REVIEW ARTICLE

Single-Subject Research in Rehabilitation: A Review of Studies Using AB, Withdrawal, Multiple Baseline, and Alternating Treatments Designs

Catherine L. Backman, MS, OT(C), Susan R. Harris, PhD, PT, Jo-Anne M. Chisholm, BSR, OT(C), Angela D. Monette, BSc, OT(C)

ABSTRACT. Backman CL, Harris SR, Chisholm JM, Monette AD. Single-subject research in rehabilitation: a review of studies using AB, withdrawal, multiple baseline, and alternating treatments designs. Arch Phys Med Rehabil 1997;78-1145-53.

Objective: To review the "methodologic rules" for using single-subject research designs (SSRDs) and to review the use of SSRDs in rehabilitation research of the past decade.

Data Sources: CINAHL and MEDLINE searches using "single subject" and "single system" as key words for the period 1985-1995 yielded 61 articles related to rehabilitation.

Study Selection: Studies were selected for review if they described one of four commonly used SSRDs, specifically AB, withdrawal, multiple baseline, or alternating treatments.

Data Extraction: Studies cited were identified by consensus and either exemplify adherence to the experimental rules of SSRDs or illustrate errors that result in threats to the validity of stated findings.

Data Synthesis: All four types of SSRDs have been reported in rehabilitation studies, sometimes incorrectly.

Conclusions: SSRDs, with their client-centered focus, are ideally suited for researching human behavior in the rehabilitation practice environment. Although numerous sources clearly identify the methodologic requirements for single-subject experiments, several studies violate the basic rules, threatening the validity of the results of these studies. Other properly applied SSRDs illustrate the strengths of this approach, which can produce empirical support for rehabilitation interventions.

© *1997 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation*

SINGLE-SUBJECT research design (SSRD) is a valuable clinical research tool. Properly used, it can inform and illuminate clinical practice. Single-subject research designs can provide concrete data to validate existing theories in rehabilitation as well as formulate new ones. In the past several years, single-subject research has become increasingly acknowledged in rehabilitation literature. However, some misconceptions regarding both the design and its implementation still exist. The purpose of this article is to review the basic concepts of SSRD, then to review rehabilitation literature that reports the use of SSRD as the experimental method.

From the School of Rehabilitation Sciences, The University of British Columbia, Vancouver, BC, Canada.

Submitted for publication August 5, 1996. Accepted in revised form February 19, 1997.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

Reprint requests to Catherine Backman, MS, OT(C), School of Rehabilitation Sciences, The University of British Columbia, T325-2211 Wesbrook Mall, Vancouver, BC, Canada V6T 2B5.

© 1997 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

0003-9993/97/7810-4142$3.00/0

Single-subject research is not the same as a case study. A case study provides a detailed description of a client and the client's responses to treatment, but there is no attempt to define and manipulate an independent variable in order to examine its effects on a dependent variable, as is required in experimental studies.[1] In contrast, the single-subject experimental paradigm adopts the assumptions of the quantitative research paradigm, except that the unit of study is an individual rather than a group.[2] See table 1 for a comparison of the two approaches.

The process of single-subject research consists of systematic, repeated measurement of a target behavior (dependent variable) through one or more baseline and intervention phases. Data are gathered for a minimum of three sessions to establish a baseline, followed by the introduction of an intervention (independent variable). Continued repeated measures are taken throughout the intervention phase, which permits cause-effect inferences to be made. If a clinically significant change occurs as a result of the intervention, then the study must be replicated across subjects, settings, and practitioners to accumulate evidence to strengthen external validity.[1]

Single-subject research design also requires that only one independent variable be changed at a time and that stability of the response (or target behavior) be achieved before introducing the intervention.[3] In practical terms, stability of response is difficult to judge without at least 5 data points because target behaviors tend to have some day-to-day fluctuations in most clinical situations.

Visual analysis of graphed data is the traditional method used in SSRD.[4] It allows for a dynamic, interactive research process through the ongoing plotting of data. Clinical interpretation of visually displayed data can be inconsistent between raters[5]; however, it is usually the first level of analysis in SSRD and may suggest the need for additional techniques. Statistical analyses, such as the split-middle method (celeration lines) or the two standard deviation band method, can quantify and strengthen visual findings.[6] A minimum of 10 data points is recommended before statistical analysis is performed.

All SSRDs should adhere to the basic design criteria listed in table 2. This article will explicate these criteria by (1) describing four different SSRDs, and (2) reviewing selected rehabilitation studies from the past decade that reportedly used one of the four SSRDs.

CINAHL and MEDLINE searches from 1985 to 1995 using the key words *single subject* and *single system* produced a total of 61 articles pertaining to rehabilitation. Articles describing a case study or a single-subject design other than one of the four types of design under discussion were not reviewed. A total of 40 papers were retrieved and are listed in table 3, along with the topic and stated type of SSRD. Seven studies using the AB, withdrawal, multiple baseline, and alternating treatments single-subject designs were selected for critical appraisal because they illustrate common strengths and weaknesses of the single-subject experimental method. An eighth study from 1996 was subsequently added to maintain the currency of this review and because it met the selection criteria.

Table 1: Characteristics of Case Studies Versus Single-Subject Research Designs

Case Study	Single-Subject Research Design
Subjective description of an individual's behavior	Objective measurement of an individual's behavior
Typically, responses to treatment are described and target behaviors may or may not be specifically defined	Target behaviors are operationally defined
Anecdotal	Precise methods of measurement
Frequently longitudinal	Repeated measures over time
Cannot document relationship between independent and dependent variables	Can document relationships between independent and dependent variables
No manipulation of variables	One intervention strategy manipulated at a time; careful control of extraneous variables
Can be used to generate hypotheses for future research	Can be used both to generate and to test hypotheses

SINGLE-SUBJECT RESEARCH: FOUR BASIC DESIGNS

AB Design

The AB design is the foundation of SSRD. The baseline (A) is established through repeated measures of a client's target behavior (dependent variable). When the baseline is stable, an intervention (or independent variable) is introduced, with the intention of affecting the target behavior (outcome). The target behavior continues to be measured throughout the intervention (B) phase. The data obtained through measurement of client responses across the no-treatment phase (A) and the intervention phase (B) are analyzed for changes. In a well-controlled study, changes in the outcome behavior can be clearly attributed to the introduction of the independent variable. The selected outcome measure should be one that can be consistently recorded by an individual unaware of the purpose of the study. Confidence in the results of an AB design is dependent on baseline stability: a long, unbroken pattern of response before intervention followed by a dramatic change in level, trend, or variability of the plotted data points suggests that the treatment effected a change in the subject.[1]

A change in level is seen when there is "an abrupt change in performance" between two phases, resulting in noticeably different levels of data on the y-axis.[1] Trend changes occur when there are changes in the direction in which the data pattern is moving.[4] Variability is noted when there are wide fluctuations in a series of data points.[1] Examples of changes in level, trend, and variability are illustrated in figure 1.

The AB design is the weakest of the single-subject designs because it is unable to control for some potential sources of internal validity, such as history or maturation.[3] A hypothetical example of an AB design is illustrated in figure 2, where minutes of independent ambulation is the dependent variable, and lower-limb strength training is the independent variable in an individual after amputation and prosthetic fitting. In this example it appears that strength training has had the desired effect on the person's ability to walk for increasing periods of time. This design has not controlled for other contributing factors, however, such as the natural recovery process (ie, perhaps the person's walking time would increase at a similar rate even without a strengthening program).

Withdrawal Design

The withdrawal design (also known as ABA) is an expansion and improvement on the AB design, since it can control for history and maturation. It requires the introduction and removal of the independent variable; thus, changes in the dependent variable can be more confidently related to the presence or absence of the independent variable.[1] According to a study by Wolery and associates,[7] a withdrawal design must contain a minimum of two baseline phases and one treatment phase. There are no limits to the number of treatment and withdrawal phases in a study; for example, ABAB or ABABABA are also examples of the withdrawal design. As well, different treatments may be applied or withdrawn, eg, ABACA, where C denotes a second treatment. Each application and withdrawal of the independent variable strengthens the case for a relationship between the variables, providing there is a corresponding change in performance on the dependent variable.[1] The application and withdrawal, even if done rapidly, and the ABABA notation for this design, should not be confused with the alternating treatments design, described later in this article.

A withdrawal design is characterized by the removal of the independent variable in a second baseline phase after the treatment phase. The dependent variable should return to previous baseline levels during the second baseline phase. If this does not occur, threats to internal validity must be entertained as a possible explanation for changes that occurred during the treatment phase.

Two major caveats must be considered in relation to the use of this design. First, it must be ethical to withdraw the treatment. Second, it is preferable if the dependent variable exhibits the property of reversibility.[7] The withdrawal design is not suitable for evaluating treatments that provide long-lasting changes in behavior,[1] thus making it difficult for use in rehabilitation settings where treatments are often expected to produce irreversible gains in function. For example, a withdrawal design may not be appropriate for improving the hypothetical AB study illustrated in figure 2, since a program of strength training may have a lasting effect. In this case, the dependent variable is not immediately reversible. However, withdrawal designs can work well when using adaptive devices that can be sequentially applied and withdrawn, such as hand splints, tone-reducing casts, or adaptive communication devices.

Figure 3 illustrates a hypothetical ABA design, where the treatment is a trial period with an adapted computer keyboard for an individual with limitations in hand coordination, and the dependent variable is typing speed. The keyboard adaptation is a discrete treatment that can easily be removed. The target

Table 2: Design Criteria for Single-Subject Research Designs

Required Criteria for Single-Subject Research Designs
Frequent and repeated measures of the dependent variable are collected (minimum of three data points per phase).
Manipulation of the treatment variable and analysis of its effects on the outcome variable are carefully controlled.
Target behaviors are clearly specified and operationally defined. Methods of measurement are precisely defined.
Continuous measures are taken throughout each phase of the study.
One intervention strategy (independent variable) is manipulated at a time.
Extraneous variables are carefully controlled.
Methods used in data collection are both replicable and reliable.
Measurement and recording procedures are continued until requirements of the specific design have been satisfied.
Additional Recommendations
Independent ("blinded") assessment of outcome measures.
Random selection of subject(s).
Visual and statistical analyses of data.
Each phase is maintained until data demonstrate stable trend.

Table 3: List of Single-Subject Studies Reviewed (Limited to AB, Withdrawal, Multiple Baseline, and Alternating Treatments Designs)

First Author/Year	Topic	Design as Stated in Manuscript	Comment on Actual Design
Angelo[23]/1992	Comparison of three computer scanning modes in people with cerebral palsy (CP)	Single-subject replicated across 6 subjects	Three treatment phases (no baseline)
Backman[11]/1988	Effect of wrist splints on hand function in women with rheumatoid arthritis	Alternating treatments replicated with 3 subjects	
Bjornson[22]/1992	Body position and oxygen saturation of ventilated preterm infants	Alternating treatments replicated with 4 subjects	
Buning[9]/1993	Adaptive computer use for a person with visual impairment	Multiple baseline across behaviors	
Carey[20]/1993	Tactile and proprioceptive discrimination training in stroke patients	AB replicated with 4 subjects	
Casby[15]/1994	Effect of music on disruptive vocalization in dementia	3 withdrawal designs (ABA, ACA, ABCA)	
Cermak[24]/1991	Effect of lateralized tasks for unilateral neglect after right CVA	Alternating treatments with 5 subjects	
Conroy[25]/1988	Engagement level of hospitalized demented patients	Single-case experimental design on 10 consecutive patients	Withdrawal (ABA) design—data neither graphed nor statistically analyzed
Dave[26]/1992	Linear vestibular stimulation on body rocking behavior in adults with mental retardation	Multiple baseline across 3 subjects	
Diamond[27]/1990	Tone-inhibiting, dynamic ankle-foot orthosis on stride characteristics in adult with hemiparesis	Alternating treatments	
Dunbar[28]/1991	Transition from nonoral to oral feeding in children	AB replicated with 3 subjects	
Einset[29]/1989	Efficacy of the electric feeder in young adults with CP	Withdrawal replicated with 4 subjects	
Embrey[30]/1990	Effect of neurodevelopmental treatment and orthoses on knee flexion during gait	A-B-A-BC-A	Withdrawal, with 2 treatments in the 4th phase
Frank[31]/1991	Walking with assistive device and using a wheelchair in students with myelomeningocele performance at school	Alternating treatments with 3 subjects	Actually a withdrawal design with A – wheelchair, B – walking aid, A – wheelchair
Ganz[32]/1988	Decreasing tongue thrusting and tonic bite reflex with neuromotor and sensory facilitation	ABA	
Gill[33]/1992	Adverse effect of ultrasound in wrist tendonitis	Randomized clinical trial for *n* of 1	Lacks repeated measures baseline
Goodman[12]/1991	Short thumb opponens splint on hand function in CP	AB	
Hall[34]/1991	Multisensory stimulation vs nondirected stimulation in comatose patients	ABAB with 6 subjects	
Harris[21]/1986	Inhibitive ankle-foot orthoses and standing balance	Alternating treatments	
Harwood[35]/1989	Effect of occupational therapy group treatment	ABAB	
Hillis[36]/1989	Efficacy of treatment for aphasic naming errors	Multiple baseline across behaviors in 2 subjects	
Hinderer[37]/1990	Baclofen and placebo treatment of spasticity in spinal cord injury	Multiple baseline across 5 subjects	
Iammetteo[38]/1990	Effect of mouth closure on drooling and speech	ABA with 2 subjects	
Krebs[39]/1991	Exercise and gait effects on in vivo hip contact pressures	Single-subject report	Case report (no baseline)
Lilly[40]/1990	Neurodevelopmental treatment for ADL skills in children with CP	Single-subject design, alternating 2 treatments	No baseline, and does not meet criteria for alternating treatments design
Loughrey[41]/1992	Two teaching techniques on recognition and use of function words in aphasic stroke patients	Alternating treatments	
Neeman[42]/1988	Orthokinetic orthoses to treat Colles fracture hypokinesis	Single-subject time series	Withdrawal design, with baseline, sham treatment and treatment phases
Ott[18]/1991	Mealtime behaviors in institutionalized elderly	Multiple baseline across 4 subjects	
Reisman[43]/1993	Sensory integrative approach to treat self-injurious behavior in profound mental retardation	Single-subject design	AB
Reisman[44]/1992	Psychophysiologic measures of treatment effects in sensory defensiveness	Single-subject design	AB
Ross[45]/1992	Use of computers for visual scanning training	ABA multiple baseline across subjects	Actually a withdrawal design replicated with 3 subjects
Sharpe[46]/1992	Comparative effects of bilateral hand splints and elbow orthosis on stereotypical hand movements	Alternating treatments	
Smelt[47]/1989	Inhibitive weight-bearing mitt on tone reduction and function	ABA	
Smith[48]/1988	Sensory stimulation and ADL performance	AB multiple baseline across behaviors	Actually an alternating treatments (treatment vs no treatment) with 3 outcome measures
Smithers[49]/1991	Facilitation of rolling in a child with cerebral palsy	Withdrawal	
Stewart[50]/1992	Isometric joy stick control	Alternating treatments with 5 subjects	
Tona[51]/1993	Upper extremity inhibitive casting	AB	
Waagfjord[52]/1990	Treadmill training on gait in a hemiparetic patient	Withdrawal ABA	
Warren[53]/1993	Prelinguistic communication skills in young children with developmental delay	Multiple baseline across behaviors + multiple baseline across 4 subjects	
Yuen[13]/1993	Improved productivity with use of templates for a woman with cortical blindness	ABAB	

behavior (typing speed) is not permanently affected by the keyboard adaptation, nor is temporary withdrawal of the adaptation an unethical maneuver, so a withdrawal design is a useful way to measure the treatment effects in this scenario.

Multiple Baseline Design

The multiple baseline design extends the classic AB format by varying the length of the baseline phase across subjects, settings, or outcome behaviors.[1] This design is more powerful because it controls for threats to internal validity that are inherent in the AB design without the need to withdraw treatment. In clinical settings where treatment withdrawal is not possible (eg, for ethical reasons, including lack of consent to interrupt treatment) the multiple baseline design provides a viable alternative to the withdrawal design. A multiple baseline design is also appropriate when withdrawal of the intervention would not result in the outcome behavior returning to baseline levels.[8] It is an error to assume a study is a multiple baseline study just because it is replicated across subjects, behaviors, or settings. A critical feature is the varying length of baseline across the subjects, behaviors, or settings under study, in order to control for extraneous variables.

In the multiple baseline design across subjects, an intervention is applied to the first subject while the other subjects continue in the baseline phase. Each subject, in turn, enters the intervention phase on subsequent dates over time. In this way,

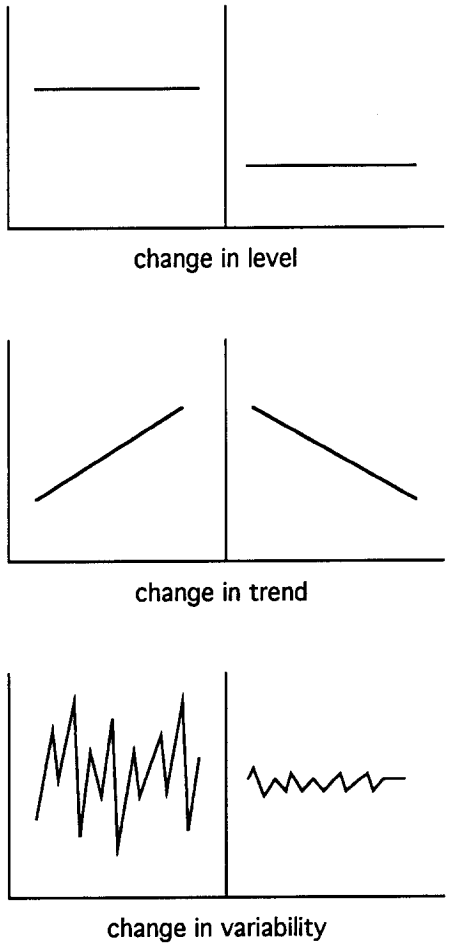

Fig 1. Examples of changes in level, trend, and variability in an AB design. (For a full discussion of possible data patterns in single-subject research, see Wolery and Harris's article.[4])

the data from the subjects remaining in baseline can be compared to the data of the subject receiving treatment, to control for the effects of history or maturation. The outcome behavior should only change in the subject who has received the intervention while the other subjects' responses remain stable. As each subsequent subject receives the intervention, there should be a corresponding change in outcome response. In this way, a stronger cause-effect inference between treatment and target behavior can be made. Wolery's study[7] recommends a return to baseline for one of the subjects if only three subjects participate. A hypothetical example of a multiple baseline design across subjects is depicted in figure 4. In this expansion of the figure 2 example, each of three subjects is sequentially introduced to a lower-limb strengthening program, and its effect on walking time is measured. The longer baseline in subsequent subjects helps to control for maturation, such as the natural recovery in walking time that could occur in this example.

The multiple baseline design can also be applied across settings for the same subject, eg, clinic, home, and workplace. Initially, baseline data are collected in all settings. The intervention is systematically introduced in one setting after another. For example, if the intervention is an adapted utensil, its efficacy might first be tested in an occupational therapy department while the client continues to use his regular utensil in two other environments (home kitchen and restaurant). After a minimum of three trials in the therapy department, the utensil is introduced

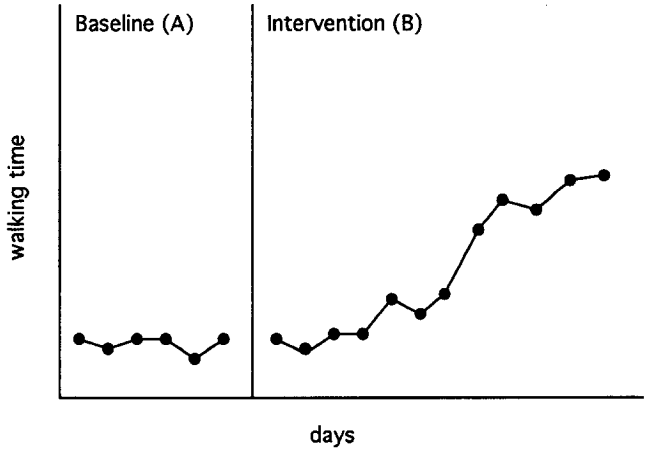

Fig 2. Data from a hypothetical example of an AB design. After a baseline phase (A) the independent variable of strength training is introduced in the intervention phase (B), and the dependent variable (walking time) is measured throughout.

in the home and, finally, the restaurant. The target response is measured for all three settings. Changes in level, trend, or variability in the graphed data are expected only during the intervention phase (after the utensil has been introduced) in each setting.

The third application of multiple baseline, across behaviors, allows the researcher to study the effects of one treatment on several functionally independent target behaviors. For example, in a study by Buning and Redditi-Hanzlik,[9] this design was used to study the effect of an adapted computer with voice synthesizer and text scanner on the reading speed of a graduate student with visual impairment. The computer was sequentially introduced to each of the three target reading behaviors (reading research articles, proofreading word-processed documents, and informational reading, eg, letters and instruction sheets). The subject's research article reading speed increased after the introduction of the computer program, while the speed of unexposed (no computer) reading behaviors of proofreading and informational reading remained stable. In the multiple baseline across behaviors design, it is imperative that the behaviors are not interdependent, and that the independent variable does not have global effects.[8] In the previous example, if the independent

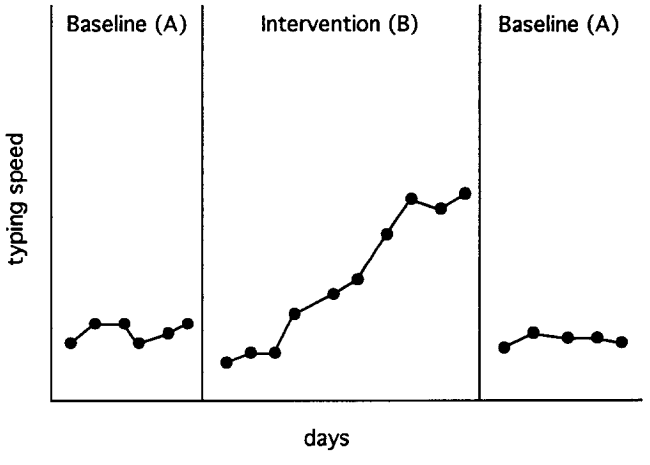

Fig 3. Data from a hypothetical example of a withdrawal design. The independent variable, an adapted computer keyboard, is introduced during the intervention phase (B) and subsequently withdrawn.

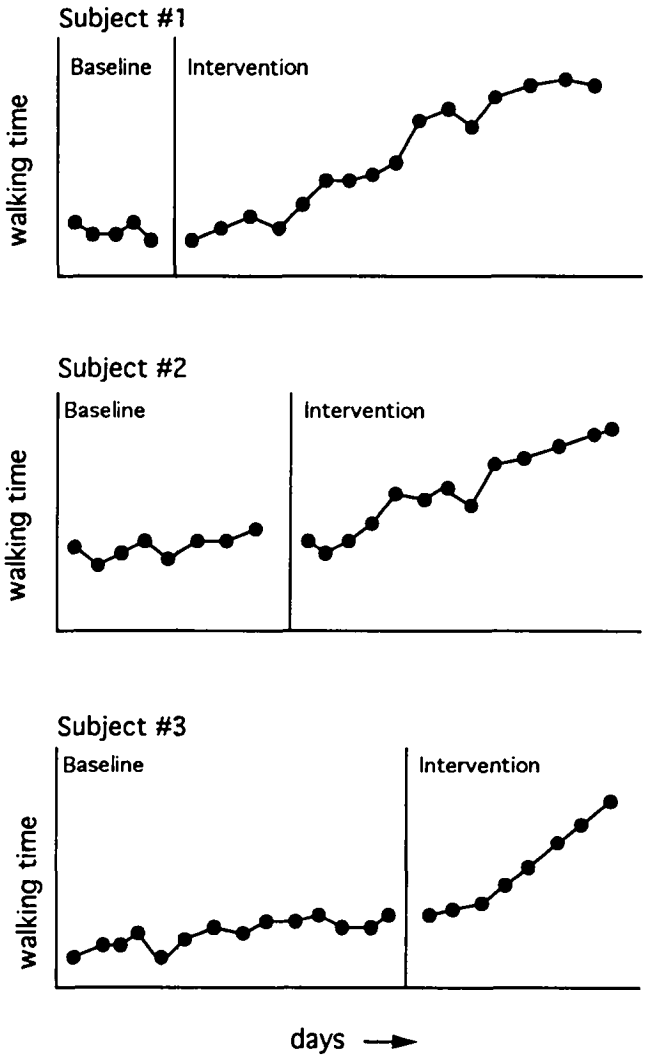

Fig 4. Data from a hypothetical example of a multiple baseline across subjects design. Each of three subjects is subsequently entered into a treatment phase of strength training. The increasing length of the baseline phase for each subject helps to control for threats to validity such as history and maturation.

variable had been a visual training program (a global effect) rather than an adapted computer (discrete effect), all the identified behaviors may have responded to the treatment at approximately the same time. This would negate the controlling effect of the multiple baseline design.

Alternating Treatments Design

The alternating treatments design is well-suited to compare the effects of two or more treatment conditions on a single subject's response.[10] Interventions are always compared with one another and can also be compared to a baseline condition. An implementation schedule is determined, and confounding factors are accounted for, before the start of the study.[7] In its most basic form, the alternating treatments design consists of two interventions presented one after the other in a rapidly alternating fashion without reference to a baseline condition. This provides systematic documentation of client responses during each intervention, but, lacking a baseline, little in the way of cause and effect relationship can be discerned.[1]

A more sophisticated and powerful alternating treatments design involves first establishing a baseline phase. Two or more alternating treatments are then randomly introduced while continuing to evaluate the baseline condition. After the alternating treatment phase, baseline data are collected for several more sessions, potentially producing a very long baseline that extends through all three phases. This stronger design permits empirical comparisons between treatments and between treatment and baseline (no-treatment) conditions.[1] Such a design is illustrated in figure 5, a hypothetical study that examines the effect of a wrist splint on hand function. After a baseline phase, repeated measures of hand function are made under two conditions: with the wrist splint on and with the wrist splint off. An example of the alternating treatments design used to measure the effect of wrist splints on several levels of hand strength and function can be found elsewhere.[11]

In the alternating treatments design, a consistent difference in level, trend, or variability between the conditions demonstrates experimental control[7] and shows the effects of different treatments on the same dependent variable. Multiple treatment interference can be a problem in the alternating treatments design.[10] The random presentation of interventions can minimize this and is an important component of a well-constructed alternating treatments study. A second caveat in using this design is that the treatments offered must produce reasonably rapid and clear behavior change for analysis to be possible.[1] It is, however, well-suited to interventions such as orthotic devices or assistive technology when these interventions are provided to enable task completion.

CRITICAL APPRAISAL OF SINGLE-SUBJECT RESEARCH IN REHABILITATION

A number of researchers have claimed, sometimes accurately and sometimes inaccurately, to have used SSRDs for applied research in rehabilitation. In this section, published studies in rehabilitation will be used to identify some of the strengths and weaknesses of the four SSRDs described previously.

Citations of AB Designs

Goodman and Bazyk's study[12] is a good example of an AB design. They evaluated the impact of an inhibitive thumb splint on hand function in a 4-year-old girl with spastic quadriplegia from cerebral palsy. The independent variable was a short opponens thumb splint and the dependent variable was hand function, operationalized by the measurement of active range of motion, grip and pinch strength, grasp patterns, cube stacking, and a standardized test of manual dexterity. Because there were several discrete categories of the dependent variable, the behaviors

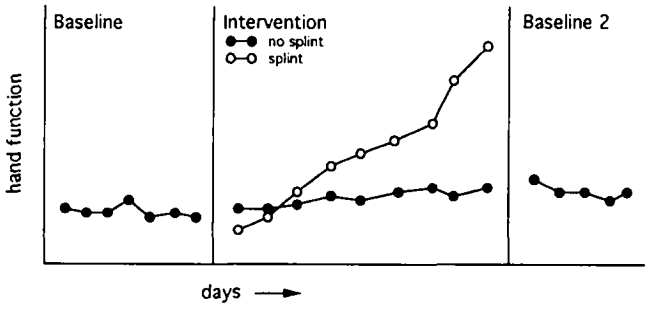

Fig 5. Data from a hypothetical example of an alternating treatments design. After a baseline phase, the treatment, a wrist splint, is introduced, and both the treatment and no-treatment conditions continue to be measured throughout the intervention phase. The dependent variable in this case is a quantitative test of hand function. A return to baseline (withdrawal of the splint) strengthens the design by lending support to a cause-effect relationship between the independent and dependent variables.

were graphed and analyzed independently of one another. There were eight data collection sessions in the baseline phase (A) and eight sessions in the intervention phase (B). This study met the essential criteria for a single-subject AB design.

The number of trials for each phase was established before beginning the study. This was a weakness of the study because data stability was not attained in baseline, before introducing the intervention. All testing, treatment, data collection, and analysis were performed by the researchers. This posed a threat to the internal validity of the study because no interrater reliability checks or blinding procedures were performed in either phase. Neither the subject nor her family was blinded to the study design, which presented an additional threat to internal validity, but one that is relatively common to SSRDs and difficult to avoid.

Despite these limitations, this was a strong AB study. The independent and dependent variables were clearly defined and replicable. Valid measures of hand function were used to operationalize the dependent variable. The setting, treatment, and testing procedures were held constant for the duration of the study. These elements were thoroughly described by the authors so that systematic replication would be possible. Analyses included tests to determine the stability and serial dependency of the data. The data were visually analyzed using celeration lines, and statistical significance was examined. The results indicated that clinically and statistically significant improvement in hand function occurred as a result of the splint; however, the authors did not discuss the functional importance of the treatment for the child.

Citations of Withdrawal Designs

Yuen[13] conducted a rigorous withdrawal design to determine if adapting a work activity would lead to an increase in productivity for a 36-year-old woman who was cortically blind. The design consisted of two baseline and two treatment phases (ABAB). This study met all single-subject and withdrawal design criteria, and all aspects of internal validity were controlled. The dependent variable was defined as the subject's productivity and operationalized as the number of usable bracket outlines traced on pine boards during a 15-minute period (this was a meaningful workshop task for the subject). Measurement of this variable did not require the subject to perform any additional procedures, so there was no threat for a testing effect. The independent variable consisted of the introduction of a physical adaptation, a second template, to the identified task. All procedures pertaining to the implementation of the independent variable and to data collection were meticulously described and, therefore, could be systematically replicated. An independent observer who was unaware of the study purpose counted and recorded the number of usable outlines produced in 19 of the 29 sessions to estimate interrater reliability. Excellent interrater reliability (96.5%) was obtained, exceeding the recommended minimum of 85% to 95% agreement.[14]

Visual analysis, augmented by the use of the C statistic, was used to assess results. A 28% increase in productivity between each baseline and each subsequent intervention phase provided convincing evidence of the treatment effect. This study is a good example of the manner in which an SSRD can be used to evaluate the gains made by a subject during the course of treatment. Although the treatment was specific to the subject studied, the author argued compellingly about the need for continued research in the use of compensatory tools for people with cognitive, perceptual, and sensory impairments, and provided an excellent example of how to evaluate such tools and strategies.

In another example of a withdrawal design, Casby and Holm[15] studied the effect of music on the number of repetitive disruptive vocalizations (RDVs) made by elderly people with dementia. The authors implemented three withdrawal designs, each with a baseline phase (A) and a treatment phase (B, classical music, or C, favorite music). Three subjects were randomly assigned to one of the following withdrawal designs: ABA, ACA, or ABCA.

The dependent variable was clearly defined as the total number of RDVs for each 10-minute interval under observation. The independent variable consisted of listening to music, but was not as well operationalized. The authors provided the titles of the music used in the study but did not specify how long the music lasted or how often it was presented. The investigators precisely described the modified event recording (time-sampling) method they used to collect the data. Although the dependent variable and the time-sampling method were well described, there was a lack of procedural information describing the intervention.

The authors did not state who collected the data, nor was there discussion indicating whether the data collectors were blind to the purpose of the study. No interrater reliability checks were performed on the outcome measures. The baseline data were highly variable and baseline stability was not achieved before introducing the interventions. This lack of experimental control reduces the confidence with which cause-effect conclusions can be made.

Visual analysis, celeration lines, and statistical analysis were performed. The results were equivocal in that a distinct change occurred during the treatment phase but the removal of the intervention did not result in a return of the data to baseline levels. This suggests that a confounding variable may have threatened the study's internal validity or that the behavior was not reversible, suggesting that a multiple baseline design might have been preferable. Systematic replication of the study by others would be difficult. The authors' goal, to identify a nonrestrictive method to decrease RDVs in persons with dementia, has important social implications. This study was a laudable attempt toward attaining that goal.

Citations of Multiple Baseline Designs

Authors of a recently published paper examining the effects of methylphenidate on attention deficits in adults with acute brain injury reported using a "prospective multiple baseline design (A-A-B-A)."[16] Subjects for the study were ten consecutive patients admitted to a brain injury rehabilitation unit. They were assessed using a battery of five standardized neuropsychologic tests on the 4th day after their admission (baseline 1). This same battery was re-administered 7 days later (baseline 2), after which the subjects received a 10-day regimen of methylphenidate (phase B). Seven days after completion of the drug trial, the subjects were again assessed using the same neuropsychologic battery (baseline 3).

Not only did this study fail to satisfy any of the criteria for a multiple baseline design, it also failed to fulfill the basic criteria required for *any* type of SSRD, ie, frequent and repeated measures of a target behavior throughout each phase of the study, visual and/or statistical analysis of graphed individual data, and the use of reliable outcome measures. Only one (group) data point was available for each of the four study "phases" and there was no indication that the outcome measure, (the battery of neuropsychologic tests), was reliably administered and scored. Rather than employing visual or statistical analysis of outcome data for each individual subject, which is one reason why SSRD is so clinically useful and meaningful, the authors grouped the data for all ten subjects and plotted the grouped means as four separate histograms.

What the authors actually conducted was a one-group pretest/

posttest design, which is a quasi-experimental design commonly used in clinical research.[17] Although they administered two "pretests" at different points in time, there is still very little control for threats to internal validity, such as history and maturation. A true multiple baseline design across subjects would have used frequent and repeated measures throughout the baseline phases (staggered across the various subjects) as well as repeated measures during the intervention phase of methylphenidate administration (see SSRD criteria in table 2). Such a design would have allowed for better control for threats to internal validity than was possible in the one-group pretest/posttest design that was used and would also have enabled the researchers to examine within-subject responses to the drug treatment.

Fortunately, more accurate representations of the multiple baseline design across subjects have also been published in the recent rehabilitation literature, and two of these will be described to illustrate how this design can be effectively employed in examining treatment effectiveness. To determine the effects of occupational therapy intervention on mealtime independence and behaviors in elderly residents of an extended care facility, Ott and colleagues[18] used a multiple baseline design across four subjects. In an effort to control for major threats to internal validity and to enhance the generalizability of the results, the four subjects were randomly selected from among all eligible subjects and randomly assigned to baselines of varying length. Data collection sessions were videotaped for future rating and reliability checks. Two independent observers viewed the videotapes and were trained to count and record mealtime behaviors. Interrater reliability ranged from 77% to 99%, depending on the specific behavior rated.

The independent variable was defined as an individualized treatment plan developed to promote independence at mealtimes. To reduce bias during treatment, a standardized sequence of prompting (less to more) was used by the treating therapist. The dependent variable was defined as mealtime behavior and was operationalized as verbal interaction (intelligible vs unintelligible), total number of self-feeding responses (correct vs incorrect), total number of inappropriate mealtime responses, and total number and type of prompts used to facilitate eating. Repeated measures of body weight and caloric intake were also taken. Each successive subject remained in the baseline phase for at least a week longer than the preceding subject, adhering to the basic feature of multiple baselines across subjects.

There were aspects of the study that could have been stronger. Although examples of each of the dependent variables were provided, the difference between incorrect self-feeding responses and inappropriate mealtime responses was not clear, thus limiting replication of the study by others. Inclusion of the score sheet used in the study might have clarified the difference between these two variables. The behaviors that produced the low (77%) interrater agreement should have been specified and an attempt should have been made to improve this level of agreement before proceeding with the study. And finally, the treatment protocol was not clear enough to permit accurate replication of the intervention.

Visual analysis revealed clinically significant improvements in all subjects. Unfortunately, no statistical analyses were performed. The Rn statistic can be used to quantify changes across four or more subjects in a multiple baseline design,[19] and would have increased the power of this study. The notable strengths in this study were the random selection of subjects (enhancing its generalizability), the use of independent raters for data collection and recording, and clearly specified phases for the study. The social importance of improved mealtimes for the institutionalized elderly justifies further investigation in this area.

Another example of a well-conceived multiple baseline design across subjects was a study that examined the effects of tactile and proprioceptive discrimination training on somatosensory loss in four subjects who had experienced strokes.[20] During phase 1 (baseline), no training occurred for either discrimination task over ten assessment sessions. In phase 2, ten treatment sessions were provided for the first discrimination task while baseline checks continued on the second task. During phase 3, ten treatment sessions were provided for the second discrimination task and subjects also received an "abbreviated version" of the training for the first discrimination task. Sequencing of tactile and proprioceptive training was alternated across the four subjects.

In addition, this paper described a study involving a series of four AB designs examining the effects of tactile discrimination training for four other subjects who were also survivors of stroke. Strengths of both the multiple baseline and AB studies included their use of a very clearly defined and replicable treatment program, involving "specific, graded discrimination tasks, attentive exploration of stimuli with vision occluded, deliberate anticipation, and quantitative feedback"[20]; well-defined descriptions of testing apparatus and procedures; both visual and statistical time series analysis of individual data, as well as meta-analyses; and 3-month and 5-month follow-up assessments to evaluate maintenance of the therapeutic effects of the intervention. Both clinically significant and statistically significant improvements were noted with maintenance of these effects apparent at each of the follow-up assessment points.

Citations of Alternating Treatments Designs

In a very simple and straightforward alternating treatments design, Harris and Riffle[21] compared the relative effects of tone-reducing ankle-foot orthoses (AFOs) to a without-orthoses condition on the duration of standing balance in a $4\frac{1}{2}$-year-old boy with cerebral palsy (moderate spastic quadriplegia). During phase 1 of the study, baseline data (without orthoses) were collected on independent standing balance during five sessions spanning a 2-week period while the child was waiting to receive his newly fabricated AFOs. After receipt of the new AFOs, standing balance data were collected during two different treatment conditions, with and without orthoses, for five measurement sessions over a second 2-week period (phase 2). The order of the two treatment conditions was randomized at each of the five sessions.

Visual analysis of the graphed data revealed a dramatic increase in duration of standing balance during the "with-orthoses" condition, both in comparison to the baseline phase (phase 1) and to the "without-orthoses" condition during phase 2. Limitations of this study included the omission of an additional baseline phase after the alternating treatments phase, and the failure to use statistical as well as visual analysis of the data.

Bjornson and associates[22] compared the effects of supine, prone, and side-lying positioning on the arterial oxygen saturation of four ventilated preterm infants. This was an excellent example of researchers selecting an SSRD to fit into a highly specialized clinical setting. The three independent variables (supine, prone, and side-lying positions) were randomly sequenced in the nine data collection sessions for each of four infants participating in the study. The dependent variable, arterial oxygen saturation, was measured by pulse oximeter at 1-minute intervals for the 20-minute period in which the position was maintained. The median score for each condition was plotted for later analysis. This prevented skewing of the data by an isolated event, such as a cough.

The researchers made rigorous attempts to maximize experimental control. The setting, time of day, and methodology were

clearly stipulated and held constant throughout the study. For example, positions were documented by photographs to ensure reproducibility. Intervals between data collection were deliberately set to minimize effects of testing and maturation. Interobserver agreement for the recording procedure was established in a pilot study and corroborated during the actual study by comparing six randomly selected sessions for an average agreement of 99%.

Results were graphed for visual analysis of oxygen saturation. Differences in level of the plotted data among the three positions indicated that the prone position consistently produced the highest oxygen saturation in all four subjects, although there was some daily variability. This was a clinically significant result. Statistical analysis was not undertaken, and the baseline phase was omitted, but these were the only shortcomings in an otherwise classic alternating treatments design.

Limitations of Single-Subject Research Design

Although SSRDs have great intuitive and practical appeal for use by rehabilitation clinicians, there are a number of limitations that must also be addressed. One of the foremost is the limited generalizability or external validity of the study results, particularly when the study is restricted to $n = 1$ subject. To counter this limitation, SSRDs should be replicated across multiple subjects in different clinical settings. Another way to enhance the generalizability of the study results is to randomly select the subject or subjects from among a pool of suitable clients. Several of the studies reviewed in table 3 included replication with more than one subject as part of the experimental design, which enhances the generalizability of findings.

Another potential limitation in the evaluation of SSRD results is the failure to employ both visual and statistical methods of data analysis. Although some results are so obvious visually that they do not necessitate additional statistical analyses, such as the marked change in trend in the study of inhibitive AFOs by Harris and Riffle,[21] many studies require that the graphed data be analyzed both visually and through a statistical or semistatistical technique, such as the split-middle procedure, the Rn statistic, or the two standard deviation band method.[1,4]

Even when employing statistical techniques in addition to visual analysis, it is important to assess the clinical significance or the social validity of the study results, ie, were the changes important and useful to the client who was involved in the study?

CONCLUSION

Single-subject research encompasses more than one design approach to measuring behavior change in a subject. Four different SSRDs were described in this paper, and examples from the rehabilitation literature were used to illustrate the strengths and weaknesses inherent both in the designs and the users' implementation of these designs. The essential criterion of all single-subject designs is the repeated measurement of a discrete, well-defined behavior that has relevance to the individual.

With the increasing emphasis on outcome-oriented, functionally relevant rehabilitation programs, SSRD is one method that permits the systematic measurement of the effects of intervention in support of this practice. Most rehabilitation settings have too few clients with similar characteristics to enable evaluation of treatment outcomes using randomized clinical trials. Single-subject research provides an alternative for the collection of empirical data to support individualized and standardized treatment protocols. Its strength lies in the use of repeated measures and in replication with additional subjects; its weaknesses in rehabilitation research tend to be due to practical limitations (such as reluctance to withhold treatment during a baseline phase), its limited generalizability if only one subject is used, or the lack of rigor imposed by some of the researchers who have attempted to use this method.

One could assume that the increasing pressures imposed by managed care might make it difficult to dictate a no-treatment baseline period before providing a specific rehabilitation intervention. However, a baseline period in SSRD simply means the absence of the particular intervention that is being studied—not an absence of all other types of rehabilitation. Furthermore, this type of "control" may be more reasonable to employ in a managed care environment than the requirement to use an entire no-treatment control group, as is necessary in a randomized controlled clinical trial.

Fortunately, the recent rehabilitation literature contains several examples of well-designed single-subject studies that can guide future practice for rehabilitation professionals.

References

1. Ottenbacher KJ. Evaluating clinical change: strategies for occupational and physical therapists. Baltimore: Williams & Wilkins; 1986.
2. Domholdt E. Physical therapy research: principles and applications. Philadelphia: WB Saunders; 1993.
3. Gonnella C. Single subject experimental paradigm as a clinical decision tool. Phys Ther 1989;69:601-9.
4. Wolery M, Harris SR. Interpreting results of single-subject research designs. Phys Ther 1982;4:445-52.
5. Harbst KB, Ottenbacher KJ, Harris SR. Interrater reliability of therapists' judgments of graphed data. Phys Ther 1991;71:107-15.
6. Ottenbacher KJ. Reliability and accuracy of visually analyzed graphed data from single subject designs. Am J Occup Ther 1986; 40:464-9.
7. Wolery M, Bailey D, Sugai G. Effective teaching: principles and procedures of applied behavior analysis with exceptional students. Needham (MA): Allyn & Bacon; 1988.
8. Hersen M, Barlow D. Single case experimental designs: strategies for studying behavior change. New York: Pergamon Press; 1976.
9. Buning ME, Redditi-Hanzlik J. Adaptive computer use for a person with visual impairment. Am J Occup Ther 1993;47:998-1008.
10. Barlow DH, Hayes SC. Alternating treatments design: one strategy for comparing the effects of two treatments in a single subject. J Appl Behav Anal 1979;12:199-210.
11. Backman CL, Deitz JC. The effect of a wrist splint on hand function in three women with rheumatoid arthritis. Arthritis Care Res 1988; 1:151-60.
12. Goodman G, Bazyk S. The effects of a short thumb opponens splint on hand function in cerebral palsy: a single subject study. Am J Occup Ther 1991;45:726-31.
13. Yuen HK. Improved productivity through purposeful use of additional template for a woman with cortical blindness. Am J Occup Ther 1993;47:105-10.
14. Campbell PH. Using a single-subject research design to evaluate the effectiveness of treatment. Am J Occup Ther 1988;42:732-8.
15. Casby JA, Holm MB. The effect of music on repetitive disruptive vocalizations of persons with dementia. Am J Occup Ther 1994; 48:883-9.
16. Kaelin DL, Cifu DX, Matthies B. Methylphenidate effect on attention deficit in the acutely brain-injured adult. Arch Phys Med Rehabil 1996;77:6-9.
17. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Norwalk (CT): Appleton & Lange; 1993.
18. Ott F, Readman T, Backman C. Mealtimes of the institutionalized elderly: a quality of life issue. Can J Occup Ther 1991;58:7-16.
19. Ottenbacher KJ. Analysis of data in idiographic research: Issues and methods. Am J Phys Med Rehabil 1992;71:202-8.
20. Carey LM, Matyas TA, Oke LE. Sensory loss in stroke patients: effective training of tactile and proprioceptive discrimination. Arch Phys Med Rehabil 1993;74:602-11.
21. Harris SR, Riffle K. Effects of inhibitive ankle-foot orthoses on standing balance in a child with cerebral palsy: a single-subject design. Phys Ther 1986;66:663-7.

22. Bjornson KF, Deitz JC, Blackburn S, Billingsley F, Garcia J, Hays R. The effect of body position on the oxygen saturation of ventilated preterm infants. Pediatr Phys Ther 1992;4:109-15.
23. Angelo J. Comparison of three computer scanning modes as an interface method for persons with cerebral palsy. Am J Occup Ther 1992;46:217-22.
24. Cermak S, Trombly C, Hausser J, Tiernan A. Effects of lateralized tasks on unilateral neglect after right cerebral vascular accident. Occup Ther J Res 1991;11:271-91.
25. Conroy M, Fincham F, Agard-Evans C. Can they do anything? Ten single subject studies of the engagement level of hospitalized demented patients. Br J Occup Ther 1988;51:129-32.
26. Dave C. Effects of linear vestibular stimulation on body rocking behavior in adults with profound mental retardation. Am J Occup Ther 1992;46:910-5.
27. Diamond M, Ottenbacher K. Effect of a tone-inhibiting dynamic ankle-foot orthosis on stride characteristics of an adult with hemiparesis. Phys Ther 1990;70:423-30.
28. Dunbar SB, Jarvis A, Breyer M. The transition from nonoral to oral feeding in children. Am J Occup Ther 1991;45:402-8.
29. Einset K, Deitz J, Billingsley F, Harris SR. The electric feeder: an efficacy study. Occup Ther J Res 1989;9:38-52.
30. Embrey D, Yates L, Mott D. Effects of neuro-developmental treatment and orthoses on knee flexion during gait: a single-subject design. Phys Ther 1990;70:626-37.
31. Frank C, Palisano R, Darbee J. The effect of walking with an assistive device and using a wheelchair on school performance in students with myelomeningocele. Phys Ther 1991;71:570-7.
32. Ganz S. Decreasing tongue thrusting and tonic bite reflex through neuromotor and sensory facilitation techniques. Phys Occup Ther Pediatr 1988;7:57-75.
33. Gill C, Stratford P, Sanford J. The use of a single subject design to evaluate a potential adverse effect. Physiol Can 1992;44(3):25-9.
34. Hall M, MacDonald S, Young G. The effectiveness of directed multisensory stimulation versus non-directed stimulation in comatose CHI patients: pilot study of a single subject design. Brain Inj 1991;6:435-45.
35. Harwood K. The effects of an occupational therapy group: a single case study. Phys Occup Ther Geriatr 1989;7:43-56.
36. Hillis A. Efficacy and generalization of treatment for aphasic naming errors. Arch Phys Med Rehabil 1989;70:632-6.
37. Hinderer S, Lehmann J, Price R, White O, DeLateur B, Deitz J. Spasticity in spinal-cord injured persons: quantitative effects of baclofen and placebo treatments. Am J Phys Med Rehabil 1990;69:311-7.
38. Iammetteo P, Trombly C, Luecke L. The effect of mouth closure on drooling and speech. Am J Occup Ther 1990;44:686-91.
39. Krebs D, Elbaum L, Riley PO, Hodge WA, Mann RW. Exercise and gait effects on in vivo hip contact pressures. Phys Ther 1991;71:301-9.
40. Lilly L, Powell N. Measuring the effects of neurodevelopmental treatment on the daily living skills of 2 children with cerebral palsy. Am J Occup Ther 1990;44:139-45.
41. Loughrey L. The effects of two teaching techniques on recognition and use of function words by aphasic stroke patients. J Rehabil Nurs 1992;17:134-7.
42. Neeman R. Orthokinetic orthoses application in treatment of patients with Colles fracture post-immobilization hypokinesia: a single-subject design efficacy study. Physio Can 1988;40:286-95.
43. Reisman J. Using a sensory integrative approach to treat self-injurious behavior in an adult with profound mental retardation. Am J Occup Ther 1993;47:403-11.
44. Reisman J, Gross AY. Psychophysiological measurement of treatment effects in an adult with sensory defensiveness. Can J Occup Ther 1992;59:248-57.
45. Ross F. The use of computers in occupational therapy for visual-scanning training. Am J Occup Ther 1992;46:314-22.
46. Sharpe P. Comparative effects of bilateral hand splints and an elbow orthosis on stereotypic hand movements and toy play in two children with Rett Syndrome. Am J Occup Ther 1992;46:134-40.
47. Smelt H. Effect of inhibitive weight-bearing mitt on tone reduction and functional performance in a child with cerebral palsy. Phys Occup Pediatr 1989;9:53-80.
48. Smith C. Sensory stimulation and ADL performance: a single case study approach. Phys Occup Ther Geriatr 1988;6:43-57.
49. Smithers J. Facilitation of rolling in a child with athetoid cerebral palsy: a single subject design. Physiother 1991;77:243-48.
50. Stewart H, Noble G, Seeger B. Isometric joystick: a study of control by adolescents and young adults with cerebral palsy. Aust Occup Ther J 1992;19:33-9.
51. Tona J, Schneck C. The efficacy of upper extremity inhibitive casting: a single-subject pilot study. Am J Occup Ther 1993;47:901-9.
52. Waagfjord J, Levanglo P, Certo C. Effects of treadmill training on gait in a hemiparetic patient. Phys Ther 1990;70:549-58.
53. Warren S, Yoder P, Gazdag G, Kim K, Jones H. Facilitating prelinguistic communication skills in young children with developmental delay. J Speech Hear Res 1993;36:83-97.