Available online at www.sciencedirect.com

ScienceDirect

Currents in Pharmacy Teaching and Learning 8 (2016) 83–89

Currents in Pharmacy Teaching & Learning

http://www.pharmacyteaching.com

Review article

Practical significance: Moving beyond statistical significance

Michael J. Peeters, PharmD, MEd, FCCP, BCPS*

College of Pharmacy and Pharmaceutical Sciences, University of Toledo, Toledo, OH

Abstract

Practical significance is an important concept that moves beyond statistical significance and *p* values. While effect sizes are not synonymous with practical significance, it is a basis for evidence of substantive significance. Investigators should find and report effect sizes whenever possible. To build evidence for practical significance in pharmacy education, three methods are discussed. First, effect sizes can be compared to *general interpretation guidelines* for practical significance. Second, using the effect sizes, investigators can *benchmark* by comparing effect sizes to external information from other studies; however, this information is not always available. Where prior data is limited, a third method after determining effect size is for investigators to calculate in their cohort an instrument's minimally important difference; the effect size could be compared to this *minimally important difference*, as opposed to a general interpretation guideline. A method to calculate the minimally important difference is described, as well as applications. Regardless, effect sizes must be determined and should be reported in articles; its comparator may vary as evidence for practical significance—so interpretation is key. Reporting effect sizes can enable benchmarking by others in the future and facilitate summaries through meta-analysis. It is clear that reporting evidence of practical significance with effect sizes is needed; simply reporting statistical significance is not enough. After reading this article, readers should be able to explain practical significance, recognize evidence of practical significance in other reports, and carry out their own analysis of practical significance using one or more of the methods described herein.

Keywords: Significance; Statistical significance; Practical significance; Effect size; Pharmacy education; Learning assessment

Situation

Finding practical significance is essential for investigating meaningful educational interventions and is often a more stringent criterion than statistical significance. One example of the importance of this distinction comes from a recent project in which faculty at our institution investigated the development of critical thinking among our PharmD students. Development of critical thinking is a broadly accepted goal of higher education,[1,2] including health professions education.[3,4] To measure development of critical thinking of PharmD students in our College of Pharmacy, we introduced periodic (approximately annual) longitudinal assessments using standardized critical thinking tests.

The director of educational assessment of the college asked that, as opposed to prior critical thinking studies, we move beyond evaluating statistically significant differences with these assessment findings. (*Note*: This director of educational assessment is already familiar with specific results and comparisons provided with standardized tests such as the Pharmacy Curriculum Outcomes Assessment and American Association of Colleges of Pharmacy surveys for PharmD programs.) Do these critical thinking tests show meaningful, practical measurement differences? If so, this College of Pharmacy may better evaluate whether the current curriculum is meaningfully helping to foster students' critical thinking development, and if not, where might curricular revisions be targeted?

* Correspondence to: Michael J. Peeters, PharmD, MEd, FCCP, BCPS, College of Pharmacy and Pharmaceutical Sciences, University of Toledo College of Pharmacy & Pharmaceutical Sciences, 3000 Arlington Ave, Mail Stop 1013, Toledo, OH 43614.

E-mail: michael.peeters@utoledo.edu

http://dx.doi.org/10.1016/j.cptl.2015.09.001

1877-1297/

Aside from this scenario, another example of the importance of distinguishing between statistical and practical significance comes from looking at a locally created annual assessment (i.e., progress examinations) of pharmacy knowledge among first-year, second-year, and third-year PharmD students and assessing whether students' knowledge development is fostered by a curriculum. The distinction might also be important if an educator sought to investigate a teaching and learning method such as a flipped classroom and evaluate students' learning using the course's final examination and comparing these results to those of last year's class.

Methodological literature review

Some statistically significant comparisons are too small for practical significance[5]; that is, some statistically significant results can be too inconsequential for meaningful, practical impact. Statistical significance is related to sample size, and p values are a common index used. Investigations with very large samples can detect very small differences, statistically, in an outcome between groups.[5] Another method to coax a statistically significant difference is to use an outcome measure with a large scale; larger scales can more easily show differences—but differences may not be meaningful either. The medical literature has numerous examples of statistically significant results that have questionable clinical significance, such as the use of topical diclofenac to improve pain control for osteoarthritis of the knee.[6] In this study, a quality-of-life instrument showed a statistically significant improvement over placebo for patients using topical diclofenac, though the change was so small numerically that other investigators have questioned its practical significance.[6]

Practical significance is contrasted with statistical significance. Practical significance in different contexts can be synonymous with substantive or clinical significance.[5,7–9] Unlike statistical significance that is simply a determinant of an inferential statistical test and has a formulaic process for interpretation using null-hypothesis significance testing (NHST), practical significance is less certain, and no single formulaic approach will always be best.[5] Teaching PharmD students about practical significance using a simple formulaic process is problematic as well. Evidence of practical significance can be sought using a few strategies. All use effect sizes. One strategy is to use effect sizes along with general interpretation guidelines. A second strategy uses effect sizes and benchmarking. A third strategy uses effect sizes and minimal important difference.

Using Method 1: General interpretation guidelines

Simply using effect sizes can suggest practical significance.[5–9] For example, Cohen's d is one popular effect size coefficient for quantifying a difference in education and other social sciences literature. It can be used to compare the means and standard deviations of two independent groups or two paired assessments from one cohort (e.g., pre- versus post-testing).[7] Cohen provided a general framework for interpreting these Cohen's d effect sizes [small (0.2), medium (0.5), and large (0.8)] and, similarly, for Pearson correlations [small (0.1), medium (0.3), and large (0.5)]; other significance test interpretations were provided as well.[10]

Using Method 2: Benchmarking

This is another strategy for determining practical significance. General interpretation guidelines have limits due to context. Thus, benchmarking uses comparative data from similar sources to set standards (i.e., expectations) based on the cumulative experiences of others. Benchmarking can provide real-world comparison instead of relying on more generic or distribution-based effect size interpretations, or even just using statistical significance. To this end, Hill et al.[11] provide an example of benchmarks for comparison within an educational setting, albeit for kindergarten through 12th grade and not focused on higher education; as such, this benchmarking reference may provide limited help. Another more rigorous example of benchmarking in pharmacy education (and briefly mentioned in this article's initial "situation") is the Pharmacy Curriculum Outcomes Assessment (PCOA), which fosters broad comparison with other colleges/schools of pharmacy. The PCOA is not without limitations either, as pharmacy students may not have fully completed all coursework included on the PCOA when they sit for this examination. The PCOA can also be limited by the broad contexts pooled from many different institutions, and so a smaller group of similar peer institutions may facilitate improved comparisons.

Using Method 3: Minimal important difference

In situations where further evidence of practical significance is desired beyond Method 1, and in the absence of literature or better context specifics with Method 2, a third method may be used. It uses standard error of measurement (SEM)[a] and is a statistical, distribution-based method to determine the *minimal important difference*.[12] Of note from the health measurement literature, use of *SEM* has been associated with clinical significance[12–14]; for health measurement instruments, the concept of statistical versus clinical significance has been investigated, and practical significance has been termed the *minimal clinically important difference*. For a multitude of health measurement instruments that will all use different scales, such as the

[a]This SEM should not be confused with "SE" (standard error) reported within many statistical programs; SEM is *different* from standard error of the mean (even though each can use SEM as their acronym).[15] Standard deviation, the standard error of the mean, the standard error of the estimate, and the standard error of the measurement are all conceptually distinct though related issues.