The Significance of Clinical Change and Clinical Change of Significance: Issues and Methods

Kenneth J. Ottenbacher, Mary Beth Johnson, Margaret Hojem

Key Words: accounting • documentation • patient outcome assessment

The emphasis on systematic methods of demonstrating accountability in health care is affecting the delivery of human services at all levels, from hospitals and community centers to individual private practitioners. New procedures are being proposed to meet the accountability demands encountered by therapists. Several recently developed methods for documenting clinical change in patient status are presented, and the relationship of this documentation to traditional research methods is briefly explored. A clinical illustration of the procedures is provided along with a discussion of advantages and limitations.

Occupational therapists practice their profession in a dynamic and evolving health care environment. Therapists are expected not only to be aware of the changes occurring in health care, but also to participate in implementing and documenting change. Accountability and documentation are areas of contemporary health care practice that are receiving a great deal of attention from health care providers, policy makers, and consumers. In relation to occupational therapy, Gillette (1982) observed that “documentation of actual occupational therapy services, leading toward the development of a data base upon which to base the practice claims of the field, is the most effective means of establishing occupational therapy as an essential, and reimbursable service” (p. 499). Gillette (1982) went on to argue that the “established procedures” of scientific inquiry and clinical research represent the usual method of demonstrating efficacy and accountability.

Although traditional methods of clinical research can provide useful information regarding the efficacy of intervention strategies used by occupational therapists, behavioral scientists have recently questioned the exclusive reliance on traditional methods to demonstrate the therapeutic utility of many practice procedures (Kazdin, 1983; Ottenbacher, 1986; Strupp, 1981). Kazdin and Wilson (1978) argued that traditional studies evaluating outcome in clinically based fields infer treatment effects based on the statistical comparison between two or more groups or conditions. They contend that these studies have two primary limitations as methods of establishing therapeutic accountability. First, results of these studies are based on the average improvement scores for all subjects and provide no information on individual patient performance. Second, the criterion of statistical significance, conventionally used as the basis for making inferential judgments regarding empirical importance, has little clinical or practical relevance.

Underlying the statistical versus clinical significance debate is a fundamental “philosophical” difference that separates traditional clinical research and practice. The goal of the practicing therapist is, and must be, to help patients get better as quickly as possible. The nature, length, and intensity of any intervention is tailored to the individual patient’s disability (cf. Yeaton & Sechrest, 1981). Therapists in clinical settings will alter treatment strategies as soon as it seems reasonable and beneficial to the patient and will base this decision on continuous monitoring of the patient’s performance.

In traditional group comparison research, patients are randomly assigned to fixed treatment conditions. Treatment is standardized and administered continuously irrespective of progress (or lack of progress), and measurements of the patients’ performance...
are taken infrequently, generally at the beginning and end of the study. Finally, in traditional group comparison designs the researcher is usually unaware of an individual patient's progress or even which condition or group the subject may be in.

The primary limitation of many traditional research procedures, from the perspective of the clinician, is the lack of emphasis on the performance of individual patients. In day-to-day clinical practice, the therapist's primary concern is the function of the patient with a specific disability. The accountability demands of the clinical therapist are focused on individual patients and whether or not they have improved during the course of intervention (Barlow, Hayes, & Nelson, 1984; Bloom & Fischer, 1982).

Research procedures based on single-subject or time series designs have been proposed as one alternative to more traditional methods of establishing therapeutic effectiveness (Kazdin, 1982; Ottenbacher, 1986). For example, Levy (1981) observed that clinical research has had little impact on clinical practice in applied fields such as social work and nursing because an often inappropriate methodology is used to conduct clinical research. Levy stated that, "while group experimental studies are valuable to answer some questions (such as the extent of change across a group), single-subject designs offer several advantages for clinical research over these traditional group methods" (p. 235). However, not all questions of efficacy and accountability can be addressed by single-subject approaches. One common limitation is the requirement to collect information on patient performance on a repeated basis. Other limitations of single-subject approaches in clinical environments were explored by Levy (1981) and Kazdin (1978).

A wide variety of methodologies and procedures are required to meet the demands of clinical accountability. Flexibility and innovation must be encouraged if therapists are to establish a therapeutic science based on clinical realities. In an effort to accommodate both clinical realities and the demand for clinical accountability, several procedures have been developed to assist therapists in documenting a therapeutic change in patients' status. The purpose of this article is to present some of these techniques in the hope that they will assist clinicians in establishing "accountability." The assumption is made in presenting these procedures that systematic documentation of patient performance is the first step in the process of developing a clinical science. It is important to note, however, that the procedures described are designed primarily to document clinical change and not to demonstrate causality.

**Measures of Clinical Change**

Sommers Graph. Sommers, McGregor, Lesh, and Reed (1980) presented a graphic method for describing the efficacy of early intervention programs for children. The procedure, referred to hereafter as the Sommers Graph, is based on the general assumption that developmental rate follows a sequential, predictable course. The course of development for a particular domain is operationally defined by the instrument used to evaluate the child's developmental progress. In developing the approach, Sommers and colleagues (1980) noted that development may be thought of as the sequential addition of skills and behaviors to an individual's repertoire. The Sommers Graph requires the use of some type of developmental checklist or test designed for children. Sommers and others (1980) noted that a wide variety of developmental assessments are available and that there is good agreement among various scales regarding the sequence of normal development. To develop the Sommers Graph for an instrument, the skills (number of items) on the developmental scale of interest are added and plotted against time. This procedure produces a graph that indicates rate of development based on the particular scale. The graph can be completed for overall development by combining items across all domains (i.e., gross motor, fine motor, etc.), or a graph can be developed for one domain. For instance, Figures 1 and 2 depict the curves computed...
for the language and the motor domains of the Developmental Programming for Infants and Young Children (Schafer & Moersch, 1977). The graphs were developed using the cumulative items across age for each developmental domain. Once the curve is computed for a particular instrument, the performance of a given patient can then be plotted against the curve. For example, Figure 3 presents the curve for the combined domains for the Gessell Developmental Schedule (Knobloch & Pasamanick, 1974) plotted at 2-month intervals. Also plotted in Figure 3 is the performance for a child initially evaluated at 8 months of age. An intervention program was initiated following the 8-month evaluation and the child's subsequent performance at 12, 18, and 24 months of age are plotted on the graph. The graph allows a visual comparison of the child's performance relative to the normal course of development as defined by this particular scale.

Sommers and colleagues (1980) described the use of the procedure with developmental checklists and scales. The technique, however, may be used with other instruments including standardized tests. Figure 4 presents the curve for the Motor-Free Visual Perceptual Test (MVPT) (Colarusso & Hammill, 1972). The raw scores are plotted using normative data from the MVPT manual (Colarusso & Hammill, 1972). A single standard deviation band is included with this curve to provide an additional parameter for assistance in making clinical decisions regarding performance.

The Sommers Graph provides an easily computed and understood method of documenting patient change over time in comparison to a standard or normal rate of performance.

Proportional Change Index. The Proportional Change Index (PCI) was introduced by Wolery (1983) as an alternative to previously proposed efficacy indexes (cf. Simeonsson & Wiegernik, 1975). The PCI is designed to measure developmental improvement in any domain (motor, cognitive, language, etc.). To compute the PCI the following information is required: (a) the child's preintervention chronological age, (b) the time spent in intervention, (c) the child's pretest (preintervention) score, and (d) the child's posttest (postintervention) score. The PCI is computed using the following formula:

\[
PCI = \frac{\text{developmental gain}}{\text{time in intervention}} = \frac{\text{posttest developmental age} - \text{pretest developmental age}}{\text{pretest chronological age}}
\]

Wolery (1983) argued that the PCI compares the child's rate of development at pretesting to his or her...
Finally, children whose rate of development during intervention is less than their (estimated) rate of preintervention development will receive a PCI score of <1.0.

The PCI is not just a gain score. Two children who demonstrate the same number of months of improvement during intervention may have different PCI scores because of different rates of estimated preintervention development. Consider the following two PCI scores:

\[
\text{PCI} = \frac{14 \text{ months}}{10 \text{ months}} \div \frac{12 \text{ months}}{12 \text{ months}} = 1.4 \quad \text{(Child 1)}
\]

\[
\text{PCI} = \frac{14 \text{ months}}{10 \text{ months}} \div \frac{8 \text{ months}}{12 \text{ months}} = 2.1 \quad \text{(Child 2)}
\]

Both children were in treatment for the same period of time (i.e., 10 months), and both made similar developmental gains (i.e., 14 months). The PCI score for Child 2 is larger, however, because her rate of gain during intervention was greater relative to her estimated rate of preintervention development.

The PCI provides an easy-to-compute numerical index of the relationship between the estimated rate of preintervention development and the rate of development during treatment. The primary disadvantages are as follows: (a) it requires that performance be measured or transformed into a developmental age equivalent and (b) it makes the assumption that development is progressing in a linear fashion. This latter assumption may or may not be valid.

**Reliability Change Index.** The two previous methods of documenting clinical change in patient performance are designed to be used with a pediatric population. The Reliability Change Index (RCI) proposed by Jacobson, Follette, and Revenstorf (1984) is a more general purpose measure that is useful with outcomes that are "nondevelopmental" in nature. Jacobson, Follette, and Revenstorf (1984) argued that for a change in performance to be clinically significant it must be "statistically" reliable (i.e., there must be some way to determine that the change is not due to chance variation or measurement error). To make this determination, Jacobson and colleagues proposed the use of the Reliability Change Index (RCI). To compute the RCI, the therapist must have the following information: (a) the patient's preintervention or pretest score, (b) the posttest score following treatment, and (c) the standard error of measurement (\(SE\)) for the test. The \(SE\) represents the spread or the distribution of repeated performances for a given individual. It is influenced by the reliability of the test and may be computed from the following formula, if not provided in the test manual: \(SE = S_1 \sqrt{1 - r}\) where \(S_1\) equals the standard deviation for the test and \(r\) is the reliability coefficient for the test. The RCI is com-
Jacobson and colleagues (1984) argued that an RCI of standardized IQ score is stable and should remain relatively constant for a given individual. Following 9 months of specialized instruction, the child's IQ may have increased to 85. The $S_e$ for the IQ test is 5. Using the formula provided above: $RCI = \frac{(X_2 - X_1)}{S_e}$, Jacobson and colleagues (1984) argued that an RCI of more than ±1.96 would be unlikely to occur ($p < .05$) without actual change.

The RCI is designed to reveal whether the pretest-to-posttest change score exceeds that which would be expected on the basis of measurement error. There is a basic problem with this, however. The $S_e$ in the formula is an index of the dispersion of an obtained score about a "true" score. The $RCI$, however, makes use of two obtained scores. The use of the $S_e$ in the RCI formula assumes that the pretest score is the "true" score and the posttest an obtained score. In reality this is not true. A more appropriate approach would be to substitute the standard error of the difference ($S_{diff}$) for the $S_e$ in the original formula. The revised formula would be: $RCI' = \frac{X_2 - X_1}{S_{diff}}$. The $S_{diff}$ can be computed from the following formula: $S_{diff} = \sqrt{2(S_e)^2}$. The $S_{diff}$, when computed as shown above, represents the amount of difference that would be expected between two scores obtained on the same test by the same individual as a function of measurement error alone. This assumption is more in line with the reality of obtaining pre- and posttest scores for a patient in a clinical setting. In the above example, the revised $RCI'$ using the $S_{diff}$ would be $RCI' = 85 - 75/7.07 = 1.41$.

It can be mathematically demonstrated that the revised $RCI'$ will consistently produce a more conservative or stringent cutoff point than that produced by the original formula (i.e., $RCI' = \sqrt{2(RCI)}$). In other words, the original $RCI$ value will always exceed the $RCI'$ by a factor of 1.41. Therapists using the above method to document the reliability of a clinically significant change in performance would be well advised to use the revised method ($RCI'$) rather than the original formula.

**Case Illustration**

To illustrate the clinical application of the procedures described above, information on a single case will be presented. The subject is a 6-year, 4-month-old boy enrolled in a public school and receiving resource room assistance. The child's IQ as measured by a standardized, individually administered IQ test is 99, and he was identified as learning disabled by a multidisciplinary team.

**Independent variable.** The child's intervention program was designed to enhance integration of tactile, proprioceptive, and vestibular input and was based primarily on the theory and work of Ayres (1973). Treatment was individually administered and included activities that stressed passive and active rotary and linear motion and included rolling games and the use of a scooter board, suspended net hammock, minitrampoline, and swings. The child was generally allowed to choose alternative activities. Attempts were made during the therapy session to structure the activities so that adaptive responses were required by the child. Treatment sessions were conducted two to three times per week for 30 minutes. Each treatment session was supervised by a therapist experienced in sensory integration assessment and treatment. Outcome measures were collected by two therapists experienced in the assessment and treatment of children with learning disorders.

**Dependent variables.** The following outcome measures were collected during the intervention period:

(a) SCPNT. Postrotary nystagmus (PRN) durations were recorded prior to treatment and during the intervention period. PRN scores were collected on a regular basis (at least once per week). The SCPNT was administered according to the standardized in-

![Figure 5: Subjects' Performance on SCPNT and Ocular Fixation Test During Baseline, Treatment, and Follow-Up Periods](image-url)
Figure 6
Developmental Profile for Design Copy Test of the SCIT

Note: Profile includes one standard deviation band and pretest and posttest performance for subjects. *Test was developed by Ayres (1980).

The weekly average PRN duration for the child was computed and is graphically presented (see Figure 5).

(b) The Saccadic Fixation Test, as described in the Pre-Con Optometric Training Manual (Vincentt, 1975) was used as a measure of eye-head coordination. This test presented the student with a sheet of paper on which a block of equally spaced letters was printed, 10 letters wide by 10 letters long. Each letter was approximately .75 cm in height. The student was asked to read aloud the first and last letter in each of the 10 lines. Both the time required to complete the task and the number of errors were recorded. A final score was obtained by multiplying task time in seconds by the ratio of 20 divided by 20 minus the number of errors. With this scoring procedure, the lower score indicates the better performance. The Saccadic Fixation Test was administered repeatedly over the study period. The weekly averages for the subject are plotted in Figure 5 along with the PRN durations.

(c) Additional Measures. To supplement the repeated assessments described above, the following preintervention and postintervention measures also were collected: Design Copy and Motor Accuracy tests of the SCSIT (Ayres, 1980), the Bruininks-Oseretsky Test of Motor Proficiency (Bruininks, 1978), and the Peabody Developmental Motor Scales (Folio & Fewell, 1983). The results of these evaluations are presented in Table 1.

Clinical Change Analysis

An example of the Sommers Graph for the Design Copy and Motor Accuracy (Right Hand) scores appear in Figures 6 and 7. The graphs depict the expected rate of performance based on normative data and also include the pre-post scores for the subject. Similar graphs could be constructed for any of the developmental measures. The Proportional Change Indexes (PCI) for the Bruininks-Oseretsky Test of Motor Proficiency subtests and the Peabody Developmental Motor Scales appear in Table 1.

The Reliability Change Index was computed using the average PRN score for the 3-week baseline period as the pretest value and the average PRN score for the last 3 weeks of treatment as the posttest score. The PCI formula using the $S_{ag}$ as the denominator was used.

$$RCI = \frac{\text{posttest PRN average} - \text{pretest PRN average}}{S_{ag}}$$

$RCI = 19.33 - 10.66/4.24 = 2.05$

The RCI score of 2.05 suggests that the difference in preintervention and postintervention values was not due to measurement error.

Table 1
Pretest, Posttest, and Proportional Change Index for Case Illustration

<table>
<thead>
<tr>
<th>Measure</th>
<th>Pretest</th>
<th>Posttest</th>
<th>PCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>6 years, 4 months</td>
<td>6 years, 9 months</td>
<td>—</td>
</tr>
<tr>
<td>Bruininks-Oseretsky Test of Motor Proficiency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper Limb Coordination</td>
<td>5 years, 8 months</td>
<td>6 years, 11 months</td>
<td>3.35*</td>
</tr>
<tr>
<td>Visual Motor Control</td>
<td>4 years, 8 months</td>
<td>4 years, 11 months</td>
<td>3.1*</td>
</tr>
<tr>
<td>Upper Limb Speed/Dexterity</td>
<td>5 years, 5 months</td>
<td>6 years, 8 months</td>
<td>3.51*</td>
</tr>
<tr>
<td>Peabody Developmental Motor Scales</td>
<td>3 years, 9 months</td>
<td>4 years, 3 months</td>
<td>2.03*</td>
</tr>
<tr>
<td>Gross Motor</td>
<td>5 years, 3 months</td>
<td>5 years, 9 months</td>
<td>1.45*</td>
</tr>
<tr>
<td>Fine Motor</td>
<td>-1.3</td>
<td>-1.6</td>
<td>—</td>
</tr>
<tr>
<td>SCSIT</td>
<td>-0.5</td>
<td>+1.1</td>
<td>—</td>
</tr>
<tr>
<td>Design Copy</td>
<td>-0.8</td>
<td>+0.0</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: *Rate of development during period of intervention was greater than estimated rate of development prior to intervention.
Conclusion

The methods illustrated in the example are intended to be used as adjuncts to documenting clinical change in patient performance. They cannot provide a demonstration of causality between the treatment and outcome measures. It is important for clinical therapists to bear in mind the distinction between documenting clinical change and demonstrating causality when using any of the described methods. To establish a causal link between treatment and outcome, a more “controlled” situation must be created (Keppel, 1973). This controlled situation is typically associated with traditional group comparison research designs (Kirk, 1968) but can also be achieved with other methods (cf. Ottenbacher, 1986).

Not every clinical therapist will have the training, resources, or motivation to produce controlled clinical research designed to demonstrate causality. Every clinical therapist, however, does have the opportunity and the obligation to document the performance of patients in a systematic and objective way. Gillette (1982) accurately noted that, “in the absence of careful and thorough documentation, members of a profession such as occupational therapy will not receive appropriate recognition nor adequate reimbursement for their services, and both the therapist and the profession will be undervalued” (p. 499).

The application of the methods described in this article will help therapists to systematically evaluate and document clinical change and, thereby, provide a mechanism for establishing therapeutic accountability. Systematic documentation and therapeutic accountability represent the initial step in the development of a clinical science of occupational therapy.

References


