How Much More Likely?
The Implications of Odds Ratios for Probabilities

Akiva M. Liberman
National Institute of Justice

Abstract: Binary outcome data are common in research and evaluation. They are often analyzed using logistic regression, and results of these analyses are often reported in the form of odds ratios (ORs). However, ORs are not directly interpretable in the metric commonly used in policy-relevant discussions, which concerns probabilities. ORs are unfamiliar to nonresearchers, and their relationship to probability implications is not well understood by researchers. For example, the common practice of taking ORs as direct estimates of changes in probabilities (i.e., as risk ratios) systematically inflates effect sizes in probability terms. Fortunately, the probability implications of ORs can be derived simply and can be presented concretely as contrasting pairs of probabilities. These probability pairs are easily understandable by both the research and the lay audiences for evaluation results. After reviewing the relationship between probabilities, odds, and their ratios, this article shows how to explore ORs' implications for probabilities, illustrating with recent examples from the literature.

Keywords: *odds ratio; risk ratio; relative risk; odds; probability*

The editors of the *New England Journal of Medicine* say they "take responsibility" for media reports which greatly exaggerated conclusions in a study about possible gender and race bias in heart care. . . . Several news organizations, including the AP [Associated Press], interpreted the study to show that doctors were 40 percent less likely to order the [cardiac] tests for women and blacks than for men and whites. However, a follow-up published in the journal recently concluded that the likelihood of women and blacks being referred for the tests was actually seven percent less than for men and whites. The follow-up . . . said the misunderstanding resulted from the original study's use of an "odds ratio" to report the differences rather than a more commonly used "risk ratio."

—Associated Press (August 14, 1999)

Interpreting quantitative research and evaluation findings for nonresearchers is central to making those results useful for policy and practice. As statistical methods become more complex, interpreting and presenting quantitative findings accurately and understandably often becomes even more difficult. This is true for communication among researchers but especially true for communicating results to applied audiences, which is critical for evaluation research. A

Akiva M. Liberman, National Institute of Justice, 810 7th St., NW, Washington, D.C. 22301; phone: (202) 514-4919; e-mail: Akiva.Liberman@usdoj.gov.

Author's Note: The views expressed in this article are those of the author and do not represent official positions or polices of the National Institute of Justice or the Department of Justice.

American Journal of Evaluation, Vol. 26 No. 2, June 2005 253-266
DOI: 10.1177/1098214005275825

variety of approaches may help clarify the meaning of quantitative results. Effective graphical presentation of results is useful, even for relatively simple statistics (e.g., Tufte, 1983); graphical presentation is a popular topic for evaluators (e.g., Henry, 1997). Statistical simulation has also been advocated for making the implications of statistical models more explicit, and software has been developed for this purpose (King, Tomz, & Wittenberg, 2000). A more basic approach has been to transform esoteric results into more commonly understood terms. For example, Rosenthal and Rubin (1982) and Rosenthal (1990) argued that the use of r^2 as a measure of effect has led to the undervaluation of behavioral science effects, and they developed the binomial effect size display (BESD) to show the implications of correlations via a contrasting pair of probabilities. In a similar vein, this article argues that the implications of odds ratio (OR) results can be rendered more understandable via probability pairs.

Dichotomous Outcomes and Logistic Regression

In many areas of evaluation, dichotomous outcomes are common. For example, recidivism is a common outcome in criminal justice, mortality (or survival) and disease are key outcomes in public health, and graduation rates are a common outcome in education. Descriptively, dichotomous outcomes are readily understandable as percentages or proportions. For example, the difference between 10% and 20% survival following alternative courses of cancer treatment is clear to most laypeople. However, contemporary analytic methods for dichotomous data often lead to results that have lost that ease of interpretation. In particular, one common analytic procedure, logistic regression, yields results that are often presented as ORs, which are not well understood. This article demonstrates how to assess the probability implications of OR results by generating pairs of probabilities that exemplify an OR. How much does arrest reduce intimate partner violence? A reported OR of 0.75 (see Example 2 below) may not be completely clear and is open to a common misunderstanding. Rendered as a reduction in revictimization rates from 54% to 46%, the result becomes much more understandable.

Linear regression of dichotomous data (using ordinary least squares [OLS]) yields unstandardized coefficients that are readily interpretable as differences in probabilities or proportions (e.g., Cohen & Cohen, 1983). However, dichotomous data violate assumptions underlying linear regression. Binary data cannot be normally distributed, the errors are heteroscedastic, and regression results may imply impossible outcome probabilities less than 0.0 or greater than 1.0.

Logistic regression (also known as logit regression) is one appropriate method for analyzing binary outcome data (e.g., DeMaris, 1992; Hosmer & Lemeshow, 1989; Long, 1997; Menard, 1995). Like OLS regression, logistic regression may involve multiple predictors, each simultaneously controlled for the others. But results are less easily understood, because logistic regression coefficients (b values) are not in the metric of proportions or probabilities. Instead, each coefficient indicates how the log odds of an outcome are increased by a 1-point increase in the predictor. Because log odds (or logits) are not readily interpretable, these coefficients are often exponentiated, which changes the metric from log odds to odds. At the same time, exponentiating changes the resulting coefficient, often labeled exp(b), to a ratio rather than an increment. The resulting ORs, then, are multiplicative and in the metric of odds, in contrast to linear regression coefficients, which are additive and in the metric of probabilities. Although perhaps somewhat more accessible than nonexponentiated logit coefficients, OR results remain mysterious to many. Rossi (1997) described the resulting quandary for evaluation research:

> Logistic regression and its variants . . . have come to be used routinely in our field in the past decade. I believe this diffusion has enriched the abilities of evaluation research to produce credible research.

Table 1
Examples With Odds Ratio (OR) = 3.0

Example			Difference	Ratio of Larger to Smaller		Inflation Factor
A. Premature infant survival	22 weeks old	23 weeks old				
Survival (p)	.10	.25	.15	2.50	(risk ratio)	1.20
Mortality (q)	.90	.75	.15	1.20	(risk ratio)	
Odds (p/q)	0.11	0.33		3.00	(OR)	
B. Parent employment	Black children	White children				
Yes (p)	.50	.75	.25	1.50	(risk ratio)	2.00
No (q)	.50	.25	.25	2.00	(risk ratio)	
Odds (p/q)	1.00	3.00		3.00	(OR)	

> Of course, there is a downside to this development. After much work, I believe we have educated our audience sufficiently that they can now understand the results of OLS regressions and ANOVA [analysis of variance]. . . . Odds ratios and relative risk are not as easily understandable by those outside the evaluation fraternity, and often have a hard time within it. (p. 66)

When improved statistical methods lead to less understandable quantitative results, which are harder to explain to practitioners or policy makers, evaluation researchers are left with difficult choices. Fortunately, an OR's implications can be made much more understandable and concrete, in terms of contrasting pairs of probabilities. However, a simple method for doing this is rarely explained in texts that discuss the statistical techniques themselves.

This article shows how to explore the implications of ORs for probabilities. I first review the relationship of probabilities, odds, and their ratios using simple bivariate results in 2 × 2 tables. I then consider multiple logistic regression results and how to explore their implications for probabilities. Finally, I work through several examples from the literature.

Probabilities, Odds, and Their Ratios

I illustrate the relationship between probabilities, odds, and their ratios with two simple examples contrasting pairs of probabilities (see Table 1).

Very premature babies' chances of survival increase with gestational age. Babies born at 22 weeks have only a 10% survival rate, but at 23 weeks, 25% survive.[1] These probabilities of survival, denoted p, are .10 and .25. The probability of the opposite outcome, here infant mortality, equals $1 - p$ and is often denoted q. Here, $q = .90$ and .75.

In 1980, 75% of non-Hispanic White children and 50% of non-Hispanic Black children lived with at least one parent who was employed full-time.[2] For White children, $p = .75$. The opposite probability, that of not residing with a working parent, is $q = .25$. For Black children, $p = .50$, and $q = .50$.

Probability Differences and Ratios

How much more likely are 23-week than 22-week babies to survive? How much more likely are White than Black children to reside with a working parent? This is the classic issue of effect size (Cohen, 1977). There are many ways to describe the magnitude of an effect, even from 2 × 2 tables (see Agresti, 1990; Fleiss, 1994).

Two obvious ways to describe the magnitudes of these effects are via the difference and ratios of the two probabilities. For infant survival, the probability difference is .25 – .10 = .15. The probability ratio, commonly known as the risk ratio or relative risk (RR), is .25 / .10 = 2.5, meaning that 23-week-old infants are 2.5 times as likely to survive as 22-week-old infants.

For parent employment, the probability difference is .75 – .50 = .25, and the RR is .75 / .50 = 1.5. Compared with Black children, White children are 25 percentage points more likely, and 1.5 times as likely, to reside with an employed parent.

Odds and ORs

Another common effect size measure is the ratio of two odds, known as the OR (or relative odds). Every probability has a corresponding odds, which equals the probability of the outcome occurring (e.g., survival) divided by the probability of the outcome not occurring (e.g., mortality):

$$\text{Odds} = p \,/\, (1 - p) = p \,/\, q, \tag{1a}$$

and

$$p = \text{Odds} \,/\, (\text{Odds} + 1). \tag{1b}$$

For example, for p = .80, the odds are .80 / .20 = 4.0, colloquially "4 to 1" odds. Note that odds and p are not linearly related. A 1-point change in odds corresponds to varying changes in p, and vice versa. For example, odds = 2.0 versus 3.0 correspond to ps = .67 versus .75, and odds = 7.0 versus 8.0 correspond to ps = .875 versus .890.

Taking the ratio of two odds (i.e., dividing one by the other) gives the OR. For White children, the odds of parent employment are .75 / .25 = 3.00, colloquially "3 to 1" odds; for Black children, the odds are .50 / .50 = 1, colloquially "1 to 1" odds. The ratio of these two odds is OR = 3.0 / 1.0 = 3.0. Thus, the odds of residing with an employed parent are 3 times as large as for White than for Black children.

$$\text{OR} = (p_1 \,/\, q_1) \,/\, (p_2 \,/\, q_2). \tag{2a}$$

For 23-week-old infants, the odds of survival are .25 / .75 = 0.33; for 22-week-old infants, the odds are .10 / .90 = 0.11. The ratio of these two odds is 0.33 / 0.11 = 3.0. Notice that the OR for the parent employment and infant survival examples are equal, even though they have different probability differences (.25 and .15) and different RRs (1.5 and 2.5).

Ratios and inverses. For any two numbers, one can compare the larger with the smaller or the smaller with the larger. When the comparison is by subtraction, the absolute difference between them is unchanged; only the sign changes. When the comparison is by division, as with ratios, then the two ratios that result are the inverses of each other. That is, doubling the smaller is equivalent to halving the larger; quadrupling one is equivalent to quartering the other. The neutral point for ratios implying no effect is 1.0.

Therefore, the inverse of a ratio has the same magnitude but opposite direction. For example, 23-week-old infants have 3 times the odds of survival as 22-week-olds; conversely, 22-week-old infants have one third the survival odds of 23-week-old infants. Any OR < 1 can be described as an OR > 1 by changing the direction of comparison and using the inverse. For simplicity, this article focuses on ORs > 1.0, or the ratio of the larger odds to the smaller.

Relationship Between RR and OR

We have seen that the OR does not describe the change in probabilities. Although OR = 3.0, White children are not 3 times as likely as Black children to reside with a working parent (RR = 1.5), nor are 23-week-old infants 3 times as likely to survive as 22-week-old (RR = 2.5). In fact, the RR is always smaller than the OR. More generally, each OR is associated with a family of probability pairs with different RRs (Figure 1 illustrates for OR = 3.0). As the probabilities involved increase, the RRs decrease (see Figure 2). With very low probabilities, the RR is close to the OR; with very high probabilities, the RR approaches 1.0. Thus, the OR is an upper bound on associated RRs. Taking ORs as estimates of RRs systematically inflates effect sizes in probability terms, and more so with larger probabilities. This general pattern is true for all ORs > 1.

In fact, each OR is actually the product of two associated RRs. Rearranging terms from Equation 2a, Equation 2b shows the OR as the product of two risk ratios, one involving two *p*s and the other involving two *q*s:

$$\mathrm{OR} = (p_1 / p_2) \times (q_2 / q_1). \tag{2b}$$

For infant survival, the RR with *p*s describes the increase in *survival* for 23-week-old infants (RR = 2.5); the RR with *q*s describes the increase in *mortality* for 22-week-old infants (RR = 1.20), as shown in Table 1. These two different RRs each describe the same effect; 23-week-old infants are 2.5 times as likely to survive, yet 22-week-old infants are only 1.2 times as likely to die.

For parent employment, the RR with *p*s describes how much more likely White children are to reside with a working parent (RR = 1.5); the RR with *q*s describes how much more likely Black children are to reside without a working parent (RR = 2.0). In each case, the OR is the product of these two RRs: OR = 3.0 = 2.5 × 1.2 = 1.5 × 2.0.

Because the OR is the product of two RRs that each appropriately describe the same effect, intuitively, ORs are in the metric of squared RRs. This suggests that $\sqrt{\mathrm{OR}}$ might be a useful indicator of the effect size in the probability metric (see The Centered Case below).

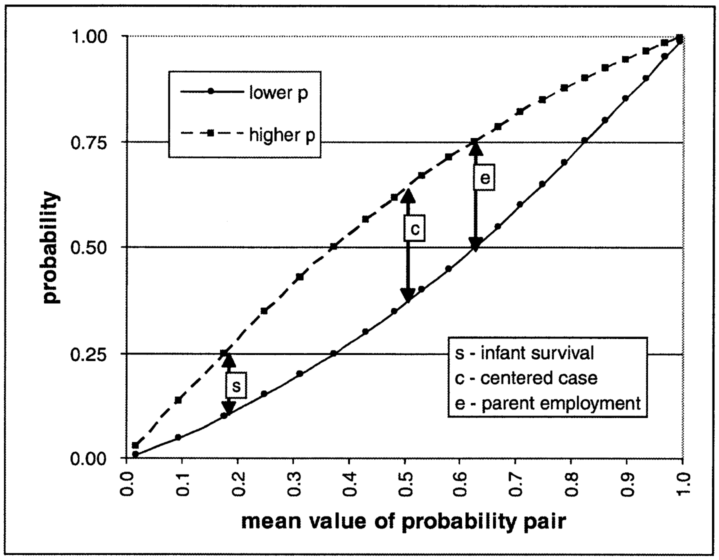
Advantages of ORs

Although the OR does not describe changes in probabilities, it does have other advantages (see Fleiss, 1981). We have seen that there are always two RRs associated with any effect, associated with the two alternative outcomes,[3] which is problematic. In contrast, because the OR uses both the *p*s and the *q*s, there is only one OR > 1 for an effect.

The OR is also unbounded, whereas the possible magnitude of RR is constrained. Because 1.00 is the upper bound on probabilities, $p = .33$ cannot be increased by more than a factor of 3.0, $p = .50$ by more than a factor of 2.0, or $p = .80$ by more than a factor of 1.25. This varying upper bound makes it problematic to compare RRs across studies, especially if samples vary on their risk for an outcome. In contrast, there is no upper bound on odds or ORs. Even with $p = .80$, treatment effects of OR = 3.0 or 4.0 are possible, which raise p to .92 and .94, respectively.

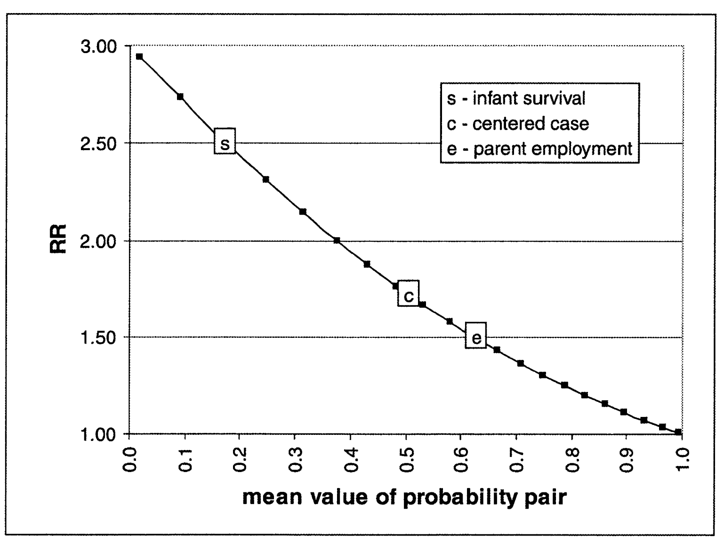
Beyond these advantages of using the OR as an effect-size measure, the statistical techniques that produce ORs have strengths that lie beyond the scope of this article (see Hosmer & Lemeshow, 1989; Long, 1997). Despite these strengths, OR results remain unclear for many until they are expressed in terms of probabilities.

Figure 1
Probability Pairs for Odds Ratio (OR) = 3.0



Note: For each value of the x-axis, the pair of probabilities shown have an OR = 3.0. Three probability pairs discussed in the text are indicated by arrows.

Figure 2
Risk Ratios (RR) Associated With Odds Ratio(OR) = 3.0



Note: RRs correspond to the probability pairs shown in Figure 1.

The Probability Implications of ORs From Models With Multiple Predictors

With 2×2 tables, one can easily examine the probability difference, the RR, and the OR and compare and contrast them. This is not as easy with results from models with multiple predictors. Assume that we had used logistic regression to analyze binary outcome data for a study with multiple predictors. The exponentiated forms of the model coefficients, exp(b), are ORs, and each OR is adjusted for the effects of other predictors. We have seen that each OR is associated with a family of probability pairs, each with different probability differences, and the OR is larger than any of the associated RRs. How then can we explore implications for probabilities? (Other issues in interpreting logistic regression results, such as confidence intervals, standardizing ORs, and assessing overall model fit, are beyond the scope of this article; see, e.g., DeMaris, 1992; Hosmer & Lemeshow, 1989; Long, 1997; Menard, 1995.)

With very low risks or probabilities, ORs are not much larger than associated RRs. Therefore, one common suggestion is to take ORs as "estimates" of RRs (e.g., Fleiss, 1981; Hosmer & Lemeshow, 1989, p. 42). However, as can be seen in Figures 1 and 2, low probabilities are a special case rather than the rule. Routinely interpreting ORs as RRs systematically inflates effect sizes in probability terms. If such probability estimates are used in program planning or cost-benefit analyses, they will inflate program expectations and may feed into a cycle of overpromising and subsequent disappointment.

Using Probability Pairs to Illustrate an OR's Implications

Fortunately, we can explore the probability implications relatively simply, without the primary data or even the full statistical models. One simply calculates how an OR affects chosen probability values. The resulting pairs of probabilities exemplify the OR. This is valuable for evaluation researchers when presenting their results and also for readers attempting to understand (or check) the implications of reported OR results.

One can do this in three steps, as follows: (a) convert a chosen p to its odds (Equation 1a), (b) multiply the odds by the OR,[4] and (c) convert the resulting odds back to a p (Equation 1b). For example, how would OR = 3.0 affect $p = .10$? (a) odds (.10) = .10 / .90 = 0.11; (b) $0.11 \times 3.0 = 0.33$; and (c) $p = 0.33 / 1.33 = .25$.

Equivalently, one can calculate the higher value in a probability pair (denoted p_H) given the lower (denoted p_L), using Equation 3, or calculate the lower probability given the higher, using Equation 4.[5] In the rest of this section, I work through using these formulas with our running examples with OR = 3.0, before discussing a special probability pair called the centered case.

$$p_H = (p_L \times \text{OR}) / (1 - p_L + p_L \times \text{OR}), \tag{3}$$

and

$$p_L = p_H / (\text{OR} + p_H - p_H \times \text{OR}). \tag{4}$$

With OR = 3.0, Equation 3 raises $p_L = .50$ to $p_H = .75$; conversely, Equation 4 lowers $p_H = .75$ to $p_L = .50$. We can then calculate differences and ratios from this pair of probabilities. Given p_L, one can also directly calculate the RR:

$$\text{RR} = \text{OR} / (1 - p_L + p_L \times \text{OR}) \tag{5}$$

Table 2
Exploring Odds Ratio (OR) = 3.0

Probability pair	p_L	p_H	Difference	Risk Ratio	Inflation Factor
$p_H = .70$	.44	.70	.26	1.60	1.875
$p_L = .35$	.35	.62	.27	1.76	1.70
$p_L = .60$	.60	.82	.22	1.37	2.18
$p_H = .60$	.33	.60	.27	1.80	1.67
Centered case					
p	.37	.63	.26	1.73	1.73
q	.63	.37	.26	1.73	
Odds	1.73	1.73		3.00	

In addition, dividing the OR by the resulting RR indicates how much the OR inflates the RR for those probabilities, the "inflation factor" (IF; Zhang & Yu, 1998):

$$IF = OR / RR. \tag{6}$$

To illustrate, imagine OR = 3.0 as the coefficient for race from a multiple logistic regression of parent employment, which also controlled for other demographic factors related to employment (e.g., urbanicity, region). Assume further that the observed rates were .70 for White children and .35 for Black children. (Note that for these observed rates, the OR = [.70 / .30] / [.35 / .65] = 4.33, is larger than the modeled result. This parallels the common situation in which covariates reduce the size of an effect.) To explore probability implications, we apply OR = 3.0 to various *p*s. One might take the observed rates for each group and see how they would be affected by this effect (see Table 2). Applying Equation 4 to the observed rate for White children of .70, we would expect .44 parent employment for comparable Black children. These two probabilities have a difference of .26, and a RR = 1.60. If we apply Equation 3 to the observed rate for Black children of .35, we would expect a .62 rate for White children, with a difference of .27, and a RR = 1.76.

Another reasonable value to explore may be the base rate or sample grand mean, which is generally representative of the data, especially when observed group rates are not available. Assume that we had a base rate of parent employment of .60. Taking .60 as p_L, we find that if 60% of Black children lived with a working parent, we would expect 82% parent employment for comparable White children (Equation 3), with a difference of .22, and RR = 1.37. (Exploration of .60 as the value of p_H is left to Table 2.) All of these resulting RRs are considerably below 3.00, and the discrepancy between the OR and the RR increases with larger probability pairs. In the last case, the inflation of the effect (a factor of 2.18) is actually larger than the effect itself (1.37).

The Centered Case

Because RRs decrease as the probability pairs increase, the probability pair at the center of the probability range, with *p*s centered around .50, would seem to be of special interest. I call this the centered case. It is particularly useful when one lacks a strong basis for choosing representative probabilities to explore.

Remember that the OR is the product of two RRs (see Equation 2b) and that the RR for *p*s decreases as the probability values increase (Figure 1). Intuitively, this suggests that setting the two RRs to be equal, by taking the square root of the OR, will generate a case with $p_1 = q_2$ and vice versa. The centered risk ratio provides a sort of average of all possible RRs for an OR.[6] Thus, the centered case is useful if one wants to make a general statement across the probability range. This centered case is simple to generate.[7]

$$\text{Centered RR} = \sqrt{\text{OR}}, \tag{7a}$$

$$\text{Centered } p_L = 1 / (1 + \sqrt{\text{OR}}), \tag{7b}$$

and

$$\text{Centered } p_H = 1 - p_L. \tag{7c}$$

The centered case for OR = 3.0 has $p_L = .366$ versus $p_H = .634$, with a difference of .268, RR = 1.73 (see Table 2). The centered case also has the maximum probability difference for a given OR (see Figure 1). We find, then, that OR = 3.0 generates at most an increase in the outcome of .268.

This centered case is similar to the BESD, which illustrates the probability difference implied by a correlation (Rosenthal & Rubin, 1982).[8]

An Alternative Approach: Exploring Illustrative Scenarios Defined by Predictor Variables

The approach discussed above begins with one specified probability and uses the OR to generate the other. In a complementary approach, neither probability is specified directly. Rather, the logistic regression equation is used to calculate both probabilities, for a scenario defined by specific values on all predictor variables (e.g., Menard, 1995). Substituting these values into the equation, one obtains log odds of the outcome and then transforms them to probabilities. To generate a broadly representative scenario, all variables (except the variable of interest) are often set to their means (e.g., Long, 1997). This approach uses model coefficients directly (i.e., *b* values), rather than their exponentiated forms (ORs). I do not work through this method in detail here because it does not directly involve the OR, is somewhat more complex computationally, and is covered in detail elsewhere (e.g., Hosmer & Lemeshow, 1989; Long, 1997; Menard, 1995).

Examples From the Literature

Example 1: Defense Counsel and Out-of-Home Placement

Burruss and Kempf-Leonard (2002) evaluated how representation by defense counsel affects dispositions of out-of-home placement in Missouri juvenile court. I discuss their logistic regression results for suburban circuits, for which the overall out-of-home placement rate is .15. They find that defense counsel actually increased out-of-home placement ($b = 1.042$; $\exp[b]$ = OR = 2.83). The model includes nine additional predictors, three of which are significant.

How much does defense counsel increase out-of-home-placement? Burruss and Kempf-Leonard (2002) went to unusual lengths to explore their OR's probability implications. They applied the full logistic regression equation to scenarios defined using the nine predictor vari-

Table 3
Effects of Defense Counsel in Juvenile Court on Out-of-Home Placement at Disposition, Suburban Circuits

	Any Counsel $b = 1.042$, $\exp(b) = \text{OR} = 2.83$				Public Defender $b = 1.672$, $\exp(b) = \text{OR} = 5.32$			
Probability Pair	p_L	p_H	RR	IF	p_L	p_H	RR	IF
BR as lower rate	.15	.33	2.22	1.27	.15	.48	3.22	1.65
BR as higher rate	.06	.15	2.56	1.11	.03	.15	4.66	1.14
Centered case	.37	.63	1.68	1.68	.30	.70	2.34	2.27
.33 as lower rate	.33	.58	1.74	1.63	.33	.72	2.19	2.43
.85 as lower rate	.85	.94	1.10	2.57	.85	.97	1.13	4.71

Source: Burruss and Kempf-Leonard (2002, Table 5, Figure 1).
Note: OR = odds ratio; RR = relative risk; IF = inflation factor; BR = base rate.

ables. (They also used Monte Carlo methods beyond the scope of this article.) Five variables (here a to e) were set to describe a (a) 14-year-old, (b) White, (c) male (d) living with a single parent and (e) detained during the case proceeding. For such a defendant, they then explored five different instant offense and delinquency history scenarios, as defined by four other variables (here f to i). In the least serious scenario, the arrestee had (f) one prior offense and (g) no abuse and neglect history and was charged with (h) one (i) nonviolent offense. In the most serious scenario, he had (f) multiple priors, (g) a history of abuse, and (h) more than three instant offenses, (i) including a crime against a person.

I derive equivalent results, using Equation 3, taking p_L from their least and most serious scenarios (see the left half of Table 3).[9] Defense counsel would raise out-of-home placement from .33 to .58 (RR = 1.74) or from .85 to .94 (RR = 1.10), respectively, for the least and most serious scenarios. Although these scenarios seem meaningful and worth exploring, they do not generally represent the data, which had an overall placement rate of only .15. With higher probabilities than typical, these scenarios generate smaller RRs.

To generate more representative probability pairs, I set p_L or p_H to the grand mean and apply Equations 3 and 4. We find that defense counsel would raise a .15 placement rate to .33 (RR = 2.22) and .06 to .15 (RR = 2.56). I also generate the centered case (Equations 7a to 7c), which contrasts rates of .37 to .63, with a difference of .26 (RR = 1.68). Because the probability pair in the centered case has the maximum probability difference for an OR, this means that counsel raises the probability of out-of-home placement by up to 26 percentage points. We find, then, that in typical cases, the placement rate more than doubles, raising a 6% rate to 15%, and 15% to 33%, with the base rate set to p_H and p_L, respectively. (A separate model finds that public defenders raised placements even more dramatically, OR = 5.32; probability implications are shown in the right half of Table 3.)

Example 2: Arrest and Intimate Partner Violence

Maxwell, Garner, and Fagan (2002) evaluated the preventive impact of arrest on intimate partner violence by conducting logistic regression analyses on pooled data from five experiments. On the basis of victim interviews, they reported that "arrest reduced the prevalence of new victimizations by 25%" (p. 64), describing an OR = 0.75. In the interest of brevity, I limit exploration to the centered case and the base rate (Table 4).

Note that the OR is less than 1.0. We generate the centered case as usual (equations 7a to 7c) and get RR = $\sqrt{\text{OR}}$ = 0.866, which contrasts *ps* of .46 and .54. To use Equations 3 and 4 with an

Table 4
Effect of Arrest on New Intimate Partner Violence

	Probability Pair	p_L	p_H	RR	Inverse of RR	IF
OR = 0.75, inverse OR = 1.33	Centered case	.464	.536	1.155	0.864	1.155
	BR as higher rate	.357	.425	1.19	0.84	1.12
	BR as lower rate	.425	.496	1.17	0.85	1.14

Source: Maxwell, Garner, and Fagan (2002).
Note: RR = relative risk; IF = inflation factor; OR = odds ratio; BR = base rate.

Table 5
Effect of Gang Membership on Violence at Age 18

	Probability Pair	p_L	p_H	RR	IF
Gang membership at age 14: OR = 3.39	Centered	.35	.65	1.84	1.84
	.10 as lower rate	.10	.27	2.74	1.24
Gang membership at age 16: OR = 4.58	Centered	.32	.68	2.14	2.14
	.10 as lower rate	.10	.34	3.37	1.36

Source: Hawkins et al. (1998).
Note: RR = relative risk; IF = inflation factor; OR = odds ratio.

OR > 1, I first take the inverse, getting OR = 1 / .75 = 1.33. Setting p_H to the reported base rate of .425, Equation 4 gives p_L = .357 (RR = 1.19). Setting p_L to .425, Equation 3 gives .496 as the higher rate (RR = 1.17). To see how much arrest *reduces* victimization, we take the inverse of these RRs. For RR = 1.19, the inverse is 1 / 1.19 = 0.84, which means 84% victimization compared with no arrest, or a reduction of 16%. For RR = 1.17, the inverse is 1 / 1.17 = 0.85, or a reduction in victimization of 15%. We find, then, that arrest reduces the prevalence of new victimization by about 15%, important to be sure, but not 25%.

Example 3: Gang Membership as a Risk Factor for Later Violence

The OR has advantages for summarizing results involving varying samples and base-rates (Fliess, 1981; Haddock, Rindskopf, & Shadish, 1998; Lipsey & Wilson, 2001) and is sometimes used as the effect size in quantitative literature reviews. This is another common use of the OR besides logistic regression results. For example, Hawkins et al. (1998) reviewed predictors of youth violence from longitudinal studies and often described ORs as RRs. Thus, ORs of 3.39 and 4.58 were described as follows: "Gang membership at age 14 more than tripled the risk of violence at age 18, and gang membership at age 16 more than quadrupled the risk of violence at age 18" (p. 143).

I explore the probability implications of gang membership at age 14 here and in Table 5. (I leave gang membership at age 16 to the table.) In the centered case, RR = $\sqrt{3.39}$ = 1.84, which contrasts violence rates of .35 to .65 (Equations 7a to 7c). This centered case implies a maximum probability difference of .30. Because these are fairly high violence rates, we might also wish to explore how lower violence rates are affected.[10] Setting p_L = .10 yields p_H = .27 (RR = 2.74; see Table 5). We find, then, that gang membership at age 14 would raise a 35% risk almost twofold to 65%; a 10% violence risk would almost triple to 27%.

Discussion

Although some statistical methods for analyzing dichotomous outcomes generate ORs, their probability implications are not immediately obvious. There is a mismatch between the assumptions of nonlinear models and the probability metric in which many applied research results are discussed.

From ORs to Probabilities

Although the OR provides an upper bound on associated RRs, it is approached only for very low probabilities. Moreover, no single RR most accurately describes an OR, because the probability implications vary for different subsets of data and at different parts of the probability range.

Fortunately, an OR's probability implications can be explored rather simply, by generating pairs of probabilities, from which differences or ratios can be computed. These probability pairs are intuitively meaningful illustrations of the effect. One can generate probability pairs representative of the data set (e.g., around the grand mean), relevant to particular cases of theoretical interest, or in other parts of the probability range. The centered case is generally representative of the entire probability range; conveniently, the centered RR = $\sqrt{\text{OR}}$. Which probability pairs best illustrate an OR's implications largely depends on the probability range considered substantively relevant.

Exploring Scenarios Defined by Predictors

One can also generate probability pairs from full logistic regression equations by specifying values on all predictor variables. Statistical packages increasingly help generate such estimates. These two approaches are complementary. Each approach derives straightforwardly from the definitions of model coefficients, but in different forms (i.e., exp[b] vs. b). The conceptual difference is between defining a scenario of interest using only a probability value on the outcome (and the OR) or by specifying values on all predictors (and using the full model equation).

Using the OR is especially useful when one does not have specific theoretically meaningful values for all predictors. In addition, because full models are not necessary for exploring the OR's implications, readers or secondary researchers can easily explore those implications, even when primary researchers have not (or have inappropriately described the OR as the RR).

Conclusion

When appropriate statistical techniques yield results in a form that is not readily understandable or that is difficult to explain to nonresearchers, applied researchers often choose among suboptimal analysis, no interpretation at all, or simple but inappropriate description of their findings. In analyses of dichotomous data, this dilemma seems more paradoxical because dichotomous outcomes themselves are so readily understandable.

Logistic regression results of binary data often leave the meaning of findings obscure. OR results are often simply interpreted as RRs. Although reasonable in the special case of very low probabilities, this generally exaggerates effects, which can feed into overpromising the results of programs and interventions. Some researchers simply avoid interpreting OR results in probability terms, discussing only statistical significance and the direction of effects. This leaves real-world implications to guesswork, often by untrained research consumers, and is particularly unsuited to evaluation results.

Alternatively, some researchers may avoid presenting statistically appropriate analyses with multiple predictors. Instead, researchers may present bivariate results for one predictor at a time, failing to control for other predictors.

The present approach allows one to easily show the probability implications of OR results, in the concrete terms of contrasting pairs of probabilities or proportions. For binary data, this should help evaluators alleviate the tension between the scientific need to use rigorous analytical techniques and the pragmatic need to communicate results meaningfully in real-world terms understandable to practitioner and policy audiences.

Notes

1. Actual numbers are as follows: at 22 weeks, 0% to 10%; at 23 weeks, 10% to 40%; and at 24 weeks, 40% to 70% (University of Wisconsin Medical School, Department of Pediatrics, 2004).

2. Figures for 1980 are used for illustration; comparable 2001 figures are 84% and 65% (Federal Interagency Forum on Child and Family Statistics, 2003, p. 90).

3. I refer here to risk ratios (RRs) > 1. For each of these RRs, there are also two associated RRs < 1, which are their inverses and have the opposite direction of comparison.

4. With odds ratios (ORs) and multivalued predictors, the effect of x scale points is not OR × x, as in OLS, but rather OR^x.

5. These equations are for OR > 1.0. For ORs < 1.0, these equations can also be used, but p_L is the higher probability and p_H the lower.

6. In the centered case, the two associated RRs discussed earlier, one for p and one for q, are identical. For all other probability pairs associated with the OR, the centered RR lies between these two RRs.

7. Equation 7c is definitional for the centered case. To derive Equation 7b, substitute Equation 7c into the definitional formula $OR = [p_H / (1 - p_H)] / [p_L / (1 - p_L)]$, and rearrange terms. To derive Equation 7a, start with $p_L = p_H / RR$ (true by definition of the centered RR); substitute $p_H = 1 - p_L$ on right side, and solve for p_L.

8. In the BESD, the probability difference simply equals r. For example, for $r = .30$, the BESD is .35 versus .65. The two centered cases are not quite the same, because one derives from the OR and the other from r. The OR and r are not related in a simple linear fashion, and unequal *ns* (i.e., different marginal distributions) affect r but not the OR.

9. Burruss and Kempf-Leonard (2002) reported .57 rather than .58. In general, Equation 1 generated results within .01 of their results, with differences presumably due to rounding errors.

10. Note, however, that about 30% of high school seniors report involvement in serious violence (U.S. Department of Health and Human Services, 2001, p. 27).

References

Agresti, A. (1990). *Categorical data analysis*. Hoboken, NJ: John Wiley.

Burruss, G. W., & Kempf-Leonard, K. (2002). The questionable advantage of defense counsel in juvenile court. *Justice Quarterly, 19*, 37-67.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences, revised edition.* London: Academic Press.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

DeMaris, A. (1992). *Logit modeling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-086). Newbury Park, CA: Sage.

Federal Interagency Forum on Child and Family Statistics. (2003). *America's children: Key national indicators of well-being, 2003.* Washington, DC: U.S. Government Printing Office.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions.* New York: John Wiley.

Fleiss, J. L. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245-260). New York: Russell Sage.

Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3*, 339-353.

Hawkins, J. D., Herrenkohl, T., Farrington, D. P., Brewer, D., Catalano, R. F., & Harachi, T. W. (1998). A review of predictors of youth violence. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors, and successful interventions* (pp. 106-146). Thousand Oaks, CA: Sage.
Henry, G. T. (Ed.). (1997). Creating effective graphs: Solutions for a variety of evaluation data. *New Directions for Evaluation, 73.*
Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York: John Wiley.
King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science, 44,* 341-355.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* New York: Sage.
Long, S. J. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.
Maxwell, C. D., Garner, J. H., & Fagan, J. A. (2002). The preventive effects of arrest on intimate partner violence: Research, policy, and practice. *Criminology and Public Policy, 2,* 51-80.
Menard, S. (1995). *Applied logistic regression analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-106). Thousand Oaks, CA: Sage.
Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 45,* 775-777.
Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166-169.
Rossi, P. H. (1997). Advances in quantitative evaluation, 1987-1996. *New Directions in Evaluation, 76,* 57-68.
Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.
University of Wisconsin Medical School, Department of Pediatrics. (2004). *For parents of preemies: Anticipating the birth of a premature infant.* Retrieved December 14, 2004, from http://www.pediatrics.wisc.edu/patientcare/preemies/anticipating.html
U.S. Department of Health and Human Services. (2001). *Youth violence: A report of the surgeon general.* Rockville, MD: Author.
Zhang, J., & Yu, K. F. (1998). What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *Journal of the American Medical Association, 280,* 1690-1691.