Letters to the Editor

Am. J. Hum. Genet. 72:496, 2003

Simulation-Based *P* Values: Response to North et al.

To the Editor:
North et al. (2002) discussed the estimation of a *P* value on the basis of computer (i.e., Monte Carlo) simulations. They emphasized that such a *P* value is an estimate of the true *P* value. This is essentially their only point with which we agree. The letter from North et al. is more likely to confuse than enlighten.

Consider an observed test statistic, x, that under the null hypothesis follows some distribution, f. Let X be a random variable following the distribution f. We seek to estimate the *P* value, $p = \Pr(X \geqslant x)$. Let $y_1, \ldots, y_n$ be independent draws from f, obtained by computer simulation. Let $r = \#\{i{:}y_i \geqslant x\}$ (i.e., the number of simulated statistics greater than or equal to the observed statistic). Let $\hat{p} = r/n$ and $\tilde{p} = (r+1)/(n+1)$.

North et al. (2002) stated that $\hat{p}$ is "not strictly correct" and that $\tilde{p}$ is "the most accurate estimate of the *P* value." They further called $\tilde{p}$ "the true *P* value."

We strongly disagree with this characterization. First, minor differences in *P*-value estimates on the order of Monte Carlo error should not be treated differently in practice, and so it is immaterial whether one uses $\hat{p}$ or $\tilde{p}$. Second, $\hat{p}$ is a perfectly reasonable estimate of p. Indeed, in many ways $\hat{p}$ is superior to $\tilde{p}$. Given the observed test statistic, x, r follows a binomial (n,p) distribution, and so $\hat{p}$ is unbiased, whereas $\tilde{p}$ is biased. (The bias of $\tilde{p}$ is $(1-p)/(n+1)$.) Further, $\hat{p}$ has smaller mean square error (MSE) than $\tilde{p}$, provided that $p < n/(1+3n) \approx 1/3$. (The MSE of $\hat{p}$ is $p(1-p)/n$, whereas that of $\tilde{p}$ is $(1-p)(np+1-p)/(n+1)^2$.)

These results are contrary to those of North et al. (2002) because they evaluate the performance of $\tilde{p}$ under the joint distribution of both the observed and Monte Carlo data, whereas we prefer to condition on the observed value of the test statistic. Evaluating *P*-value estimates conditionally on the observed data is widely accepted when the estimation is performed via analytic approximations.

Regarding the question of how many simulation replicates to perform, we recommend consideration of the precision of the estimate, $\hat{p}$, using the properties of the binomial distribution, rather than adherence to a rule such as $r \geqslant 10$. Standard statistical packages, such as R (Ihaka and Gentleman 1996), allow one to calculate a CI for the true *P* value and to perform a statistical test, such as whether the true *P* value is <.01.

Karl W. Broman and Brian S. Caffo
Department of Biostatistics
Johns Hopkins University
Baltimore

References

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comp Graph Stat 5:299–314

North BV, Curtis D, Sham PC (2002) A note on the calculation of empirical *P* values from Monte Carlo procedures. Am J Hum Genet 71:439–441

Address for correspondence and reprints: Dr. Karl W. Broman, Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205. E-mail: kbroman@jhsph.edu

0002-9297/2003/7202-0032$15.00

Am. J. Hum. Genet. 72:496–498, 2003

On Estimating *P* Values by Monte Carlo Methods

To the Editor:
North et al. (2002) propose a new formula for the empirical estimation of *P* values by Monte Carlo methods to replace a standard conventional estimator. They claim that their new formula is "correct" and "most accurate" and that the conventional formula is "not strictly correct," repeating this claim many times in their letter. The claim, however, is incorrect, and the conventional formula is the correct one.

The North et al. claim arises when a test statistic (called here "t") takes a certain numerical value (called here "t^*") when calculated from data from some experiment, and it is required to find an unbiased estimate of the *P* value corresponding to t^* by Monte Carlo simulation. This is done by performing n Monte Carlo simulations, all performed under the null hypothesis tested