Journal of Pelvic, Obstetric and Gynaecological Physiotherapy, Autumn 2023, **133,** 47–52

GETTING THE MOST OUT OF RESEARCH

Why understanding study methods matters

J. A. LaCross
School of Physical Therapy, Texas Woman's University, Dallas, TX, USA

Abstract
A significant barrier to a clinician's ability to translate research into physiotherapy practice is an incomplete understanding of how to analyse and interpret study methods. The aim of this commentary is to unpack each of the individual elements included in the methods section of a research article. These include the research design, participants, instrumentation, procedures and data analysis. In addition to a breakdown of the individual elements, examples and clinical applications for each are provided. Using the information presented to improve skills in critically appraising the research literature will improve the delivery of evidence-based practice.

Keywords: evidence-based practice, methods, research, translation.

Introduction

The translation of published research into evidence-based practice presents an ongoing challenge for physiotherapists. While this may be because of a variety of factors, an incomplete understanding of the components of the methods section of a study and what these tell a clinician is a significant barrier for many. Luckily, the unpacking of these methods is "low-hanging fruit" and the focus of this commentary. Before exploring fundamental statistical elements and what these mean (which will be covered in a future contribution to this series), it is helpful to gain a solid understanding of how the data in a methods section inform both a study's results and critical reasoning for clinical application.

Broadly speaking, the information included in a methods section tells the reader what to expect in the remainder of the article. Imagine that the methods are a road map. The more accurate and detailed the map is, the easier it is to get where you are going. What if half of it is missing? What if it is upside down? These are issues that may prevent you from getting from where you are (i.e. a clinician wanting to find out more information) to where you need to go (i.e. clinical guidance on how to help a patient). While researchers take rigorous steps to minimize and account for methodological issues, flawed application of research can contribute to disappointing clinical outcomes. This happens when you attempt to apply research to a patient that was conducted on a different set of participants, asked a slightly different question, employed a small sample size (Kamper 2022) and resulted in a very close to not-significant finding (Martínez-Camblor *et al.* 2019). As a clinician, it helps to notice if the research applies to the question at hand. This is why skipping directly from reading an abstract to the results and conclusion is not recommended. There are five main components in the methods section of a quantitative research study: research design, participants, instrumentation, procedures and data analysis. Each of these are explored below and translational applications are suggested.

Correspondence: Jenny LaCross PT DPT PhD ATC CLT, School of Physical Therapy, Texas Woman's University T. Boone Pickens Institute of Health Sciences – Dallas Center, 5500 Southwestern Medical Avenue, Dallas, TX 75235-7299, USA (e-mail: jalacros@umich.edu).

Research design

The research design is based on the nature of the research question(s) asked. There are three categories, each with a unique purpose, that encompass multiple more-specific study designs (see Table 1) (Portney & Watkins 2015):

(1) *Descriptive research* aims to help characterize a certain condition or patient population in order to assist a reader's understanding. Since such studies can be informative in many ways, these tend to be the building blocks of future studies.
(2) *Exploratory research* seeks to identify relationships between a condition of interest

Table 1. Types of research design [adapted from Portney & Watkins (2015, p. 21), © 2015 F. A. Davis Company, reprinted with permission]

	Research design		
Variable	Descriptive	Exploratory	Experimental
Study design	Case study Case series Descriptive surveys Qualitative research Normative studies Developmental studies	Cohort studies Case-control studies Methodological studies*	Randomized controlled trials Quasi-experimental studies Single-subject design studies
Purpose	Describe	Correlation/prediction	Cause and effect (causal)

*Reliability and validity.

Table 2. Recruitment considerations for clinical application

Participant recruitment	What to consider for clinical application
How?	*Probability techniques (Portney & Watkins 2015)* *Simple random.* Participants are randomly picked from all possible participants *Systematic.* The researcher divides the number of desired participants by the number of available participants (usually on a list), and then uses that interval for selection *Stratified random.* The researcher divides the participants into unique groups (e.g. nursery, primary and secondary school), and then randomly selects a certain sample from the people available in each group *Non-probability techniques* *Convenience (most common).* Participants are picked based on availability to the researcher *Purposive.* The researcher picks the participants *Snowball.* A few participants are initially identified who meet the criteria, and then share the names of others who also meet the criteria, creating a "snowball" effect
When?	Time of day Time of year/season Before/after surgery Pre-/post-pregnancy, menopause, etc.
Where?	Urban/suburban/rural Hospital/clinical/gym/community Different regions of the same country Different countries
How many?	Was an *a priori* power analysis performed? What was the α, β and expected effect size? Does the number of actual participants meet the predetermined threshold?

(e.g. constipation) and other factors (e.g. diet). These studies can be used to predict relationships of interest. It is important to note that this type of research does not tell the reader about causation.

(3) *Experimental research* includes comparisons between two or more factors, and is used for determining causality.

The statistical tests that can be employed in a study are determined by the research design. The latter dictates what set of guidelines the authors should follow in order to report their findings. For example, if you are reading an experimental randomized controlled trial, this probably follows the Consolidated Standards of Reporting Trials and checklist (Butcher *et al.* 2022). These guidelines are akin to the rules of a game, and provide a standard framework to ensure structural uniformity, fairness and transparency.

Clinician application

Match the clinical question you aim to answer through reading a research article to the appropriate study design. If you want to know more about a particular condition, consider a descriptive research study. If you are more interested in identifying an efficacious intervention, consider an experimental research study.

Participants

Information about the participants in a study helps you to determine how applicable the findings are to a specific patient or clinical population. This section should describe how, when, where and how many participants were recruited (Portney & Watkins 2015) (see Table 2). Additionally, it should include the criteria used for participant selection, which are commonly referred to as the inclusion and exclusion criteria. Statements about informed consent and

institutional review board approval are sometimes included in this section as well, but can also be reported elsewhere in the methods. These statements reflect the researchers' intent to protect the participants. The how, when and where details provide information that helps to determine both the validity and applicability of the findings.

For example, let us say that you were going to conduct a study about the effects of running on mood. You decide to recruit participants from a local suburban running club on a Saturday morning by speaking to the group at the end of their run. Alternatively, consider the following recruitment strategy: you stand outside a small rural community centre on a Saturday morning, and talk to people individually as they enter the facility. The results from the first group of participants demonstrate a strong positive relationship between running and mood, but those from the second demonstrate no relationship between the two variables. Which set of results is correct? Both recruitment strategies introduce bias into the participant selection process, but do so in very different ways, and this should be considered when deciding if the findings apply to your patient or patient population. However, if a participant recruitment strategy reflects your patient demographic (e.g. you treat a lot of avid runners in a suburban location who like to wake up earlier and run in a group), then the results may be more clinically relevant for you. This also applies to the inclusion and exclusion criteria. If the clinical question is about cardiovascular considerations in females who are ageing, and the participants are healthy male university students, you probably want to search for a different study!

Perhaps most importantly, the sample size must be considered. Participant numbers in the rehabilitation literature are frequently small, which can lead to flawed results and conclusions. Inappropriate sample sizes can lead to two types of errors: (type I) finding an effect in a population when one does not exist; or (type II) finding no effect in a population even though one exists (Field 2018). The probability of committing a type I error is statistically represented by the alpha level (α), and is usually set at 0.05. The probability of committing a type II error is statistically represented by the beta level (β), and is usually set at 0.2 (Field 2018; Kamper 2022). This information is generally presented in the form of an *a priori* power analysis, a calculation that uses this information, in addition to other information such as effect size, to determine the sample size needed in each study to maximize the chance of finding an effect if one exists. Type I errors are considered more serious. As mentioned above, type II errors are more common in the rehabilitation literature, usually because of the small sample sizes involved, and lead to a study lacking enough power to find a significant effect even if one exists. However, if sample sizes are extremely large (e.g. as in epidemiological studies), differences may be found that are too small to be clinically relevant. For additional clinically applicable information on sample size, refer to Kamper (2022).

Clinician application

Make sure that the condition or population that you are interested in is represented by the sample used. Check to see if the final number of participants included in a study met the calculated *a priori* threshold. If not, try to decipher why and interpret the results cautiously.

Instrumentation

This section provides information about the tools used to conduct a study. There are two types of instruments that should be described: physical tools and patient-reported outcome measures (Portney & Watkins 2015) (see Table 3). Physical tools and patient-reported outcome measures are sometimes described together under the uniform heading of "instrumentation"; however, they can also be separated in a variety of ways. For example, a research article may have one heading for instrumentation and another for self-reported outcome questionnaires (see Lewis *et al*. 2021). An alternative organizational structure is to provide a heading for each individual tool.

Physical tools include any piece of equipment used to conduct a study. The specific brand name of the tool used, and the company name and location of its headquarters should be listed. The validity and reliability of the tool should also be discussed if this information is available. If not, many authors will provide a statement such as, "While this patient-reported outcome measure has been validated in patients with pelvic pain caused by vaginismus, its validity has not been established for patients with pelvic pain related to endometriosis." This type of statement promotes transparency and alerts the reader to consider this limitation when interpreting the results. Patient-reported outcome measures should include a detailed description of the measure,

Table 3. Types of instrumentation (Portney & Watkins 2015): (MCID) minimal clinically important difference; and (MDC) minimal detectable change

Type of instrumentation	Examples	Information to be included about each tool*
Physical tool	Hand-held dynamometer	Brand name and model number
	Perineometer	Name and location of company
	Instrumented speculum	Validity
	Magnetic resonance imaging	Reliability (inter- and intra-rater, test–retest, and internal consistency)
	Electromyographic biofeedback	
	Rehabilitative ultrasound imaging	MCID
	Force plates	MDC
Patient-reported outcome measure	International Consultation on Incontinence Questionnaire – Urinary Incontinence Short Form	Description
	Pelvic Floor Distress Inventory Questionnaire – 20	Scoring
	Numerical Pain Rating Scale	Validity (face, content, construct and criterion)
	International Hip Outcome Tool	Reliability (inter- and intra-rater, test–retest, and internal consistency)
		MCID
		MDC

*All items listed may not be available for a given tool.

how it is scored and the related psychometric properties. If an experimental study found large improvements in a particular patient-reported outcome measure, but used tools (i.e. a physical or patient-reported outcome measure) that were not valid or reliable, the results are meaningless. This holds true for clinical practice as well, and highlights the importance of implementing validated and reliable objective measures and patient-reported outcome measures, as opposed to ones that are self-designed.

Although a deep dive into validity and reliability is beyond the scope of the present commentary, a brief overview follows. Reliability is the extent to which a measure is repeatable and free from error, whereas validity indicates that a test measures what it sets out to gauge (Kamper 2019). Validity cannot exist without reliability, but reliability can exist in the absence of validity. The classic example is a bathroom scale. If someone weighing 66 kg steps on a scale 10 times and the reading is 66 kg each time, then the measurement from the scale is reliable. If the scale is properly zeroed, then the measurement from the scale is also valid. However, what if the scale is measuring pounds, but displaying kilograms as the unit of measurement? The readings will all still be the same if the same person steps on the scale, but the scale no longer provides accurate information about weight measured in kilograms. In this scenario, the scale is reliable but not valid.

Clinician application

If you are trying to clinically implement a new tool or appraise a current tool, make sure that it is both reliable and valid. When reading a study, use this section for guidance on the tools that you may wish to implement. Additionally, use the information in this section to frame how you interpret the study results.

Procedures

To return to the analogy of a game used above (see "Research design"), the procedures serve as the general directions that ensure that everyone involved plays it the same way every time. This section will provide you with an explanation of what occurred and in what order, along with enough detail to allow the study to be replicated. The order in which events occurred and detailed descriptions of all events also allow you to compare findings across studies. This is especially helpful if studies of your clinical topic of interest have conflicting results. See Table 4 for examples of items that should be explained (Portney & Watkins 2015). Differences in any of these items may account for variations between studies.

Clinician application

If you find conflicting or mixed results in the literature when attempting to answer a clinical question, then the procedures section is a good place to check for variations between the results of different studies. This may help to clarify why studies investigating similar questions and topics differ in terms of their findings. You should also consider whether the procedures used in a study are similar to or different from how you practice clinically.

Data analysis

This final component of the methods section details the plan for both the descriptive and

Table 4. Procedure items (Portney & Watkins 2015) and related considerations for standardization

Procedure items	Examples
Participant/patient	Positioning Cueing and directions provided Clothing worn
Testing and treatment (if intervention)	Setup: marker placement (e.g. motion capture) height of equipment (e.g. bike seat) Timing/duration When? Who administered it? How were they trained? Who provided the training? Was the training standardized? Were they blinded? If there were multiple testers, how was reliability established? Repetitions/sets/frequency If multiple repetitions were tested (e.g. for strength testing), how many? Was the mean used? Lowest score? Highest score?
Data collection	Who performed it? How were they trained? Who provided the training? Was the training standardized? Were they blinded? How was it collected (paper versus electronic)? How frequently was it collected? When was it collected? How was it stored (e.g. lock secured or encrypted)?

inferential statistical analyses (Field 2018) (see Table 5). Descriptive statistics explain and summarize the characteristics of a data set (e.g. participant height). Inferential statistics use hypothesis testing to make conclusions or predictions about a population (e.g. individuals with pelvic organ prolapse) based on the representative sample used in the study. Planning the analyses to be used before data are collected, and sticking to this strategy, helps to minimize *P*-value hacking or data dredging. While slightly different techniques, *P*-value hacking and data-dredging are both poor research practices in which the data and/or the statistical analyses are manipulated if the desired results are not obtained from the originally planned assessments. Other items reported in this section are the type of software (i.e. model and version) used to perform the analyses, the acceptable significance level (usually $P \leq 0.05$) and how missing data will be handled.

In addition to reporting the plan, this component of the methods must provide enough information to allow you to determine if the statistical tests selected fit the data. Different statistical analyses can only be performed on certain types of data (i.e. nominal, ordinal, interval and ratio), and when certain criteria have been met (i.e. parametric versus non-parametric). For example, if a researcher compares the mean value of hip flexion (interval data) between two groups after the delivery of an intervention designed to increase range of motion, and the data meet all the assumptions of parametric testing (i.e. linearity, homoscedasticity, independence and normality),

Table 5. Types of statistics (Field 2018)

Type of statistics	Examples
Descriptive	Measures of central tendency (mean, median and mode) Measures of variability (range, variance and standard deviation) Measures of distribution (histograms, box plots and scatter plots)
Inferential	Correlation coefficient (Pearson's *r* and Spearman's ρ) Regression (linear, logistic and multivariate) Analysis of variance Factor analysis χ^2 test

then an independent samples *t*-test would be an appropriate statistic. However, if the parametric assumptions were violated, the non-parametric Mann–Whitney *U*-test should be used instead. Many data analysis plans will include all this information. Using the previous situation as an example:

> "Independent *t*-tests were used to examine differences in hip flexion range of motion between groups following the intervention. If parametric assumptions were not met, a non-parametric Spearman's ρ correlation analysis was used."

Informing readers of plan A and plan B helps to improve transparency.

Clinician application

It is all right if you are unfamiliar with the terms and statistical rules used here. However, if information about the statistical plan is missing,

or the findings sound too good to be true, your alarm bells should ring. In that instance, consider reaching out to a colleague who is more familiar with statistics for assistance in interpreting the information presented.

Conclusion

According to Dawes *et al.* (2005, p. 4):

> "Evidence-Based Practice [. . .] requires that decisions about health care are based on the best available, current, valid and relevant evidence. These decisions should be made by those receiving care, informed by the tacit and explicit knowledge of those providing care, within the context of available resources."

These three pillars are all important and interrelated. The ability to translate published research into clinical practice relies on the clinician's ability to effectively interpret the literature. This commentary is intended to serve as a resource bridge. The next time that you go to the literature for assistance with a clinical question, do not skip from an article's abstract to its results. Take time to stop at the methods and see where the road map leads.

Acknowledgements

The present author would like to thank Dr Cara Lewis for her invaluable help with concept refinement and technical support, and Dr Andrea Wood for her technical assistance.

Ethical approval

Ethical approval from the institutional review board was not obtained because this commentary did not include the use of human subjects.

Funding

This project was not supported by funding.

Conflicts of interest

The present author does not have any conflicts of interest to declare.

References

Butcher N. J., Monsour A., Mew E. J., *et al.* (2022) Guidelines for reporting outcomes in trial reports: the CONSORT-Outcomes 2022 extension. *JAMA* **328** (22), 2252–2264.

Dawes M., Summerskill W., Glasziou P., *et al.* (2005) Sicily statement on evidence-based practice. *BMC Medical Education* **5** (1), 1. DOI: 10.1186/1472-6920-5-1.

Field A. (2018) *Discovering Statistics Using IBM SPSS Statistics*, 5th edn. SAGE Publications, Thousand Oaks, CA.

Kamper S. J. (2019) Reliability and validity: linking evidence to practice. *Journal of Orthopaedic & Sports Physical Therapy* **49** (4), 286–287.

Kamper S. J. (2022) Sample size: linking evidence to practice. *Journal of Orthopaedic & Sports Physical Therapy* **52** (8), 563–564.

Lewis C. L., Halverstadt A. L., Graber K. A., *et al.* (2021) Individuals with pre-arthritic hip pain walk with hip motion alterations common in individuals with hip OA. *Frontiers in Sports and Active Living* **3:** 719097. DOI: 10.3389/fspor.2021.719097.

Martínez-Camblor P., Pérez-Fernández S. & Díaz-Coto S. (2019) The role of the *p*-value in the multitesting problem. *Journal of Applied Statistics* **47** (9), 1529–1542.

Portney L. G. & Watkins M. P. (2015) *Foundation of Clinical Research: Applications to Practice*, 3rd edn. F. A. Davis Company, Philadelphia, PA.

Jenny LaCross PT DPT PhD ATC CLT (she/her) completed her doctorate in physical therapy at Texas Woman's University in June 2023. She is a board-certified clinical specialist in women's health physical therapy, and works as adjunct faculty for South College and University of Michigan – Flint in their respective clinical doctorate of physical therapy programmes. Having completed her PhD, Jenny will begin a postdoctoral research fellowship at University of Michigan – Ann Arbor with the pelvic floor research group. As a clinician scientist, she understands the difficulty of translating research into practice, and aims to help other clinicians bridge this gap and improve the delivery of care.