An Estimate of the Odds Ratio That Always Exists

Michael Parzen, Stuart Lipsitz, Joseph Ibrahim, and Neil Klar

This article proposes an estimate of the odds ratio in a $(2 \times 2)$ table obtained from studies in which the row totals are fixed by design, such as a phase II clinical trial. Our estimate, based on the median unbiased estimate of the probabilities of success in the $(2 \times 2)$ table, will always be in the interval $(0, \infty)$. Another estimate of the odds ratio which has such properties is obtained when adding .5 to each cell of the table. Using simulations, we compared our proposed estimate to that obtained by adding .5 to every cell, and found that our estimate had smaller finite sample bias, and larger mean square error. We also propose the use of the bootstrap to form a confidence interval for the odds ratio based on our proposed estimate. Instead of a Monte Carlo bootstrap, one can easily calculate the “exact” bootstrap distribution of our estimate of the odds ratio, and use this distribution to calculate confidence intervals.

Key Words: Median unbiased estimator; Phase II clinical trials; Small samples.

1. INTRODUCTION

Phase II cancer clinical trials are designed to determine if a new treatment produces favorable results (proportion of success), when compared to a known, “standard treatment.” For a given subject, the outcome of the phase II trial is success or failure. If the new treatment produces favorable results, then further testing will be done in a phase III study, in which patients will be randomized to the new treatment or the “industry standard.” Often, instead of one new promising treatment, there are two new promising treatments. In an effort to reduce the time necessary to determine if either or both of the new treatments are effective, a randomized phase II trial is often conducted. In the randomized phase II trial, patients are randomized to receive one of the two new therapies. The data can be arranged in a $(2 \times 2)$
Table 1. Data for a Randomized Phase II Clinical Trial

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Outcome</th>
<th>Success</th>
<th>Failure</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Y₁</td>
<td>n₁ - Y₁</td>
<td>n₁</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Y₂</td>
<td>n₂ - Y₂</td>
<td>n₂</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Y₁ + Y₂</td>
<td>(n₁ + n₂) - (Y₁ + Y₂)</td>
<td>n₁ + n₂</td>
<td></td>
</tr>
</tbody>
</table>

table as in Table 1, with the rows representing treatment, and the columns representing the outcome (success or failure). In Table 1, there are n₁ subjects on treatment 1, with Y₁ successes, and n₂ subjects on treatment 2, with Y₂ successes.

Because phase II trials are pilot studies, most have small samples, many with sample sizes of less than 20 subjects per treatment. Thus, the sample sizes in phase II randomized studies are not large enough to estimate with precision the odds ratio for success between the two treatments. Furthermore, the nature of the cancer is such that the probability of success is often very small, even for promising new therapies. Nevertheless, even though the sample sizes and the probability of successes on each treatment arm may be small, in a final report on the trial, investigators usually report an estimate of this odds ratio between the two treatments. With such small sample sizes and probabilities of success, such an estimate of the odds ratio will have a large standard error, and thus the investigator will get only a rough idea of the value of the true odds ratio. Furthermore, besides estimation of the odds ratio, investigators are also interested in obtaining a confidence interval to get a range of possible values for the odds ratio.

Our example is a randomized phase II clinical for the evaluation of two new chemotherapy treatments in patients with advanced large bowel cancer. The two treatments were Homoharringtonine and Caracemide. The clinical trial was developed by the United States’ Eastern Cooperative Oncology Group (ECOG), and open for patient accrual from 1987 through 1990. In this clinical trial, the investigators were interested in two binary outcomes for each treatment. The first binary outcome of interest is the “tumor shrinkage,” with the “successful” outcome defined as the tumor shrinking at least 50%; a successful outcome in this context is called a “response.” The second binary outcome of interest is the toxicity, or side effects of the treatment, with the “successful” outcome defined as “life-threatening” toxicity. In order to present a complete picture of the effectiveness of a treatment, one must look at both the “tumor shrinkage” and “toxicity.” If, for example, the tumor shrinks in all subjects, but all subjects also die as a result of toxicity, then the treatment will not warrant further study in a phase III trial. Thus, we are interested in estimating the odds ratio for response and well as the odds ratio for toxicity.

Prior to the trial, the commonly accepted treatment was 5-Flurouracil, which gave a response rate of 15–25% on previous trials (David 1982; Heal and Schein 1977). Homoharringtonine is a cephalotoxine ester from the bark of an evergreen tree in China. It was used mostly for its anti-leukemia properties in China, and was also shown to have anti-leukemia properties in phase II trials. Caracemide is a water soluble compound which functions as a
nonspecific inhibitor of macromolecular synthesis. The study entered 25 patients: \(n_1 = 14\) on the Homoharringtonine arm and \(n_2 = 11\) on the Caracemide arm. The accrual goal was 30 patients on each arm, but the study was terminated early after being open for 43 months due to poor accrual. Unfortunately, there were no responses on either arm, that is, \(Y_1 = Y_2 = 0\). Furthermore, there were \(Y_1 = 2\) subjects with life-threatening toxicities on the Homoharringtonine arm and \(Y_2 = 1\) subject with life-threatening toxicities on the Caracemide arm. Thus, the probability of a life-threatening toxicity appears small, but not as small as the probability of a response. Thus, this dataset illustrates the estimation of the odds ratio with 0 cells, as well as the estimation of the odds ratio with small, but no zero cells.

The usual maximum likelihood estimate of the odds ratio from a \((2 \times 2)\) table,

\[
\hat{\text{OR}}_{\text{MLE}} = \frac{Y_1(n_2 - Y_2)}{Y_2(n_1 - Y_1)},
\]

(1.1)
can equal 0 or \(\infty\) if one of the cells in Table 1 is 0 (\(\hat{\text{OR}}_{\text{MLE}} = 0\) if the 0 cell is in the numerator, and \(\hat{\text{OR}}_{\text{MLE}} = \infty\) if the 0 cell is in the denominator of (1.1)); if there is a 0 in both the numerator and denominator of (1.1), then \(\hat{\text{OR}}_{\text{MLE}}\) is undefined. Then, to estimate the odds ratio with such data, Haldane (1955) and Gart and Zweifel (1967) suggested adding .5 to each cell and the calculation of the odds ratio from the resulting \((2 \times 2)\) table, which gives the modified maximum likelihood estimate (MMLE)

\[
\hat{\text{OR}}_{\text{MMLE}} = \frac{(Y_1 + .5)(n_2 - Y_2 + .5)}{(Y_2 + .5)(n_1 - Y_1 + .5)}.
\]

(1.2)
Haldane (1955) and Gart and Zweifel (1967) showed that it has the smallest first-order finite sample bias of any estimate of the form,

\[
\hat{\text{OR}}_{\text{MMLE}}(c) = \frac{(Y_1 + c)(n_2 - Y_2 + c)}{(Y_2 + c)(n_1 - Y_1 + c)},
\]

for any constant \(c\). Agresti (1999) compared many different methods for obtaining confidence intervals for the odds ratio with small samples, and found that a confidence interval based on the Gart and Zweifel (1967) method had good coverage when the true odds ratio is less than 4. Here, we are interested in both a point estimate that always exists, and a confidence interval that has endpoints in the interval \((0, 1)\). Of the confidence intervals considered by Agresti (1999), only the confidence interval based on (1.2) has the property that the endpoints will be between 0 and \(\infty\). Even though (1.2) is between 0 and \(\infty\), regardless of the number of 0’s in the \((2 \times 2)\) table, Bishop, Fienberg, and Holland (1975) and Agresti and Yang (1987) discouraged adding .5 to each cell, because of the appearance of adding “fake data.” Several similar alternatives to this modified maximum likelihood estimator have been proposed. For instance Walter and Cook (1991) and Whaley (1991) considered adding .5 only to those cells having zero observations. Any such modification to the maximum likelihood estimator used to ensure that the estimated odds ratio is between 0 and \(\infty\), again has the appearance of adding “fake data,” and is discouraged. Note, when
Estimate of the odds ratio

423

both $Y_1$ and $Y_2$ equal 0, the estimate of the odds ratio in (1.2) equals

$$\hat{OR}_{\text{MLE}} = \frac{(n_2 + .5)}{(n_1 + .5)},$$

(1.3)

so that, basically, the estimate of the odds ratio depends on the ratio of the sample sizes in the two groups. Intuitively, suppose $n_2$ is much larger than $n_1$, then, we know the probability of success in group 2 is likely to be closer to 0 than the probability of success in group 1, so that we would expect the estimate of the odds ratio to be large; in this case (1.3) would be large. Thus, when both $Y_1 = Y_2 = 0$, one can get a rough estimate of the odds ratio using (1.2) or the estimate proposed in this article.

Alternatively, for small samples, a popular approach is to obtain estimates and confidence intervals from a conditional likelihood, which is formed by conditioning on the column totals of the $(2 \times 2)$ table (by design of the clinical trial, the number on each treatment, i.e., the row totals, are assumed fixed). With both margins of the $(2 \times 2)$ fixed, the table then follows a noncentral hypergeometric distribution which is a function only of the odds ratio. Using this conditional noncentral hypergeometric distribution, one can obtain a conditional maximum likelihood estimate of the odds ratio (Breslow and Day 1990) and an exact conditional confidence interval (Breslow and Day 1990); both the conditional maximum likelihood estimate and the exact conditional confidence interval are implemented in the software StatXact (StatXact-4 for Windows, 1998) or LogXact (LogXact for Windows, 1996). Unfortunately, if $Y_1$ equals 0 or $n_1$ or $Y_2$ equals 0 or $n_2$, then the conditional maximum likelihood estimate is undefined, and one of the endpoints of the confidence interval will be 0 or $\infty$. Another possibility is a median unbiased estimate (MUE) of the odds ratio (Hirji, Tsiatis, and Mehta 1989), which is also obtained from the conditional noncentral hypergeometric distribution, and also implemented in LogXact. Unfortunately, this median unbiased estimate of the odds ratio from the conditional noncentral hypergeometric distribution still has the problem of being undefined for certain tables; if $Y_1 = n_1$ and $Y_2 = n_2$ or $Y_1 = Y_2 = 0$, then the MUE is undefined. For the Eastern Cooperative Oncology Group bowel cancer study discussed earlier, since $Y_1 = Y_2 = 0$, the MUE of the odds ratio is undefined. This article proposes an estimate of the odds ratio that exists even if $Y_1 = 0$ and/or $Y_2 = 0$

For our proposed estimate, we first obtain the median unbiased estimate of the probabilities of success for each of the two rows (treatments) of the $(2 \times 2)$ table. Basically, we have a binomial distribution for each row of the $(2 \times 2)$ table, and we obtain the MUE of the success probability for each row. The median unbiased estimate of the probability of success is always in the interval (0, 1), even if there are 0 or $n_t$ successes for treatment $t$. We then estimate the odds ratio as the ratio of the estimated odds of success for these two probabilities. Since the MUE of the probability of success for the two treatments will always be between 0 and 1, the estimated odds ratio will always be between 0 and $\infty$. Our estimate of the odds ratio differs from the MUE of Hirji et al. (1989) in that they use the conditional noncentral hypergeometric distribution to get the MUE of the odds ratio, whereas we use the two binomial distributions to obtain the MUE of the two success probabilities, and then we form the estimated odds ratio from these two estimated success probabilities.
We also propose the use of the bootstrap to form a confidence interval for the odds ratio based on our proposed estimate. Instead of a Monte Carlo bootstrap, one can easily calculate the “exact” bootstrap distribution of our estimate of odds ratio, and use this distribution to calculate confidence intervals. This confidence interval will always have endpoints that are finite. This is opposed to the exact conditional confidence interval for the odds ratio based on the exact noncentral hypergeometric distribution (Breslow and Day 1990), which can have endpoints which are 0 or $\infty$. For example, for the ECOG study with outcome “response,” the exact conditional confidence interval is $[0, \infty]$.

2. THE MEDIAN UNBIASED ESTIMATOR

As discussed by Read (1985) and Hirji et al. (1989), the median unbiased estimator (MUE), $\hat{\theta}$, of a parameter $\theta$, satisfies

$$\Pr(\hat{\theta} \leq \theta) \geq .5 \quad \text{and} \quad \Pr(\hat{\theta} \geq \theta) \geq .5.$$  

For discrete data, a range of values usually satisfy this criterion, and one usually takes the midpoint as the median unbiased estimate. First, we obtain the MUE of $p_t$, the probability of success on treatment $t$, $t = 1, 2$, which we denote $\hat{p}_t$. Then, our estimate of the odds ratio is the modified MUE

$$\tilde{OR}_{MUE} = \frac{\hat{p}_1(1 - \hat{p}_2)}{\hat{p}_2(1 - \hat{p}_1)}.$$  

To obtain $\hat{p}_t$, we use the distribution of $Y_t$, which is binomial

$$\Pr(Y_t = y_t | p_t) = \binom{n_t}{y_t} p_t^{y_t} (1 - p_t)^{n_t - y_t}. \quad (2.1)$$

As discussed by Hirji et al. (1989), for binary data problems, the MUE can be computed from the distribution of the sufficient statistics for the parameters. For binomial data, the sufficient statistic is $Y_t$. In particular, we compute the MUE of $p_t$ as follows. First, we compute the values $\hat{p}_t^L$ and $\hat{p}_t^U$ to be those values of $p_t$ for which

$$\Pr(Y_t \geq y_t | p_t = \hat{p}_t^L) \geq .5 \quad \text{and} \quad \Pr(Y_t \leq y_t | p_t = \hat{p}_t^U) \geq .5. \quad (2.2)$$

Here, $\hat{p}_t^L$ and $\hat{p}_t^U$ are the smallest (denoted $L$ for lower) and largest (denoted $U$ for upper) values of $p_t$, respectively, that satisfy (2.2). The MUE is defined as

$$\hat{p}_t = (\hat{p}_t^L + \hat{p}_t^U)/2.$$  

When $0 < Y_t < n_t$, we can actually find values $\hat{p}_t^L$ and $\hat{p}_t^U$ which satisfy

$$\Pr(Y_t \geq y_t | p_t = \hat{p}_t^L) = \Pr(Y_t \leq y_t | p_t = \hat{p}_t^U) = .5.$$  

The lower value, $\hat{p}_t^L$, is obtained by solving,

$$\frac{1}{2} = \Pr(Y_t \geq y_t | p_t = \hat{p}_t^L) = \sum_{j=y_t}^{n_t} \binom{n}{j} (\hat{p}_t^L)^j (1 - \hat{p}_t^L)^{n_t-j},$$

$$= \sum_{j=y_t}^{n_t} \binom{n}{j} (\hat{p}_t^L)^j (1 - \hat{p}_t^L)^{n_t-j},$$

$$= \binom{n}{y_t} (\hat{p}_t^L)^{y_t} (1 - \hat{p}_t^L)^{n_t-y_t} + \sum_{j=y_t+1}^{n_t} \binom{n}{j} (\hat{p}_t^L)^j (1 - \hat{p}_t^L)^{n_t-j}.$$
and the upper value, $\tilde{p}_t^U$, is obtained by solving

$$
.5 = \Pr(Y_t \leq y_t | p_t = \tilde{p}_t^U)
$$

$$
= \sum_{j=0}^{y_t} \binom{n_t}{j} (\tilde{p}_t^U)^j (1 - \tilde{p}_t^U)^{n_t-j}.
$$

The values $\tilde{p}_t^L$ and $\tilde{p}_t^U$ can actually be obtained noniteratively by using the following relationship between the cumulative beta-distribution function and the cumulative binomial distribution function (Daly 1992; Johnson and Kotz 1969). Let $F(p_t | \alpha, \beta)$ be the cumulative distribution function of the beta-distribution,

$$
F(p_t | \alpha, \beta) = \int_0^{p_t} \frac{\Gamma(\alpha + \beta)u^{\alpha-1}(1-u)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}
$$

$$
= \sum_{j=\alpha}^{n_t} \binom{n_t}{j} (p_t)^j (1 - p_t)^{n_t-j},
$$

where $\Gamma(\cdot)$ is the gamma-function, and $\alpha$ and $\beta = n_t - \alpha + 1$ are integers. Thus, we need to find $\tilde{p}_t^L$ such that

$$
F(\tilde{p}_t^L | \alpha = y_t, \beta = n_t - y_t + 1) = .5
$$

and $\tilde{p}_t^U$ such that

$$
F(\tilde{p}_t^U | \alpha = y_t + 1, \beta = n_t - y_t + 2) = .5.
$$

In particular,

$$
\tilde{p}_t^L = F^{-1}(.5 | \alpha = y_t, \beta = n_t - y_t + 1)
$$

and

$$
\tilde{p}_t^U = F^{-1}(.5 | \alpha = y_t + 1, \beta = n_t - y_t + 2),
$$

where $F^{-1}(Q | \alpha, \beta)$ is the $Q$th quantile of the beta-distribution with parameters $\alpha$ and $\beta$. The $Q$th quantile of the beta-distribution can be obtained in most statistical packages.

Now suppose $y_t = 0$, then

$$
\Pr(Y_t \geq y_t | p_t = \tilde{p}_t^L) = \Pr(Y_t \geq 0 | p_t = \tilde{p}_t^L) = 1.
$$

Then, any value of $\tilde{p}_t^L$ in the interval $[0, 1]$ satisfies

$$
\Pr(Y_t \geq 0 | p_t = \tilde{p}_t^L) \geq .5;
$$

here, $\tilde{p}_t^L = 0$, the smallest possible value of $\tilde{p}_t^L$. Also, if $y_t = 0$, $\tilde{p}_t^U$ satisfies

$$
\Pr(Y_t \leq y_t | p = \tilde{p}_t^U) = \Pr(Y_t = 0 | p = \tilde{p}_t^U) = (1 - \tilde{p}_t^U)^n = .5.
$$
or
\[ \tilde{p}^U_t = 1 - .5^{(1/n)}. \]

Then, when \( y_t = 0 \), \( \tilde{p}_t \) equals
\[ \tilde{p}_t = (\tilde{p}^L_t + \tilde{p}^U_t)/2 = (1 - .5^{(1/n)})/2. \]

Using similar calculations, if \( y_t = n \), \( \tilde{p}^U_t = 1, \tilde{p}^L_t = .5^{(1/n)} \), and \( \tilde{p}_t \) equals
\[ \tilde{p}_t = (\tilde{p}^L_t + \tilde{p}^U_t)/2 = [.5^{(1/n)} + 1]/2. \]

Looking at the outcome “response” for the Eastern Cooperative Oncology Group bowel cancer study discussed earlier, in which \((Y_1, Y_2) = (0, 0)\) and \((n_1, n_2) = (14, 11)\), our proposed estimate of the odds ratio is \( \hat{OR}_{MMUE} = .787 \), and the estimate of the odds ratio after adding .5 to every cell is \( \hat{OR}_{MMLE} = .793 \). These estimates are very similar. For the outcome “toxicity,” in which \((Y_1, Y_2) = (2, 1)\) and \((n_1, n_2) = (14, 11)\), our proposed estimate of the odds ratio is \( \hat{OR}_{MMUE} = 1.53 \), and the estimate of the odds ratio after adding .5 to every cell is \( \hat{OR}_{MMLE} = 1.40 \). Again, these estimates are similar. Section 4 compares the finite sample properties of these two estimators via simulation.

### 3. Bootstrap Confidence Interval

This section describes a bootstrap confidence interval for the odds ratio based on the bootstrap distribution of \( \hat{OR}_{MMUE} \). To do this, we describe the “exact” parametric bootstrap here. We call it “exact” because we can enumerate the full bootstrap distribution, and no Monte Carlo approaches are needed.

Consider the \( t \)th treatment group, with binomial distribution given in (2.1). We pose a “parametric” bootstrap, which replaces \( p_t \) in (2.1) with the MUE \( \tilde{p}_t \) to get
\[
p(y_t|\tilde{p}_t) = \text{pr}(Y_t = y_t|p_t = \tilde{p}_t) = \begin{pmatrix} n_t \\ y_t \end{pmatrix} \tilde{p}^{y_t}_t (1-\tilde{p}_t)^{n_t-y_t}. \tag{3.1}
\]

A parametric bootstrap sample is defined as one random binomial sample from (3.1), which we denote \( y^*_t \). If we draw a random bootstrap sample \( y^*_1 \) from \( p(y_1|\tilde{p}_1) \) and a random bootstrap sample \( y^*_2 \) from \( p(y_2|\tilde{p}_2) \), we can obtain the median unbiased estimates \( \tilde{p}^*_1 \) and \( \tilde{p}^*_2 \), and thus the modified median unbiased estimate of the odds ratio \( \hat{OR}^*_{MMUE} \). Note, if we had replaced \( p_t \) with \( y_t/n_t \) in (3.1), then we get the nonparametric bootstrap of Efron (1982), but this causes problems when \( y_t = 0 \) or \( y_t = n_t \).

To calculate the “exact” parametric bootstrap distribution of \( \hat{OR}^*_{MMUE} \), for each point \((y^*_1, y^*_2)\) in the sample space of
\[
p(y_1, y_2|\tilde{p}_1, \tilde{p}_2) = \begin{pmatrix} n_1 \\ y_1 \end{pmatrix} \begin{pmatrix} n_2 \\ y_2 \end{pmatrix} \tilde{p}^{y_1}_{1} (1-\tilde{p}_1)^{n_1-y_1} \tilde{p}^{y_2}_{2} (1-\tilde{p}_2)^{n_2-y_2}, \tag{3.2}
\]
one calculates the MMUE of the odds ratio, \( \hat{OR}^*_{MMUE}(y^*_1, y^*_2) \), and its associated probability given by \( p(y^*_1, y^*_2|\tilde{p}_1, \tilde{p}_2) \). The \( \hat{OR}^*_{MMUE}(y^*_1, y^*_2) \)'s and their associated probabilities
Table 2. Bootstrap Distribution for $\tilde{\text{OR}}_{\text{MMUE}}$ from ECOG Data

<table>
<thead>
<tr>
<th>Sample space point</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$\tilde{\text{OR}}_{\text{MMUE}}^*$</th>
<th>Probability</th>
<th>Cumulative probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>7</td>
<td>0.0144</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>11</td>
<td>0.0178</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>6</td>
<td>0.0207</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>10</td>
<td>0.0209</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>9</td>
<td>0.0214</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
<td>11</td>
<td>0.0238</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>5</td>
<td>0.0295</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
<td>11</td>
<td>0.0315</td>
<td>0.00000</td>
<td>0.00001</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>10</td>
<td>0.0332</td>
<td>0.00000</td>
<td>0.00001</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>8</td>
<td>0.0350</td>
<td>0.00000</td>
<td>0.00001</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td>11</td>
<td>0.0417</td>
<td>0.00000</td>
<td>0.00001</td>
</tr>
<tr>
<td>140</td>
<td>12</td>
<td>5</td>
<td>6.6701</td>
<td>0.00000</td>
<td>0.99701</td>
</tr>
<tr>
<td>141</td>
<td>9</td>
<td>2</td>
<td>7.4682</td>
<td>0.00000</td>
<td>0.99701</td>
</tr>
<tr>
<td>142</td>
<td>7</td>
<td>1</td>
<td>8.5679</td>
<td>0.00000</td>
<td>0.99701</td>
</tr>
<tr>
<td>143</td>
<td>3</td>
<td>0</td>
<td>9.0219</td>
<td>0.00279</td>
<td>0.99980</td>
</tr>
<tr>
<td>144</td>
<td>11</td>
<td>3</td>
<td>9.5855</td>
<td>0.00000</td>
<td>0.99980</td>
</tr>
<tr>
<td>145</td>
<td>13</td>
<td>6</td>
<td>9.3000</td>
<td>0.00000</td>
<td>0.99980</td>
</tr>
<tr>
<td>146</td>
<td>14</td>
<td>9</td>
<td>9.5931</td>
<td>0.00000</td>
<td>0.99980</td>
</tr>
<tr>
<td>147</td>
<td>12</td>
<td>4</td>
<td>9.6063</td>
<td>0.00000</td>
<td>0.99980</td>
</tr>
<tr>
<td>148</td>
<td>10</td>
<td>2</td>
<td>10.2657</td>
<td>0.00000</td>
<td>0.99980</td>
</tr>
<tr>
<td>149</td>
<td>8</td>
<td>1</td>
<td>11.3438</td>
<td>0.00000</td>
<td>0.99980</td>
</tr>
<tr>
<td>150</td>
<td>4</td>
<td>0</td>
<td>13.0261</td>
<td>0.00019</td>
<td>0.99999</td>
</tr>
<tr>
<td>151</td>
<td>13</td>
<td>5</td>
<td>13.2417</td>
<td>0.00000</td>
<td>0.99999</td>
</tr>
<tr>
<td>152</td>
<td>12</td>
<td>3</td>
<td>14.3886</td>
<td>0.00000</td>
<td>0.99999</td>
</tr>
<tr>
<td>153</td>
<td>11</td>
<td>2</td>
<td>14.8219</td>
<td>0.00000</td>
<td>0.99999</td>
</tr>
<tr>
<td>154</td>
<td>9</td>
<td>1</td>
<td>15.1924</td>
<td>0.00000</td>
<td>0.99999</td>
</tr>
<tr>
<td>155</td>
<td>14</td>
<td>8</td>
<td>15.6966</td>
<td>0.00000</td>
<td>0.99999</td>
</tr>
<tr>
<td>156</td>
<td>5</td>
<td>0</td>
<td>17.9056</td>
<td>0.00001</td>
<td>1.00000</td>
</tr>
<tr>
<td>157</td>
<td>13</td>
<td>4</td>
<td>19.0708</td>
<td>0.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>158</td>
<td>10</td>
<td>1</td>
<td>20.8834</td>
<td>0.00000</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

comprise the bootstrap distribution. Since $y_1^*$ can take on $(n_1 + 1)$ values and $y_2^*$ can take on $(n_2 + 1)$ values, one will need to calculate $(n_1 + 1)(n_2 + 1)$ probabilities and MMUE's in order to calculate the bootstrap distribution. For example, if $n_1 = n_2 = 100$, one will need to calculate 10,201 probabilities. Even though one would probably use maximum likelihood to obtain a confidence interval with the large sample sizes of $n_1 = n_2 = 100$, it is still very feasible to do the bootstrap for $n_1 = n_2 = 100$ even on most old Pentium PC's. For the Eastern Cooperative Oncology Group bowel cancer study discussed earlier, in which $n_1 = 14$ and $n_2 = 11$, we only need to calculate 180 probabilities and MMUE's.

To obtain a two-sided $100(1 - \alpha)%$ confidence interval, one then orders the $(n_1 + 1)(n_2 + 1)$ values of $\tilde{\text{OR}}_{\text{MMUE}}^*(y_1^*, y_2^*)$, calculates the bootstrap distribution function, and from this bootstrap distribution function calculates the $(100\alpha/2)$th and $[100(1 - \alpha/2)]$th percentiles. Table 2 gives a subset of the bootstrap distribution for $\tilde{\text{OR}}_{\text{MMUE}}^*$ for the ECOG clinical trials example for the outcome “response” with $n_1 = 14$, $n_2 = 11$, $Y_1 = Y_2 = 0$, $\tilde{p}_1 = 0.0242$ and $\tilde{p}_2 = 0.03055$. The Appendix gives SAS commands for calculating the bootstrap distribution in Table 2.

Since the bootstrap distribution is discrete, it is very unlikely that the $P$th percentile will
occur at a point of the discrete distribution. As such, we suggest to use linear interpolation to calculate the percentiles. For example, suppose one wants to calculate the 2.5th percentile. In Table 2, the closest percentiles to the 2.5th percentile are the $P_U = 0.032$ percentile ($\tilde{\text{OR}}_U = 0.1042$) and the $P_L = 0.277$ percentile ($\tilde{\text{OR}}_L = 0.1041$). Then, using linear interpolation, we estimate the 2.5th percentile to be

$$
\tilde{\text{OR}}_{2.5} = \frac{\tilde{\text{OR}}_L(P_U - 0.025) + \tilde{\text{OR}}_U(0.025 - P_L)}{(P_U - P_L)}.
$$

In general, if one wants to calculate the $P$th percentile, and $P_U$ is the closest percentile greater than $P$ (with the corresponding value of the odds ratio equal to $\tilde{\text{OR}}_U$) and $P_L$ is the closest percentile less than $P$ (with the corresponding value of the odds ratio equal to $\tilde{\text{OR}}_L$), then the $P$th percentile obtained using linear interpolation is

$$
\tilde{\text{OR}}_P = \frac{\tilde{\text{OR}}_L(P_U - P) + \tilde{\text{OR}}_U(P - P_L)}{(P_U - P_L)}.
$$

Using linear interpolation, a 95% confidence interval is the 2.5th and 97.5th percentiles from the bootstrap distribution from Table 2, or $[0.10421, 5.59886]$.

Instead of using linear interpolation, another method to calculate a 100$(1 - \alpha)$ confidence interval is to choose the closest observed percentile that is less than 100$\alpha/2$ and the closest observed percentile that is greater than 100$(1 - \alpha)/2$. In the ECOG example, we get $[0.10410, 5.6799]$. We call this the conservative method.

After using linear interpolation or the conservative method, one can refine the confidence interval further using the “bias-corrected” percentile method (Efron 1982). In the simulations in the following section, the conservative method and/or the “bias-corrected” method did little to change the coverage, so we present the linear interpolation method without using the bias correction. For the ECOG clinical trial with outcome “toxicity” with $(Y_1, Y_2) = (2, 1)$ and $(n_1, n_2) = (14, 11)$, the 95% confidence interval using linear interpolation is $[0.10419, 12.4456]$.

4. SIMULATIONS AND DISCUSSION

Using the results of Read (1985) and Hirji et al. (1989), the MUE of $p_t$ is consistent, asymptotically fully efficient, and asymptotically normal with mean $p_t$. Since $\tilde{\text{OR}}_{\text{MMUE}}$ is a continuous function of $\tilde{p}_1$ and $\tilde{p}_2$, it is consistent, asymptotically fully efficient, and has an asymptotic normal distribution with mean equal to the true odd ratio. The estimate $\tilde{\text{OR}}_{\text{MMLE}}$ has the same asymptotic properties as the MLE in (1.1), that is, it is also consistent, asymptotically fully efficient, and has an asymptotic normal distribution with mean equal to the true odd ratio. Thus, both the MMLE and MMUE have the same asymptotic properties. However, we are also interested in comparing the finite sample bias and the finite sample mean square error of these estimators. This is important since our proposed estimator is a viable alternative to $\tilde{\text{OR}}_{\text{MMLE}}$ that does not use “fake” data, and which exists for all tables of the form given in Table 1.
We consider the scenario in which $p_2 = .05$, and $p_1$ varies from .05 to .35, giving odds ratios that range from 1 to 10.23, and log-odds ratios from 0 to 2.23. To look at a scale from $-\infty$ to $+\infty$ instead of 0 to $+\infty$, we considered the bias and mean square error for estimating the log-odds ratio (instead of the odds ratio), as estimated by the log of $\text{OR}_{\text{MMLE}}$ and the log of $\text{OR}_{\text{MMUE}}$. Here, we just report results for $n_1 = n_2 = 10$; as $n_t$ gets larger, the bias and mean square error decrease, and are not reported here. Figure 1 gives the simulation bias, calculated as the average over the simulations minus the true value. Figure 2 gives the mean square error, calculated as the average of the squared bias over all simulation runs. To form Figures 1 and 2, we performed 15,000 simulations at each of .01 increments of $p_1$ from .05 to .35.

From Figure 1, we see that, since the biases are negative, both $\text{OR}_{\text{MMUE}}$ and $\text{OR}_{\text{MMLE}}$ tend to underestimate the true odds ratio. However, the bias is uniformly smaller for $\text{OR}_{\text{MMUE}}$ for the whole interval, with the difference in the bias of $\text{OR}_{\text{MMUE}}$ and $\text{OR}_{\text{MMLE}}$ increasing as the log-odds ratio gets larger, especially with a log-odds ratio greater than .5. On the other hand, from Figure 2, the mean square error is uniformly smaller for $\text{OR}_{\text{MMLE}}$. In Figure 2, the smallest value of the ratio of the mean square error of $\text{OR}_{\text{MMLE}}$ to the mean square error of $\text{OR}_{\text{MMUE}}$ is .8. This indicates that the mean square error is not much larger for $\text{OR}_{\text{MMUE}}$ as compared to $\text{OR}_{\text{MMLE}}$. In simulations not reported here, using the same configuration as Table 2, except with $n_1 = n_2 = 20,000$, we get 0 bias and identical variance for

![Figure 1. Plot of simulation bias versus log-odds ratio, $\log \frac{p_2/(1-p_2)}{p_1/(1-p_1)}$, with $p_1 = .05$ and $n_1 = n_2 = 10$.](image-url)
both $\hat{\text{OR}}_{\text{MMUE}}$ and $\hat{\text{OR}}_{\text{MMLE}}$, agreeing with the asymptotic results of consistency and full efficiency for both.

Using the same simulations as above, Table 3 gives the coverage probabilities (the percentage of confidence intervals that contain the true value) and the average lengths of the 95% confidence intervals. The confidence interval using the MMLE is the estimate plus or minus 1.96 times the estimated standard error, that is,

$$\text{exp}\left\{ \log \frac{(Y_1 + .5)(n_2 - Y_2 + .5)}{(Y_2 + .5)(n_1 - Y_1 + .5)} \right\} \pm 1.96 \sqrt{\frac{1}{Y_1 + .5} + \frac{1}{n_1 - Y_1 + .5} + \frac{1}{Y_2 + .5} + \frac{1}{n_2 - Y_2 + .5}}.$$ 

The confidence intervals using both the MMLE and the bootstrapped MMUE are conservative. The MMLE is more conservative for small values of the log-odds ratio, and the bootstrapped MMUE is slightly more conservative for larger values of the log-odds ratio. The average lengths of the confidence intervals are much smaller for the bootstrapped MMUE. Since the coverage is adequate using the bootstrapped MMUE, and the length of the confidence interval much smaller, we suggest using the bootstrapped MMUE. Note, neither the confidence interval based on the usual MLE nor the exact confidence interval based on the conditional noncentral hypergeometric distribution (Breslow and Day 1990)
Table 3. Coverage Probabilities and Average Lengths of 95% Confidence Intervals From Simulations

<table>
<thead>
<tr>
<th>Log odds-ratio</th>
<th>MMUE</th>
<th>MMLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>99.7(^a)</td>
<td>99.9</td>
</tr>
<tr>
<td></td>
<td>9.8(^b)</td>
<td>31.2</td>
</tr>
<tr>
<td>0.19</td>
<td>96.9</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>10.6</td>
<td>34.4</td>
</tr>
<tr>
<td>0.36</td>
<td>97.8</td>
<td>99.9</td>
</tr>
<tr>
<td></td>
<td>12.2</td>
<td>37.2</td>
</tr>
<tr>
<td>0.50</td>
<td>97.0</td>
<td>99.8</td>
</tr>
<tr>
<td></td>
<td>13.9</td>
<td>42.1</td>
</tr>
<tr>
<td>0.63</td>
<td>96.7</td>
<td>99.2</td>
</tr>
<tr>
<td></td>
<td>15.1</td>
<td>45.3</td>
</tr>
<tr>
<td>0.75</td>
<td>97.7</td>
<td>99.0</td>
</tr>
<tr>
<td></td>
<td>16.2</td>
<td>46.1</td>
</tr>
<tr>
<td>1.21</td>
<td>97.2</td>
<td>98.7</td>
</tr>
<tr>
<td></td>
<td>23.6</td>
<td>65.8</td>
</tr>
<tr>
<td>1.56</td>
<td>97.4</td>
<td>97.2</td>
</tr>
<tr>
<td></td>
<td>32.0</td>
<td>90.8</td>
</tr>
<tr>
<td>1.85</td>
<td>96.8</td>
<td>95.3</td>
</tr>
<tr>
<td></td>
<td>41.5</td>
<td>119.1</td>
</tr>
<tr>
<td>2.10</td>
<td>96.8</td>
<td>95.8</td>
</tr>
<tr>
<td></td>
<td>50.7</td>
<td>140.8</td>
</tr>
<tr>
<td>2.33</td>
<td>96.9</td>
<td>95.3</td>
</tr>
<tr>
<td></td>
<td>63.2</td>
<td>173.0</td>
</tr>
</tbody>
</table>

\(^a\) the first entry in each cell is the coverage probability
\(^b\) the second entry in each cell is the average length of the confidence interval.

have finite endpoints if at least one cell in Table 1 is 0. We note that other possibilities for confidence intervals for the odds ratio are given in Agresti (1999), but all of these confidence intervals can have endpoints that equal 0 or \(\infty\), so that we did not consider them here.

The MMUE proposed here performs well with respect to bias in small samples, and is an alternative to adding .5 to every cell in the table, a practice that some investigators discourage. In closing, it is also possible to define other “hybrid” estimators that may also have good properties, and may deserve further exploration. For example, one could use the median unbiased estimate of the odds ratio proposed by Hirji et al. (1989) when it is finite, and our MMUE when Hirji, et al.’s estimate is infinite. Even further, one could use the MLE when it is finite, Hirji et al.’s estimate when the MLE is 0, \(\infty\), or undefined, and our MMUE when Hirji’s estimate is 0, \(\infty\), or undefined. This “hybrid” estimate may have the smallest variance of any estimate that is always finite without having to add .5 to every cell. This is an avenue for future research. Another avenue for further research is extending our technique to estimation of a common odds ratio from a set of \((2 \times 2)\) tables when each \((2 \times 2)\) table has small samples; early research on this topic is given in McKinlay (1975) and McKinlay (1978).
APPENDIX: SAS COMMANDS FOR OUTPUT IN TABLE 2

data or;
  input y1 n1 y2 n2;
cards;
0 14 0 11
;
/* FIND MMUE */

data or(keep=or_mue p1_mue p2_mue n1 n2);
  set or;
/* calculate MUE for p1 */

if (y1=n1) then do;
  p_1u = 1;
end;
else do;
  F_u = (y1+1)/(n1-y1) * FINV(.5,2*(y1+1),2*(n1-y1));
  p_1u = F_u/(1 + F_u);
end;

if (y1=0) then do;
  p_1ll = 0;
end;
else do;
  F_l = (n1-y1+1)/y1* FINV(.5,2*(n1-y1+1),2*y1);
  p_1ll = 1/(1+F_l);
end;

p1_mue= .5*(p_1ll+p_1u);
/* calculate MUE for p2 */

if (y2=n2) then do;
  p_2u = 1;
end;
else do;
  F_u = (y2+1)/(n2-y2) * FINV(.5,2*(y2+1),2*(n2-y2));
  p_2u = F_u/(1 + F_u);
end;
if (y2=0) then do;
  p_2l = 0;
end;
else do;
  F_1 = (n2\cdot y2+1)/y2\cdot \text{FINV}(.5,2\cdot (n2\cdot y2+1),2\cdot y2 )
  p_2l = 1/(1+F_1);
end;

p2_mue= .5\cdot (p_2l+p_2u);

or_mue = p1_mue\cdot (1-p2_mue)/( p2_mue\cdot (1-p1_mue) ) ; /* MMUE */

run;

proc print;
  var or_mue ;
run;

/* Creating Table 2 */

data or ;
  set or;

/* Create all possible values of (y1,y2) for bootstrap */

do y1 = 0 to n1;
  do y2 = 0 to n2;
    output;
  end;
end;

run;

/* Find bootstrap probabilities and MMUE's */

data or;
  set or;

/* Binomial probabilities for y1 */

if y1>0 then L_1=probnnml(p1_mue,n1,y1) - probnnml(p1_mue,n1,y1-1) ;
if y1=0 then L_1 = probbnml(p1_mue, n1, y1);
/* Binomial probabilities for y2 */
if y2>0 then L_2 = probbnml(p2_mue, n2, y2) * probbnml(p2_mue, n2, y2-1);
if y2=0 then L_2 = probbnml(p2_mue, n2, y2);
prob = L_1*L_2; /* Bootstrap probabilities */
/* Calculate MMUE's for bootstrap */
if (y1=n1) then do;
  p_1u_j = 1;
end;
else do;
  F_u = (y1+1)/(n1-y1) * FINV(.5, 2*(y1+1), 2*(n1-y1));
  p_1u_j = F_u/(1 + F_u);
end;
if (y1=0) then do;
  p_1l_j = 0;
end;
else do;
  F_l = (n1-y1+1)/y1 * FINV(.5, 2*(n1-y1+1), 2*y1);
  p_1l_j = 1/(1+F_l);
end;

p1_mue_j = .5*(p_1l_j + p_1u_j);

if (y2=n2) then do;
  p_2u_j = 1;
end;
else do;
  F_u = (y2+1)/(n2-y2) * FINV(.5, 2*(y2+1), 2*(n2-y2));
  p_2u_j = F_u/(1 + F_u);
end;
if (y2=0) then do;
  p_2l_j = 0;
end;
else do;
  F_l = (n2-y2+1)/y2 * FINV(.5, 2*(n2-y2+1), 2*y2);
\[
p_{2l_j} = \frac{1}{1+F_{l1}}; \quad \text{end};
\]

\[
p_{2\text{mue}_j} = 0.5(p_{2l_j} + p_{2u_j});
\]

/* Bootstrap MMUE*/

\[
or_{\text{mue}_j} = p_{1\text{mue}_j}*(1-p_{2\text{mue}_j})/(p_{2\text{mue}_j}*(1-p_{1\text{mue}_j})
\]

run;

proc sort data=or;
  by or_mue_j;
run;

data or;
  set or;
  cum + prob; /* cumulative probabilities */
run;

proc print;
  var y1 y2 or_mue_j prob cum ;
run;

ACKNOWLEDGMENTS

We are grateful for the support provided by the following grants from the United States' Institutes of Health: CA 57253, CA 55576, CA 70101, CA 23318, and CA 74015.

[Received August 1999. Revised March 2001.]

REFERENCES


Efron, B. (1982), *The Jackknife, the Bootstrap, and Other Resampling Plans*, SIAM, monograph 38, CBMS–NSF.


