A critique of the false-positive report probability

Joseph Lucke
A Critique of the False-Positive Report Probability

Joseph F. Lucke

Center for Clinical Research and Evidence-based Medicine, University of Texas Medical School at Houston, Houston, Texas

The false positive report probability (FPRP) was proposed as a Bayesian prophylactic against false reports of significant associations. Unfortunately, the derivation of the FPRP is unsound. A heuristic derivation fails to make its point, and a formal derivation reveals a probabilistic misrepresentation of an observation. As a result, the FPRP can yield serious inferential errors. In particular, the FPRP can use an observation that is many times more likely under the null hypothesis than under the alternative to infer that the null hypothesis is far less probable than the alternative. Contrary to its intended purpose, the FPRP can promote false positive results. It should not be used. A modified FPRP is derived, but it appears to have limited application and does not address the problem of false reports of significant associations. The conditional error probability is a possible replacement for the FPRP. "Genet. Epidemiol. 33:145–150, 2009. © 2008 Wiley-Liss, Inc.

Key words: Bayesian analysis; conditional error probability; false-positive report probability (FPRP); hypothesis tests; P-value

INTRODUCTION

Wacholder et al. [2004] (WCGER) proposed a new method for assessing the probable truth of a reported scientific result by combining the frequentist concepts of significance and power with the Bayesian concept of subjective probabilities of hypotheses to obtain, via Bayes’s theorem, the false-positive report probability (FPRP)—"the probability of no association given a statistically significant finding" [Wacholder et al., 2004, p 434]. By assigning a small prior probability to the hypothesis of an association, skepticism can be formally incorporated into assessing its posterior probability given a significant result. Such skepticism appears essential in disciplines such as molecular genetic epidemiology and genome-wide association studies (GWAS) in which the number of single nucleotide polymorphisms (SNPs) or haplotypes that could be tested for associations is approaching $10^6$ [Wellcome Trust Case Control Consortium, 2007; Ziegler et al., 2008]. The FPRP is intended to separate noteworthy discoveries from spurious reports [Wacholder et al., 2004, p 439].

To understand how the FPRP is determined, WCGER’s example given in their appendix is recapitulated. Suppose an investigator has observed a log odds ratio $x_0 = \log 1.516 = 0.2746$ with standard error $s = 0.1003$. (The extra digits of precision are included in case the reader wants to replicate the example provided in WCGER’s Excel® spreadsheet program.) The computation of the FPRP requires four steps.

1. Two statistical hypotheses are postulated. The null hypothesis of spuriousness claims that $x_0$ arose from a normal distribution with mean 0 and standard deviation $\sigma$. The alternative hypothesis of noteworthy requires the specification of a mean representing a value of scientific importance, $\pm \mu_i$, for some $\mu_i > 0$. For log odds ratios in GWAS, WCGER proposed values between $\log 1.2$ and $\log 2$ [Wacholder et al., 2004, pp 436, 441]. Upon choosing the larger, the alternative claims that $x_0$ arose from a normal distribution either with mean $\mu_1 = \log 2 = 0.6931$ or $\mu_1 = -\log 2$. An estimate of $\sigma$ must be retrieved from the report. If the observed standard deviation $s$ and sample size $n$ are given, then we use $\sigma = s$. If only the standard error $\sqrt{n}/\sigma$ is given without $n$ itself being given, then, without loss of generality, we set $\sigma = s/\sqrt{n}$ and $n = 1$. The standard error is often extracted from the confidence interval of the log odds estimate.

2. The prior probabilities of the null hypothesis, $P_0$, and the alternative, $1 - P_0$, are determined. This is a subjective evaluation of the probable truth of the previously determined, alternative hypothesis derived from prior research, theoretical plausibility, and scientific consensus. Here $P_0 = 0.9$ is chosen, although it could, in some circumstances, be as large as 0.99999 [Wacholder et al., 2004, p 436].

3. The investigator determines a criterion for noteworthyness, here called $v$. This criterion is a probability value no more than 0.5 and perhaps as small as 0.2 [Wacholder et al., 2004, p 436]. If the FPRP, which remains to be calculated, is less than $v$, then the report will be considered noteworthy at level $v$. Here $v = 0.2$ is chosen.

4. For any $x$, the observed test quantity $t$ is

$$t = \sqrt{n}x/\sigma,$$  \hspace{1cm} (1)

which will be compared to the reference quantity

$$\delta = \sqrt{n}\mu/\sigma,$$  \hspace{1cm} (2)
for any \( \mu \). The FPRP is computed from two quantities denoted here as \( p_0(t) \) and \( p_1(t) \). The function \( p_0(t) \) is the probability observing a value greater than \( |t| \) or less than \(-|t|\) under the null hypothesis of no association. This quantity is the familiar \( P \)-value, the formula for which being presented later. The given values here yield \( t_1 = 2.378 \) and \( p_0(2.378) = 0.0062 \). The function \( p_1(t) \) is the probability of observing a value greater than \( |t| \) or less than \(-|t|\) under the alternative hypothesis of \( \pm \delta_1 \) from equation (2). This quantity, the formula for which likewise being presented later, may be construed as the “observed power” of the observation for the alternative hypothesis. The given values yield \( \delta_1 = 6.911 \) and \( p_6.911(2.378) = 1 \). The FPRP at \( t \) is obtained as

\[
\text{FPRP} = \frac{p_0(t) \Pi}{p_0(t) \Pi + p_1(t) (1 - \Pi)}. \tag{3}
\]

Substituting the obtained values yields an observed FPRP = 0.053. As this value is less than \( \nu = 0.2 \), this reported result is declared noteworthy.

Had the observation instead been, say, \( x_1 = \log 1.22 \), keeping the same standard error and mean for the alternative hypothesis, then \( t_1 = 1.983 \), \( p_0(1.983) = 0.047 \), \( p_6.911(1.983) = 1 \), and FPRP = 0.30. The report would have been significant at \( \nu = 0.05 \), but spurious at \( \nu = 0.2 \).

The FPRP opened to mostly favorable editorial reviews [Rebeck et al., 2004; Thomas and Clayton, 2004; Matullo et al., 2005], being promoted as “ingenious and more convincing than other approaches” [Matullo et al., 2005, p 550]. The FPRP has largely been accepted as a valid inferential procedure, and a Web of Science citation search shows that as of June, 2008, the FPRP had been cited in over 300 mostly substantive but occasionally methodological articles [Kraft, 2006; Wakefield, 2007; Strömbäck et al., 2008].

The FPRP received only light criticism, most of which focused on Bayesian issues. The most salient problem, noted by WCGER themselves, has been the determination of prior probabilities for the two hypotheses [Rebeck et al., 2004; Thomas and Clayton, 2004; Wacholder et al., 2004; Matullo et al., 2005]. However, the elicitation of prior probabilities is a standard, if difficult, problem in Bayesian inference, is viewed as an advantage by most Bayesians, and in any case, does not invalidate the Bayesian approach [Spiegelhalter et al., 1994; Kadane, 1995; Garthwaite et al., 2005; Greenland, 2006]. Thus, the problem of eliciting priors applies to Bayesian statistics in general and not to the FPRP specifically.

A second problem with the FPRP is its extremely restricted setting [Thomas and Clayton, 2004; Wakefield, 2007]. WCGER considered only normal distributions with equal and known variances. They further restricted the setting by allowing only two discrete points symmetric about zero to represent nonzero associations. To make matters worse, they equated the population standard error to the sample standard error and employed computationally inaccurate methods to obtain the sample standard error. Although these restrictions seriously hamper the application of the FPRP in realistic settings, they presumably could be lifted with further developments. The assumed normal distributions with equal, known variances could be replaced with Student’s \( t \) distributions with unknown, unequal variances. The two-point alternative hypothesis could perhaps be replaced with an arbitrary finite set of hypotheses or a set of normal- or \( t \)-distributed continuous alternatives. Once the FPRP has proved its mettle, these and other generalizations would easily follow.

These reviews overlooked the most serious problem: The FPRP is incorrectly derived. Accepting the admittedly simplistic model as adequate, the FPRP nevertheless is not correct. The likelihood of the FPRP uses tail-area probabilities \( p_0 \) and \( p_1 \). It should instead use probability densities. Recently several investigators have noted that the use of tail-area probabilities in the FPRP is an error. Because of this error, the FPRP wastes information and underestimates the posterior distribution of the null [Wakefield, 2007, 2008]. Some have proposed to modify the FPRP by using probability densities, but have not presented any details [Samani et al., 2007; Ziegler et al., 2008].

To illustrate this flaw in the FPRP as a Bayesian method, consider again the two examples. In the first, the prior probability of a false-positive result was \( \Pi = 0.9 \) and the posterior was \( \text{FPRP} = 0.053 \). However, the observation, \( X_1 = \log 1.22 \), is closer to the mean, 0, of the null distribution than it is to the mean, \( \log 2 \), of the alternative (2.7 versus 4.2 standard errors, respectively). Thus, an observation more likely to have arisen from the null distribution than from the alternative is inferred to favor the alternative. This result violates common sense. If the data are more likely to have arisen from the null than from the alternative, the posterior distribution should yield an increase, not a decrease, in probability of the null. In the second example, the probability is reduced from the prior of 0.9 to a posterior of 0.3, even though the observation, \( X_1 = \log 1.22 \), is likewise closer the null mean than it is to the alternative mean (2.0 versus 4.9 standard errors, respectively). Again, an observation more likely to have arisen from the null distribution than from the alternative reduces the probability of the null.

These counterintuitive inferences are symptoms of the incorrect derivation of the FPRP. It also has another, more general source of counterintuitive inferences: The FPRP can never increase the posterior probability of the null hypothesis. For any observation, \( \text{FPRP} \) will be at most \( \Pi \), and for every nonzero observation, even if it is more likely to have arisen from the null distribution than the alternative, the FPRP will be less than \( \Pi \). No matter how successfully the null hypothesis predicts an observation, the FPRP cannot give it additional credence for having done so.

**THE BASIC SCENARIO**

Consider an outcome random variable \( x \) that follows a normal distribution with a random mean \( \mu \) and known standard deviation \( \sigma \) in a sample of size \( n \). Consider two hypotheses. The null hypothesis of spuriousness claims that \( \mu = 0 \). The alternative hypothesis of noteworthiness claims that it can take the values \( \mu = \mu_1 \) or \( \mu = -\mu_1 \) with equal probability (\( \mu_1 > 0 \)). The mean \( \mu \) is to be interpreted as a level of association, with 0 being no association and \( \pm \mu_1 \) being a nonzero association irrespective of direction.

An investigator wishes to conduct a Bayesian comparison of the hypotheses \( \mu = 0 \) versus \( \mu \in \{ \mu_1, -\mu_1 \} \). Her prior probability for the null hypothesis is

\[
\Pr(\mu = 0) = \Pi. \tag{4}
\]
Her prior for the alternative hypothesis is
\[
\Pr(\mu \in [\mu_1, -\mu_1]) = 1 - \Pi \quad \text{with} \quad \Pr(\mu = \mu_1) = (1 - \Pi)/2 = \Pr(\mu = -\mu_1).
\] (5)
The test statistic and reference value are given in equations (1) and (2), but with \(x\), \(\mu\), \(t\), and \(\delta\) being random variables.

**A HEURISTIC DERIVATION**

WCGER gave a heuristic derivation of the FPRP that combined Neyman-Pearson Type I and II error rates, \(\alpha\) and \(\beta\), with the prior probability of the association being null to obtain via Bayes theorem, the posterior probability of the null hypothesis given a significant result [Wacholder et al., 2004, Appendix].

Let \(x\) be the outcome random variable with its test statistic given in the section “The Basic Scenario.” Now, let \(s_2\) be the binary random variable denoting “significance” such that \(s_2(t) = 1\) if \(t\) is significant at level \(\alpha\) and \(s_2(t) = 0\) if not. The probability of a significant result is postulated to be \(\alpha\) when the null hypothesis is true and \(1 - \beta\) when the alternative is true. The reporting of significant result then follows a Bernoulli process.

\[
\Pr[s_2(t) = s|\mu = 0] = \alpha(1 - \alpha)^{1-s},
\] (6)
and
\[
\Pr[s_2(t) = s|\mu \in [\mu_1, -\mu_1]] = (1 - \beta)^{1-s},
\] (7)
where \(s = 0,1\). Given the prior probabilities in equations (4) and (5), the posterior probability of no association given a report of significance or nonsignificance is

\[
\Pr[\mu = 0] = \frac{\alpha^s(1 - \alpha)^{1-s}}{\alpha^s(1 - \alpha)^{1-s} + (1 - \beta)^{1-s}(1 - \Pi)}.
\] (8)
which yields WCGER’s Table 1 [Wacholder et al., 2004, p 440]. The posterior probability of no association given a significant outcome is

\[
\text{FPRP}^\ast = \Pr[\mu = 0|s_2(t) = 1] = \frac{\alpha\Pi}{\alpha\Pi + (1 - \beta)(1 - \Pi)},
\] (9)
which corresponds to WCGER’s equation (1) [Wacholder et al., 2004, p 440]. WCGER then considered \(p_0(t)\) and \(p_0(t)\) to be estimators of \(\alpha\) and \(1 - \beta\), respectively, thereby yielding FPRP given by (3) as an estimator of FPRP” given by (9) [Wacholder et al., 2004, p 441].

Treating \(p_0\) as an estimator of \(\alpha\) and \(p_0\) as an estimator \(1 - \beta\) has been noted as problematic [Thomas and Clayton, 2004; Wakefield, 2007]. More than problematic, this assumption commits a major conceptual mistake by conflating the operating characteristics of a statistic with a measure of evidence [Goodman, 1999; Hubbard and Bayarri, 2003]. On one hand, \(\alpha\) and \(\beta\) are pre-study quantities giving the anticipated performance of the test over an infinite set of identical replications. On the other, \(p_0\) and \(p_0\) are post-study random variables, in which \(p_0\) follows the standard uniform distribution and \(p_0\) follows a distribution parameterized only by \(\delta\) [Hung et al., 1997]. Neither \(\alpha\) nor \(\beta\) nor functions of the two are parameters for the distributions of either \(p_0(t)\) or \(p_0(t)\). Realizations of \(p_0(t)\) or \(p_0(t)\) cannot provide any information regarding \(\alpha\) or \(\beta\). Therefore, neither \(p_0(t)\) nor \(p_0(t)\) can be estimators of \(\alpha\) or \(\beta\), and the heuristic derivation cannot be an estimator of the FPRP”. The heuristic derivation fails to make its point.

**A DERIVATION OF THE FPRP**

WCGER did not present a derivation of the FPRP as given in equation (3). Here is a derivation. Consider the relevant outcome for inference to be the interval

\[
U(t) = (-\infty, -|t|] \cup [|t|, \infty),
\] (10)
i.e., the union of the tail-area intervals comprising the observed point \(|t|\) along with unobserved outcomes greater than \(|t|\) and all unobserved outcomes at most as large as \(-|t|\). For any \(x\) and \(\mu\) and therefore \(t\) and \(\delta\), the probability of the random interval \(U(t)\) is

\[
p_0(t) = \Pr[U(t)|\delta] = 2 - \Phi(|t| - \delta) - \Phi(|t| + \delta),
\] (11)
where \(\Phi\) is the standard normal distribution function.

With the priors given in equations (4) and (5), Bayes’s Theorem yields the posterior odds of the null hypothesis to the alternative as

\[
\Omega_W(t) = \frac{\Pr[\mu = 0|U(t)]}{\Pr[\mu \in [\mu_1, -\mu_1]|U(t)]} = \frac{W(t)}{1 - W(t)}
\] (12)
where

\[
W(t) = \frac{p_0(t)}{p_0(t) + 2[1 - \Phi(|t|)] - \Phi(|t| - \delta) - \Phi(|t| + \delta)}
\] (13)
is the Bayes factor for the null against the alternative given outcomes more extreme than \(|t|\) or \(-|t|\). The FPRP at the observation \(t\) is then

\[
\text{FPRP}(t) = \Pr[\mu = 0|U(t)] = \frac{\Omega_W(t)}{1 + \Omega_W(t)}
\] (14)
which in turn yields (3). Equation (11) does not appear in WCGER’s article, nor do the ensuing derivations. Instead this equation was derived from the ideas presented in WCGER’s appendix and partially confirmed by the formulas used in their Excel spreadsheet. The function \(p_0\) appears in their spreadsheet but \(p_0\) does not, as they used the formally incorrect but numerically similar \(1 - \Phi(|t|)\).

This derivation, which explicates WCGER’s reasoning, is not sound. The error occurs in equation (10). The interval \(U(t)\) is not the correct representation of an observation \(t\). The interval \(U(t)\) represents an unobserved outcome to be found at random in the complement of the open interval \((-|t|, |t|)\), with \(t\) itself merely determining the bounds for this interval. Equation (10) models the outcome as if it were subject to censoring at \(t\) [Wakefield, 2007]. The likelihood given by (11) is correct for \(U(t)\), but incorrect for \(t\). The posterior odds ratio, (12), Bayes factor, (13), and FPRP, (14), all correctly follow from the previous equations, but use the wrong model for the observation.

WCGER’s formulation of an interval observation yields an especially peculiar result regarding no association. The interval observation \(U(t) = (-\infty, \infty)\) is the universal event and has probability 1 under any model, which can be confirmed by substituting \(t = 0\) into equation (11). It follows from equation (13) that \(W(0) = 1\) and from (12) and (14) that \(\text{FPRP} = 1\). According to WCGER’s reasoning, the observation of no association (\(x = 0\)) is equivalent to “any outcome could have occurred” and entirely uninformative.
The use of the interval $U(t)$ rather than the observation $t$ generates counterintuitive inferences. The Bayes factor $W$ is restricted to
\[ p_0(t) \leq W(t) \leq 1. \] (15)

Proof is given in the Appendix. An informal understanding of this result is easily comprehended. For $\delta = 0$, $W(t) = 1$. For $\delta$ very large, $\Phi(|t| + \delta) \approx 1$ and $\Phi(|t| - \delta) \approx 0$, so that the denominator of (13) is approximately 1 and $W(t) = p_0(t)$. The restriction of range in $W$ forces the FPRP in (14) to be bounded above by the prior probability of the null hypothesis $\Pi$:
\[ \frac{p_0(t)\Pi}{1 - \Pi[1 - p_0(t)]} \leq \text{FPRP}(t) \leq \Pi. \] (16)

The FPRP can never increase the probability the null hypothesis beyond its prior. For any nonzero observation, even one more likely under the null distribution than under the alternative, the FPRP will always be less than the prior.

### A MODIFIED FPRP

This section presents the correct Bayesian analysis of the model with the scenario given in the section “The Basic Scenario.” The result is presumably the same as the undocumented “modified FPRP” reported elsewhere [Samani et al., 2007; Ziegler et al., 2008].

The probability density for the test $t$ is
\[ f(t; \delta) = \frac{1}{2}[\phi(t - \delta) + \phi(t + \delta)], \] (17)
where $\phi$ is the standard normal density function. The posterior odds for the null against alternative at $t$ are, via Bayes’s Theorem,
\[ \Omega(t) = \frac{\Pr(\mu = 0 | t)}{\Pr(\mu \in \{\mu_1, -\mu_1\} | t)} = B(t) \frac{\Pi}{1 - \Pi}, \] (18)

where
\[ B(t) = \frac{2\phi(t)}{\phi(t - \delta_1) + \phi(t + \delta_1)} \] (19)
is the Bayes factor at $t$ for $\mu = 0$ against $\mu \in \{\mu_1, -\mu_1\}$. The modified FPRP, the posterior probability of the null hypothesis given the observation $x$, is
\[ \text{mFPRP} = \Pr(\mu = 0 | x) = \frac{\Omega(t)}{1 + \Omega(t)}. \] (20)

Table I presents the comparison of FPRP and modified FPRP for various prior distributions.

<table>
<thead>
<tr>
<th>Prior</th>
<th>FPRP</th>
<th>mFPRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>0.0061</td>
<td>0.997</td>
</tr>
<tr>
<td>0.75</td>
<td>0.018</td>
<td>0.999</td>
</tr>
<tr>
<td>0.90</td>
<td>0.053</td>
<td>0.9996</td>
</tr>
<tr>
<td>0.99</td>
<td>0.38</td>
<td>0.99996</td>
</tr>
<tr>
<td>0.999</td>
<td>0.86</td>
<td>0.999997</td>
</tr>
<tr>
<td>0.9999</td>
<td>0.98</td>
<td>0.99999996</td>
</tr>
</tbody>
</table>

Given a prior probability of 0.5 for the null hypothesis, the FPRP is 0.0061. Similarly, given a prior probability of 0.9 for the null, the FPRP is 0.053. The FPRP reduces the initial skepticism in every case. The FPRP indicates the observation to be noteworthy at $\nu < 0.5$ for $\Pi \leq 0.99$ and noteworthy at $\nu < 0.2$ for $\Pi \leq 0.90$.

This conclusion is absurd. The observation is $B = 285$ times more likely under the null than the alternative. The properly derived modified FPRP supports the null hypothesis with posterior probabilities exceeding 0.99 for all $\Pi > 0.26$. The modified FPRP properly credits the null hypothesis as accounting for the observation. The FPRP always credits the alternative. The FPRP reduces initial skepticism when it should be increasing it, thereby promoting noteworthiness where none exists.

### ONE-SIDED TESTS

The above results are based on two-sided tests in accordance with WCGER’s presentation. Here, a digression is undertaken to show that the above results are similar for one-sided tests. The notation is recycled and confined to this section. Assume $\mu_1 > 0$. The two hypotheses are $\mu = 0$ versus $\mu = \mu_1$ with the prior for the null being $\Pi$ and that for the alternative being $1 - \Pi$. The test statistic $t$ and reference $\delta$ remain as in equations (1) and (2).

For the WCGER approach, the interval in equation (10) is replaced with
\[ U(t) = \langle t, \infty \rangle, \] so that equation (11) is replaced with
\[ p_0(t) = \Pr[U(t) | \delta] = 1 - \Phi(t - \delta). \] Thus, (13) becomes
\[ W(t) = \frac{1 - \Phi(t)}{1 - \Phi(t - \delta)}. \] The posterior odds and FPRP are still given by (12) and (14), respectively.

As in the two-sided case, this derivation is unsound because the interval $U$ does not correctly represent the observation $t$. Furthermore, $p_0(x) \to 0$ as $\delta \to \infty$, and $\Phi(\delta) \geq \Phi(t - \delta)$, so that equations (15) and (16) still obtain. Thus, $W$ and the FPRP carry the same flaws in the one-sided case as found in the two-sided case.

For the modified FPRP, the density function given by (17) is replaced with
\[ f(t; \delta) = \phi(t - \delta). \]

Genet. Epidemiol.
The Bayes’s factor given by (19) is thus replaced with
\[ B(t) = \frac{\phi(t)}{\phi(t - \delta)}. \]
The posterior odds and probability of the null are still given by (18) and (20), respectively.

**THE CONDITIONAL ERROR PROBABILITY**

Although the FPRP was derived in error, the idea behind the FPRP has considerable merit. There should be a simple Bayesian procedure that combines evidence from the study with additional, relevant information to determine, if only approximately, the probable truth of a reported result. Fortunately, there is such a statistic. Let \( p \) be the ordinary \( p \)-value, which heretofore was denoted \( p_0 \), and let
\[ B(p) = -cp \log(p), \]
where \( p < e^{-1} \). This quantity is approximately the lower bound to the Bayes factor for the point-null hypothesis against a wide class of continuous alternatives [Sellke et al., 2001]. Unfortunately, its justification and derivation are too complicated to present here. Let \( \Pi \) be the prior probability of the null hypothesis. The conditional error probability, \( \alpha(p, \Pi) \), is the lower bound to the posterior probability of the point-null hypothesis, that is \( \Pr(\mu = 0 | p) \geq \alpha(p, \Pi) \). The conditional error probability is then
\[ \alpha(p, \Pi) = \left( 1 + \frac{1}{B(p)} \frac{1 - \Pi}{\Pi} \right)^{-1}. \]

Of the two examples given in the introduction with \( \Pi = 0.9 \), the first yields \( \alpha(0.0062, 0.9) = 0.44 \) and the second yields \( \alpha(0.047, 0.9) = 0.78 \). Thus, according to the conditional error probability with the specified prior, the first observation, \( x_0 = \log 1.316 \), has at least a 44% chance of being spurious and the second, \( x_1 = \log 1.22 \), at least a 78% of being spurious. The conditional error probability has recently been introduced to genetic epidemiology [Wakefield, 2008]. Perhaps it can become a suitable replacement for the FPRP.

**DISCUSSION**

WCGER proposed the FPRP as a Bayesian prophylactic against false reports of significant associations [Wacholder et al., 2004, p 439]. Unfortunately, it uses a wrong probabilistic representation of an observation, so that its derivation is not sound. It does not yield, as was claimed, the posterior probability of the null hypothesis given a significant report. Furthermore, it can yield substantial inferential errors. The FPRP can use an observation that is more likely under the null hypothesis than the alternative to infer the null hypothesis is probably false. WCGER’s own example is a case in point. Contrary to its intent, the FPRP promotes false-positive results. Given the increased prominence being placed on its use [e.g., Hung et al., 2005; Engels et al., 2007; Dong et al., 2008; Tobin et al., 2008], the risk of incorrectly inferred “noteworthy” results is no longer an academic possibility. The FPRP should henceforth be proscribed.

The modified FPRP correctly gives the posterior probability of the null hypothesis given the observation. However, there is very little to recommend it for general use. The modified FPRP is but an extreme special case of Bayesian hypothesis testing. It should only be used when the alternative hypothesis is a specific, symmetric two-point normal distribution that is seriously proposed and not merely offered as a convention. The type of hypothesis for which the modified FPRP is derived does not appear realistic for most studies in GWAS. Bayesian hypothesis tests for a point-null hypothesis against a continuous set of alternatives are well documented for several important distributions [Lee, 2004; Wakefield, 2007].

Even though there is no substitute for a Bayesian analysis derived specifically for the substantive hypotheses of interest, there does remain a use for a Bayesian method that can be used to quickly and automatically review set of results to identify, if only approximately, probably true results. One possible method is the conditional error probability [Sellke et al., 2001]. It remains to be exploited by GWAS.

**REFERENCES**


APPENDIX

Proof of Equation (15). Without loss of generality, let $\sigma = 1$. For the left inequality of (15), fix $x \geq 0$. As $|\delta| \to \infty$, $\Phi(x - \delta) \to 0$ and $\Phi(x + \delta) \to 1$, so that $p_0(x) \to 1$. In practice, $W(x)$ differs negligibly from $p_0(x)$ for $|\delta| > 6$.

Turning to the right inequality, again fix $x \geq 0$ and $\delta > 0$. Let $a = x - \delta$ and $b = x + \delta$, noting that $a + b \geq 0$. Also note that $\Phi$ is concave for $0 \leq u < \infty$ because $\Phi'(u) = -u \Phi(u) \leq 0$ [Bartle and Sherbert, 2000, p 188, Theorem 6.4.6]. For $a \leq u \leq b$, let

$$f(u) = \frac{b - u}{b - a} \Phi(a) + \frac{u - a}{b - a} \Phi(b)$$

be the line connecting $(a, \Phi(a))$ with $(b, \Phi(b))$.

Assume $a \geq 0$. Then because $\Phi$ is concave on $[a, b]$,

$$\Phi\left(\frac{a + b}{2}\right) \leq f\left(\frac{a + b}{2}\right).$$

Otherwise assume $a < 0$. For $0 \leq u \leq b$, let

$$g(u) = \frac{b - u}{2b} + \frac{a}{b} \Phi(b)$$

be the line connecting $(0, 1/2)$ with $(b, \Phi(b))$. Then

$$g\left(\frac{a + b}{2}\right) = \frac{b[1 - \Phi(a)] - a[1 - \Phi(b)]}{4b} > 0.$$  

Because $\Phi$ is concave on $[0, b]$,

$$\Phi\left(\frac{a + b}{2}\right) \leq g\left(\frac{a + b}{2}\right) > f\left(\frac{a + b}{2}\right).$$

Therefore, in either case

$$\Phi(x) = \Phi\left(\frac{a + b}{2}\right) \geq f\left(\frac{a + b}{2}\right) = \frac{\Phi(x - \delta) + \Phi(x + \delta)}{2}.$$  

Equation (13) follows immediately, and the right inequality of (15) follows from (13).