



9-2001

# Application and Interpretation of Simple Odds Ratios in Physical Therapy-Related Research

Pamela Levangie  
*Sacred Heart University*

Follow this and additional works at: [http://digitalcommons.sacredheart.edu/pthms\\_fac](http://digitalcommons.sacredheart.edu/pthms_fac)



Part of the [Physical Therapy Commons](#)

---

## Recommended Citation

Levangie, Pamela. "Application and Interpretation of Simple Odds Ratios in Physical Therapy-Related Research." *JOSPT* 31.9 (2001): 496-503.

This Peer-Reviewed Article is brought to you for free and open access by the Physical Therapy & Human Movement Science at DigitalCommons@SHU. It has been accepted for inclusion in All PTHMS Faculty Publications by an authorized administrator of DigitalCommons@SHU. For more information, please contact [ferriby@SacredHeart.edu](mailto:ferriby@SacredHeart.edu), [lysobeyb@SacredHeart.edu](mailto:lysobeyb@SacredHeart.edu).

# Application and Interpretation of Simple Odds Ratios in Physical Therapy-Related Research

Pamela K. Levangie, PT, DSc

Over the past several decades, physical therapists have demonstrated an increasing responsiveness to the profession's obligation to generate objective evidence for examination and intervention strategies employed in physical therapy practice. This trend is evident, not only in the increasing number of journals that are publishing physical therapy research, but in the growing sophistication of research design and analytic options used by investigators. At the same time, physical therapists are held increasingly accountable for adopting an evidence-based approach to practice. The result for many of us is a growing concern about our ability to interpret study findings. The ability to independently weigh the importance to our own practice of evidence reported in a study requires that we understand the strengths and potential weaknesses of the sample, design, and analyses being used. The odds ratio (OR) is one of the analytic measures that has only recently appeared in the physical therapy literature. Because the OR may be unfamiliar to physical therapists, the goal of this paper is to provide a description of the simple OR and a discussion of its uses, interpretation, and potential limitations. *J Orthop Sports Phys Ther* 2001;31:496-503.

**Key Words:** confidence intervals, odds ratios, relative risk analysis, statistical analysis

## MEASURES OF ASSOCIATION

In their simplest form, most clinical questions are related to 1 of 2 issues: whether a particular outcome differs between groups who receive different interventions or whether the outcome is associated with some other factor. The research designs and analytic methods used to assess differences between groups receiving different interventions may be relatively familiar to physical therapists; those used to identify the association between 2 factors and to ascertain the probability that the association is likely to have occurred by chance alone may be less familiar to therapists. If, for example, one is interested in the association between the time it takes to climb 10 stairs and isokinetic knee extensor strength in some patient population, the magnitude of the relationship can be determined using statistical procedures, such as the Pearson product moment correlation coefficient or the coefficient of determination ( $r^2$ ). However, these statistics are based on the assumption that the outcome variable (eg, the time it takes to climb 10 stairs) is a continuous variable and that the data meet certain assumptions (ie, normal dis-

tribution and equal variance between groups). If the variable is not a continuous variable but is measured on at least an ordinal level, the Spearman rank-order correlation coefficient can be used. Frequently, however, we are interested in ascertaining the association between 2 variables that are measured on a nominal scale consisting of only 2 levels (referred to as dichotomous variables). For example, we may be interested in whether a risk factor is associated with the presence or absence of low back pain or whether the use of a new mobility aid is associated with an increased likelihood that a patient will be discharged to home as opposed to discharged to a supervised care facility. For these types of questions, the chi-square ( $\chi^2$ ) statistic is commonly used.

The probability associated with the  $\chi^2$  statistic reflects the likelihood that the association between 2 dichotomous (categorical) variables is due to chance. When a  $\chi^2$  value is statistically significant at a criterion alpha level of 0.05, we can be at least 95% sure that the observed association did not occur by chance alone. Neither the value of the  $\chi^2$  nor the probability level, however, estimate the magnitude of association between the 2 variables. The magnitude of an as-

Send correspondence to Pamela Levangie, Sacred Heart University, 5151 Park Avenue, Fairfield, CT 06432-1000. E-mail: levangiep@sacredheart.edu

sociation is a key element in understanding the clinical relevance of the relationship between 2 variables. Just as we would be remiss if we simply determined that some treatment is statistically better than another without also asking "How much better?" we would also be remiss if we only tested whether or not 2 variables were related without also asking, "To what degree?" One measure that permits estimation of the magnitude of association between 2 dichotomous variables is the odds ratio (OR).

### Estimating Magnitudes of Association Between Dichotomous Variables

The OR and the relative risk (RR) analysis from which it evolved are used in epidemiology to assess the magnitude of association between a negative exposure (risk factor) and a disease. We can apply the same concept to physical therapy research questions when we explore relations, such as the association of a risk factor (poor balance) to a "disease" (a fall or a hip fracture). The common usage in epidemiology and the example just given assume that the outcome is a negative (unwanted) event and that the exposure may *increase* the likelihood of the negative event. The conceptual framework, however, has broader applicability. One can also estimate the magnitude of association: (1) when the "disease" is a negative outcome but the "exposure" is thought to *decrease* the likelihood of that negative outcome, and (2) when the "disease" is a positive rather than a negative outcome, and the "exposure" is intended to increase the likelihood of that outcome. Consider 2 examples:

**EXAMPLE 1.** You wish to estimate the association of an ergonomically sound workspace (subjects have that positive exposure or do not), and what you hope to find is a decreased likelihood of a negative outcome like carpal tunnel syndrome (CTS). Here, CTS (the disease) is a negative outcome, but it is anticipated that exposure to an ergonomically sound workspace will decrease the likelihood of the disease.

**EXAMPLE 2.** You wish to estimate the magnitude of the association between a new intervention for athletes with anterior cruciate reconstruction (treated or untreated) and the likelihood that the athlete will "return-to-sport" or not. Here, the positive outcome of interest that takes the place of disease is return-to-sport (or some other measure of success). The exposure that we believe will positively influence the outcome of return-to-sport is the new intervention component.

While OR and RR analyses have broad and flexible applicability to physical therapy research, an understanding of ORs is best developed if we first

		Disease	
		Cases (diseased)	Controls (not diseased)
Exposure	Exposed	A	B
	Not exposed	C	D

$$RR = \frac{A/A+B}{C/C+D} \quad \hat{OR} = \frac{(A)(D)}{(B)(C)}$$

FIGURE 1. Both an odds ratio ( $\hat{OR}$ ) and a relative risk (RR) are calculated from values in a 2 x 2 contingency table. A, B, C and D represent the number of subjects that fall into each cell.

examine the exposure/disease concept and RR analysis from which ORs emerged. We will then turn to the physical therapy-related examples and research literature to apply that understanding.

### Relative Risk

Relative risk is a measure of association between the presence or absence of disease and the presence or absence of exposure to a potential risk factor. One typically finds a RR analysis in a cohort study, where the exposure for all members of the cohort is ascertained, and the cohort is followed forward in time to ascertain later disease status. The RR is calculated by counting the number of individuals in the cohort with and without the exposure and counting the number of individuals who develop the disease (cases) and who do not develop the disease (noncases). The resulting frequency data typically are presented in a 2 x 2 contingency table as shown in Figure 1. Referring to the counts in Figure 1, the RR may be computed as:

$$RR = \frac{\text{Incidence of disease among exposed individuals (or } A/A + B)}{\text{Incidence of disease among unexposed individuals (or } C/C + D)}$$

If the exposure does not affect disease occurrence, the incidence of disease will be similar for both exposed and unexposed individuals; that is, the RR will be 1.0 under the null hypothesis of no association between exposure and disease. If exposure does increase the incidence of disease, the RR will be greater than 1.0 because the numerator will be larger than the denominator. If the exposure provides some

protection from disease, the RR will be less than 1.0 because the denominator will be less than the numerator.

## Limitations and Alternatives

In physical therapy research, as in epidemiologic research, it is both time-consuming and expensive to ascertain the exposure in an entire population (cohort) of interest and then follow that cohort forward in time to monitor outcome status. A simpler and less expensive alternative is to obtain a sample of persons, some with disease (cases) and some without disease (controls), and then determine their exposure status either concomitantly or retrospectively. In these case-control designs, the actual incidence of disease cannot be calculated because an entire population of interest (intact cohort) is not available and because subjects are selected on the basis of their disease status. While the RR cannot be calculated, the OR can be calculated and can serve as an estimate of the RR. A case-control design and OR analysis are particularly useful in providing preliminary evidence of an association before undertaking a more expensive and lengthier cohort or experimental study. However, due to differences in the way subjects are identified for the case-control design (ie, disease status) compared to the prospective cohort design (ie, exposure status), you will see in the next section that the wording of the expressions is slightly different for the RR than for the OR.

## Odds Ratio

The OR is a ratio of the odds of exposure in cases (those with the disease) and the odds of exposure in controls (those without the disease). The 'odds' for each group (cases or controls) is the proportion of individuals in the group who had been exposed, divided by the proportion of those from that group who had not been exposed. Referring again to the 2 × 2 contingency table in Figure 1, the OR mathematically simplifies as follows:

$$\frac{\left. \frac{A/(A + C)}{C/(A + C)} \right\} \text{Odds of exposure among cases}}{\left. \frac{B/(B + D)}{D/(B + D)} \right\} \text{Odds of exposure among controls}} = \frac{A/C}{B/D} = \frac{AD}{BC}$$

As long as cases and controls have been chosen *independent* of exposure status, the OR calculation has been shown to be conceptually and mathematically similar to the RR<sup>14</sup> and conceptually can be considered equivalent to:

$$\text{OR} = \frac{\text{(Odds of exposure among individuals with disease)}}{\text{(Odds of exposure among individuals without disease)}}$$

$$\div \frac{\text{(Odds of exposure among individuals without disease)}}{\text{(Odds of exposure among individuals without disease)}}$$

From this formula, we can see that the OR will be 1.0 if the odds of exposure are similar among subjects with and without disease. An OR > 1.0 indicates an increased likelihood of exposure among diseased subjects (a positive association between exposure and disease), while an OR < 1.0 indicates a decreased likelihood of exposure among diseased individuals (a negative association between exposure and disease).

It should be noted that the value of the OR obtained from the 2 × 2 contingency table will be mathematically the same whether you are estimating the risk of exposure given disease or estimating the risk of disease given exposure. Conceptually, however, the interpretation of the OR should be based on study design. In a case-control study where ORs must be used because a RR cannot be computed, subjects are entered into the study based on disease status; consequently, the OR estimates the likelihood of exposure *given disease*. If one uses an OR in a prospective follow-up study (where either a RR or an OR can also be used), subjects are entered into the study based on exposure status; consequently, the OR estimates the likelihood of disease *given exposure*.<sup>11,13</sup> The previous examples can be used to illustrate the different interpretations.

**EXAMPLE 1.** In this example, subjects are chosen on the basis of whether they have CTS or not (disease), and then their ergonomic environment (exposure) is ascertained. Consequently, the OR would be computed as below and interpreted as the likelihood that subjects were exposed to an ergonomically sound workspace given they have CTS.

$$\text{OR} = \frac{\text{Odds of exposure to ergonomically sound workspace given CTS}}{\text{Odds of exposure to ergonomically sound workspace given no CTS}}$$

**EXAMPLE 2.** In this example, subjects are chosen on the basis of whether they received the intervention or not (exposure), and then return-to-sport (disease) status is ascertained. Consequently, the OR would be computed as below and interpreted as the likelihood that subjects returned to sport given they received the intervention.

$$\text{OR} = \frac{\text{Odds of return-to-sport among those receiving the new intervention}}{\text{Odds of return-to-sport among those not receiving the new intervention}}$$

When reading a paper using an OR as a measure of association, you must first determine which factor

the authors labeled as the disease and which factor they labeled as the exposure. This is not always immediately evident. If the 2 factors are arranged in the  $2 \times 2$  format as depicted in the Figure 1, you can begin describing the OR in the context of the upper left cell, referred to as Cell A. Using Figure 1, now consider how our 2 clinical examples would be depicted and interpreted. In Example 1, Cell A would contain the number of individuals who received the ergonomic intervention *and* who had CTS. In this case, we would hope that the OR would be  $< 1.0$  because this would indicate that having CTS decreased the likelihood that subjects were in an ergonomically sound workspace (noting that subjects were selected based on CTS status, with exposure subsequently identified). In Example 2, Cell A would contain the number of individuals who received the intervention (the exposure) and who also returned to sport (the disease). Consequently, we would hope that the OR would be  $> 1.0$  because this would indicate that exposure to the new intervention increased the likelihood of returning to sport (noting that subjects were assigned to the intervention or no intervention groups and the outcome subsequently ascertained).

Using data from an actual case-control study as another example, the estimated OR for the association between low back pain and smoking among patients receiving physical therapy was 2.21.<sup>6</sup> Cell A in the  $2 \times 2$  contingency table in this example would include the number of subjects who had low back pain and were smokers. Given the hypothesis that smoking and low back pain are positively related, an OR  $> 1.0$  was expected. Because subjects were recruited based on low back pain status and smoking status was ascertained afterwards, the OR indicates that those with low back pain were 2.21 times more likely to be smokers than those without low back pain. Another way to express the magnitude of effect is that there is a 1.21 or 121% *increased* risk of smoking among those with low back pain (2.21 minus the null value of 1.0) compared to those without low back pain.

Another finding from the same study was an estimated OR of 0.89 for the association between men with low back pain and daily lifting of  $\geq 35$  pounds.<sup>6</sup> Because Cell A in the  $2 \times 2$  contingency table included men who had low back pain and lifted  $\geq 35$  pounds regularly, an OR of  $< 1.0$  indicates an *inverse* association between these factors. That is, those with low back pain were 0.89 times as likely to regularly lift 10–20 pounds compared to subjects who routinely lifted little or no weight. Alternatively, we could state that those with low back pain were 11% less likely to regularly lift  $\geq 35$  pounds than to regularly lift little or no weight (0.81 minus the null value of 1.0).

## 95% Confidence Intervals and Odds Ratios

When an estimated OR (designated by  $\hat{OR}$ ) is derived from a sample, rather than from population data, the value is expected to have some degree of error associated with it. If the estimate is going to be applied to appropriate groups outside the study (ie, generalized), the amount of error in the estimate must be characterized. One common way to characterize the amount of error that may exist around the  $\hat{OR}$  is to compute a 95% confidence interval (CI) for the value. The 95% CI identifies the range of values within which the 'true' OR will lie 95% of the time given the laws of probability.

### Calculation of 95% Confidence Intervals

For those who wish to understand confidence intervals quantitatively, it is easiest to begin with a generic formula, such as:  $X \pm 1.96 (SD_x)$ , where X is the point estimate (eg, a mean or an OR),  $SD_x$  is the standard deviation of the point estimate, and 1.96 is the z score associated with 95% of a normal curve.<sup>8</sup> The SD is an index of the variability (also referred to as error) of the point estimate. Because the SD represents how "scattered" values are around the point estimate, the more scatter there is, the larger the SD is. A z score of 1.96 represents slightly less than 2 SDs from the point estimate, assuming there is a normal distribution. Consequently, the 95% CI represents values that range from approximately 2 SD less than and 2 SD more than the point estimate. The larger the SD, the wider the 95% CI and the less precisely the  $\hat{OR}$  represents the true population value of the RR.

The computation of a 95% CI for an  $\hat{OR}$  differs slightly from the generic formula because the OR is based on dichotomous (categorical) disease and exposure variables and does not meet the assumptions upon which the generic 95% CI formula is based (ie, an SD cannot be computed for dichotomous variables).<sup>10</sup> The formula for computing the 95% CI for the OR is:

$$\exp\{\ln\hat{OR} \pm 1.96(SD[\ln\hat{OR}])\}$$

While a complete explanation of the basis of the computation is beyond the scope of this paper, the calculation of the 95% CI for the OR is based on a normalization of the data using a logarithmic (natural log or ln) transformation. While the log transformation (ln) and exponentiation (exp) make the formula look more complicated, the general form of 95% CI =  $X \pm 1.96 (SD_x)$  is retained.

### Interpretation of Odds Ratio and 95% Confidence Interval

In our previous example of the association between low back pain and smoking, consider how the

95% CI values aid our interpretation of the  $\hat{OR}$ . The  $\hat{OR}$  of 2.21 had a corresponding 95% CI of 1.09–4.46.<sup>6</sup> Typically, the values are reported as using the format of  $\hat{OR} = 2.21(1.09, 4.46)$  or  $\hat{OR} = 2.21(1.09-4.46)$ . The literal interpretation of the results can be expressed as follows: the best estimate from these data is that those with low back pain were 2.21 times more likely to be smokers than those without low back pain; however, we can be 95% confident that the true likelihood of being a smoker, if not 2.21 times greater, is at least 1.09 times greater and may be as much as 4.46 times greater in those with low back pain than in those without low back pain. The practical interpretation of these findings might be phrased somewhat differently. Specifically, from these data low back pain appears to double the likelihood of being a smoker; however, it is possible that there is no actual increase in likelihood ( $\hat{OR} \approx 1.0$ ) or that there may be a three-and-a-half-fold increase in likelihood of smoking among those with low back pain.

Moreland and Thomson<sup>7</sup> used ORs in their meta-analysis to summarize the outcomes of several studies that assessed the likelihood of improvement on a functional test (disease) given exposure to a biofeedback intervention among patients with stroke. The summary  $\hat{OR}$  of 2.16 across analyzed studies had a 95% CI of 0.85–5.79. The  $\hat{OR}$  indicates that patients receiving biofeedback were more than twice as likely to improve on the functional test, whereas the CI indicates that we can be 95% sure that the actual value may be as low as 0.85 (indicating a 15% reduction in success with biofeedback) or as high as a nearly five-fold increase in success with biofeedback. While it is generally acceptable to say that we are “95% sure that the true value lies in the CI,” it should be pointed out that it is most correct to say that the true value will lie in the calculated interval in 95 out of 100 repetitions of the study. The 95% CI of 0.85–5.79 is fairly wide, indicating that the  $\hat{OR}$  of 2.16 is not very precise. As the data in any  $2 \times 2$  contingency table (Figure 1) become sparser (1 or more of the frequencies within the cells have small numbers), the OR will have more error because it is based on less data and, consequently, the data are likely to be more variable under repeated sampling. As the amount of error in the  $\hat{OR}$  increases, the CI must be widened to ensure with 95% certainty that the true value lies in the interval 95% of the time. Conversely, a narrower 95% CI indicates a more precise  $\hat{OR}$  because we are more certain that the true OR lies either at or close to the estimated OR.

### Odds Ratios, 95% Confidence Interval, and Hypothesis Testing

The 95% CI also can be used to estimate statistical probability for hypothesis testing. For the purpose of

hypothesis testing, the value of 1.0 is the value of the OR that is considered to be consistent with the null hypothesis. Therefore, if the value of 1.0 for the  $\hat{OR}$  lies within the 95% CI, the corresponding  $P$  value for the  $\hat{OR}$  will be *greater than* 0.05, and the null hypothesis of no association must be accepted. To further clarify, if 1.0 falls within the CI, then 1.0 is one of the possible estimates that the true value may take on with repeated testing, and we cannot be at least 95% sure that the true value is different than 1.0. For example, Moreland and Thomson<sup>7</sup> calculated an  $\hat{OR}$  and 95% CI of 2.16 (0.82, 5.79). Thus, we can be 95% sure that the true value lies within the interval, but because the interval includes the possibility of no association, we cannot be sure that the true value is something other than 1.0. When Moreland and Thomson tested for significance, they determined the corresponding  $P$  value for the  $\hat{OR}$  to be 0.07. Because 0.07 is not less than or equal to the criterion level of 0.05 that is always implicit in a 95% CI, the null hypothesis (no association) must be accepted. We cannot be 95% confident that biofeedback resulted in improvement as measured on functional tests.

Gadsby and Flowerdew<sup>2</sup> conducted a Cochrane-type review of the available evidence on transcutaneous electrical nerve stimulation and acupuncture-like transcutaneous electrical nerve stimulation (AL-TENS) for subjects with chronic back pain using improvement in pain as the outcome. Because AL-TENS was the exposure and improvement in pain was the “disease,” an  $\hat{OR} > 1.0$  was anticipated. They calculated an  $\hat{OR}$  and 95% CI of 7.22 (2.60, 20.01). Given the data and the fact that outcome ascertainment followed exposure ascertainment, the interpretation is that AL-TENS is 7.22 times more likely than a placebo to result in an improvement in pain. In addition, we can be 95% certain that the true effect of AL-TENS is no less than 2.6 times more likely to improve pain or may be as much as 20 times more likely to improve pain than a placebo. Although the 95% CI is quite wide (indicating an imprecise OR), the 1.0 value does not lie within the 95% CI for the  $\hat{OR}$ . Thus, the corresponding  $P$  value will be less than 0.05, and the null hypothesis can be rejected.

If you assessed the  $\hat{OR}$  based exclusively on whether the null value were in the 95% CI, there would be no benefit to 95% CIs over  $P$  values alone. You would simply accept or reject the  $\hat{OR}$ . The added value of the 95% CI is in the information gained about the precision of the estimated OR, especially when the estimate is not statistically significant ( $P > 0.05$ ) but is of potential clinical importance. In Moreland and Thomson’s study,<sup>7</sup> the  $\hat{OR}$  of 2.16 may indicate a clinically relevant advantage to biofeedback over conventional therapy alone. While that interpretation must be weighted by other considerations in that study (and in the contributing studies), it might be



short-sighted to rule out biofeedback as an adjunct treatment for stroke patients because the finding was not “statistically significant” (the null value was in the 95% CI of 0.82–5.79). By using the 95% CI, without regard to accepting or rejecting the null, we can say that the worst case in repeated testing would mean an 18% reduction ( $1.0 - 0.82 = 0.18$ ) in successful outcome with use of biofeedback. Alternatively, it is possible that the actual answer will be as much as a 479% increase in successful outcomes among those using biofeedback ( $5.79 - 1.0$ ) converted to percent. Understanding this range of outcomes seems preferable to dismissing biofeedback based on significance testing alone.

## Limitations to Use of Simple Odds Ratios

The OR is one of a very limited number of options that will allow assessment of a magnitude of association when either the outcome of interest is (or must be treated as) a dichotomous variable. The outcome may be naturally dichotomous (eg, discharged to home or not), but it may also make conceptual sense to dichotomize a continuous (eg, improved or not based on a 0–100 visual pain scale) or ordinal variable (eg, improved or not based on a functional score). As we saw for the outcome, the exposure may be naturally dichotomous (eg, delivery of an intervention or not). Often, however, the exposure can be measured as continuous or ordinal data. Logistic regression can be used to obtain an OR and a 95% CI to estimate the magnitude of association between a continuous exposure and a dichotomous outcome. However, this analytic method is unfamiliar, conceptually complex, and includes an assumption that the continuous exposure variable is exponentially related to the odds of the disease or outcome (which is frequently not the case).<sup>10</sup> In this article, we will continue to emphasize calculation of simple ORs because of their relative simplicity; however, use of simple ORs requires that exposure data measured on a continuous scale be reduced to dichotomous data to meet the requirements of the  $2 \times 2$  contingency table. In other words, there are trade-offs to consider in using both a logistic regression approach and a contingency table approach to calculating ORs when exposure data are continuous. While the issues with logistic regression are beyond the scope of this paper, we will look more closely at the issues involved in reducing continuous exposure variables too dichotomous to fit a  $2 \times 2$  contingency table.

## Reducing Continuous Variables to Dichotomous Variables

When either the outcome or the exposure data are reduced from categorical or continuous levels to dichotomous data so they fit in a  $2 \times 2$  contingency

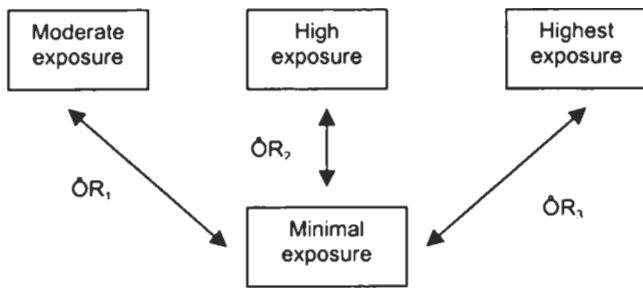
table as a preliminary step to computing a simple OR, some potential problems emerge. First, the cut-point for dichotomizing the variable is generally somewhat arbitrary (and arguable). Second, information is lost (is less precise) when the natural variability of the data is reduced into simple categories of positive or negative outcome and exposed or not exposed. The OR model implies an abrupt change in risk at the estimated cut-point, with the risk being constant within the defined categories.<sup>12</sup> This, of course, is rarely true.

Several approaches have been proposed that attempt to minimize the loss of information and bias toward the null that occur with dichotomizing an exposure variable in an OR analysis. However, once the decision is made to dichotomize data, the loss of information that occurs must be acknowledged as a limitation of the analysis because it cannot be directly controlled or offset. The strategy used to reduce a variable to a dichotomy is important to both the outcome of the analysis and to the reader when attempting to independently evaluate the clinical meaningfulness of the ORs estimated in the study.

## Categorization of Exposure Variables

One of the simplest strategies used to minimize the loss of information that accompanies dichotomization of a continuous exposure variable is categorization of the exposure variable into more than 2 levels (or categories) of exposure. Rather than having just 2 categories, such as exposed or not exposed, we could have a set of categories, such as unexposed, moderately exposed, and very exposed. This strategy essentially reduces the continuous variable to an ordinal rather than a dichotomous variable. We would expect the approach to yield more homogeneity within each category than would be likely if only 2 categories of the exposure were used. The strategy works particularly well if the sample size supports enough categories to produce narrow exposure ranges of biologically homogenous response groups.<sup>3</sup> As the number of categories is increased, the number of subjects in each category tends to decrease. As a result of the smaller number of subjects in each category, there will be tendency for the 95% CIs around the ORs to become wider because the error of the estimate is larger for smaller sample sizes.

When data from an exposure variable is categorized into multiple levels, several ORs are calculated to assess the association of the exposure with the outcome. The lowest category of exposure is typically used as the referent category. The referent category may be absence of exposure (eg, nonsmokers or those with no intervention). The referent may also be the lowest category of the exposure in situations where absence of exposure is conceptually impossible (eg, age or activity level). Two or more ORs are then



**FIGURE 2.** Categorization of a continuous exposure variable into several levels allows estimation of an odds ratio between exposure and disease for each higher level of exposure with the lowest level of exposure (referent).

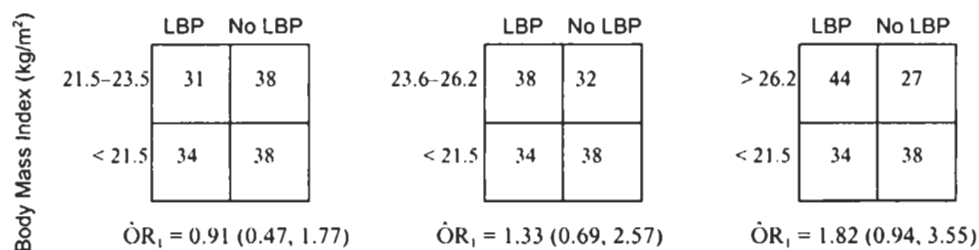
calculated by comparing each of the higher categories of exposure to the referent category (Figure 2).

When 2 or more  $\hat{OR}$ s are calculated for increasing levels of exposure, you can look for trends in the category-specific  $\hat{OR}$ s ( $\hat{OR}_1$ – $\hat{OR}_3$  in Figure 2) that would indicate a dose-response relation between the outcome and the exposure (eg, there is an increasing likelihood of the expected outcome with increased exposure). Unlike  $\hat{OR}$ s resulting from logistic modeling of a continuous exposure variable, this approach can be used to examine the linearity of the association across levels of exposure. For example, if, instead of increasing across all levels of exposure, the association increases across 2 levels of exposure but plateaus or decreases across others, the association is described as nonlinear rather than linear.

As is true for dichotomizing data, categorizing the exposure variable still requires specification of values for cut-points, and those values are likely to be somewhat arbitrary and arguable. A common strategy is to use quartiles or quintiles as the cut-points between categories.<sup>14</sup> If 4 exposure categories are developed based on quartiles, 3  $\hat{OR}$ s will result because the lowest quartile will be the referent for each of the 3 higher levels of exposure. While use of strategies like quartiles or quintiles reduces some of the arbitrariness of determining cut-points for the exposure data, subjects in the study must be sufficient in number and distribution to have subjects in each cell of each of the  $2 \times 2$  tables, or the  $\hat{OR}$  and 95% CI computations cannot be done. Again, it is particularly important that the rationale for any cut-point(s) be speci-

fied by the authors so that the reader can evaluate the strategy.

For an example of how an exposure variable can be categorized, let's consider estimation of the association between body mass index (BMI) as the exposure and low back pain as the outcome.<sup>6</sup> If we dichotomized BMI, Cell A in the  $2 \times 2$  contingency table would indicate the number of individuals with high BMI and low back pain. Consequently, an  $OR > 1.0$  would indicate that low back pain is associated with an increased likelihood of high BMI. We would ascertain the self-reported height and weight in a sample of physical therapy patients being treated either for low back pain or for an upper extremity problem.<sup>6</sup> BMI would be calculated for each subject, but we would be concerned that dichotomizing the data would "wash out" a possible positive association between BMI and low back pain because there would be a fairly broad range of BMIs represented in both the 'low' and 'high' BMI group. While logistic regression is an option, we would have to assume that there is an exponential relationship between BMI and the odds of low back pain, an assumption that we would not be comfortable making. Consequently, we would ascertain the quartile distribution for BMI values for patients with and without low back pain as a preliminary step to computing simple  $OR$ s within levels of BMI. Each of the categories would include individuals who are relatively similar in BMI, or at least more similar than if the data had been dichotomized. The category with the lowest values is used as the referent because this is the group that is *least* likely to be an increased risk for low back pain based on BMI. The referent BMI values are indicated at the bottom row in each of the  $2 \times 2$  contingency tables in Figure 3. The ranges for the moderate, high, and highest levels of BMI are found in the upper row of each of the 3 tables. The  $\hat{OR}$ s and 95% CIs comparing each of the upper quartiles (with increasing magnitudes of overweight) to the lowest quartile are calculated and reported in Figure 3. Because the upper 3 quartiles are compared to the referent separately, the estimates are less likely to be biased by mixing people with different levels of overweight and different potential risks. The 3 estimates also allow assessment of the association between BMI and



**FIGURE 3.** Association of low back pain (LBP) with levels of self-reported body mass index: frequencies and estimated odds ratios ( $\hat{OR}$ s) with 95% confidence intervals.



low back pain across increasing levels of exposure without making any assumptions about their relationship to each other. As can be seen, the ORs show some evidence of a linear trend for increased risk of low back pain with increasing levels of BMI. However, the 95% CIs are wide (the estimates are not very precise), and each of the CIs includes the null value of 1.0 (the estimates are not significant at the  $P < 0.05$  level). These data are not, therefore, conclusive of an association between BMI and low back pain, although the importance of the trend can be left to individual judgment based on what the reader believes to be the strengths and weaknesses of the analyses. There are other strategies beyond the scope of this paper that permit further analysis of the association between an outcome and an exposure when ORs are used. These include methods for assessing effect modification (where the OR differs across categories of a covariate) and for assessing confounding (where the potential effects of a confounding covariate on the OR can be ascertained), either by performing stratified analyses or by using logistic regression. The interested reader may pursue these more complex methods and the use of logistic regression in textbooks of epidemiology.<sup>5,9,11,13</sup>

## CONCLUSION

Simple ORs and 95% CIs can be used when you wish to find an estimate of the magnitude of association between a dichotomous outcome variable and a dichotomous exposure variable. Given the range of applications for the OR, physical therapists can anticipate seeing such analyses increasingly in the physical therapy research literature. It is important, therefore, that the practitioners develop sufficient understanding of the application, interpretation, and limitations of the method so that they have some ability to inde-

pendently assess the weight of evidence presented in such research.

## REFERENCES

1. Fletcher R, Fletcher S, Wagner E. *Clinical Epidemiology—the Essentials*. Baltimore, Md: Williams & Wilkins; 1982.
2. Gadsby J, Flowerdew M. Transcutaneous electrical nerve stimulation and acupuncture-like transcutaneous electrical nerve stimulation for chronic low back pain. *Cochrane Database Syst Rev*. 2000: CD000210.
3. Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology*. 1995;6:356–365.
4. Hennekens CH, Buring JE. *Epidemiology in Medicine*. In: Mayrent SL, ed. Boston/Toronto: Little, Brown and Co; 1987.
5. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. New York, NY: Van Nostrand Reinhold Co; 1982.
6. Levangie P. Association of low back pain with self-reported risk factors among patients seeking physical therapy services. *Phys Ther*. 1999;79:757–766.
7. Moreland J, Thomson M. Efficacy of electromyographic biofeedback compared with conventional physical therapy for upper-extremity function in patients following stroke: a research overview and meta-analysis. *Phys Ther*. 1994;74:534–547.
8. Portney L, Watkins M. *Foundations of Clinical Research: Applications to Practice*. 2nd ed. Upper Saddle River, NJ: Prentice Hall Health; 2000.
9. Rothman K, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia, Pa: Lippincott Williams and Wilkins; 1998.
10. Rothman KJ. *Modern Epidemiology*. Boston: Little, Brown and Co; 1986.
11. Schlesselman JJ. *Case-Control Studies*. New York, NY: Oxford University Press; 1982.
12. Schulgen G, Lausen B, Olsen JH, Schumacher M. Outcome-oriented cutpoints in analysis of quantitative exposures. *Am J Epidemiol*. 1994;140:172–184.
13. Selvin S. *Statistical Analysis of Epidemiological Data*. 2nd ed. New York, NY: Oxford University Press; 1996.
14. Thompson WD. Statistical analysis of case-control studies. *Epidemiol Rev*. 1994;16:33–50.