January 2019

# Reporting Number Needed To Treat In Clinical Trials Published In Physical Therapy Specific Literature 1989 - 2018

Susan Ann Talley
*Wayne State University*, s.a.talley.pt@gmail.com

# REPORTING NUMBER NEEDED TO TREAT IN CLINICAL TRIALS PUBLISHED IN PHYSICAL THERAPY SPECIFIC LITERATURE 1989 – 2018

by

**SUSAN ANN TALLEY**

**DISSERTATION**

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

In partial fulfillment of the requirements

For the degree of

**DOCTOR OF PHILOSOPHY**

2019

MAJOR: EDUCATION EVALUATION AND

RESEARCH

Approved by:

________________________________________
Advisor Date

________________________________________

________________________________________

________________________________________

________________________________________

# DEDICATION

*I dedicate this dissertation to my husband, Joe, and our children, Jason (my favorite son) and Katy (my favorite daughter). Your unending love, support and belief in me made what seemed to be an impossible goal, doable. It's great having you as my personal cheerleading team.*

*I also dedicate this dissertation to my physical therapist students, former students and colleagues. Translating evidence into practice is no easy task but striving to do so can only improve the quality of life of our patients/clients*

## ACKNOWLEDGMENTS

I owe a debt of gratitude to many individuals who have inspired, encouraged and supported me along this path and allowed me to be successful.

Thank you, Dr. Donald Marcotte, one of my first EER instructors. On the first day he asked us to write down why we were taking that course and I wrote “for fun.” Dr. Marcotte’s masterful teaching and passion, indeed made it fun and propelled me forward.

Thank you, Dr. Gail Fahoome, an original member of my dissertation committee and mentor. You demonstrated what it meant to “do what you love and love what you do.” We had many discussions about program evaluation, research and the state of the academy.

Thank you to numerous friends, family and co-workers who would not let me derail myself in the pursuit of this degree... my husband Joe, children, Jason and Katy, for their undying love and belief in me; John and Margaret, Mark and Judy and Donna who all provided space “away” where I could go on writing retreats when I just could not get it done otherwise; my friends at church, Judy and Fred, Pat and Mike, Sharon to name a few, for their multitude of prayers; and my work colleagues and friends, especially Nancy, Allon, Stephanie, Beverly, Frank and Donna. This is only the tip of the iceberg. I couldn’t have done it without you!

Thank you to my committee members (Drs. Irwin Jopps, Monte Piliawsky, Monica Tracey and Sujay Galen). You were each able to give me just what I needed at a given point in time. Thank you to my research assistants for assisting with the coding of variables making a daunting task manageable.

Thank you to my mentor and advisor, Dr. Shlomo Sawilowsky, who likely doesn’t remember that I took one of the first courses he taught at Wayne State University in the 1980’s, long before I ever thought about pursuing a degree in EER. Thank you for challenging me to think and question and consider the context in understanding... whether in interpreting research results or the evolution of our profession and the figures who have played key roles in the process. I value your contributions to my learning and not giving up on me. Thank you.

**TABLE OF CONTENTS**

**TABLE OF CONTENTS continued**

**TABLE OF CONTENTS continued**

## LIST OF TABLES

## LIST OF FIGURES

# INTRODUCTION

## Clinical Decision Making in Physical Therapy

Physical therapists are required to make many clinical decisions about the best plan of care or intervention to use when providing physical therapy to patients or clients. Evidence-based practice is the foundation for making decisions that reflect best practices in the care of patients with impairments and activity limitations. D. L. Sackett, Rosenberg, Gray, Haynes, and Richardson (1996) defined evidence-based practice as "integrating individual clinical expertise with the best available external clinical evidence from systematic research" (p. 71). Clinically based research informs the decision to choose an intervention with the intent of improving a patient/client's ability to perform activities and to participate, as desired, in life activities and roles.

## Null Hypothesis Statistical Testing as the Basis for Clinical Decisions

The effectiveness of physical therapy interventions is commonly evaluated by comparing means of two or more groups, ideally through randomized control trials. At the least complex level, in a comparative study the mean of a treatment group is compared with a control group or means of two treatment groups could be compared. By controlling confounding variables, the only difference expected between groups is the intervention. This permits judgements to be made about the efficacy of the treatment relative to the specific measured outcomes.

Thompson (1996) opined effect sizes are interpretable when the null hypothesis is retained. However, Sawilowsky and Yoon (2002) and Sawilowsky (2003) noted it is futile to discuss effectiveness of a given treatment if the null-hypothesis is found to be non-significant (see also Sawilowsky (2007); Sawilowsky, Sawilowsky, & Grissom, 2011). Similarly, (Cohen, 1988) indicated:

> Whatever the manner of representation of a phenomenon … the null hypothesis always means the effect size is zero… [but] when the null hypothesis is false, it is false to some specific degree, i.e., the effect size (ES) is some specific nonzero in the population (p. 10).

It is difficult for a clinician to translate the results of group comparison studies to the clinical decision-making process required to select an intervention for an individual patient/client. The problem is how to determine which intervention will be the most effective for a patient/client. In too many instances individuals attempt to interpret the results inappropriately, such as estimating the value of a treatment based on the magnitude of the p value or on the difference between raw scores (Sawilowsky, 2007; Sawilowsky, 2009; Sawilowsky et al., 2011) For example, a calculated *p* value that is very small might be inappropriately interpreted as meaning that the effect of the treatment is very strong when there is little clinical difference. Similarly, the mean raw score difference is simplistic in nature and is not robust as the finite break-down point is 1/n.

**Magnitude of Differences or Associations as the Basis for Clinical Decisions**

Alternative methods for evaluating the relative effectiveness of potential interventions include calculating effect size utilizing one of over 40 effect size measures such as Cohen's *d*. However, the underlying assumptions such as normality, homogeneity of variance, outliers and heteroscedasticity is problematic (Knapp & Sawilowsky, 2001; Sawilowsky (2018); Sawilowsky & Blair, 1992; Vacha-Haase & Thompson, 2004). The number needed to treat (NNT) is an alternative effect size method to interpret the effectiveness of an intervention compared to other interventions when measuring dichotomous outcome variables. NNT reflects the effect size of the treatment and indicates the number of patients who would need to be treated to ensure one successful outcome (Portney & Watkins, 2009).

NNT was first introduced by Laupacis, Sackett, & Roberts in 1988. Although multiple articles on NNT were published in the medical literature during the 1990's (Cook & Sackett, 1995; McQuay & Moore, 1997; D. L. Sackett et al., 1996), the first article about NNT in the physical therapy literature did not occur until 2000 in the journal *Physical Therapy*, the flagship journal of the American Physical Therapy Association (Dalton & Keating). Only one article utilizing NNT was found in *Physical Therapy* by (Dalton & Keating, 2000) searching Medline back to 1991. Subsequent articles encouraging the use of NNT in physical therapy intervention studies were published in 2004 (Weeks & Noteboom), 2006 (Hilton, Reid, & Paratz) and, most recently, 2016 (Hancock & Kent, 2016). Although using and reporting NNT to assist in interpreting the clinical importance of the results of an intervention study will help translate research into clinical practice, no studies were found that have examined the use of NNT in the physical therapy literature during the three decades since it was introduced in 1988. The effectiveness of publishing "encouraging articles" to increase use of NNT is unknown.

## Purpose of the Study

The purpose of the study is to examine the methods of reporting research results of intervention studies in the physical therapy literature. Specifically, the purpose of this study is to explore the reporting of null hypothesis statistical tests, effect size and number needed to treat in physical therapy intervention studies over time in the physical therapy literature. Bibliometric studies to identify core journals in physical therapy, primarily through citation analyses and content analyses, have been conducted to describe content at a single point in time and longitudinally. There are no bibliometric analyses that explored trends in reporting results to inform treatment selection in evidence-based physical therapist practice.

**Assumptions**

It is assumed investigators intended to publish results of intervention studies in a manner which facilitates the use of the results to inform clinical decision making in physical therapist practice including the use of statistical methods. Consequently, it is assumed that investigators are aware of NNT, or at least have had increasing awareness over the past 3 decades, so that reporting (or not reporting) NNT is an active choice. It is further assumed that investigators were free to choose statistical methods for publication in the physical therapist literature without publication bias. Finally, it is assumed that interpretation and categorization of bibliometric variables are accurate and appropriate as the variables are clearly defined.

**Limitations**

A limitation is the sampling strategy. Articles will be limited to intervention studies published in select journals of the physical therapy professional association in the United States, the American Physical Therapy Association (*Journal of Neurologic Physical Therapy*, *Journal of Orthopedic and Sports Physical Therapy, Physical Therapy*) limiting generalizability of the findings. A different sampling strategy may result in different results. Bibliometric content analysis is a historical descriptive study of published physical therapy literature. It is beyond the scope of this study to assess the quality of the research methodology for each article. The results reflect the published physical therapy literature in the sample which is not the same as the broader state of current physical therapy research. It is possible that articles may have been submitted for review, accepted for publication and/or published in different years due to the lag time for publication which may influence the outcomes.

**Operational Definitions**

*Alpha*: nominal value determined a priori to indicate the acceptable maximum probability of a Type I Error (Portney & Watkins, 2009).

*Bracketed Interval:* commonly referred to as a confidence interval. The determination of a range of values for an outcome, for which the value of a population parameter is located between the upper and lower limits at a given probability.

*Clinical trial*: “A research study in which one or more human participants are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes” (National Institutes of Health, 2014).

*Effect size*: “A statistical expression of the magnitude of the difference between two treatments or the magnitude of a relationship between two variables, based on proportional relationship of the difference to the variance” (Portney & Watkins, 2009, p. 867); “the magnitude of a treatment (or naturally occurring) effect when the null hypothesis is false” (Sawilowsky, p. 1).

*Number Needed to Treat* (NNT): “The number of patients that need to be treated to prevent one adverse outcome or achieve one successful outcome” (Portney & Watkins, 2009, p. 872).

*p value*: calculated value in inferential statistics to evaluate a null hypothesis; the probability of the available (or even less likely) data, given that the null hypothesis is true.

*Statistical non-significance*: when the calculated p value is greater than the predetermined nominal alpha the null hypothesis is not rejected, indicating no statistically significant effect.

*Statistical significance*: when the calculated p value is less than or equal to the predetermined nominal alpha the hull hypothesis is rejected, indicating a statistically significant effect. "The term indicating that the results of an analysis are unlikely to be the result of chance at a specified probability level" (Portney & Watkins, 2009, p. 877).

*Type I error* (α): the probability of rejecting a null hypothesis when the null hypothesis is true.

*Type II error* (β): the probability of failing to reject a null hypothesis when the null hypothesis is false.

Table 1

NHST Decision Making and Type I and Type II Error

| | | Truth | |
|---|---|---|---|
| | | $H_0$ is True | $H_0$ is False |
| DECISION | Reject $H_0$ | Type I Error *α* | Correct |
| | Fail to Reject $H_0$ | Correct | Type II Error *β* |

*Note*. Adapted from Portney, L.G. & Watkins, M.P. (2009). Statistical inference. In L.G. Portney & M.P. Watkins (Eds). *Foundations of clinical research: Applications to practice* (2nd ed., p. 418). Upper Saddle River, ND: Pearson Education, Inc.

# CHAPTER 2 REVIEW OF LITERATURE

## Limitations of Null Hypothesis Statistical Testing

Null hypothesis significance testing (NHST) is a statistical approach frequently used in quantitative research in the social, behavioral and health sciences to help answer a research hypothesis. NHST is one of several approaches to interpret the outcome of a clinical trial investigating the comparative effectiveness of an intervention, compared to a control or another intervention. However, there is longstanding controversy about the appropriate use and interpretation of NHST. Thompson (1998) identified five "methodological errors" that occur commonly in educational research including "the incorrect interpretation of *statistical significance* and the related failure to report *effect sizes* present in all quantitative analyses" (Thompson, 1998, p. 6). Thompson further stated "…even today some researchers still do not understand what their statistical significance tests do and do not do" (Thompson, 1998, p. 39). Campo and Lichtman (2008) wrote a position paper published in *Physical Therapy* on the limitations of NHST in interpreting physical therapy research, identifying issues and suggesting alternative measures to consider. Cohen (1994) was widely credited with having written the seminal article at the base of the NHST controversy, which at the time had already spanned four decades, and suggested NHST be replaced with other methods, such as examining the data graphically as in Exploratory Data Analysis (Cox, 2017; John W Tukey, 1977) and the reporting of effect sizes by using bracketed (confidence) intervals. Others defended use of NHST when used appropriately (Compton & Sawilowsky, 2003; Cortina & Landis, 2011; Hagen, 1997; Knapp & Sawilowsky, 2001). However, Cohen (1994) advised "don't look for a magical alternative to NHST, some other objective mechanical ritual to replace it. It doesn't exist" (p.1001). The NHST discussion continues for more than 60 years without real change in the arguments and counterarguments. A review of several of these arguments are particularly relevant to utilizing outcomes from clinical trials to inform clinical decision making in physical therapist practice.

A common assertion fueling the question regarding the utility of NHST is the null is always false (Cohen, 1994; Hays, 1981; Meehl, 1978; Thompson, 1993; Thompson, 1996; Tukey, 1991), and therefore there is no justification for NHST. Thompson (1993) stated "Virtually all null hypotheses will be rejected at some sample size" (p.362). Gross (2015) echoed this argument 20 years later, stating that NHST:

> …compels us to engage in sort of Kabuki theater, going through the motions of what Rozeboom (1960) has called our "tribal ritual" of rejecting $H_0$, when we know that with a large enough sample, a point null hypothesis will almost surely be rejected. (p.777)

However, Knapp and Sawilowsky (2001) stated there are clearly circumstances in which the null hypothesis is indeed true, for example when testing a null hypothesis for an experiment where there is a dichotomous outcome. They further stated that a Monte Carlo simulation with two groups randomly selected with replacement from a given population with a Gaussian, or normal, distribution would result in the null hypothesis being rejected 5% of the time as predicted because the nominal alpha was set to .05. Sawilowsky and Blair (1992) demonstrated this in a Monte Carlo simulation testing the robustness of the *t* test for Type I errors using authentic social or behavioral data sets. They stated "thus, under the truth of the null hypothesis, the notion that there must be some large sample size that will reject a true null hypothesis, aside from committing a Type I error, is false" (Sawilowsky & Blair, 1992, p. 72). Knapp and Sawilowsky (2001) stated clinically trivial effects may become statistically significant if the sample size is sufficiently large.

Another alleged concern raised against NHST is that it does not lead to the cumulation of scientific knowledge or discoveries (Cohen, 1994; Thompson, 1996). Cohen (1994) stated NHST "has not only failed to support the advance of psychology as a science, but has seriously impeded it" (p. 997). It may be that the problem lies in depending on a NHST in isolation to make decisions about the clinical relevance and importance of an observed difference necessitating the need for additional analysis such as confidence intervals, effect size or number needed to treat, as well as interpretation by subject matter experts (Campo & Lichtman, 2008; Cortina & Dunlap, 1997; Cortina & Landis, 2011; Gross, 2015; Knapp & Sawilowsky, 2001). However, NHST contributes to the body of scientific knowledge, particularly with purposeful replication of experiments (Frick, 1996; Gross, 2015; Hagen, 1997; Robinson & Levin, 1997). Knapp and Sawilowsky (2001) advised statistical testing must be understood as separate, in the sense being just a tool, from scientific discovery.

Misinterpretation of the results of NHST, specifically the *p* value, is widespread (for example see Campo & Lichtman, 2008; Cohen, 1994; Falk & Greenbaum, 1995; Gross, 2015; Haller & Krauss, 2002; Hubbard & Lindsay, 2008; Nickerson, 2000; Sawilowsky, 2011). Ninety percent of participants in a study at six German universities (students and faculty/scientists) held at least one misconception of the meaning and interpretation of *p* values (Haller & Krauss, 2002). Examples of common misconceptions include (a) treating the *p* value as the probability that the null hypothesis is true (or false), (b) the complement of the *p* value ($1 - p$) is the probability that the study could be replicated with the same outcome, (c) statistical significance implies the results are also clinically important, (d) interpreting the magnitude of the *p* value as a measure of the magnitude of the treatment effect, and (e) a *p* value equal to or less than a nominal alpha of .05 provides conclusive evidence against the null hypothesis or in support of the alternative hypothesis, among others. Although there is agreement that misinterpretation occurs, there is not agreement on the appropriate response to those misinterpretations ranging from abandoning NHST (Cohen, 1994, 1995) to asserting that NHST is a necessary first step to interpreting research results (Knapp & Sawilowsky, 2001). Falk and Greenbaum (1995, p. 93) stated "To be fair, the fact that people misinterpret significance tests is not the tests' fault and is no reason to discard them altogether. Misconceptions may, after all, be clarified and the right meaning restored" (p. 93). Gross (2015) stated misinterpretations of the *p* value are "merely the most recognizable trappings of an overall framework that overemphasizes minor details. It is not so much their inclusion in analyses that is objectionable as much as their outsized role" (p. 777).

**Effect Size**

An effect size is an estimate of the magnitude and direction of a relationship (mean differences or associations) (Campo, Eckardt, Findley, Cardinale, & Shiyko, 2017; Sawilowsky). It was recommended that effect size be used in addition to NHST to interpret and make decisions about the clinical relevance and importance of an observed difference in addition to including interpretation of results by subject matter experts (Campo & Lichtman, 2008; Cortina & Dunlap, 1997; Cortina & Landis, 2011; Gross, 2015; Knapp & Sawilowsky, 2001; Sawilowsky). Campo et al. (2017) stated that "Effect sizes offer an important way to move beyond the limitations of significance testing, because they offer estimates of the magnitude of treatment effects, between-group differences, and associations between variables" (p. 67). Effect sizes may be easier for most people, including physical therapists, to understand than NHST, thereby helping the clinician decide if statistically significant results are also clinically important outcomes (Campo et al., 2017; Tracey, 2000; Vacha-Haase & Thompson, 2004). Sink and Stroh (2006) stated "if researchers fail to report [effect sizes], and only include the research findings' derived significance levels, key information is missing that assists in understanding the practical value of the results" (p. 402). Although it was advocated that effect size be reported regardless of the result of NHST (Carver, 1978, 1993; Rosnow & Rosenthal, 1989; Thompson, 1996, 1998; Vacha-Haase & Thompson, 2004; Wilkinson, 1999), it is illogical to compute an effect size when the null hypothesis is not rejected. "Trivials are effect sizes associated with statistically non-significant results" and are problematic (Sawilowsky & Yoon, 2002, p. 143). Sawilowsky et al. (2011) stated:

> Under the truth of the null hypothesis observed results are not statistically different from zero, and thus the magnitude of the observed result is meaningless. Hence, effect sizes are only meaningfully reported in conjunction with a statistically significant hypothesis test. (p. 1413)

Sawilowsky and Yoon (2002) demonstrated the lack of meaning of effect sizes generated when the hull hypothesis was true through a Monte Carlo simulation which generated effect sizes ($|d|$ = .34) even though the effect size was modeled as zero ($n_1 = n_2 = 0$), Gaussian Distribution, Nominal α = 0.05, (Sawilowsky, 2003). When interpreting the outcomes of a study physical therapists should heed the admonition by Robinson and Levin (1997) to "First convince us that a finding is not due to chance, and only then, assess how impressive it is" (p. 23).

More than 40 indices of effect magnitude have been developed (Kirk, 1996; Vacha-Haase & Thompson, 2004) which may give differing results depending on the measure used (Knapp & Sawilowsky, 2001). Effect size measures have been grouped to facilitate understanding although the number of groupings has varied from two to four (Campo et al., 2017; Ferguson, 2009; Kline, 2013; Sawilowsky; Sink & Stroh, 2006; Vacha-Haase & Thompson, 2004). Two groups of effect size measures are identified consistently: (a) group mean differences/ standardized mean differences and (b) strength of association indices. However, effect size methodology is relatively young. Consequently, effect size must be interpreted with caution. Violations of assumptions (normality, homogeneity of variance, heteroscedasticity and outliers) can distort the derivation of effect size (Knapp & Sawilowsky, 2001).

There are multiple resources that explored the calculation of effect size in depth (for example, see Campo et al., 2017; Ferguson, 2009; Sink & Stroh, 2006; Vacha-Haase & Thompson, 2004). Commonly used effect size measures to assess mean differences include Cohen's *d* (Cohen, 1988), Hedge's g (Λ) (Hedges, 1982), Glass' Δ (Campo et al., 2017; Sawilowsky, 2018). Other approaches to measure effect size include the point serial $r_{PB}$, $r^2$, partial $\eta^2$, $\eta^2$, odds ratio, and number needed to treat (Campo et al., 2017; Sawilowsky et al., 2011). Each effect size measure has strengths and weakness in both the denominator and numerator, which in actuality are generally shared mathematically, i.e. one can be translated from one to another (Sawilowsky et al., 2011). For example, Glass' Δ uses the standard deviation of the control group for the denominator instead of the pooled standard deviation, with the intent to compensate for differences in variability between the control and intervention groups. Campo et al. (2017) advocated that physical therapist education curricula include a wide variety of approaches to measuring effect size including those for mean difference, proportions, ANOVA and regression and correlation.

Cohen (1988) created guidelines, or rules of thumb to help interpret the meaning of Cohen's *d*, recognizing that these are guidelines that require the context to be considered when attempting to discern the practical or clinical importance of an effect size (Table 2)

Table 2

Effect Size "Rules of Thumb"

| Magnitude | Description | Source |
|---|---|---|
| 0.01 | Very Small | Sawilowsky (2009) |
| 0.2 | Small | Cohen (1988) |
| 0.5 | Medium | Cohen (1988) |
| 0.8 | Large | Cohen (1988) |
| 1.2 | Very Large | Sawilowsky (2009) |
| 2.0 | Huge | Sawilowsky (2009) |

Osborne (2008) reflected on those guidelines and later stated "It is unclear whether these [ES = .20, .50 or .80] accurately reflect effect sizes observed in our (or any other) field…I have yet to find published reports of average effect sizes reported in various fields" (p. 154).

The development of an encyclopedia of effect sizes in psychology and education was proposed (but not funded) by Sawilowsky (2009, p. 9) with widespread support from leaders in the field. This resource would have been useful for sample size estimation and power analysis. Sawilowsky (2009) recognized the thresholds suggested by Cohen (1988) were useful, but they could not reflect the range of effect sizes that observed in the social sciences. In lieu of the encyclopedia, Sawilowsky (2009) developed new rules of thumb for effect size to expand those suggested by Cohen (Table 2). A limitation of these rules of thumb is they do not translate beyond Cohen's *d*. Sink and Stroh (2006, pp. 404-405) published a table of effect size magnitudes for multiple effect size measures to two decimal places, although only for small, medium and large effects based on the work of Green and Salkind (2004). Similarly, Sawilowsky (personal communications, 2017) transformed effect size magnitudes for a variety of effect size measures across the expanded rules of thumb to four decimal places (Table 3).

Table 3

Transformation of Effect Size Magnitudes for Multiple Effect Size Measures

| Description | *d* | *r* | $\eta^2$ | *f* | OR | NNT |
|---|---|---|---|---|---|---|
| Very Small | 0.01 | 0.005 | 0 | 0.005 | 1.0183 | 177.2364 |
| Small | 0.2 | 0.0995 | 0.0099 | 0.1 | 1.4373 | 8.8919 |
| Medium | 0.5 | 0.2425 | 0.0588 | 0.25 | 2.4766 | 3.6189 |
| Large | 0.8 | 0.3714 | 0.1379 | 0.4 | 4.2675 | 2.3343 |
| Very Large | 1.2 | 0.5145 | 0.2647 | 0.6 | 8.8159 | 1.656 |
| Huge | 2.0 | 0.7071 | 0.5 | 1 | 37.6224 | 1.1867 |

*Note. d* = Cohen's *d*; *r* = Pearson *r*; $\eta^2$ = eta squared; *f* = ANOVA *f* ratio; OR = odds ratio; NNT = number needed to treat. Adapted from personal communications by Sawilowsky, 2017.

**Number Needed to Treat**

One method of measuring effect size that may be particularly useful when translating the results of intervention research to clinical practice is the Number Needed to Treat (NNT) (Cook & Sackett, 1995; Dalton & Keating, 2000). Pinson and Gray (2003) defined NNT as "the number of patients who would need to be treated with a specified intervention in order to obtain one additional positive outcome that would not have occurred had the patient not received the comparison treatment" (p. 146) in a given period of time (Weeks & Noteboom, 2004). A positive outcome is interpreted as either the prevention of an adverse effect or the occurrence of a desirable effect (Herbert, 2000; Hilton et al., 2006; Laupacis, Sackett, & Roberts, 1988; Weeks & Noteboom, 2004). Laupacis et al. introduced NNT as a measure of effect size in 1988 and stated that it "expresses efficacy in a manner that incorporates both the baseline risk without therapy and the risk reduction with therapy" (p. 1730). Cook and Sackett (1995) add that "it is more meaningful to use the measure 'number needed to treat'…it has the advantage that it conveys both statistical and clinical significance" (p.452). Although NNT was initially used for studies of drug therapy, surgical procedures, immunization, diagnosis and risk factors (Laupacis et al., 1988) it is appropriate for intervention studies in other disciplines including physical therapy.

NNT is a measure of effect size for dichotomous (binary) outcome variables (Dalton & Keating, 2000; Herbert, 2000). It is the reciprocal of absolute risk reduction which is the difference in risk between the experimental and comparison groups (Cook & Sackett, 1995; Dalton & Keating, 2000; Laupacis et al., 1988; McQuay & Moore, 1997; Portney & Watkins, 2009):

$$NNT = \frac{1}{\frac{Pi}{Ti} - \frac{Pc}{Tc}}$$

NNT = number needed to treat
Pi = number of positive outcomes in the intervention group
Ti = total number of participants in the intervention group
$P_C$ = number of positive outcomes in the comparison group
$T_C$ = total number of participants in the comparison group

The NNT always refers to outcomes relative to a comparison group (McQuay & Moore, 1997). The magnitude of NNT is impacted by both the effectiveness of the intervention and the baseline risk in the comparison group (Hancock & Kent, 2016;

Laupacis et al., 1988). If the control/comparison group has better outcomes than the experimental group, the NNT will be negative and the intervention may be interpreted as potentially ineffective or harmful. NNT is typically rounded to the nearest whole number (Weeks & Noteboom, 2004). A NNT of 1 indicates that one patient would need to be treated to experience a positive outcome. Hence, the closer the NNT is to 1 the more likely a patient will benefit from the intervention compared to the alternate (control or comparison) intervention (McQuay & Moore, 1997; Portney & Watkins, 2009). Sawilowsky transformed effect size magnitudes for a variety of effect size measures, including NNT, based on effect size rules of thumb (Table 3) (Cohen, 1988; Sawilowsky, 2009). McQuay and Moore (1997) stated that an NNT = 2-3 would indicate an intervention that was effective although Weeks and Noteboom (2004) stated an NNT = 2-5 would indicate an effective intervention. Pinson and Gray (2003) found that psychiatric therapies had reported NNTs between 3 and 6 which is comparable to those reported for other medical therapies (D. Sackett, Straus, & Richardson, 2000). In a study of 9 high quality (PEDro score $\geq$ 6) randomized control trials selected from the *Physiotherapy Evidence Database* (PEDro) to demonstrate NNT, six studies had a reported NNT between 2 and 6, two had a reported NNT of 7-8 and one study had a reported NNT = 34 (Hilton et al., 2006). Sawilowsky stated that an NNT = 2 indicates a large to very large effect size and an NNT = 5-7 could be interpreted as a small to medium effect size. A large NNT does not necessarily rule out the use of an intervention particularly if the positive outcome is the prevention of a serious undesirable outcome such as death or permanent disability (Weeks & Noteboom, 2004).

It is not clear how often NNT is calculated and reported in the literature to help translate the results of intervention studies for clinical decision making, particularly in the physical therapy literature. Cook and Sackett (1995) reported that NNT was “becoming widely used as a tool for therapeutic decision making and bedside teaching” (p. 453). Weeks and Noteboom (2004) stated that NNT was gaining attention as a method of reporting the results of clinical trials with dichotomous outcome measures. However, multiple authors reported that NNT is not widely used (Dalton & Keating, 2000) or is underused (Hilton et al., 2006; Nuovo, Melnikow, & Chang, 2002; Pinson & Gray, 2003). The CONsolidated Standards of Reporting Trials (CONSORT) 2010 guidelines state “For both binary and survival time data, expressing the results also as the number needed to treat for benefit or harm can be helpful” (Moher et al., 2010, Section 17a). This CONSORT recommendation for improving the reporting of the results of randomized controlled trials, including reporting effect size, were first made in 1996 (Begg et al.). Nevertheless, Nuovo et al. (2002) reported that only 8 of 359 (2.2%) eligible papers in five major biomedical journals (*Annals of Internal Medicine*, *British Medical Journal*, *Journal of the American Medical Association*, the *New England Journal of Medicine*, *Lancet*) published in four discrete years at 3-year intervals (1989, 1992, 1995, 1998) reported the NNT. The first year, 1989, was selected because it was one year following the introduction of NNT by Laupacis et al. (1988). Nuovo et al. (2002) stated that guidelines, such as CONSORT, may not be sufficient motivation to increase the reporting of NNT and that “additional measures to ensure compliance with reporting standards may be needed” (p. 2814). In contrast, Naing, Aung, and Mak (2012) found that 7 of 8 (87.5%) systematic reviews accessed through PUBMED on a single date in 2012 included NNT in the results, perhaps an indication of increasing reporting of NNT in recent years as there is a 14-year difference between the end of the Nuovo et al. (2002) study and the Naing et al. (2012) study.

Over the past 18 years five articles have been published in the physical therapy specific literature encouraging the utilization of NNT when reporting outcomes of clinical trials. In 2000, *Physical Therapy (PTJ)*, the journal of the American Physical Therapy Association, published the first article which introduced and advocated for inclusion of NNT in the reporting of physical therapist intervention studies (Dalton & Keating). The authors stated that “NNT provides results in a way that is directly transferrable to the clinical setting” (Dalton & Keating, 2000, p. 1216). The purpose of the paper, although not explicitly stated, was to introduce the potential usefulness of NNT to report and interpret outcomes of clinical trials to readers of *PTJ*. Dalton and Keating (2000) conducted a MEDLINE search dating back to 1991 using the search terms “number needed to treat” OR “NNT” and identified 121 citations which reported NNT. Of those, only three involved physical therapist outcomes and only one of those three was published in a physical therapist specific journal (*PTJ*) in 1994 (Moreland & Thomson), six years after Laupacis et al. (1988) first introduced NNT.

Although the journal *Archives of Physical Medicine and Rehabilitation (Arch PM&R)* is not a physical therapy specific journal, it is considered a core physical therapy journal (Wakiji, 1997). In 2004 *Arch PM&R* published a special communication to describe the NNT statistic and how it can be used for the selection of clinical interventions (Weeks & Noteboom). The authors stated that “there is a growing application of the NNT in the rehabilitation literature, both in single studies and meta-analysis of multiple studies”

(Weeks & Noteboom, 2004, p. 1730) and cited a physical therapist intervention study for acute low back pain as an example (Fritz, Delitto, & Erhard, 2003). The next article published on the use of NNT specific to physical therapy was published in 2006 in *Physiotherapy,* the journal of the physiotherapy professional association in the United Kingdom (Hilton et al., 2006). The purpose of this study was to demonstrate how "the NNT can help clinicians to converse with patients to convey details about the likelihood of benefit with treatment and/or likelihood of risk, in order that a decision may be made with respect to therapy" (Hilton et al., 2006, p. 240) with the aim to "provide practical examples to demonstrate the utility of this statistic in the interpretation of findings in the physiotherapy literature"(Hilton et al., 2006, p. 241). As discussed earlier, Campo and Lichtman (2008) published an article on uses and limitations of NHST and recommended that physical therapist students and educators consider using other measures including NNT. Finally, in 2016 Hancock & Kent published a paper in the *Journal of Physiotherapy* (journal of the Australian Physiotherapy Association) with the intent to "describe the correct interpretation of commonly used methods of reporting dichotomous outcomes" (p.172) which included risks, odds, absolute and relative risk reduction and NNT. Despite the interest in the utilization of NNT in physical therapy research as evidenced by the aforementioned articles, there is a paucity of research on the frequency of reporting of NNT in the physical therapy literature nor is there research on trends of reporting NNT over time.

**Bibliometric Analysis of Physical Therapy Literature**

The definition of bibliometric analysis in not easily conveyed as there is no consistent, satisfactory definition. Pritchard (1969) referred to "bibliometrics", defined as "…the application of mathematics and statistical methods to books and other media of concern" (p.348). A more specific definition was offered by de Glas (1986) "generally speaking bibliometrics could be defined as the search for systematic patterns in comprehensive bodies of literature" (p. 40). Pritchard and Wittig (1981) expanded the definition: bibliometrics "includes all studies which use or discuss statistical analysis of data relating to printed communication, e.g. citation studies, abstracts journals studies, publication counts, some circulation studies…and studies of individual elements within papers" (p.3). There are numerous other histories of the field of bibliometrics available (see, for example Broadus, 1987a, 1987b; Hertzel, 1987; Hood & Wilson, 2001). In short, bibliometrics is a scientific method to explore the content and meaning of scientific literature. Bibliometric analysis has increased in prevalence in the literature of many fields (Borgman, 1989). Bibliometric citation analysis has been used to identify core journals or map the literature of a given field. Bibliometric citation analysis can also be used to quantify productivity of individuals or groups of investigators. It has been used to quantify characteristics or describe trends over time in scientific literature (Smith & Rivett, 2009).

The first bibliometric analysis specific to the physical therapy literature was a citation analysis of contributors to the journals *Physical Therapy* and *Physiotherapy Canada* over a two-year period published in *Physical Therapy (Dean & Davies, 1986)*. Since 1986, more than 19 bibliometric analyses of the physical therapy literature were published: five citation analyses focused on identifying core physical therapy journals

(Bohannon, 1999; Bohannon & Gibson, 1986; Bohannon & Tiberio, 1989; Fell, Burnham, Buchanan, Horchen, & Scherr, 2011; Wakiji, 1997), two analyses of core journals used the *Physiotherapy Evidence Database* (*PEDro*) as the source (Costa et al., 2010; C. Maher, Moseley, Sherrington, & Herbert, 2001), two content analyses focused on physical therapy clinical trials (Babu, Veluswamy, Rao, & Maiya, 2014; Hoderlein, Moseley, & Elkins, 2017), three content analyses focused on statistical methods used in physical therapy literature to inform physical therapist education curriculum (Bandy, 2003; Roush et al., 2015; Tilson, Marshall, Tam, & Fetters, 2016), one analysis utilized inclusion in Medline as an indicator of quality (Roberts, 1992), five bibliometric content analyses explored characteristics and trends over time (Coronado, Riddle, Wurtzel, & George, 2011; Miller, McKibbon, & Haynes, 2003; Paci, Cigna, Baccini, & Rinaldi, 2009; Simon, Coronado, Wurtzel, Riddle, & George, 2014; Wiles, Matricciani, Williams, & Olds, 2012), and one content analysis focused specifically on neurologic physical therapy (Fell et al., 2015). Three bibliometric analyses related to physical therapist practice were more broadly focused on rehabilitation (Bohannon & Roberts, 1991; Franchignoni & Munoz Lasa, 2011; Tesio, Gamba, Capelli, & Franchignoni, 1995), and one explored the research productivity of physical and occupational therapy faculty in Canada (MacDermid, Fung, & Law, 2015).

Bibliometric citation analysis was used to map physical therapy literature with the purpose of identifying core journals. Bohannon and Gibson (1986) stated “citation analyses, which assess the frequency with which specific journals are cited in the scientific periodical literature, were performed to assist librarians, authors, practitioners, and others in identifying important journals for acquisition, publication, and reference” (p.

540). Costa et al. (2010) used a different approach to identify core physical therapy journals. The *Physiotherapy Evidence Database* (*PEDro*) was used to identify the journals which published the most randomized control trials (RCTs). The key components of these studies can be found in Table 4.

TABLE 4

Bibliometric Analyses of Physical Therapy Literature: Core Journals

| Authors | Source | Time | Sampling | no. | PT Specific |
|---|---|---|---|---|---|
| Bohannon and Gibson (1986) | *PTJ*[a] | 4 years<br>1980-84 | All citations<br>Articles<br>Editorials<br>Commentary | 67 | *PTJ*[a]<br>*Physio*<br>*Physio Can* |
| Bohannon and Tiberio (1989) | *Austr Physio*<br>*PTJ*[a]<br>*Physio*<br>*Physio Can*<br>*Physio Pract* | 1 year<br>1987 | All citations<br>Articles<br>Editorials<br>Commentary | 64 | *PTJ*[a]<br>*Physio*<br>*Physio Can*<br>*Austr Physio*<br>*Physio Pract* |
| Wakiji (1997) | *Arch PM&R*<br>*PTJ*[a] | 3 years<br>1991-93 | All citations<br>Articles<br>Commentary<br>Letter<br>Lecture<br>Study Guide | 14 | *PTJ*[a] |

TABLE 4 continued

Bibliometric Analyses of Physical Therapy Literature: Core Journals

| Authors | Source | Time | Sampling | no. | PT Specific |
|---|---|---|---|---|---|
| Bohannon (1999) | *Austr Physio*<br>*PTJ*[a]<br>*PT Sci*<br>*Physio*<br>*Physio Can*<br>*Physio Res Int*<br>*Physio Theory* | 1 year<br>1997-98 | All citations<br>Articles<br>Commentary<br>Letter | 47 | *PTJ*[a]<br>*Physio*<br>*Physio Can*<br>*Austr Physio*<br>*JOSPT*[a] |
| C. Maher et al. (2001) | *PEDro* | June 2, 2000 | Journals that published 5+ RCTs | 75 | *Austr Physio*<br>*Physio Pract*<br>*PTJ*[a]<br>*Physio*<br>*Physio Can* |
| Costa et al. (2010) | *PEDro* | Sept. 7, 2009 | Journals that published 80+ RCTs | 22 | *PTJ*[a]<br>*Physio*<br>*JOSPT*[a] |
| Fell et al. (2011) | *Austr Physio*<br>*PTJ*[a]<br>*Physio*<br>*Physio Can* | 3 years<br>2005-07 | All Articles | 16 | *PTJ*[a]<br>*Physio*<br>*Physio Can*<br>*Austr Physio* |

*Note*: *Arch PM&R* = *Archives of Physical Medicine & Rehabilitation*; *Austr Physio* = *Journal of Physiotherapy*; *JOSPT* = *Journal of Orthopedic and Sports Physical Therapy; PEDro* = *Physiotherapy Evidence Database; Physio* = *Physiotherapy*; *Physio Can* =

*Physiotherapy* Canada; *Physio Pract* = *Physiotherapy Practice*; *Physio Res Int* = *Physiotherapy Research International*; *Physio Theory* = *Physiotherapy Theory and Practice*; *PTJ* = *Physical Therapy*; *PT Sci* = *Journal of Physical Therapy Science;* no. = number of core journals identified
[a]Journal of the American Physical Therapy Association

Each of the bibliometric citation analyses used source journals from which to map to the most frequently cited journals, ranging from a single journal to seven journals specific to physical therapist practice. Only one journal, *Physical Therapy*, was included as a source journal in all the studies. Bohannon and Gibson (1986) stated "Although analyses can be conducted using a large number of source journals, 'a good approximation' can be determined by starting with a journal or set of journals relevant to a particular field" (p. 540). The citation analyses studies varied in the time span for which citations were collected, including one year (Bohannon, 1999; Bohannon & Tiberio, 1989), three years (Fell et al., 2011; Wakiji, 1997), and four years (Bohannon & Gibson, 1986). Costa et al. (2010) and C. Maher et al. (2001) accessed the PEDro database on a single day to identify all indexed RCTs.

There is an inverse relationship between number of source journals and the time frame, likely a practical solution to manage the amount of data generated. Journals were ranked according to frequency of citation from the source journals. Core journals were identified as some portion of all the journals cited. Bradford's Law of Scattering directs that for a given field "there are a few very productive periodicals, a larger number of more moderate producers, and a still larger number of constantly diminishing productivity" (Bradford, Egan, and Shera (1953) as cited in Nash-Stewart, Kruesi, and Del Mar (2012, p. 135).

Practically applied, the top one-third of the most frequently cited journals are considered the core journals for the field. Wakiji (1997) applied Bradford's law directly using it to identify 14 core journals in physical therapy. Despite the variations in sources, time frames and methods, the lists of core journals in physical therapy, while varied, have many journals in common. *Physical Therapy* was consistently identified as the top ranked core journal specific to physical therapy. *Physiotherapy* was identified as a core journal in five studies (Bohannon, 1999; Bohannon & Gibson, 1986; Bohannon & Tiberio, 1989; Costa et al., 2010; Fell et al., 2011), *Physiotherapy Canada* in four studies (Bohannon, 1999; Bohannon & Gibson, 1986; Bohannon & Tiberio, 1989; Fell et al., 2011) and *Australian Journal of Physiotherapy* was included in three studies (Bohannon, 1999; Bohannon & Tiberio, 1989; Fell et al., 2011). *Physiotherapy* and *Physiotherapy Canada* were identified as core journals in only one study where they were not a source journal (Bohannon & Gibson, 1986). The *Journal of Orthopedics and Sports Physical Therapy* was identified as a core journal in two studies (Bohannon & Leveau, 1986; Costa et al., 2010) but was not a source journal for any of the citation analyses. Costa et al. (2010) found that *Physical Therapy* had about twice as many RCTs indexed in PEDro (161) as *Physiotherapy* (84) and the *Journal of Orthopedic and Sports Physical Therapy* (78).

Several physical therapy specific content analyses have been published utilizing a variety of sampling strategies, varying on time period and article selection method (Table 5). Coronado et al. (2011) reported an "increased emphasis on publishing articles consistent with evidence-based practice and clinically based research" (p. 642) and the findings were similar to other reviews such as C. G. Maher, Moseley, Sherrington, Elkins, and Herbert (2008) and Moseley, Herbert, Sherrington, and Maher (2002). Wiles et al.

(2012) reported the results were similar to other health professions (Gore, Nordberg, Palmer, & Piorun, 2009; Potter, 2010; Shadgan, Roig, HajGhanbari, & Reid, 2010). Common trends included an increasing total number of articles published (Coronado et al., 2011; Paci et al., 2009; Simon et al., 2014; Wiles et al., 2012), an increasing number of research articles with a concomitant decrease in the number of topical reviews (Coronado et al., 2011; Simon et al., 2014; Wiles et al., 2012), an increased use of symptomatic or patient populations as participants (Coronado et al., 2011; Simon et al., 2014), and an increased number of authors including more with international affiliations (Wiles et al., 2012).

**TABLE 5**

Bibliometric Content Analyses of Physical Therapy Literature

| Authors | Journal(s) | Time | Sampling | no. Articles |
|---|---|---|---|---|
| Miller et al. (2003) | *Austr Physio*<br>*PTJ*[a]<br>*Physio*<br>*Physio Can* | 6 months<br>Jan – June 2001 | 6 consecutive issues<br>All article types | 179 |

**TABLE 5** continued

Bibliometric Content Analyses of Physical Therapy Literature

| Authors | Journal(s) | Time | Sampling | no. Articles |
|---|---|---|---|---|
| Paci et al. (2009) | *Austr Physio*<br>*Geriatric PT*[a]<br>*Neuro PT*[a]<br>*JOSPT*[a]<br>*Ped PT*[a]<br>*PTJ*[a]<br>*Physio*<br>*Physio Res Int*<br>*Physio Theory* | 5 years<br>2003-07 | All issues<br>Research articles<br>Review articles | 1,627 |
| Coronado et al. (2011) | *PTJ*[a] | 30 years<br>1980-2009 | All issues<br>Research report<br>Topical review<br>Case report | 2,519 |
| Wiles et al. (2012) | *PTJ*[a] | 65 years<br>1945-2016 | 4 issues every 5th year at 3-month intervals within the year<br>All article types | 337 |

**TABLE 5** continued

Bibliometric Content Analyses of Physical Therapy Literature

| Authors | Journal(s) | Time | Sampling | no. Articles |
|---|---|---|---|---|
| Simon et al. (2014) | *JMMT* | 20 years 1993-2012 | All issues Research report Topical review Case report | 375 |
| Hoderlein et al. (2017) | *PEDro* | 2 years 2001 and 2015 | 10% of RCTs randomly selected in 2001 and in 2015 | 2001: n = 70<br>2015: n = 151 |

*Note: Austr Physio = Journal of Physiotherapy Australia*; *Geriatric PT = Geriatric Physical Therapy; JMMT = Journal of Manual and Manipulative Therapy; JOSPT = Journal of Orthopedic and Sports Physical Therapy; Neuro PT = Journal of Neurologic Physical Therapy; PEDro = Physiotherapy Evidence Database; Ped PT = Journal of Pediatric Physical Therapy; Physio = Physiotherapy*; *Physio Can = Physiotherapy Canada*; *Physio Res Int = Physiotherapy Research International*; *Physio Theory = Physiotherapy Theory and Practice*; *PTJ = Physical Therapy*.
[a]Journal of the American Physical Therapy Association

Coronado et al. (2011) found no change in the number of random control trials in *Physical Therapy* during the period 1980 – 2009, although others found an increase in RCTs in physical therapy literature. The percentage of RCT's varied from 10% in 2009 (Coronado et al., 2011) to 24.3% in 2010 in *Physical Therapy* (Wiles et al., 2012). Paci et al. (2009) reported 12.6% RCTs in 2007 across nine physical therapy specific journals. Less than 10% of articles were RCTs in the *Journal of Manual & Manipulative Therapy* (Simon et al., 2014). Differences, particularly between 2009 and 2010 in *Physical*

*Therapy*, may be due to varying sampling strategies or classification methods. The *Journal of Manual & Manipulative Therapy* had the lowest number of RCTs (less than 10%) and is the newest and most narrowly focused journal. Using the PEDro database to identify clinical trials, the number of clinical trials doubled comparing 2001 to 2015 (Hoderlein et al., 2017). Although the number of RCT's is increasing, the proportion of RCTs of all research articles remains small. Systematic reviews, although also trending upward, represent an even smaller proportion of articles ranging from less than 5% (Coronado et al., 2011) to 8.1% (Wiles et al., 2012).

The level of evidence is increasing in physical therapy specific journals but there is a paucity of evidence on the quality of the research studies. Miller et al. (2003) used the Hedges Project criteria to assess research articles for high quality evidence suitable for application to patient/client care in four physical therapy specific journals. Only 19 of 179 articles met the standard for sufficient rigor. All the assessed intervention studies in *Physical Therapy* (n = 7) met the Hedges Project standards compared to only 36% - 80% in the other three physical therapy related journals included in the study. None of the bibliometric content analyses examined trends in how outcomes were reported for application to patient/client clinical decision making, specifically the reporting of null hypothesis statistical testing, effect size and number needed to treat.

In the past 15 years, three physical therapy specific bibliometric analyses focused on the use of statistics in the physical therapy literature (Bandy, 2003; Roush et al., 2015; Tilson et al., 2016). The earliest study (Bandy, 2003) utilized content analysis to identify the type and frequency of statistical techniques used by articles identified as Research Reports in a single journal, *Physical Therapy*, during a two-year span of publication (2000-

2002; 90 articles). The intent was to inform educators about which commonly used statistical techniques should be included in the physical therapist research curriculum. The 10 most frequently occurring statistical techniques accounted for 82.4% of all statistical techniques used during the 2-year period. Five of the top 10 statistical techniques identified utilized NHST. However, EST and NNT were mentioned as included amongst the 307 statistical techniques identified in the study.

Similarly, Roush et al. (2015) used content analysis to identify commonly used statistical techniques in articles during a two-year period (2009-2010; 5,546 articles) in the 16 journals identified as core physical therapy or physiotherapy journals by Fell et al. (2011). Articles included in the analysis included those identified as research reports, scientific articles, original contributions, clinical investigations or brief reports. Journals that were considered of interest, but not specific to, physical therapists such as the British Medical Journal, and the Clinical Journal of Pain were included in this study as they were among the most frequently cited journals as mapped from four physical therapy or physiotherapy specific journals (*Physical Therapy*, *Physiotherapy*, *Physiotherapy Canada*, and *Australian Journal of Physiotherapy*).

However, only 6.0% of the articles reviewed were from physical therapy or physiotherapy specific journals. Despite the increased number and breadth of journals used in this study, the results were very similar to Bandy (2003). The top 10 statistical methods or categories of statistical methods (e.g. descriptive statistics were considered a category of statistical methods) accounted for 82.56% of the statistical methods used. (Bandy, 2003) included confidence intervals with descriptive statistics while Roush et al. (2015) listed them separately, ranking second in frequency (9.29%). Effect size was the

15th most frequently occurring statistical method (1.04%). NNT was not included in the top 25 statistical methods and may have been among the 534 statistical methods not reported but could not have occurred more frequently than 0.06%.

The third bibliometric analysis related to the use of statistical methods in the physical therapy literature expanded the definition of statistical methods. Tilson et al. (2016) stated that focusing specifically on statistical techniques did not address the level of understanding that physical therapists require to understand, interpret and apply the results of a research study. Consequently, Tilson et al. (2016) asks the question “What are the most common statistical terms and research concepts physical therapists are likely to encounter in the physical therapy literature that need to be included in professional education curricula?” (p.119). The method for identifying statistical terms is not clear, initially beginning with 532 terms which were collapsed into 321 representative terms.

Examples of statistical terms unique to this study compared to the previous two statistical bibliometric studies included statistical significance, *p*-value, significance level, minimal detectable change, minimally clinically important difference and degrees of freedom among many others. The sample included all research, case series and case report articles published in the 14 peer-reviewed journals associated with the American Physical Therapy Association during a 12-month period (2011-2012; 391 articles). Confidence intervals and effect size were included in a category labeled “Clinically Meaningful Statistics” and were referred to in 34.8% and 11.5% of the articles included in this study respectively. However, Cohen’s *d* was listed in a separate category (3.6%) even though it is an effect size. NNT was reported in supplemental materials (Additional File 1 – Statistical Terms) as having been reported in four articles (1.0%)

Common to these three studies is the limitation that using a statistical technique in a study does not necessarily make it the appropriate statistic. These studies report frequencies of occurrence which cannot be interpreted to mean that they are also the most important. Recommendations for modifications to physical therapist education research curriculum need to be interpreted in light of these limitations. Interestingly, all three bibliometric analyses found little to no use of effect size or number needed to treat in the articles reviewed. The low frequency of occurrence may reflect the breadth of research designs included in the sample as effect size and number needed to treat are primarily limited to intervention studies.

# CHAPTER 3 METHODS

## Design

This study is a bibliometric content analysis of clinical trial/intervention studies published in physical therapy specific literature.

## Human Subjects

No human subjects were involved in this bibliometric analysis of published research.

## Sample

### *Journal Selection.*

Three journals of the American Physical Therapy Association (APTA) were used as the source journals for articles included in this bibliometric study: (a) *Physical Therapy*, the (b) *Journal of Orthopaedic and Sports Physical Therapy* and the (c) *Journal of Neurologic Physical Therapy). Physical Therapy* has been consistently included as a source journal or target journal for citation analysis or content analysis of the physical therapy literature (see Tables 4, 5). It is a well-established international journal with the largest circulation of all physical therapy specific journals. Editorial policy is consistent across APTA journals. The *Journal of Orthopaedic and Sports Physical Therapy* (*JOSPT*) and the *Journal of Neurologic Physical Therapy* represent the two largest areas of physical therapist practice (Human Resources Research Organization, 2017). Prior to 2003, the *Journal of Neurologic Physical Therapy* was known as *Neurology Report*, but for the purposes of this project it was labeled as its current name for the entire final dataset.

***Article Selection.***

All research reports that are clinical trials as defined by the NIH and published in *Physical Therapy,* the *Journal of Orthopaedic and Sports Physical Therapy* or the *Journal of Neurologic Physical Therapy* between July 1989 and July 2018 qualified for potential inclusion in the final dataset sample. NIH defines a clinical trial as “A research study in which one or more human participants are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes” (National Institutes of Health, 2014). NIH clarified the definition in 2018 by adding that researchers:

> apply the following four questions to determine whether NIH would consider the research study to be a clinical trial:
>
> - Does the study involve human participants?
> - Are the participants prospectively assigned to an intervention?
> - Is the study designed to evaluate the effect of the intervention on the participants?
> - Is the effect being evaluated a health-related biomedical or behavioral outcome?
>
> If the answer to all four questions is “yes” then the clinical study would be considered a clinical trial according to the NIH definition. (National Institutes of Health, 2018)

See Figure 1 for the decision flow chart for including an article in the study.

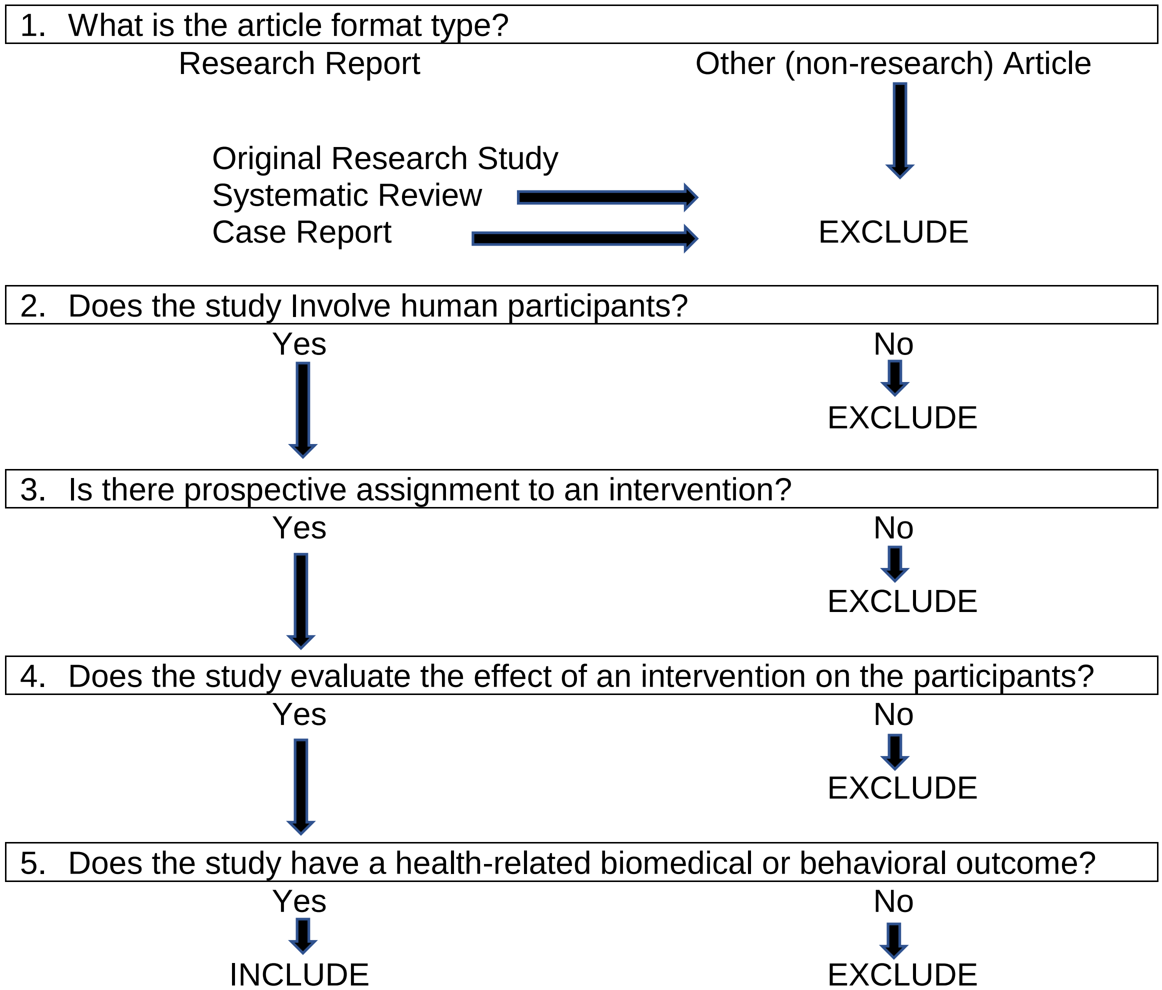


*Figure 1.* Article inclusion decision flow chart. Adapted from “Notice of Revision of ‘NIH Definition of Clinical Trial Case Studies” by National Institutes of Health, 2018, https://grants.nih.gov/policy/clinical-trials/case-studies.htm

Most articles that qualified as clinical trials under the NIH definition also qualified for inclusion in this study. However, because the purpose of this study is to investigate effect size, specifically NNT, within clinical trials, certain trials that intrinsically preclude calculation of NNT, such as case studies or other trials where n = 1, were excluded from this study. In addition, secondary analyses, e.g. systematic reviews, were not included to avoid the potential for clinical trials to be included multiple times. Abstracts and article briefs from clinical trials were excluded as these were typically reporting on articles published previously in other journals.

**Bibliometric Variables**

Bibliometric variables were divided into three categories: (a) characteristics related to the publication (Appendix A), (b) characteristics related to author and institution (Appendices B) and (c) characteristics related to the study design and data analysis (Appendix C). Information from each article was coded based on the variables and entered into a database spreadsheet. In cases where more than one outcome variable was reported in an article only one was recorded. If one of the outcome variables was a dichotomous outcome variable and NNT was reported it was chosen to be coded. If there was not a dichotomous outcome variable the first outcome variable reported was chosen to be coded.

**Data Analysis**

An a priori sample size analysis was conducted to determine the minimum required sample size for the study. A chi-squared goodness of fit test was calculated estimating a small to medium effect size ($r$ = 0.3), $df$ = 4 and a nominal alpha (α) = 0.05. A minimum sample size of 133 qualified clinical trial journal articles were required to achieve a statistical power of at least 0.80. The critical chi-squared value for this study size was calculated to be $\chi^2 = 9.49$.

Descriptive statistics (count, percent, cumulative percent) were calculated for variables describing the sample in total and over time. Variables relevant to characterizing the reporting of research design and outcomes were analyzed using descriptive statistics annually and for 5-year time intervals beginning with 1993 and ending with 2017 (1993–1997; 1998–2002; 2003–2007; 2008–2012; 2013–2017) in total and individually for each journal. Specifically, variables described over time included (a) number of articles that were a clinical trial as defined by the NIH, (b) number of articles in which an effect size was reported, (c) number of articles in which a dichotomous outcome variable was reported, and (d) the number of articles in which NNT was reported. Additionally, contingency tables displaying the expected and observed distribution of reporting type (NHST, effect size and NNT) over 5-year periods and separated by journal are presented along with the chi-squared distribution tests. These tests were performed to determine if the incidence of each reporting type increases or decreases over time, and if the distribution of reporting type is statistically significantly different among the three journals.

Descriptive statistic tables, contingency tables and accompanying chi-squared distribution tests were calculated and organized in JASP version 0.9.1.0 (JASP Team, 2018). Pearson correlation was calculated using SPSSv24. All primary figures were created in R version 3.5.1 ‘feather spray’ (R Core Team, 2018) with the package Kendall installed and using the graphical package ‘ggplot2’ (Wickham, 2016). A priori and post-hoc power analyses were conducted using G*Power 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007).

# CHAPTER 4 RESULTS

## Journals

Three journals of the American Physical Therapy Association, the *Journal of Neurologic Physical Therapy*, the *Journal of Orthopaedic and Sports Physical Therapy* and *Physical Therapy,* were used as source journals for this study. Characteristics of these journals can be found in Table 6. They are published in the USA, are highly ranked among rehabilitation journals, include authors from international institutions, and have Impact Factors that increased during the span of this study, which is 1989 – 2018 (Clarivate & Institute for Scientific, 1997).

TABLE 6

Source Journals

| | *JNPT* | *JOSPT* | *PTJ* |
|---|---|---|---|
| Inception | 1976 | 1979 | 1931 |
| Issues per Year | 4 | 12 | 12 |
| Publisher | Lippincott, USA | APTA, USA | Oxford University Press, USA |
| No. Contributing Countries (2017) | 21 | 34 | 50 |
| No. Contributing Organizations (2017) | 50 | 50 | 50 |
| Rank in Rehab Journals (2017) | 5/65 Q1 | 9/65 Q1 | 13/65 Q1 |
| Rank in Orthopedic Journals (2017) | NA | 12/77 Q1 | 23/77 Q2 |

TABLE 6 continued

Source Journals

| | *JNPT* | *JOSPT* | *PTJ* |
|---|---|---|---|
| Rank in Clinical Neurology Journals | 51/187 | NA | NA |
| (2017) | Q1 | | |
| 1-year Impact Factor (1997) | NA | 0.576 | 0.833 |
| 1-year Impact Factor (2017) | 3.633 | 2.090 | 2.587 |
| 5-year Impact Factor (2017) | 3.743 | 4.061 | 3.343 |
| Medline Indexed | since 2005 | Yes | Yes |

*Note*. *JNPT = Journal of Neurologic Physical Therapy*; *JOSPT = Journal of Orthopedic and Sports Physical Therapy*; *PTJ = Physical Therapy*; Rehab = rehabilitation; Q1 = ranked in 1st quartile. Adapted from Clarivate and Institute for Scientific (1997).

**Articles**

**Descriptive statistics.**

A total of 448 articles met the inclusion criteria for this study across the three physical therapy specific journals, the *Journal of Neurologic Physical Therapy,* the *Journal of Orthopaedic and Sports Physical Therapy* and *Physical Therapy.* The distribution frequency table for the included articles from the three journals can be found in Table 7. The most clinical trials which met the inclusion criteria were published by the J*ournal of Orthopaedic and Sports Physical Therapy* followed by *Physical Therapy*. The *Journal of Neurologic Physical Therapy* published only 11% of the articles meeting the inclusion criteria.

Table 7

Frequencies of Included Articles by Journal

| Journal Name | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|
| *Journal of Neurologic Physical Therapy* | 50 | 11.2 | 11.2 | 11.2 |
| *Journal of Orthopaedic & Sports Physical Therapy* | 215 | 48.0 | 48.0 | 59.2 |
| *Physical Therapy* | 183 | 40.8 | 40.8 | 100.0 |
| Missing | 0 | 0.0 | | |
| Total | 448 | 100.0 | | |

Post-hoc power analysis was conducted on the final acquired sample size for this dataset. The a priori minimum sample size for a chi-squared goodness-of-fit test estimating a small to medium effect size $r$ = 0.3, a nominal alpha (α) = 0.05, and 4 degrees of freedom was calculated to be 133 qualified journal articles to achieve a statistical power of at least 0.80; the final sample size of 448 articles meeting the inclusion criteria resulted in a statistical power of 0.99.

Frequency tables displaying total number of articles reporting NHST, effect size (EST) and NNT measurements for each journal are summarized in Tables 8-10. Total percentage of qualified articles that reported NHST, effect size and NNT separated by journal is presented in Table 11 and graphically in Figure 2. Annual number of articles in which NHST, EST and NNT are reported, separated by journal, is shown in Figure 3. Total number of qualified articles, separated by journal are displayed annually in Figure 4, and over 5-year periods in Figure 5.

Table 8

Frequencies of Articles in which NHST is Reported by Journal

| **Journal Name** | **NHST** | **Frequency** | **Percent** | **Valid Percent** | **Cumulative Percent** |
|---|---|---|---|---|---|
| *Journal of Neurologic Physical Therapy* | R | 47 | 94.0 | 94.0 | 94.0 |
| | NR | 3 | 6.0 | 6.0 | 100.0 |
| | Missing | 0 | 0.0 | | |
| | Total | 50 | 100.0 | | |
| *Journal of Orthopaedic & Sports Physical Therapy* | R | 209 | 97.2 | 97.2 | 97.2 |
| | NR | 6 | 2.8 | 2.8 | 100.0 |
| | Missing | 0 | 0.0 | | |
| | Total | 215 | 100.0 | | |
| *Physical Therapy* | R | 171 | 93.4 | 93.4 | 93.4 |
| | NR | 12 | 6.6 | 6.6 | 100.0 |
| | Missing | 0 | 0.0 | | |
| | Total | 183 | 100.0 | | |

*Note*. NHST = Null Hypothesis Statistical Test; R = Reported; NR = Not Reported.

Table 9

Frequencies of Articles in which EST is Reported by Journal

| Journal Name | EST | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| *Journal of Neurologic Physical Therapy* | R | 13 | 26.0 | 26.0 | 26.0 |
| | NR | 37 | 74.0 | 74.0 | 100.0 |
| | Missing | 0 | 0.0 | | |
| | Total | 50 | 100.0 | | |
| *Journal of Orthopaedic & Sports Physical Therapy* | R | 32 | 14.9 | 14.9 | 14.9 |
| | NR | 183 | 85.1 | 85.1 | 100.0 |
| | Missing | 0 | 0.0 | | |
| | Total | 215 | 100.0 | | |
| *Physical Therapy* | R | 32 | 17.5 | 17.5 | 17.5 |
| | NR | 151 | 82.5 | 82.5 | 100.0 |
| | Missing | 0 | 0.0 | | |
| | Total | 183 | 100.0 | | |

*Note:* EST = Effect Size Test; R = Reported; NR = Not Reported.

Table 10

Frequencies of Articles in which NNT is Reported by Journal

| Journal Name | NNT | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| *Journal of Neurologic Physical Therapy* | R | 0 | 0.0 | 0.0 | 0.0 |
| | NR | 50 | 100.0 | 100.0 | 100.0 |
| | Missing | 0 | 0.0 | | |
| | Total | 50 | 100.0 | | |
| *Journal of Orthopaedic & Sports Physical Therapy* | R | 3 | 1.4 | 1.4 | 1.4 |
| | NR | 212 | 98.6 | 98.6 | 100.0 |
| | Missing | 0 | 0.0 | | |
| | Total | 215 | 100.0 | | |
| *Physical Therapy* | R | 5 | 2.7 | 2.7 | 2.7 |
| | NR | 178 | 97.3 | 97.3 | 100.0 |
| | Missing | 0 | 0.0 | | |
| | Total | 183 | 100.0 | | |

*Note.* NNT = Number Needed to Treat; R = Reported; NR = Not Reported.

Table 11

Total Percentage of Reporting Type by Journal

| Type | Journal Name | Percentage |
|---|---|---|
| NHST | *Journal of Neurologic Physical Therapy* | 94.00 |
| | *Journal of Orthopaedic & Sports Physical Therapy* | 97.21 |
| | *Physical Therapy* | 93.44 |
| EST | *Journal of Neurologic Physical Therapy* | 26.00 |
| | *Journal of Orthopaedic & Sports Physical Therapy* | 14.88 |
| | *Physical Therapy* | 17.49 |
| NNT | *Journal of Neurologic Physical Therapy* | 0.00 |
| | *Journal of Orthopaedic & Sports Physical Therapy* | 1.40 |
| | *Physical Therapy* | 2.73 |

*Note.* NHST = Null Hypothesis Statistical Test; EST = Effect Size Test; NNT = Number Needed to Treat.

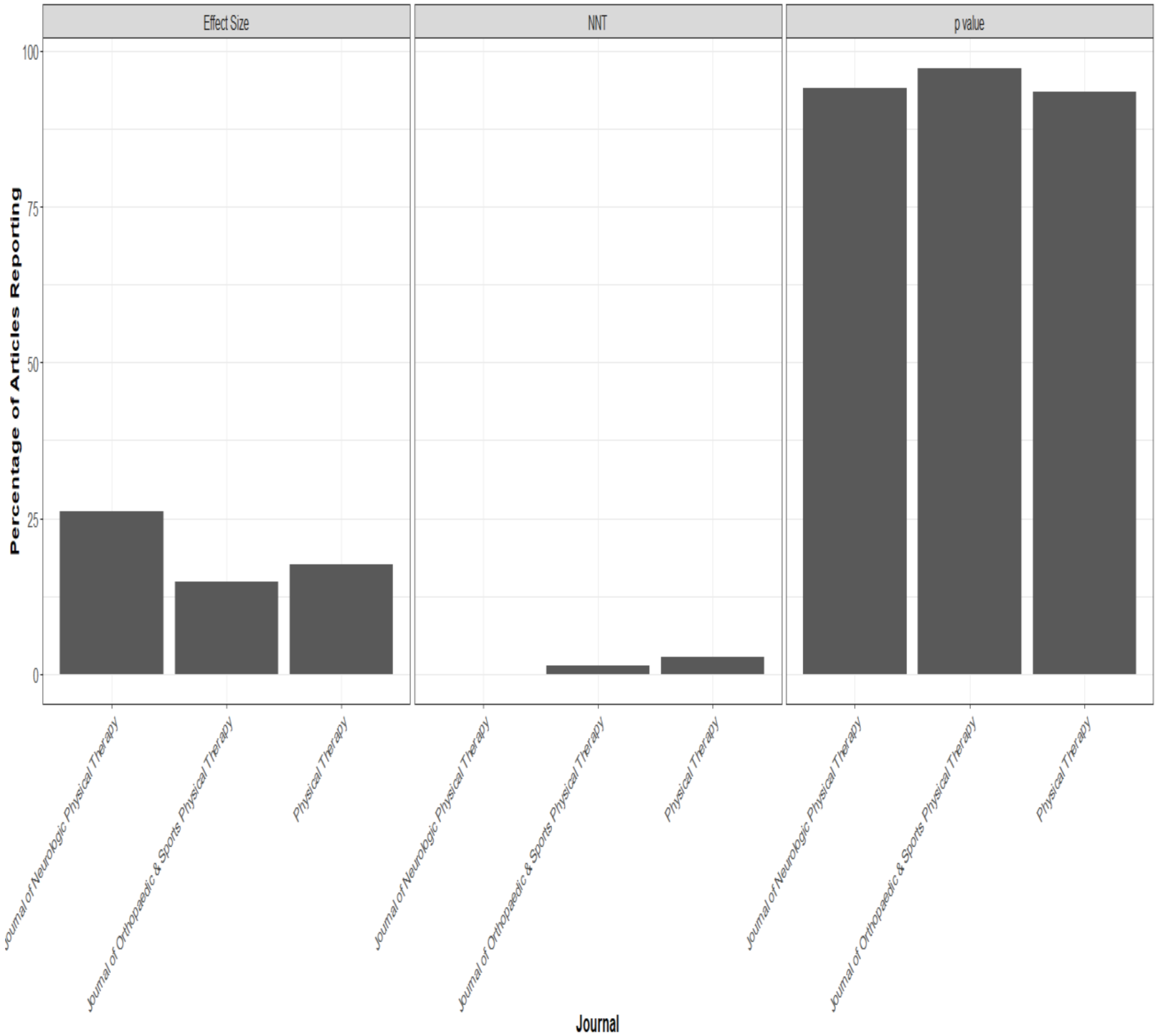


*Figure 2.* Total percentage of reporting type by journal. NNT = number needed to treat; p value = report from a null hypothesis statistical test.

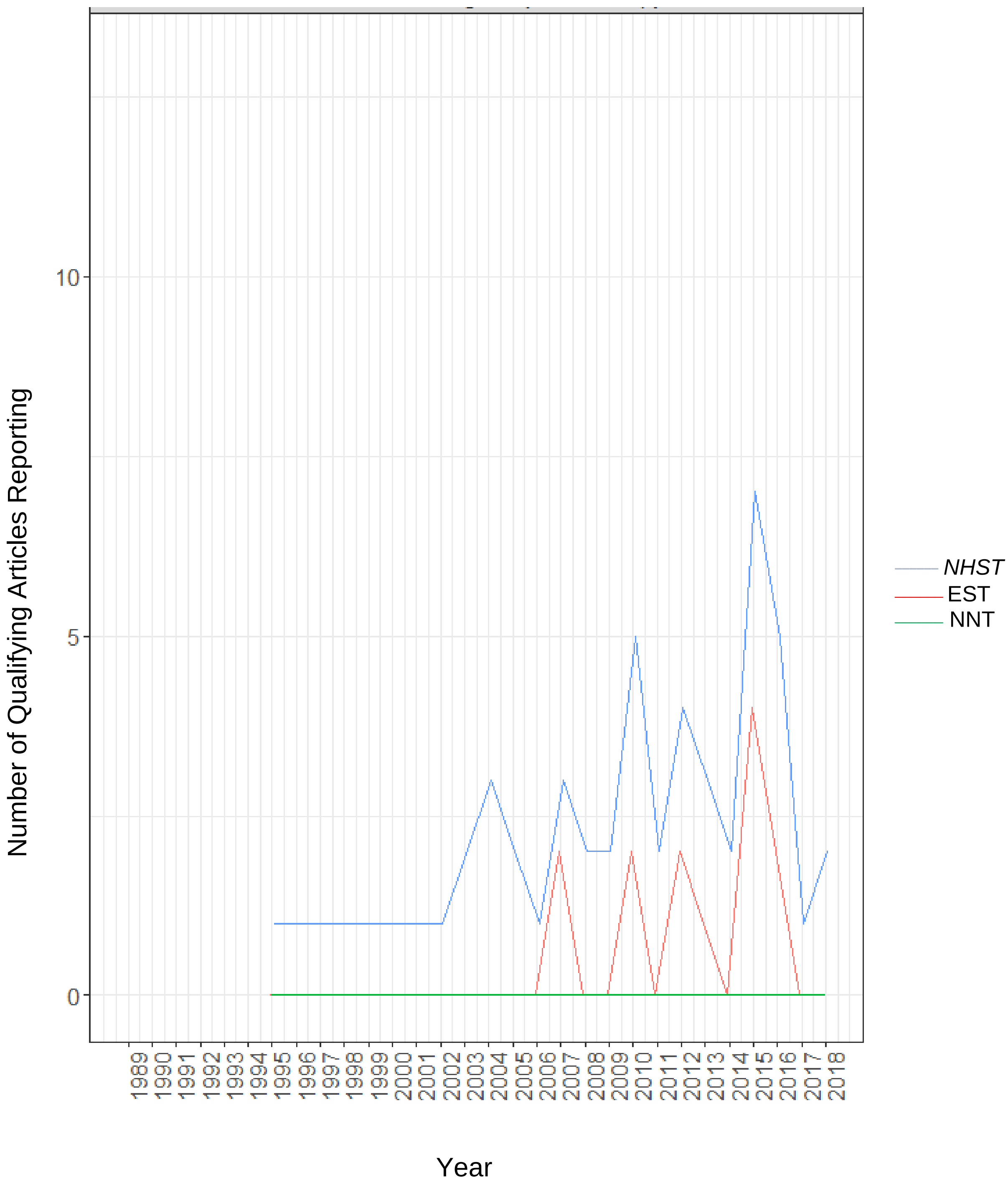


*Figure 3.* Frequency of reporting NHST, EST and NNT in the *Journal of Neurologic Physical Therapy*. NHST = null hypothesis statistical testing; EST = effect size; NNT = number needed to treat.

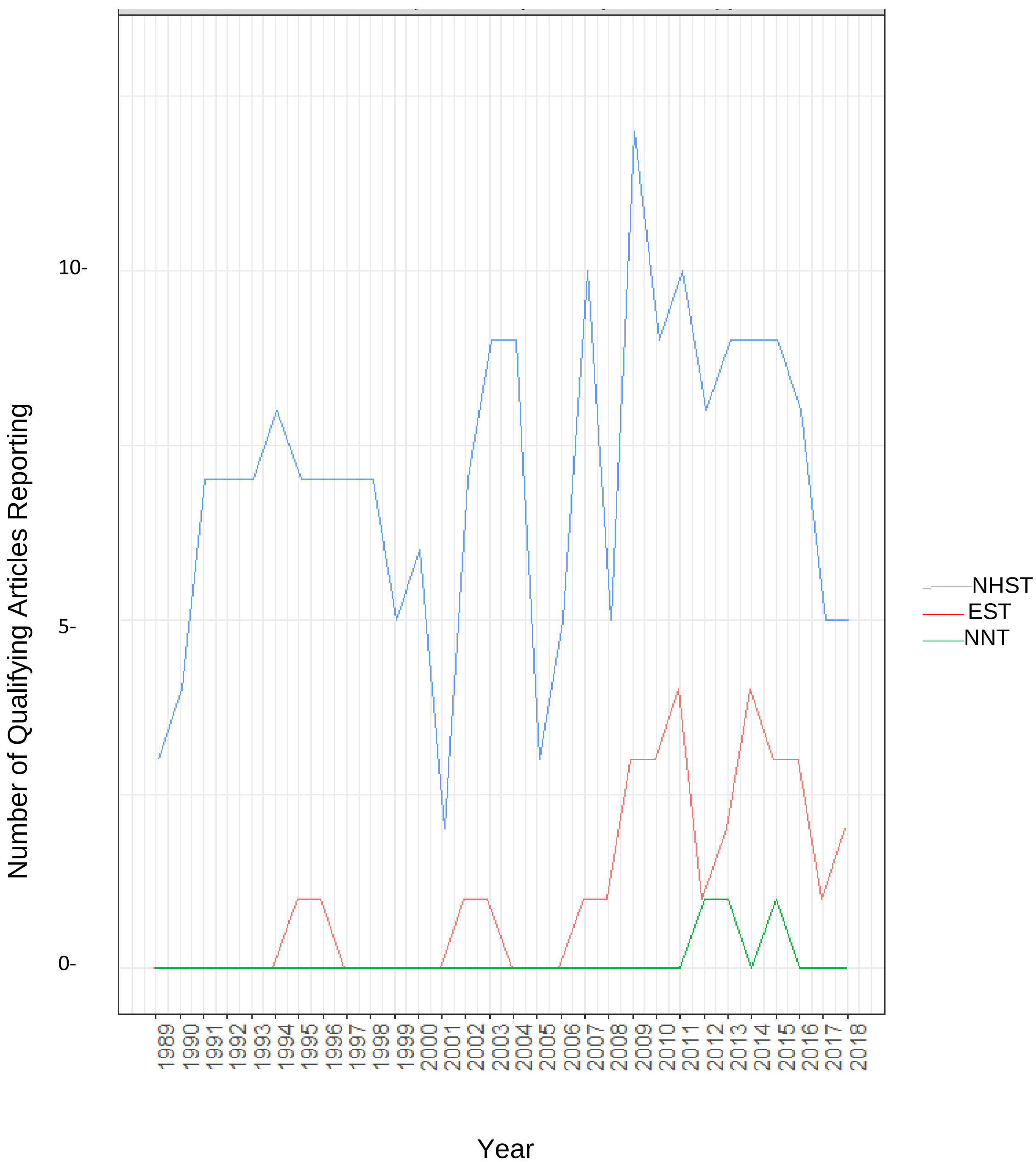


*Figure 4.* Frequency of reporting NHST, EST and NNT in the *Journal of Orthopaedic and Sports Physical Therapy*. NHST = null hypothesis statistical testing; EST = effect size; NNT = number needed to treat.

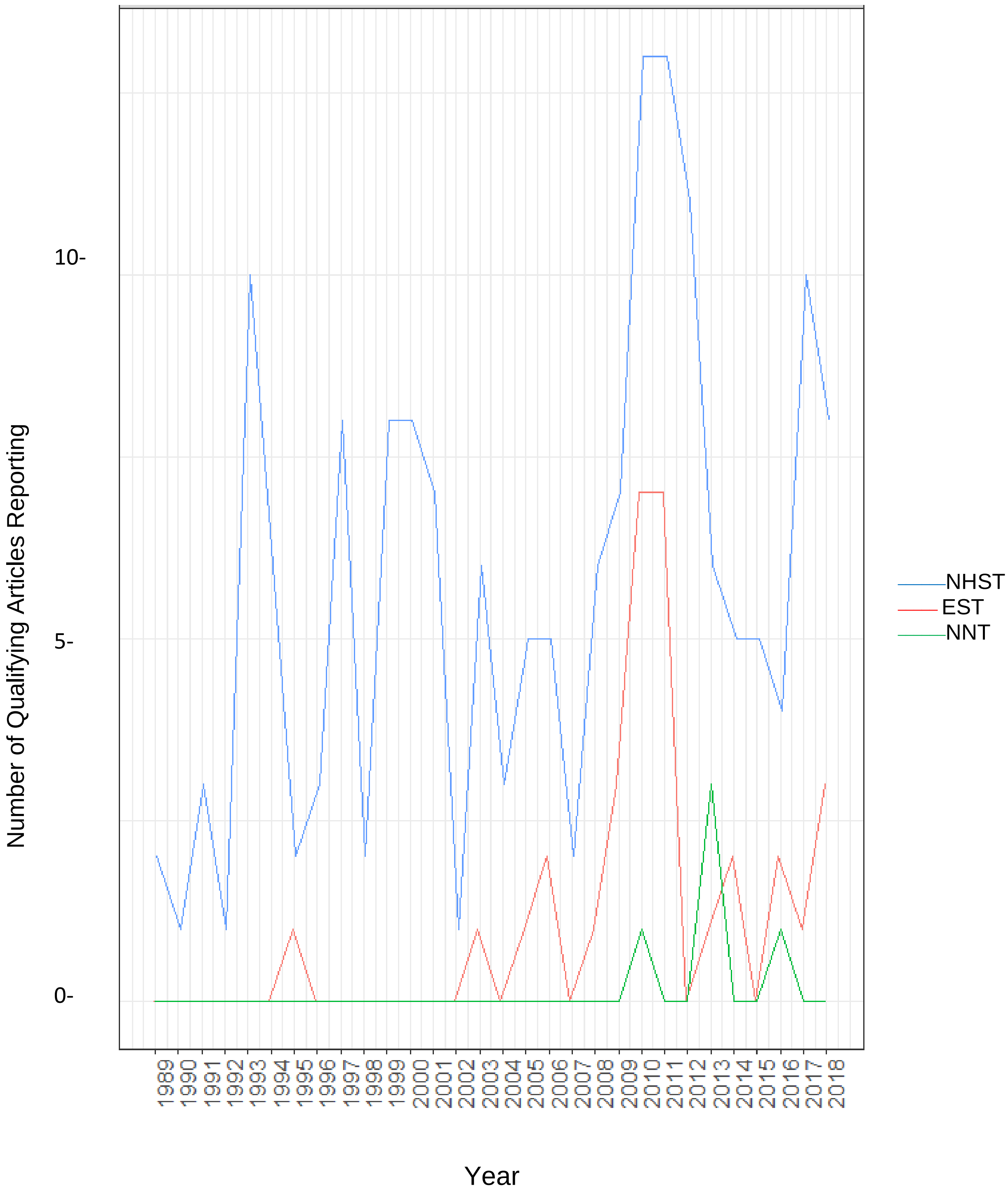


*Figure 5:* Frequency of reporting NHST, EST and NNT in *Physical Therapy*. NHST = null hypothesis statistical testing; EST = effect size; NNT = number needed to treat.

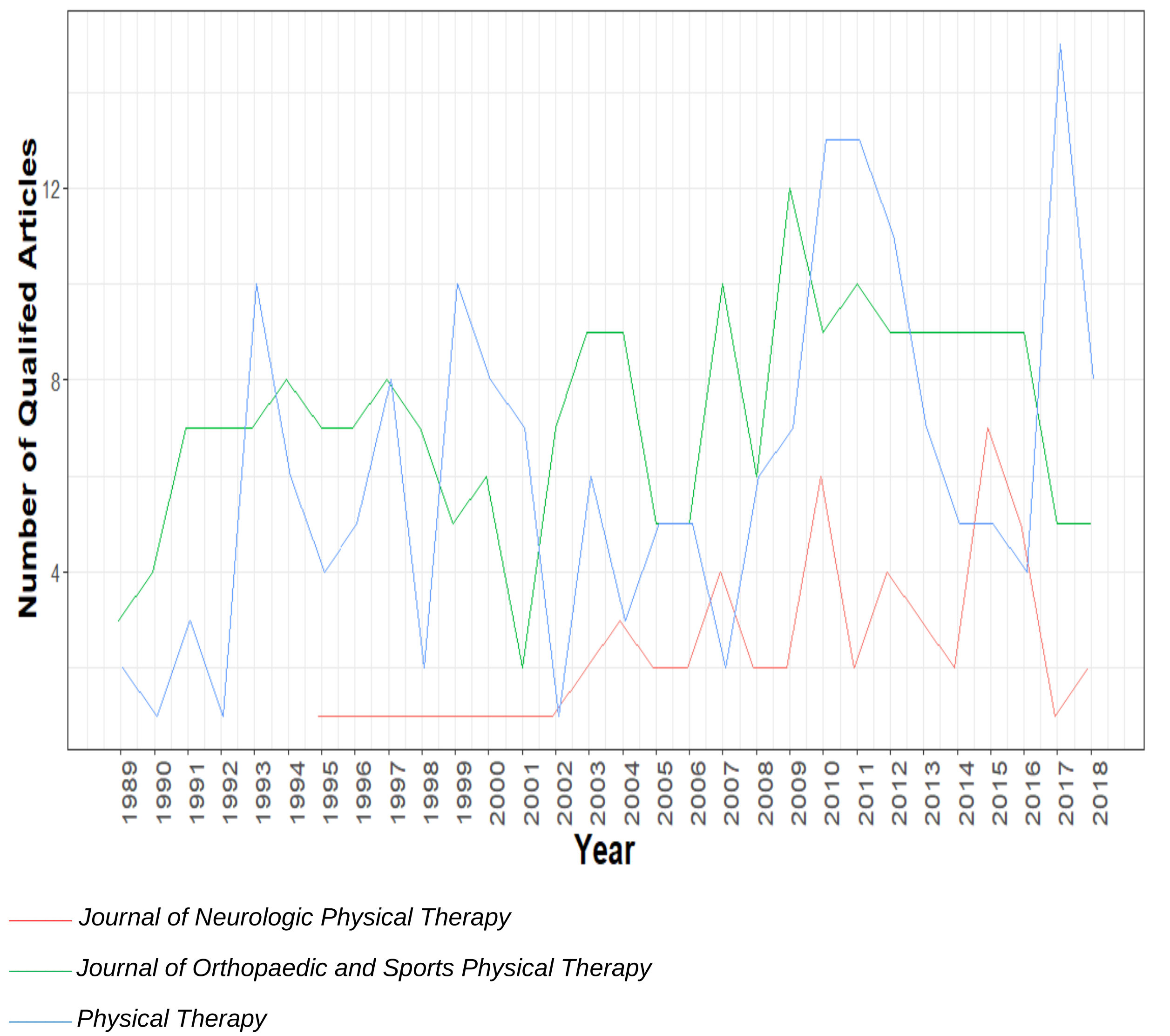


*Figure 6*. Total number of qualified articles published by year, separated by journal.

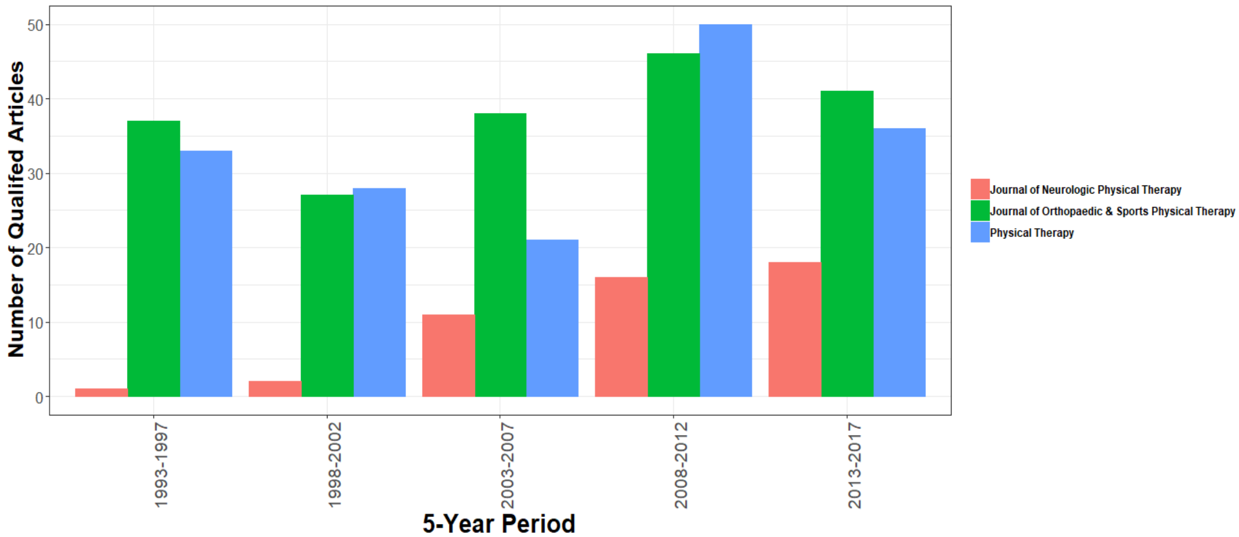


*Figure 7*. Total number of qualified articles by 5-Year Period, separated by journal

**Statistical analysis.**

Contingency tables displaying the expected and observed distribution of reporting type (NHST, EST and NNT) over 5-year periods are presented along with their chi-squared distribution tests and likelihood ratios in Tables 12-14. Contingency tables displaying the expected and observed distribution of reporting separated by journal are presented along with their chi-squared distribution tests and likelihood ratios in Tables 15-17.

Table 12

5-Year Contingency Table and Chi-Squared for Report of NHST

| **5-Yr Period** | | **NHST** R | **NHST** NR | **Total** |
|---|---|---|---|---|
| 1993-1997 | Count | 66.00 | 5.00 | 71.00 |
| | Expected count | 67.32 | 3.68 | 71.00 |
| 1998-2002 | Count | 55.00 | 2.00 | 57.00 |
| | Expected count | 54.04 | 2.96 | 57.00 |
| 2003-2007 | Count | 66.00 | 4.00 | 70.00 |
| | Expected count | 66.37 | 3.63 | 70.00 |
| 2008-2012 | Count | 109.00 | 3.00 | 112.00 |
| | Expected count | 106.19 | 5.81 | 112.00 |
| 2013-2017 | Count | 88.00 | 7.00 | 95.00 |
| | Expected count | 90.07 | 4.93 | 95.00 |
| Total | Count | 384.00 | 21.00 | 405.00 |
| | Expected count | 384.00 | 21.00 | 405.00 |

Chi-Squared Tests

| | **Value** | ***df*** | ***p*** |
|---|---|---|---|
| $X^2$ | 3.216 | 4 | 0.522 |
| Likelihood ratio | 3.399 | 4 | 0.493 |
| N | 405 | | |

*Note.* NHST = Null Hypothesis Statistical Test; R = Reported; NR = Not Reported.

Table 13

5-Year Contingency Table and Chi-Squared for Report of EST

| 5-Yr Period | | EST R | EST NR | Total |
|---|---|---|---|---|
| 1993-1997 | Count | 3.00 | 68.00 | 71.00 |
| | Expected count | 12.62 | 58.38 | 71.00 |
| 1998-2002 | Count | 1.00 | 56.00 | 57.00 |
| | Expected count | 10.13 | 46.87 | 57.00 |
| 2003-2007 | Count | 8.00 | 62.00 | 70.00 |
| | Expected count | 12.44 | 57.56 | 70.00 |
| 2008-2012 | Count | 34.00 | 78.00 | 112.00 |
| | Expected count | 19.91 | 92.09 | 112.00 |
| 2013-2017 | Count | 26.00 | 69.00 | 95.00 |
| | Expected count | 16.89 | 78.11 | 95.00 |
| Total | Count | 72.00 | 333.00 | 405.00 |
| | Expected count | 72.00 | 333.00 | 405.00 |

Chi-Squared Tests

| | **Value** | ***df*** | ***p*** |
|---|---|---|---|
| $X^2$ | 38.97 | 4 | < .001 |
| Likelihood ratio | 45.39 | 4 | < .001 |
| N | 405 | | |

*Note:* EST = Effect Size Test; R = Reported; NR = Not Reported

Table 14

5-Year Contingency Table and Chi-Squared for Report of NNT

| | | NNT | | |
|---|---|---|---|---|
| **5-Yr Period** | | **R** | **NR** | **Total** |
| 1993-1997 | Count | 0.00 | 71.00 | 71.00 |
| | Expected count | 1.40 | 69.60 | 71.00 |
| 1998-2002 | Count | 0.00 | 57.00 | 57.00 |
| | Expected count | 1.13 | 55.87 | 57.00 |
| 2003-2007 | Count | 0.00 | 70.00 | 70.00 |
| | Expected count | 1.38 | 68.62 | 70.00 |
| 2008-2012 | Count | 2.00 | 110.00 | 112.00 |
| | Expected count | 2.21 | 109.79 | 112.00 |
| 2013-2017 | Count | 6.00 | 89.00 | 95.00 |
| | Expected count | 1.88 | 93.12 | 95.00 |
| Total | Count | 8.00 | 397.00 | 405.00 |
| | Expected count | 8.00 | 397.00 | 405.00 |

Chi-Squared Tests

| | **Value** | ***df*** | ***p*** |
|---|---|---|---|
| $X^2$ | 13.25 | 4 | 0.010 |
| Likelihood ratio | 13.81 | 4 | 0.008 |
| N | 405 | | |

*Note:* NNT = Number Needed to Treat; R = Reported; NR = Not Reported.

Table 15

Contingency Table and Chi-Squared for Report of NHST by Journal

| Journal Name | | NHST R | NHST NR | Total |
|---|---|---|---|---|
| *Journal of Neurologic Physical Therapy* | Count | 47.00 | 3.00 | 50.00 |
| | Expected count | 47.66 | 2.34 | 50.00 |
| *Journal of Orthopaedic & Sports Physical Therapy* | Count | 209.00 | 6.00 | 215.00 |
| | Expected count | 204.92 | 10.08 | 215.00 |
| *Physical Therapy* | Count | 171.00 | 12.00 | 183.00 |
| | Expected count | 174.42 | 8.58 | 183.00 |
| Total | Count | 427.00 | 21.00 | 448.00 |
| | Expected count | 427.00 | 21.00 | 448.00 |

Chi-Squared Tests

| | **Value** | ***df*** | ***p*** |
|---|---|---|---|
| $X^2$ | 3.356 | 2 | 0.187 |
| Likelihood ratio | 3.472 | 2 | 0.176 |
| N | 448 | | |

*Note:* NHST = Null Hypothesis Statistical Test; R = Reported; NR = Not Reported.

Table 16

Contingency Table and Chi-Squared for Report of EST by Journal

| Journal Name | | EST R | EST NR | Total |
|---|---|---|---|---|
| *Journal of Neurologic Physical Therapy* | Count | 13.00 | 37.00 | 50.00 |
| | Expected count | 8.59 | 41.41 | 50.00 |
| *Journal of Orthopaedic & Sports Physical Therapy* | Count | 32.00 | 183.00 | 215.00 |
| | Expected count | 36.95 | 178.05 | 215.00 |
| *Physical Therapy* | Count | 32.00 | 151.00 | 183.00 |
| | Expected count | 31.45 | 151.55 | 183.00 |
| Total | Count | 77.00 | 371.00 | 448.00 |
| | Expected count | 77.00 | 371.00 | 448.00 |

Chi-Squared Tests

| | **Value** | ***df*** | ***p*** |
|---|---|---|---|
| $X^2$ | 3.541 | 2 | 0.170 |
| Likelihood ratio | 3.279 | 2 | 0.194 |
| N | 448 | | |

*Note:* EST = Effect Size Test; R = Reported; NR = Not Reported.

Table 17

Contingency Table and Chi-Squared for Report of NNT by Journal

| Journal Name | | NNT R | NNT NR | Total |
|---|---|---|---|---|
| *Journal of Neurologic Physical Therapy* | Count | 0.00 | 50.00 | 50.00 |
| | Expected count | 0.89 | 49.11 | 50.00 |
| *Journal of Orthopaedic & Sports Physical Therapy* | Count | 3.00 | 212.00 | 215.00 |
| | Expected count | 3.84 | 211.16 | 215.00 |
| *Physical Therapy* | Count | 5.00 | 178.00 | 183.00 |
| | Expected count | 3.27 | 179.73 | 183.00 |
| Total | Count | 8.00 | 440.00 | 448.00 |
| | Expected count | 8.00 | 440.00 | 448.00 |

Chi-Squared Tests

| | **Value** | ***df*** | ***p*** |
|---|---|---|---|
| $X^2$ | 2.031 | 2 | 0.362 |
| Likelihood ratio | 2.809 | 2 | 0.245 |
| N | 448 | | |

*Note:* NNT = Number Needed to Treat; R = Reported; NR = Not Reported.

As indicated in Tables 12-14, although there is a statistically even distribution of NHST being reported in qualified articles over every 5-year period ($p = 0.522$, Table 12), the distribution of EST and NNT are uneven ($p < 0.001$, Table 13, $p = 0.010$, Table 14 respectively). Specifically, although NHST was evenly reported over every 5-year period, EST was under-represented from 1993-2007 and over-represented from 2008 onward

(Table 13). Similarly, NNT was under-represented from 1993-2012, but over-represented in the 2013-2017 period (Table 14). This indicated rather than a consistent distribution of reporting EST and NNT from 1993-2017, they become statistically more prevalent in the literature.

As indicated in Tables 15-17, this phenomenon was not dependent on journal type. There was no statistically significant difference in reporting frequency among journals for NHST ($p$ = 0.187, Table 15), EST ($p$ = 0.170, Table 16) or NNT ($p$ = 0.332, Table 17). This occurred despite *The Journal of Neurologic Physical Therapy* not having a single qualified article report NNT from 1989-2017. It may have been due to the frequency of *Physical Therapy* and *The Journal of Orthopaedic and Sports Physical Therapy* with low incidences of reporting NNT as well (2.7% & 1.4% respectively, Table 10).

**Articles in which NNT was reported**

The authors of only eight (1.79%) of the 448 articles meeting the inclusion criteria for this study reported NNT for at least one outcome variable. Citations for these articles are found in Appendix C. The patient/client population for all eight studies was orthopedics. Intervention was provided for neck pain in three (37.5%) of the eight studies (Cleland et al., 2010; Dunning et al., 2012; Masaracchio, Cleland, Hellman, & Hagins, 2013) and were multi-center studies conducted exclusively in the United States. The *Global Rating of Change* (GROC) (Jaeschke, Singer, & Guyatt, 1989) was used as an outcome measure in each of these studies and was used to determine the NNT. Dunning et al. (2012) also reported NNT using the *Neck Disability Index* (MacDermid et al., 2009). One author was common to these three studies (Cleland). These three studies were published over the course of four years, 2010 – 2012. The Cleland et al. (2010) study was published in *Physical Therapy* and the authors were the first to report NNT in any of the three source physical therapy specific journals in this study.

Intervention was provided for chronic low back pain in three (37.5%) of the eight studies (Garcia et al., 2013; Miyamoto, Leonardo Oliveira Pena, Galvanin, & Cristina Maria Nunes, 2013; Siemonsma et al., 2013). These were conducted exclusively outside of the United States, two in Brazil (Garcia et al., 2013; Miyamoto et al., 2013) and the third in the Netherlands and United Kingdom (Siemonsma et al., 2013). One author (Costa) was common to the two studies conducted in Brazil. All three studies were published in 2013 in *Physical Therapy* and each used different outcome measures to determine the NNT. The remaining two studies were the most recently published of the eight articles (Abbott et al., 2015; Christiansen et al., 2016). Published in the *Journal of Orthopaedic and Sports Physical Therapy* and *Physical Therapy*, intervention was provided for knee and shoulder diagnoses respectively. Christiansen et al. (2015) was conducted internationally (Denmark, Germany, United Kingdom) and Abbott et al. (2015) was conducted both in the United States and New Zealand. The outcome measures used to determine the NNT in these studies were unique to these studies. The computed NNT(s) reported in the eight studies can be found in Table 18.

Table 18

Outcome Measures Reported using NNT

| Article | System<br><br>Diagnosis | Outcome Measure | NNT | NNT 95% CI |
|---|---|---|---|---|
| Christiansen et al. (2016) | Orthopedics<br><br>Subacromial Impingement Syndrome post-surgery | OSS | 5.0 | [2.6, 48.6] |
| Abbott et al. (2015)[@] | Orthopedics | WOMAC | 2.8* | [1.7, 50.5] * |
| | Knee Osteoarthritis | | 2.7* | [1.7, 3.4] ** |
| Garcia et al. (2013) | Orthopedics | RMDQ | 4 | NR |
| | Chronic LBP | | | |
| Siemonsma et al. (2013) | Orthopedics<br><br>Chronic LBP | PSC | 4 | NR |
| Miyamoto et al. (2013) | Orthopedics<br><br>Chronic LBP | GPE | 4 | [2.0, 32.0] |
| Masaracchio et al. (2013) | Orthopedics<br><br>Mechanical Neck Pain | GROC | 2 | [1.0, 3.0] |
| Dunning et al. (2012) | Orthopedics | GROC | 1.8 | [1.4,  2.6] |
| | Mechanical Neck Pain | NDI | 2.3 | [1.7, 3.5] |
| Cleland et al. (2010) | Orthopedics | GROC | 15^ | [-4.6, 18.9] |
| | Neck Pain | | 6^^ | [-1.9, 34.8] |
| | | | 4^^^ | [2.1, 7.5] |

*Note:* NNT = Number Needed to Treat; OSS = *Oxford Shoulder Score*; WOMAC = *Western Ontario and McMaster Universities Osteoarthritis Index;* RMDQ = *Roland-Morris Disability Questionnaire*; NR = Not

reported; PSC = *Patient-Specific Complaints Questionnaire*; GPE = *Global Perceived Effect Scale*; GROC = *Global Rating of Change*; NDI = *Neck Disability Index.*
@ four groups *group 2; **group 3; ^ at 1 week; ^^ at 4 weeks; ^^^ at 6 months.

# CHAPTER 5 DISCUSSION

The purpose of this study was to examine the methods of reporting research results of intervention studies in the physical therapy literature. The reporting was considered regarding the number needed to treat (NNT) relative to null hypothesis statistical tests (NHST) and effect size (EST) in physical therapy clinical trials over time published in physical therapy specific journals. The methods used to report the results of a clinical trial impact the ability of physical therapists to translate the results into clinical importance and practice.

## Summary of Findings

The frequency of reporting the result of NHST, EST and NNT was examined in 448 clinical trials published in three physical therapy specific journals (*Journal of Neurologic Physical Therapy*, *Journal of Orthopaedic Physical Therapy* and *Physical Therapy*) between 1989 and 2018. More than 90% of clinical trials included a report of the result of NHST, ranging from 93.4% (*Physical Therapy*) to 97.2% (*Journal of Orthopaedic and Sports Physical Therapy*). The reporting of EST in the clinical trials was much less frequent than for NHST, ranging from 14.9% (*Journal of Orthopaedic Physical Therapy*) to 26.0% (*Journal of Neurologic Physical Therapy*). The reporting of NNT in clinical trials was non-existent in the sample prior to 2010. NNT was reported in eight clinical trial articles 2010 – 2018. None of these were published in the *Journal of Neurologic Physical Therapy.*

To determine if there was a change in reporting frequency of NHST, EST and/or NNT over time, the articles were combined into 5-year time span groups, enabling analysis of expected and actual counts of reporting for each 5-year period. There was no statistically significant difference in reporting of NHST for the full sample, nor for any of the individual journals. EST was reported more often than expected beginning with the

2008-2012 5-year period and remained high in 2013-2017 5-year period. NNT was reported more frequently than expected only during the last 5-year period in this study, 2013- 2017. In summary, NHST has remained the most frequently and consistently reported statistic in the clinical trials included in this study. An increase in reporting of EST did not occur until the fourth of the five 5-year periods, followed by an increase in reporting of NNT, albeit a small percent of all the included clinical trials (1.7%), in the most recent 5-year period, 2013-2017.

Translation of research into practice takes many years. Although a common lag time cited is 17 years (Morris, Wooding, & Grant, 2011), there were multiple influencing factors such as the adoption of new statistical methods. There were several events occurring from the time NNT was first introduced to when NNT was reported in the source journals. In Figure 6, the left side of the time line represents published articles where the use of NNT were introduced or encouraged since the introduction of NNT (Laupacis et al.) in 1988. Note the 12-year lag until the first article aimed at physical therapy was published in *Physical Therapy* in 2000 (Dalton & Keating). Three additional articles in physical therapy supporting the use of NNT in clinical trials (Campo & Lichtman, 2008; Hilton et al., 2006; Weeks & Noteboom, 2004) were published before the first article in which NNT was reported was published in Physical Therapy in 2010, 22 years after the introduction of NNT. The eight articles in which NNT was reported in the source journals are marked on the right side of the time line as are the points where EST and NNT reporting exceeded expectations.

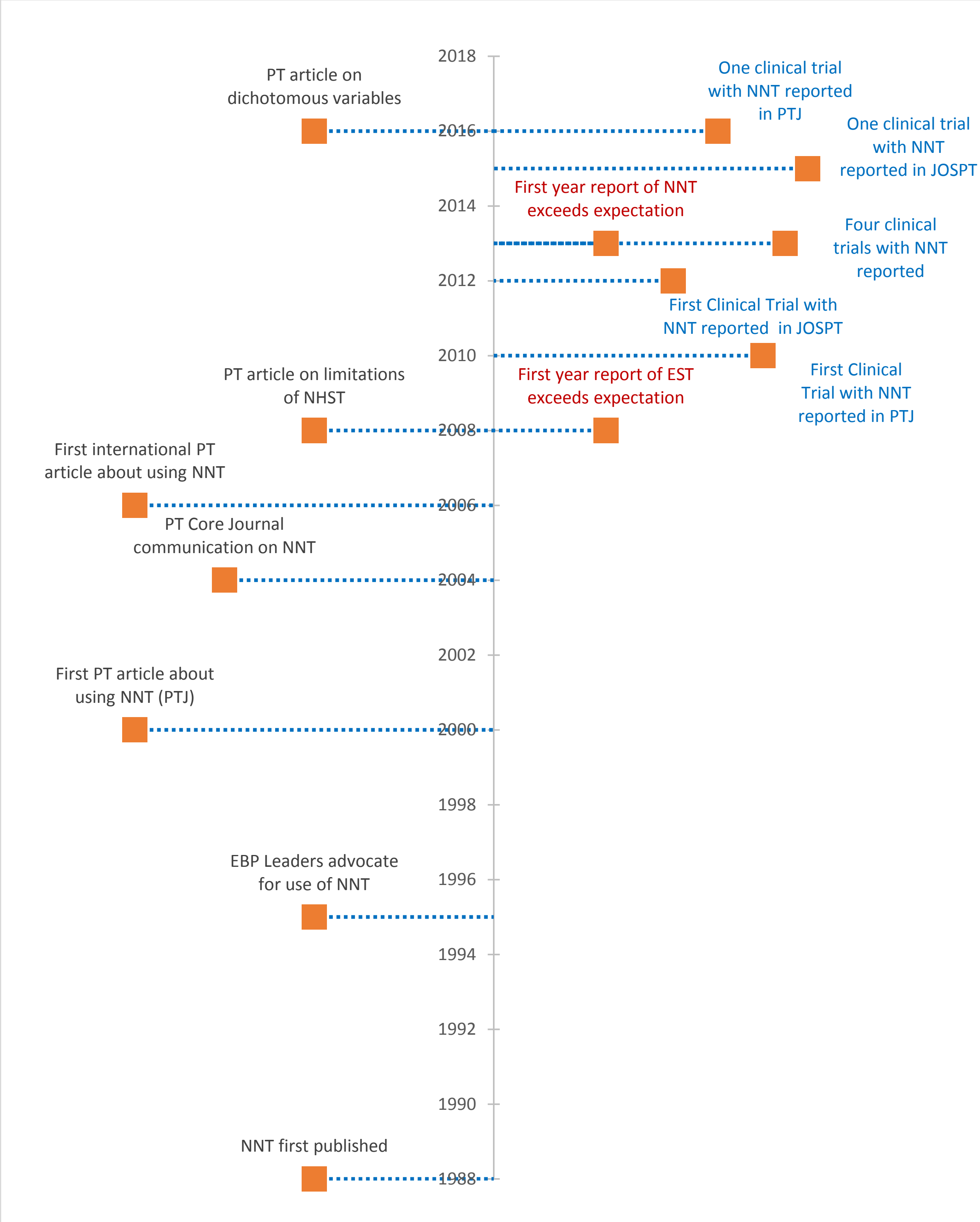


*Figure 7:* Number Needed to Treat (NNT) Timeline. *PTJ = Physical Therapy*; *JOSPT = Journal of Orthopaedic and Sports Physical Therapy*; EST = Effect Size.

## Interpretation of Findings

The reporting of NNT in articles published in physical therapy specific journals increased. However, the number of articles is a very small proportion (n = 8, 1.79%) of the 448 clinical trials published since the introduction of NNT and all were published only in the past 9 years. The correlation between the number of NNT per year during the nine-year period from 2010 to 2018 with the number of clinical trials reported was not statistically significant, $r = 0.18$, $p = 0.964$. The Mann-Kendall test for linear trend was conducted on the total number of NNT and clinical trials. Although there was a statistically significant increase in linear trend (tau = 0.684, $p = 0.016$) of the number of clinical trials reported during these nine years, there was no similar increasing trend in the number of NNT (tau = 0.272, $p = 0.354$). Although the number of articles in which NNT was reported increased during the most recent 5- year period included in this study there is no evidence of a positive linear trend during the past nine years. The number of clinical trials published in the source journals did increase but there was no concomitant increase in NNT. This may indicate that there was only a short-lived time period of increased NNT which may be attributed to two authors (Cleland, Costa) who were authors of five of the eight articles in which NNT was reported.

Interestingly, EST reporting increased above expectations during the 5-year period immediately preceding the 5-year period in which NNT increased above expectations. Editorial boards for many journals, including the source journals in this study, have increasingly required the reporting of effect size. NNT is one of over 40 types of effect size measures. Increased use of EST may have influenced an increased awareness of the many effect size measures, including NNT. However, it was beyond the scope of this study to identify any causal relationships.

It was not clear, across multiple disciplines, how often NNT was calculated and reported in the literature. It was stated that awareness and use of NNT has been increasing (Cook & Sackett, 1995; Weeks & Noteboom, 2004). Conversely, others have reported that NNT is not widely used or is underused (Hilton et al., 2006; Nuovo et al., 2002; Pinson & Gray, 2003). The results of this study provided the first evidence that the use of NNT is increasing in the physical therapy specific literature, but it does not represent a positive linear trend 2010 - 2018

**Contextual Considerations**

The computation of NNT requires that the outcome variable be dichotomous. Outcome measures reported in physical therapy clinical trials reflect all levels of measurement, continuous, ordinal and nominal, including dichotomous variables. Dichotomous outcome variables may occur naturally or may be derived from continuous or ordinal measurement scales. Many outcome measures used in physical therapist practice are naturally continuous (e.g. time, distance, repetitions, degrees of movement) or ordinal (manual muscle tests, 11-point pain scales, fear of falling ratings) levels of measurement. Consequently, conversion into a dichotomous variable would be required in order compute NNT.

Some physical therapy outcome measures which use continuous/ordinal scales have an identified cut score. The cut score is used to define a positive or negative outcome. For example, *Timed Up and Go* (*TUG*) is a timed test that involves rising from a chair, walking 3 meters, turning around, walking back to the chair and sitting down

(Shumway-Cook, Brauer, & Woollacott, 2000). The cut score for fall risk in community dwelling older adults is 13.5s. Although typically recorded as time to complete, a dichotomous variable, at risk for falls, could be derived using scores above 13.5s representing at risk for falls and other scores representing no increased risk for falls. Cut scores can be similarly derived for many physical therapy outcome measures.

The Global Rating of Change (GROC) is a self-report outcome measure of perceived change in a health condition over time and was used in three of the eight articles in this study in which NNT was reported. Dunning et al. (2012) stated that

> we dichotomized patients as having experienced as having a successful outcome using...greater than or equal to +4 on the GRC (Cleland, Glynn, et al., 2007). It has been reported that scores of +4 are indicative of moderate changes in patient status and have been previously used as a measure of success in clinical research (Cleland, Childs, Fritz, Whitman, & Eberhart, 2007; Whitman et al., 2009). (p. 10)

Unfortunately, in a different study Cleland et al. (2010) used +5 (also indicative of moderate change) on the GROC to define a successful outcome (Jaeschke et al., 1989). Using a different cut score to dichotomize an outcome variable invalidates the ability to compare NNT across studies. Another method that has been used to identify a cut point to dichotomize a continuous or ordinal level outcome measure is to use the Minimal Clinically Important Difference (MCID) sometimes referred to as the Minimal Clinically Important Change (MCIC). For example, the MCIC was determined to be a score of 6 for the Oxford Shoulder Score (Christiansen et al., 2015; van Kampen et al., 2013). Christiansen et al. (2016) used this MCIC to define the cut score to define a successful outcome. Although this intuitively makes sense, Siemonsma et al. (2013), after using clinically relevant change to define a successful outcome, stated:

> The best method to define and determine a clinically relevant change, however, is under debate. (Frost, Lamb, & Stewart-Brown, 2008) Fundamental statistical

> issues currently cloud the precise estimation of clinically relevant changes in general. (Terwee et al., 2010) Therefore, some caution in the interpretation of our results is warranted. (p. 444)

**Implications of Findings**

The findings of this study may advance research methodology by increasing awareness and understanding of the usefulness of NNY in translating research findings into clinical practice. Although the frequency of authors reporting NNT in clinical trials published in physical therapy specific journals has increased recently it is not known if physical therapist practitioners have the knowledge of how to use the NNT when making clinical decisions when developing a plan of care, specifically when selecting an intervention for a specific patient/client. As previously referenced, Dalton and Keating (2000, p. 1216) stated “NNT provides results in a way that is directly transferable to the clinical setting” which reinforces its role in clinical decision making as well as helping patients/clients make better informed consent decisions.

**Limitations**

There are several limitations of this study which should be considered. The first relate to sample selection which was limited to three physical therapy specific source journals (*Journal of Neurologic Physical Therapy*, *Journal of Orthopaedic Physical Therapy* and *Physical Therapy*). These journals of the American Physical Therapy Association were selected because they are highly regarded, large circulation and readily available core physical therapy journals. Orthopedics and neurology are primary practice areas in physical therapy. However, there are other physical therapy specific journals both in the United States and internationally that could have been source journals. Similarly, physical therapy clinical trials may have been identified by using the

*Physiotherapy Evidence Database* (*PEDro*)(The PEDro Partnership, 2019). The PEDro database includes a wide variety of physical therapy core journals, many of which are not physical therapy specific. Authors of physical therapy clinical trials may choose to submit to journals that are not specific to physical therapy for many reasons. Varying the sampling strategy may have resulted in different outcomes and limits the generalizability of the results.

Although the post hoc power for this study was 0.99, the total number of articles in which NNT was reported very small and limited to the orthopedic physical therapy patient/client population. The small number of articles in which NNT is reported limits any broader interpretation of the data such as identifying factors that may increase the likelihood of NNT being reported in a study. The frequency of articles published with NNT reported did not become statistically greater than expected until the 5$^{th}$ of the five 5-year periods considered.

The increased reporting of NNT does not imply appropriate nor accurate use. Methods for calculating bracketed intervals for NNT have been suggested or recommended by some authors (Cook & Sackett, 1995; Laupacis et al., 1988; Weeks & Noteboom, 2004). However, the use of bracketed intervals in NNT is not well supported and there is not agreement on how to, or if it is meaningful, to do so (Hancock & Kent, 2016; Julious, 2005). Despite this, 6 of the 8 (75%) studies included 95%CI for the computed NNT (Table 18). Cleland et al. (2010) reported NNT even though the results of the NHST at one-week post intervention were non-significant. As stated previously, Sawilowsky and Yoon (2002) and Sawilowsky (2003) noted it is futile to discuss effectiveness of a given treatment if the null-hypothesis is found to be non-significant (see also Sawilowsky (2007); Sawilowsky et al., 2011).

**Future Research**

There are many directions for future research related to the use of NNT in physical therapy clinical trials. Future studies should incorporate different sampling strategies, expanding to other physical therapy specific journals nationally or internationally or to other core physical therapy journals accessed through a database such as PEDro. Future research should explore the knowledge/utilization of various methods of reporting effect size of various stakeholders including student physical therapists, physical therapist educators in both entry-level and post-professional programs, physical therapist researchers and physical therapist practitioners to barriers to the implementation of NNT in clinical trials. It is important to utilize one or more common outcome measure(s) for specific or similar diagnoses such as the GROC which was used in each of the 3 clinical trials for neck pain. Future research should focus on developing a consistent set of outcome measures, exploring valid and reliable methods of dichotomizing continuous or ordinal outcome variables.

There were no articles in the sample which used NNT to identify the risk for adverse effects in physical therapist interventions, only for what would be considered positive outcomes. Newman and Allison (2007) wrote an editorial in *Journal of Orthopaedic and Sports Physical Therapy* encouraging researchers to investigate risk using measures including absolute risk reduction, relative risk reduction, NNT and number needed to harm. Although physical therapy interventions tend to be conservative, understanding risk is essential for risk management.

## Conclusion

In the 30 years since NNT was introduced by Laupacis et al. (1988) it has only recently been included in the results of clinical trials published in two of the three source journals (*Journal of Orthopaedic Physical Therapy* and *Physical Therapy*) in this study. This study is the first to report increased reporting of NNT in the physical therapy specific literature but there is no evidence of a positive trend during the past nine years. Stakeholders, including physical therapist students, educators, researchers and practitioners would be well served to improve their understanding of how to include NNT in clinical trial research designs, for making clinical decisions about the physical therapist plan of care including the selection of interventions and to explain intervention selection and effectiveness at the level of patient/client numbers to referring healthcare practitioners and patients. It is recommended that this process include the development of an agreed upon set dichotomous outcome variables used consistently across studies of similar health conditions that would result in easier translation of research results into physical therapist clinical practice. Additionally, NNT cannot be interpreted in isolation. Generalizability and importance of the results of this study need to be considered by the physical therapist as the subject matter expert.

# APPENDIX A

Bibliometric Variables: Article Publication Information

| Variable Name (Level) | Variable | Values |
|---|---|---|
| Article_No (Nominal) | Article Number | N/A |
| Article_Yr (Nominal) | Year Published | N/A |
| Article_Vol (Nominal) | Volume Number Published | N/A |
| Article_Iss (Nominal) | Issue Published | N/A |
| Article_Mo (Nominal) | Month Published | N/A |
| Article_Cit (Scale) | Number of times article cited | N/A |

## APPENDIX B

Bibliometric Variables: Author and Institution Information

| Variable Name (Level) | Variable | Values |
|---|---|---|
| 1stAuthor_L_N (Nominal) | 1st Author Last Name | N/A |
| 1st Author F_N (Nominal) | 1st Author First Name | N/A |
| 1st_Author_Inst (Nominal) | Name of home institution of the first author | N/A |
| 1st_Author_Country (Nominal) | Country of 1st author's home institution | N/A |
| 1st_Author_Inst_Type (Nominal) | Type of institution of 1st author | 1. Education<br>2. Research<br>3. Health Provider |

**APPENDIX B** (continued)

Bibliometric Variables: Author and Institution Information

| Variable Name (Level) | Variable | Values |
|---|---|---|
| 1st_Author_Rank (Nominal) | Academic Rank of 1st Author | 1. Assist Prof TT<br>2. Assist Prof non-TT<br>3. Assist Prof research<br>4. Assoc Prof tenured<br>5. Assoc Prof non-TT<br>6. Assoc Prof research<br>7. Prof tenured<br>8. Prof non-TT<br>9. Prof research<br>10. Lecturer/Instructor<br>11. Post-doc<br>12. Other<br>13. Not applicable |
| 1st_Author_PT (Nominal) | Is first author a physical therapist? | 1. Yes<br>2. No |

**APPENDIX B** (continued)

Bibliometric Variables: Author and Institution Information

| Variable Name (Level) | Variable | Values |
|---|---|---|
| 1st_Author_Prof (Nominal) | What is the first author's profession? | 1. Physical Therapist<br>2. Medical Doctor<br>3. Doctor of Osteopathy<br>4. Biologic Sciences<br>5. Behavioral/Social Sciences<br>6. Other |
| Corr Author_L_N (Nominal) | Corresponding Author (if different from 1st author) Last Name | N/A |
| Corr Author F_N (Nominal) | Corresponding author (if different from 1st author) First Name | N/A |
| Corr Author_Inst (Nominal) | Corresponding author (if different from 1st author) Name of home institution | N/A |
| Corr_Author_Country (Nominal) | Corresponding author (if different from 1st author) country of home institution | N/A |

**APPENDIX B** (continued)

Bibliometric Variables: Author and Institution Information

| Variable Name (Level) | Variable | Values |
|---|---|---|
| Corr_Author_Inst_Type (Nominal) | Corresponding author (if different from 1st author) type of institution | 1. Education (e.g. university, college)<br>2. Research<br>3. Health Provider (e.g. hospital, clinic) |
| Corr_Author_Inst_Carn (Nominal) | Corresponding author (if different from 1st author) Carnegie classification if education institution | 1. R1 Doctoral University - highest<br>2. R2 Doctoral University - higher<br>3. R3 Doctoral University - moderate<br>4. M1 Master's College larger<br>5. M2 Master's College medium<br>6. M3 Master's College smaller<br>7. Special Focus<br>8. Not Applicable |

**APPENDIX B** (continued)

Bibliometric Variables: Author and Institution Information

| Variable Name (Level) | Variable | Values |
|---|---|---|
| Corr_Author_Dept (Nominal) | Corresponding author (if different from 1st author) home unit (department, program) | N/A |
| Corr _Author_College (Nominal) | Corresponding author (if different from 1st author) name of unit that the dept is housed in (e.g. college, school). | N/A |

**APPENDIX B** (continued)

Bibliometric Variables: Author and Institution Information

| Variable Name (Level) | Variable | Values |
|---|---|---|
| Corr_Author_Rank (Nominal) | Corresponding author (if different from 1st author) academic rank | 1. Assist Prof TT<br>2. Assist Prof non-TT<br>3. Assist Prof research<br>4. Assoc Prof tenured<br>5. Assoc Prof non-TT<br>6. Assoc Prof research<br>7. Prof tenured<br>8. Prof non-TT<br>9. Prof research<br>10. Lecturer/Instructor<br>11. Post-doc<br>12. Other<br>13. Not applicable |
| Corr_Author_PT (Nominal) | Is corresponding author (if different from 1st author) a physical therapist? | 1. Yes<br>2. No |

**APPENDIX B** (continued)

Bibliometric Variables: Author and Institution Information

| Variable Name (Level) | Variable | Values |
|---|---|---|
| Corr_Author_Prof (Nominal) | Corresponding author (if different from 1st author) profession? | 1. Physical Therapist<br>2. Medical Doctor<br>3. Doctor of Osteopathy<br>4. Biologic Sciences<br>5. Behavioral/Social Sciences<br>6. Other |
| N_Authors (Scale) | Number of authors | N/A |
| N_Inst (Scale) | Number of unique institutions represented by authors | N/A |
| N_Dept (Scale) | Number of unique departments represented by authors | |
| N_Countries (Scale) | Number of unique countries represented by institutions | N/A |

**APPENDIX B** (continued)

Bibliometric Variables: Author and Institution Information

| Variable Name (Level) | Variable | Values |
|---|---|---|
| N_Educ_Inst (Scale) | Number of unique educational institutions represented | N/A |
| N_Res_Inst (Scale) | Number of unique research institutions represented | N/A |
| N_Health_Inst (Scale) | Number of unique health institutions represented | N/A |

# APPENDIX C

Bibliometric Variables: Research Design and Data Analysis

| Variable Name (Level) | Variable | Values |
|---|---|---|
| Research Design | Is there randomized assignment to groups | 1. Randomized<br>2. Non-Randomized |
| N_Grps (Scale) | Number of Groups | N/A |
| Sample_Size (Scale) | Number of participants | N/A |
| Control_Group (Nominal) | Is there a control group? | 1. Yes<br>2. No |
| Control_Size (Scale) | Number participants in control group | N/A |
| TxGrp_Size (Scale) | Number participants in treatment group 1 | N/A |
| TxGrp2_Size (scale) | Number participants in treatment group 2 | N/A |
| TxGrp2_Size (Scale) | Number participants in treatment group 2 | N/A |
| TxGrp3_Size (Scale) | Number participants in treatment group 3 | N/A |

**APPENDIX C** (continued)

Bibliometric Variables: Research Design and Data Analysis

| Variable Name (Level) | Variable | Values |
|---|---|---|
| Pt_Pop (Nominal) | What is the patient condition category | 1. Cardiovascular<br>2. Lymphatic<br>3. Oncology<br>4. Orthopedics<br>5. Neurology<br>6. Pulmonary<br>7. Sports<br>8. Women's health/pelvic floor<br>9. 0ther |
| Sample Age Cent Tend Measure (Nominal) | How is central tendency measured? | 1. Mean<br>2. Median<br>3. Mode |
| Sample Age Cent Tend (Scale/Ordinal) | Reported central tendency of sample age | N/A |
| Sample Age Dispersion Measure (Nominal) | How is dispersion of the sample age reported | 1. SD<br>2. Range |

**APPENDIX C** (continued)

Bibliometric Variables: Research Design and Data Analysis

| Variable Name (Level) | Variable | Values |
|---|---|---|
| Sample Age dispersion (Scale/Ordinal) | Reported dispersion of sample age | N/A |
| NHST (Nominal) | Is a NHST reported? | 1. Yes<br>2. No |
| NHST_Type (Nominal) | What category of NHST was reported | 1. Independent t test<br>2. Paired t test – equal variance<br>3. F Test |
| Alpha (Scale) | What is the reported nominal alpha? | N/A |
| Stat_Sig (Nominal) | Was the result reported as statistically significant? | 1. Yes<br>2. No |
| EST (Nominal) | Is an effect size reported? | 1. Yes<br>2. No |

**APPENDIX C** (continued)

Bibliometric Variables: Research Design and Data Analysis

| Variable Name (Level) | Variable | Values |
|---|---|---|
| EST_Type (Nominal) | What type of EST was computed? | 1. Absolute mean difference<br>2. Cohen's d<br>3. Hedge's g<br>4. Glass's Δ<br>5. Cohen's f<br>6. Eta<br>7. Partial Eta<br>8. Omega<br>9. Partial Omega<br>10. Other |
| EST_Mag (Scale) | Magnitude of the effect size reported? | N/A |
| EST_CI (Nominal) | Is a confidence interval reported for the ES? | 1. Yes<br>2. No |
| EST_CI_Level (Scale) | What is the confidence level for the ES CI? | N/A |
| NNT (Nominal) | Is number needed to treat reported? | 1. Yes<br>2. No |

**APPENDIX C** (continued)

Bibliometric Variables: Research Design and Data Analysis

| Variable Name (Level) | Variable | Values |
|---|---|---|
| NNT_Mag (Scale) | Magnitude of the number needed to treat | N/A |
| NNT_CI (Nominal) | Is a confidence interval reported for NNT? | 1. Yes<br>2. No |
| NNT_CI_Level (Scale) | What is the confidence level for the NNT CI? | N/A |
| Dichot (Nominal) | Is there a dichotomous outcome variable? | 1. Yes<br>2. No |

## APPENDIX D

### Eight Articles in which NNT was Reported

Abbott, J. H., Chapple, C. M., Fitzgerald, G. K., Fritz, J. M., Childs, J. D., Harcombe, H., & Stout, K. (2015). The incremental effects of manual therapy or booster sessions in addition to exercise therapy for knee osteoarthritis: A randomized clinical trial. *Journal of Orthopaedic and Sports Physical Therapy, 45*(12), 975-983. doi:10.2519/jospt.2015.6015

Christiansen, D. H., Frost, P., Falla, D., Haahr, J. P., Frich, L. H., Andrea, L. C., & Svendsen, S. W. (2016). Effectiveness of standardized physical therapy exercises for patients with difficulty returning to usual activities after decompression surgery for subacromial impingement syndrome: Randomized controlled trial. *Physical therapy, 96*(6), 787-796. doi:10.2522/ptj.20150652

Cleland, J. A., Mintken, P. E., Carpenter, K., Fritz, J. M., Glynn, P., Whitman, J., & Childs, J. D. (2010). Examination of a Clinical Prediction Rule to Identify Patients With Neck Pain Likely to Benefit From Thoracic Spine Thrust Manipulation and a General Cervical Range of Motion Exercise: Multi-Center Randomized Clinical Trial. *Physical therapy, 90*(9), 1239-1250. doi:10.2522/ptj.20100123

Dunning, J. R., Cleland, J. A., Waldrop, M. A., Arnot, C., Young, I., Turner, M., & Sigurdsson, G. (2012). Upper cervical and upper thoracic thrust manipulation versus nonthrust mobilization in patients with mechanical neck pain: A multicenter randomized clinical trial. *Journal of Orthopaedic and Sports Physical Therapy, 42*(1), 5-18. doi:10.2519/jospt.2012.3894

Garcia, A. N., Lucíola da Cunha Menezes, C., Tatiane Mota da, S., Francine Lopes Barreto, G., Cyrillo, F. N., Costa, R. A., & Leonardo Oliveira Pena, C. (2013). Effectiveness of Back School Versus McKenzie Exercises in Patients With Chronic Nonspecific Low Back Pain: A Randomized Controlled Trial. *Physical therapy, 93*(6), 729-747. doi:10.2522/ptj.20120414

Masaracchio, M., Cleland, J., Hellman, M., & Hagins, M. (2013). Short-term combined effects of thoracic spine thrust manipulation and cervical spine nonthrust manipulation in individuals with mechanical neck pain: A randomized clinical trial. *Journal of Orthopaedic and Sports Physical Therapy, 43*(3), 118-127. doi:10.2519/jospt.2013.4221

Miyamoto, G. C., Leonardo Oliveira Pena, C., Galvanin, T., & Cristina Maria Nunes, C. (2013). Efficacy of the Addition of Modified Pilates Exercises to a Minimal Intervention in Patients With Chronic Low Back Pain: A Randomized Controlled Trial. *Physical therapy, 93*(3), 310-320. doi:10.2522/ptj.20120190

Siemonsma, P. C., Stuive, I., Roorda, L. D., Vollebregt, J. A., Walker, M. F., Lankhorst, G. J., & Lettinga, A. T. (2013). Cognitive Treatment of Illness Perceptions in Patients With Chronic Low Back Pain: A Randomized Controlled Trial. *Physical therapy, 93*(4), 435-446. doi:10.2522/ptj.20110150

# REFERENCES

Abbott, J. H., Chapple, C. M., Fitzgerald, G. K., Fritz, J. M., Childs, J. D., Harcombe, H., & Stout, K. (2015). The incremental effects of manual therapy or booster sessions in addition to exercise therapy for knee osteoarthritis: A randomized clinical trial. *Journal of Orthopaedic and Sports Physical Therapy, 45*(12), 975-983. doi:10.2519/jospt.2015.6015

Babu, A. S., Veluswamy, S. K., Rao, P. T., & Maiya, A. G. (2014). Clinical trial registration in physical therapy journals: a cross-sectional study. *Phys Ther, 94*(1), 83-90. doi:10.2522/ptj.20120531

Bandy, W. D. (2003). Use of Statistics in Physical Therapy Over a 2-Year Period-2000-2002: Implications for Educators. *Journal of Physical Therapy Education, 17*(2), 67.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., . . . Simel, D. (1996). Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA, 276*(8), 637-639.

Bohannon, R. W. (1999). Core Journals of Physiotherapy. *Physiotherapy, 85*(6), 317-321. doi:10.1016/S0031-9406(05)67135-0

Bohannon, R. W., & Gibson, D. F. (1986). Citation analysis of Physical Therapy. A special communication. *Phys Ther, 66*(4), 540-541.

Bohannon, R. W., & Leveau, B. F. (1986). Clinician Use of Research Findings - a Review of Literature with Implications for Physical-Therapists. *Physical therapy, 66*(1), 45-50.

Bohannon, R. W., & Roberts, D. (1991). Core journals of rehabilitation: identification through index analysis. *Int J Rehabil Res, 14*(4), 333-336.

Bohannon, R. W., & Tiberio, D. (1989). Physiotherapy literature in medical indexes: How comprehensive is index coverage of journals cited frequently by five physiotherapy journals? *Physiotherapy Practice, 5*(4), 201-205.

Borgman, C. L. (1989). Bibliometrics and Scholarly Communication:Editor's Introduction. *Communication Research, 16*(5), 583-599. doi:10.1177/009365089016005002

Bradford, S. C., Egan, M. E., & Shera, J. H. (1953). Documentation. In (2 ed.). London, UK: Lockwood.

Broadus, R. N. (1987a). Early Approaches to Bibliometrics. *Journal of the American Society for Information Science (1986-1998), 38*(2), 127-129. doi:10.1002/(SICI)1097-4571(198703)38:2<127::AID-ASI6>3.0.CO;2-K

Broadus, R. N. (1987b). Toward a definition of “bibliometrics”. *Scientometrics, 12*(5-6), 373-379.

Campo, M., Eckardt, P., Findley, J. L., Cardinale, J. A., & Shiyko, M. P. (2017). A Practical Guide to Teaching Measures of Effect Size in Physical Therapist Education. *Journal of Physical Therapy Education, 31*(2), 67.

Campo, M., & Lichtman, S. W. (2008). Interpretation of Research in Physical Therapy: Limitations of Null Hypothesis Significance Testing. *Journal of Physical Therapy Education, 22*(1), 43.

Carver, R. (1978). The Case Against Statistical Significance Testing. *Harvard Educational Review, 48*(3), 378-399. doi:10.17763/haer.48.3.t490261645281841

Carver, R. (1993). The Case against Statistical Significance Testing, Revisited. *The Journal of Experimental Education, 61*(4), 287-292. doi:10.1080/00220973.1993.10806591

Christiansen, D. H., Frost, P., Falla, D., Haahr, J. P., Frich, L. H., Andrea, L. C., & Svendsen, S. W. (2016). Effectiveness of standardized physical therapy exercises for patients with difficulty returning to usual activities after decompression surgery for subacromial impingement syndrome: Randomized controlled trial. *Physical therapy, 96*(6), 787-796. doi:10.2522/ptj.20150652

Christiansen, D. H., Frost, P., Falla, D., Haahr, J. P., Frich, L. H., & Svendsen, S. W. (2015). Responsiveness and Minimal Clinically Important Change: A Comparison Between 2 Shoulder Outcome Measures. *The Journal of orthopaedic and sports physical therapy, 45*(8), 620-625. doi:10.2519/jospt.2015.5760

Clarivate, A., & Institute for Scientific, I. (1997). Journal citation reports. In. Philadelphia, Pa.;New York, NY;London; U6 - ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rfr_id=info%3Asid%2Fsummon.serialssolutions.com&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.atitle=Journal+citation+reports&rft.date=1997-01-01&rft.pub=Institute+for+Scientific+Information&rft.externalDocID=004107284&paramdict=en-US U7 - Web Resource: Institute for Scientific Information.

Cleland, J. A., Childs, J. D., Fritz, J. M., Whitman, J. M., & Eberhart, S. L. (2007). Development of a Clinical Prediction Rule for Guiding Treatment of a Subgroup of

Patients With Neck Pain: Use of Thoracic Spine Manipulation, Exercise, and Patient Education. *Physical therapy, 87*(1), 9-23. doi:10.2522/ptj.20060155

Cleland, J. A., Glynn, P., Whitman, J. M., Eberhart, S. L., MacDonald, C., & Childs, J. D. (2007). Short-Term Effects of Thrust Versus Nonthrust Mobilization/Manipulation Directed at the Thoracic Spine in Patients With Neck Pain: A Randomized Clinical Trial. *Physical therapy, 87*(4), 431-440. doi:10.2522/ptj.20060217

Cleland, J. A., Mintken, P. E., Carpenter, K., Fritz, J. M., Glynn, P., Whitman, J., & Childs, J. D. (2010). Examination of a Clinical Prediction Rule to Identify Patients With Neck Pain Likely to Benefit From Thoracic Spine Thrust Manipulation and a General Cervical Range of Motion Exercise: Multi-Center Randomized Clinical Trial. *Physical therapy, 90*(9), 1239-1250. doi:10.2522/ptj.20100123

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J: L. Erlbaum Associates.

Cohen, J. (1994). The earth is round (p<. 05). *American psychologist, 49*(12), 997.

Cohen, J. (1995). The earth is round (p < 05): Rejoinder. *American psychologist, 50*(12), 1103-1103. doi:10.1037/0003-066X.50.12.1103

Compton, S., & Sawilowsky, S. (2003). Do not discourage the use of P values. *Annals of Emergency Medicine, 41*(4), 584-584. doi:10.1067/mem.2003.130

Cook, R. J., & Sackett, D. L. (1995). The number needed to treat: a clinically useful measure of treatment effect. *BMJ, 310*(6977), 452-454.

Coronado, R. A., Riddle, D. L., Wurtzel, W. A., & George, S. Z. (2011). Bibliometric analysis of articles published from 1980 to 2009 in Physical Therapy, journal of the

American Physical Therapy Association. *Phys Ther, 91*(5), 642-655. doi:10.2522/ptj.20100267

Cortina, J. M., & Dunlap, W. P. (1997). On the logic and purpose of significance testing. *Psychological Methods, 2*(2), 161.

Cortina, J. M., & Landis, R. S. (2011). The Earth Is Not Round (p = .00). *Organizational Research Methods, 14*(2), 332-349. doi:10.1177/1094428110391542

Costa, L. O. P., Moseley, A. M., Sherrington, C., Maher, C. G., Herbert, R. D., & Elkins, M. R. (2010). Core Journals That Publish Clinical Trials of Physical Therapy Interventions. *Physical therapy, 90*(11), 1631-1640. doi:10.2522/ptj.20090419

Cox, V. (2017). Exploratory data analysis. In *Translating Statistics to Make Decisions* (pp. 47-74): Springer.

Dalton, G. W., & Keating, J. L. (2000). Number needed to treat: a statistic relevant for physical therapists. *Phys Ther, 80*(12), 1214-1219.

de Glas, F. (1986). Fiction and Bibliometrics: Analyzing a Publishing House's Stocklist. *Libri, 36*, 40.

Dean, E., & Davies, J. (1986). Frequency of citation and reputational assessment of contributors in physical therapy. *Phys Ther, 66*(6), 961-966.

Dunning, J. R., Cleland, J. A., Waldrop, M. A., Arnot, C., Young, I., Turner, M., & Sigurdsson, G. (2012). Upper cervical and upper thoracic thrust manipulation versus nonthrust mobilization in patients with mechanical neck pain: A multicenter randomized clinical trial. *Journal of Orthopaedic and Sports Physical Therapy, 42*(1), 5-18. doi:10.2519/jospt.2012.3894

Falk, R., & Greenbaum, C. W. (1995). Significance Tests Die Hard: The Amazing Persistence of a Probabilistic Misconception. *Theory & Psychology, 5*(1), 75-98. doi:10.1177/0959354395051004

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191.

Fell, D. W., Britton, R., Burnham, J., Dickens, N., Josey, M., Peterson, T., & Wamsley, J. (2015). Bibliometric analysis of the journal of neurologic physical therapy: 1993–2013. *Physiotherapy, 101*, e378. doi:https://doi.org/10.1016/j.physio.2015.03.597

Fell, D. W., Burnham, J. F., Buchanan, M. J., Horchen, H. A., & Scherr, J. A. (2011). Mapping the core journals of the physical therapy literature. *J Med Libr Assoc, 99*(3), 202-207. doi:10.3163/1536-5050.99.3.007

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532.

Franchignoni, F., & Munoz Lasa, S. (2011). Bibliometric indicators and core journals in physical and rehabilitation medicine. *J Rehabil Med, 43*(6), 471-476. doi:10.2340/16501977-0821

Frick, R. W. (1996). The Appropriate Use of Null Hypothesis Testing. *Psychological Methods, 1*(4), 379-390. doi:10.1037/1082-989X.1.4.379

Fritz, J. M., Delitto, A., & Erhard, R. E. (2003). Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. *Spine, 28*(13), 1363-1371.

Frost, H., Lamb, S. E., & Stewart-Brown, S. (2008). Responsiveness of a patient specific outcome measure compared with the Oswestry Disability Index v2.1 and Roland and Morris Disability Questionnaire for patients with subacute and chronic low back pain. *Spine, 33*(22), 2450-2457. doi:10.1097/BRS.0b013e31818916fd

Garcia, A. N., Lucíola da Cunha Menezes, C., Tatiane Mota da, S., Francine Lopes Barreto, G., Cyrillo, F. N., Costa, R. A., & Leonardo Oliveira Pena, C. (2013). Effectiveness of Back School Versus McKenzie Exercises in Patients With Chronic Nonspecific Low Back Pain: A Randomized Controlled Trial. *Physical therapy, 93*(6), 729-747. doi:10.2522/ptj.20120414

Gore, S. A., Nordberg, J. M., Palmer, L. A., & Piorun, M. E. (2009). Trends in health sciences library and information science research: an analysis of research publications in the Bulletin of the Medical Library Association and Journal of the Medical Library Association from 1991 to 2007. *J Med Libr Assoc, 97*(3), 203-211. doi:10.3163/1536-5050.97.3.009

Green, S. B., & Salkind, N. J. (2004). *Using SPSS for Windows and Macintosh: Analyzing and understanding* (4 ed.). Upper Saddle Hill, NJ: Prentice Hall.

Gross, J. H. (2015). Testing What Matters (If You Must Test at All): A Context-Driven Approach to Substantive and Statistical Significance. *American Journal of Political Science, 59*(3), 775-788. doi:10.1111/ajps.12149

Hagen, R. L. (1997). In Praise of the Null Hypothesis Statistical Test. *American psychologist, 52*(1), 15-24. doi:10.1037/0003-066X.52.1.15

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers. *Methods of Psychological Research, 7*(1), 1-20.

Hancock, M., & Kent, P. (2016). Interpretation of dichotomous outcomes: risk, odds, risk ratios, odds ratios and number needed to treat. *J Physiother, 62*(3), 172-174. doi:10.1016/j.jphys.2016.02.016

Hays, W. L. (1981). Statistics New York: Holt. *Rinehart and Winston*.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92*(2), 490-499. doi:10.1037/0033-2909.92.2.490

Herbert, R. D. (2000). How to estimate treatment effects from reports of clinical trials. II: Dichotomous outcomes. *Aust J Physiother, 46*(4), 309-313.

Hertzel, D. H. (1987). History of the development of ideas in bibliometrics. In A. Kent (Ed.), *Encyclopedia of Library and information sciences* (Vol. 42, pp. 144-219). New York, NY: Marcel Dekker.

Hilton, D. J., Reid, C. M., & Paratz, J. (2006). An under-used yet easily understood statistic: the number needed to treat (NNT). *Physiotherapy, 92*(4), 240-246. doi:10.1016/j.physio.2006.06.003

Hoderlein, X., Moseley, A. M., & Elkins, M. R. (2017). Citation of prior research has increased in introduction and discussion sections with time: A survey of clinical trials in physiotherapy. *Clinical Trials, 14*(4), 372-380. doi:10.1177/1740774517699821

Hood, W. W., & Wilson, C. S. (2001). The Literature of Bibliometrics, Scientometrics, and Informetrics. *Scientometrics, 52*(2), 291-314. doi:10.1023/A:1017919924342

Hubbard, R., & Lindsay, R. M. (2008). Why P Values Are Not a Useful Measure of Evidence in Statistical Significance Testing. *Theory & Psychology, 18*(1), 69-88. doi:10.1177/0959354307086923

Human Resources Research Organization. (2017). Analysis of Practice for the Physical Therapy Profession: Entry-Level Physical Therapists.

Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials, 10*(4), 407-415. doi:10.1016/0197-2456(89)90005-6

JASP Team. (2018). JASP (Version 0.9) [computer softfware]. Retrieved from https://jasp-stats.org/

Julious, S. A. (2005). Issues with number needed to treat. *Statistics in Medicine, 24*(20), 3233-3235. doi:10.1002/sim.2150

Kirk, R. E. (1996). Practical Significance: A Concept Whose Time Has Come. *Educational and Psychological Measurement, 56*(5), 746-759. doi:10.1177/0013164496056005002

Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences*: American Psychological Association.

Knapp, T. R., & Sawilowsky, S. S. (2001). Constructive criticisms of methodological and editorial practices. *The Journal of Experimental Education, 70*(1), 65.

Laupacis, A., Sackett, D. L., & Roberts, R. S. (1988). An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med, 318*(26), 1728-1733. doi:10.1056/NEJM198806303182605

MacDermid, J. C., Fung, E. H., & Law, M. (2015). Bibliometric Analyses of Physical and Occupational Therapy Faculty across Canada Indicate Productivity and Impact of Rehabilitation Research. *Physiotherapy Canada. Physiotherapie Canada, 67*(1), 76-84. doi:10.3138/ptc.2013-71BC

MacDermid, J. C., Walton, D. M., Avery, S., Blanchard, A., Etruw, E., McAlpine, C., & Goldsmith, C. H. (2009). Measurement properties of the neck disability index: a systematic review. *The Journal of orthopaedic and sports physical therapy, 39*(5), 400-C412. doi:10.2519/jospt.2009.2930

Maher, C., Moseley, A., Sherrington, C., & Herbert, R. (2001). Core journals of evidence-based physiotherapy practice. *Physiotherapy Theory and Practice, 17*(3), 143-151. doi:10.1080/095939801317077605

Maher, C. G., Moseley, A. M., Sherrington, C., Elkins, M. R., & Herbert, R. D. (2008). A description of the trials, reviews, and practice guidelines indexed in the PEDro database. *Phys Ther, 88*(9), 1068-1077. doi:10.2522/ptj.20080002

Masaracchio, M., Cleland, J., Hellman, M., & Hagins, M. (2013). Short-term combined effects of thoracic spine thrust manipulation and cervical spine nonthrust manipulation in individuals with mechanical neck pain: A randomized clinical trial. *Journal of Orthopaedic and Sports Physical Therapy, 43*(3), 118-127. doi:10.2519/jospt.2013.4221

McQuay, H. J., & Moore, R. A. (1997). Using Numerical Results from Systematic Reviews in Clinical Practice. *Annals of Internal Medicine, 126*(9), 712. doi:10.1059/0003-4819-126-9-199705010-00007

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*(4), 806-834. doi:10.1037/0022-006X.46.4.806

Miller, P. A., McKibbon, K. A., & Haynes, R. B. (2003). A quantitative analysis of research publications in physical therapy journals. *Phys Ther, 83*(2), 123-131.

Miyamoto, G. C., Leonardo Oliveira Pena, C., Galvanin, T., & Cristina Maria Nunes, C. (2013). Efficacy of the Addition of Modified Pilates Exercises to a Minimal Intervention in Patients With Chronic Low Back Pain: A Randomized Controlled Trial. *Physical therapy, 93*(3), 310-320. doi:10.2522/ptj.20120190

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., . . . Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *International Journal of Surgery, 10*(1), 28-55. doi:10.1016/j.ijsu.2011.10.001

Moreland, J., & Thomson, M. A. (1994). Efficacy of electromyographic biofeedback compared with conventional physical therapy for upper-extremity function in patients following stroke: a research overview and meta-analysis. *Physical therapy, 74*(6), 534-543.

Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question: understanding time lags in translational research. In (Vol. 104, pp. 510-520). London, England: SAGE Publications.

Moseley, A. M., Herbert, R. D., Sherrington, C., & Maher, C. G. (2002). Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). *Aust J Physiother, 48*(1), 43-49.

Naing, C., Aung, K., & Mak, J. W. (2012). Reporting 'number needed to treat' in meta-analyses: A cross-sectional study. *Journal of Evidence-Based Medicine, 5*(4), 232-237. doi:10.1111/jebm.12002

Nash-Stewart, C. E., Kruesi, L. M., & Del Mar, C. B. (2012). Does Bradford's Law of Scattering predict the size of the literature in Cochrane Reviews? *Journal of the Medical Library Association: JMLA, 100*(2), 135.

National Institutes of Health. (2014, October 23, 2014). Notice of Revised Definition of "Clinical Trial'. Retrieved from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-015.html

National Institutes of Health. (2018, January 17, 2018). NIH Definition of Clinical Trial Case Studies. Retrieved from https://grants.nih.gov/policy/clinical-trials/case-studies.htm

Newman, D., & Allison, S. C. (2007). Risk and physical therapy? *The Journal of orthopaedic and sports physical therapy, 37*(6), 287-289. doi:10.2519/jospt.2007.0106

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods, 5*(2), 241-301.

Nuovo, J., Melnikow, J., & Chang, D. (2002). Reporting Number Needed to Treat and Absolute Risk Reduction in Randomized Controlled Trials. *JAMA, 287*(21), 2813-2814. doi:10.1001/jama.287.21.2813

Osborne, J. W. (2008). Sweating the small stuff in educational psychology: how effect size and power reporting failed to change from 1969 to 1999, and what that means for the future of changing practices. *Educational Psychology, 28*(2), 151. doi:10.1080/01443410701491718

Paci, M., Cigna, C., Baccini, M., & Rinaldi, L. A. (2009). Types of article published in physiotherapy journals: a quantitative analysis. *Physiotherapy research* *international : the journal for researchers and clinicians in physical therapy, 14*(4), 203-212. doi:10.1002/pri.447

Pinson, L., & Gray, G. E. (2003). Psychopharmacology: Number Needed to Treat: An Underused Measure of Treatment Effect. *Psychiatric Services, 54*(2), 145-154. doi:10.1176/appi.ps.54.2.145

Portney, L. G., & Watkins, M. P. (2009). Foundations of clinical research: applications to practice.

Potter, J. (2010). Mapping the literature of occupational therapy: an update. *J Med Libr Assoc, 98*(3), 235-242. doi:10.3163/1536-5050.98.3.012

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of documentation, 25*(4), 348-349.

Pritchard, A., & Wittig, G. R. (1981). *Bibliometrics, A Bibliography and Index* (Vol. 1). Watford, Hertz, United Kingdom: ALLM Books.

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Roberts, D. (1992). The journal literature of physiotherapy: quality through peer review and access through MEDLINE. *Physiotherapy, 78*(1), 29-33.

Robinson, D. H., & Levin, J. R. (1997). Reflections on Statistical and Substantive Significance, with a Slice of Replication. *Educational Researcher, 26*(5), 21-26. doi:10.2307/1176544

Rosnow, R. L., & Rosenthal, R. (1989). Statistical Procedures and the Justification of Knowledge in Psychological Science. *American psychologist, 44*(10), 1276-1284. doi:10.1037/0003-066X.44.10.1276

Roush, J., Farris, J. W., Bordenave, L. M., Sesso, S., Benson, A. M., & Millikan, C. (2015). Commonly used statistical methods in the journals associated with physical therapy and physioltherapy. *Journal of Physical Therapy Education, 29*(1), 5.

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin, 57*(5), 416-428. doi:10.1037/h0042040

Sackett, D., Straus, S., & Richardson, W. (2000). Evidence-based Medicine: how to practice and teach EBM, 2nd. *London: Churchill Livingstone*, 169-182.

Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ, 312*(7023), 71-72.

Sawilowsky, S. (2018). [Primer on effect size].

Sawilowsky, S. (2003). Trivials: The birth, sale, and final production of meta-analysis. *Journal of Modern Applied Statistics, 2*(1), 5. Retrieved from http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?article=1748&context=jmasm

Sawilowsky, S. (2007). Effect sizes, interaction vs. main effects, and modified ANOVA table. In S. Sawilowsky (Ed.), *Real Data Analysis* (pp. 191-212). Washington, DC: InfoAge Publishing.

Sawilowsky, S. (2007). *Real data analysis*. Charlotte, N.C.: Information Age Pub.

Sawilowsky, S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistics, 8*(2), 3.

Sawilowsky, S. (2011). Statistical Fallacies: Misconceptions, and Myths. In (Vol. 3, pp. 1412-1417).

Sawilowsky, S., & Blair, R. C. (1992). A More Realistic Look at the Robustness and Type-Ii Error Properties of the T-Test to Departures from Population Normality. *Psychological Bulletin, 111*(2), 352-360. doi:Doi 10.1037/0033-2909.111.2.352

Sawilowsky, S., Sawilowsky, J., & Grissom, R. (2011). Effect Size. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (Vol. 1, pp. 426-429). Berlin Heidelberg: Springer-Verlag.

Sawilowsky, S., & Yoon, J. S. (2002). The trouble with trivials (p > .05). *Journal of Modern Applied Statistical Methods, 1*(1), 143-144.

Shadgan, B., Roig, M., HajGhanbari, B., & Reid, W. D. (2010). Top-cited articles in rehabilitation. *Archives of Physical Medicine and Rehabilitation, 91*(5), 806-815.

Shumway-Cook, A., Brauer, S., & Woollacott, M. (2000). Predicting the Probability for Falls in Community-Dwelling Older Adults Using the Timed Up & Go Test. *Physical therapy, 80*(9), 896. doi:10.1093/ptj/80.9.896

Siemonsma, P. C., Stuive, I., Roorda, L. D., Vollebregt, J. A., Walker, M. F., Lankhorst, G. J., & Lettinga, A. T. (2013). Cognitive Treatment of Illness Perceptions in Patients With Chronic Low Back Pain: A Randomized Controlled Trial. *Physical therapy, 93*(4), 435-446. doi:10.2522/ptj.20110150

Simon, C. B., Coronado, R. A., Wurtzel, W. A., Riddle, D. L., & George, S. Z. (2014). Content and bibliometric analyses of the Journal of Manual & Manipulative Therapy. *The Journal of manual & manipulative therapy, 22*(4), 181-190. doi:10.1179/2042618614Y.0000000075

Sink, C., & Stroh, H. (2006). Practical significance: The use of effect sizes in school counseling research. *Professional School Counseling, 9*(4), 401-411.

Smith, D. R., & Rivett, D. A. (2009). Bibliometrics, impact factors and manual therapy: Balancing the science and the art. *Manual Therapy, 14*(4), 456-459. doi:10.1016/j.math.2008.11.004

Terwee, C. B., Roorda, L. D., Dekker, J., Bierma-Zeinstra, S. M., Peat, G., Jordan, K. P., . . . de Vet, H. C. W. (2010). Mind the MIC: large variation among populations and methods. *Journal of Clinical Epidemiology, 63*(5), 524-534. doi:10.1016/j.jclinepi.2009.08.010

Tesio, L., Gamba, C., Capelli, A., & Franchignoni, F. P. (1995). Rehabilitation: the Cinderella of neurological research? A bibliometric study. *The Italian Journal of Neurological Sciences, 16*(6), 473-477. doi:10.1007/BF02229325

The PEDro Partnership. (2019). Physiotherapy Evidence Database. Retrieved 04/06/2019, from University of Sydney https://www.pedro.org.au/english/about-us/cebp/

Thompson, B. (1993). The Use of Statistical Significance Tests in Research: Bootstrap and Other Alternatives. *The Journal of Experimental Education, 61*(4), 361-377. doi:10.1080/00220973.1993.10806596

Thompson, B. (1996). Research news and Comment: AERA Editorial Policies Regarding Statistical Significance Testing: Three Suggested Reforms. *Educational Researcher, 25*(2), 26-30. doi:10.3102/0013189X025002026

Thompson, B. (1998). *FIve Methodology Errors in Educational Research: The Pantheon of Statistical Significance and other faux pas*. Paper presented at the Annual Meeting of the AMerican Educational Research Association, San DIego, CA.

Tilson, J. K., Marshall, K., Tam, J. J., & Fetters, L. (2016). A bibliometric analysis of statistical terms used in American Physical Therapy Association journals (2011-2012): evidence for educating physical therapists. *BMC medical education, 16*(1), 118. doi:10.1186/s12909-016-0641-1

Tracey, T. J. (2000). Issues in the analysis and interpretation of quantitative data: Deinstitutionalization of the null hypothesis test.

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2): Reading, Mass.

Tukey, J. W. (1991). The Philosophy of Multiple Comparisons. *Statistical Science, 6*(1), 100-116. doi:10.1214/ss/1177011945

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of counseling psychology, 51*(4), 473.

van Kampen, D. A., Willems, W. J., van Beers, L. W., Castelein, R. M., Scholtes, V. A., & Terwee, C. B. (2013). Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). *Journal of orthopaedic surgery and research, 8*(1), 40. doi:10.1186/1749-799X-8-40

Wakiji, E. M. (1997). Mapping the literature of physical therapy. *Bull Med Libr Assoc, 85*(3), 284-288.

Weeks, D. L., & Noteboom, J. T. (2004). Using the number needed to treat in clinical practice. *Archives of Physical Medicine and Rehabilitation, 85*(10), 1729-1731. doi:10.1016/j.apmr.2004.03.025

Whitman, J. M., Cleland, J. A., Mintken, P. E., Keirns, M., Bieniek, M. L., Albin, S. R., . . . McPoil, T. G. (2009). Predicting short-term response to thrust and nonthrust manipulation and exercise in patients post inversion ankle sprain. *The Journal of orthopaedic and sports physical therapy, 39*(3), 188-200. doi:10.2519/jospt.2009.2940

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*: Springer.

Wiles, L., Matricciani, L., Williams, M., & Olds, T. (2012). Sixty-five years of Physical Therapy: bibliometric analysis of research publications from 1945 through 2010. *Phys Ther, 92*(4), 493-506. doi:10.2522/ptj.20110013

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American psychologist, 54*(8), 594.

## ABSTRACT


**REPORTING NUMBER NEEDED TO TREAT IN CLINICAL TRIALS PUBLISHED IN PHYSICAL THERAPY SPECIFIC LITERATURE 1989-2018**

by

**SUSAN ANN TALLEY**

**August 2019**

**Advisor:** Dr. Shlomo S. Sawilowsky

**Major:** Education Evaluation and Research

**Degree:** Doctor of Philosophy

Evidence-based practice requires physical therapists to make clinical decisions about the best intervention to use when providing services to patients/clients. Although null hypothesis significance testing (NHST) is frequently used to interpret the outcome of a clinical trial investigating the comparative effectiveness of an intervention, statistical significance does not directly translate into clinical importance. Number needed to treat (NNT) is a measure of effect size (ES) that may be particularly useful when translating the results from clinical trials to PT clinical practice. The purpose of this study was to conduct a bibliometric content analysis of the methods of reporting research results of clinical trials published in the physical therapy specific literature, specifically NHST, ES and NNT.

The frequency of reporting the result of NHST, EST and NNT was examined in 448 clinical trials published in three physical therapy specific journals (*Journal of Neurologic Physical Therapy*, *Journal of Orthopaedic Physical Therapy* and *Physical Therapy*) between 1989 and 2018. More than 90% of clinical trials included a report of the result of NHST but less than 30% reported effect size. NNT was reported.in only eight (1.79%)

articles. The number of articles in which NNT was reported during 2013-2018 was statistically greater than the previous four 5-year periods. However, there was no positive linear trend of the frequency of NNT during the last nine years, 2010 – 2018. This is the first study in which evidence is presented indicating increased reporting of NNT in the physical therapy specific literature however there is no evidence of a positive trend during the past nine years. Physical therapist students, educators, researchers and practitioners would be well served to improve their understanding of how to include NNT in clinical trial research designs to improve decisions about the clinical importance of an intervention.

## AUTOBIOGRAPHICAL STATEMENT

Susan Ann Talley was born in Highland Park, MI and grew up in Detroit in the house her father lived in from the time he was 10 years old, blocks from the house where her mother grew up. She attended the Detroit Public Schools, graduating from Pershing High School in 1972. Her father was a Detroit police officer for 31 years after serving in the Navy and having earned a GED. Her mother, who also graduated from Pershing High School (does this make Sue a DPS legacy student?), was an office worker. Growing up in the late 60's and early 70's was a riot (literally in 1967) and she leaned toward humanism, the anti-war movement, the Beatles and folk music.

In 10th grade Sue decided that she wanted to become a pediatric physical therapist (PT). Although no one in her family had graduated from college she was determined to earn her BS in PT. Wayne State was her only option and she earned her PT degree in 1976. PT school was the most intense educational experience in her life (until her PhD Qualifying exams!) and she swore she would never go to school again, ever! One year later she started taking French courses and then decided she may as well work on a degree. In 1982 she earned a MA degree in Human Development and Relationships with a focus on infants and toddlers from Wayne State University.

Her first professional position was as a pediatric physical therapist at the MSCCA, working with children with cerebral palsy and with sensory processing disorders for three years. She was asked to become a TA in her master's program and thus began her affinity for academia. At the age of 25 she joined the PT faculty at Wayne State, teaching human development, neurologic and pediatric physical therapy and... research. Fortunately for the students, a mentor was hired to teach the research course that first year, mentored her the following year, and then she was on her own. Recognizing her woeful preparation to teach research, she pursued professional development courses, leading to her first exposure to Education Evaluation and Research. At the time faculty were not allowed to earn degrees at Wayne State so she took a smattering of courses. Along the way she earned a Doctor of PT degree from Creighton University in 2005. A few years later she realized that she had taken nearly enough courses to earn her PhD in EER and officially enrolled in the doctoral program.

After 34 years as a member of the PT faculty at Wayne State, including 10 years as the academic director, she retired – but only for a blink of an eye. The next day she started as Associate Director for Professional Education for the PT department at the University of Michigan-Flint in 2014. Along the way she practiced as a PT with a variety of patients, including adults and children with neurological conditions, cardiac rehabilitation, and palliative care.

Dr. Talley has co-authored eight peer-reviewed articles largely focusing on physical performance in community dwelling older adults, most recently looking at the contribution of genomics. She has also authored four chapters on a variety of physical and occupational therapy topics. Throughout her career she has been very active in the Michigan Physical Therapy Association, including serving as President for four years. She was awarded the Marjorie L Stamm Outstanding Service award and the Jane Murdock Legislative Advocacy award by her colleagues.

She has been married to Joe Olesnavage for 30+ years. Her children, Jason and Katy, live in Chicago, IL and Hermosa Beach, CA. respectively. She has hiked on the Appalachian Trail, canoed in the Boundary Waters and performed folkdance on Moscow TV. Life is good!!!