
For Debate

The statistical basis of public policy: a paradigm shift is overdue

R J Lilford, D Braunholtz

The recent controversy over the increased risk of venous thrombosis with third generation oral contraceptives illustrates the public policy dilemma that can be created by relying on conventional statistical tests and estimates: case-control studies showed a significant increase in risk and forced a decision either to warn or not to warn. Conventional statistical tests are an improper basis for such decisions because they dichotomise results according to whether they are or are not significant and do not allow decision makers to take explicit account of additional evidence—for example, of biological plausibility or of biases in the studies. A Bayesian approach overcomes both these problems. A Bayesian analysis starts with a “prior” probability distribution for the value of interest (for example, a true relative risk)—based on previous knowledge—and adds the new evidence (via a model) to produce a “posterior” probability distribution. Because different experts will have different prior beliefs sensitivity analyses are important to assess the effects on the posterior distributions of these differences. Sensitivity analyses should also examine the effects of different assumptions about biases and about the model which links the data with the value of interest. One advantage of this method is that it allows such assumptions to be handled openly and explicitly. Data presented as a series of posterior probability distributions would be a much better guide to policy, reflecting the reality that degrees of belief are often continuous, not dichotomous, and often vary from one person to another in the face of inconclusive evidence.

Every five to 10 years a “pill scare” hits the headlines. Imagine that you are the chairperson of the Committee on Safety of Medicines. You have been sent the galley proofs of four case-control studies showing that the leading brands of oral contraceptive, which have been widely used for some five years, are associated with a doubling of the risk of venous thromboembolism. You are surprised; you seem to remember that these new brands contain an “improved” progesterone which has been shown to have no adverse effects on clotting factors—indeed the widespread acceptance of this treatment was predicated on the favourable metabolic effects of the new compound. A literature search and telephone call to local experts confirms your memory. You are aware that case-control studies are often biased. What do you do?

On the one hand you do not wish to over-react. After all, even if the newer brands do carry a higher risk of thrombosis, the risk arising from pregnancy is higher still. Thus widespread alarm may precipitate contraceptive withdrawal in mid-cycle and hence do more harm

than good. On the other hand, if you fail to issue a statement advising the profession that a statistically significant doubling of the risk of deep vein thrombosis has been measured then you lay yourself (and others) open to public criticism when, sooner or later, reports of a serious medical mishap are brought to public attention. “Why did you not warn the public so that individuals could make an informed choice? After all, there was a ‘statistically significant’ doubling of thrombosis rates in the study.”

The scenario painted here has an obvious similarity to the recent controversy surrounding oral contraceptives containing new third generation gestagens. Four case-control studies (one nested in a cohort study) have recently been reviewed by McPherson.¹ Taken together they show a statistically significant doubling in the risk of venous thromboembolism. We are not experts in this subject and do not want to add to this particular debate: we want to make a general point about the interpretation of new data in the context of a treatment (or prophylaxis) of which the clinical community has had considerable experience and about which other data exist.

Our thesis is that conventional statistical tests and estimates are an improper basis for public policy for two reasons. Firstly, they dichotomise results according to whether or not they are “significant,” thereby tending to produce an off/on response by decision makers. Secondly, they do not take account of additional evidence (generated outside or within the index study) in an explicit way. Such evidence must then be handled implicitly, and this makes it much less useful in defending decisions. The statistically significant result seems “hard” and is explicit, while the notion that our conclusions should be tempered by knowledge of the biochemistry and plausible biases seems “soft” and that knowledge is handled in an implicit manner. Since the statistical analysis does not incorporate these additional factors, they cannot impact explicitly on the conclusions. The chairperson of the Committee on Safety of Medicines is placed on the defensive: she may be seen to be “explaining away” the observed effect if she does not act decisively in the direction predicated by the statistically significant result.

Confronting the difficulty: the Bayesian alternative

But is there another way to proceed: how else can statistics be used to guide policy on an issue of private and public concern? Clearly, if clear cut answers are available then an unambiguous official statement should follow. The effects of the sun’s rays on skin cancer and of posture on sudden infant death may be examples where epidemiology has produced sufficiently clear cut answers to provoke specific recommendations. When the situation is less clear cut, however, as in the case of third generation oral contraceptives, conventional

University of
Birmingham,
Birmingham B16 9PA
R J Lilford, *professor of health
services research*

Nuffield Institute for
Health, Leeds University,
Leeds LS2 9PL
D Braunholtz, *medical
statistician*

Correspondence to:
Professor Lilford.

BMJ 1996;313:603-7

Bayesian statistics

The key difference between Bayesian and conventional (or frequentist) statistics is the view of what probability is. Frequentists view probability as a relative frequency, or proportion. Thus the probability P of a fair coin landing heads up is 0.5 because in a long series of tosses it lands heads up half the time. Frequentists should not therefore estimate probabilities for one off events—like the probability of President Clinton winning a second term. Strictly, of course, all events are one off, but many events are similar enough to satisfy frequentists' requirements. Bayesians, on the other hand, view probability as a degree of personal belief. Personal belief changes as evidence (data) accrues, but no data at all are necessary. A Bayesian might judge the value of P to be close to 0.5, without the need for any previous experience of coin tossing—on the basis of the physics involved. In fact he or she would want to give a probability distribution for the true value of P . This would be a prior distribution for P , which could then be updated via coin tossing (by means of Bayes's law) to produce a posterior distribution of probabilities.

Bayes's law in itself is uncontroversial and is used by frequentists as well as Bayesians, but frequentists use it in much more restricted circumstances. The classic examples are Mendelian genetics and computerised diagnosis, such as that popularised in the UK by the late professor Tim deDombal. Bayes's law as used by Bayesians simply states that the posterior probability dis-

tribution is formed by weighting the prior probability distribution by the likelihood.

One practical advantage of the Bayesian approach is that it provides probability distributions for parameters—which is exactly what is needed to inform decisions. As we show in this paper, it also makes the synthesis of new data, and other kinds of evidence, relatively straightforward. Frequentists would argue that the disadvantage is that prior beliefs, being personal, can vary—and conclusions may therefore differ from person to person. Bayesians would respond that that is what real life is like. Also, by carefully doing sensitivity analyses, researchers can assess how robust conclusions are to changes in prior probability distributions, or indeed to changes in the model used to create the likelihood.

Other than in very simple cases (such as that presented here) calculating the posterior probability distribution becomes impossible analytically, and it has to be approximated—for instance, using "Monte-Carlo" methods on computers. This involves generating a large, random, sample from the posterior probability distribution (each number generated may involve substantial computations), and the properties of the posterior probability distribution are "discovered" by analysing this sample.

The advent of fast, cheap computers now makes this feasible for almost anyone, and programs such as BUGS (available from ftp.mrc-bsu.cam.ac.uk) are making it easier to do.

statistics may drive decision makers into a corner (resulting in either false reassurance or excessive caution) and produce sudden, large (and hence potentially harmful) changes in prescribing. The problem does not lie with any of the individual decision makers, but with the very philosophical basis of scientific inference. We propose that conventional statistics should not be used in such cases and that the Bayesian approach is both epistemologically and practically superior.²

Here we start with prior belief, which is measured and made explicit. We then incorporate the new data but in so doing we may adjust for the likely extent of bias. We then combine the prior with the adjusted data to obtain a "posterior" probability distribution, using the mathematical theorem associated with the name of the eighteenth century clergyman, Thomas Bayes (see box). Lastly, we carry out a sensitivity analysis, to see what effects different prior beliefs and different assumptions about possible bias might have. Given the data, almost everyone will now have a stronger belief that third generation pills cause clots in the venous system than they had before, but everybody does not have to believe the same thing. Even without considering possible beneficial effects on the risk of heart attack, the health care system can respond incrementally and not precipitate a large scale shift in prescribing practice. There would be little reason for a scare story causing a surge in demand for consultations and in unwanted pregnancies. The principles of Bayesian inference are described in more detail in the box.

Bayesian inference: how it works

We give a worked example, based on McPherson's summary, which shows an odds ratio of 2 for the risk of deep venous thrombosis when the third generation pills were compared with others. Since the risks are small, we can think of the odds ratio as a relative risk. The 95% confidence interval ranges from a relative risk of 1.4 to 2.7. Clearly the 95% confidence interval excludes 1 and the results are therefore significant at the usual $P < 0.05$ level. P here is the proportion of times that an effect of this size (or greater) would be measured in an infinite repetition of studies if the true effect was 1—that is, both third generation and older pills were associated with the same risk.

However, decision makers want to know the probabilities of thrombosis for the next patient who is eligible for either treatment. A decision maker might ask: "What is the probability that the third generation pills increase the risk when compared to the others; what is the probability that they at least double the risk—as measured in the case-control study; and what is the 'median estimate' (as likely to be too small as too large)?"

The calculations require a prior probability distribution for the true effect. We could obtain this by measuring the collective prior belief of experts. We could contact, say, 25 randomly selected members of the Faculty of Family Planning, probably before they knew about the new data. We would interrogate them to see what their thoughts were on: (a) the best estimate of the true relative risk—the effect of the third generation pills on the risk of clotting when compared with the standard pills; (b) what values they thought were unlikely for the true relative risk—such that an effect of that size or more extreme would have a chance of being true of less than 0.025. The answers are those that respondents would give if they were forced to set odds and accept any bets while wishing to minimise their losses. For example, they might set odds of 19:1 that the true relative risk would lie within the interval specified at (b) above. Imagine that our average respondent thinks that the true relative risk is as likely to be above as below 0.8 (corresponding to a 20 percentage point reduction in risk (relative risk=0.8)) and that a relative risk of 1.6 or greater, or of 0.40 or less, are unlikely to be true. In that case, their prior distribution of probability estimates could be represented on a log relative risk scale as a normal curve—prior distribution 1 in fig 1.

Bayes's theorem allows us to update this prior distribution to take account of McPherson's data, which are converted into a likelihood—likelihood A in fig 2. This updating of the prior distribution by the likelihood would give us the posterior distribution of probabilities referred to as posterior 1A in fig 1. The middle of the posterior distribution corresponds to a relative risk of about 1.69 and the 95% interval (now referred to as a *credible* interval rather than a confidence interval) for the relative risk ranges from 1.3 to 2.3. If asked to state the most likely effect an observer with prior 1 would give a relative risk close to 1.69 (it is not exactly 1.69 because

of the mathematical point made in the legend to fig 1). For the mathematically minded, likelihood is discussed in more detail in the box below.

Taking into account different beliefs and likely bias: sensitivity analysis

The above figures represent the probabilities for an observer who agrees with the prior distribution of probabilities. We discussed these prior probability distributions with two eminent Leeds gynaecologists with an interest in family planning. Dr Nicholas Johnson agreed with these probability estimates and hence with the posterior probability distributions. Professor James Drife, however, was more sceptical: he was in absolute equipoise³ before the new data—that is, he thought it equally likely that the third generation or standard oral contraceptives had a higher risk of causing deep vein thrombosis. However, like Johnson, his prior probability distribution was vague, admitting of an equally wide range of plausible values, with a 95% probability that the true relative risk was between 0.5 and 2.0 (curve prior 2 in fig 1). For Drife, the middle relative risk, when both the data and prior belief are taken into account, is 1.76 and the 95% credible interval extends from 1.3 to 2.4—posterior 2A in fig 1. The comparison of Johnson (who was cautiously enthusiastic to start with), Drife (who was sceptical), and yet other experts who may hold more extreme views constitutes a sensitivity analysis.

Sensitivity analysis can be extended to take into account evidence that case-control and other observational studies are often biased and that in this particular case we have reasons to suspect that the measured effect has been overestimated.

Firstly, we could suppose that the particular design and implementation of the studies contributing to McPherson's summary may result in a bias but that this bias is as likely to be positive as negative. We could further suppose that the distribution of this bias was normal on a log relative risk scale, with a standard deviation (SD) of 0.2624 (corresponding to a multiplying, or dividing, factor of 1.3 on the relative risk scale) so that the biased relative risk being estimated from McPherson's summary would be in the range of 60% to 167% of the true relative risk, with probability 0.95. This weakening of the evidence provided by the data results in likelihood B (fig 2) and in a posterior probability distribution closer to the prior distribution, as illustrated in fig 3. Posterior 2A is as in fig 1 (no bias), but posterior 1B and posterior 2B (from Johnson and Drife's prior probability distributions respectively) assume a bias in the included studies distributed as just described.

Secondly, however, it appears that non-randomised studies typically overestimate treatment effects by about 30%,^{4,5} and in this instance we have reason to suspect an overestimate. Firstly, third generation pills may have been given preferentially to higher risk women, and it is never possible to be certain that this has been fully accounted for by statistical adjustment.⁶ Secondly, more "modern" general practitioners may both preferentially prescribe newer brands of pill and be especially vigilant in investigating symptoms which could result from venous thromboembolism. Thirdly, women using oral contraceptives which have been in use for a long time are biased with respect to those on newer brands, because many of those with venous thromboembolism (which typically occurs within a few months of starting the pill) will have been screened out—the so called "healthy user effect."^{7,8} If we assume a median bias of 30%, given the above, and make no other new assumptions, then the biased relative risk estimated from the summary would be in the range 78% to 217% of the true relative risk, with probability 0.95. The evidence from the data is thus both weakened and shifted—see likelihood C in fig 2. The resulting posterior probability distributions are shown in fig 4, where posteriors 1C and 2C were derived from Johnson's and Drife's prior distributions respectively. The middle of Drife's posterior probability distribution now corresponds to a relative risk of 1.27, while for Johnson a true relative risk of above or below a central value of only 1.16 is equally likely. The probabilities that the relative risks of venous thrombosis are not increased at all with the third generation pills are 15% for Drife and 27% for Johnson. A relative risk of 1.27, calculated on the basis of Drife's original prior probability distribution (which was both equiposed and fairly vague), the data, and (arguably) modest assumptions of bias, translate into 0.4 to 0.8 additional cases of venous thromboembolisms per 10 000 women years (assuming a background risk of between 1.5 and 3 venous thromboembolisms per 10 000 women years on the previous generation of pills).

Manipulation or simply recognising reality?

Some people will feel very uneasy about these and other adjustments in a sensitivity analysis: the judgmental manipulation of "real" figures may seem wrong. Wrong that is, until we examine the alternative, which is uncritically to accept data which we suspect to be less reliable than, say, the results of a randomised controlled trial. If there is reason to suspect systematic bias then it seems inappropriate not to allow for this in the analysis.^{9,10} In this case not only is there empirical

Likelihood

When trying to understand the implications of a dataset researchers usually focus on a few parameters of special interest, which in some way summarise the interesting facets of the data. In this case the parameter of interest is the relative risk. Note that this is not directly observable in the data, but is an intangible idea that we find useful.

Parameters are linked to the data via a model, which describes the sort of data associated with particular values of the parameters. In this case the model we have assumed specifies that the probability distribution for the "observed" log relative risk will be normal with a mean of log (true relative risk) and a known standard deviation. In fact the standard deviation really depends on the sample size and the value of the true relative risk, but in our simple analysis we estimate the standard deviation from the data and then pretend we know it. Of course we have only one dataset, and we do not know the true parameter (relative risk) values. We consider all possible true

parameter values, and for each calculate the probability of getting the data actually obtained. These probabilities can be plotted on a graph, and, when thought of as providing information on the likely true value of the parameter given the data, this plot is called the likelihood. Bayesians adhere to the intuitively attractive likelihood principle, which states that information arising from studies or experiments should be based only on the actual data observed. Frequentists often find themselves in conflict with this—for instance, when calculating P values, which take into account the probability of observations more extreme than the actual observations. However, in the case of the normal distribution conventional methods in effect use the likelihood to calculate confidence intervals, and we have used this in converting McPherson's summary into a likelihood: the likelihood is normal on the log (relative risk) scale, centred around log (2.0), and with a standard deviation such that log (2.7) – log (1.4) is 2×1.96 SD.

evidence that observational studies in general may be biased; there are plausible reasons to suspect bias in a particular direction. Thus any bias would be replicated across studies if the confounding factor was typical of the "treatment" in question. An advantage of explicit manipulation of the data, before statistical analysis, is that the process is transparent and hence open to challenge and recalculation on the basis of different assumptions.

Data presented as a series of posterior probability distributions (each based on a respective prior probability distribution and assumption of likely bias) would be a much better guide to policy than results analysed in the conventional way. They would reflect the reality that degrees of belief (a) are continuous or incremental, but not dichotomous, and (b) vary (quite properly) from one person to another in the face of inconclusive evidence.

On the above scenarios some clinicians might change prescribing habits, while others would be "sensitised" (have a new, more cautious, prior distribution) against the day when yet more data may become available. Women themselves could see that evidence regarding venous thromboembolism was moving against the new pills but would not be alarmed by the notion that harm was proved by "statistics." They would understand the new data (correctly) as merely one more piece of evidence in a complex array. This would encourage women to derive their own estimate of likely risk in con-

sultation with their clinician and make any trade off required by perceptions of countervailing benefit.¹¹

In the case of some of the third generation pills there is reason to believe that the risk of heart attack is reduced, in comparison with earlier brands. The newer pills have more favourable effects on blood fats than their second generation cousins. On the basis of this information alone, many rational observers may have formed a prior probability distribution which, while vague, was shifted in the direction of net benefit—that is, many may have had a prior distribution with respect to heart attack similar to that which Johnson had with respect to venous thromboembolism. One of the studies quoted by McPherson does, in fact, give results for heart attacks: the odds ratio is 0.36, suggesting that the risk is indeed lower with third generation pills, but the confidence interval is wide (0.1 to 1.2).¹² Thus, although the latter results are not statistically significant, perhaps because the number of adverse events is still small, they could be used to update a Bayesian prior probability distribution. With any reasonable prior belief and assumptions about bias the posterior probability distribution will be centred on a large reduction in relative risk, but will be widely spread. The uncertainty (corresponding to non-significance in frequentist terms) is, however, no reason to ignore the effect of the newer pills on heart attacks, since that is essentially to assume with complete certainty that there is no effect.

Fig 1—Probability distributions, on a log (relative risk) scale, of relative risk of venous thromboembolism in third generation contraceptive pills compared with second generation pills. All prior distributions and likelihoods (and hence, owing to the mathematics of Bayes's theorem, posterior distributions) are assumed to be normally distributed on the log scale.

Priors 1 and 2 are Johnson and Drife's respectively. Both are fairly wide, indicating considerable doubts about the value of the true relative risk. Drife's prior is centred on $\log(1.0)$ as (before learning of the new case-control study data) he believed that third generation pills were as likely to be better as to be worse than second generation pills. Johnson was more optimistic that the new pills would have a lower risk of venous thromboembolism, his prior distributions being centred on $\log(0.80)$. If McPherson's summary of the various studies is taken at face value (likelihood A in fig 2) and is used to update the experts' prior distributions via Bayes's theorem, posterior distributions 1A and 2A result. These are much narrower than the prior distributions, indicating less doubt about the value of the true relative risk. The data (with an observed odds ratio of 2.0) has influenced the posterior distributions more than the rather vague prior distributions, with the result that they are centred on $\log(1.69)$ and $\log(1.76)$ respectively, and the probability of the true relative risk being greater than 1 is more than 0.999 in both cases.

Note: The most probable value for the true log (relative risk) is not equal to log (most probable value of relative risk)—that is, the position of the highest points of the probability distributions drawn on log (relative risk) and on relative risk scales do not correspond. For instance the most probable value of log (relative risk) for prior 2 is $\log(1.0)=0$, but the most probable value of relative risk for prior 2 is 0.89, not 1.0. This is because the whole of the negative log (relative risk) axis (and its probability) is "squashed" into the interval (0,1) on the relative risk scale, while the positive log relative risk axis is increasingly stretched out. The centres of distributions are not affected by this problem and have for this reason been used in this paper.

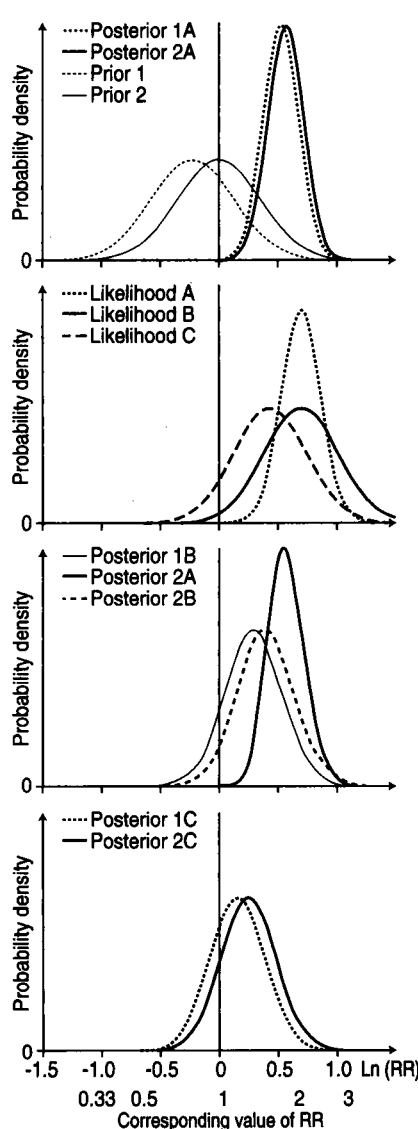


Fig 2—Posterior distributions are calculated by "weighting" the prior distributions by the data likelihood. The likelihood can be calculated as the probability of the data given varying true values of the parameter (in this case log (relative risk)) but is viewed as the likelihood of the various parameter values given the data. The likelihoods shown here correspond to McPherson's summary of the various studies (relative risk of 2.0, 95% confidence interval 1.4 to 2.7): (A) taken at face value; (B) assuming the summarised data may be biased, with bias drawn at random from a normal distribution on a log (relative risk) scale with mean zero, and SD of $\log(1.3)$; (C) as (B) but with mean log bias $\log(1.3)$. Clearly an assumption that the data may be biased reduces the information from the data, and if the mean bias is thought to be non-zero, the information is also shifted accordingly.

Fig 3—Posterior 2A is as in fig 1, deriving from Drife's prior distribution and the data taken at face value (likelihood A). Posterior 2B again derives from Drife's prior distribution, this time weighted by likelihood B. The information conveyed by the data is thus much reduced, and the posterior distribution correspondingly wider, and closer to the prior distribution. It is centred on $\log(1.48)$, and the probability that the true relative risk is less than 1—that is, that the new pills have reduced risk of venous thromboembolism—is now much increased (though still small) at 0.048. Posterior distribution 1B derives from Johnson's prior distribution and likelihood B. As would be expected, it produces a much higher probability (0.11) that the true relative risk is less than 1.

Fig 4—Both posterior distributions derive from the data summary adjusted by the assumption that the studies which produced the data may have been biased, with the summary bias (on log relative risk scale) sampled from a normal distribution with mean $\log(1.3)$ and SD $\log(1.3)$ —that is, likelihood C. The information conveyed by the data is thus reduced and shifted. Johnson and Drife's prior distributions, when weighted by likelihood C, result in posterior distributions 1C and 2C respectively centred on $\log(1.16)$ and $\log(1.27)$. The probabilities that the true relative risk is less than 1—that is, that the new pills actually reduce risk of venous thromboembolism—are further increased, to 0.27 and 0.15, and are now far from negligible in both cases.

Thomas Bayes

Bayes was a member of the first secure generation of English religious non-conformists. His father, Joshua Bayes FRS, was a respected theologian of dissent; he was also one of the group of six ministers who were the first to be publicly ordained as non-conformists. Privately educated, Bayes became his father's assistant at the presbytery in Holborn, London; his mature life was spent as minister at the chapel in Tunbridge Wells. Despite his provincial circumstances, he was a wealthy bachelor with many friends. The Royal Society of London elected him a fellow in 1742. He wrote little: *Divine Benevolence* (1731) and *Introduction to the Doctrine of Fluxions* (1736) are the only works known to have been published during his lifetime. The latter is a response to Bishop Berkeley's *Analyst*, a stinging attack on the logical foundations of Newton's calculus; Bayes' reply was perhaps the soundest retort to Berkeley then available.

Bayes is remembered for his brief "Essay towards solving a problem in the doctrine of chances" (1763), the first attempt to establish a method to calculate a probability distribution (the probabilities of different events occurring) given a set of data. In so doing he laid the foundations for statistical inference.

Before Bayes there was some understanding of how to reject statistical hypotheses in the light of data, but no one had shown how to measure the probability of statistical hypotheses in the light of data. Bayes began his solution of the problem by noting that sometimes the probability of a statistical hypothesis is given before any

particular events are observed; he then showed how to compute the probability of the hypothesis after some observations are made. Bayes was himself too modest to claim that he had solved the basis for the whole of statistical inference, and it was left to Richard Price to submit his work to the Royal Society. However, the great Laplace had no qualms about Bayes's argument; his enormous influence made Bayes's ideas almost unchallengeable until George Boole protested in his *Laws of Thought* (1854). Since then Bayes's technique has been a constant subject of controversy. The controversy relates to deriving the probability of statistical hypotheses (prior probability distributions), especially before any data of the type we want to analyse have been observed.

In *Foundations of Statistics* Leonard J Savage interprets probability in a personal way, as reflecting a person's personal degree of belief; hence, a prior probability distribution is a person's belief before the new observations become available, and a posterior probability distribution is a person's belief after the observations are made available. In the past 10 years or so there has been a sharp revival of interest in Bayes's work, especially its application to medical problems. Researchers in the UK have been in the forefront of this resurgence: they include David Spiegelhalter at the MRC Biostatistics Unit Cambridge, Adrian Smith at Imperial College, London, Deborah Ashby at the University of Liverpool, and, from a philosophical perspective, Peter Urbach at the London School of Economics.

A reasonable approach to answer the relevant question—Are third generation pills preferable to second generation pills?—needs to deal in absolute risks and explicit "costs" to women. The absolute risk of heart attack in users of second generation pills is even lower than that of venous thromboembolism,¹² but a heart attack is typically more serious—so the overall mortality and morbidity due to both may be similar. The combined posterior distribution for the difference between third and second generation pills in total mortality may thus be quite spread out, with a substantial proportion of the area—that is, probability—on both sides of the origin. A summary would conclude that, although it looks fairly probable that venous thromboembolism occurs somewhat more frequently with third generation pills, there is still considerable doubt as to which is safer overall. Such a statement would not have been likely to initiate large scale changes in prescribing, except for women with risk factors for venous thromboembolism. The possibility of collecting more useful data on the safety of third generation pills would not have been all but removed, as McPherson suggests it has been in his editorial.¹ The importance of collecting more data on the safety of third generation pills—to tighten up the posterior distributions—would be emphasised.

Acknowledging imperfections: a better basis for public policy

Bayesian techniques allow all our current knowledge to be explicitly represented and synthesised with new data. If there is little knowledge this is reflected in vague prior probability distributions. If explicit costs and benefits can be assigned to outcomes decision analysis¹¹ can then be used to trade off the best available estimates of benefit and harm, incorporating preferences for health in the short over the long term. Conventional statistics do not include all the evidence within the calculations. They therefore dichotomise results and tend to result in sensationalism. Faced with data presented in Bayesian and decision analysis terms journalists would have to communicate with the public in a more sophisticated way to show how probabilities vary according to different interpretations of the "starting" information and that the final decision can take account of personal

trade offs. Practical actions are based on (often unrecognised) philosophical assumptions. A move from standard to Bayesian statistics would represent a fundamental change in how we think about knowledge and this in turn would affect policy making.

Health issues are now much more complex and the amount of disparate evidence that impacts on belief has increased. Only the Bayesian approach can do justice to all this information and provide the probabilistic basis for action when the results of a particular type of study have not (yet) reached statistical significance or, indeed, for not acting when they have. Sheldon and Smith have advocated this method in the context of environmental effects on health,¹³ and a change in approach is overdue in this and other areas of public policy.

We thank Professor Zephne Van Der Spuy and Dr Victoria Lilford, whose dinner party conversation provided the inspiration for this article, and Professor James Drife and Dr Nicholas Johnson for the helpful discussions alluded to in the text.

- 1 McPherson K. Third generation oral contraception and venous thromboembolism. *BMJ* 1995;312:68-9.
- 2 Spiegelhalter D, Freedman LS, Parmar MKB. Bayesian approaches to randomised trials. *Journal of the Royal Statistical Society* 1994;157:357-416.
- 3 Lilford R J, Jackson J. Equipose and the ethics of randomization. *J R Soc Med* 1995;88:552-9.
- 4 Schulz, K F, Chalmers I, Grimes, D A, Altman, D G. Assessing the quality of randomization from report of controlled trials published in obstetrics and gynaecology journals. *JAMA* 1994;272:125-8.
- 5 Sacks H, Chalmers T, Smith H. Randomised versus historical controls for clinical trials. *Am J Med* 1982;72:233-40.
- 6 Buring JE, Glynn RJ, Hennekens CH. Calcium channel blockers and myocardial infarction: a hypothesis formulated but not yet tested. *JAMA* 1995;274:654-5.
- 7 Reijnen HBM, Atsma WJ. Risk is highest in the first months of use. *BMJ* 1995;311:1637.
- 8 Arrighi HM, Hertz-Picciotto. The evolving concept of the healthy worker survivor effect. *Epidemiology* 1994;5:189-96.
- 9 Ashby D, Hutton JL, McGee MA. Simple Bayesian analysis for case-control studies in cancer epidemiology. *The Statistician* 1993;42:385-97.
- 10 Eddy DM, Hasselbald V, Shachter R. A Bayesian method for synthesising evidence: the confidence profile method. *International Journal of Health Technology Assessment* 1990;6:31-55.
- 11 Lilford R J, Thornton J. Decision logic in medical practice. *J R Coll Phys* 1992;26:1-20.
- 12 Lewis MA, Spitzer WO, Heinemann AJ, MacRae KD, Bruppacher r, Thorogood M. Third generation oral contraceptives and risk of myocardial infarction: an international case-control study. *BMJ* 1996;312:88-90.
- 13 Sheldon TA, Smith D. Assessing the health effects of waste disposal sites: issues in risk analysis and some Bayesian conclusions. In: Clark M, Smith D, Blowers A. Eds. *Waste location*. London: Routledge, 1990.

(Accepted 17 June 1996)