

Letter to the Editor

Using the False Discovery Rate Approach in the Genetic Dissection of Complex Traits: A Response to Weller *et al.*

Dmitri V. Zaykin,^{*,†} S. Stanley Young[†] and Peter H. Westfall^{†,‡}

^{*}Programs in Statistical Genetics and Biomathematics, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203, [†]GlaxoWellcome Inc., Research Triangle Park, North Carolina 27709 and [‡]Texas Tech University, Lubbock, Texas 79409

Manuscript received August 4, 1999
Accepted for publication December 13, 1999

WELLER *et al.* (1998) proposed controlling the “false discovery rate” (FDR) or the expected proportion of false rejections within the class of rejected null hypotheses when performing preliminary genome scans, adopting the method of Benjamini and Hochberg (1995; hereafter BH). The BH procedure is as follows: for the ordered set of P values, $P_{(1)} \leq \dots \leq P_{(L)}$ determine a largest j such that $P_j \leq j\alpha/L$, where α is the declared FDR. Then reject all hypotheses H_i that correspond to $P_{(i)}$, $i = 1, \dots, j$. The method controls the average proportion of false positives at 5% (say) across multiple studies, including those where no significances are found. However, this procedure does not provide information about the expected proportion of false positive results for a given experiment where some of the null hypotheses are rejected.

Benjamini and Hochberg (1995) write that “a desirable error rate to control may be the expected proportion of errors among the rejected hypotheses, which we term the false discovery rate (FDR).” The actual method that they propose controls this false discovery rate in an *unconditional* manner, and they readily acknowledge that their method cannot control FDR, *conditional upon having rejected one or more hypotheses*.

Weller *et al.* (1998) state that “A further advantage of the FDR is that an accurate prediction has been made of the proportion of hypotheses rejected in the first analysis that represent true effects” and consequently that seven or eight marker-trait combinations in the analysis of actual data are likely to be true effects. Such interpretation of the FDR control is, however, wrong, because experiments with multiple tests and with successful rejections usually have a greater expected proportion of false positives than α .

To be specific, let us consider the conditional FDR, given one or more rejections, defined as

$$\text{FDR}^* = E\left\{\frac{F}{T+F} \mid T+F \geq 1\right\}, \quad (1)$$

where T, F are the numbers of true and false rejections. This quantity is not controlled using the BH method.

The unconditional FDR, which is controlled by the BH method, is

$$\begin{aligned} \text{FDR} &= E\left\{\frac{F}{T+F} \mid T+F \geq 1\right\} \Pr(T+F > 0) \\ &\quad + 0 \Pr(T+F = 0) \end{aligned} \quad (2)$$

$$= \text{FDR}^* \Pr(T+F > 0). \quad (3)$$

The difference between the conditional and unconditional false discovery rates, FDR^* and FDR, can be very substantial.

To illustrate this point, we conducted a series of simulations with $\alpha = 0.25$ and $L = 896$, which correspond to the number of tests performed on the actual data in Weller *et al.* (1998). We included estimates of family-wise error rate (FWER) within the class of rejected hypotheses (FWER'), probabilities of rejection, and estimates of FDR^* for the BH method (Table 1) and the FWER-controlling method of Hochberg (1988; Table 2).

We assumed a continuous test statistic, so that P values corresponding to true null hypotheses were generated from the uniform (0, 1) distribution; P values corresponding to N_A true effects were generated as

$$P_A = 1 - \Phi(\Phi^{-1}(1 - u) + \gamma), \quad (4)$$

where u is a uniform (0, 1) random number and Φ is the standard normal cumulative distribution function. The parameter γ was set to ensure 100 β % power for the individual 5%-level tests. All estimates were obtained by averaging over at least 10,000 simulation experiments.

Weller *et al.* (1998) state the following on p. 1703: “If all 10 hypotheses are rejected, q , and thus FDR, are still < 0.25 , even though $\text{FWER} = 0.88$. Thus, seven or eight marker-trait combinations should represent “true” effects and can be expected to repeat on a second popu-

Corresponding author: Dmitri Zaykin, Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203. E-mail: zaykin@statgen.ncsu.edu

TABLE 1

Simulations of the conditional false discovery rate with Benjamini and Hochberg's (1995) method

| β | N_A | FWER* | FDR* | $\Pr(T + F \geq 1)$ |
|---------|-------|-------|-------|---------------------|
| 0.50 | 1 | 0.925 | 0.891 | 0.280 |
| 0.80 | 1 | 0.748 | 0.648 | 0.393 |
| 0.95 | 1 | 0.561 | 0.419 | 0.598 |
| 0.99 | 1 | 0.495 | 0.328 | 0.782 |
| 0.50 | 2 | 0.868 | 0.804 | 0.307 |
| 0.80 | 2 | 0.655 | 0.496 | 0.499 |
| 0.95 | 2 | 0.537 | 0.319 | 0.778 |
| 0.99 | 2 | 0.535 | 0.272 | 0.943 |

$L = 896, \alpha = 0.25.$

lation sample." Weller *et al.* found, however, that "only two F values have a FWER < 0.05 ." Table 1 shows that assuming that these two are the only true effects and that the power of corresponding tests is 80%, the proportion of false discoveries *in their data* is expected to be as large as 50%. If there is only one true effect, power $> 99\%$ is needed to maintain the declared FDR.

Similar results are found for the FWER-controlling method of Hochberg (Table 2); therefore, neither method controls the FDR, when conditioned upon the occurrence of one or more rejections. But FWER-controlling methods neither advertise nor require conditional error rate control. When using an FWER-controlling method, one may claim that *all* significances obtained in the study are real, gambling upon the occurrence that the given study was not one of the 25% (or whatever FWER level that is used) that will produce a false positive.

FDR control allows that false positives will occur, in fact they are *expected* in any given study. However, given that a significance has been found, the implied operational interpretation that 95% of the claimed results

TABLE 2

Simulations of the conditional false discovery rate with Hochberg's (1988) method

| β | N_A | FWER* | FDR* | $\Pr(T + F \geq 1)$ |
|---------|-------|-------|-------|---------------------|
| 0.50 | 1 | 0.889 | 0.874 | 0.250 |
| 0.80 | 1 | 0.633 | 0.582 | 0.356 |
| 0.95 | 1 | 0.391 | 0.311 | 0.568 |
| 0.99 | 1 | 0.299 | 0.199 | 0.761 |
| 0.50 | 2 | 0.802 | 0.775 | 0.271 |
| 0.80 | 2 | 0.477 | 0.405 | 0.457 |
| 0.95 | 2 | 0.295 | 0.192 | 0.748 |
| 0.99 | 2 | 0.240 | 0.115 | 0.928 |

$L = 896, \alpha = 0.25.$

will replicate is wrong, since a smaller percentage is expected to replicate in reality. The problem is more pronounced as the total number of true null hypotheses increases (data not shown). Thus, the interpretation of Weller *et al.* is incorrect, and we suggest that no conclusions about the likely proportion of false positives in the given data should be made on the basis of either FWER or FDR controlling methods.

This research was partly supported by National Institutes of Health grant GM-43544 to North Carolina State University and was performed while P.H.W. was Research Fellow at GlaxoWellcome Inc.

LITERATURE CITED

- Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**: 289–300.
- Hochberg, Y., 1988 A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**: 800–802.
- Weller, J. I., J. Z. Song, D. W. Heyen, H. A. Lewin and M. Ron, 1998 A new approach to the problem of multiple comparisons in the genetic dissection of complex traits. *Genetics* **150**: 1699–1706.

Communicating editor: C. Haley