The single-subject experimental paradigm, a research paradigm, can also be a powerful decision-making tool in clinical evaluation. To support this thesis, a brief review of the single-subject paradigm is presented, including a discussion of its 1) place in the continuum of methods to develop knowledge, 2) essential characteristics, 3) similarities to the therapeutic process, 4) validity, and 5) generality of findings. Usefulness of designs in this paradigm is illustrated with examples from the physical therapy literature, the ABA and the multiple-baseline designs. Also illustrated is the precision of information gained through individual data analysis.

The single-subject experimental paradigm (intensive study of individuals) has much to offer physical therapy as a clinical research tool. I propose also that it has great potential as a tool for clinicians in evaluating effectiveness of treatments. The intensive study of individuals has a long and honored tradition of contributions to knowledge. It was supplanted in the early years of this century with developments in experimental group designs and statistical analyses as the scientific method. Unfortunately, a perception accompanied these developments that only experimental group designs could produce scientific data. This perception still persists among some persons. Concepts, procedures, and designs in the intensive study of the individual have been evolving over the past several decades. These developments have led to a wider acceptance of this paradigm as a valid research and evaluation tool.

The single-subject paradigm is known by a variety of labels (eg, time-series methods, within-subject comparisons). Each label highlights a particular characteristic of the paradigm; no label encompasses its several unique features. A special feature of this tool is its client-centered focus. This focus is also characteristic of the practice of physical therapy; thus, my preference remains the label emphasizing the paradigm’s single-subject aspect.

So that ideas are not lost in jargon or unclear terms, let me briefly define my use of the labels single-subject paradigm and single-subject design. Single-subject paradigm is the generic label. I will use it interchangeably with single-subject study. Single-subject design refers to design options within the paradigm.

Continuum of Research

Until recently, physical therapists and other professional specialties in the rehabilitation sciences limited themselves primarily to case reports, case studies, and group-comparison studies. I initially attempted to illustrate the relationship of these various methods in a linear sequence. This sequence, from left to right, began with the case report, proceeded to the case study and then to single-subject experimental designs, and ended with group-comparison designs. The ordering principle was the degree of control in these strategies to ensure valid data. I began to question the validity of this ordering principle, as I used it, with my continued study of the group-comparison paradigm and the single-subject paradigm. In my revised diagram, the two paradigms are parallel with each other (Fig. 1). The designs in each of these paradigms will vary in their...
degree of control over internal and external validity. One paradigm is not better than the other, as a linear sequence would imply. Which paradigm and more specifically which design one uses will depend on the question or problem of interest and other aspects. The critical difference between these two paradigms is in the unit of analysis. As their labels indicate, the group is the unit of analysis in one; the individual, in the other.

Note the addition, “single-subject clinical evaluation,” in the revised linear diagram shown in Figure 1. It reflects my belief that single-subject study could be integrated into the daily regimen as an evaluation tool. For example, a common dilemma is the treatment or device of choice given two or more options. The clinical strategy to answer this question would be the alternating-treatment design. With an alternating-treatment strategy, one can proceed with treatment and at the same time evaluate the comparative effectiveness without compromising patient treatment. Once one treatment method is shown to be most effective, which could be determined in as few as four sessions, one would of course proceed with the most effective method. An excellent example of the use of this strategy to answer a different clinical question is the study by Harris and Riffle.

The qualifier, clinical evaluation, is intended to highlight the distinction between the use of this paradigm as a research tool and its use as a clinical decision-making tool. As an evaluation tool, the independent variable or the therapeutic intervention is not manipulated experimentally as it is in research. In such instances, informed consent is necessary only when clinical data will be published or will be used for educational purposes. The feedback loops in the diagram indicate complementary and supplementary relationships.

Until now, most clinicians not involved in research have reported their findings primarily in the prescientific mode, that is, case reports and case studies. This is not to say that these research methods should be eliminated, given the option of single-subject designs. Case reports and case studies can contribute to developing knowledge and to clinical decisions. Each documents and summarizes a patient’s demographic status, diagnosis, clinical problem, treatment, and response to treatment. A case study also includes a review of the literature and brief comparison of the results with the reported literature. I identify case reports and case studies as prescientific because they lack controls to eliminate or minimize other possible sources of explanations for the results observed. From both, however, come the exceptions to the rule, the intuitions, and hypotheses. These methods, of course, are important to pursue in advancing knowledge and its corollary, the care given to patients.

In many instances, the case report and the case study may be easily converted to a single-subject design, strengthening the suggested results and implications.

**Characteristics of Single-Subject Experimental Paradigm**

With the focus of this paradigm on the individual, several requirements have evolved to ensure valid inferences. Modifications of these requirements are sometimes necessary in adapting the paradigm to environmental conditions; examples are given throughout the remainder of this article.

The process of intensive study of individuals consists of a series of repeated observations, the introduction of the experimental or therapeutic variable into this series of observations, and evaluation of the changes during and following intervention. The interval in which measurements are taken before intervention is the baseline, or “A” phase. The A phase is usually a “no treatment” phase. One can, however, extend the baseline phase to include treatment data that have shown little or no effects. The intervention, or treatment, phase is the “B” phase. The conventional method of displaying these phases is shown in Figure 2. One looks for changes between and within phases to assess what is happening. In this paradigm, patients serve as their own controls. In some designs (e.g., multiple baseline), patients also serve as controls for each other.

The single-subject paradigm requires a minimum of two sets of measurements: 1) a series taken before intervention and 2) a series taken during the intervention. If other therapeutic interventions are added, they are also labeled alphabetically and sequentially. For example, one may choose to compare ultrasound with ultrasound and exercise; these variables would be labeled “B” and “BC,” respectively.
The basic building block of the single-subject paradigm is an A phase and a B phase, a simple phase change. In varying combinations of this fundamental unit, one can answer a variety of questions on treatment effects. Even with a simple phase change, one can answer questions such as: Does this treatment work? How well does it work (by assessing, for example, the slope and rate of change)? and Is treatment necessary every day? With a series of phase changes, other assessments can be made, such as the contributions of different components in a treatment package. The idea of a simple phase change should be a familiar one, for that is our current mode of practice. We evaluate (phase A) and then intervene therapeutically (phase B), albeit not with the controls required when one is deliberately applying a single-subject strategy. The AB design is the weakest of all design options. Nevertheless, even this simple phase change allows one to be accountable, if one follows the rules of the paradigm. One has an accurate report of what was done and the progress of the patient. With some caution, one may be able to say that changes are likely because of the intervention. This cautious interpretation is particularly valid if differences between phases A and B are large and immediate. Usually, however, replication of some kind is necessary. Otherwise, all one can say for sure with an AB design is that a change occurred. Other possible explanations are coincidental events, time alone, an assessment effect, or changes in the instrument itself. With each replication, however, one strengthens the likelihood that the intervention was indeed responsible for changes in performance.

The process in applying the single-subject paradigm parallels that of physical therapy in the administration of therapeutic procedures. In both, the target problem is specified, baseline behavior is established, appropriate outcome behaviors (or dependent variables) are identified, a treatment plan is initiated (independent variable), and changes are evaluated. Some dissimilarities, however, exist. Differences are not in the intent or the basic elements that comprise each model. The differences are in the manner these elements are implemented.

The single-subject study model differs from the therapeutic process in three respects: 1) establishing baseline performance with a minimum of three data points, 2) applying a specific strategy or design that will permit cause-effect inferences, and 3) measuring performance repeatedly (Fig. 3). These conditions can be met in the clinical environment. When not feasible (eg, three data points in the baseline phase), other alternatives are available that do not compromise the intent in establishing baseline behavior. An example of available alternatives is the use of archival or normative data to establish, a priori, the stability of the patient's current status. In this instance, one is assuming that the one data point represents the patient's current status. This assumption, of course, needs to be validated.

<table>
<thead>
<tr>
<th>THERAPEUTIC MODEL</th>
<th>SINGLE-SUBJECT EXPERIMENTAL PARADIGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Evaluate patient’s status</td>
<td>Evaluate patient’s status</td>
</tr>
<tr>
<td>One baseline measure (BL)</td>
<td>*Three or more BLs</td>
</tr>
<tr>
<td>Specify problem(s)</td>
<td>Same</td>
</tr>
<tr>
<td>Measure problem(s)</td>
<td>Same</td>
</tr>
<tr>
<td>Specify goal(s)</td>
<td>Same</td>
</tr>
<tr>
<td>Specify therapeutic intervention (TI)</td>
<td>Same</td>
</tr>
<tr>
<td>2. Apply TI</td>
<td>*Specify design to assess efficacy</td>
</tr>
<tr>
<td>Adjust TI to patient’s response</td>
<td>*Apply TI within design strategy</td>
</tr>
<tr>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td>3. Evaluate patient’s response</td>
<td>Evaluate patient’s response</td>
</tr>
<tr>
<td>*Repeated measures</td>
<td>*Indicates absence or difference in approach from the current model.</td>
</tr>
</tbody>
</table>
and can be validated with repetition of the design strategy with similar patients. Other requirements of this paradigm, in addition to repeated measures and establishing baseline behavior, are 1) changing one variable at a time (otherwise, one does not know whether the removal or addition of the variable is responsible for changes observed), 2) demonstrating stability of behavior before one intervenes, 3) arranging data collection in phases because one assesses changes within and between phases, and 4) conducting replications to confirm internal and external validity. The phrase “stability of behavior” commonly used to describe this requirement can be misleading. A better descriptor is “typical behavior.” To illustrate, consider trying to determine whether a modality is effective in reducing spasticity. Baseline measurements would produce a varying response pattern, typical of spasticity, not a stable pattern. With the introduction of the therapeutic intervention, one would look for reduced variability or stability of performance. Graphing of data is not a requirement, but it is characteristic of the single-subject paradigm because visual analysis is often the preferred mode for interpreting the data between and within phases.

**Data Analysis**

The single-subject paradigm evolved out of the need of psychologists to understand the change behavior in the treatment environment. The traditional method of analyzing data has been visual analysis. Clinicians have argued that for changes in performance to be clinically important, they must be visible. Significance or success, in this context, is a therapeutic criterion; that is, success is defined by the clinical goal or expected outcomes of treatment. Tests of statistical significance were not used initially. Instead, descriptive statistics such as percentage of improvement and increase in the frequency of the targeted behavior have been the common mode for reporting results. More recently, statistical methods for assessing significance have been appearing in the literature as a result of increasing interest in and use of this paradigm. In some instances, the usual tests of statistical significance may apply.

**Issues of Validity and Generality of Results**

Whether one is conducting a formal study or evaluating effects of treatment on a patient, one is looking for causal relationships and patterns of relationships. We must be concerned then with variables other than treatment intervention that could explain the results. We must be concerned also with the effects of other sources of variance on the generality of findings to other patients. The various threats to validity are well described elsewhere. These threats to validity (or other sources of explanations for changes observed), such as history, maturation, testing, and statistical regression, can occur with either paradigm (group comparisons or single subject). Their potential influence is more obvious in the single-subject paradigm. In group designs, one attempts to control threats to validity by such procedures as random selection, random assignment, use of control groups, and averaging techniques. These procedures are not exclusive to group studies; they can also be used in the single-subject studies. In the single-subject paradigm, one relies also on combinations of other conditions to increase the likelihood that the treatment was the causal variable (eg, covariation of changes with the introduction of treatment, stability of performance before initiating treatment, immediacy and magnitude of change following the introduction of treatment). Most important are replications with other patients and the design strategy. Threats to internal validity can readily be ruled out by the single-subject paradigm; how this is done varies with the design used.

External validity is a problem for both paradigms. What makes generalization difficult is the variability within and between patients. Generality of results in the single-subject paradigm is determined through systematic replications with individuals or design strategy. Because the single-subject paradigm focuses on one subject or patient, variability from sources other than treatment is identified at the level of the individual. With subsequent replication of single-subject designs with many clients, one can generate more reliable rules that relate particular client, therapist, and environmental characteristics to outcome. Interestingly, Bernard commented in 1885 on the potential for misleading information in the use of group averages. He stressed the importance of knowing individual variations and the conditions for these variances to arrive at “illuminating and fertile generalizations.” Identifying reasons for variability and controlling for them promote the advancement of science. The single-subject paradigm, by its structure, elicits these individual variations. One obtains these variations patient by patient until one has a sufficient data pool from which to derive the specifics of variability and also of what can be generalized.

**Single-Subject Design Examples**

Some of the possibilities of this paradigm have already appeared in Physical Therapy. I will draw on three of these studies to illustrate the usefulness and validity of single-subject designs. A clinical problem was the focus in each study. Two of the three studies used a single-subject design. All were conducted in either a clinical or home environment. The third illustration is a critique of a single-subject study in which analysis was at the level of the group. As a result, some potentially helpful information was overlooked.

**The ABA Single-Subject Design**

The basic experimental design—the ABA—has a minimum of three phases. The second A phase is one in which treatment is usually withdrawn. Measurements, however, continue through the second A phase. The ABA research design is a stronger design than the simple phase change of an
AB design. Consider how much more substantial the evidence for internal validity is if a distinct change in behavior occurs when treatment is begun, when it is withdrawn, and again, when it is resumed. With each successive change in the pattern of data, coincidental effects are less likely as the reason for the change.

Yes, a withdrawal phase can pose ethical problems, and yes, we look for irreversible effects. Thus, in some instances, the ABA design may not be an option for the clinician. Here again, legitimate ways of incorporating withdrawal periods and other suitable alternatives exist. Among the examples are natural withdrawal periods such as weekends when treatment is not given, vacations, and missed treatments. During this type of withdrawal period, the patient is not available for measurements. A representative measure for that interval is an assessment of the target behavior before treatment is resumed.

An excellent example of an ABA design is the study by Laskas et al.\textsuperscript{10} in which they evaluated a neurodevelopmental procedure in a child with spastic quadriplegia. The authors hypothesized an increase in the frequency of heel contact during movement to a standing position. The authors also hypothesized increases in dorsiflexor muscle activity. Two measures were used: 1) frequency of heel contact during movement to a standing position and 2) changes in electromyographic responses between the initial and the final score of each treatment session. Looking at the results of heel contact, we can clearly see the difference in the phase changes from A (baseline) to B (treatment) and back again to A (Fig. 4). These changes were also supported by the EMG recordings.

For this patient, the procedure worked. Also suggested in the data (Fig. 4) is a carry-over effect. For clinicians, however, other questions follow. For example, can this behavior be strengthened so that it is maintained beyond the treatment phase? Does it work for other patients? Regrettably, the authors missed an opportunity to strengthen the internal validity of their findings in this study. Despite the marked changes that lend credence to a cause-effect inference, some other variable may have come into play at the exact time of the intervention and withdrawal. With each successive pattern change, however, this hypothesis of another source of explanation for the treatment effect becomes less likely. The addition of a second B phase would have strengthened the internal validity of the findings.

\textbf{Multiple-Baseline Design}

The multiple-baseline design is a series of two or more AB units with staggered introduction of the intervention phase (Fig. 5). This design can be used to study different behaviors within a patient, the same behavior in several patients, or the same behavior across different environments. As treatment data are collected for the first behavior, baseline measurement continues for the second behavior, patient, or setting. If a shift in performance measurements occurs with each intervention, one has strong evidence of an intervention effect, confirming internal validity. These shifts over several patients or conditions also support external validity. The multiple-baseline strategy, because of its staggered durations of the baseline phase, can provide a clear demonstration that behavior
The use of this design appeared in our literature in 1976 with a fine study by Fowler et al. These clinicians assessed the effectiveness of a color-matching technique for training six children to alternate their feet in stair climbing. The children, four severely mentally retarded and two learning disabled, varied in mental capacity and age; some also had physical deficits; none had learned to climb a stairway with alternating steps. This inability, despite considerable prior therapy, is shown in the repeated measurements in the baseline phases (Fig. 6).

The first child (Patty) immediately demonstrated the desired behavior and maintained it throughout the phase of removal of cues and at follow-up (Fig. 6). The next three children (Mary, Don, and Eve) also showed a marked change in behavior. Note how the children differed in acquiring the desired behavior. Given such clinically documented evidence, one can proceed with this procedure with other children who have similar characteristics. Two other children (Jane and Tom), however, did not respond similarly (Fig. 7). Differing as they did in response to treatment, they cannot be put into the same sample.

The fifth child (Jane) did not respond to color cuing. With no change in performance, those data were added to the baseline phase. The addition of a new stimulus—physical prompting—eventually helped her to acquire the targeted behavior (Fig. 7). The sixth child (Tom) was responding to the color cue training procedure; the response, however, was quite variable, dropping off to former behavior. For this child, color cuing and physical cuing increased the consistency and frequency of alternate stepping in stair climbing.

This study, besides showing a successful use of the multiple-baseline design, is instructive in other ways.
Data showing no treatment effect or little effect were added to the baseline data. Other information beyond the success of the therapeutic strategy was also generated, that is, the range of treatment sessions to acquire the skill. This characteristic is true of other designs. Note also part of the logic of the single-subject paradigm in this design. There is within-patient control (from phase A to phase B) and between-patients control with the staggered intervention of therapy. One also has the opportunity to change treatment strategies, using earlier data as control data, as one attempts to change failures into successes. Such cumulative data and noting the differences between successes and failures provide information from which one can predict for whom a particular intervention would be effective, or one can predict approximate time or frequency of treatment necessary to achieve the treatment goal. This kind of information is typical also of other designs in the single-subject paradigm.

The multiple-baseline design is stronger than the ABA design in controlling for internal validity. By its structure, it also provides data for assessing the generality of results.

**Information Differences in Group and Individual Analysis**

The last example illustrates the use of an ABA design with five infants studied in each session. The statistical significance of the data was tested with an analysis of variance. Those of you who work with high-risk neonates must have welcomed the interesting study by Leonard et al. The purpose of the study was to determine whether perioral stimulation would increase the rate of nutritive sucking in five high-risk neonates. The morning feeding provided baseline behavior, the noon feeding was the intervention phase, and the evening feeding was the second baseline period.

Often, grouped data conceal individual differences that are important to understanding the phenomena and for implementing appropriate therapy.

**Fig. 8.** Number of sucks per minute in high-risk neonates. Typical response: Baby 3. (A = baseline phase, B = intervention phase.) (Graphed from data in Leonard et al.)

**Fig. 9.** Number of sucks per minute in high-risk neonates. Atypical response: Baby 4. (A = baseline phase, B = intervention phase.) (Graphed from data in Leonard et al.)
Such appears to be the case in this study. To illustrate, graphic profiles were drawn from data reported in the article (Figs. 8-10). Two of these profiles, although supportive of the authors' conclusion that perioral stimulation is beneficial in managing newborns with feeding dysfunction, bring out individual differences. The strong similar response of three babies is illustrated in the graph of baby 3 (Fig. 8). The difference between the baseline phases and the therapeutic phase are visually quite evident. A carry-over effect of the stimulation technique into the second A phase did not occur. The babies' sucking rates were lower in the evening feeding than in the morning feeding. Perhaps the increase at the noon meal was enough for them. If a control baseline period for all three feedings the day before the study could have been added to this design, it would have helped to clarify possible reasons for the lower sucking rate in the evening.

The response patterns for babies 4 and 5 are interestingly different from those of the other infants in the study and from each other (Figs. 9, 10). They support the hypothesis of increased nutritive sucking after perioral stimulation, not as strongly but certainly in the right direction. Nevertheless, these individual differences suggest other questions, not obvious from the group analysis, that may affect treatment.

If one looks not at the aggregate data but at the individual data, then other information of clinical importance is available, allowing clinicians to be even more precise in developing treatment strategies. This kind of analysis illustrates one of the strong features of the single-subject paradigm. The data generated on individual responses and characteristics of these responses are immediately helpful in proceeding with treatment for the individual patient. The patient benefits from this strategy. As such data accumulate, we begin to expand and strengthen the knowledge base that supports treatment practices and development of health-related policies.

This study was well conceived. Five patients were observed in this ABA design, and the study was completed in one day, controlling nicely for maturation as an alternative explanation.

**Conclusion**

The clinician makes multiple decisions daily on choice of treatments, procedural aspects of treatments, and effectiveness. Some of these decisions are based on research findings; other decisions, experientially based, remain to be validated scientifically. The single-subject paradigm, developed originally for clinical research, presents us with an opportunity to become scientific practitioners not only with greater use of this paradigm as a research tool but also as a clinical decision-making tool. Collecting objective clinical data is not enough; neither are reliable and valid data. The additional critical element is the strategy one uses to collect clinical data and thus control for alternative sources of explanations for observed changes. Just as the researcher must ask whether the experimental variable was truly responsible for the changes, so also must the clinician ask the same question of the treatment protocol. The single-subject paradigm offers data-collection strategies to answer such questions. It is a problem-driven, client-centered, flexible process that fits the clinical mode of operation. A treatment protocol may be changed if the patient is not responding without compromising the scientific acceptability of the data. At the same time, the patient's best interests are advanced. With this paradigm, decisions to continue or modify treatment become data driven as do subsequent decisions to select or recommend treatment. The use of this paradigm is limited only by the creativity of the clinician. Despite my enthusiasm for this paradigm, the paradigm itself needs further research to determine its best uses in physical therapy. The many design options that are available and useful to us now were developed by clinicians of other disciplines as they applied the principles of this paradigm within the contraints of their
disciplines. What modifications or even new designs that may be needed for additional applications in physical therapy remain to be determined. Because the paradigm is a problem-driven system, one may well find oneself developing a new design strategy for a clinical problem that produces data as scientifically acceptable as those designs already developed.

We have a rich resource in clinical data. Where this paradigm can be integrated into the clinical routine, data can be produced that equal data collected through research and that are immediately available for the decisions facing clinicians. Accumulation of such data, beyond their immediate usefulness to the therapist and the patient, would contribute substantially to the knowledge base supporting the practice of physical therapy and related health care policies.

**Acknowledgments**

In the development of my ideas on clinical research, I wish to acknowledge the stimulating exchanges with Steven B Hayes, PhD, Janie Wise, MSc, PT, Barbara Jerry, MSc, PT, and my colleagues in physical therapy and rehabilitation medicine at the Karolinska Institutet and the University of Lund, Sweden.

**References**