Supplement to Chapter 21

Clinical Significance Assessment With the Jacobson–Truax Approach

In this Supplement, we discuss the approach to clinical significance developed by Jacobson and Truax (1991). This approach—sometimes called the **J-T approach—**has been widely used in the field of psychotherapy but was not adopted in other health fields until recently. This approach appears to be gaining ground in medical research.

In an important paper in the health literature that compared the utility of different approaches to establishing benchmarks for clinical significance, Beaton and colleagues (2011) described the J-T method as a "combination" method. Similarly, Mann and colleagues (2012) referred to this approach as a "hybrid" approach. The terms used in both papers reflect the fact that the J-T approach is a two-step process that relies on an assessment of both measurement error and a change from a dysfunctional to a functional state. This is different than the "triangulation" efforts we discussed in the book, which involve taking multiple pieces of information and integrating them into a single estimate of the minimal important change (MIC). Using the J-T approach, a person's change score on an outcome has to pass two tests to be considered clinically significant.

STEP 1: RELIABLE CHANGE

As discussed in Chapter 15 of the textbook, the **reliable change index (RCI)** is a method of evaluating the reliability of a change score and is similar to another index, the *smallest detectable change (SDC).* Jacobson and his colleagues argued that, to be clinically meaningful, a change score on an outcome measure must pass the test of being "real"—that is, a change beyond measurement error. They proposed the RCI as the standard for real change.

Calculation of the RCI

The RCI requires the calculation of the *standard error of measurement* (*SEM*) for the focal measure. We noted in Chapter 15 that the SEM can be computed using information about the measure's reliability. Specifically, the formula for the SEM is:

$$\text{SEM} = SD\sqrt{(1 - R)}$$

The *SD,* in the context of the RCI, is the standard deviation of a "control group, normal population, or pretreatment experimental group" (Jacobson & Truax, 1991, p. 14).

There has been a debate about which index to use as *R* in this SEM formula. Jacobson and Truax (1991) originally used a test–retest reliability coefficient, which is the coefficient advocated by the COSMIN group (DeVet et al., 2011) and by Polit and Yang (2016). Others, however, have used coefficient alpha as the estimate of reliability for computing the SEM.

Once the SEM has been estimated, the RCI "value-to-exceed" to classify a change score as

probably real, at the 95% confidence level, can be computed with the following formula:

$$RCI = 1.96 \times \sqrt{2 \times SEM^2}$$

Suppose, for example, the SEM for an outcome measure was 2.0. The RCI would be the square root of 8 ($2 \times 2^2 = 8$) times 1.96, or $2.83 \times 1.96 = 5.54$. In this case, a person's change in score greater than 5.54 (e.g., from baseline to a postintervention follow-up) would be deemed a reliable change, according to the RCI. Once a person's change score has been classified as probably real, the second part of the J-T approach can be applied.

The 95% CI Standard for the RCI

Before going on to describe the second step of the J-T approach, we pause briefly to mention some debate about the RCI. The detection of reliable change based on the standard RCI (or the SDC) relies on a strict criterion for a change as "probably real"—a 95% CI. Several writers in the field of measurement have noted that the 95% CI sets a very stringent standard for individual-level change. For example, Cella and others participating in a clinical significance meeting group (2002) commented that "a reliable change index is conservative, only allowing classification as changed if the change exceeds a degree that would *only rarely* be due to chance" (p. 386, emphasis added).

A study by Jordan and colleagues (2006) illustrates how such a stringent standard can result in seemingly anomalous results. These researchers studied changes on the Roland Morris Disability Questionnaire (RMDQ) among patients with low back pain at baseline and 6 months after a treatment regimen. The researchers found that nearly two-thirds of the patients who described themselves as "completely recovered" on a global rating scale (GRS) for back pain at the 6-month point did *not* meet 95% RCI criterion for reliable change on the RMDQ.

Because of concerns such as those by Jordon et al. (2006), several researchers have used less conservative standards than the 95% criterion. For example, several have adopted a 90% CI for calculating the SDC or RCI, which translates to using 1.65 rather than 1.96 in the RCI formula. In our example in which the SEM of a measure was 2.0, the 90% CI would translate to an RCI of 4.67 rather than 5.54—a more lenient threshold. And, as mentioned in the textbook, some researchers have proposed that a more liberal standard—1 SEM—be used in interpreting change scores. Wyrwich (2004) maintained that "the 1 SEM threshold is well beyond the necessary 'more than likely' or 51 percent level of confidence that change occurred at the individual level" (p. 586).

STEP 2: NORMALIZATION

The second part of the J-T approach to ascertaining a clinically significant change is based on the goal of helping patients who are not functioning at a "healthy" or "normal" level on an outcome to attain a "normal" or "functioning" state. This conceptualization arose within the context of psychotherapy research, where patients with a clinical diagnosis (e.g., clinical depression) are seeking to recover from that diagnosis. This view on clinical significance could also apply to several health problems.

The Normality Criterion

Jacobson and colleagues (1991, 1999) recognized the challenge of classifying patients based on the second criterion of "normalcy." They proposed three alternative methods of operationalizing whether a patient had achieved a "normal" classification, all based on an underlying conceptualization of movement from a dysfunctional population to a normally functioning population:

- *Method A.* The final score should fall outside the range of the dysfunctional population, defined as being at least two standard deviations (*SD*s) from the mean for that population, in the direction of a better outcome.
- *Method B.* The final score should fall within the range of the functional (normal) population, defined as being within two *SD*s of the mean for that population.
- *Method C.* The final score should place the person closer to the mean of the functional population than to the mean of the dysfunctional population.

Jacobson and colleagues concluded that when the distributions on the focal measure for the functional and dysfunctional populations overlap, then Method C is the best criterion. Method C is also a simple cutoff value to compute, when relevant norms are available:

$$\text{Cutoff}_C = \left(\text{Mean}_{\text{Dysfunctional}} + \text{Mean}_{\text{Functional}}\right) \div 2$$

When norms for a healthy population are not available, however, Method A is the only viable alternative. Using the combined criteria (RCI plus attainment of a final state of normality), the J-T approach results in classifying patients into one of three groups: (1) recovered (reliably changed and in a positive final state); (2) improved but not recovered (reliably changed, but failed to meet the final-state criterion); and (3) unchanged or not reliably changed.

Example of Using the J-T Approach

As described in the textbook, Da Mata and colleagues (2018) presented a fully worked-out example of applying the J-T approach. No norms were available for their measure of patient knowledge, and so they used Method A to establish a benchmark for clinical significance.

The J-T Approach Used in Health Studies

Beaton and colleagues (2011) undertook a comparative analysis of 13 approaches to setting clinical significance benchmarks for the Disability of the Arm, Shoulder, and Hand (DASH) scale—a measure of symptoms and functional status for patients with upper extremity musculoskeletal conditions. The researchers used diagnostic assessment methods to classify individuals undergoing physical therapy as "responders" or "nonresponders" to treatment, based on change scores on the DASH scale and using the 13 benchmarks. Their receiver operating characteristic (ROC) analysis used two patient-reported criteria for the responder classification: whether or not treatment goals were met, and whether or not there was important improvement, based on patients' rating on a GRS as "much better." The J-T approach was one of the 13 methods they tested. The researchers used J-T methods B and C based on normative data for the US general population on the DASH to define the boundaries of a healthy population. They noted that their study results "point to a combination of change greater than error and/or a final score within general population norms as being the most clinically sensible with strong diagnostic accuracy" (p. 487). Replication of their findings with other populations and other measures is needed, but their work in suggesting how to evaluate various clinical significance benchmarks is instructive.

TIP Each of the two steps of the J-T approach has been used independently by some nurse researchers to assess clinical significance—that is, the researchers used one step but not the other. For example, Bond and colleagues (2016) used the RCI value as their benchmark for clinical change in neurocognitive performance in patients with cancer, and Lindseth et al. (2014) used norms for measures of anxiety and depression as the threshold for clinical significance.

REFERENCES CITED IN CHAPTER 21 SUPPLEMENT

Beaton, D. E., van Erd, D., Smith, P., van der Velde, G., Cullen, K., Kennedy, C., & Hogg-Johnson, S. (2011). Minimal change is sensitive, less specific to recovery; a diagnostic testing approach to interpretability. *Journal of Clinical Epidemiology, 64*, 487–496.

Bond, S., Dietrich, M., Gilbert, J., Ely, E., Jackson, J., & Murphy, B. (2016). Neurocognitive function in patients with head and neck cancer undergoing primary or adjuvant chemoradiation treatment. *Supportive Care in Cancer, 24*, 4433–4442.

Cella, D., Bullinger, M., Scott, C., Barofsky, I., & the Clinical Significance Consensus Meeting Group, (2002). Group vs. individual approaches to understanding the clinical significance of differences or changes in quality of life. *Mayo Clinic Proceedings, 77*, 384–392.

Da Mata, L., Bernardes, M., Azevedo, C., Chianca, T., Pereira, M., & de Carvalho, E. (2018). Jacobson and Truax method: Evaluation of the clinical effectiveness of a home care program after prostatectomy. *Revista Latino-Americana de Enfermagem, 26*, e3003.

DeVet, H. C. W., Terwee, C., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide*. Cambridge: Cambridge University Press.

Jacobson, N. S., Roberts, L., Berns, S., & McGlinchey, J. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology, 67*, 300–307.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.

Jordan, K., Dunn, K., Lewis, M., & Croft, P. (2006). A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. *Journal of Clinical Epidemiology, 59*, 45–52.

Lindseth, G. N., Coolahan, S., Petros, T., & Lindseth, P. (2014). Neurobehavioral effects of aspartame consumption. *Research in Nursing & Health, 37*, 185–193.

Mann, B. J., Gosens, T., & Lyman, S. (2012). Quantifying clinically significant change: A brief review of methods and presentation of a hybrid approach. *The American Journal of Sports Medicine, 40*, 2385–2393.

Polit, D. F. & Yang, F. M. (2016). *Measurement and the measurement of change: A primer for health professionals*. Philadelphia: Lippincott.

Wyrwich, K. W. (2004). Minimal important difference thresholds and the standard error of measurement: Is there a connection? *Journal of Biopharmaceutical Statistics, 14*, 97–110.