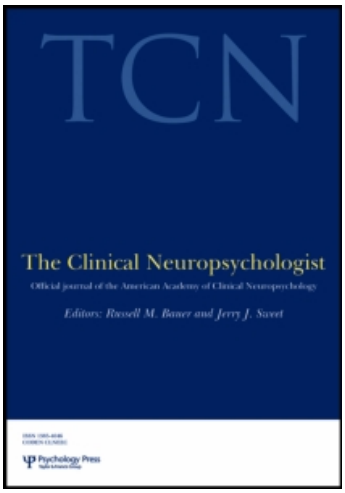

The Clinical Neuropsychologist
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713721659

Bayes' theorem and diagnostic tests in neuropsychology: Interval estimates for post-test probabilities

John R. Crawford [a]; Paul H. Garthwaite [b]; Karolina Betkowska [a]
[a] School of Psychology, University of Aberdeen, [b] Department of Mathematics and Statistics, The Open University, Milton Keynes, UK

First Published on: 20 February 2009

To cite this Article Crawford, John R., Garthwaite, Paul H. and Betkowska, Karolina(2009)'Bayes' theorem and diagnostic tests in neuropsychology: Interval estimates for post-test probabilities',The Clinical Neuropsychologist,23:4,624 — 644
To link to this Article: DOI: 10.1080/13854040802524229
URL: http://dx.doi.org/10.1080/13854040802524229

The Clinical Neuropsychologist, 23: 624–644, 2009
http://www.psypress.com/tcn
ISSN: 1385-4046 print/1744-4144 online
DOI: 10.1080/13854040802524229

Psychology Press
Taylor & Francis Group

BAYES' THEOREM AND DIAGNOSTIC TESTS IN NEUROPSYCHOLOGY: INTERVAL ESTIMATES FOR POST-TEST PROBABILITIES

John R. Crawford[1], Paul H. Garthwaite[2], and Karolina Betkowska[1]

[1]*School of Psychology, University of Aberdeen and* [2]*Department of Mathematics and Statistics, The Open University, Milton Keynes, UK*

Most neuropsychologists are aware that, given the specificity and sensitivity of a test and an estimate of the base rate of a disorder, Bayes' theorem can be used to provide a post-test probability for the presence of the disorder given a positive test result (and a post-test probability for the absence of a disorder given a negative result). However, in the standard application of Bayes' theorem the three quantities (sensitivity, specificity, and the base rate) are all treated as fixed, known quantities. This is very unrealistic as there may be considerable uncertainty over these quantities and therefore even greater uncertainty over the post-test probability. Methods of obtaining interval estimates on the specificity and sensitivity of a test are set out. In addition, drawing and extending upon work by Mossman and Berger (2001), a Monte Carlo method is used to obtain interval estimates for post-test probabilities. All the methods have been implemented in a computer program, which is described and made available (www.abdn.ac.uk/~psy086/dept/BayesPTP.htm). When objective data on the base rate are lacking (or have limited relevance to the case at hand) the program elicits opinion for the pre-test probability.

Keywords: Post-test probabilities; Bayesian methods; Diagnostic testing; Interval estimates; Quantitative methods.

INTRODUCTION

There are many circumstances in neuropsychology in which neuropsychological tests are used in an attempt to diagnose a condition of interest (COI). This is particularly the case with the dementias as there are few definitive biological markers with which to establish a diagnosis. A common strategy when evaluating neuropsychological tests is to compare the performance of a sample of patients with a COI to a healthy control sample using group statistics. If a highly significant difference is observed then it is often concluded that the test in question provides a useful means of identifying the COI. However, a diagnosis is applied to individual cases and so this form of validity data is limited in its utility (Smith, Ivnik, & Lucas, 2008). A group difference may be

Address correspondence to: Professor John R. Crawford, School of Psychology, College of Life Sciences and Medicine, King's College, University of Aberdeen, Aberdeen AB24 2UB, UK. E-mail: j.crawford@abdn.ac.uk

Accepted for publication:September 28, 2008. First published online: February 20, 2009.

highly significant but there may still be a very high degree of overlap between patients and controls, particularly if power is high (such that a relatively small difference between the groups will be statistically significant).

Reporting of effect sizes, although informative and important for other purposes (Becker, Knowlton, & Anderson, 2005; Henry & Crawford, 2004), is also insufficient. Cohen (1988) produced tables that can be used to estimate the degree of overlap between two samples from an effect size (such as Cohen's *d*) but (a) these assume that scores in both samples are normally distributed (this assumption will often be violated), and (b) such information does not provide a basis for selecting cutting scores.

SENSITIVITY AND SPECIFICITY OF TESTS

A much more satisfactory approach to examining the diagnostic utility of tests is to calculate their sensitivity and specificity. The sensitivity of a test is a proportion: The number of people with a COI who are correctly identified by the test is divided by the total number of people with the COI. Thus, if a test used to identify dementia of the Alzheimer type (DAT) correctly identifies 60 out of 80 cases of DAT, on average, then the sensitivity of the test is 0.75.

The specificity of a test is also a proportion: The number of people free of the COI who test negative is divided by the total number of people who are free of the COI. To continue the previous example, if for every 100 healthy controls who are administered the DAT test, 90 of them test negative on average, then the specificity of the test is 0.90.

There is always a trade-off between the sensitivity and specificity of a test. Suppose for example that using a particular cut-off for a test results in a disappointingly low sensitivity. The sensitivity can be increased by lowering the cut-off score so that more people with the COI will test positive. However, this will almost inevitably reduce the test's specificity: more people who are free of the COI will then also test positive. The choice of cut-off score will depend on the seriousness of the effects of a false negative versus a false positive. If a false negative has particularly serious consequences (e.g., potentially life-saving treatment will be withheld) then high specificity will be sacrificed in the interests of achieving high sensitivity. If the consequences of a false positive are particularly serious (e.g., a treatment with potentially serious side effects is administered unnecessarily), then high sensitivity will be sacrificed in favor of high specificity.

Fortunately it is becoming more common for studies of diagnostic tests in neuropsychology to report sensitivity and specificity information (Smith et al., 2008). Such information is central to an evidence-based approach to diagnostic assessment in the individual case. However, the sensitivity statistic tells us the probability of a positive test given that an individual has the COI, whereas, in practice, neuropsychologists need to know the probability of the COI given that an individual obtained a positive test (it is this inverse probability that is crucial). Similarly, specificity tells us the probability of a negative test, given that an individual is free of the COI. Again, the neuropsychologist is interested in the inverse probability; that is, the probability of the absence of the COI, given

a negative test. These post-test probabilities can be obtained using Bayes' theorem.

BAYES' THEOREM: OBTAINING POST-TEST PROBABILITIES FOR A COI

To obtain post-test probabilities for a COI requires the sensitivity and specificity of the test in question but also the base rate of the condition in the relevant population. Like sensitivity and specificity, the base rate, which is the *pre*-test probability of the COI, is expressed as a proportion. Thus if 5% of a relevant population (e.g., the general population of people over 60) suffer from a COI then the base rate is 0.05. Bayes' theorem combines this pre-test probability for the COI with the data on sensitivity and specificity to yield the *post*-test probability. The post-test probability of the COI given a *positive* test result (PTPP) is obtained as follows:

$$\text{PTPP} = \frac{\text{sensitivity} \times \text{base rate}}{(\text{sensitivity} \times \text{baserate}) + ([1 - \text{specificity}] \times [1 - \text{base rate}])}. \quad (1)$$

The post-test probability of the *absence* of a COI given a *negative* test result (PTPN) is obtained as follows:

$$\text{PTPN} = \frac{\text{specificity} \times [1 - \text{base rate}]}{(\text{specificity} \times [1 - \text{base rate}]) + ([1 - \text{sensitivity}] \times \text{base rate})}. \quad (2)$$

Take the post-test probability for a positive test as an example: Suppose that the base rate (pre-test probability) of a disorder in a particular setting is 0.40, and that the sensitivity and specificity of a test designed to diagnose this disorder are 0.80 and 0.85 respectively. Then, if an individual tests positive, the post-test probability of the disorder is 0.78. It is important to appreciate that the base rate exerts a strong influence on the post-test probability. Suppose that in the previous example the base rate of the disorder was 0.04 rather than 0.40: Then the post-test probability is only 0.18.

QUANTIFYING THE UNCERTAINTY OVER DIAGNOSTIC TEST STATISTICS AND BASE RATES

In the foregoing discussion of sensitivity, specificity, base rates, and post-test probabilities these statistics have all been treated as fixed, known quantities. However, these statistics must always be derived from sample data and so they are only *point estimates* of the true quantities. For example, the sensitivity of a test may have been estimated from a sample of 100 patients. In all probability this estimate would be different if another sample of 100 patients had been used to estimate it. That is, there will be considerable uncertainty over the true sensitivity of a test, particularly if the sample providing the data is modest in size. Exactly the same holds for specificity, and for estimates of the base rate of a COI.

Because of these unavoidable uncertainties it would be useful to supplement point estimates of these quantities with interval estimates: the width of these interval estimates will vary with the size of the sample providing the relevant data.

Provision of such interval estimates is relatively straightforward because all three quantities are proportions. Bayesian or classical (frequentist)[1] methods for inferences regarding a proportion can therefore be applied. In keeping with the general emphasis on a Bayesian approach in the present work, a Bayesian method will be employed. However, classical and Bayesian approaches to this problem usually give similar results in any case (Phillips, 1973).

The provision of interval estimates for sensitivity, specificity, and the base rate is in keeping with the contemporary emphasis on the use of interval estimates in psychology (American Psychological Association, 2001). Interval estimates serve the generally useful purpose of reminding us that test statistics are fallible, and the specific, concrete purpose of quantifying this fallibility (Crawford & Garthwaite, 2002, 2007, 2008).

QUANTIFYING THE UNCERTAINTY OVER POST-TEST PROBABILITIES

It was noted earlier that there is uncertainty over the true sensitivity and specificity of a test, and uncertainty over the true base rate of a disorder in the population. It follows from this that there will also be uncertainty over the post-test probabilities obtained from Bayes' theorem, as these three quantities are used in its computation. Although it is straightforward to obtain interval estimates for the former three quantities, there has been very little previous work on quantifying the uncertainty over the resultant post-test probabilities. This is unfortunate as it is this very quantity that is the crucial one for the neuropsychologist. However, Mossman and Berger (2001) have recently developed a method for obtaining interval estimates for post-test probabilities. Their approach will be adopted and extended in the present work. Details of the extensions to their method are provided in a later section but a brief overview of these is provided as part of the summary of aims stated below.

SUMMARY OF AIMS

The aims of the present study can be summarized as follows:

(1) To set out methods for obtaining interval estimates for the sensitivity and specificity of tests, thereby encouraging researchers to supplement point estimates of these quantities with interval estimates (the same methods can be used to obtain interval estimates for base rates)
(2) To implement and extend upon methods set out by Mossman and Berger (2001) for obtaining interval estimates for post-test probabilities in the individual case. Specifically, (a) to provide interval estimates for post-test probabilities of a COI following a positive or negative test (Mossman and Berger only explicitly provided intervals for post-test probabilities following a positive test result); (b) to offer neuropsychologists the choice of obtaining one-sided lower or upper intervals on post-test probabilities, as well as the

[1]The term frequentist is used to denote a view of probability as the long-run expected frequency of occurrence.

two-sided intervals presented in Mossman and Berger; (c) to allow neuropsychologists to obtain intervals on post-test probabilities based on subjective opinion concerning the base rate of the COI; and (d) to help neuropsychologists assimilate the results of a Bayesian analysis by providing a visual representation of the pre-test and post-test probability distributions.

(3) To implement all of the foregoing methods in a single computer program so that neuropsychologists have a fast, convenient, reliable, and flexible means of applying the methods in practice.

(4) To briefly examine the effects of varying the sensitivity, specificity, and the base rate, and the sample sizes providing these data, on the post-test probabilities for positive and negative tests. The aim here is to provide neuropsychologists with an insight into the effects of these factors on the degree of uncertainty attached to post-test probabilities.

OBTAINING INTERVAL ESTIMATES FOR THE SPECIFICITY AND SENSITIVITY OF A DIAGNOSTIC TEST

Sensitivity and specificity are simple binomial proportions and therefore methods of obtaining interval estimates of a proportion can be used to set limits on these quantities. There is a wide array of Bayesian and classical (frequentist) statistical methods available for obtaining interval estimates of a binomial proportion. However, typically these intervals show a high degree of convergence, except when the proportions are very large, or very small, or sample sizes are small. As noted, in keeping with the general emphasis on Bayesian methods in the present work, a Bayesian method is used to provide interval estimates of these quantities.

In Bayesian statistics a *prior* distribution is used to convey any information about model parameters that was available before the sample data were gathered. This is combined with the information supplied by the data, which is contained in the *likelihood*, to yield a *posterior* distribution. Inferences and estimates are based on the posterior distribution. In the most common Bayesian approach to obtaining interval estimates of a proportion, the prior distribution for the proportion is represented by a beta distribution. The treatment offered here is designed to provide a largely non-mathematical overview of the approach (additional information on beta distributions is introduced, as required, in later sections).

A beta distribution has two parameters, denoted a and b, and the beta distribution is denoted beta(a, b). For the present problem the prior distribution is a non-informative prior distribution. That is, we do not have prior knowledge of the sensitivity or specificity of the test ahead of obtaining the data from the appropriate sample (or, although we may have a prior opinion over these quantities, we do not wish to incorporate this opinion into the analysis).

Three beta distributions have been used to represent a lack of prior knowledge or opinion, the Jeffrey's prior: beta$(0.5, 0.5)$; a uniform prior: beta$(1, 1)$; and a beta$(0, 0)$ prior. In the present case we use a beta$(0, 0)$ prior—this prior has the advantage of simplicity, it also represents *no* prior data (as opposed to a little in the case of the other two priors), and has the attractive feature that its use ensures that the Bayesian *posterior* point estimate of sensitivity and specificity is then identical to

the value that would be obtained by the usual formulas (i.e., if 90 cases out of 100 with a COI test positive then the Bayesian point estimate is 0.90). The choice among these three priors makes little practical difference to the analysis: The influence of the prior on the posterior distribution is swamped by the influence of the data unless the sample size is very small, or if virtually all the test results are in one direction (either positive or negative).[2]

As in any Bayesian analysis, the prior is combined with the data to yield the posterior distribution. For the present problem, conveniently, the posterior is another beta distribution, thereby avoiding the need for numerical integration (when prior and posterior distributions are of the same distributional form they are referred to as conjugate distributions). When an interval estimate for the sensitivity of a test is required, the posterior distribution is obtained by denoting the number of participants who have the COI and tested positive as "successes" (s) and the number who have the COI and tested negative as "failures" (f). Then the posterior distribution is a beta(s,f) distribution.

To obtain a 95% two-sided, equal-tailed interval estimate of the sensitivity of the test simply requires finding the quantiles corresponding to the 0.025 and 0.975 percentile points of this beta distribution. These quantiles can be obtained from tables (e.g., Phillips, 1973) or, generally more accurately (as the need for any interpolation is avoided), by a computer algorithm. The same procedure is used to obtain an interval estimate of the specificity of the test except that the number of participants who do not have the COI and tested negative are denoted as "successes" and those who do not have the COI and tested positive as "failures." Bayesians use the term "credible interval" rather than "confidence interval" for these interval estimates.

Table 1 illustrates the effects of varying the sample size providing the data on the 95% credible intervals for sensitivity (or specificity, as values matching those for sensitivity will have identical credible intervals): In these examples the sample size was varied from 10 to 1000 and the sample point estimate of sensitivity was set at one of two values, 0.70 or 0.90. It can be seen that the uncertainty over the true sensitivity of a test is considerable when modestly sized samples provide the data: For example, the 95% credible interval for an estimated sensitivity of 0.70, based on a sample size of 50, ranges from 0.567 to 0.817. Table 1 provides a vivid illustration of the dangers of limiting attention to point estimates of the sensitivity and specificity of neuropsychological tests and underlines the advantages of using tests that have large validation samples.

Finally, it was noted earlier that the interval estimates on proportions (sensitivity and specificity in the present case) obtained using Bayesian methods usually exhibit convergence with classical (frequentist) methods. This convergence is reassuring, regardless of whether one is classical, Bayesian, or eclectic in orientation. For example, the mid-p variant of the Clopper-Pearson method

[2]The exception to our use of a beta(0, 0) prior occurs at the extremes; that is when there are 0 successes or failures in the sample providing the data on sensitivity or specificity (the posterior distribution cannot then be formed). This will arise rarely in practice but when it does one possibility is to substitute the Jeffrey's prior, i.e., beta(0.5, 0.5). This rule is implemented in the computer program that accompanies this paper. An alternative would be to use expert opinion, an approach we are currently exploring.

Table 1 Illustration of the uncertainty over a test's sensitivity (or specificity) as a function of the size of the sample providing the data

	Sample estimates of sensitivity (or specificity)	
Sample size	0.70	0.90
10	(0.400, 0.925)	(0.719, 0.997)
50	(0.567, 0.817)	(0.804, 0.966)
100	(0.607, 0.785)	(0.834, 0.950)
200	(0.635, 0.761)	(0.855, 0.938)
500	(0.659, 0.739)	(0.872, 0.925)
1000	(0.671, 0.728)	(0.881, 0.918)

The intervals are 95% Bayesian credible intervals.

(Clopper & Pearson, 1934) is regarded by many as the most satisfactory of the classical methods. As an illustration of the convergence between the methods, the 95% two-sided Clopper-Pearson mid-*p* confidence interval for a sample sensitivity of 0.70, based on a sample size of 50, is (0.563, 0.814); which is very close to the corresponding Bayesian 95% credible interval recorded earlier.

Although the numerical results provided by these methods converge, the interpretations placed upon them differ. As Antelman (1997) notes, the classical (frequentist) conception of a *confidence* interval is that, "It is one interval generated by a procedure that will give correct intervals 95% of the time. Whether or not the one (and only) interval you happened to get is correct or not is unknown" (p. 375). Thus, in the present context, the frequentist interpretation is as follows: "If we could compute confidence intervals for a large number of samples collected in the same way as the sample in hand, about 95% of these intervals would contain the true sensitivity of the test."

In contrast, the Bayesian interpretation of the corresponding *credible* interval is that "there is a 95% probability that the true sensitivity of the test lies within the stated interval". This latter statement is not only less convoluted but, we suggest, it also captures what a neuropsychologist would wish to infer from an interval estimate. Indeed, as Howell (2002) observes, most psychologists who use frequentist confidence limits probably construe these in what are essentially Bayesian terms.

OBTAINING INTERVAL ESTIMATES OF THE BASE RATE

As previously discussed, to apply Bayes' theorem to diagnostic tests in the individual case requires an estimate of the base rate. To our knowledge, the base rate has also been treated as a fixed, known quantity in all previous treatments of this topic in neuropsychology but this is no more realistic (indeed is even less realistic) than treating sensitivity or specificity as fixed, known quantities. When the base rate data are obtained from a prevalence study, then it is possible to obtain a credible interval on the proportion using exactly the same methods as those just set out for interval estimates of the sensitivity and specificity of the test: The number with the COI are denoted as "successes" and those free of the COI as "failures" and thus the posterior beta distribution is beta(s,f). All that is required is the total

number of participants N in the prevalence study and the number with the disorder ($f = N - s$). To illustrate: if, in a prevalence study of dementia of the Alzheimer type (DAT), 20 out of 300 participants were diagnosed as having DAT (i.e., the base rate is 0.067), then the posterior beta distribution is beta(20, 280) and the 95% credible interval on the proportion is (0.041, 0.097). In contrast, if the prevalence study consisted of only 30 participants, two of whom were DAT cases (i.e., the point estimate is 0.067, the same as the previous example), then the posterior beta distribution is beta(2, 28) and the 95% interval is (0.008, 0.178).

OBTAINING INTERVAL ESTIMATES FOR POST-TEST PROBABILITIES

Although a few studies in neuropsychology have supplemented point estimates of the sensitivity and specificity of neuropsychological tests with interval estimates for these quantities, none has presented a means of obtaining credible intervals on post-test probabilities. This is unfortunate because, as noted, it is the post-test probability for the presence or absence of a COI that is of primary relevance to the neuropsychologists when they engage in diagnostic decision making.

Of course most neuropsychologists will be aware in a general sense that the degree of confidence to be placed in a point estimate of a post-test probability must vary with the size of the samples used to obtain estimates of the sensitivity and specificity of the test and the estimate of the base rate (i.e., the pre-test probability). Moreover, their ability to evaluate the confidence that should be placed in the point estimate of the post-test probability will be aided if interval estimates of these three quantities have been provided. Nevertheless, it is not at all realistic to suppose that a neuropsychologist could accurately integrate and weight this latter information. Thus an interval estimate for post-test probabilities is even more important from a practical point of view than interval estimates for sensitivity, specificity, and the base rate. The next section tackles this key problem of obtaining credible intervals for post-test probabilities.

CREDIBLE INTERVALS FOR POST-TEST PROBABILITIES

One factor that simplifies the problem is that the three quantities that determine post-test probabilities are independent of each other. That is, the sensitivity of a test is estimated from a sample of participants with the COI, and specificity from a sample who are free of the COI: The base rate is estimated from a third sample such as that employed in a prevalence study of the disorder in which the presence or absence of the disorder is established by some gold standard or by following up patients to establish their diagnostic status.

Mossman and Berger (2001) have recently tackled this important problem using Bayesian Monte Carlo methods. In the first stage of their procedure beta distributions are used to specify non-informative *prior* distributions for the sensitivity, specificity, and base rate. These are then combined with the observed data for each of these quantities to obtain *posterior* distributions. (Note that, in the present case, these posterior beta distributions have already been specified and used to obtain credible intervals for the three quantities involved.) In the second stage of the procedure (the Monte Carlo stage) a large number of random draws are then

made from each of these three posterior distributions (we opt for 100,000 draws in the present application) and these observations are passed through Bayes' theorem, using formula (1). This provides a *distribution* of post-test probabilities given a positive test (rather than the scalar quantity obtained in the standard application of Bayes' theorem to diagnostic tests). Although not dealt with explicitly by Mossman and Berger, it is just as straightforward to obtain a distribution of post-test probabilities for the absence of a COI given a negative test result: Formula (2) is simply substituted for formula (1).

Mossman and Berger used Jeffrey's priors, i.e., beta (0.5, 0.5), as prior distributions for sensitivity, specificity, and the base rate. Such priors contain a small amount of information: if, for example, 40 out of 50 cases give positive test results, then with a beta (0.5, 0.5) prior the posterior mean for sensitivity is 40.5/51, rather than 40/50. For this reason we prefer to use beta (0, 0) as the prior, which does give 40/50 as the posterior mean. Of course, as this example illustrates, the results obtained from beta (0.5, 0.5) and beta (0, 0) priors will normally be very similar.

As stated, the mean value of the 100,000 probabilities generated by the Monte Carlo stage of the analysis is the Bayesian point estimate of the post-test probability of having the disorder given a positive test (or point estimate of the post-test probability of the absence of a COI given a negative test). The post-test probabilities can also be sorted to provide 95% credible intervals on the post-test probability. That is, if a two-sided credible interval is required (and assuming 100,000 random draws have been made) the 2500th smallest and 2500th largest observations from the sorted posterior distribution provide the two-sided 95% credible interval.

To illustrate: Suppose that the base rate of a condition is 0.10 and that this base rate was obtained from a prevalence study of 100 people. Suppose also that the sensitivity of a test is estimated at 0.75, based on a sample of 80 people with the disorder (true positives = 60) and that the specificity is estimated at 0.90 based on a sample of 150 people who were free of the disorder (true negatives = 135). Finally, suppose that an individual is tested and obtains a positive result. Then the post-test probability of the disorder is 0.45 and the 95% two-sided credible interval is (0.26, 0.65). Alternatively suppose an individual obtains a negative result. Then the post-test probability of being free of the disorder is 0.97 and the 95% two-sided credible interval is (0.94, 0.99). A later section will briefly examine the effects on the widths of the credible intervals as a function of sensitivity, specificity, and the base rate, as well as the sample sizes used to provide these statistics.

Note that the focus in the present paper is primarily on obtaining *interval* estimates for post-test probabilities. However, it is worth noting that using the Bayesian posterior mean as the *point* estimate of the post-test probability is preferable to the standard practice of obtaining a point estimate by plugging sample estimates of the sensitivity, specificity, and the base rate into Equations (1) and (2). The two point estimates will usually be similar but not identical. Normally the advantage of the latter "plug-in" point estimate would be its convenience. However, this does not apply in the present case because the computer program accompanying this paper provides Bayesian posterior means and does this more reliably and in less time than using a calculator or spreadsheet to obtain the latter plug-in estimate.

ONE-SIDED VERSUS TWO-SIDED CREDIBLE INTERVALS FOR POST-TEST PROBABILITIES

Mossman and Berger (2001) focused exclusively on two-sided intervals for post-test probabilities. However, there will be circumstances in which a neuropsychologist may be interested in obtaining a one-sided interval. For example, a neuropsychologist may be interested in the possibility that the post-test probability following a positive test is lower than the point estimate suggests, but may be relatively uninterested in whether it is higher. That is, if the lower one-sided endpoint is still high (for example, if the neuropsychologist can be 95% confident that the post-test probability is not below 0.9) then they can have a reasonably high degree of confidence that the patient does have the COI. Similarly, if an individual tests negative then the neuropsychologist may also be primarily interested in the lower endpoint (i.e., they want to determine how confident they can be that the person does not have the COI).

It is straightforward to extend Mossman and Berger's approach to generate one-sided intervals. All that is required is to specify the desired level of confidence (95%) and then "spend" all of this in setting a one-sided endpoint. For example, again assuming 100,000 draws, the 95% lower endpoint is found by finding the 5000th lowest observation from among the sorted posterior distribution. Similarly, a one-sided *upper* endpoint is obtained by finding the 5000th *highest* observation.

USING SUBJECTIVE OPINION TO ESTIMATE THE BASE RATE

Mossman and Berger's (2001) approach assumes that the neuropsychologist has objective data from a prevalence study to estimate the pre-test probability (i.e., the base rate). However, formal data of this type will often be unavailable, or may not be appropriate for the case at hand. For example, suppose a neuropsychologist is examining a 65-year-old man referred to a memory clinic because of suspected dementia of the Alzheimer type (DAT). The neuropsychologist may have formal data on the prevalence of DAT in the general population of persons over 60. However, they are also aware that the prevalence rate for DAT in the population of people over 60 *and* referred to the memory clinic will be considerably higher. Furthermore, if the patient has a family history of early DAT, then the rate may be higher still. Thus it would be useful if the approach set out here could be broadened to allow calculation of interval estimates for post-test probabilities using subjective opinion over the pre-test probability.

A variety of methods of eliciting prior opinion over binomial quantities have been proposed; see Garthwaite, Kadane, and O'Hagan (2005) for a review. Although the methods of eliciting initial opinion can differ substantially, most methods share the common feature that they are iterative: After the initial attempt to quantify opinion, the user is given feedback (in the form of summary statistics, a graphical representation of their opinion, or both) and the process is repeated until the user is satisfied that their opinion has been captured adequately.

In the present case subjective opinion is elicited by first asking the user to provide a point estimate of the quantity and the standard deviation of this estimate. The point estimate provided is taken as the mode of the beta distribution. In this case

the quantity is the estimate of the pre-test probability of the patent having the COI and is expressed as a percentage (proportions are often used rather than percentages but, in the view of the present authors, percentages are easier for users to work with, particularly when base rates are low). Thus, for example, if a neuropsychologist estimated that the most likely value for the base rate for a COI in people of the same age, background, and medical circumstances as their patient was 5%, then the point estimate is 5. By definition, the most likely value is the mode. The user's level of certainty/uncertainty over their point estimate is quantified by requiring them to specify an accompanying standard deviation (suppose for illustrative purposes that a value of 3.9 is specified). These two bits of information are all that are required to specify a beta distribution to represent their prior opinion.

The parameters of the beta distribution are easily obtained. The mode of a beta distribution is:

$$m = \frac{a-1}{a+b-2}, \tag{3}$$

and the standard deviation is:

$$\sigma = \sqrt{\frac{ab}{(a+b+1)(a+b)^2}}. \tag{4}$$

Solving these equations using a simple search procedure gives the values of a and b (the computer program accompanying this paper automates this process but for completeness a suitable search procedure is set out in the Appendix). Thus, in the foregoing example (where the mode was 5 and standard deviation was 3.9), the values for the two parameters of the beta distribution representing the neuropsychologist's initial prior opinion are $a = 3.02$ and $b = 39.45$; that is, the distribution is beta(3.02, 39.45).

This beta distribution can be represented graphically and the area lying between designated percentile points can be shaded to differentiate it from the areas below and above the lower and upper points respectively. In the present application the 25th and 75th percentile points are used; therefore the area between these points exactly matches the areas outwith them. The quantiles corresponding to these percentile points are also presented so that the user has two sources of information to consult when assessing the adequacy with which the distribution has captured their opinion. In the present example, these quantiles are 4.2% and 9.3%. The computer program accompanying this paper implements this method and a screen capture showing the graphed beta distribution for the current example is presented as Figure 1.

If the beta distribution has successfully captured a user's opinion over the base rate then they would be comfortable that there is a 50% chance that the true base rate is between these percentile points: In the present example they should be comfortable that there is a 50% chance that the true base rate is between 4.2% and 9.3%. In other words they would be indifferent regarding a bet as to whether the true base rate lies between these points or outwith them. If on inspection, they think the probability that the base rate between these points is appreciably greater than 50%, then the distribution has underestimated

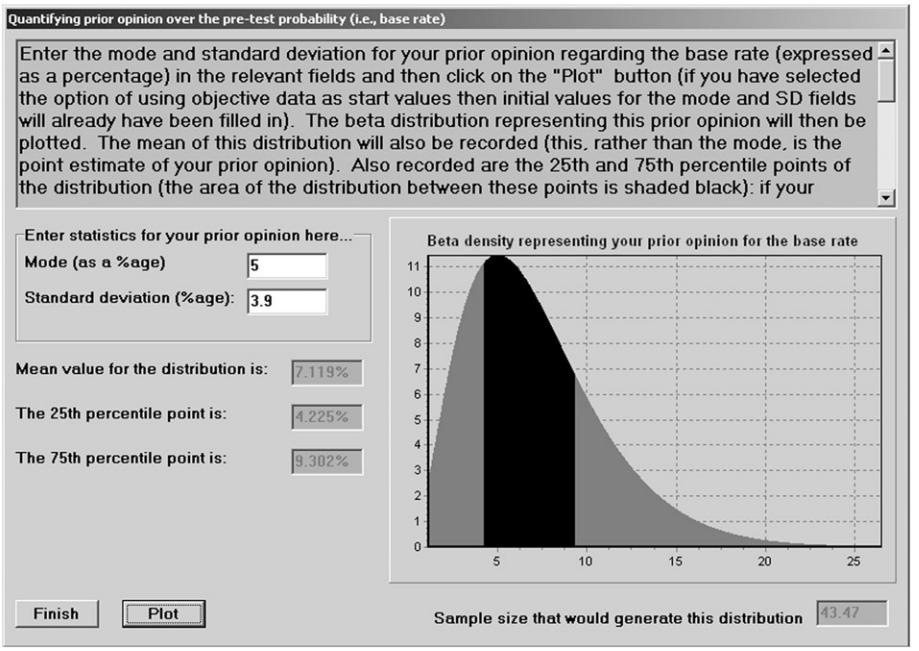

Figure 1 Illustration of the use of beta distributions to elicit prior opinion over the base rate of a condition of interest. The figure is a screen capture from the computer program that accompanies this paper and follows the example provided in the text in which the user's estimated median for the base rate for a COI was 5% with a standard deviation of 3.9%.

their level of confidence and they would repeat the process after decreasing the standard deviation. Alternatively, if they believed the chances are appreciably below 50%, they would increase the standard deviation. The process is therefore an iterative one: the user continues until they consider the beta distribution captures their opinion adequately. Once a user has calibrated their prior opinion to their satisfaction the procedure is identical to that used when objective data are available for the base rate.

Psychological research has found that people are generally quite poor at quantifying their opinions as standard deviations and are much better at relating their opinions to quantiles (Hogarth, 1975; Lichtenstein, Fischhoff, & Phillips, 1982). The reader may therefore wonder why opinion was not elicited by directly asking the user to specify quartiles instead of a standard deviation. The explanation is that a beta distribution has only two parameters so, after fixing the mode, one quartile cannot be changed without also changing the other. Modifying the quartiles via an assessment of the standard deviation is one way of simultaneously changing both quartiles consistently. Where some compromise is needed between the values of these quartiles (if the user feels one quartile is too near the mode while the other is too far from it), the conservative approach is to choose a value for the standard deviation that is too large rather than too small.

USING OBJECTIVE PREVALENCE DATA AS A STARTING POINT FOR QUANTIFYING SUBJECTIVE OPINION OVER THE BASE RATE

It will be quite common for the neuropsychologist to have some quantitative data on the base rate for a condition of interest, but to also be aware that it needs to be modified in the light of an individual patient's history and presentation. For example, if the objective estimate of the pre-test probability comes from a study of the prevalence of a COI in the general (elderly) adult population, it is liable to seriously underestimate the pre-test probability for a patient who has been referred to a memory clinic and has a family history of the COI.

In such circumstances it may be useful to use the prevalence data as a starting point when calibrating subjective opinion. That is, the parameters of the initial beta distribution can be determined using the prevalence data (N and the number of "successes") and subsequently modified by the neuropsychologist to factor in the individual circumstances of the patient. For example, suppose that a study, based on a sample of 400, reports a prevalence rate of 5% (thus 20 individuals had the COI). Then, using formula (3) the mode for the beta distribution representing these data is $19/398 = 4.77$ and the standard deviation is 1.09. The neuropsychologist would increase the mode from its start value of 4.77 and also increase the standard deviation to reflect the increased uncertainty stemming from the use of subjective opinion. The iterative elicitation process outlined earlier would then be followed until the neuropsychologist was satisfied that their opinion had been adequately represented. The computer program accompanying this paper offers this option (see later section).

The process just described is very flexible. For example, in the previous example the neuropsychologist had objective base rate information on the COI for persons of the same age as their case but wanted to modify this to account for the fact that the case was referred to a memory clinic (the prevalence data were for the general population). The situation could be reversed: a neuropsychologist may have objective base rate data for the COI for referrals to a memory clinic but the data are not stratified by age. If age is a significant risk factor for the COI and the case is very elderly, then the neuropsychologist may wish to modify the base rate data to take account of this. As a further example, a neuropsychologist could have objective base rate data on a COI for overall referrals to their service but may also be aware that the risk of the COI will vary with the source and circumstances of referral.

A prior distribution that is beta(a, b) can be interpreted as equivalent to having seen a people with the COI in a sample of $a + b$ people. Hence, $a + b$ represents the sample size of the amount of information on which a neuropsychologist's opinion is based. This equivalence might in principle be used as part of the process of selecting a prior distribution to represent opinion (Bolstad, 2007) but, in practice, people tend to substantially overestimate the sample size that corresponds to their knowledge (Schaefer & Borcherding, 1973). This implies that users should revise their assessments if they assess a beta distribution and then find that $a + b$ is *larger* than corresponds to their level of confidence, but this will rarely be the case. However, when forming a prior distribution that takes account of objective base rate data from a similar (but slightly different) population, the value of $a + b$ should relate to the sample size that gave the base rate data for that population.

For example, if the base rate data are from a population that is very similar to the population of current interest, then $a+b$ should be only slightly smaller than that sample size, while it should be significantly smaller if differences between the populations are greater.

Comprehensive, objective data on a COI are of immense value to the neuropsychologist and is clearly preferable to the use of subjective opinion. However, the reality is that such data are rarely available, given that they would need to be stratified by a number of variables such as age, possibly gender depending on the COI, presence of risk factors, referring agency, referring physician within agency, circumstances of referral (i.e., is the neuropsychologist the first port of call or being used to "confirm" a diagnosis etc.), to name but a few. Therefore, there is clearly a role for quantifying a neuropsychologist's specific knowledge in order to modify the estimate of the pre-test probability.

Some neuropsychologists may be uneasy about using subjective opinion to specify or modify a distribution for the pre-test probability. However, it is simply a formalization and extension of the process a neuropsychologist undertakes in a more casual fashion. Furthermore, by offering neuropsychologists feedback (in the form of the graphical display and percentile points etc.), and providing the opportunity to recalibrate in the light of this feedback, the proposed method should help them clarify their opinion (Garthwaite et al., 2005).

Our view is that, if the neuropsychologist is confident that the available objective base rate data are not appropriate for the case in hand, then they should use their knowledge and experience to modify the base rate estimate: if they are confident only that (say) the base rate is too low but less clear as to how high it may be then this can be modeled by specifying a relatively large standard deviation for the beta distribution.

EFFECTS ON CREDIBLE INTERVALS FOR POST-TEST PROBABILITIES OF VARYING SENSITIVITY, SPECIFICITY, AND BASE RATES, AND THE SAMPLE SIZES PROVIDING THESE DATA

In diagnostic testing the distribution of post-test probabilities is determined by a complex combination of factors: the sensitivity and specificity of the test, the base rate of the COI, and the precision with which each of these three quantities have been measured. A detailed elaboration of the effects of these factors is beyond the scope of the present paper. In this section we limit ourselves to illustrating some basic features, starting with the effects on interval estimates of the sample sizes that provide the data for sensitivity, specificity, and the base rate. Table 2 presents 95% interval estimates when the sample sizes for all three quantities are varied from 50 to 1000. The sensitivity, specificity, and base rate in these examples are all held constant at 0.90, 0.90, and 0.30 respectively so that the point estimate of the post-test probability is the same (0.794) in all examples. It can be seen from Table 2 that the limits narrow considerably in going from modest sample sizes of 50 to large sample sizes (i.e., 500). However, it can also be seen that there are diminishing returns when sample sizes are very large: the differences in the width of the intervals for sample sizes of 500 versus 1000 is very modest.

Table 2 Illustration of the effects of differing sample size on the 95% interval estimates for post-test probabilities

Sample size	Lower limit	Upper limit
50	0.60	0.93
100	0.66	0.90
200	0.70	0.87
500	0.74	0.84
1000	0.76	0.83

The sample estimates of sensitivity, specificity, and base rate are fixed at 0.90, 0.90, and 0.30 respectively in this example, hence the point estimate of the post-test probability given a positive test result is 0.794 in all cases.

Sample size is, of course, far from being the only factor that influences the degree of uncertainty over the post-test probability. For example, suppose that the point estimate of the post-test probability of a COI following a positive test is 0.79 in two cases (Case A and Case B) and that the sample sizes used to estimate sensitivity, specificity, and the base rate were identical in both cases. Despite these commonalities, the uncertainty over the post-test probability could still be vastly different for the two cases. If, as is currently the norm in neuropsychological practice, only the point estimate of the post-test probability is calculated using Bayes' theorem, i.e., using Equation (1) or (2), these differing levels of uncertainty are entirely hidden from view.

To continue with the example, suppose that data on the sensitivity, specificity, and the base rate were all obtained from samples of 100 participants in both cases. A point estimate of the post-test probability of 0.79 could be obtained if the samples gave a sensitivity of 0.71, specificity of 0.85 and base rate of 0.44 (Case A), but also if they gave a sensitivity of 0.98, specificity of 0.98, and base rate of 0.07 (Case B). In the former case the interval estimate accompanying the point estimate of 0.79 is (0.67, 0.88), whereas in the latter case the interval is much wider (0.49, 0.97). Figure 2a plots the distributions of the pre-test and post-test probabilities for Case A, and Figure 2b plots the distributions of the pre-test and post-test probabilities for Case B. (Distributions of the pre-test probabilities are given by the base rates.) These plots vividly demonstrate the different levels of uncertainty in the two cases.

Case B illustrates that, when the base rate is low, the post-test probability can still be modest despite high sensitivity and specificity. Those with experience of estimating post-test probabilities using the standard scalar version of Bayes' theorem will be familiar with this scenario. However, the present example also illustrates that not only will the point estimate of the post-test probability be modest, there will also be a high degree of uncertainty attached to it.

Neuropsychologists will also be aware of the converse situation: The point estimate of the post-test probability for the *absence* of a COI following a negative test is typically very high when the base rate is low. In addition, however, it is also the case that there will be very little uncertainty attached to this point estimate. Take the sensitivity, specificity, and base rate given for Case B, for example, and suppose

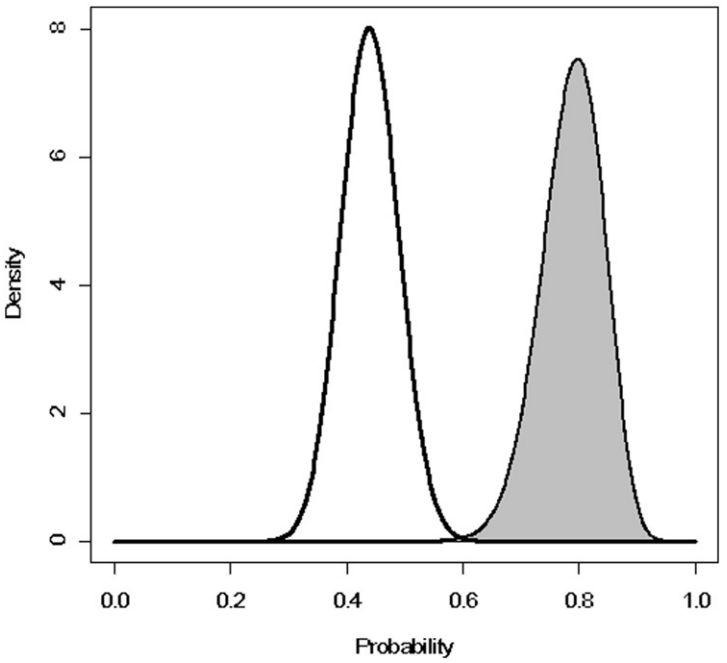

Figure 2a Plot of the pre-test (unshaded) and post-test (shaded) probability distributions for Case A.

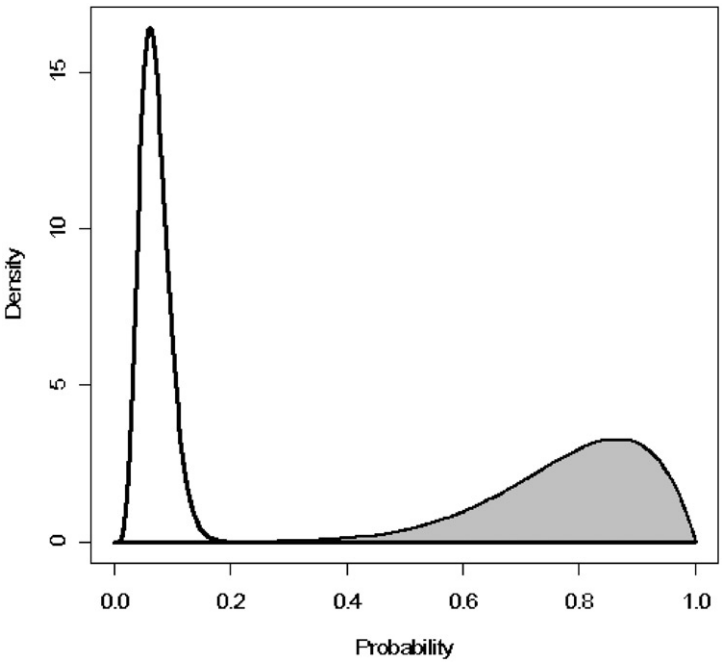

Figure 2b Plot of the pre-test (unshaded) and post-test (shaded) probability distributions for Case B.

that the case tests negative rather than positive. Then the point estimate of the post-test probability (0.998) for the absence of the COI is very high and, moreover, the 95% interval estimate is very narrow (0.995 to 1.000). This is an extreme example but the principle will hold, albeit in a slightly attenuated fashion, when values for sensitivity and specificity are not as high as the present ones. For example, if the sensitivity and specificity were both 0.80, then the point estimate is 0.981 and the 95% interval is still very narrow (0.962, 0.993).

This section only touches on the factors influencing the degree of uncertainty over post-test probabilities. The program that accompanies this paper, although primarily designed for direct use in neuropsychological practice, can also serve as a convenient means of exploring these factors. Details of the program are provided in the next section.

COMPUTER PROGRAM FOR QUANTIFYING THE UNCERTAINTIES ATTACHED TO POST-TEST PROBABILITIES

The method employed to provide credible intervals on post-test probabilities uses simulation and therefore requires the use of a computer (it is also much more convenient to use a program to obtain credible intervals for sensitivity, specificity, and the base rate). It would be relatively simple to implement the methods in a generic statistics package or spreadsheet package. However some technical ability would be required, particularly in order to implement the methods for calibrating subjective opinion over the base rate. In view of this a computer program for PCs was written (in the Delphi programming language) to implement all of the methods set out in the present paper.

The program, Post_Test_Probabilities.exe, has three main options. The first option is selected when a neuropsychologist has *objective* data on the base rate of a COI from a prevalence study. The second option is for use when a neuropsychologist does not have access to formal objective data from a prevalence study. It therefore elicits subjective opinion over the base rate and offers feedback to assist in calibrating this opinion: As illustrated in Figure 1, it graphs the beta distribution, records the distribution's mean and 25th and 75th percentile points, and records the sample size that would generate it. The third option is intended to be used to elicit subjective opinion when a neuropsychologist wishes to use objective data as a starting point A further set of options allows the user to select between generating two-sided 95% interval estimates of the post-test probability or one-sided 95% lower or upper interval estimates.

The inputs required for the program are the raw data on sensitivity and specificity for the test in question (and raw data on the base rate if objective data are employed). The raw data for sensitivity are prompted for in the form of the sample size providing the data and the number of true positives. Thus if, in a sample of 120 people with the COI, 90 tested positive, then 120 and 90 are entered. If the data available to the neuropsychologist are in the form of the point estimate of the test's sensitivity (0.75 in the present example) and sample size, then the number of true positives can be obtained simply by multiplying the sensitivity by the sample size. The equivalent data are prompted for in the case of specificity: the size of the sample and the number of true negatives (again if the available data are in the form of the

test's specificity, this value should be multiplied by the sample size to obtain the true negatives).

The output of the program varies according to which options have been selected but in all cases it records point and interval estimates for sensitivity and specificity. Most importantly, it reports the point and interval estimates for the post-test probability of the presence of a COI following a positive test, or the post-test probability of the absence of a COI following a negative test.

The program also reports the point and interval estimates for the absence of the COI following a positive test (and the same information for the presence of a COI following a negative test). Most treatments of Bayes' theorem in diagnostic testing limit attention to the probability of a COI following a positive test (PTPPT) and the probability of the absence of a COI following a negative test (PTPNT). This is understandable since these probabilities are where the primary interest will normally lie. Moreover, the probability of the absence of a COI following a positive test is simply 1 – PTPPT (similarly the probability of the presence of a COI following a negative test is 1 – PTPNT). Thus if, following a positive test, the post-test probability of the COI is 0.90, then the probability of the absence of the COI is 0.10. However, these latter probabilities can still be useful in some circumstances and, in the present case, the point estimates for the post-test probabilities are supplemented with interval estimates. Therefore, three calculations would need to be performed to get these latter point and interval estimates (admittedly these calculations are very simple but the possibility of clerical error remains). In view of the foregoing considerations we considered it worth reporting these additional quantities.

If either of the two subjective opinion options have been selected, the output records details of the beta distribution that captured this opinion (mean, median, standard deviation, and the distribution's parameters); otherwise it provides point and interval estimates for the base rate based on the objective prevalence data entered by the user. The program output can be viewed on screen, printed, and saved to a file. The program also has the option of adding User Notes (e.g., to keep a record of details of the patient or the source of the prevalence data etc.); these notes are reproduced in the output from the program.

The program also plots the distributions of the pre-test and post-test probabilities. This plot resembles those featured in the present paper (Figures 2a and 2b) and is an informative and effective way of conveying the results of the analysis to the user. A compiled version of the program (together with a zipped version) can be downloaded from the following web page: www.abdn.ac.uk/~psy086/dept/BayesPTP.htm.

If the pre-test probability is estimated using subjective opinion, it is advisable to use the program to quantify this opinion *ahead* of testing the patient. Otherwise there is the risk that knowing the outcome of testing will influence what is firmly required to be the user's *prior* opinion concerning the probability that the patient has the COI.

The program was written with the primary aim of providing neuropsychologists with a tool to improve the rigor with which they approach diagnostic decisions with individual cases. However, as noted, it can also serve the more general purpose of helping neuropsychologists develop a fuller understanding of the

factors influencing post-test probabilities. A previous section attempted to illustrate the influence of some of these factors. However, we believe that a deeper understanding can be gained by a direct, hands-on approach in which the neuropsychologist explores the factors influencing these probabilities for themselves by varying the various parameters (i.e., sensitivity, specificity, and pre-test probabilities together with different combinations of sample sizes). The program's ease of use, combined with the speed with which the calculations are performed, means that it is well suited to this purpose. The plot of the pre-test and post-test probability distributions is a particularly effective aid to understanding. The fuller appreciation gained by engaging in this learning process will lead to more informed evaluations of the results obtained for actual cases.

Finally, the emphasis of the present paper has been on quantifying the uncertainty over post-test probabilities because it is this quantity that is at the heart of diagnostic decision making. However, as was noted, researchers who are either developing new neuropsychological tests or evaluating existing tests should be encouraged to accompany point estimates of the sensitivity and specificity of the tests in question with interval estimates of these quantities. The program developed here provides a convenient means of obtaining these interval estimates. That is, although it was designed for use with individual cases, it provides point and interval estimates of the sensitivity and specificity of the tests as part of its output. When used for this latter purpose (i.e., when used to obtain intervals estimates for a research report on a test), arbitrary base rate data should be entered and all output, other than the 95% interval estimates for sensitivity and specificity, can simply be ignored

CLOSING REMARKS

Diagnostic testing in neuropsychology is an inherently uncertain process. When the standard version of Bayes' theorem is applied to neuropsychological test results, the post-test probabilities of a COI (or its absence) rarely approach 0 or 1. This may be sufficiently unsettling in itself for many neuropsychologists without having to also acknowledge to themselves that the uncertainty surrounding these point estimates can be considerable. Crucially, of course, ignoring this latter uncertainty does not make it go away: It is better to acknowledge its existence and take solace in the fact that its effects can at least be quantified.

The advantages of incorporating interval estimates into the decision making process are perhaps best appreciated by reminding oneself of the limitations inherent in the simple point estimates of post-test probabilities: They are blind as to whether their inputs come from very modest or very large samples. Moreover, they are silent on the fact that the uncertainty attached to two identical post-test probabilities (say, point estimates *A* and *B*), even when based on the same sample sizes, may be very different because they were arrived at through different combinations of values for the base rate, sensitivity, and specificity. For example, if the distance between the pre- and post-test probabilities is much larger for *A* than for *B* then the uncertainty will also be much greater.

REFERENCES

American Psychological Association (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington DC: APA.

Antelman, G. (1997). *Elementary Bayesian statistics*. Cheltenam, UK: Elgar.

Becker, J. T., Knowlton, B., & Anderson, V. (2005). Editorial. *Neuropsychology*, *19*, 3–4.

Bolstad, W. M. (2007). *Introduction to Bayesian statistics* (2nd ed.). Hoboken, NJ: Wiley.

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*, 404–413.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, *40*, 1196–1208.

Crawford, J. R., & Garthwaite, P. H. (2007). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. *Cognitive Neuropsychology*, *24*, 343–372.

Crawford, J. R., & Garthwaite, P. H. (2008). On the ''optimal'' size for normative samples in neuropsychology: Capturing the uncertainty associated with the use of normative data to quantify the standing of a neuropsychological test score. *Child Neuropsychology*, *14*, 99–117.

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, *100*, 680–700.

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology*, *18*, 284–295.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, *70*, 271–294.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Belmont, CA: Duxbury Press.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). London: Cambridge University Press.

Mossman, D., & Berger, J. O. (2001). Intervals for posttest probabilities: A comparison of 5 methods. *Medical Decision Making*, *21*(6), 498–507.

Phillips, L. D. (1973). *Bayesian statistics for social scientists*. London: Nelson.

Schaefer, R. E., & Borcherding, K. (1973). The assessment of subjective probability distributions: A training experiment. *Acta Psychologica*, *37*, 117–129.

Smith, G. E., Ivnik, R. J., & Lucas, J. (2008). Assessment techniques: Tests, test batteries, norms and methodological approaches. In J. E. Morgan & J. H. Ricker (Eds.), *Textbook of clinical neuropsychology* (pp. 41–60). New York: Taylor & Francis.

APPENDIX 1

Finding beta(*a*, *b*) from mode and standard deviation assessments

The search procedure is complicated a little by the fact that, in their initial attempt to quantify their opinion, users may select impossible combinations of the mode and standard deviation.

Let σ be the assessed standard deviation and let m be the assessed mode (m is between 0 and 1). Also let m^* be the larger of m and $1-m$.

Method:

(1) Let U be an upper bound for a and let L be a lower bound. Put $L=0.501$. We first look for an upper bound U and check that σ is not impossibly large.
(2) Set a equal to $2L$.
(3) Estimate b by putting $b=2-a+(a-1)/m^*$
(4) Calculate an estimate (s) of the assessed standard deviation (σ) using Equation (4) in the text.
(5) If $a=1.002$ and $s<\sigma$, then the assessed value σ is too large and must be re-assessed. Return to step (1) with a new assessed value of σ.
(6) If $s\geq\sigma$, set L equal to a and return to step (2).
(7) If $a>1.002$ and $s<\sigma$, then set U equal to a and continue to step (8).
(8) Set a equal to $(U+L)/2$.
(9) Estimate b by putting $b=2-a+(a-1)/m^*$
(10) Calculate an estimate (s) of the assessed standard deviation (σ) using Equation (4) in the text.
(11) If $s<\sigma$, set U equal to a; if $s>\sigma$, set L equal to a. [If $s=\sigma$, then we have found a and should go to step (13).]
(12) Return to step (8) and repeat steps (8)–(11) fifteen times (this gives sufficient precision).
(13) If $m\geq 0.5$ then the final estimates of a and b are the required parameters of the beta prior. If $m<0.5$ then a and b must be switched to give the parameters of the beta prior.