

RESEARCH FORUM--How to Critically Read a Journal Research Article

Thomas R. Lunsford, MSE, CO
Brenda Rae Lunsford, MS, MAPT

ABSTRACT

The purpose of this article is to present the concepts involved in reader evaluation of research literature. Knowledge gained through research is valuable only if it is shared with others. It is the duty of the members of a profession to critically review the profession's literature, challenge claims that seem unreasonable and champion those that elevate the profession. Better patient care and increased professional growth will result when clinicians learn to evaluate and make use of the research literature.

The essential elements of a research article are the title, abstract, introduction, method, results, discussion and conclusion. The introduction should contain the following elements: statement of the problem, literature review, purpose and expected results (hypothesis). The method should define the subjects, instrumentation and apparatus, procedure, and data analysis. The data analysis is divided into statistical tests for continuous data and discrete data. The results section should succinctly present the results with no interpretation of their meaning. The discussion is where the knowledge and insight of the author(s) are allowed to bloom.

This article identifies the polarization between clinician and researcher, and readers are enjoined to help educate both calling of the O&P profession. The profession will be elevated with the development of research-oriented clinicians and clinically oriented researchers.

Introduction

Developing a foundation based on research has long been a goal of the orthotics and prosthetics (O&P) profession. Substantially more O&P research is being published today than 10 years ago. The quality and validity of O&P research articles also have improved significantly. Therefore, O&P professionals must make it one of their collective objectives to become educated readers and interpreters of professional research literature.

Clinicians often are critical of studies of a purely research nature with very little relevance to the clinic, and researchers often are critical of clinicians who are not aware of the latest research developments. Journal editors are criticized for publishing manuscripts deemed as too research-oriented with little clinical relevance or articles that have clinical implications with no scientific basis. As long as O&P practitioners see themselves as either researchers or clinicians, this crossfire of criticism will continue. However, if the O&P profession is to grow, it must develop a foundation that is both unique and based on research.

Researchers are not required to be clinically involved, but those who are have more credibility. Similarly, clinicians are not required to conduct research, but the growth of the O&P profession depends on clinicians learning how to critically read research. The two activities can become integrated if clinicians critically read O&P research and clinically apply the findings.

Documentation of scientific evidence supporting the efficacy of devices must be published. Research results must be communicated so the findings can be integrated into the expanding body of O&P knowledge. Researchers are obligated to share their information with others who might wish to replicate the studies. Published reports provide a vehicle for igniting sparks among other interested investigators and informing them of completed work. A published study will be valuable to readers if a new twist has been given to an old idea, if literature has been reviewed critically, if an author's experiences are shared and if illuminating situations are cited (1).

Why would anyone devote time and resources to undertake a study if the outcome is not shared? The researcher is derelict in responsibilities to colleagues and peers if results of a study are not published. Also, clinicians owe it to both their patients and themselves to independently and critically review pertinent literature.

One common roadblock to reviewing O&P research articles is

One common roadblock to reviewing O&P research articles is

learning to become a critic without formal training in research methods. At the current stage of development of the O&P profession, the skills and competencies required to engage in critical analysis must be self-developed; there are no "how to conduct, read and critique O&P research" courses or books available. O&P professionals will have to pull themselves up by the bootstraps and develop expertise quietly and purposefully until the knowledge is pervasive. The research evolution in the O&P profession affords an exciting opportunity for practitioners to participate in and contribute to the advancement of the O&P profession.

The main criticism of increasing research efforts is that research is too technical and does not have much clinical relevance. Human nature tends to reject the misunderstood. Rather than continue to reject, perhaps it would be better for O&P clinicians to learn to read research reports; the purpose of this article is to provide them with a guide or framework for reviewing research literature.

Elements

Published studies are expected to meet the literary standards of organization and precision that apply to all forms of scientific and technical writing (2).

The standard journal format of published research reports contains the following subheadings or elements: title, abstract, introduction, method, results, discussion and conclusion.

Each of the elements is discussed, and the essential ingredients are described. A readers' guideline is provided in [Table 1](#) as checklist for evaluating a research article.

Title

The title of an article is very important since initially it is the only exposure readers have to the article. As readers peruse the table of contents of a journal, they will appreciate titles that are both short and informative. A good title should give insight into what (was done), whom (it was done to) and how (it was done).

For example, consider the title "A Comparison of Walking Velocity in Adult Male Transtibial Amputees Using SACH versus Dynamic Response Feet." What was done: a comparison of walking velocities. To whom was it done: adult male transtibial amputees. How was it done: by measuring walking velocity.

Try to determine the what, whom and how for the following title: "Effects of Inhibitory Orthoses on Bony Alignment of the Foot and Ankle During Weightbearing in Children with Spasticity." In this case, the "what was done" involved the measurement of the bony alignment of the foot and ankle during weightbearing. The "to whom was it done" were children with spasticity, and the "how was it done" was by intervening with inhibitory orthoses. Practitioners have precious little time to review the literature and cannot afford to expend time fishing through minutiae to identify important articles. If a title is too long or loaded with complex technical jargon, chances are the article will be skipped. This is unfortunate since the article may contain significant findings. However, this is a problem for the authors and the journal's editorial board, not the readers. It is not reasonable to expect readers to take the time to read an article that is improperly or inappropriately titled. If the title hooks the readers, they will be motivated to read the abstract.

Abstract

The abstract should contain a brief statement about the study's purpose, method, results, conclusion and clinical relevance. Reading the abstract is a time-efficient way for readers to determine if the article suits their needs. If an abstract is well written, some hurried readers may choose to read only the abstract, then return to the article later as more in-depth information or convincing evidence is needed.

A well-written abstract gives readers a good idea of what the study is about, how it was conducted, and the findings and recommendations of the author. Readers should remember not to accept the conclusions before critically reading the entire article. The abstract should pique readers' interest, but they should reserve judgment until they read the more substantial

evidence presented in the body of the paper.

Given the increasing number of journals available for review, determining which ones deserve readers' full attention requires a screening system. Using the abstract as a screening device is a recommended starting point.

Introduction

The introduction to a research article should contain the following major elements: statement of the problem, literature review, purpose of the study and expected results (hypothesis).

Statement of the Problem

The statement of the problem should describe the questions and concerns that led the author to undertake the investigation. Readers should ask themselves, "Why did the author conduct this study? What question did the author try to answer?" Readers should get a sense of the answers to these questions early in the introduction.

Literature Review

The literature review should establish a theoretical and historical basis for the research paper and should provide support for construct validity. Construct validity is the theoretical conceptualization of intervention and response (3). It is a type of measurement validity that informs readers of the degree to which a theoretical construct is measured by an instrument (3).

The author should attempt to identify a "gap of knowledge" between what is known (or previously documented) and what is desired to be known. Readers perusing the introduction should try to identify the "gap" as well as find information in the literature that supports the concept and approach of the study. In this section, the author should explain how his/her work is an attempt to close the gap by explaining why the study was conducted.

Another way the author can close the gap is to critically review the published work of others and point out flaws, inconsistencies or areas where no conclusions can be drawn. For example, Sutherland helped narrow the gap in defining the function of the gastrocnemius-soleus during walking (4). In the introduction of his article, he states the primary role of the plantarflexor muscles is to stabilize the tibia on the tarsus as had been demonstrated by several investigators; he postulates a secondary role is to stabilize the femur on the tibia. Sutherland states, "This secondary knee-stabilizing activity of the plantarflexor muscles, however, remains in dispute. Does such action occur? If so, how is the knee joint affected?" For Sutherland, this was the "gap" in the continuum of knowledge about the role of the plantarflexor muscles during walking. Prior to his study, it was known that these muscles begin to contract in early stance, but their role was not clear (the gap). Sutherland provided an answer through an EMG study using invasive wire electrodes in normal subjects while they were walking on level surfaces.

A literature review should be current; i.e., cited references should not be more than five to seven years old unless they are "classics."

Readers should determine whether the author has failed to cite references on any crucial points. The literature review should be sufficient to meet the objectives stated above, but the author should avoid "overkill" in reviewing the literature. Very little is to be gained from citing 10 references to make one point.

A general statement should be made identifying the type of study (e.g., experimental, correlational or descriptive). This statement alerts readers to expect certain information in a particular format. For example, if the study is experimental or correlational, the author should delineate the expected results or the null hypothesis (this subject will be given more attention later). If the study is descriptive, the author should identify the need to collect the descriptive data or report the findings.

The literature review should provide readers with a clear idea of what has been done in the past and provide conceptual support to the method. Readers can easily tell the author has spent a reasonable amount of time reviewing the literature if the review is a synthesis of reports logically arranged in sequential and chronological order.

Purpose of the Study

The purpose of the study should be described in a direct, clear statement. For example, the purpose of the following study is to compare the walking velocity of adult male transtibial amputees using two different prosthetic feet, or as stated in "Orthotic Management of Scoliosis in Familial Dysautonomia" (5): The purpose of this study [is] to describe the characteristics of Familial Dysautonomia that give rise to treatment modifications of accepted orthotic intervention (TLSOs and CTLSOs).

The author who cannot clearly state the purpose of his/her research will most likely produce results that are not applicable in clinical situations.

Expected Results

Ideally, the author of a research article should frame the research question in the form of a hypothesis. For these purposes, a hypothesis is defined as a tentative theory or supposition provisionally adopted to explain certain facts and to guide readers into further investigation (6).

A report of a study should include an explicit statement of the study's hypothesis or expected research results. A research hypothesis states the researcher's true expectation of results; it is a statement that guides the interpretation of outcomes and conclusions. However, the statistical analysis of data is based on testing a statistical or null hypothesis, which differs from the research hypothesis in that it will always express no difference, or no relationship between the independent and dependent variables is expected.

For example, assume a study has been designed to compare the energy expenditure exhibited by a group of transtibial amputees who are fitted with two types of prosthetic feet, F_1 and F_2 . If F_1 is an older design that has been the industry standard for years and F_2 is a newly designed prosthetic foot with improved materials and biomechanics, then one would expect the energy expenditure with F_2 to be better (less) than with F_1 . If the energy expenditures are denoted as E_{F1} and E_{F2} and the research hypothesis as H_0 , then the hypothetical statement is written as follows:

$$H_0: E_{F1} > E_{F2}$$

This says, "It is hypothesized (expected) that the energy expended with the older, industry-standard prosthetic foot is greater than that with the newly designed, high-technology prosthetic foot." The statistical or null hypothesis would state no difference exists between the two types of prosthetic feet. The analysis of the data could confirm (prove) the null hypothesis, thereby disproving the hypothesis.

The hypothesis could be disproved two ways. First, the energy expended with the two feet could be the same, i.e., not significantly different:

$$H_0: E_{F1} = E_{F2}$$

Alternatively, the energy expended with the older prosthetic foot could be less than that expended with the newly designed prosthetic foot:

$$H_0: E_{F1} < E_{F2}$$

After the statement of the problem, literature review, purpose of the study and expected results have been examined in the introduction, the method used in the study is described.

Method

The method section of the research report should clearly explain how the study was conducted. Critical readers should prefer they are going to replicate the study: Is there sufficient detail in the method to conduct the study and obtain similar results?

For clarity and convenience, the method can be divided into the following subsections: subjects, instrumentation and apparatus, procedure, and data analysis.

Subjects

The author should summarize and describe the subjects who participated in the study in terms of age, sex, diagnosis and other pertinent demographic characteristics. If a particular diagnosis or characteristic is required for inclusion in the study, the criteria should be explained. Readers appreciate authors who summarize the characteristics of their subjects in a table (see [Table 2](#)).

The extent to which readers are able to use the results of the study depends on how the sample of subjects was selected and how many subjects were included in the sample. Ideally, the subjects should be selected randomly so each individual in a larger population has the same chance of being included in the sample as anyone else. For more details on sampling, the reader is referred to the last two issues of the JPO (7,8).

In all likelihood, the sample in a research article will be a "sample of convenience." That is, the subjects will be comprised of institutionalized clients, patients in a particular clinic or students attending a certain program. The results from a research project using a convenience sample are not as easily generalized as the results from a study where the subjects are randomly selected.

If the effect of a rigid AFO on walking velocity were measured on a convenience sample of middle-aged hemiplegics from an intensive rehabilitation institution, readers should be wary of expecting the result to apply to more elderly, homebound hemiplegics.

Frequently, O&P research is conducted to compare two or more similar devices. Moreover, the control group may have received no devices, and experimental groups may have worn competing, similar devices. Critical readers should determine if the control group that received no device and the experimental groups that did receive devices were randomly selected from the same pool of subjects. This is an area of serious weakness in retrospective studies where very little information is available to compare subjects in the various groups.

Instrumentation and Apparatus

The instruments used to measure variables should be described in such a way that readers can replicate the study. Footnotes specifying model numbers, corporate names and addresses, and other pertinent details about the instruments should be included here. If standardized questionnaires are used, they must be referenced.

Any apparatus designed and developed by the researcher should be fully described with a drawing, photograph and description. If a questionnaire is developed by the researcher, it also should be presented.

Readers should rely on their natural curiosity when evaluating the instrumentation's appropriateness for measuring the study variables. Were the instruments calibrated? How were they calibrated? Are they reliable? Are they repeatable day-to-day? Is the instrument measuring what it is purported to measure? One common measurement error occurs when the author intends to measure pressure but measures force or torque instead. These are entirely different physical entities and cannot be interchanged without impunity.

Some researchers refer to reliability when describing the instrumentation or apparatus. Reliability refers to the reproducibility of results at a different time or by a different investigator. Readers should be wary of fickle instruments that only a well-trained technician familiar with all their idiosyncrasies can operate; in someone else's hands, different results may occur. Some research projects are undertaken solely for the purpose of establishing the reliability of an instrument. If this is the case, the author is obligated to reference that in his or her article.

Procedure

The procedure section of the method should explain exactly how and when the steps of the study were applied and how the data were collected. Readers who have a clear idea of how the research was conducted also will have a clear idea of how to apply the results or determine if they can accept the author's conclusions.

Readers should be satisfied the changes noted during the study are the result of the devices being studied and not the result of a sloppy procedure. The concern for what actually causes the changes is called internal validity (9). Internal validity is

concerned with the following question: Did the experimental treatment cause the observed change in the dependent variable? In other words, could other (extraneous) factors be responsible for that change? Other factors that offer competing explanations for the observed relationship between the independent and dependent variables threaten internal validity; that is they interfere with cause-and-effect inferences.

Sometimes studies are conducted over such a long period of time that it is unclear whether the treatment being studied caused the change or if the change was due to other events, such as healing, fatigue, growth or aging. A list of the most common threats to internal validity is shown in [Table 3](#). (The reader is referred to pages 135-139 of Reference 3 for a comprehensive explanation of these confounding factors that threaten internal validity.)

The testing procedure itself can cause changes in the results. For example, readers should be aware that a subject being tested repeatedly with the same instrument may improve without any concomitant improvement in the functional skill that the device being tested purports to cause. In certain cases, subject familiarity with the newly designed orthosis or prosthesis causes improvement in function. This potential problem is thwarted when the investigator uses a comparable control group.

Some investigators improve their internal validity by alternating the order in which the subjects perform certain tasks. Readers should try to assess if the investigator has taken steps to control sources of secondary variance.

Data Analysis

The data analysis section should describe all testing applied to the data. Readers must assess if the author chose the appropriate statistical tests for the type of study and design. This part of the method should not contain any results.

When analyzing data, arithmetic operations too frequently are misapplied to data based on nominal and ordinal levels of measurement (10). The most common error is analyzing ordinal data as though they were quantitative (interval or ratio). Ordinal data often are obtained from questionnaires where the answer to a question may be "always, most often, usually, infrequently or never." The magnitude between the differences in adjacent categories on the ordinal scale is not measurable. The inclination to assign numbers to the answers is irresistible, e.g., always 5, most often 4, etc. There is nothing wrong with this procedure for sorting the answers and performing tallies. However, a problem occurs when the arbitrary numerical assignments are analyzed with conventional statistical tests as though the answers were measured with a calibrated instrument.

Even experienced investigators sometimes fail to realize, or to remember, that arithmetic operations (addition, subtraction, multiplication, division, squaring) cannot be performed legitimately on numbers associated with nominal or ordinal measurements. Ordinal scores merely reflect "greater than" or "less than" values, and the differences between the scores are not equal.

Different from ordinal and nominal data are continuous data for which mathematical manipulation is valid. There are two types of continuous data: interval and ratio. The difference between interval and ratio data is the zero-value for interval data is arbitrary (e.g., temperature) whereas the zero-value for ratio data is absolute (e.g., height, velocity, etc.). The types of data are summarized in [Figure 1](#).

An important and often missed step in the treatment of the data is screening (11). Readers can have more confidence in the statistical analysis when the author mentions the data were screened for errors in data entry, outliers and distribution. In computerized data management, there are numerous opportunities to err.

Conventional parametric statistical analyses are conducted on continuous data as described in Figure 1. It is helpful to categorize four types of analyses: descriptive, comparative, associative and predictive. These common tests are shown in Figure 2 for both continuous and discrete types of data. As shown, it is common for continuous data to use means and standard deviations to summarize data sets whereas it is common practice to use frequencies, counts or percentages to summarize ordinal or nominal data. The comparative tests are a little more complicated; for continuous data, authors should use the t-test when comparing one or two devices and the ANOVA (analysis of variance) when comparing more than two devices. Associative tests are used to establish relationships between variables, and predictive tests are used to fit curves through data and extrapolate beyond the range of measured data.

Results

The results section of a journal article should include the findings of the data analysis without commentary. Two groups of statistics, descriptive and inferential, may be included. Descriptive statistics summarize the raw data such as means and standard deviations. Inferential statistics are more complex and allow readers to infer conclusions from the data. It's not necessary to publish raw data in its entirety. Charts, graphs, tables and histograms are welcome additions when attempting to develop an overall summary of the results.

Most clinicians are not highly skilled at interpreting statistical analyses. As a result, the statistics included in research articles can be intimidating. The following concepts may help readers gain an understanding of two basic inferential statistical concepts.

The first concept is the level of significance (12). Statistical tests have what is known as a level of significance. For example, if two groups are wearing different prosthetic feet and the investigator is measuring energy cost during walking, then a 5-percent level of significance would imply that the difference in energy cost has a 5-percent chance of being due to chance, not to differences in the prosthetic feet. Alternatively, the chance the measured difference in energy cost is to the prosthetic feet is 9 percent. This level of significance is written ($p < .05$) and is referred to as the p-value.

If the calculated p-value falls at or below the specified level of significance, the result is considered statistically significant. If the p-value is greater than the preset level, the result is considered not significantly different or statistically insignificant. Sometimes the level of significance is referred to as the *alpha* level or as the criterion for rejection of a hypothesis. Understanding this concept of statistical significance, readers should be able to review the results of the majority of the articles in research journals and understand the significance without knowledge of the statistical test itself.

The second major statistical concept concerns statistical versus clinical significance (12). If a large group of relatively homogeneous subjects is used in a research study, a very small difference in their test scores will cause statistical significance. If a small group of more divergent individuals is studied, a very large absolute difference must be seen before a result is considered statistically significant.

The effect of sample size is built into the statistical probability and is reflected in the p-value. Consequently, readers must be cautious about assigning clinical significance to minute, though statistically significant, differences in large group experiments. Similarly, large differences that are statistically insignificant in small group studies may prove to be both clinically and statistically significant when replicated on a larger scale.

Even without sophisticated backgrounds in statistics, astute readers should be able to understand the information in the results section by reading the text carefully and studying the graphs and tables. If the graphs and tables seem incomprehensible, the problem is probably not with the readers but with the author's presentation of the data. Research critics must decide if a study that identifies significant differences is clinically relevant.

Discussion

Readers will be able to judge the knowledge and insight of the investigator in the discussion section. Has the author tied the results to the material presented in the introduction? Is the research question answered? Has the author given meaning to the results? While reviewing this section, readers should think back to the logic of the arguments presented and consider the issues related to the original problem. Is there a succinct reference to the original hypothesis? Has the author considered broader implications of his/her findings?

One common pitfall readers should watch for is a discussion of insignificant results described as though they were significant. Imputing meaning to data that may reflect chance differences is misleading because it suggests significance where none exists.

Readers also should be wary of unsupported conclusions. For example, an unsupported conclusion results when an author admits there is no significant difference between two prosthetic treatments, but explains that if more data were collected, the results would surely become significant. Drawing conclusions from future experiments is fraught with suspicious bias. Most research is not of a dramatic, profound, profession-changing nature and usually creates more questions than it answers. Readers should ask themselves, has the author made suggestions for future studies to expand upon his/her lead? Finally,

critical readers must judge if the researcher has conducted fair and objective research.

Conclusion

The conclusion section of a research article contains a brief restatement of the experimental results and describes the implications of the study. Because the abstract summarizes the entire article, only key points are given in the conclusion.

Quality patient care results when existing knowledge is combined creatively with new knowledge. It is not enough for a clinician to be just a clinician or a researcher to be just a researcher. The O&P profession cannot help but be elevated by a preponderance of research-oriented clinicians and clinically oriented researchers.

A well-written article should be comprehensible by knowledgeable clinicians. If this is not the case, it is likely the readers are not the problem-but rather the article is poorly written. Clinicians can enhance their comprehension by becoming educated readers. Becoming educated reviewers of O&P literature is a positive step in providing better patient care.

THOMAS R. LUNSFORD, MSE, CO, is director of the orthotic department at The Institute for Rehabilitation and Research in Houston, Texas, and assistant professor of physical medicine and rehabilitation at Baylor College of Medicine.

BRENDA RAE LUNSFORD, MS, MAPT is visiting assistant professor at Texas Woman's University in Houston, Texas, and physical therapist II at The Institute for Rehabilitation and Research.

References:

1. Fishbein M. Medical writing: the technician and the art, 4th ed. Springfield, Ill.: Charles C. Thomas, 1978.
2. Currier DP. Elements of research in physical therapy, 2nd ed. Baltimore: Williams and Wilkins, 1984:298.
3. Portney LG, Watkins, MP. Foundations of clinical research- applications and practice. Norwalk, Calif.: Appleton & Lange, 1993:680.
4. Sutherland DH. An electromyographic study of the plantarflexors of the ankle in normal walking on the level. JBJS January 1966;48A:1 :66-71.
5. Cappa AJ, Burke SB, Axlerod FB, Levine DB. Orthotic management of scoliosis in familial dysautonomia. JPO Summer 1994;6:3:74-8.
6. Webster's new collegiate dictionary, 2nd ed. Springfield. Mass.: G & C Merriam Co.
7. Lunsford TR, Lunsford BR. The research sample, part I: sampling. JPO Summer 1995;7:3:105-12.
8. Lunsford TR, Lunsford BR. The research sample, part II: sample size. JPO Fall 1995;7:4:137-41.
9. Lunsford TR. Types of clinical studies. JPO October 1993;5:4:105-11.
10. Lunsford BR. Methodology: variables and levels of measurement. JPO October 1 993;5:4:1 21-4.
11. Lunsford BR. Statistics: screening and data summary. JPO October 1993;5:4:125-30.
12. Domholdt E, Malone T. Evaluating research literature: the educated clinician. Phvs Ther April 1985;65:4:487-91.



Contact the Academy
(703) 836-0788
(703) 836-0737 Fax

526 King Street
Suite 201
Alexandria, VA 22314

Copyright © American Academy of Orthotists & Prosthetists (AAOP)
All rights reserved. [See disclaimer](#)



Website built