

Review Article

TREATMENT EFFECTS 101

Thelma J. Mielenz ^{*1}, Janet K. Freburger ², Timothy S. Carey ^{2,3}.

^{*1} Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York 10032.

² Cecil G. Sheps Center for Health Services Research, University of North Carolina, Chapel Hill, North Carolina 27599.

³ Departments of Medicine and Social Medicine, University of North Carolina, Chapel Hill, North Carolina 27599.

ABSTRACT

Physical therapy researchers are interested in how beneficial an intervention is or the "treatment effect." There are many measures of treatment effect that are applicable for understanding the efficacy and effectiveness of health interventions. Given that each treatment effect has its own set of advantages and disadvantages, understanding these characteristics can help guide which measure is most appropriate for a specific study. This article presents the more common treatment effects for both dichotomous and continuous outcomes. The overall aim is to serve as a guide to newer physical therapy researchers on using and interpreting treatment effects.

KEY WORDS: Treatment effects, Relative measures of effect, Absolute measures of effect.

Address for correspondence: Thelma J. Mielenz, PhD, Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 West 168th Street, New York, NY 10032, Phone: 212-342-0169, Fax: 212-342-5168, **E-Mail:** tjm2141@columbia.edu

Access this Article online

Quick Response code



DOI: 10.16965/ijpr.2015.173

International Journal of Physiotherapy and Research

ISSN 2321- 1822

www.ijmhr.org/ijpr.html

Received: 31-07-2015

Accepted: 09-08-2015

Peer Review: 31-07-2015

Published (O): 11-10-2015

Revised: None

Published (P): 11-10-2015

INTRODUCTION

When physical therapy (PT) researchers assess the results of studies on the efficacy or effectiveness of interventions, they are most interested in determining whether the intervention is beneficial, and if so, how beneficial? They are essentially interested in the effect of the intervention or the "treatment effect." The treatment effect is defined as the amount of change in the outcome that is caused by the intervention [1].

Traditionally, PT researchers have reported the results of studies on the efficacy or effectiveness of interventions as "statistically significant" or "statistically non-significant"

based on an arbitrary p-value threshold (e.g., $p < .05$) [2]. Just because a p-value is reported as statistically significant does not mean the effect of the intervention is large, nor does a p-value reported as non-significant imply that the effect of the intervention is trivial [3]. Studies with large samples may have statistically significant findings, but the effects may be small. Alternatively, studies with small samples may have statistically non-significant findings, but the effects may be large. Today, PT researchers are moving away from the idea that results are statistically significant or non-significant. Although the actual p-value may still be used to interpret study results (i.e., $p=0.003$ versus $p<0.05$),

more emphasis is being placed on measures of treatment effect [2,4].

There are many measures of treatment effect that can be used to report the results of a study on the efficacy or effectiveness of an intervention [5,6]. All of these measures have advantages and disadvantages. The choice of which measure of treatment effect to use is based on the research question, the research design, and how the intervention and outcome variables are measured [5]. The treatment effect measure should also be relevant to the underlying goal of the research project, intelligible to the intended audience, and appropriate from a statistical perspective (i.e., assumptions of the statistical test used to generate the treatment effect measure are not violated).

The purpose of this article is to present some of the more common ways to characterize treatment effects and to discuss some of the advantages and disadvantages of using these measures. This article should help guide researchers in determining which measures of treatment effect may be appropriate for studies on PT interventions. Many of the measures that are covered come from the epidemiology literature where the exposure is often a behavior, characteristic, or risk factor of a subject, and the outcome is commonly the occurrence of a disease. This document focuses on the intervention or treatment as the exposure and on the more generic term of "outcome."

The remainder of this article is organized as follows. We first present basic definitions of epidemiology terminology (Appendix A). Understanding these terms are necessary for understanding treatment effect definitions. Next, measures of treatment effect that are used when the outcome is dichotomous are presented (i.e., the outcome does or does not occur). Definitions, advantages, and disadvantages are provided. Measures of treatment effect that are used when the outcome is continuous are then presented in a similar manner.

MEASURES OF TREATMENT EFFECT FOR DICHOTOMOUS OUTCOMES

In epidemiology, there are multiple ways to consider treatment effects, from relative

measures to absolute measures. Relative measures are ratios of occurrence measures while absolute measures are differences in occurrence measures [7]. Relative effects give information on the strength of association between exposure and disease and help in understanding the causes of disease. Absolute effects provide information on the magnitude of the effect of the exposure on the outcome from a population perspective (e.g., the number of people affected). The following sections illustrate ways to characterize treatment effect in relative and absolute terms when the outcome is dichotomous.

Relative Measures of Effect: Relative measures of effect are ratios of prevalence, incidence proportions, and incidence rates [8]. These measures can be used to characterize the effects of a treatment by comparing those individuals who received the treatment under study (treatment group) to those individuals who received another treatment or no treatment (reference group) [8]. Tables 1 and 2 illustrate these relative measures.

Risk Ratio or Relative Risk (Table 1)

$$\frac{\text{risk or cumulative incidence in treatment group}}{\text{risk or cumulative incidence in reference group}}$$

Table 1: 2 X 2 Table for Risk.

	Achieved Outcome	Did Not Achieve Outcome	Total # of Subjects
Tx Group	a	b	a + b
Ref Group	c	d	c + d
	a+c	b+d	

$$\text{Risk Ratio (RR)} = \frac{a/(a+b)}{c/(c+d)}$$

Prevalence Ratio (same formula as the Risk Ratio but uses prevalent cases)

$$\text{Odds Ratio (OR)} = \frac{(a/b)}{(c/d)} = \frac{ad}{bc}$$

Prevalence Odds Ratio (same formula as the Odds Ratio but uses prevalent cases)

$$\text{Exposure Odds Ratio (EOR)} = \frac{(a/c)}{(b/d)} = \frac{ad}{bc}$$

$$\text{Risk Difference (RD)} = a/(a+b) - c/(c+d)$$

Risk ratios cannot be calculated in case-control and cross-sectional studies because incidence cannot be measured directly [7]. In these types of studies, the odds ratio is calculated as an alternative to the risk ratio.

Odds Ratio (Table 1): The odds ratio is the odds of having the outcome in the treatment group divided by the odds of having the outcome in the reference group; or in case-control studies, the odds of being exposed in the cases (i.e., subjects with the outcome) divided by the odds of being exposed in the controls (i.e., subjects without the outcome) [7]. Odds ratios are frequently calculated using multivariable logistic regression modeling [9].

Odds Ratio formula for cohort studies and clinical trials [10]:

$$\frac{\text{odds of incident outcome in treatment group}}{\text{odds of incident outcome in reference group}}$$

Exposure Odds Ratio formula for case-control study [10]:

$$\frac{\text{odds of exposure among cases}}{\text{odds of exposure among controls}}$$

Prevalence Odds Ratio formula for cross-sectional studies [10]:

$$\frac{\text{odds of prevalent outcome in treatment group}}{\text{odds of prevalent outcome in reference group}}$$

Prevalence Ratio for Cross-sectional Studies (Table 1): The prevalence ratio is another measure of treatment effect that can be calculated for a cross-sectional study when the outcome has a short period of risk (i.e., acute diseases) [11]. The prevalence odds ratio mentioned above is preferred when the period of risk is longer (i.e., chronic diseases).

$$\frac{\text{prevalence of outcome in treatment group}}{\text{prevalence of outcome in reference group}}$$

Rate Ratio or Relative Risk [11] (Table 2)

$$\frac{\text{incidence density in treatment group}}{\text{incidence density in reference group}}$$

Table 2: 2 X 2 Table for Rates.

	Did Not Achieve Outcome	Did Achieve Outcome	Total Amount of Time Contributed by Subjects
Tx Group	a	b	
Ref Group	c	d	Person-time in Tx Group
			Person-time in Ref Group

Rate Ratio = $\frac{a/\text{Person-time in Tx Group}}{c/\text{Person-time in Ref Group}}$

Rate Difference = $\frac{a/\text{Person-time in Tx Group} - c/\text{Person-time in Ref Group}}$

Risk Ratio (RR) = $\frac{3}{80} = 0.0375 = 4\% = .5 \text{ or } 50\%$

Difference Between Relative Risk and Odds Ratio: Relative risk and odds ratios are different measures [7]. There are instances, however, when the odds ratio approximates the relative risk [7]. For example, in case-control studies, where risk cannot be directly estimated, the odds ratio is usually a good estimate of the relative risk even when the incidence of the outcome is greater than 10% [12].

Because multivariable logistic regression is a common statistical method used in cohort studies, odds ratios are usually reported [9]. The odds ratio approximates relative risk in cohort studies when the incidence of the outcome is infrequent (i.e., < 10%) [12]. Nevertheless, when the incidence of the outcome is common (i.e., >10%) and the size of the treatment effect is moderate to large; the odds ratio does diverge from the relative risk (according to Zhang et al. [13], a moderate to large treatment effect is defined as an odds ratio greater than 2.5 or a decrease in risk of .5 or less). The odds ratio will overestimate the relative risk when the odds ratio is greater than one and it will underestimate the relative risk when the odds ratio is less than one. There are different methods to estimate relative risk from an adjusted odds ratio derived from multivariable logistic regression when cohort studies have a common outcome [9], but alternative regression techniques are preferred and are beyond the scope of this document.

Relative Risk Reduction: The proportion or percent by which the incidence of the outcome in the treatment group decreases relative to the incidence of the outcome in the reference group [14].

$$\frac{\text{Incidence in the Treatment Group} - \text{Incidence in the Reference Group}}{\text{Incidence in the Reference Group}}$$

Hazard Ratio: A hazard ratio is the outcome in a time to event analysis or a survival analysis using a proportional hazards regression model [15]. It is the ratio of two hazard functions. A hazard function is a measure of the number of outcomes per interval of time [15]. The hazard ratio is sometimes described as the instantaneous incidence densities between two groups.

Absolute Measures of Effect: Absolute measures of effect are the differences in incidence rates or proportions [8]. These measures can also be used to characterize the effects of a treatment by comparing those individuals who received the treatment under study (treatment group) to those individuals who received another treatment or no treatment (reference group) [8].

Risk Difference or Absolute Risk Reduction (Table 1): Risk difference (RD) or absolute risk reduction is defined as the difference in cumulative incidence in the treatment group and the reference group [11]. It describes the absolute change in risk that is attributable to the treatment intervention. For example, if successful outcomes were achieved in 42% of treatment participants and 23% of control participants, the RD would be 19%.

$$RD = 42\% - 23\% = 19\%$$

In other words, this means that we can estimate that 19%, or about one in five, participants are expected to have a successful treatment outcome that is directly attributable to the treatment above the success rate that we expect from the placebo.

Number Needed to Treat (NNT): NNT is defined as the inverse of the risk difference or absolute risk reduction (one divided by risk difference) [8,16]. The NNT is the number of patients that would need to be treated in order to derive one positive outcome [8,17]. NNT is considered a useful, clinically meaningful way for reporting binary outcomes from randomized control trials. For example, in a clinical trial assessing the efficacy of physical therapy for the treatment of low back pain, if 30 of the 50 patients in the PT treatment group had pain relief and 10 of the 50 patients in the control group (i.e., reference group) had pain relief, the NNT would be 2.5.

$$NNT = 1/(30/50 - 10/50) = 1/(0.6 - 0.2) = 1/0.4 = 2.5$$

NNT's are rounded up to a whole number. In this example, a NNT of 2.5 would mean that 3 patients with low back pain would have to be treated with PT for one to have a positive outcome. An NNT of 1 would be ideal because it would mean that every patient benefited from the treatment [16].

NNT can also be calculated by taking the inverse of the quantity of the risk in the control group multiplied by the relative risk reduction [8]. If the risk in the control group is 0.2 and the relative risk reduction is 2, then the NNT would be 2.5.

$$NNT = 1/(0.2)(2) = 1/0.4 = 2.5$$

Number Needed to Harm (NNH): NNH is also the inverse of the absolute risk reduction, but is interpreted as the number of patients that would need to be treated in order to have one negative outcome [8]. For example, in a clinical trial with cervical manipulation as the treatment, if 3 out of 5000 patients in the treatment group had a vertebral artery dissection and 1 out of 5000 patients in the control group had this outcome, the NNH would be 2500.

$$NNH = 1/(3/5000 - 1/5000) = 1/(0.0006 - 0.0002) = 1/0.0004 = 2500$$

In this example, the NNH of 2500 would mean that treating 2500 patients with cervical manipulation would lead to one additional patient being harmed by vertebral artery dissection compared to patients not receiving cervical manipulation. This is a hypothetical example.

Rate Difference (Table 2): The rate difference is the difference in incidence density in the treatment group and reference group [14]. For example, in a 12-week fall prevention intervention study, falls incidence in the exercise (treatment) group is 6.0 falls/1000 activity hours compared to 16.2 falls/1000 activity hours for the control group. The rate difference is 10.2/1000 activity hours.

$$RD = (6/1000) - (16.2/1000) = 10.2/1000 \text{ activity hours}$$

In this example, if exercise and falls were not associated, we would expect the two incidences to be very close, and thus the difference between the incidence rates would be near zero. A rate difference not equal to zero indicates an some degree of association between the exposure and the outcome. The larger the difference, the stronger the association.

Advantages/Disadvantages of Relative and Absolute Measures

Relative measures of treatment effect can be misleading because they do not tell us the magnitude of the absolute risk reduction [5].

Consider a study reporting that the relative risk of falling decreased by 50% for persons who received PT. This could mean that the treatment decreased the risk of falling from 8% to 4% or from 90% to 45%. Had an absolute risk reduction been reported, 4% or 45% respectively, the clinical implications of the findings would be apparent. Tables 3 and 4 illustrate this example.

Table 3: Calculation of Relative and Absolute Measure of Effect (RR 50%, RD 4%)

	Falls	No Falls	Total Number of Subjects
Tx Group	3	77	80
Ref Group	6	74	80

$$6/80, 0.075, 8\%$$

$$\text{Risk Difference (RD)} = 4\%$$

Table 4: Calculation of Relative and Absolute Measure of Effect (RR 50%, RD 45%).

	Falls	No Falls	Total Number of Subjects
Tx Group	36	44	80
Ref Group	72	8	80

$$\text{Risk Ratio (RR)} = \frac{36/80}{72/80} = \frac{0.45}{0.90} = \frac{45\%}{90\%} = .5 \text{ or } 50\%$$

$$\text{Risk Difference (RD)} = 45\%$$

Using relative measures to report differences in treatment effects across different subgroups can be also be misleading for the same reason [5]. The baseline risk of the outcome could vary across subgroups. A 50% decrease from a baseline risk of 80% is much greater than a 50% decrease from a baseline risk of 5%. Because of these shortcomings, relative measures have limited usefulness in making treatment decisions. Nevertheless, relative measures of effect are often reported in the literature. Although there has been a recent uptake of statistical programs with computational capabilities which can convert these relative measures to absolute measures. Odds ratios, in particular, are commonly reported. One reason for this may be that multiple logistic regression analysis is often used when the study outcome is dichotomous [9,10,18]. This type of analysis, which generates odds ratios, is one method that

can be used to control for confounders while assessing the effect of the intervention on the outcome [18,19].

The biggest advantage of absolute measures of treatment effect is that they reflect the magnitude of the change in risk and, therefore, have greater clinical utility [20]. The consequence of giving no treatment or a standard/comparison treatment is estimated with absolute measures. Absolute measures of treatment effect are also easy to calculate and understand by both clinicians and patients. A 20% decrease in risk, for example, is clear and unambiguous. As with relative measures, absolute measures do not reflect the magnitude of the baseline risk [12]. For example, a risk difference of 10% may represent a decrease in the risk of the outcome from 20% to 10% or from 90% to 80%.

NNT, which has received a lot of attention in the literature, also has some specific advantages and disadvantages. NNT allows for relatively simple cost analyses [21]. For example, with an NNT of 2, the cost of achieving one positive outcome would be equal to the cost of treating two patients. NNT is also easy for patients to understand. For example, with an NNT of 3, a clinician could tell a patient that on average one in three patients will have a positive outcome. Unfortunately, the NNT alone cannot determine whether the patient being seen by the clinician will be the one who has the positive outcome. Research that determines the characteristics of patients most likely and least likely to benefit from an intervention may help to address this issue.

Calculating an NNT from a study comparing one intervention to another (versus comparing an intervention to no intervention) can also be problematic [17]. For example, a NNT of 100 for a multicomponent exercise intervention on fall prevention versus stabilization strengthening exercise intervention on fall prevention may be acceptable if the NNT for the multicomponent exercise intervention relative to no exercise is 10 and the NNT for the strengthening exercise intervention relative to no exercise is 15. Therefore, when looking at the effectiveness of one intervention relative to another, reporting the risk difference may be a more useful way to assess the treatment effect [17].

Calculating confidence intervals for the NNT are also cumbersome for mathematical reasons [22]. To circumvent this problem, one could calculate confidence intervals for the risk difference (the inverse of which is the NNT) and present this information along with the NNT.

MEASURES OF TREATMENT EFFECT FOR CONTINUOUS OUTCOMES

The measures of treatment effect presented in the previous section are applicable when the outcome is dichotomous. Oftentimes, however, the outcome of interest is continuous. Below are some ways to characterize treatment effects when the outcome is continuous.

Absolute Mean Difference: The absolute mean difference is defined as the difference between the treatment and reference group means or between pre-treatment and post-treatment means [7]. One disadvantage of this measure is that it does not take into account the standard deviation of the means. A difference of 5 from mean scores with large standard deviations [e.g., 20 (12) and 15(10)] is different than when the means scores have a small standard deviation [e.g., 20 (2) and 15 (2)].

Effect Size: Effect size is a name given to a family of indices that measure the magnitude of a treatment effect [23,24]. The most common measure of effect size is the standardized difference between two means (i.e., difference between the means divided by the standard deviation) [24]. Various formulas are available to calculate the standardized difference between two means. These formulas vary based on the derivation of the standard deviation [24]. Three measures of standardized differences between means for independent groups are presented below and are based on ANOVA methods [25]. In theory, Cohen's *d* is a conceptual formula because the population standard deviation of a group is rarely known [24].

$$\text{Cohen's } d = \frac{\text{treatment group mean} - \text{reference group mean}}{\text{population standard deviation}^*}$$

* control group, treatment group or pooled

$$\text{Glass' } \Delta = \frac{\text{treatment group mean} - \text{reference group mean}}{\text{standard deviation of reference group}}$$

$$\text{Hedges' } g = \frac{\text{treatment group mean} - \text{reference group mean}}{\text{pooled standard deviation of treatment and reference group}}$$

When computing effect sizes for group means that are not independent (e.g., pre-treatment and post-treatment means) the original standard deviation (e.g., standard deviation of pretreatment scores) should be used [24].

Advantages/Disadvantages of Absolute Mean Difference and Effect Size:

The calculation of both the absolute mean difference and the effect size are straightforward and simple. When the outcome measure is familiar, reporting the absolute mean difference is likely to be a sufficient way to characterize the effect of an intervention. For example, most clinicians would be able to interpret a 20-degree difference in knee range-of-motion (ROM) between the treatment group and the reference group. If the outcome measure is unfamiliar, reporting the absolute mean difference between groups may be difficult to interpret. For instance, a clinician might not understand what a 10-point difference means on a newly developed scale used to measure balance. The use of effect size is controversial. Some believe that effect size is useful in cases when the scale or the meaning of the measure is not familiar because it can be interpreted independent of the actual values of the outcome measure [24]. For example, an effect size of 0.8 can be interpreted as follows: 79% of the subjects in the reference group have an outcome measure below the average person in the treatment group. Others argue that because effect size is a unitless measure, it lacks clinical relevance and interpretability [25].

Calculation of an effect size is based on the assumption that the outcome is normally distributed and the standard deviation used to calculate the effect size is a good estimate of the standard deviation for the population being studied [24-26]. Because these two assumptions are often not met in studies on the efficacy or effectiveness of an intervention, some argue that effect size is not a useful measure of treatment effect and that comparisons of effect sizes across studies are inappropriate [24,27]. If the standard unit used to compute an effect size (i.e., the standard deviation of the treatment group,

the control group, or the pooled standard deviation) varies across studies, comparisons across studies will be invalid [27]. For example, consider two clinical trials that both reported the mean improvement in Timed-Up-and-Go scores was 12 seconds for patients who received physical therapy relative to patients who did not. The effect sizes for these two studies would differ if the standard deviations for the scores in the two studies were different. For example, if the standard deviation of the Timed-Up-and-Go scores was 12 seconds in study A and 6 seconds in study B, the effect sizes would be 1.0 and 2.0 respectively. The fact that the standard deviation of the outcome measure in study A was two times greater than that in study B implies that the populations in the two studies are different and illustrates the danger of comparing effect sizes, or any measures of treatment effects, across studies. The population of study A may be more heterogeneous and thus "different" than the population in study B.

Comparing effect sizes on multiple outcome measures within a single study may be problematic for a similar reason [24]. For example, in a study that reports an effect size of 0.6 for range-of-motion and 0.4 for gait velocity, a potentially incorrect interpretation of these findings would be that the intervention had a greater effect on range-of-motion. This may not actually be the case if the outcome measures are not normally distributed. For similar reasons, classifying effect sizes as small (≤ 0.2), medium (>0.2 and ≤ 0.8), or large (>0.8), as some suggest, may be questionable.

Effect size represents one way to characterize the benefits of an intervention. In order for this measure of treatment effect to be valid, the assumptions detailed above (i.e., normal distribution of the outcome variable, good estimate of the standard deviation of the population) must be met. From a statistical perspective, the assumptions for conducting an ANOVA should also be met. Even if these assumptions are met, the clinical relevance and interpretability of effect size is questionable because it is a unitless measure. Despite these limitations, effect size is often reported in the medical research literature and is commonly used in meta-analyses [24].

Regression Coefficients (unstandardized and standardized): Regression models can be used to characterize the effect of a treatment when both the treatment and outcome variables are continuously distributed (e.g., number of PT treatments and ROM values) or when the treatment variable is dichotomous (e.g., whether individual received PT treatment) and the outcome variable is continuous [12,28]. For example, in a simple bivariate model of the relationship between the treatment (x_1) and the outcome (y) [29]:

$y = \beta_0 + \beta_1 x_1 + \varepsilon$ (where β_0 = intercept, β_1 = slope; ε = error term)

The treatment effect is represented by the regression coefficient, or β_1 . The regression coefficient indicates how much the outcome measure would increase or decrease (e.g., change in ROM) for an increment increase in the treatment variable (e.g., a one unit increase in the number of PT treatments) [29]. In the situation where the treatment variable is dichotomous (i.e., PT treatment or no PT treatment), the regression coefficient indicates how much the outcome measure would increase or decrease if the treatment was received. For example, if b_1 was 8.7, then patients receiving PT treatment would increase their ROM 8.7 degrees compared to patients not receiving PT treatments. Although a bivariate model was used for illustrative purposes, the interpretation of the regression coefficient in a multivariable analysis is similar.

Regression coefficients are standardized in order to make meaningful comparisons between different models, or between independent or explanatory variables in the same model [30]. Standardized regression coefficients have been used to represent treatment effects just as unstandardized regression coefficients do [30]. The Pearson correlation coefficient (r) is the standardized regression coefficient or standardized slope for the bivariate regression model. In the case of a PT intervention study, r would be a measure of the association between the treatment and the outcome. The value of r lies between -1 and $+1$ and when squared indicates the proportion of the variation in the outcome variable that is "explained" or "accounted for" by the treatment variable.

Regression coefficients in multivariable models can also be standardized by multiplying the regression coefficient (i.e., b_k) by the standard deviation of the independent or explanatory variable (i.e., x_k) and dividing by the standard deviation of y . These standardized coefficients are sometimes referred to as beta weights or path coefficients and are essentially a measure of the relative weights attached to the various independent variables in contributing to the mean of the dependent variable. For example, consider the hypothetical data below:

Variable	Mean	Standard Deviation
Y	22.16	7.08
X ₁	43.15	21.23
X ₂	57.80	28.76

If $b_1 = 0.114$, the estimated standardized regression coefficient would be equal to $[(0.114) \cdot (21.23/7.08)]$ or 0.34. Likewise, if $b_2 = 0.098$, the estimated standardized regression coefficient would be equal to $[(0.098) \cdot (28.76/7.08)]$ or 0.39. So, the estimated change in the mean of Y for a one standard deviation increase in X₁, controlling for X₂, is approximately the same magnitude as for a one standard deviation increase in X₂, controlling for X₁.

Advantages/Disadvantages of Regression Coefficients: Using a regression coefficient to characterize the effects of a treatment is particularly useful when the outcome variable is continuous. Multiple regression analysis, in particular, is commonly used in PT research when there is a need to control for other variables (e.g., age, pre-injury functional status) before estimating the effects of an intervention. As with any statistical test used to estimate a treatment effect, the assumptions of the regression analysis must be met for the treatment effect measure to be valid.

The use of standardized regression coefficients as a measure of treatment effect is controversial. Although standardized regression coefficients are commonly reported in the health care research literature, some argue that standardizing a regression coefficient distorts the measure of effect because the standardization is based on the distribution (i.e., the standard deviations) of the independent and dependent variables. Comparisons of standardi-

-zed regression coefficients across studies, therefore, would not be valid if the standard deviations for the variables of interest varied across studies. Likewise, the comparison of standardized regression coefficients within the same study (i.e., comparing the effects of two independent variables) would not be valid because the standard deviations of two independent variables within the same study would likely be different.

Number Needed to Treat for Continuous Outcomes: A NNT can be calculated for continuous outcomes by dichotomizing the outcome measure in a clinically meaningful way. One approach would be to dichotomize based on whether subjects had a minimally important change in the outcome measure, referred to by some as a minimally important difference (MID) [16]. For example, a PT may consider a 2cm decrease in pain intensity (measured on a 10cm visual analog scale) to be the minimally important difference or change in the outcome. Patients with a decrease in pain of 2cm or more would be classified as having a positive outcome. Those with less than a 2cm decrease in pain would be classified as having a negative outcome. Another approach would be to dichotomize based on whether subjects reached a relevant threshold value for the outcome measure [31]. Again, using the outcome measure of pain intensity, a PT may consider a pain intensity rating of 1cm or less to be a clinically relevant threshold. In other words, patients who rated their pain intensity as 1cm or less on a 10cm visual analog scale would be classified as having a positive outcome. Those patients with a higher pain intensity rating would be classified as having a negative outcome. Regardless of the method of dichotomization, once a continuous outcome is dichotomized, the NNT can be calculated.

Advantages/Disadvantages of NNT for Continuous Outcomes: In addition to the advantages and disadvantages of NNT outlined in the dichotomous outcomes section, there are a number of issues related to dichotomizing a continuous outcome. Perhaps the most important issue is determining the MID or clinically relevant threshold for the outcome variable [16]. There may be little information to

help guide this determination. Dichotomizing a continuous outcome variable may also decrease the precision of the results [16]. Finally, dichotomizing a continuous outcome does not take advantage of the full spectrum of information included in the measure. Important relationships, such as a dose-response relationship may be missed.

ACKNOWLEDGEMENTS

This work was supported in part by the American Physical Therapy Association through consultation with TJ Mielenz and JK Freburger. The funding source did not play a role in conducting the review and did not review or approve the manuscript. It was supported in part by the Foundation for Physical Therapy *New Investigator Fellowship Training Initiative*, the American College of

Rheumatology *Research and Education Foundation Health Professional New Investigator Award*, a National Arthritis Foundation *New Investigator Award*. It was supported in part by Grant 1 R49 CE002096 01 from the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention to the Center for Injury Epidemiology and Prevention at Columbia University. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention. This research was supported in part by Contract Number 11P2PI000797-0 from PCORI. All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of PCORI, its Board of Directors or Methodology Committee. The authors would also like to acknowledge Kimberly J. Alvarez, MPH, Patrick R. Hlavacek, MPH and Laura L. Durbin, MPH for their assistance with manuscript preparation while with the Columbia University Mailman School of Public Health.

APPENDIX A: Definitions of Epidemiology Terminology

Term	Definition	Additional Information
Ratio	A comparison of two measures in which one distinct quantity is divided by another distinct quantity [11].	Proportion, odds, and rate are all types of ratios [11].
Proportion	A ratio where the numerator and denominator are measured in the same units and where the value of the numerator is less than or equal to the value of the denominator [11].	Proportions are dimensionless and range from 0 to 132. A proportion can be expressed in terms of a decimal, a fraction, or a percentage.
Risk	The probability that an outcome will occur [11]. The numerator is the number of patients developing the outcome during a period of time and the denominator is the number of patients followed for the time period [20].	Risk is a proportion. Although the term risk implies a negative outcome, the term can also be used to describe the probability that a positive outcome will occur [11]. For example, if 50 patients out of 100 had a positive outcome following a physical therapy intervention, the risk of the outcome would be 50/100 or 0.5 for the specified period of time.
Odds	A ratio of the probability of the outcome occurring to the probability of the outcome not occurring [33].	For example, an odds is the probability of a patient walking independently over the probability of the patient not walking independently. A risk of 50/100 or 0.5 is equal to an odds of 50:50 or 1.0.
Rate	A rate is one way to measure the frequency of an outcome in a defined population [34]. The numerator is the number of persons with the outcome in a given time period and the denominator is the number of persons "at risk " for the outcome in a given period of time [34].	"At risk" means those persons that do not already have the outcome, but are susceptible to getting it (i.e., being counted in the numerator). The denominator is usually expressed in person-time.
Person-Time	A measurement commonly used in the denominator of incidence rates that combines persons and time [7]. Person-time is the time a person(s) in the study population was observed and was free of the outcome [7].	If a study is 5 years long and a patient has the outcome or is lost to follow-up after 1 year then he has contributed 1 person-year, if he had the outcome after 4 years he would have contributed 4 person-years. Person-years is commonly used in the epidemiological literature, but other units of time can be used to calculate person-time [7].
Prevalence	The proportion of a population at risk that has the outcome at a given point in time or during a period of time [11]. Prevalence is the probability that a population member will have the outcome in question.	It is generally calculated as the total number of persons in the population who have the outcome divided by the population size.
Point prevalence	The number of persons in the population who have the outcome in a given period of time divided by the number of persons in the population at risk for the outcome at a point in time [11].	For example, the point prevalence for low back pain (defined as pain on the day of the interview) is 15-30%. This means that on the day of the interview 15-30% of the population had low back pain.
Period prevalence	The number of persons in a population who have the outcome at some point in a given period of time divided by the number of persons in a population at risk for the outcome during that period of time [11].	For example, the lifetime prevalence of low back pain, which is a type of period prevalence, is 50-80%. This means that approximately 50-80% of the population can remember having low back pain at some point in their lifetime.
Incidence	The number of persons in a defined population who develop the outcome during an observed period of time [11].	Incidence quantifies the risk of the outcome in a population and can be measured either as a proportion or a rate.
Cumulative Incidence (CI) or Incidence Proportion	CI is the number of persons who develop the outcome during the period of time divided by the number of persons in a population at risk for the outcome at the beginning of the observation period.	A measure of the risk of developing the outcome over time [11]. Cumulative incidence is a direct estimate of risk.
Incidence Density (ID) or Incidence Rate	ID is the number of persons who develop the outcome during a period of time divided by the sum of person-time observation that individuals were free of the outcome. The sum of person-time observation is the time that each person spent in the population and was free of the outcome.	A measure of the rate at which persons in a population at risk develop the outcome [11]. The measure is usually expressed as the number of persons per person-time [7]. ID is not a direct estimate of risk, but rather an average rate of outcome occurrence. For example, five people are followed for the incidence rate of hip fracture. Three people fracture their hip during the follow-up period. Their person-time observation ends when they have the outcome (i.e., hip fracture). The fourth person's person-time observation ends when she is lost to follow-up. The fifth person does not fracture his hip and is followed for the complete observation period.

Conflicts of interest: None

REFERENCES

- [1]. Streiner DL, Norman GR. *PDQ Epidemiology*. 3rd ed. Shelton: People's Medical Publishing House; 2009.
- [2]. Sterne J A, Davey Smith G. Sifting the evidence—what's wrong with significance tests? *BMJ*. Jan 27 2001;322(7280):226-231.
- [3]. Rosenthal R, Rosnow RL, Rubin DB. *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, United Kingdom: Cambridge University Press; 2000.
- [4]. Walter SD. Choice of effect measure for epidemiological data. *J Clin Epidemiol*. Sep 2000;53(9):931-939.
- [5]. Guyatt G H, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. *BMJ*. Feb 28 1998;316(7132):690-693.
- [6]. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. *J Clin Epidemiol*. Aug 1994;47(8):881-889.
- [7]. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
- [8]. Rothman KJ GS. *Modern Epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven; 1998.
- [9]. Fleiss JL, Levin B, Cho Paik M. *Statistical Methods for Rates and Proportions*. 3rd ed. New York: John Wiley & Sons; 2003.
- [10]. Pagano M, Gauvreau K. *Principles of Biostatistics*. 2nd ed. Pacific Grove: Duxbury; 2000.
- [11]. Last JM. *A Dictionary of Epidemiology*. 4th ed. New York: Oxford University Press, Inc.; 2001.
- [12]. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev*. 1987;9:1-30.
- [13]. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*. Nov 18 1998;280(19):1690-1691.
- [14]. Sackett DL SS, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based Medicine: How to Practice and Teach EBM*. 2nd ed. London: Churchill Livingstone; 2000.
- [15]. Le CT. *Applied Survival Analysis*. New York: John Wiley & Sons; 1997.
- [16]. Walter SD, Irwig L. Estimating the number needed to treat (NNT) index when the data are subject to error. *Stat Med*. Mar 30 2001;20(6):893-906.
- [17]. Chatellier G, Zapletal E, Lemaitre D, Menard J, Degoulet P. The number needed to treat: a clinically useful nomogram in its proper context. *BMJ*. Feb 17 1996;312(7028):426-429.
- [18]. Davies HT, Crombie IK, Tavakoli M. When can odds ratios mislead? *BMJ*. Mar 28 1998;316(7136):989-991.
- [19]. Deeks J. When can odds ratios mislead? Odds ratios should be used only in case-control studies and logistic regression analyses. *BMJ*. Oct 24 1998;317(7166):1155-1156; author reply 1156-1157.
- [20]. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology: The Essentials*. 3rd ed. Baltimore: Williams & Wilkins; 1996.
- [21]. Jaeschke R, Guyatt G, Shannon H, Walter S, Cook D, Heddle N. Basic statistics for clinicians: 3. Assessing the effects of treatment: measures of association. *CMAJ*. Feb 1 1995;152(3):351-357.
- [22]. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses—sometimes informative, usually misleading. *BMJ*. Jun 5 1999;318(7197):1548-1551.
- [23]. Coe R. What is 'effect size'? 2000; <http://www.cem.dur.ac.uk/ebeuk/research/effsize/ESguide.htm>, July 1, 2011.
- [24]. Becker LA. Effect Size. 2000; <http://www.web.uccs.edu/becker/psy590.es.htm>. Accessed July 1, 2011.
- [25]. Rosnow RL, Rosenthal R, Rubin D. B. Contrasts and correlations in effect-size estimation. *Psychol Sci*. Nov 2000;11(6):446-453.
- [26]. Rosnow RL, R. R. Computing contrasts, effect sizes, and counterfactuals on other people's published data: General procedures for research consumers. *Psychological Methods*. 1996;1:331-340.
- [27]. Rosenthal R, Rubin DB. Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*. 1986;99:400-406.
- [28]. Garrett J. *Bios 141: Quantitative methods for healthcare professionals I*. Chapel Hill: University of North Carolina at Chapel Hill; 1999.
- [29]. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol*. May 1987;125(5):761-768.
- [30]. Greenland S, Schlesselman JJ, Criqui MH. The fallacy of employing standardized regression coefficients and correlations as measures of effect. *Am J Epidemiol*. Feb 1986;123(2):203-208.
- [31]. Schulzer M, Mancini GB. 'Unqualified success' and 'unmitigated failure': number-needed-to-treat-related concepts for assessing treatment efficacy in the presence of treatment-induced adverse events. *Int J Epidemiol*. Aug 1996;25(4):704-712.
- [32]. Gordis L. *Epidemiology*. 3rd ed. Philadelphia: W.B. Saunders Company; 2004.
- [33]. Merrill RM. *Introduction to Epidemiology*. 5th ed. Sudbury: Jones and Bartlett Publishers, LLC; 2010.
- [34]. Last JM. *A Dictionary of Epidemiology*. 3rd ed. New York: Oxford University Press, Inc.; 1995.

How to cite this article: Thelma J. Mielenz, Janet K. Freburger, Timothy S. Carey. TREATMENT EFFECTS 101. *Int J Physiother Res* 2015; 3(5): 1198-1207. **DOI:** 10.16965/ijpr.2015.173