International Journal of Nursing Studies 51 (2014) 795–806

Contents lists available at ScienceDirect

International Journal of Nursing Studies

journal homepage: www.elsevier.com/ijns

Power, effects, confidence, and significance: An investigation of statistical practices in nursing research

Cadeyrn J. Gaskin [a,b], Brenda Happell [c,d,*]

[a] Gaskin Research, Melbourne, Victoria, Australia
[b] Central Queensland University, Institute for Health and Social Science Research, Centre for Mental Health Nursing Innovation, School of Nursing and Midwifery, Melbourne, Victoria, Australia
[c] Central Queensland University, Institute for Health and Social Science Research, Centre for Mental Health Nursing Innovation, Rockhampton, Queensland, Australia
[d] Central Queensland University, School of Nursing and Midwifery, Rockhampton, Queensland, Australia

ARTICLE INFO

Article history:
Received 6 January 2013
Received in revised form 30 September 2013
Accepted 30 September 2013

Keywords:
Research design
Nursing research
Review literature as topic
Statistics
Statistics as topic

ABSTRACT

Objectives: To (a) assess the statistical power of nursing research to detect small, medium, and large effect sizes; (b) estimate the experiment-wise Type I error rate in these studies; and (c) assess the extent to which (i) a priori power analyses, (ii) effect sizes (and interpretations thereof), and (iii) confidence intervals were reported.
Design: Statistical review.
Data sources: Papers published in the 2011 volumes of the 10 highest ranked nursing journals, based on their 5-year impact factors.
Review methods: Papers were assessed for statistical power, control of experiment-wise Type I error, reporting of a priori power analyses, reporting and interpretation of effect sizes, and reporting of confidence intervals. The analyses were based on 333 papers, from which 10,337 inferential statistics were identified.
Results: The median power to detect small, medium, and large effect sizes was .40 (*interquartile range* [*IQR*] = .24–.71), .98 (*IQR* = .85–1.00), and 1.00 (*IQR* = 1.00–1.00), respectively. The median experiment-wise Type I error rate was .54 (*IQR* = .26–.80). A priori power analyses were reported in 28% of papers. Effect sizes were routinely reported for Spearman's rank correlations (100% of papers in which this test was used), Poisson regressions (100%), odds ratios (100%), Kendall's tau correlations (100%), Pearson's correlations (99%), logistic regressions (98%), structural equation modelling/confirmatory factor analyses/path analyses (97%), and linear regressions (83%), but were reported less often for two-proportion *z* tests (50%), analyses of variance/analyses of covariance/multivariate analyses of variance (18%), *t* tests (8%), Wilcoxon's tests (8%), Chi-squared tests (8%), and Fisher's exact tests (7%), and not reported for sign tests, Friedman's tests, McNemar's tests, multi-level models, and Kruskal–Wallis tests. Effect sizes were infrequently interpreted. Confidence intervals were reported in 28% of papers.
Conclusion: The use, reporting, and interpretation of inferential statistics in nursing research need substantial improvement. Most importantly, researchers should abandon the misleading practice of interpreting the results from inferential tests based solely on

* Corresponding author at: Central Queensland University, Institute for Health and Social Science Research, Centre for Mental Health Nursing Innovation, Rockhampton, Queensland, Australia. Tel.: +61 409756775.
E-mail address: b.happell@cqu.edu.au (B. Happell).

0020-7489/$ – see front matter. Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijnurstu.2013.09.014

whether they are statistically significant (or not) and, instead, focus on reporting and interpreting effect sizes, confidence intervals, and significance levels. Nursing researchers also need to conduct and report a priori power analyses, and to address the issue of Type I experiment-wise error inflation in their studies.
Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.

What is already known about the topic?

- A review conducted over two decades ago showed that nursing studies were sufficiently powered to detect large effect sizes only.
- The findings of a subsequent review of recent research in one mental health nursing journal suggest that the statistical power in nursing studies may not have changed during this time period.

What this paper adds

- The findings suggest that a priori power analysis does not appear to be routine practice when designing nursing studies.
- This study highlights the poor control of experiment-wise Type I error in nursing research.
- The findings show that researchers who publish work in nursing journals rarely report effect sizes for some statistical tests and infrequently interpret them when they do.

1. Introduction

Nurses are increasingly expected to ground their practices within evidence-based frameworks (Polit and Beck, 2012). To maintain and enhance their professional practices, nurses are strongly encouraged to read research literature, which invariably means having to become familiar with inferential statistics (Moore, 2011). Nurses are not only coming to depend on research for improving their practices, they also rely on researchers correctly using inferential statistics and appropriately interpreting the results they produce. Pertinent issues in this regard are whether studies have sufficient statistical power and whether researchers provide ample information (e.g., effect sizes, confidence intervals) to facilitate accurate interpretations of findings.

Statistical power is the long-term probability of rejecting the null hypothesis when it is false (Cohen, 1992b); it is the likelihood that an effect, if present, will be statistically significant. Three factors influence the power of a statistical test: the sample size (N), the significance criterion (α), and the population effect size (incorporating both the magnitude of the difference or relationship between outcome measures and the variability within each of the measures). Given the interplay between these factors, one of the main tasks when designing a study is to calculate how many participants will need to be recruited (or how large datasets need to be) for effects of expected magnitudes to be statistically significant. Unfortunately, evidence of a priori power analyses being conducted in published studies is rare (Ferrin et al., 2007; Kosciulek and Szymanski, 1993). In nursing, a review of the 2010 and 2011 volumes of the *International Journal of Mental Health Nursing* found that a priori analyses were only reported in 17% of studies (Gaskin and Happell, 2013). It is unknown whether this limited attention paid to a priori power analyses is widespread across nursing research.

Conducting multiple tests can also affect the power within a study. Multiple tests using the same dataset increases the risk of Type I errors (Cohen, 1990). If only one test is conducted, with α set at .05, then the probability that a significant result will be obtained purely by chance is also .05. When multiple independent tests are performed, however, the experiment-wise (across the whole study) error rate increases as demonstrated in the following equation: $1-(1-\alpha)^N$, where N is the number of tests. If 10 tests are performed, for example, the likelihood of finding a significant result by chance is .40, which is much greater than the .05 many researchers might expect they are using. Fortunately, several strategies (e.g., modified Bonferroni procedures; Hochberg, 1988; Hochberg and Benjamini, 1990; Hommel, 1988) have been developed to address the problems that arise through using multiple tests. Many strategies come at the cost of statistical power, however, which means that, if using them, researchers would need to recruit more participants for effects of a given magnitude to remain statistically significant. Based on the findings of the review in the *International Journal of Mental Health Nursing*, it would seem that nursing researchers typically do not control Type I error in their studies and, as a consequence, the experiment-wise error rate is grossly inflated (Gaskin and Happell, 2013). To make more conclusive statements about the practices of nursing researchers, a broader selection of journals in this discipline would need to be reviewed.

Analyses of the power within studies to detect effect sizes of certain magnitudes have been undertaken in several fields including applied psychology and management (Mone et al., 1996), behavioural ecology and animal behaviour (Jennions and Møller, 2003), health psychology (Maddock and Rossi, 2001), rehabilitation counselling (Kosciulek and Szymanski, 1993), and sport science (Speed and Andersen, 2000). Studies have also been conducted in nursing (Polit and Sherman, 1990), mental health nursing (Gaskin and Happell, 2013), and paediatric nursing (Beck, 1994). In a review of nursing studies conducted over 20 years ago, the power to detect small, medium, and large effects was .26, .71, and .95, respectively (Polit and Sherman, 1990). A recent study in mental health nursing yielded similar findings, with the power to detect small, medium, and large effects being .34, .79, and .94, respectively (Gaskin and Happell, 2013). With current conventions in the social sciences suggesting that researchers aim for power of .80 (Cohen, 1992a), nursing studies conducted over 20 years ago were, on average, sufficiently powered to detect large effect sizes, and

Download English Version:

https://daneshyari.com/en/article/7515989

Download Persian Version:

https://daneshyari.com/article/7515989

Daneshyari.com