A Statistic for Inferences Based upon Negative Results

To the Editor—The letter by Benefiel et al. calls attention to an important concept of probability related to the strength of a clinical inference based upon the nonoccurrence of a disease in a series of patients. However, the equation for such is incorrect, as printed. Hence, I would like to derive the correct formula while presenting the simple underlying concepts.

If \( R \) is the rate of occurrence or incidence of the condition in a general population, \( 1 - R \) is the fraction of the population free of the condition and represents the probability that observation of a single patient will be negative. Then, the probability \( P \) of a negative observation in \( n \) consecutive patients is \( 1 - R \), or

\[
(1 - R)^n = P. \tag{1}
\]

The \( P \) in equation 1 is indeed our familiar \( P \) value, usually taken at 0.05 or 5\%, representing the probability that the observed result occurred by chance alone. The corresponding level of confidence, \( 1 - P \), is 0.95 or 95\%.

Depending upon the question asked, equation 1 may be rearranged into several useful forms, as shown below.

\[
(1 - R)^n = P \tag{1}
\]

Taking the \( n \)-th root on both sides of the equality yields:

\[
1 - R = P^{1/n}. \tag{2}
\]

Solving for \( R \):

\[
R = 1 - P^{1/n}, \tag{3}
\]

or

\[
R = 1 - \sqrt[n]{P}, \tag{4}
\]

which is the correct equation for the letter of Benefiel et al. using his notation for the \( n \)-th root of \( P \). The form of equation 3 or 4 is useful for calculating, in the general population, a condition's expected rate of occurrence corresponding to a desired \( P \) value and the number \( n \) of consecutive negative observations made.

Another useful form of equation 1 is obtained as follows:

\[
(1 - R)^n = P. \tag{1}
\]

Taking the logarithm on both sides of the equality

\[
n \cdot \log (1 - R) = \log P. \tag{5}
\]

Solving for \( n \)

\[
n = \frac{\log P}{\log (1 - R)}. \tag{5}
\]

The form of equation 5 enables calculation of the number of consecutive negative observations (i.e., condition absent) needed to confirm a known or assumed rate of occurrence \( R \) at a chosen \( P \) value.

To one who has not used these relationships quantitatively, it is surprising to learn the number of consecutive negative observations needed to confirm or infer a low rate of occurrence in the population of a condition under scrutiny. To familiarize the reader with the relation between the rate of occurrence \( R \) and the number \( n \) of consecutive negative observations, table 1 presents several milestone values of \( R \) and the corresponding values of \( n \) for the commonly-used \( P = 0.05 \) (confidence level 0.95).

For example, if we adopt a new procedure or administer a drug, and we wish to show that the rate of untoward effects is not greater than 10\%, we need to observe 28 consecutive cases where the undesirable side effects have not occurred. Observation of 13 consecutive cases without an occurrence only permits the inference of a maximum rate of 20\% in the general population of patients. And, as Benefiel et

<table>
<thead>
<tr>
<th>( P = 0.05 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
</tr>
<tr>
<td>0.20 or 20%</td>
</tr>
<tr>
<td>0.15 or 15%</td>
</tr>
<tr>
<td>0.10 or 10%</td>
</tr>
<tr>
<td>0.05 or 5%</td>
</tr>
</tbody>
</table>