International Journal of Nursing Studies 73 (2017) 17–23

Contents lists available at ScienceDirect

International Journal of Nursing Studies

journal homepage: www.elsevier.com/locate/ijns

Clinical significance in nursing research: A discussion and descriptive analysis

Denise F. Polit[a,b,*]

[a] Humanalysis, Inc.,75 Clinton Street, Saratoga Springs, NY, 12866 USA
[b] Centre for Health Practice Innovation, Griffith University, 170 Kessels Road, Nathan, Brisbane, QLD, 4111 Australia

ARTICLE INFO

Keywords:
Change scores
Clinical significance
COSMIN
Effect size index
Evidence-based practice
Minimal important change
Minimal clinically important difference
Reliable Change Index
Responder analysis
Statistical significance

ABSTRACT

Background: It is widely understood that statistical significance should not be equated with clinical significance, but the topic of clinical significance has not received much attention in the nursing literature. By contrast, interest in conceptualizing and operationalizing clinical significance has been a "hot topic" in other health care fields for several decades.

Objectives: The major purpose of this paper is to briefly describe recent advances in defining and quantifying clinical significance. The overview covers both group-level indicators of clinical significance (e.g., effect size indexes), and individual-level benchmarks (e.g., the minimal important change index). A secondary purpose is to describe the extent to which developments in clinical significance have penetrated the nursing literature.

Methods: A descriptive analysis of a sample of primary research articles published in three high-impact nursing research journals in 2016 was undertaken. A total of 362 articles were electronically searched for terms relating to statistical and clinical significance.

Results: Of the 362 articles, 261 were reports of quantitative studies, the vast majority of which (93%) included a formal evaluation of the statistical significance of the results. By contrast, the term "clinical significance" or related surrogate terms were found in only 33 papers, and most often the term was used informally, without explicit definition or assessment.

Conclusions: Raising consciousness about clinical significance should be an important priority among nurse researchers. Several recommendations are offered to improve the visibility and salience of clinical significance in nursing science.

What is already known about the topic?

- In research, statistical significance is achieved when the results of a statistical test have a low probability of being spurious. Statistical significance does not guarantee that results will be clinically significant—that is, that they could have genuine, palpable effects on patients' health or on health care decisions made on their behalf.
- Researchers in medicine and other health care fields have made substantial progress in defining and operationalizing clinical significance in the past few decades.

What this paper adds

- In the nursing literature, there has not been much coverage of conceptual and operational breakthroughs on clinical significance. This paper provides an overview of statistical indexes that can be used as benchmarks in drawing conclusions about clinical significance, most often about clinically significant change.
- Group-level benchmarks include effect size indexes and number needed to treat. At the individual level, clinical significance is most often operationalized using an index of whether a person's change score on an outcome corresponds to clinically meaningful improvement or deterioration—the minimal important change (MIC) index.
- The recent advances in measuring clinical significance have not penetrated to any great extent in nursing. In a sample of 261 articles reporting quantitative analyses published in three high-impact general nursing research journals in 2016, it was found that the topic of clinical significance was mentioned in only 33 papers, usually informally. A change score benchmark for individual study participants was referenced in only four papers.

1. Introduction

Nurse researchers and students learning research methods are

* Correspondence to: Humanalysis, Inc 75 Clinton Street, Saratoga Springs, NY 12866, USA.
E-mail address: denisefpolit@gmail.com.

http://dx.doi.org/10.1016/j.ijnurstu.2017.05.002
Received 19 January 2017; Received in revised form 2 May 2017; Accepted 4 May 2017

routinely told that statistical significance does not mean that quantitative study results are clinically important. For example, nearly 40 years ago, Polit and Hungler (1978) stated the following in the first edition of their research methods textbook: "The fact that statistical significance was attained in testing the hypothesis does not necessarily mean that the results were important or of value to the nursing community and their clients" (p. 600). Polit and Hungler did not, however, offer advice on how to assess the clinical significance of results.

The purpose of this paper is to provide an overview of recent advances in defining and operationalizing clinical significance. A secondary purpose is to explore the degree of penetration of these advances in the nursing literature. First, however, statistical significance is briefly discussed to provide context.

2. Statistical significance

Statistical significance is the research term used to indicate that the results of a statistical analysis probably are not attributable to chance or serendipity, at a specified level of probability. By convention, researchers typically accept that a relationship or group difference is likely to be "real" if the probability of a spurious result is less than 0.05 (or, less often, 0.01). Many commentators have criticized this threshold as arbitrary (e.g., Greenland et al., 2016; Sterne and Smith, 2001), but it likely will continue as the standard criterion used within the standard hypothesis-testing framework.

An important reason for not equating statistical significance with clinical importance is that statistical significance is strongly affected by sample size. With a large sample, statistical power is high and the risk of a Type I error (wrongly concluding there is no relationship between variables) is low. For example, with a sample size of 500, a modest correlation of $r = 0.10$ is statistically significant at $p < 0.05$, even though such a weak relationship may have little practical importance.

The "bias against the null hypothesis" (e.g., Conn et al., 2003; Song et al., 2010) reflects decisions on the part of authors, reviewers, and editors to not publish papers when results are not statistically significant. The risk of nonsignificant results can be reduced by undertaking a priori power analyses. Perhaps as a result of enhanced awareness of power analysis, sample size appears to be increasing in published nursing studies. For example, Polit and Sherman (1990) found that 4% of nursing studies had samples of 500 or more, compared to 27% of samples in the recent study by Gaskin and Happell (2014). In the 1990 analysis, the average power to detect moderate effects in nursing studies was only 0.71, compared to 0.98 in the 2014 study.

Nursing studies may be less likely to be underpowered today than they were decades ago, but an unintended consequence might be that more published studies have effects that are too modest to be meaningful clinically. With the recent emphasis on evidence-based practice, it is important for nurses to base their practice on evidence that is not only "real" (in the statistical sense) but also clinically important. The question facing researchers is how to ensure that both standards are met.

In 1993, LeFort (LeFort, 1993) wrote an excellent article in a nursing journal about the distinction between statistical and clinical significance. However, that article appears not to have had much of an impact on nursing research. LeFort commented on the "recent interest" in the topic of clinical significance, but that interest has developed primarily in health disciplines other than nursing.

3. Clinical significance

For nearly three decades, medical researchers and clinical epidemiologists have grappled with how to define and measure clinical significance. This interest likely reflects the fact that patient-reported outcomes (e.g., pain, quality of life) are playing an increasing role in evaluating the effectiveness of medical and pharmacologic interventions. Clinicians can grasp whether changes in, say, blood pressure or cholesterol have clinical importance. But interpreting the clinical relevance of a change of, say, 4 points on a quality of life scale is challenging.

Despite the efforts of many in the medical field to define and quantify clinical significance, no consensus on either front has been reached. Moreover, consensus may be unachievable because of the complexity of the issue. For example, in some cases clinical significance may be defined in terms of the degree of *improvement* experienced by patients (e.g., the amount of pain reduction among patients with cancer). In other cases, a clinically significant outcome might be the *absence* of change (e.g., the absence of deterioration in spirometry test values among patients with idiopathic pulmonary fibrosis). Another thorny issue is that the amount of change that would be clinically significant might depend on what the "baseline" value is. For example, a modest pain reduction (e.g., a mean change of 2 points on a numeric 0-to-10 point rating scale) might be more meaningful to those with average initial pain scores of 5.0 than to those with an average initial score of 9.0. Thus, there will probably never be an analog to $p < 0.05$ for deciding on the clinical significance of study results.

An important distinction for evaluating clinical significance concerns group-level versus individual-level approaches.

3.1. Group-Level clinical significance

Statistical significance is a conclusion based on group results—for example, the statistical significance of a mean group difference on an outcome for an intervention group versus a control group. Clinical significance is sometimes evaluated based on group-level statistical indexes. The most often-used indicators of clinical significance at the group level are effect size indexes, confidence intervals, and number needed to treat (often abbreviated as NNT).

Effect size indexes (such as the *d* statistic for comparing group means) summarize the magnitude of effects and can be used to evaluate how much a group, on average, can benefit from a treatment. The Publication Manual of the American Psychological Association (APA), the style manual used by many nursing journals, encourages the inclusion of effect size information in research articles: "For the reader to appreciate the magnitude *or importance* (emphasis added) of a study's findings, it is almost always necessary to include some measure of effect size in the Results section" (APA, 2010, p. 34). Reporting effect size estimates is also advocated in the 2010 Consolidated Standards of Reporting Trials or CONSORT (Moher et al., 2010).

Confidence intervals, which show the precision of estimates, are recommended by some as a tool for interpreting the clinical significance of group results. Fethney (2010), for example, presented an example of how confidence intervals were used to evaluate clinical significance in a randomized clinical trial testing weight gain interventions for premature infants. The APA manual (2010) and CONSORT (Moher et al., 2010) recommend presenting confidence intervals around effect size estimates.

Number needed to treat is an index favored by many clinicians because it is easy to understand in real-world situations (Kraemer et al., 2003). The number needed to treat represents the number of patients who would need to be treated with an intervention for 1 patient to derive benefit or avoid harm. For example, a number needed to treat of 3.0 means that about one out of every three patients would be expected to benefit from the treatment. A number needed to treat of 1.0 means that the intervention would be expected to result in positive effects for everyone who received it.

At the group level, as well as at the individual level, the operationalization of clinical significance shares a feature with statistical significance: they both involve using a benchmark (threshold) to reach a conclusion about significance. Without an a priori threshold, researchers rely on subjective judgments about whether their results are significant. For statistical significance, the traditional threshold is $p < 0.05$. For clinical significance at the group level, researchers