Informative allocation and consistent treatment selection

Tamás Rudas

Department of Statistics, Faculty of Social Sciences
Eötvös Loránd University, Budapest

Abstract

When data in the form of a $2 \times 2$ treatment by response table are available, the better out of the two treatments is often selected using the odds ratio (cross product ratio). Such decisions do not depend on the allocation of the observations in the different treatment categories and may exhibit a counterintuitive reversal property, called Simpson’s paradox. In cases when those receiving different treatments are potentially different, as in observational studies or in designed experiments with dropout or noncompliance, decisions taking into account the difference in observed allocations may be useful. Using an approach that postulates certain desirable properties of decision functions and derives further characteristics from these, a new decision procedure based on the cross sum ratio is investigated. This procedure is not only sensitive to allocation but also turns out to be the only selection procedure that avoids Simpson’s paradox. In addition to these logical advantages, the probability of wrong decision when using the cross sum ratio tends to be smaller than when using the cross product ratio. The application of the new decision procedure is illustrated by the re-analysis of data sets, some of which exhibit Simpson’s paradox when analyzed using the cross product ratio. Finally, generalizations of the decision procedures to $2 \times J$ decision tables are considered.

Key words: Cross product ratio, Cross sum ratio, Decision functions, Simpson’s paradox.

1. Introduction

This paper deals with the comparison of two treatments, $Tr_1$ and $Tr_2$, based on data in the form of a $2 \times 2$ contingency table or decision table $T$:
Such data arise in a variety of contexts as the result of different data collection procedures, including designed experiments (clinical trials), case-control studies (Dawson & Trapp, 2004) and observational studies (Rosenbaum, 1995). The goal of the analysis is to evaluate the treatments relative to each other, either in the context of causal decision theory or that of evidential decision theory (see, e.g., Meek & Glymour (1994), Curley & Browne (2001)). From a practical perspective, several diagnostic (Goldstein & Simel, 2005) and treatment decisions (Davies et al, 1998) in medicine and also other policy decisions (see, e.g., Wainer & Brown (2004)) are based on the analysis of data as in T.

The comparison is often based on the value of the odds ratio (Rudas, 1998) or cross product ratio (cpr)

\[
\frac{a/b}{c/d} = \frac{ad}{bc}
\]

and on the value of the relative risk (rr), or risk ratio that, depending on the context, is also called success rate (see, e.g., Dawson & Trapp (2004)),

\[
\frac{a}{a+b} \frac{c}{c+d}
\]

and deems Tr1 better than Tr2 if the cpr or the rr is greater than 1.

A formal discussion of treatment comparisons may be given by using decision functions. For all tables T, these functions are defined to yield 1 or -1 or 0, telling which treatment is better (the first one, the second one, none of them). For example, a decision function based on the cpr may defined as

\[
CPR = \text{sgn}(\log\left(\frac{ad}{bc}\right))
\]

where sgn is the sign function. Two such decision functions \(\gamma_1\) ans \(\gamma_2\) are considered equal, if for all tables \(T\), \(\gamma_1(T) = \gamma_2(T)\). Therefore, the decision functions are, in fact, equivalence classes and the same decision functions may be represented by different functional forms. This is similar to considering the functions \(f(x) = x\) and \(g(x) = (x^3)^{1/3}\) equal. For example, it is easy to see that the \(CPR\) is the same decision function as sgn(log(rr)). In
Section 2, six basic properties of such decision functions are postulated and all possible decision functions with these properties are investigated.

The developments in this paper are motivated by two properties of decisions based on the CPR, which may be seen as problematic, at least in certain situations. The first such property is illustrated by the fact that it selects the same treatment (namely $Tr1$) based on the following tables:

<table>
<thead>
<tr>
<th>$T_1$</th>
<th>Response</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Tr1$</td>
<td>450</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>$Tr2$</td>
<td>280</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T_2$</th>
<th>Response</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Tr1$</td>
<td>90</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>$Tr2$</td>
<td>360</td>
<td>150</td>
<td></td>
</tr>
</tbody>
</table>

The two sets of data are identical with respect to the conditional distribution of response, given treatment, but they are different with respect to the treatment marginals, that is, with respect to allocation. In other words, $cpr$ and $CPR$ are invariant against changes in allocation, if these changes leave the conditional distribution of response, given treatment, unaffected (see (4)).

Whether arriving at the same conclusion from these two sets of data is appropriate or not, depends on whether the difference in allocation is the result of the action of the experimenter, as in a designed experiment, and in this case reaching the same conclusion seems justified based on an exchangeability argument (Lindley & Novick, 1981), or rather is the result of some procedure outside of the control of the experimenter, as in an observational study, and in this case it is not obvious whether reaching the same conclusion is justified or not. In the former case, allocation into treatment categories is noninformative and in the latter case it is, at least potentially, informative. This difference needs to be taken into account when choosing the decision function. Note, that even in the case of an originally noninformative allocation, as in a clinical trial, different dropout rates (Diggle et al, 2002) and different compliance rates (Diggle et al, 2002; Schechtman, 2000) often make the observed treatment marginal informative.

The second characteristic of decisions based on the CPR is that the so-called Simpson’s paradox (Blyth, 1972; Lindley & Novick, 1981; Wagner, 1982; Meek & Glymour, 1994; Aldrich, 1995; Pearl, 2000) may occur. Simpson’s paradox is said to have occurred, when one treatment is better in the data than the other one, but if the data are split according to another variable, the other treatment appears better in both subgroups (briefly referred to as reversal):

$$CPR(T_1) = CPR(T_2) \neq CPR(T_1 + T_2).$$
where the sum of two tables is the table with the sum of the individual frequencies.

It is proved in Section 2 that the two issues related to the use of the CPR to select a treatment are, in fact, related. Simpson’s paradox may occur with any decision function that is invariant against changes in allocation (including also the risk ratio).

The main formal result of the paper is that there exists only one consistent (see (3)) decision function, that is such that Simpson’s paradox cannot occur with it, and this is

\[ CSR = \text{sgn}(\log(\frac{a + d}{b + c})) \]

This decision function is based on the cross sum ratio

\[ csr = \frac{a + d}{b + c} \]

The main message of the paper is that when allocation is potentially informative, a decision function, like the CSR that is sensitive to allocation may be preferable to the CPR. Obviously, there are data sets where the CPR and the CSR lead to different conclusions. For example, with the above data, CPR(T_2) = 1 and CSR(T_2) = -1, a conclusion that may appear surprising to one’s intuition – an intuition that has been educated using the CPR. The decision based on the CSR reflects the fact that in T_2, Tr2 was beneficial to 450 subjects in excess of those to whom it was detrimental, while the same quantity is only 70 for Tr1, because the CSR depends on how \( a - b \) compares to \( c - d \). In fact, for \( g_i = (g_{i1}, g_{i2}) \), and \( \delta(g_i) = g_{i1} - g_{i2} \),

\[ \text{sgn}(\delta(g_1) - \delta(g_2)) = CSR \text{ if } g_1 = (a, b) \text{ and } g_2 = (c, d) \] (1)

and

\[ \text{sgn}(\delta(g_1) - \delta(g_2)) = CPR \text{ if } g_1 = (\frac{a}{a + b}, \frac{b}{a + b}) \text{ and } g_2 = (\frac{c}{c + d}, \frac{d}{c + d}) \] (2)

That is, the CSR and the CPR are the same function applied to the observed conditional frequency distributions or to the observed conditional probability distributions in the decision table. The difference between the two decisions is exactly whether or not the \( a + b \) and \( c + d \) marginal distributions, that is the allocations into treatment categories, are taken into account.
It has to be emphasized again, that the CSR is investigated here as a decision function and it is a simple representative of an equivalence class of decision functions. Therefore, the csr is not considered a measure of effect strength and issues related to the cpr in that context (see, e.g., Craemer (2004), Rudas & Bergsma (2004)) need not be considered.

Section 3 deals with sampling properties of the CSR, in particular the probability of wrong decision based on a sample. In this comparison, the CSR presents itself as a more reliable decision function than the CPR. Further, confidence intervals for the true value of the CSR are proposed.

Section 4 discusses real examples when CPR ≠ CSR, including the Berkeley admissions data (Bickel et al, 1975), data concerning medical school applications (Wainer & Brown, 2004) and data regarding administering penicillin to children with meningococcal disease (Harnden et al, 2006). The analysis based on the CSR removes the paradoxical elements present in the published analyses of these data sets and gives new insight into the relationships among the variables involved. The practice of using the net approval rate in opinion polls is also discussed and is shown to be equivalent to using the CSR.

Section 5 points to possible extensions of the CPR and of the CSR beyond 2 × 2 decision tables. When comparing several treatments, both the CPR and the CSR are transitive. Extensions of the CSR and of the CPR based on (1) and on (2) to situations with more than two different responses show that Simpson’s paradox may occur with the generalized CPR but the generalized CSR remains free from it.

Section 6 concludes the paper with a brief discussion.

2. Decision functions

This section gives an axiomatic treatment of decision functions. Obvious properties that decision functions have to possess are postulated. Based on these properties, the relationship between invariance against changes in allocation and the possibility of Simpson’s paradox to occur, and the roles played by the CPR and the CSR among decision functions will be clarified. To simplify presentation, all cell entries will be assumed to be positive.

The first two properties describe noninformative sets of data, that is data that imply that no treatment is preferable to the other. A data set is not informative for treatment selection if the number of positive and negative responses is the same within both treatments.

Property 1
\[ \gamma \left( \begin{array}{cc} a & a \\ b & b \end{array} \right) = 0. \]

A data set is also not informative if the same frequencies are observed in the response categories of both treatments.

**Property 2**

\[ \gamma \left( \begin{array}{cc} a & b \\ a & b \end{array} \right) = 0. \]

The next two properties postulate simple antisymmetries: if the rows are interchanged, then the decision changes to the opposite,

**Property 3**

\[ \gamma \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) = -\gamma \left( \begin{array}{cc} c & d \\ a & b \end{array} \right), \]

and if the positive and negative responses are interchanged, then the decision changes to its opposite.

**Property 4**

\[ \gamma \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) = -\gamma \left( \begin{array}{cc} b & a \\ d & c \end{array} \right). \]

The last two properties describe data sets that imply that the first treatment is better than the second one. This is the case if \( Tr_1 \) has more positive than negative responses and \( Tr_2 \) does not have more positive than negative responses,

**Property 5**

\[ a > b, \ c \leq d \Rightarrow \gamma \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) = 1, \]

and this is also the case when the two treatments have the same number of negative responses but \( Tr_1 \) has more positive responses than \( Tr_2 \):

**Property 6**

\[ a > c, \ d = b \Rightarrow \gamma \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) = 1. \]

In what follows, only decision functions with the above properties will be considered. Three further properties, not necessarily possessed by a decision function, will play a central role in the analysis of decision functions. The first one is related to combining data sets and means that Simpson’s paradox cannot occur.
Consistency

if \( \gamma(T_1) = \gamma(T_2) \), then \( \gamma(T_1 + T_2) = \gamma(T_1) \). \hspace{1cm} (3)

The second one addresses a related issue, namely the combination of decisive data (that is, data where one treatment appears to be better than the other one) with noninformative data in Properties 1 and 2.

Indifference

If \( \gamma(T) = 0 \) by Property 1 or Property 2, then \( \gamma(T_1 + T) = \gamma(T_1) \), for all \( T_1 \).

Note that indifference does not assume that adding any \( T \) with \( \gamma(T) = 0 \) leaves the decision unchanged.

An important aspect of the argument in this paper is the distinction between decision functions that are and that are not sensitive to changes in allocation into treatment categories.

Invariance against changes in allocation

\[
\gamma \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) = \gamma \left( \begin{array}{cc} ta & tb \\ uc & ud \end{array} \right)
\] \hspace{1cm} (4)

for every table and all positive \( t \) and \( u \).

The following result may be verified directly.

**Proposition 1.** Properties 1 to 6 hold true for both CPR and CSR. Further, CPR is not consistent and not indifferent but is invariant against changes in allocation, CSR is consistent and indifferent but is not invariant against changes in allocation.

**Proposition 2.** If a decision function is invariant against changes in allocation, then it is equal to the CPR.

The proofs of the results of the paper are given in the Appendix.

Propositions 1 and 2 readily imply

**Proposition 3.** If a decision function is consistent, then it cannot be invariant against changes in allocation.

The variation independence of the odds ratio from the marginals of the two-way contingency table, which is a very desirable property of it when used as a measure of association (see, e. g., Rudas (1998)), turns out to be the ‘reason’ that Simpson’s paradox may occur, when used as a decision function.

**Proposition 4.** Any indifferent decision function is equal to the CSR.
Propositions 1 and 4 imply

**Proposition 5.** If \( \gamma \) is indifferent then it is also consistent.

The next important result is

**Proposition 6.** Any consistent decision function is equal to the CSR.

The foregoing results are summarized in

**Proposition 7.** The following three statements are equivalent

(a) \( \gamma \) is consistent
(b) \( \gamma \) is indifferent
(c) \( \gamma = CSR \).

Indifference and consistency are equivalent and every decision function with these properties is equal to the CSR, if Properties 1 – 6 are assumed to hold. Invariance against changes in allocation, on the other hand, implies that the reversal may occur and every such decision function is equal to the CPR. However, if there are equal allocations into the two treatment categories of \( T \), then the CSR and the CPR reach the same conclusion.

**Proposition 8.** If \( a + b = c + d \), then \( CSR(T) = CPR(T) \).

To complete this section, we note that the indifference property used so far implies a more general indifference.

**Proposition 9.** If \( \gamma \) is indifferent, then \( \gamma(T_1) = 0 \) implies that

\[
\gamma(T_1 + T_2) = \gamma(T_2)
\]

for all \( T_2 \).

3. Statistical properties of the **CSR and the CPR**

Because the CSR is proposed here as a decision function, out of its statistical properties, the probability of wrong decision is of central interest. A wrong decision is to choose \( Tr1 \) (\( Tr2 \)) based on the observed value of the CSR when its population value is equal to -1 or 0 (+1 or 0). This probability is easily evaluated under multinomial sampling. Suppose that the treatment by response table in the population has the following distribution

<table>
<thead>
<tr>
<th>( p_a )</th>
<th>( p_b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_c )</td>
<td>( p_d )</td>
</tr>
</tbody>
</table>

8
and the relevant random variables behind observations \(a, b, c, d\) are \(A, B, C, D\), respectively. When \((A, B, C, D)\) has a multinomial joint distribution with parameters \(n = a + b + c + d\) and \((p_a, p_b, p_c, p_d)\), then \(A + D \sim B(n, p_a + p_d)\).

**Proposition 10.** The probability of wrong decision under multinomial sampling with the CSR if \(p_a + p_d > p_b + p_c\) is

\[
\Phi\left(\frac{1}{2} \sqrt{n} \frac{1 - \text{csr}}{\text{csr}}\right) + o(1),
\]

where \(\Phi\) is the standard normal distribution function.

Exact probabilities of wrong decision are given for selected values of \(n\) and of \(p_a + p_d\) in Table 1. As one would expect, the probability of choosing \(Tr2\), when \(Tr1\) is better, reduces with increasing values of the \(\text{csr}\) and with increasing sample sizes. For values of \(p_a + p_d\), or of the \(\text{csr}\), larger than those reported in Table 3, the probability of wrong decision is zero for the sample sizes investigated. The same table may be used to read off the probability of wrong decision when \(\text{CSR} = -1\): one simply has to use the entry \(1/\text{csr}\) in these cases.

*** Table 1 around here ***

Threshold values of \(\text{csr}\), for different sample sizes, such that if the true \(\text{csr}\) is greater than the respective threshold value, the probability of wrong decision (concluding from the sample that \(\text{CSR} = -1\) or \(\text{CSR} = 0\)) is less than 0.05, are given in Table 2. The reciprocals of these values are similar thresholds in cases when \(\text{csr} < 1\).

*** Table 2 around here ***

Comparable results are difficult to obtain for the CPR, mainly because its variance, even asymptotically that could be used to calculate an approximation, does not only depend on the value of the \(\text{cpr}\), but rather on the individual cell probabilities.

For purposes of comparison, for each sample size investigated, 1000 2×2 tables were generated that had the \(\text{csr}\) values given in Table 2. These tables had uniformly distributed \(p_a\) and \(p_b\) values and the \(p_c\) and \(p_d\) values were implied. Then the probability of wrong decision was approximated using the asymptotic distribution of the logarithm of the \(\text{cpr}\) (see e.g., Agresti (2002) for a review of the large sample properties of the \(\text{cpr}\)). In each case, the probability of wrong decision was determined with reference to
the correct decision according to the CPR which, in some cases, was the opposite of the decision suggested by the CSR. Table 3 reports the average approximate error probabilities. Note that the average is taken here as a simple descriptive statistic and no assumption concerning a superpopulation of 2×2 tables is made. The average error probabilities for the CPR were considerably higher than those for the CSR but the advantage of the CSR seemed to decrease with increasing sample size. These results indicate that decisions based on the CSR are more stable in the statistical sense than those based on the CPR.

*** Table 3 around here ***

Because the CSR takes on three values only, the construction of confidence intervals simplifies to the question whether or not a confidence interval with a given level of confidence contains any value of the CSR, other than the observed one.

**Proposition 11.** If CSR = 1 is observed under multinomial sampling, an asymptotic 1 − α level confidence interval for the true value of CSR does not contain 0 or -1 if

\[ 1 - \Phi\left(\frac{2(a + b) - n}{\sqrt{n}}\right) < \alpha. \] (5)

\[ \square \]

When the value of CSR = 1 is observed in the data, one can say that a 95\% confidence interval does not contain 0 or -1, if the csr is such that the left hand side of (5) is less than 0.05. Threshold values of the csr, such that the left hand side of (5) is equal to 0.05, for various values of n, are given in Table 4.

*** Table 4 around here ***

When the sampling scheme specifies the row marginals of the decision table, allocation in the treatment categories is not informative and the use of the CPR is not suggested. When the column marginals are fixed, say \( N_+ \) and \( N_- \), as in a case-control study, and sampling is product multinomial, the probability of wrong decision with the CSR if \( p_a + p_d > p_b + p_c \) is

\[
\sum_{0 \leq A \leq N_+} \left( \binom{N_+}{A} p_a^A p_c^{N_+-A} \right) \sum_{A+(N_- - N_+)/2 \leq B \leq N_-} \left( \binom{N_-}{B} p_b^B p_d^{N_- - B} \right),
\]
where $\lfloor \cdot \rfloor$ means integer part. Using $X = A - C$ and $Y = B - D$, similarly to Proposition 10, a normal approximation of this probability may be obtained. **Proposition 12.** Under product multinomial sampling, with fixed $N_+\text{ and } N_-$, the probability of wrong decision with the CSR, when $p_a + p_d > p_b + p_c$, is

$$\Phi\left(\frac{E_Y - E_X}{\sqrt{V_X + V_Y}}\right) + o(1),$$

where

$$E_X = N_+(p_a - p_c), \quad E_Y = N_-(p_b - p_d)$$

$$V_X = 4N_+ \frac{p_a}{p_a + p_c} \frac{p_c}{p_a + p_c}$$

$$V_Y = 4N_- \frac{p_b}{p_b + p_d} \frac{p_d}{p_b + p_d}$$


4. Applications of the CSR

In this section we revisit three data sets where inconsistent findings using the CPR have been discussed in the literature and compare the standard analysis to that of obtained by using the CSR, and also discuss the practice of using the net approval rates in opinion polls that is closely related to the CSR.

4.1. Berkeley admissions

The first example we consider is the famous Berkeley admissions data (Bickel et al, 1975). Here, the gender by admission status table shows that male applicants had a higher chance of being admitted to the graduate school than female applicants had, and this observation raised the question of gender discrimination. When the data are studied at the departmental level, it turns out that female applicants had an advantage in all departments where there was a considerable difference between the admission rates. The marginal admission rate was favorable to male applicants because they applied in larger fractions to departments where the overall admission rate was high, while female applicants tended to apply to departments with lower overall admission rates. Because admission decisions were made at the departmental level, the suspicion of gender discrimination was refuted. This analysis is widely accepted today and is also reported in textbooks (see, e.g., Freedman et al (1998)).
The foregoing analysis used, essentially, the CPR and gave an explanation of the paradox that relies on different allocations. In other words, an analysis that is not sensitive to allocation, suggests that allocation should be taken into account. If so, the analysis based on the CPR is not appropriate, rather, an analysis using the CSR, which is sensitive to allocation, is needed. Table 5 contains the values of these decision functions for the six largest departments based on the data as reported in Freedman et al (1998).

The analysis based on the CSR shows that in four out of the six departments and in both totals, men have an advantage, including Department A, where the CPR found the largest female advantage. In the remaining two departments (D and F) women have an advantage according to the CSR, although a 95% level confidence interval for the true value of the CSR contains 1 (that is, male advantage). The CSR shows an overwhelming male advantage in the entire admission procedure, including application and selection. Although there was no male advantage when selection was looked at only using the CPR, the fact that men tended to apply to programs with high admission rates (or programs where men applied tended to have high admission rates), points to a male advantage indicated by the CSR that is sensitive to allocation and, therefore, can incorporate the effects of both application and selection in the admission process. Of course, whether or not this advantage has a discriminatory nature, cannot be decided based on the data only.

4.2. Medical school applications

Wainer & Brown (2004) reported medical school application percentages for MCAT Biological Sciences test takers with various scores. In all score groups, black test takers applied to medical schools in larger fractions than white test takers did. In spite of this, among all test takers, whites applied to medical school in larger fraction than blacks did. Again, the paradox goes away when the CSR is used to decide whether blacks or whites tend to apply more typically to medical school. For test results up to 7, blacks tend to apply to medical school and for scores higher than 7, white test takers apply more typically to medical school, and the same is true when the test scores are disregarded. The cpr and csr values are given in Table 6.
4.3. Administering penicillin to meningococcal disease patients

The effect of administering parenteral penicillin to children with meningococcal disease (MC) before admission to hospital has been the subject of considerable debate. Harnden et al (2006) reported a $cpr$ of 5.96 indicating a sizable association between administering penicillin and death for children with MC diagnosed prior to hospital admission. Although they warned that it might very well be the case that severity of the disease increased the chances of administering penicillin and the increased odds of death among those who received penicillin might be a consequence of this selection, a British Medical Journal editorial, based on the $CPR$, went as far as saying that 'Prehospital penicillin for MC may be harmful...'. Other contributions supported the practice of administering penicillin (York, 2006). In a commentary (Perera, 2006), the contributing statistician described that the odds ratio obtained depended strongly on whether all children with MC or only those for whom the GP diagnosed the disease were taken into account. For the former group the odds ratio of death as opposed to survival, for those who did versus those who did not receive penicillin, was 0.86. Thus, overall, there appeared to be a small protective effect but for those with a diagnosed MC, there was harm associated with administering penicillin, a conclusion that seemed somewhat paradoxical. Given the relatively limited information about covariates, in conclusion, Perera (2006) warned that 'strong associations are not necessarily causal'.

The two data sets (Perera, 2008) are reported here as Tables 7 and 8. They are examples of the situation when $CSR=-CPR$. The data came essentially from an observational study. Allocation is informative, not only concerning the numbers treated but also because of the best judgment applied by the GP's. Consequently, the application of the $CPR$ may not be appropriate. For children with MC diagnosed before hospital admission, $csr=0.79$ ($CSR=-1$, not significantly different from $CSR=0$ or 1). For all children with MC, $csr=1.73$ ($CSR=1$, significantly different from $CSR=0$ or -1).

*** Table 7 around here ***

*** Table 8 around here ***

For children with MC diagnosed before hospital admission, the $CSR$ suggests a beneficial effect of penicillin, in conformity with expectation, however this effect is weak. For all children, the $CSR$ suggests harm associated with penicillin. The conclusion based on the $CSR$ confirms the intuition of Harnden et al (2006), who assumed that those who did receive penicillin could
be more seriously ill than those who did not receive penicillin. However, as suggested by the CSR, this applies not within the group of those with an early MC diagnosis but rather to all children who later turned out to have MC. Our results, based on the CSR, support the practice of administering penicillin to children who are diagnosed with MC.

4.4. Net ratings in opinion polls

It is very common in the polling industry to use net ratings (or net approval rates) to compare the public’s view on two politicians (e.g., two candidates for an office). The net rating is the difference between the percentage of those who say a certain characteristic applies minus the percentage of those who say it does not apply to the politician (see, e.g., Gallup & Newport (2006)).

If the following data are observed for a given characteristic

<table>
<thead>
<tr>
<th>Response</th>
<th>Applies</th>
<th>Does Not Apply</th>
<th>No Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politician A</td>
<td>a</td>
<td>b</td>
<td>e</td>
</tr>
<tr>
<td>Politician B</td>
<td>c</td>
<td>d</td>
<td>f</td>
</tr>
</tbody>
</table>

then the net approval rates of the two politicians are $a - b$ and $c - d$, respectively. A comparison of the two politicians’ approvals based on the net approval rates is exactly the same as their comparison based on the CSR in the following table:

<table>
<thead>
<tr>
<th>Response</th>
<th>Applies</th>
<th>Does Not Apply</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politician A</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Politician B</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Note, that in the first table $a + b + e = b + c + f$ (the same sample is interviewed about the politicians) but in the second table $a + b \neq c + d$ (because of differences in the numbers of people who know them). The use of the net ratings is suggested to reduce or eliminate the effect of different levels of name recognition that is, different allocations of those who have an opinion. On the other hand, there seems to be no trace in the published literature of the ratio of positive to negative ratings being used as a means of comparing the overall approval rates, which would be parallel to a decision based on the CPR.

5. More general decision tables

A reviewer asked whether any of the above results can be extended beyond $2 \times 2$ tables. This section gives a brief discussion of such possibilities.
First note that Simpson’s paradox is a lack of collapsibility property in $2 \times 2 \times 2$, or more generally, $K \times 2 \times 2$ decision tables, where the two treatments are observed under 2 (or $K$) different scenarios. Here, decisions based on the CSR are collapsible in the sense that if the same treatment is better in each of the 2 (or $K$) conditions, then the same treatment is also better overall. This is not always true for decisions based on the CPR.

When several treatments are compared in an $I \times 2$ decision table, both the CSR and the CPR lead to decisions that are transitive. More precisely, it is easy to see that

**Proposition 13.** Consider three treatments $T_1, T_2, T_3$ and decision tables $T_{ij}$ comparing $T_i$ to $T_j$. For $\gamma = CSR$ and $\gamma = CPR$, if $\gamma(T_{12}) = 1$ and $\gamma(T_{23}) = 1$, then $\gamma(T_{13}) = 1$, that is if $T_1$ is better than $T_2$ and $T_2$ is better than $T_3$, then $T_1$ is also better than $T_3$.

When two treatments are compared using polytomous outcomes, like categories very good, good, bad, very bad, an extension of the decision procedure from $2 \times 2$ tables to $2 \times J$ tables is needed.

If in the $2 \times J$ decision table the observed response frequencies of $T_i$ are $f_i = (f_{i1}, \ldots, f_{ij})$, in a descending order of desirability, define $\delta_j, j = 1, \ldots, J - 1$, for an arbitrary nonnegative vector $v = (v_1, \ldots, v_J)$ of dimension $J$, as

$$\delta_j(v) = (v_1 + \ldots + v_j) - (v_{j+1} + \ldots + v_J),$$

so $\delta_j(f_i)$ is the difference of the number of responses in the $j$ best categories and of the number of responses in the $J - j$ worst response categories for treatment $i$. A generalization of the CSR (see (1)) is obtained by considering $T_1$ better than $T_2$ if

$$\text{sgn}(\delta_j(f_1) - \delta_j(f_2)) = 1 \text{ for all } j = 1, \ldots, J - 1,$$

and a generalization of the CSR (see (2)) is obtained by considering $T_1$ better than $T_2$ if

$$\text{sgn}(\delta_j(f_{1+}) - \delta_j(f_{2+})) = 1 \text{ for all } i = 1, \ldots, J - 1,$$

where $f_{i+}$ is the sum of all observations for $T_i$. Again, the generalizations of the CSR and of the CPR are based on the same function but the CSR uses the observed conditional frequencies and the CPR uses the observed conditional probabilities, just like in (1) and (2). Thus, the CSR remains sensitive to allocation and the CPR remains not sensitive to allocation in
these generalized forms. Also, Simpson’s paradox may occur with the
generalized CPR and does not occur with the generalized CSR.

The generalizations suggested above are related to the concept of stochas-
tic dominance (of responses to $Tr_1$ versus those to $Tr_2$) but this aspect is
not going to be pursued in this paper. For $Tr_1$ to be better than $Tr_2$, these
generalizations require the same relation to hold in all $2 \times 2$ decision tables
that are obtained from the $2 \times J$ decision table by a monotone combina-
tion of responses into two groups. Whether or not this is an appropriate
definition, depends on the actual decision problem, of course.

6. Discussion

This paper considered the problem of selecting the better one out of two
treatments. Such decisions are mostly obtained by using functions of the
data closely related to the odds ratio. The possibility or the actual occur-
rence of Simpson’s paradox in such decisions has been the subject of much
scholarly work, some explaining under what conditions the paradox may
or may not occur, some making suggestions for choosing the better treat-
ment if it did occur. This paper takes the position that a real evaluation of
various decision procedures is not possible unless at least minimal require-
ments for the behavior of decision procedures are postulated. Based on six
such 'axioms', various decision procedures were investigated. It was argued
that when allocation into the treatment categories is informative (observa-
tional studies or designed experiments with dropout or noncompliance), it
is desirable to use decision functions which are sensitive to allocation. In
addition, it turned out that all decision procedures which are not sensitive
to allocation may exhibit Simpson’s paradox, so a good way to avoid it is to
use decision procedures which are sensitive to allocation, at least when the
allocation is informative. The main results of the paper included that all
decision procedures which are not sensitive to allocation always reach the
same decision as the CPR and all decision procedures which never exhibit
Simpson’s paradox always reach the same conclusion as the CSR does. The
paper also discussed statistical properties of the CSR and suggested possible
generalizations to more complex decision problems.

The illustrative applications of the CSR to data that had been ana-
lyzed by the CSR showed that the conclusions are sometimes different. Our
statistical education relies strongly on the CPR and we all may feel the
conclusions reached by the CSR counter-intuitive in some cases. The fre-
quent controversies around decisions reached by the CPR should suggest to
consider using, in the case of informative allocation, the CSR that is a logically sounder decision procedure. However, the CSR and the CPR are not that different: they are based on the same function applied to conditional frequency or to conditional probability distributions of the responses in the two treatment categories.

7. Acknowledgments

The author is indebted to Antonio Forcina, Adrian Raftery and Joe Shafer for helpful discussions and to Rafael Perera for kindly allowing to use the data in Tables 7 and 8. Part of this research was done while the author was a visitor in the Center for Statistics and the Social Sciences and the Department of Statistics, University of Washington, where he is now an Affiliate Professor. The author is also a Recurrent Visiting Professor at the Central European University and the moral support received is acknowledged.

References


Appendix

Proof of Proposition 2. Invariance against changes in allocation implies that
\[
\gamma\left(\begin{array}{cc}
a & b \\
c & d \\
\end{array}\right) = \gamma\left(\begin{array}{cc}
a & b \\
\frac{b}{c} & \frac{b}{d} \\
\end{array}\right) = \gamma\left(\begin{array}{cc}
a & b \\
\frac{b}{c} & b \\
\end{array}\right)
\]
and by Property 6, if \(CPR(T) = 1\), then \(\gamma(T) = 1\), and by Property 3, if \(CPR(T) = -1\) then \(\gamma(T) = -1\). Finally, by Property 2, if \(CPR(T) = 0\) then \(\gamma(T) = 0\).

Proof of Proposition 4. One has to see that for indifferent decision functions \(\gamma\)
(a) if \(a + d > b + c\) then \(\gamma(T) = 1\)
(b) if \(a + d < b + c\) then \(\gamma(T) = -1\)
(c) if \(a + d = b + c\) then \(\gamma(T) = 0\)

Part (a): If \(a = b\) then \(d > c\) and because of Property 5
\[
\gamma\left(\begin{array}{cc}
d & c \\
b & a \\
\end{array}\right) = 1
\]
and Properties 3 and 4 imply that
\[
\gamma\left(\begin{array}{cc}
a & b \\
c & d \\
\end{array}\right) = 1.
\]

If \(a > b\) then if \(d \geq c\), Property 5 implies the required result. If \(c > d\), then \(a - b > 0\) and \(c - d > 0\), therefore \(T = T_1 + T_2\) with
\[
T_1 = \frac{a - b + \frac{1}{2}\min(b, d)}{c - d + \frac{1}{2}\min(b, d)} \quad \frac{\frac{1}{2}\min(b, d)}{\frac{1}{2}\min(b, d)} 
\]
\[
T_2 = \frac{b - \frac{1}{2}\min(b, d)}{d - \frac{1}{2}\min(b, d)} \quad \frac{\frac{1}{2}\min(b, d)}{\frac{1}{2}\min(b, d)}
\]
and \(\gamma(T_1) = 1\) because of Property 6, for the condition in (a) may be written as \(a - b > c - d\), and \(\gamma(T_2) = 0\) because of Property 1. Then indifference implies that \(\gamma(T) = 1\).
If, finally, $a < b$ then $d > c$ and for the table

\[
\begin{array}{cc}
d & c \\
b & a \\
\end{array}
\]

one has the same situation as for the

\[
\begin{array}{cc}
a & b \\
c & d \\
\end{array}
\]

table in the last case and, therefore

\[
\gamma \left( \begin{array}{cc} d & c \\ b & a \end{array} \right) = 1
\]

and Properties 3 and 4 imply that $\gamma(T) = 1$ also in this case.

Part (b): Apply part (a) to

\[
\begin{array}{cc}
c & d \\
a & b \\
\end{array}
\]

to obtain that if $a + d < b + c$ then for any indifferent $\gamma$,

\[
\gamma \left( \begin{array}{cc} c & d \\ a & b \end{array} \right) = 1,
\]

and then Property 3 implies the required result.

Part (c): In this case $a - b = c - d$. If $a \geq b$ then $T = T_1 + T_2$ with

\[
T_1 = \begin{array}{c} a - b + \frac{1}{2}\min(b,d) \\ c - d + \frac{1}{2}\min(b,d) \end{array} \quad T_2 = \begin{array}{c} b - \frac{1}{2}\min(b,d) \\ d - \frac{1}{2}\min(b,d) \end{array}
\]

and $\gamma(T_1) = 0$ because of Property 2 and $\gamma(T_2) = 0$ because of Property 1. Then indifference implies the required result.

If $a < b$, then $T = T_1 + T_2$ with

\[
T_1 = \begin{array}{c} \frac{1}{2}\min(a,c) \\ \frac{1}{2}\min(a,c) \end{array} \quad T_2 = \begin{array}{c} a - \frac{1}{2}\min(a,c) \\ c - \frac{1}{2}\min(a,c) \end{array}
\]

and $\gamma(T_1) = 0$ because of Property 2 and $\gamma(T_2) = 0$ because of Property 1. Then indifference implies the required result.

□

Proof of Proposition 6. We have to prove that for consistent decision functions $\gamma$
(a) if \( a + d > b + c \) then \( \gamma(T) = 1 \),
(b) if \( a + d < b + c \) then \( \gamma(T) = -1 \),
(c) if \( a + d = b + c \) then \( \gamma(T) = 0 \).

Part (a): Assume first that \( a > b \) and \( c < d \). In this case Property 5 implies that \( \gamma(T) = 1 \). If \( a \leq b \) and \( c < d \) then let \( x \) be any number such that \( d - c > x > b - a \). Then \( T \) splits into the following tables

\[
T_1 = \begin{array}{cc}
\frac{a - \frac{1}{2}\min(a, c)}{c - \frac{1}{2}\min(a, c)} & \frac{a - \frac{1}{2}\min(a, c)}{d - x - \frac{1}{2}\min(a, c)} \\
\frac{\frac{1}{2}\min(a, c)}{d - x - \frac{1}{2}\min(a, c)} & \frac{\frac{1}{2}\min(a, c)}{d - x - \frac{1}{2}\min(a, c)}
\end{array} 
\]

\[
T_2 = \begin{array}{cc}
\frac{\frac{1}{2}\min(a, c)}{b - a + \frac{1}{2}\min(a, c)} & \frac{\frac{1}{2}\min(a, c)}{x + \frac{1}{2}\min(a, c)} \\
\frac{\frac{1}{2}\min(a, c)}{x + \frac{1}{2}\min(a, c)} & \frac{\frac{1}{2}\min(a, c)}{x + \frac{1}{2}\min(a, c)}
\end{array} 
\]

Then \( \gamma(T_1) = 1 \) because of Properties 3, 4, 5 and \( \gamma(T_2) = 1 \) because of Properties 3, 4, 6. Then by consistency, \( \gamma(T) = 1 \). If \( c \geq d \) then let \( x \) be any number such that \( a - b > x > c - d \). Then \( T \) splits into the following tables

\[
T_1 = \begin{array}{cc}
\frac{a - \frac{1}{2}\min(b, d)}{d - \frac{1}{2}\min(b, d)} & \frac{b - \frac{1}{2}\min(b, d)}{d - \frac{1}{2}\min(b, d)} \\
\frac{\frac{1}{2}\min(b, d)}{d - \frac{1}{2}\min(b, d)} & \frac{\frac{1}{2}\min(b, d)}{d - \frac{1}{2}\min(b, d)}
\end{array} 
\]

\[
T_2 = \begin{array}{cc}
\frac{\frac{1}{2}\min(b, d)}{c - d + \frac{1}{2}\min(b, d)} & \frac{\frac{1}{2}\min(b, d)}{c - d + \frac{1}{2}\min(b, d)} \\
\frac{\frac{1}{2}\min(b, d)}{c - d + \frac{1}{2}\min(b, d)} & \frac{\frac{1}{2}\min(b, d)}{c - d + \frac{1}{2}\min(b, d)}
\end{array} 
\]

Here \( \gamma(T_1) = 1 \) because of Property 5 and \( \gamma(T_2) = 1 \) because of Property 6. Consistency implies that \( \gamma(T) = 1 \).

Part (b): Swapping the two columns of \( T \) and applying Property 4 shows that \( \gamma(T) = -1 = CSR(T) \).

Part (c): If \( a < c \), choose a positive number \( x \) such that \( x < min(a, b, c - a) \). Then the split \( T \) into

\[
T_1 = \begin{array}{cc}
x & x \\
c - a + x & c - a + x
\end{array} 
\]

\[
T_2 = \begin{array}{cc}
a - x & b - x \\
a - x & b - x
\end{array} 
\]

With reference to Properties 1 and 2, consistency implies that \( \gamma(T) = 0 \). If \( a > c \), apply Property 3 and the previous result to obtain that \( \gamma(T) = 0 \). If \( a = c \), apply Property 1.

\[\square\]

**Proof of Proposition 7.** Proposition 1 is that (c) implies both (a) and (b). Proposition 4 is that (b) implies (c). Proposition 5 is that (b) implies (a). Proposition 6 is that (a) implies (c).

\[\square\]

**Proof of Proposition 8.** Let \( e = a + b = c + d \) so that \( b = e - a \) and \( d = e - c \). Then

\[
ad = a(e - c) > / = / < (e - a)c = bc
\]

if and only if

\[
a > / = / < c,
\]

21
which happens if and only if
\[ a + d = a + e - c > / = / < e - a + c = b + c. \]

\[ \square \]

**Proof of Proposition 9.** By Proposition 4, \( \gamma = CSR \) and the statement is true for the CSR.

\[ \square \]

**Proof of Proposition 10.** When \( p_a + p_d > p_b + p_c \), a wrong decision occurs if \( A + D \leq B + C \) or, equivalently, if \( A + D \leq n / 2 \).

Using the normal approximation to the binomial distribution, the probability of this event is
\[
P(X \leq \frac{n}{2} - n(p_a + p_d)}{\sqrt{n}(p_a + p_d)(1 - p_a - p_d)}) + o(1) = \tag{6}
\]

\[
P(X \leq \sqrt{n} \frac{0.5 - p_a - p_d}{\sqrt{(p_a + p_d)(1 - p_a - p_d)})} + o(1),
\]

where \( X \) is a standard normal random variable.

The true (population) value of the csr is \( (p_a + p_d)/(p_b + p_c) \) and \( p_a + p_d = csr/(1 + csr) \), thus (6) may be rewritten as
\[
P(X \leq \frac{1}{2} \sqrt{n} \frac{1 - csr}{\sqrt{csr}}) + o(1).
\]

\[ \square \]

**Proof of Proposition 11.** A \( 1 - \alpha \) level confidence interval for the true value of CSR does not contain 0 or -1, when CSR = 1 was observed, if
\[
t\left(\frac{a + d}{b + c}, u\right) = P\left(\frac{A + D}{B + C} > \frac{a + d}{b + c} | \text{sgn} \left(\log \frac{p_a + p_d}{p_b + p_c}\right) = u\right) < \alpha,
\]

for \( u = 0 \) and \( u = -1 \). It is easily seen that
\[
t\left(\frac{a + d}{b + c}, u\right) = P(A + D > a + d | \text{sgn} \left(\log \frac{p_a + p_d}{p_b + p_c}\right) = u).
\]

The hypothesis CSR = 0 is simple but the hypothesis CSR = -1 is composite and asymptotically
\[
\sup_{\text{csr} < 1} t\left(\frac{a + d}{b + c}, -1\right) = t\left(\frac{a + d}{b + c}, 0\right)
\]

22
and it is sufficient to determine the value of $t$ for $u = 0$. Under the hypothesis of $CSR = 0$, that is $p_a + p_d = 0.5$, the distribution of $A + D$ is $B(n, 0.5)$ and asymptotically

$$t\left(\frac{a + d}{b + c}, 0\right) = 1 - \Phi\left(\frac{2(a + b) - n}{\sqrt{n}}\right).$$

□

**Proof of Proposition 12.** $X$ and $Y$ are independent and asymptotically normal. Therefore, asymptotically, $X = 2A - N_+$ is normal with $E_X = N_+ (p_a - p_c)$ and variance $V_X = 4N_+ p_a p_c / (p_a + p_c)$ and $Y = 2B - N_-$ is normal with expectation $E_Y = N_- (p_b - p_d)$ and variance $V_Y = 4N_- p_b p_d / (p_b + p_d)$. Using independence, $X - Y$ is asymptotically normal, with expectation $E_X - E_Y$ and variance $\sqrt{V_X + V_Y}$. The probability of wrong decision is

$$P(B + C > A + D) = P(B - D > A - C) = P(Y > X)$$

$$= P(X - Y < 0)$$

and $P(X - Y < 0)$ is the same, asymptotically, as the probability that a standard normal variable is less than

$$\frac{E_Y - E_X}{\sqrt{V_X + V_Y}}$$

□
Table 1: Exact probabilities of wrong decision ($CSR = -1$ or $CSR = 0$) for various sample sizes and population values of the $csr$, when $CSR = 1$

<table>
<thead>
<tr>
<th>$p_a + p_d$</th>
<th>$p_b + p_c$</th>
<th>$csr$</th>
<th>$n = 100$</th>
<th>$n = 200$</th>
<th>$n = 500$</th>
<th>$n = 1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.51</td>
<td>0.49</td>
<td>1.04</td>
<td>0.4599</td>
<td>0.4158</td>
<td>0.3436</td>
<td>0.2739</td>
</tr>
<tr>
<td>0.52</td>
<td>0.48</td>
<td>1.08</td>
<td>0.3816</td>
<td>0.3100</td>
<td>0.1975</td>
<td>0.1086</td>
</tr>
<tr>
<td>0.53</td>
<td>0.47</td>
<td>1.13</td>
<td>0.3078</td>
<td>0.2178</td>
<td>0.0970</td>
<td>0.0309</td>
</tr>
<tr>
<td>0.54</td>
<td>0.46</td>
<td>1.17</td>
<td>0.2409</td>
<td>0.1437</td>
<td>0.0402</td>
<td>0.0062</td>
</tr>
<tr>
<td>0.55</td>
<td>0.45</td>
<td>1.22</td>
<td>0.1827</td>
<td>0.0887</td>
<td>0.0140</td>
<td>0.0008</td>
</tr>
<tr>
<td>0.56</td>
<td>0.44</td>
<td>1.27</td>
<td>0.1341</td>
<td>0.0511</td>
<td>0.0040</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.57</td>
<td>0.43</td>
<td>1.33</td>
<td>0.0950</td>
<td>0.0273</td>
<td>0.0010</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.58</td>
<td>0.42</td>
<td>1.38</td>
<td>0.0650</td>
<td>0.0136</td>
<td>0.0002</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.59</td>
<td>0.41</td>
<td>1.44</td>
<td>0.0428</td>
<td>0.0062</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.60</td>
<td>0.40</td>
<td>1.50</td>
<td>0.0271</td>
<td>0.0026</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.61</td>
<td>0.39</td>
<td>1.56</td>
<td>0.0165</td>
<td>0.0010</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.62</td>
<td>0.38</td>
<td>1.63</td>
<td>0.0096</td>
<td>0.0004</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.63</td>
<td>0.37</td>
<td>1.70</td>
<td>0.0054</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.64</td>
<td>0.36</td>
<td>1.78</td>
<td>0.0029</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.65</td>
<td>0.35</td>
<td>1.86</td>
<td>0.0015</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.66</td>
<td>0.34</td>
<td>1.94</td>
<td>0.0007</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.67</td>
<td>0.33</td>
<td>2.03</td>
<td>0.0003</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.68</td>
<td>0.32</td>
<td>2.13</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.69</td>
<td>0.31</td>
<td>2.23</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.70</td>
<td>0.30</td>
<td>2.33</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 2: Minimum required population values of the $csr$ to have wrong decision ($CSR = -1$ or $CSR = 0$) probabilities not exceeding 0.05 for various samples sizes

<table>
<thead>
<tr>
<th>Sample size</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum value of the $csr$</td>
<td>1.4178</td>
<td>1.2743</td>
<td>1.1631</td>
<td>1.1119</td>
</tr>
</tbody>
</table>

24
Table 3: Average approximate error probabilities for 1000 tables each having the \( csr \) values given in Table 2, when the \( CPR \) is used for decision. In all tables, \( CSR = 1 \) and the probability of wrong decision is 0.05 when the \( CSR \) is used. In some tables, \( CPR = -1 \), in others \( CPR = 1 \) and wrong decision is always the opposite or \( CPR = 0 \)

<table>
<thead>
<tr>
<th>Sample size</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average error probability of the ( CPR )</td>
<td>0.12</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 4: One-sided 95% confidence bounds for various sample sizes for the true value of the \( CSR \). When the observed value of \( csr \) exceeds (is less than) the value given, a one-sided 95% confidence interval does not contain \( CSR = 0 \) and \( CSR = -1 \) \( (CSR = 0 \) and \( CSR = 1) \)

<table>
<thead>
<tr>
<th>Sample size</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>3000</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed ( csr &gt; 1 )</td>
<td>1.61</td>
<td>1.39</td>
<td>1.26</td>
<td>1.16</td>
<td>1.11</td>
<td>1.09</td>
<td>1.08</td>
<td>1.06</td>
<td>1.05</td>
</tr>
<tr>
<td>Observed ( csr &lt; 1 )</td>
<td>0.62</td>
<td>0.71</td>
<td>0.79</td>
<td>0.86</td>
<td>0.90</td>
<td>0.92</td>
<td>0.93</td>
<td>0.94</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Table 5: Berkeley admissions: \( cpr \) and \( csr \) values for the six largest departments

<table>
<thead>
<tr>
<th>Dept A</th>
<th>Dept B</th>
<th>Dept C</th>
<th>Dept D</th>
<th>Dept E</th>
<th>Dept F</th>
<th>Total Depts A-F</th>
<th>Total Grad School</th>
</tr>
</thead>
<tbody>
<tr>
<td>( cpr )</td>
<td>0.36</td>
<td>0.80</td>
<td>1.14</td>
<td>0.91</td>
<td>1.23</td>
<td>0.85</td>
<td>1.84</td>
</tr>
<tr>
<td>( csr )</td>
<td>1.32</td>
<td>1.61</td>
<td>1.26</td>
<td>0.93</td>
<td>1.52</td>
<td>0.91</td>
<td>1.21</td>
</tr>
</tbody>
</table>

Table 6: Medical school applications: \( cpr \) and \( csr \) values for different MCAT-BS scores

<table>
<thead>
<tr>
<th>Score 3 or less</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10 or more</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( cpr )</td>
<td>1.75</td>
<td>1.81</td>
<td>2.00</td>
<td>2.42</td>
<td>2.40</td>
<td>1.86</td>
<td>1.55</td>
<td>0.93</td>
</tr>
<tr>
<td>( csr )</td>
<td>1.17</td>
<td>1.57</td>
<td>1.72</td>
<td>1.62</td>
<td>1.20</td>
<td>0.87</td>
<td>0.61</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Table 7: Patients with meningococcal disease diagnosed before hospital admission (Perera, 2008). \( CPR = 1 \) \( (cpr = 5.96) \) and \( CSR = -1 \) \( (csr = 0.79) \)

<table>
<thead>
<tr>
<th>Response</th>
<th>Died</th>
<th>Survived</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penicillin</td>
<td>22</td>
<td>83</td>
</tr>
<tr>
<td>No penicillin</td>
<td>2</td>
<td>45</td>
</tr>
</tbody>
</table>

Table 8: All patients with meningococcal disease (Perera, 2008). \( CPR = -1 \) \( (cpr = 0.86) \) and \( CSR = 1 \) \( (csr = 1.73) \)

<table>
<thead>
<tr>
<th>Response</th>
<th>Died</th>
<th>Survived</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penicillin</td>
<td>22</td>
<td>83</td>
</tr>
<tr>
<td>No penicillin</td>
<td>81</td>
<td>262</td>
</tr>
</tbody>
</table>