness the peptic ulcer story.[3]

As Greenland says, raw data tells us nothing about the risk of an exposure without a statistical model. Even making inferences from a sample mean assumes an implicit model. Models are built on assumptions about the data (e.g. observations from different individuals are treated as independent) and relate the data to a parameter representing a quantity of interest. Suppose, as is common, we fit a model by maximum likelihood. Then the information about the parameter is in the likelihood; the question is how to interpret, or calibrate, this information.

Frequentist and 'pure likelihood'[7] inference looks at the ratio of the likelihood of the data at the maximum likelihood estimate to the likelihood of the data at the null value. The frequentist paradigm differs from the pure likelihood paradigm in its interpretation of this information. Frequentists ask: given there is really no exposure effect, what is the chance of seeing a likelihood ratio as extreme as this if we repeat our study lots of times? Roughly, pure likelihood inference interprets the evidence in the likelihood ratio by reference to likelihood ratios from familiar experiments in other settings. In this respect, I believe the 'pure likelihood' approach has an implicit subjective element and disagree with Greenland's sweeping statement that the likelihood paradigm 'replicates the pretense of objectivity that render frequentist methods so misleading'. If the likelihood ratio is 'too large to be believed', I submit the model is quite badly wrong.

Could this be the case here? As Greenland argues, issues about the appropriateness of additive or multiplicative models are secondary. But this should not let us move on, ignoring the elephants in the room. One is surely selection bias, on the basis of the '*P*-value'—whose presence is established in the submission[8], publication,[9,10] and citation of[11,12] randomized controlled trials, and surely bedevils epidemiology. Others are data contamination processes, from measurement error to missing data.

Frequentist inference may well coincide with poor modelling, or encourage its adherents not to think hard about the issues because they are being 'objective'. However, this does not mean it causes it; neither can a Bayesian approach average its effects away. Further, modelling can be improved without recourse to a Bayesian approach, although it may often be convenient.

On one reading, then, this paper[5] can be summarized as follows. It is hard to justify the data models often used in epidemiology. So, without quite knowing how these models are wrong, we are best advised to shrink the results by our vaguely formulated scepticism. In so doing, we dethrone the *P*-value. Better, though, to think hard in advance about what our priors should be, and inform them with concrete evidence. For measurement error, we need additional information to establish accuracy. For selection bias, a register of studies, to which new studies were added when they received ethical approval, would then enable us to read of the number of published studies and unpublished studies about a particular exposure. A simple formula[13] then gives us a bound on the publication bias. We can then decide if we wish to be less sceptical than this.

Unfortunately, studies are often poorly reported[14,15]; the further difficulty of accurate communication to the wider public is well known.[16] With regard to public presentation, epidemiological findings are often given to the press in black and white terms[1], with predictable consequences. The truth is never clear-cut. Why not summarize the results by the odds of the relative risk being, say, 1.5 vs 1. Expressing uncertainty through odds is already familiar to a large proportion of the public. Moreover, using such odds a journal could 'bet' on the studies it publishes: the degree to which it stays in the black (as opposed to the impact factor) is a measure of success much closer to the aims of epidemiology!

I also think odds may be a more accessible way into a Bayesian approach for students. We do not need to think of priors as expressing prior bets and then turn our priors into pseudo data to include in a statistical analysis. The latter process is useful, but not always easy, and a prior does not always lead to unique prior data. Furthermore, if the model is badly wrong, can the prior data really fix the problem?

Turning to the methods presented in the paper, information weighted averaging is a very useful tool and should be taught to epidemiological students. However, I do not share Greenland's views about the limited usefulness of MCMC methods in epidemiology. These provide a natural, and often computationally most feasible, general framework to model both the intrinsically interesting multilevel nature of much epidemiological data and data contamination processes [see e.g. Ref. (17) discussion].

In summary, I believe teaching epidemiology students to think more deeply about inference, modelling, and their implicit assumptions is vital, and I endorse teaching and using the Bayesian paradigm as part of this. However, I believe that to think it alone will save epidemiology from the fate some have predicted is wishful thinking.

References

[1] Taubes G. Epidemiology faces its limits. *Science* 1995;**269:**164–69.

[2] Tricopolous D. The future of epidemiology. *BMJ* 1996;**313:**436–37.

[3] Davey Smith G, Ebrahim S. Epidemiology—is it time to call it a day? *Int J Epidemiol* 2001;**30:**1–11.

[4] Ionnidis JPA. Why most published research findings are false. *PLoS Med* 2005;**2:**e124.

[5] Greenland S. Bayseian perspectives for epidemiologic research. I. Foundations and basic methods. *Int J Epidemiol* 2006;**35**:765–75.

[6] McPearson G. The Devil's drug development dictionary. Available at: www.senns.demon.co.uk/wdict.html (Accessed February 2, 2006).

[7] Royall R. *Statistical Inference: A Likelihood Paradigm*. New York: Chapman and Hall, 1997.

[8] Chan A, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials. *J Am Med Assoc* 2004;**20:**2457–65.

[9] Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technology Assess* 2000;**4:**1–115.

[10] Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decision revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *Am Stat* 1995;**49:**108–12.

[11] Nieminen P, Rucker G, Miettunen J, Schumacher M. Empirical evidence for preferential citation based on statistical significance. Submitted for publication, 2006.

[12] Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992;**305:**15–19.

[13] Copas J, Jackson D. A bound for publication bias based on the fraction of unpublished studies. *Biometrics* 2004;**60:**146–53.

[14] Pocock SJ, Collier TJ, Dandreo KJ *et al*. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* 2004; **329:**883.

[15] Altman D, Egger M, Pocock S, Vandenbrouke JP, von Elm E. Strengthening the Reporting of Observational studies in Epidemiology (STROBE). Available at: http://www.strobe-statement.org/ (Accessed February 6, 2006).

[16] Cox DR, Darby SC. The communication of risk. *J R Stat Soc [Ser A]* 2003;**166:**203–04.

[17] Greenland S. Multiple-bias modeling for analysis of observational data (with discussion). *J R Stat Soc [Ser A]* 2005;**168:**267–308.

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2006;35:777–778
doi:10.1093/ije/dyl081

Advance Access publication 27 April 2006

Response: Bayesian perspectives for epidemiological research

Sander Greenland[1,2]

Dr Carpenter[1] and I agree on the value of Bayesian perspectives and the inappropriateness of Neyman–Pearsonian testing for epidemiology. Unfortunately, he misrepresents several of my positions and misunderstands the import of data priors.

To prevent misuse of Bayesian methods, the meaning of priors must be made clear. Shockingly to me, Carpenter dismisses data priors with 'We don't need to think of priors as expressing prior bets, and then turn our priors into pseudo data to include in a statistical analysis. The latter process is useful, but not always easy, and a prior does not always lead to unique prior data.' This passage overlooks every advantage of data priors:

(i) We need not think at all to apply any statistical method, and that is *not* a virtue of statistics: One can just cram data into software, give incorrect (if standard) interpretations of the output, and get those published. Translating priors into data makes painfully clear how much prior information is assumed by a Bayesian analysis, in a way odds do not. Data priors reveal the enormous knowledge claimed by some writers, as shown in the magnetic field/leukaemia example in my paper.

(ii) Data priors demonstrate that the same approximate methods taught for frequentist analysis serve just as well for Bayesian analysis (this fact is apparently upsetting to Bayesians who have laboured over specialized Bayesian techniques).

(iii) Creation of data priors is much less difficult than what is routinely done to torture data, such as logistic regression. It requires only a moment with a calculator to solve the equations given in my paper; whether the solution is unique is irrelevant, since all solutions lead to approximately the same posterior. It does not begin to approach the computing or convergence issues needed for Markov Chain Monte Carlo.

As I will describe in part II of this series,[2] posterior computation via data priors extends far more easily to regression and multilevel modelling than do other methods. It has the remarkable property of *not* increasing in complexity with the underlying model, because it involves only adding data records and using ordinary regression software, and can employ highly non-normal priors. In contrast, information weighting for regression requires normal priors and additional matrix computations.

Dr Carpenter goes on to ask 'if the model is badly wrong, can the prior data really fix the problem?' I never claimed data priors fix model problems, because they do not; neither do Markov-Chain Monte-Carlo, propensity scores, inverse-probability weighting, or other modern tools, for they are all

[1] Department of Epidemiology, University of California, Los Angeles 90095-1772, USA.

[2] Department of Statistics, University of California, Los Angeles 90095-1772, USA.

E-mail: lesdomes@ucla.edu

methods that assume models and compute from there. Model-expansion methods like bias modelling addresses model deficiencies; data priors address contextual understanding and computational transparency. Sadly, context and transparency are largely neglected by statistical research, which instead focuses on generality and precision beyond any relevance to most health or social scientists.

Dr Carpenter closes his commentary with 'However, I believe that to hope [the Bayesian paradigm] alone will save epidemiology from the fate some have predicted is wishful thinking.' I never said and do not believe that 'the Bayesian paradigm alone' is sufficient for 'saving epidemiology' or for epidemiological inference; I only argue that Bayesian thinking is necessary as part of a broad and well-rounded approach. A major flaw shared by most statistical methods (be they frequentist, likelihoodist, or Bayesian) is that they pretend the data came from an ideal study in which all important influences on the data (including those of the investigators and the subjects) can be approximated by a known model. In observational epidemiology this can be a fatal assumption, resulting in far too much certainty placed on statistical results. Others and I have discussed how statistics can approach this problem in a contextually informed manner by using bias models with explicit priors for bias parameters.[3]

Note well: There is no single Bayesian paradigm[4] just as there is no single frequentist paradigm.[5] I argue that a *subjective* Bayesian perspective is needed. Neither 'objective' Bayesian methods[6] nor 'pure-likelihood' methods[7] address the pseudo-objectivity that frequentism perpetuates, and like frequentist methods they contain hidden subjective elements. I do not hold that subjective Bayesian methods should replace these methods; instead I concluded that 'they should take their place alongside frequentist approaches'[3] in a Bayesian/frequentist dualism.[4,8,9]

A dualistic approach is needed because Bayesianism and frequentism address different questions. Bayesianism addresses questions of the form 'Having seen the data, what odds should I place on this hypothesis versus another?' and seeks methods that use contextual information to improve the bet; it cares about the observed data, not counterfactual data as might arise under a hypothetical long run. In contrast, frequentism addresses questions of the form 'if I applied this method to a hypothetical long run of studies like this one, how would it behave?' and seeks methods with desirable long-run behaviour; it does not care about odds of hypotheses given the data that were actually observed, or even whether a particular decision produced by applying its methods to those data is better than alternatives. Importantly, Bayesian methods exhibit desirable frequentist (long-run) properties when both they and the evaluation are well informed by the scientific context.[10]

Most researchers are not interested in evaluating statistical methods, however, but instead are interested in contextual hypotheses. Smart researchers understand that an epidemiological study cannot by itself form a sound basis for accepting or rejecting hypotheses but can only provide evidence within a larger context. Contextual statements about hypotheses are sought and are what Bayesian statistics can provide. It is thus unsurprising that, having been given only frequentist methods, researchers consistently misinterpret frequentist outputs as if those were Bayesian (e.g. they write their discussions as if their two-sided *P*-value is the probability of the null hypothesis given the data).

The mismatch between the methods researchers are taught and the questions they actually ask has produced a chronic psychosis in study reporting, in which 'nonsignificance' is taken as evidence for the null, or (as Dr Carpenter notes) even as grounds to not report or cite the result, thus distorting entire literatures and reviews. It is technically correct that the blame for such nonsense is with the user, not with frequentism; emphasis on the width and limits of confidence intervals rather than on statistical significance might avoid such problems.[11,12] But blaming users is like blaming consumers for eating junk food when that is most promoted to them. We should blame elementary statistics textbooks and teachers for giving users such poor conceptual foundations, and providing only tools inappropriate for the most user's questions. In the US at least, most statistics for non-statisticians still fails to give Bayesian perspectives any meaningful time. Fortunately, applied statistics has rediscovered the importance of those perspectives for answering scientific questions.[8,9] It is time for basic statistical education—and epidemiology—to catch up, and data priors are the simplest way to do so.

References

[1] Carpenter JR. Commentary: On Bayesian perspectives for epidemiological research. *Int J Epidemiol* 2006;**35**:775–77.

[2] Greenland S. Bayesian perspectives for epidemiologic research. II. Extensions to non-normal priors and regression analysis. Submitted.

[3] Greenland S. Multiple-bias modeling for analysis of observational data (with discussion). *J R Stat Soc Ser A* 2005;**168:**267–308.

[4] Good IJ. *Good Thinking*. Minneapolis: University of Minnesota Press, 1983.

[5] Goodman SN. *P* values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate (with discussion). *Am J Epidemiol* 1993;**137:**485–501.

[6] Berger JO. The case for objective Bayesian analysis. *Bayesian Analysis*, to appear 2006 (Available at: http://www.stat.cmu.edu/bayesworkshop/2005/panel.html).

[7] Royall R. *Statistical Inference: A Likelihood Paradigm*. New York: Chapman and Hall, 1997.

[8] Rubin DB. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics* 1991;**47:**1213–34.

[9] Efron B. Bayesians, frequentists, and scientists. *J Am Stat Assoc* 2005;**100:**1–5.

[10] Gustafson P, Greenland S. The performance of random coefficient regression in accounting for residual confounding. *Biometrics* 2006;**62**:In press.

[11] Poole C. Low P-values or narrow confidence intervals: which are more durable? *Epidemiology* 2001;**12:**291–94.

[12] Altman DG, Machin D, Bryant TN, Gardner MA (eds). *Statistics with Confidence*, 2nd edn. London: BMJ Publishing Group, 2000.