Available online at

ScienceDirect
www.sciencedirect.com

Elsevier Masson France

EM|consulte
www.em-consulte.com

Physical & Rehabilitation Medicine

ELSEVIER

CrossMark

Update article

Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide

Agata Krasny-Pacini [a,b,c,*], Jonathan Evans [d]

[a] Institut universitaire de réadaptation Clemenceau-Strasbourg, 45, boulevard Clemenceau, 67082 Strasbourg, France
[b] Service de chirurgie orthopédique infantile, hôpital de Hautepierre, CHU de Strasbourg, avenue Molière, 67098 Strasbourg, France
[c] GRC handicap cognitif et réadaptation (HanCRe), hôpitaux universitaires Pitié-Salpêtière, 75013 Paris, France
[d] Institute of Health and Wellbeing, University of Glasgow, The Academic Centre, Gartnavel Royal Hospital, 1055 Great Western Road, Glasgow G12 0XH, United Kingdom

A B S T R A C T

Single-case experimental designs (SCED) are experimental designs aiming at testing the effect of an intervention using a small number of patients (typically one to three), using repeated measurements, sequential (± randomized) introduction of an intervention and method-specific data analysis, including visual analysis and specific statistics. The aim of this paper is to familiarise professionals working in different fields of rehabilitation with SCEDs and provide practical advice on how to design and implement a SCED in clinical rehabilitation practice. Research questions suitable for SCEDs and the different types of SCEDs (e.g., alternating treatment designs, introduction/withdrawal designs and multiple baseline designs) are reviewed. Practical steps in preparing a SCED design are outlined. Examples from different rehabilitation domains are provided throughout the paper. Challenging issues such as the choice of the repeated measure, assessment of generalisation, randomization, procedural fidelity, replication and generalizability of findings are discussed. Simple rules and resources for data analysis are presented. The utility of SCEDs in physical and rehabilitation medicine (PRM) are discussed.

© 2018 Elsevier Masson SAS. All rights reserved.

1. Introduction

1.1. What are SCEDs?

The term single-case experimental designs (SCEDs) refers to a set of experimental methods that can be used to test the efficacy of an intervention using a small number of patients (typically one to three), and involve repeated measurements, sequential (± randomized) introduction of an intervention, specific data analysis and statistics. SCEDs are not case reports but studies carefully designed prior to the start of an intervention and are therefore truly "experimental" designs.

Different names have been given to SCEDs (see column 1 of Table 1), and many different types of SCEDs have been used in the literature (see column 2 of Table 1), which will be described later in this paper. Regardless of the terminology, the design framework is essentially the same:

- studying prospectively and intensively a single person or small group of persons over time;
- measuring repeatedly and frequently the outcome in all phases of the study;
- sequentially applying and/or withdrawing the intervention [1].

What distinguishes SCEDs from group designs is that individual behavior is repeatedly measured both in the absence and presence of a specified intervention. These repeated measures allow patients and participants to serve as their own controls by reflecting each individual's performance at baseline (i.e. before the intervention is introduced), then with intervention. Individuals are studied during multiple discrete phases—at minimum two phases, generally baseline (by convention designated with the letter, A) and treatment or intervention phase (designated with the letter, B) [2].

SCEDs have been used for 50 years, especially in the field of education and psychology. In the medical setting, the term "N-of-1 trial" arose in the mid-1980s in response to limitations that were apparent in applying the findings of randomized controlled trials (RCTs) to the individual patient when making treatment decisions [3]. In psychology, SCEDs have a long history of use in the evaluation of behavior management interventions and in the

* Corresponding author at: Institut universitaire de réadaptation Clemenceau-Strasbourg, 45, boulevard Clemenceau, 67082 Strasbourg, France.
E-mail addresses: Agata.krasny@ugecam-alsace.fr, agatakrasny@yahoo.com (A. Krasny-Pacini).

Table 1
Different names give to SCEDs.

Different names given to SCEDs	Different types of SCEDs
Single-case experimental designs (SCED)	Reversal/withdrawal = ABAB trial
Single-subject experimental designs (SSED)	N-of-1 trial[a]
Single-subject research design (SSRD)	Multiple baseline design: across participants; across settings; across behaviours
N-of-1 trial[a]	
Small N designs	Mixed multiple baseline design
Multiple-case design	Alternating treatment design
Single-case design (SCD)	Changing criterion design
Single-systems designs	Changing intensity design [1]

SCED: single-case experimental designs.

[a] N of 1 trial is the term usually used for SCED in medicine (research on drugs using single cases especially). Although the term "N of 1 trial" is sometimes used for different types of SCEDs, Guyatt et al. [3] have proposed to limit the term N of 1 trial to introduction/withdrawal designs i.e. ABAB designs with multiple cross-overs, blinding of patient and therapist, and randomization.

context of learning disability, whilst in rehabilitation, most SCED papers examine cognitive interventions (especially in aphasiology, neuropsychological rehabilitation and special education) with a number of tutorials and didactic papers presenting SCED use in cognitive rehabilitation and behavioral interventions [4–6]. Introductory papers on SCEDs and reviews have been published in motor areas as well, such as sports [7], adapted physical activity [8] and domains important to rehabilitation such as pain treatments [9]; technology-based health interventions [10]; music therapy [11]. Graham et al. wrote an excellent special communication presenting SCED use across numerous fields of rehabilitation [1].

A recent resurgence of interest in SCEDs has been noted by Smith [12] and by Tate et al. [2], and is reflected in a number of journal special issues on SCEDs, including in rehabilitation journals [*Aphasiology* Volume 29, 2015, Issue 5; *Neuropsychological Rehabilitation* 2014, 42; *Evidence-Based Communication Assessment and Intervention* (Volume 2, Issue 3) in 2008, *Remedial and Special Education* (Volume 34, Issue 1) in 2013]. Evans et al. [13] identified three possible reasons for this recent resurgence:

- the Oxford Centre for Evidence-Based Medicine (http://www.cebm.net/) now rank the randomised N-of-1 trial as Level 1 evidence for treatment decision purposes in individual patients, alongside systematic reviews of RCTs;
- the development of quality assessment tools and reporting guidelines, aimed at improving the methodological quality, and consistency in reporting, of SCEDs;
- the development of methods of analysis suitable for SCED data.

SCEDs enable high-quality research with small numbers of patients, in the clinical setting, in populations that are small, too heterogeneous, or too atypical to constitute a group in RCTs. They allow an intervention to be tailored to the unique needs of a patient, and to assess its effectiveness through a rigorous methodology. Because one to three subjects are sufficient to draw reliable conclusions, SCEDs are less influenced by recruitment problems. They have, therefore, a lower risk of type 2 error, often caused in group studies by insufficient number of included subjects [14] – in SCEDs power comes from the number of repeated measures and not from the number of patients. Studying less subjects but more intensely and comprehensively allows insight into intervention mediating effects and better knowledge of the studied subjects [15]. Furthermore, SCEDs can detect an intervention effect within the (often large) variability of a subject's performance (due to pain, fatigue, etc.). RCTs on the other hand, measure a patient's performance a limited number of times (most often: pre, post and at follow-up) and have a risk of obtaining a score that is not representative of the individual (e.g. if the patient was particularly in pain/tired on the day of the evaluation).

Compiling a list of advantages in using SCEDs is beyond the aim of this paper; readers can refer to excellent papers [1,6,16–18] that

comprehensively outline the numerous positive aspects of SCED methodology.

1.2. Aim of the paper

The aim of this paper is to familiarise professionals working in all fields of rehabilitation with SCED methodology and provide practical advice on how to design and implement a SCED in clinical rehabilitation practice. It does not aim to be an exhaustive tutorial on SCEDs, but rather to be a practical guide for clinicians who are beginners in SCEDs wishing to use this methodology in their daily practice.

2. When to use SCED methodology

SCED methodology aims to test the effectiveness of an intervention or to compare the relative effectiveness of two or more interventions. "In general, small-N designs (i.e. SCEDs) are practical complements to larger N trials. They can be useful in the early developmental phase of research as well as in refining the application of research findings to individual patients." (p. 115) [1]

Situations that particularly lend themselves to SCEDs are:

- evaluating the efficacy of a current intervention for one particular patient in daily clinical practice to provide the best treatment based on evidence rather than clinical impressions;
- conducting research in a clinical rehabilitation setting (outside a research team) with a single or few patients;
- piloting a novel intervention, or application/modification of a known intervention to an atypical case or other condition/type of patients that the intervention was originally designed for;
- investigating which part of an intervention package is effective;
- working with rare conditions or unusual target of intervention, for which there would never be enough patients for a group study;
- impossibility to obtain a homogenous sample of patients for a group study;
- time limitation (e.g. a study needing to be completed within 8 months, e.g. for a master degree research...) or limited funding not allowing recruitment of a group.

Having decided that a SCEDs is, in principle, appropriate and preferable to a group design, the next questions is whether a SCED is feasible? The main reasons that may prevent use of a SCED is the difficulty in choosing a valid and reliable outcome measure that can be measured repeatedly.

3. Repeated outcome measures in SCEDs

One of the most challenging aspect of SCED methodology is finding an adequate outcome measure to assess intervention

Table 2
Differences in outcome measures between group studies and SCEDs.

Traditional group research	SCEDS
Use of validated measurement tools with known psychometrics/clinimetrics, especially inter-rater reliability	Create your own outcome measure [a]
	Corresponding to an objective and measurable behavior
	That can be repeated in time (without major practice effect) [b]
Measurement is repeated not more than two or three times (before, after and at follow-up)	Need to examine and report its reliability in your SCED paper
	Relatively short to administer, as administration will be repetitive
	Sensitive to change/with good responsiveness [c]

SCED: single-case experimental designs.

[a] The repeated measure can, however be a validated test as long as it can be repeated without major retest effect, and is not too long to administer. For example, Gharebaghy et al. [19] used a standardized measure of motor skills, repeated 14 times, to assess the effectiveness of a Cognitive Orientation to Occupational Performance (CO-OP) intervention in children, Nikles et al. [20] used the Conners questionnaires for parents and teachers (assessing attention problems experimented by children at school and at home) to assess the effectiveness of methylphenidate after brain injury.

[b] The repeated measure can also use many parallel forms of a task to avoid the retest effect (see for example McKerracher et al. [21]), although this faces the additional requirement of having parallel forms with the same difficulty level.

[c] The issue of sensitivity to change is important not only for SCEDs but for any study testing an intervention.

effectiveness. Contrary to group trials and clinical practice where norm-references, standardized tests of known clinimetrics are used, SCED methodology usually requires the clinician to create an outcome measure that is relevant to the function being addressed, specifically for the patient and for the intervention being tested that can be assessed reliably multiple times.

Table 2 gives the differences in outcome measures between group studies and SCEDs.

Following behavioral sciences literature, the primary outcome measure in SCED methodology is referred to as the "target behavior" This variable is measured repetitively in baseline(s) and intervention phase(s) and is therefore also simply called the "repeated measure" or the "target variable". The Single-Case Reporting guideline In Behavioural Interventions (SCRIBE) statement [22] provides the following recommendation on the target variable:

- target behaviours should be: "relevant to the behavior in question and that best match the intervention, as well as accurate in their measurement";
- in order to enhance quality of the study and minimize bias: they should be specific, observable and replicable (p. 20) [22];
- "Because the target behaviours are highly specific to the presenting case in SCEDs, formal psychometric evaluation of the measures will generally not have been established. It is therefore recommended practice that evaluation of inter-observer agreement on the target behavior is conducted and reported." [22].

Although it may seem at first extremely challenging to find an outcome measure that fulfils the above requirements, the rehabilitation literature has shown great creativity in the repeated outcome measures used. Examples from different rehabilitation domains are provided in Table 3.

There are three main ways of acquiring the repeated measures:

- independently of intervention sessions (e.g. in an intervention consisting of a weekly one-hour rehabilitation session, the repeated measures consisted of text messages sent by patients three times a week outside the rehabilitation centre [37]);
- during each intervention sessions (e.g.: number of correctly named verbs at the end of each speech therapy session [27]);
- during the intervention session, but on some intervention days only (e.g. once a week during a daily intervention).

This last option is preferable to the second option when:

- the intervention is expected to show slow changes;
- the intervention is intensive (e.g. everyday) and a similar frequency of measures is not desirable;

- the repeated measure is time consuming and administering it at each session would decrease the actual intervention time.

4. Other measures in SCEDs

Other measures used in SCEDs may include control variables, generalization measures, implementation data and standardized tests.

4.1. Control measures

These are measures evaluating untrained behaviors that are not expected to change as a result of the implementation of the target intervention. Their stability after the implementation of the intervention shows that the patient is not making progress on the target behavior due to spontaneous recovery, practice effects, general stimulation, developmental maturation or time spent with therapist but due to the specific effect of the intervention on the target variable. Conversely, if the patient is making a spontaneous recovery or reacting to time spent with therapist, both the target behavior and the control measure are expected to show progress alike. Example of control measures include speech intelligibility in an intervention for dysphagia [26], naming of untrained words in an aphasia training [27], prospective memory tasks not prompted by the assistive technology being tested [31].

4.2. Generalization measures

While it is important that an intervention shows a specific effect, the main quest for all rehabilitation programs is to improve patients beyond trained items, which is often called generalization. Generalization measures are increasingly recognized for their important role in contributing to the external validity of a study testing an intervention. Generalization measures assess the intervention's effect on untrained tasks/items. These measures should be repeated but they do not have to be repeated as often as the target variable. There are four main timing options for assessing generalization (see Table 4). Because the aim of rehabilitation is to improve daily life skills, a special effort should be taken to use outcome measures that are ecologically valid: this is true for the repeated measure but even more for the generalisation measure. It is not an issue specific to SCEDs, but a key issue in all rehabilitation research.

4.3. Implementation data

Implementation data demonstrates that implementing the intervention and progress of the patient go together. For example

Table 3
Examples of repeated measures used in rehabilitation SCEDs.

Example of study	Repeated measure/target behavior
Constraint-induced movement therapy for a baby [23]	Video-taped analysis of the number of times the baby reached for an object, stabilized weight and approached midline with affected limb
Comparing two types of memory book [21]	Number of prospective tasks completed each week. Prospective memory tasks chosen were things that could be verified by a member of staff at the rehabilitation centre (e.g., telephone key worker on Tuesday by 3.30, bring letter in on Wednesday for the occupational therapist). The types of tasks or themes remained the same each week, although the details varied, in order to have eight sets of tasks that were parallel
Strength and proprioceptive training program after ankle sprains [24]	Number of times a balance board made contact with the floor
Web-based navigation system for topographical disorientation following brain injury [25]	Mean navigation errors per route, fifteen route pairs were identified (e.g., OT to PT, ST to OT), any of which, in varying orders, the participant might follow on a particular therapy day
Use of ultrasound feedback in dysphagia rehabilitation [26]	Frequency of observed anterior spillage, of observed penetration-aspiration signs and of swallows self-initiated
Semantic-phonologic treatment for noun and verb retrieval [27]	Number of nouns and verbs correctly named, including trained and untrained noun and verb lists
Evaluation of a thumb opens splint on hand function in cerebral palsy [28]	Use of Goal Attainment Scales (i.e. patient-specific goals rated on a scale between −2 and +2)
Treatment of insomnia associated with traumatic brain injury [29]	Daily diary reports of total wake time
Dressing disability after stroke [30]	6-point rating from the shirt section of the Nottingham Stroke Dressing Assessment (selects correct hole for paretic arm; pulls over paretic elbow; selects correct hole for non-paretic arm; pulls over elbow; pulls overhead; pulls down AND time needed to dress)
Use of Google Calendar for severe memory problems [31]	Number of target actions remembered to perform among patient chosen goals (attending the mosque, attending rehabilitation activities and going to medical appointments)
Impact of typical peer interactions in children with severe profound multi disabilities [32]	Happiness behavior, defined pre-intervention individually for each child, presented on 10-seconds video sequences to an external judge in a random order
Location-based prompting for working activity for individuals with cognitive impairments [33]	Percentage of correct task steps completing snack orders
Non-pharmacological multidisciplinary care programme for persons with generalised osteoarthritis [34]	Diary measures for self-efficacy and VAS for pain
Use of smartphones for severely dysexecuitve patients [35]	Number of programmed phone calls made on time
Relearning of daily living routines (morning self-care and diabetes management) for a woman with impaired memory and initiation following haemorrhagic stroke [36]	Percentage of the steps in each of the two routines completed independently (extensive precise list of the steps of both routines provided in the paper)
Intervention for executive dysfunction and prospective memory in children [37]	"Saint Day task": punctuality score in sending the investigator the saint day name at predetermined time by text message, e mail or phone call
Management by fathers with ABI of oppositional behavior in their children [38]	Number of requests the child responds to chosen from a list of problematic behaviours tested pre-intervention
Interactive web-based cueing, to provide guidance for alarm clock setting [25]	Number of activity steps attempted

SCED: single-case experimental designs; OT: occupational therapist; PT: physiotherapist; ST: speech therapist; ABI: acquired brain injury.

Table 4
Possible timings for assessing generalization.

Timing for assessing generalization	Examples of generalization measures
(1) measured at the same frequency as the main outcome measure	Tiredness using a 5-point modified Borg scale, after a session using a novel dual-input switch [39] (the main repeated measures assessed the patient's speed and accuracy in using the switch but it was expected to decrease patient's tiredness) Percentage of completed work at school, after an intervention for executive dysfunction following brain injury targeting challenging behavior [40]
(2) "probed continuously" (i.e. measured repeatedly but not as frequently as the main outcome measure)	Recommended by SCRIBE [22], but no examples could be found in the field of rehabilitation
(3) measured at the end of the intervention	After an intervention teaching fathers with brain injuries to manage their uninjured children on a series of trained tasks [38], untrained tasks were used at the end of the intervention to evaluate if fathers could apply the strategies with their child in new situations
(4) measured pre/post-intervention	Attainment of personal goals on Goal Attainment Scales, following "Cognitive Orientation to Daily Occupational Performance" (CO-OP); intervention in children (the main repeated measure assessed global motor skill acquisition but this was expected to generalize to personal goals) [19] Multidimensional Fatigue Inventory, Beck Depression Inventory and Beck Anxiety Inventory; following a cognitive-behavioral sleep intervention for insomnia post brain injury [29] (the improvement on the main outcome measure – the total wake time, was expected to reduce daytime tiredness and depression/anxiety symptoms)

in Cosbey's and Muldoon's study [41], of a parent-delivered intervention at home for eating difficulties, the results graph shows that the progress in parents implementing the strategies at home corresponds with increasing effect on the child's eating behavior. Implementation data is part of the broader concept of procedural fidelity, that is not specific to SCEDs but to which SCEDs give a major importance. This will be described later.

4.4. Other measures

Use of repeated measures does not preclude use of other measures that are not frequently repeated. Often, standardized, well recognized measures of known clinimetrics are used before and after the intervention (and if possible at follow-up) to show the

effect of the intervention beyond the target variable, on a more general or well-known assessment tool. For example, in an intervention for sleep disorders after TBI [29], additional measures administered before and after the intervention included the Insomnia Severity index and the Dysfunctional Beliefs and Attitudes about Sleep scale.

5. Different SCED designs

In this section we aim at providing simple guidelines for beginners in SCED, but this is not a rule or recommendation.

5.1. AB designs are not true SCEDs

Before presenting the different types of SCEDs and their possible use, it should be emphasized that 'AB' designs are not SCEDs. Often clinicians and researches may consider that they have used a SCED methodology when they apply an AB design, A corresponding to a series of repeated measure taken in baseline, (without intervention) and B a series of repeated measures in a B phase corresponding to an intervention. Although leading to a stronger evidence of treatment effects than pre/post designs where the patient is tested only once before and once after an intervention, the AB design still does not have sufficient control of biases to be considered a true experimental protocol [1,2]. If a patient makes progress in phase B (as compared to phase A), this progress may be due to the intervention being tested but it may also be due to events occurring concurrently with the intervention, or just the passage of time, and causality cannot be differentiated from coincidence. The AB design is therefore NOT a true SCED [2], as reflected by the other name given to AB designs i.e. "pre-experimental" designs [4]. For this reason, its applicability is not presented here. In order to meet the standards of a SCED, the study must include at least three attempts to demonstrate an intervention effect (e.g. at least three phase change ABAB). ABA studies (two attempts to demonstrate treatment effect), however, are accepted as lower-standards SCEDs [2].

5.2. Designs for interventions with on/off effects

Introduction/withdrawal (ABA/ABAB designs) and alternating treatment designs require the intervention being tested to have immediate effects, short washout and on/off effects. Their use is particularly useful in (but not restricted to) rehabilitation to testing orthotics, prosthetics, use of drugs with on/off effects (e.g.: methylphenidate, intrathecal baclofen, anti-Parkinsonian drugs), adaptive devices and assistive technology (contactors, vocal synthesizers, house automation, smartphones...).

Alternating treatments designs consist of a rapid alternation between different treatments, typically changing intervention after a single administration. Fig. 1 gives an example of easy-to-implement alternating treatment design that allows comparison of the effectiveness of three types of ankle-foot orthosis on walking capacity of a hemiparetic patient [42]. The patient is coming every day to a physiotherapy session and with the three types of ankle-foot orthoses at every session, the order of the conditions being changed (randomized) each day. Results are typically represented by joining the points of each condition together and showing that although the patients' performance may be variable (due to fatigue, pain, motivation...), one condition shows a better effect that the other conditions. This design allows testing of more than two conditions at the same time (see [32,42] for examples of published SCEDs using alternating treatment designs). If only two conditions are tested, the predictable pattern of administration, has been criticized and it has therefore been proposed that same conditions may be administered consecutively up to two times (see Onghena and Edgington [9] p. 58 for a demonstration).

The introduction/withdrawal design, also called N of 1 trial in medicine, on the other hand, assumes that an intervention will be applied over a period of time (or in a number of sessions) and then withdrawn from the patient. For example in the ABAB study of Lui et al. [39] children with multiple disabilities were using a cord vibration switch (the "Hummer") to play a picture matching game. Mean time per item and number of errors per session, were measured repeatedly across an A phase with a single output Hummer, alternating with a B phase using the novel dual-output Hummer.

The simplest form of the introduction/withdrawal is the ABA design but the simplest design that meets the standards of showing at least 3 effects is the ABAB design (a graphical example is provided Fig. 2). However, more complex introduction/withdrawal designs, can be used, for example:

- comparing effects of multiples interventions (e.g. ABACADAEAF design in Tunnard and Wilson's paper [43] testing 5 interventions for neglect);
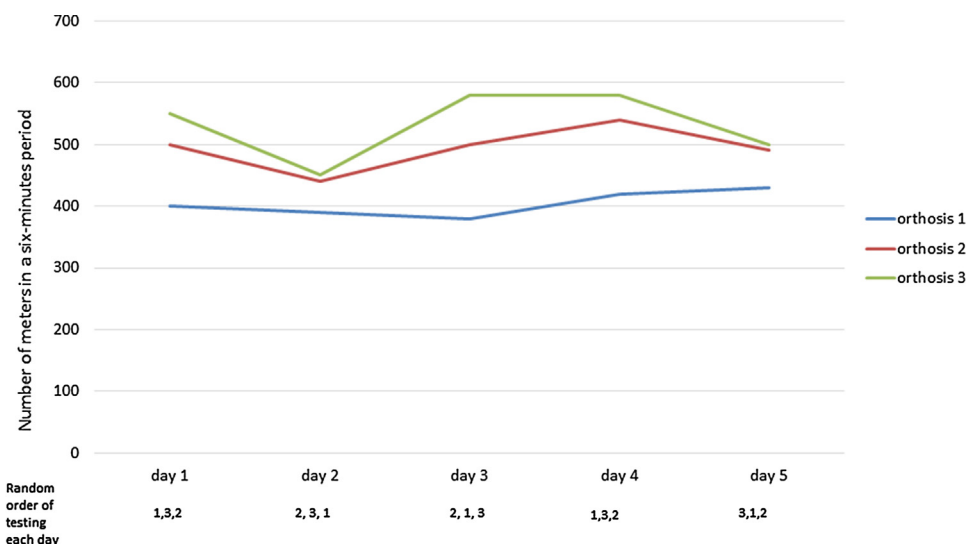

Fig. 1. Alternating treatment design: six minutes walking test using 3 types of ankle-foot orthosis for a patient with hemiparesis. Each day the patient wears all three orthosis but in a different, randomized order.
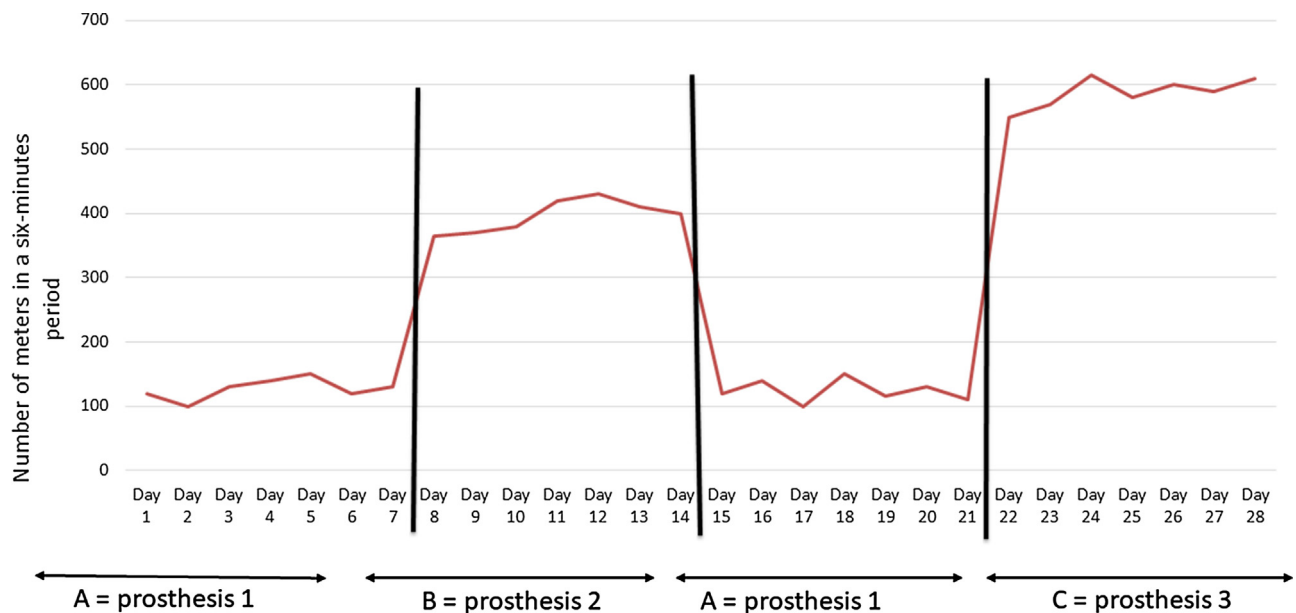
Fig. 2. Introduction/withdrawal, ABAC design: six minutes walking test using 3 types of prosthetic feet for a patient with tibial amputation. The patient wears one type of prosthetic foot for a week (seven days) and changes to another type of prosthetic foot every week. The simplest prosthetic foot 1 is taken as the baseline i.e. A phase, because a baseline condition without any prosthetic foot is not feasible.

- varying the baseline condition, to simulate an intervention and obtain a patient blinding (e.g.: Sumitani et al. [44] who used an ABAA′CBAA″B design for testing prism adaptation for chronic regional complex syndrome; A representing no intervention, A′ a neural prism and A″ a 5° (i.e. insufficient to displace the visual field) prism – all As being variants of the baseline condition);
- testing the combined use of two interventions (e.g.: ABA-CA(B + C)A).

If an introduction/withdrawal design is used on many patients, it is referred to as "Multiple N of 1 trial". This is typically the case when different patients alternate periods with a medicine and periods with a placebo, each period (= phase) typically lasting a few days or weeks. Fig. 3 shows hypothetical data of study testing a sleeping drug in patients with sleeping disorder (but see also the published multiple N of 1 trial on central nervous system stimulant medication to treat acquired attention deficits in children with brain injury [20,45]).

In rehabilitation, outside medications, adaptive devices and assistive technology, changes are usually slow and, they occur with a latency after the introduction of rehabilitation. Above all, a short washout is not desirable as the aim of rehabilitation is to obtain long-term carry-over effects, maintained after the intervention is withdrawn. Therefore, in rehabilitation, another type of SCED is often more useful: the multiple baseline design.

5.3. Designs for interventions with slow/delayed changes and no expected washout

In multiple baseline designs (MBD), the intervention is introduced sequentially to different patients or settings or behaviours. MBD can be viewed as multiple AB designs, with as many AB designs as there are target patients, settings or behaviours. The evidence of such designs comes from demonstrating that change occurs when and only when the intervention is directed at that patient that setting or that behavior. Multiple baseline designs eliminate the need to return to baseline and therefore are particularly suited for evaluation of intervention with long-lasting effects, such as rehabilitation effects.
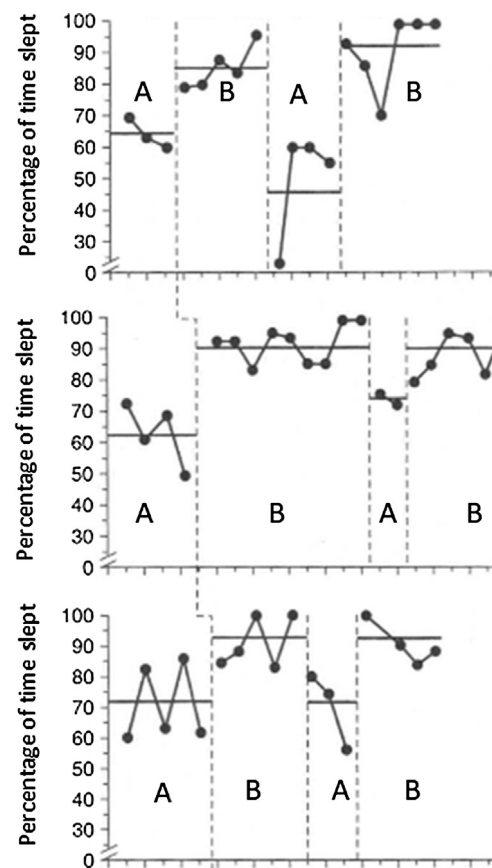

Fig. 3. Multiple N of 1 Trial: percentage of hours slept, out of all the hours spent in bed trying to sleep. A = placebo. B = sleeping drug.

The most used form of MBD is the MBD across patients/subjects: at least 3 subjects are needed. All begin the study with an A phase without any intervention (i.e. baseline), during which the target variable is measured repeatedly for each patient, until one of the patients starts the intervention, while the others continue
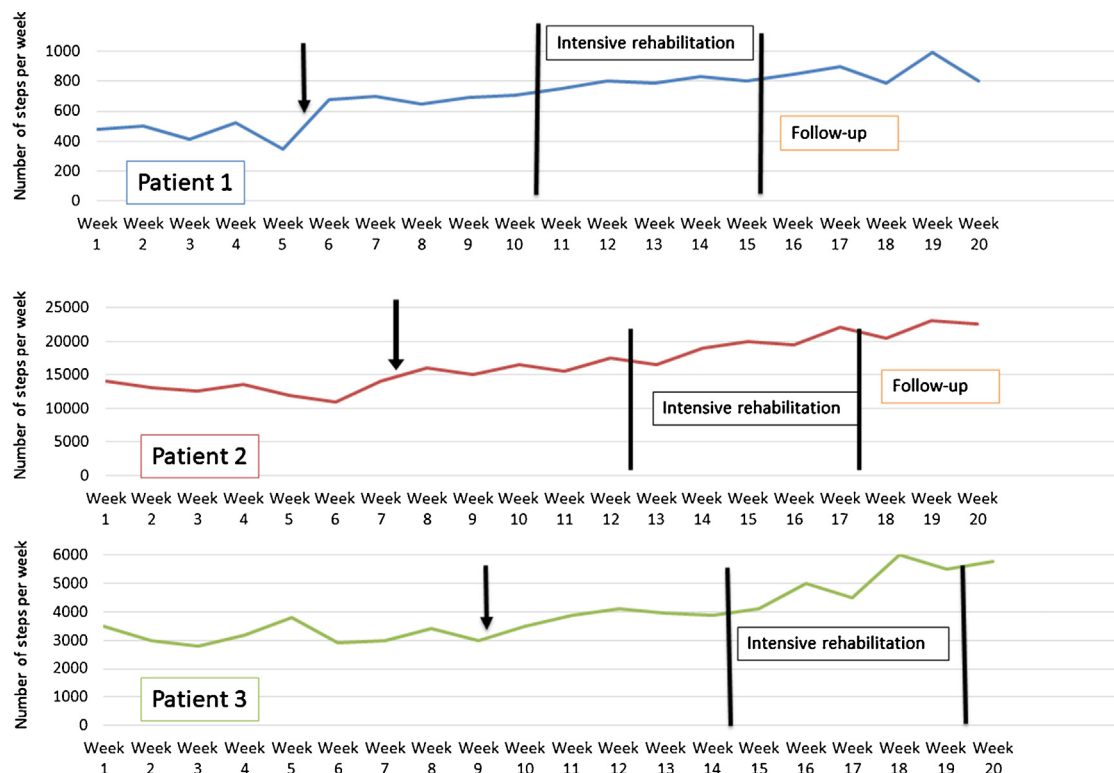
Fig. 4. Multiple baseline design across subjects: number of steps walked spontaneously per week by three hemiparetic patients, measured by podometer, before and after an intervention including botulinum toxin injection, and 5 weeks after the injections, an intensive rehabilitation. Arrows represent the moment botulinum toxin is administered. Vertical lines show the beginning and end of the intensive rehabilitation phase.

without intervention, all patients still being measured repetitively on the target variable. Ideally, choosing which patients will begin the intervention and after how many baseline measures, should be determined via randomization procedure (this will be described later). The intervention is then applied to a second patient after a delay (again if possible determined via randomization) and finally to the third patient. MBD across patients are used in many domains of rehabilitation [19,24,29,34,46]. The advantage of the sequential introduction of the intervention to is to visualize the lack of retest effect and lack of progress unrelated to intervention in the patients not having the intervention yet. This is the main reason that makes MBD across patients a stronger design than AB designs.

Note that in MBD across subjects, each subject may follow a simple AB design but more complex designs [ABC, AB(B + C), ABCD…], in which different components of an intervention are tested separately or introduced sequentially (to explore their cumulative effect) to each patient are also possible (see Fig. 4 for an example).

MBD across settings/contexts consists of implementing an intervention sequentially to different settings a patient operates in. A good example is Feeney's paper [40] on context-sensitive routines at school for children with traumatic brain injury. The intervention consists of sending an outreach team to the school of the patients and coaching teachers to provide adequate, scaffolded support for the student's executive functions. In this paper, the coaching is first introduced for the teacher of English, then after five sessions to the math teacher and finally five sessions later to the science teacher. Feeney shows that the pupil's challenging behavior drops in intensity and in frequency when and only when the teacher has received the coaching and implements it in the classroom (while on the same days, challenging behavior remains high in classrooms where teachers had not received the training).

MBD across behaviours, consist of applying an intervention (or slight variations of an intervention) to different target behaviours or goals or skills. For example:

- introducing a proprioceptive training first to wrist flexion-extension and then to wrist ulnar/radial deviation [47] and measuring repeatedly the error degree in patient estimate of wrist position;
- proving reminders for memory problems for a patient's goal one "remembering to take medication", then to goal two "remembering to have lunch", and finally to goal three "remembering to bring cell phone" [48] and measuring repeatedly the number of actions remembered for each of goal per week;
- introducing an aphasia training first to verbs then to nouns [27] and measuring repeatedly how many nouns and verbs the patient names correctly from a list composed of verbs and nouns.

Note that the outcome measures in MBD across subjects belong to the same category (respectively in the above examples: wrist position estimate, actions remembered, words correctly named), but each behavior is measured separately (respectively position estimate for flexion/extension and for ulnar radial deviation separately, for each memory goal separately, for each category: noun and verb separately).

The different types of MBD can be then combined together, which is called mixed MBD. Boman's study [48] exploring home-based electronic memory aid for persons with memory impairments following brain injury, the design was a MBD across behaviours, (each behavior being one goal to remember every day, leading to a score of actions remembered out of seven each week), but it was applied to five different patients, thus combing it with a MBD across subjects. Similarly, Raymer's study [27] of aphasia training across behaviours (nouns and verbs) was applied to eight participants.
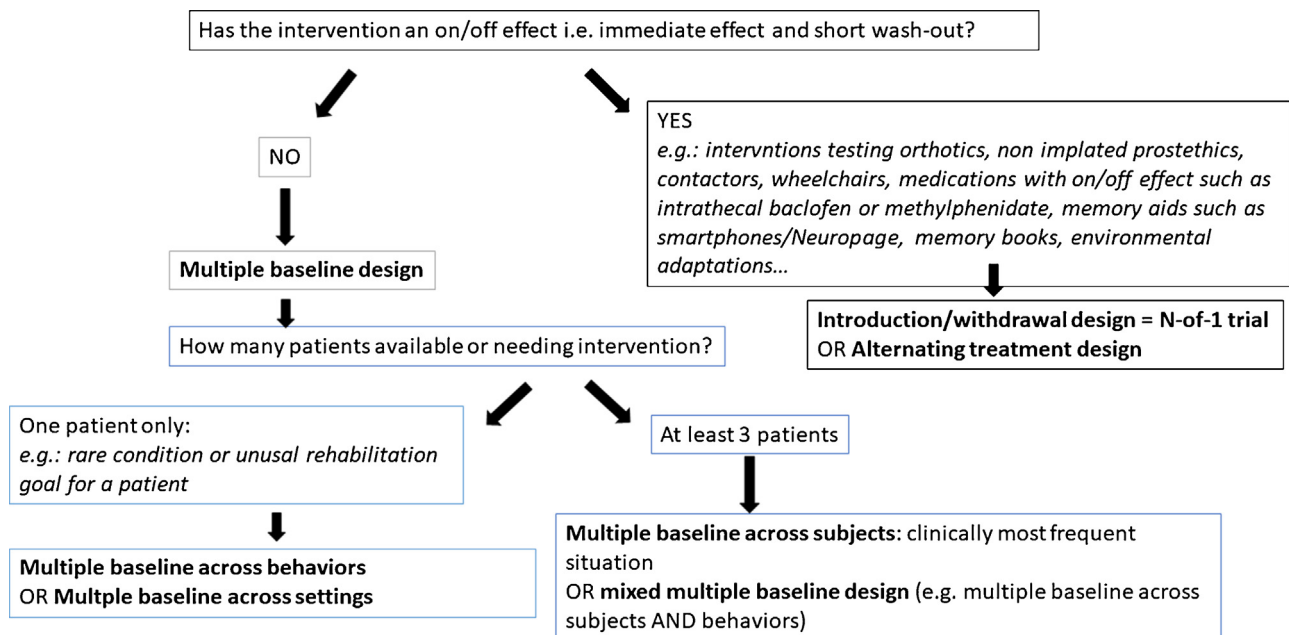
Has the intervention an on/off effect i.e. immediate effect and short wash-out?

NO

Multiple baseline design

How many patients available or needing intervention?

YES
e.g.: intervntions testing orthotics, non implated prostethics, contactors, wheelchairs, medications with on/off effect such as intrathecal baclofen or methylphenidate, memory aids such as smartphones/Neuropage, memory books, environmental adaptations…

Introduction/withdrawal design = N-of-1 trial
OR **Alternating treatment design**

One patient only:
e.g.: rare condition or unusal rehabilitation goal for a patient

At least 3 patients

Multiple baseline across behaviors
OR **Multple baseline across settings**

Multiple baseline across subjects: clinically most frequent situation
OR **mixed multiple baseline design** (e.g. multiple baseline across subjects AND behaviors)

Fig. 5. Which single-case experimental designs (SCED) design to choose.

5.4. Other types of SCEDs

Other types of SCEDs have been described such as the multiple probe, multiple treatment, concurrent schedule designs, most of them being variations of the above designs. They will not be described here but interested readers can refer to paper that listed these types of SCEDs [4,49].

The changing criterion design or changing intensity design, where the patient enters a consecutive phase after s/he reached a predetermined level of performance (see examples [35,38,50]) is interesting in rehabilitation, especially when exploring intervention dosage and timings but its ability to establish causality has been questioned [1].

5.5. Which SCED design to choose

A simple decision tree of SCED design is presented in Fig. 5. In many cases, there are different design options possible to explore the same intervention.

6. Planning the design

Having decided on an outcome measure (target behavior) and the type of design to use, the third step consists of precisely planning the experiment, bearing in mind that, a high standard SCED design should include at least three attempts to demonstrate an intervention effect (i.e. at least 3 change phase in introduction withdrawal design, or at least 3 patients, 3 setting or 3 behaviours in MBD).

Following the SCRIBE statement [22,51], each of the following should be decided prior to the beginning of the experiment and reported:

- the type of design (e.g., withdrawal/reversal);
- the number of phases (including baseline, experimental, maintenance and follow-up phases);
- the order in which the phases are sequenced (e.g., randomized, counterbalanced, data-driven);
- the number of sessions in each phase;

- the number of trials within each session in a phase (i.e., occasions when the outcome measure is being measured);
- the duration of sessions;
- the time interval between sessions;
- if inclusions were concurrent (i.e. all subjects starting baseline at the same time), or non concurrent.

Ideally, inclusion of patients, onset, and subsequent continuance, of data collection should occur concurrently in all patients (i.e., at the same points in time) i.e. concurrent inclusion. If recruitment does not allow to include 3 patients at the same time, it should be specified that inclusions (and data collection) were not concurrent;

- the duration (length) of each phase (see Table 5 and Box 1 for ways of deciding each phase length).

It has historically been recommended to collect data in baseline until the patient stabilizes (i.e. response-guided determination of baseline length). In rehabilitation, this is rarely feasible or ethical because variability in performance is common for many reasons (due to pain, fatigue…) and trends (spontaneous recovery leading to progress in baseline) exist. Clinicians can choose baseline length, based on the following:

- the more points in baseline, the more likely an intervention phase will be able to be differentiated from the baseline if an effect exists;
- standards recommend at the very least three [53] but better five points [2,54] in baseline;
- the greater the variability, the greater the need to have more than five points in baseline;
- the greater the trend towards improvement, the harder it will be to show that the intervention has an additive effect to spontaneous recovery;
- the smaller (more modest) the expected intervention effect, the greater the need to have more points in baseline.

In general, using a response-guided determination of intervention phase length is preferable in a SCED aimed at piloting a novel intervention because the latency, variability and magnitude of

Table 5
Possible ways for determining the length of each phase.

Proportional	Having in mind the time constraints for the study (e.g.: to be completed within 12 weeks) and the number of phases (e.g. ABAB = 4 phases), duration of each phase is calculated to have the same length for each phase (3 weeks each in this example)
Response-guided	In a response-guided design, the phase length depends on the emerging data. For example, a researcher who intervenes only after baseline data have stabilized is using a response-guided design. A researcher who waits until a participant treated reaches a certain level of performance in the first setting before intervening in the second setting is also using a response-guided design [52]
Randomized (see Box 1)	Phase length is determined at random, by a randomisation procedure, often using software (e.g. R program, with SCDA plugin)
Restricted randomization	Very often, due to constraints that can be either clinical (duration of hospital rehabilitation), academic (study to perform in a set amount of time for a student) or financial (funding for a set amount if session/rental of equipment), SCEDs have to be completed in a set amount of time. Planning of the design will depend on the frequency of intervention sessions, the total time required for the intervention Phase length may then be determined at random, but within pre-specified requirements (e.g. need to have at least five measures per phase) and constraints (e.g. total duration of the study, with all phases, to be kept under 16 weeks). See example in Box 2
Mixing response-guided and randomized [52]	This solution has been proposed to increase the study validity when response-guided designs are used, because in response-guided design the risk is to bias the study by introducing the intervention when the experimenter thinks the intervention has the best probability to work (which increases type 1 error) It uses random assignment of participants to intervention times and a data analyst who is blind to which participants enter treatment at which points in time (see appendix in Ferron and Jones [52] for a pragmatic guide to implement this method)

Box 1. Randomization

In group studies, randomization exclusively refers to allocation of participants to intervention groups (i.e. experimental vs. control). By contrast, in SCEDs, a number of design elements can be randomized: (a) the baseline and each phase length (number of measures in each phase); (b) the order of introduction of intervention to different subjects, setting, behaviours in multiple baselines, (c) the order of each condition in alternating treatments.
There are two main advantages of randomization:

- using randomization allows interpretation of SCED data using randomisation tests [55]. Randomization tests are permutation test based on random assignment to test a null hypothesis. This means that in order to work, there must be a sufficient number of possible assignments. In an AB design this is a minimum of 25 measures (2 phases combined) to allow 20 different intervention starting points, (with a minimum 3 points per phase). In a MBD across three subjects, it is enough to have 11 measurements per subject. Note that for a MBD the data of all subjects is used together and a single p value is obtained for all the subjects (or contexts or behaviours) the MBD entails. Different ways of performing a randomization test exist for MBD [56]. When designs do not contain randomization, the ability of randomization tests to control Type I error rates can no longer be demonstrated mathematically [57];
- using randomization increases the internal validity of the study and scientific credibility [58], in the same way as randomization increases the internal validity of a group controlled trial. A randomized experiment, is in many respects to be preferred to a nonrandomized or quasi-experiment because of the statistical control over potential unknown confounding variables related to time (e.g., history, maturation, spontaneous recovery, etc.) the participants, or the setting [9].

In a nonrandomized intervention study, one usually has no control over the variables that covary with the intervention (e.g. change in patient's medication, life stressors, pain, etc.) or with the decision to intervene (e.g. bias by beginning the intervention when the patient seems most motivated), making it very difficult to avoid response-guidance biases or regression artefacts [9].

outcome are usually unknown; both intervention and number of sessions needed may have to be adapted as the experiment runs. Randomized phase length, on the other hand, is preferable in later stages of development of an intervention, to prove its effectiveness or to conclude as to the best therapeutic option for a blinded patient.

6.1. Planning the implementation of the intervention

While researchers primarily aim at demonstrating that an intervention is effective and recruit patients for that purpose (i.e. they will choose patients according to inclusion criteria tailored to who the intervention is likely to benefit), clinicians are most often faced with a patient needing an intervention and they adapt instead the intervention to the patient's unique needs.

One of the reported strengths of single-case methodology is the flexibility of implementation of the intervention [22] because the underlying goal of SCDs is most often to determine "Which intervention is effective for this case (or these cases)?". Departing from initially planned protocol or intervention is allowed, as long as this is explained when reporting the results [22]. The researcher may actively initiate changes that are a departure from protocol or they may be thrust upon the researcher as a result of external factors. If adverse events occur or the intervention is not working sufficiently, then it is acceptable for the researcher to make alterations without necessarily compromising experimental control. For example, if a participant is not responding to an intervention, then the intervention can be manipulated (adapted) while continuing to assess the target behavior [59]. Because of the adaptive nature of SCD designs, non-responders might ultimately be considered "responders" under specific conditions that should be described, in order to allow other clinicians to know how to apply the intervention to particular patients.

6.2. Treatment fidelity/procedural fidelity

Treatment fidelity is the degree to which an intervention is implemented as planned. It helps to increase scientific confidence that the changes in the outcome measures are due to the intervention being tested.

Treatment fidelity consists of two general components:

- treatment integrity, the degree to which a treatment is implemented as intended;
- treatment differentiation, the degree to which two or more interventions or phases differ along critical dimensions.

This means in SCEDs, that between baseline and intervention phases, everything but the intervention (as described in the protocol) should be kept constant. This means for example, keeping constant: time with therapist, quantity of positive feedback, environment, number of opportunities to carry out a task, family support… It means also describing precisely the content of the baseline [60], because the baseline condition may include important ingredients that influence the effectiveness of an intervention (e.g. patients taking a drug during baseline that

influences response to a rehabilitation program or who stops a drug during a rehabilitation program confounding rehabilitation effects).

Borelli described the importance of monitoring treatment fidelity in the following terms: "Without assessment of treatment fidelity, significant results may be a function of either an effective intervention or the influence of other unknown factors added into (or omitted from) the intervention. The danger of this is type 1 error (belief that a treatment effect is significant when it is not) and the potential for dissemination of ineffective treatments. Similarly, if treatment fidelity is not measured and there are non-significant effects, it cannot be known whether these effects are due to an ineffective treatment or to the omission or addition of potentially active or inactive components. The danger of this is type 2 error (erroneous belief that a treatment effect is non-significant) and the potential for discarding effective treatments [2,6]. Thus, treatment fidelity enhances both the internal validity (the treatment is delivered as intended) and external validity (the treatment can be replicated and applied in real world settings)". (p. 1) [61]

Although the preoccupation of treatment fidelity is not specific to SCEDs (but should be essential in any trial testing an intervention), SCEDs have always put a strong emphasis on monitoring treatment fidelity. All SCED standards include the requirement to assess treatment fidelity and many ways exist to assess it [62–65].

7. SCED quality appraisal tools and standards

Reporting guidelines outline what to report in a SCED paper, in the same way as CONSORT [66] is a standard for reporting RCTs. The CENT reporting guidelines [67,68] are intended for medical *N*-of-1 trials, the SCRIBE 2016 guidelines [22,51] are intended for SCEDs in the behavioral sciences (and largely used in neuropsychological rehabilitation). They can all be found on the Equator network (http://www.equator-network.org/).

Other useful tools are scales appraising the quality of a SCED (see in particular the RoBiNT scale [2] that is a revision of the SCED scale [53]) and standards (see in particular the "What Works Clearinghouse – WWC standards of Kratochwill et al. [49,54] written for the field of education but widely used, including in rehabilitation studies). These document what constitutes a high-quality SCED. To meet the standards in WWC, the following design criteria are required:

- the intervention must have been systematically manipulated with the researcher determining when and how the intervention is implemented;
- each outcome measure must be measured systematically over time and a measure of inter-rater reliability (IRR) for no less than 20% of sessions must be reported;
- the study must include at least three attempts to demonstrate an intervention effect (i.e. at least 3 change phase in introduction withdrawal design, or at least 3 patients, 3 setting or 3 behaviours in MBD);
- for a phase to qualify as an attempt to demonstrate an intervention effect, the phase must have at least three (with a preference of at least five) data points.

Other important recommendation include:

- blinding of the patient and the assessor to the phase (which encourages that the repeated measure be performed by a person external to the therapist providing the intervention and knowing the phase the patient is currently at);

- randomization (see Box 1);
- precise description of the intervention and assessment of procedural fidelity AND precise description of the baseline condition;
- precise description of participants and of the setting of the study;
- a measure of generalization;
- provision of raw data record and statistical analysis.

It is a sensible step in designing a SCED protocol, to go through these standards, and reporting guidelines (and perhaps use the RoBiNT scale to rate the intended protocol), in order to make adjustments before the study starts.

8. Run the experiment

We have presented the first five steps required to design a SCED, before running the study. These are summarized in Table 6. After having made adjustments according to standards, the experiment may begin.

As the experiment runs, the investigator should:

- monitor unexpected events;
- keep information on any deviation or adaptation from original design and/or intervention content;
- evaluate procedural fidelity for a minimum of 20% sessions;
- do a double measure of the repeated outcome measure for at least 20% of measures in order to calculate the inter-rater reliability of the outcome measure.

Note that it would be worth assessing these last two variables before the beginning of the experiment (as well as during the experiment), because a SCED that lacks adequate IRR or procedural fidelity will be unable to provide any conclusion as to the effectiveness of the intervention being tested.

9. Representing the results – a few rules

Because SCEDs are fundamentally different from traditional group research, SCEDs use specific rules in relation to reporting results.

- Results of a SCED are typically presented as a time series graph."The metric used on the horizontal axis of graphed data should be in units of real time (i.e., days, weeks, etc.) rather than session number. Providing an exact chronology of the time interval between sessions allows the reader to accurately evaluate patterns of consistency between similar phases and effect latency following intervention onset." (p. 23) [22].

  Phase change/introduction of an intervention is typically presented as a vertical line.

  In multiple baseline designs, if the inclusion of participants was not concurrent (i.e. patients included at different times), it is better representing the time frames between patient inclusions.
- Results should not be averaged.

  Averaging means that important features of data may be lost, such as (a) stability of the initial baseline phase, (b) variability and trends within a phase, (c) degree of consistency between similar phases (e.g., intervention phases), (d) the degree of overlap between baseline and intervention phases, (e) magnitude of effect latency following intervention phase onset.
- Visual aids for graph interpretation may be represented on the graph to help visual interpretation. Note that visual aids are drawn separately for each AB comparison (so for example in a

Box 2. Example of design choice, following SCRIBE design details, using restricted randomisation

Example: planning the design for a speech therapy intervention for patients with aphasia.
Intervention: gesture either through classical therapist guided session or using mute films.
Outcome measure: number of words correctly signed and named from a list of 40 words the patient cannot name or gesture.
Design: Multiple baseline across subjects
Timing requirements: Intervention to be completed in a 12 weeks period. Patients seen 5 times a week in stroke rehabilitation centre.

SCRIBE design detail	Application to the example
The type of design	Multiple baseline (because a maintenance of effect is desirable and expected)
	Across five patients (to meet the standard of at least 3 demonstrations of effects, taking into account possible loss to follow-up: possible premature return home or difficulty in maintaining the program intensity)
The number of phases (including baseline, experimental, maintenance and follow-up phases)	Baseline (A) with no intervention (during which an extensive aphasia evaluation will be performed) Intervention 1: mute films (phase B) Intervention 2: gesture therapy (phase C) Follow-up phase (without intervention)
The duration (length) of each phase	2 weeks minimum (because interventions 1 and 2 are unlikely to be beneficial if length < 10 sessions)
The number of sessions in each phase	Five days a week program, so a minimum of $2 \times 5 = 10$ sessions per intervention phase
The number of trials within each session in a phase (i.e., occasions when the dependent variable is being measured)	Because of intervention intensity and the duration of administration of the repeated measure (10 minutes), it was decided to administer the measure at every other session.*
The order in which the phases are sequenced (e.g. randomized, counterbalanced, data-driven)	Restricted randomization in order to have at least five measures per phase and a minimum of two weeks of each intervention.
The duration of sessions	45 minutes
The time interval between sessions	One session a day from Monday to Friday, no intervention during week-end

* it would be convenient to acquire the 5 points in baseline using measures every session, but the frequency of repeated measures in baseline and intervention has to be the same.

Example of calculation of phase length randomisation restrictions:
-Calculation of the minimum time required to complete all phases of the design: Minimum of 5 measures per phase = 5 (baseline) +5 (intervention 1) +5 (intervention 2) = 15 measures, = 30 sessions (because one measure is taken every two sessions = 6 weeks minimum to complete the whole design if all phases are the same proportional and minimum length.)
-Total available time for the experiment: 12 weeks.

-Introduction of intervention can therefore be staggered for 5 different patients i.e. intervention starting after 5, 6, 7, 8, 9 or 10 measures (so 10 to 20 sessions i.e. 2 to 4 weeks of baseline) i.e. randomizing which patient starts intervention first, within the limit of starting between 5 and 10 measures.
-Baseline is followed by an intervention 1 phase of randomized length with a minimum of 5 measures (i.e. 10 sessions = 2 weeks of intervention) in the phase. Patients switch to the second intervention after 5, 6, 7, 8, 9 or 10 measures (so after 10 to 20 sessions i.e. 2 to 4 weeks of intervention 1).

-Order of interventions (Intervention 1 = mute films; Intervention 2 = gesture therapy) can also be randomized.

This design will allow to use a randomization test to interpret data.

Table 6
Steps for designing and running a SCED.

Step 1	Find an outcome measure that can be repeated many times throughout the experiment, without major retest effect, that is relatively short to administer and that reflects intervention target	See Tables 2 and 3
Step 2	Choose any other outcome measures to be used (control measures, generalisation measures, implementation measures, other standardized measures)	See Table 4
Step 3	Choose a design	See Fig. 5
Step 4	Plan design details ± randomisation	See Table 5 and Box 1
Step 5	Adapt/describe intervention details and choose a measure of procedural fidelity	
Step 6	Review protocol according to standards	
Step 7	Run the experiment, monitoring unexpected events, keeping information on any deviation or adaptation from original design and/or intervention content, evaluating procedural fidelity, and assessing IRR of the outcome measure	Keep in mind that in SCEDs at least three measures will need to be reported: a graph with the main repeated outcome measure across phases; the IRR of the main outcome measure; a measure of procedural fidelity
Step 8	Plot and analyse data visually	An aid for plotting data and performing visual analysis aids can be found at https://manolov.shinyapps.io/overlap
Step 9	Use statistics to confirm visual analysis	

SCED: single-case experimental designs; IRR: inter-rater reliability.

MBD across subjects, the data of each subject will be analysed separately).

- Statistical analysis should never standalone (as is usually the case in group studies) but should follow the graph and be interpreted in conjunction with the visual analysis of the graph.

10. Visual interpretation of SCED data

Historically, SCED data was interpreted only visually. Visual analysis aids useful in rehabilitation studies include:

- level lines (see Fig. 6a), corresponding to the median or mean of all data points of one phase, that allows to compare change in levels between phases. This is more useful in interventions with on/off effects because in interventions showing slow changes, the first measures of the intervention phase (when the intervention was not effective yet), may lower the level of intervention phase;
- trend lines (see Fig. 6b and c) i.e. the tendency of data point to go upwards or downwards within a phase. This is a key analysis in rehabilitation as patients are often not stable within a phase, but show trend towards improvement (e.g. in the acute phase post stroke) or towards decline (e.g. degenerative disease). Trend

lines can be used to project baseline trend into the intervention phase allowing visualization of whether the trend continues the same (reflecting a progress likely to be due to the follow-up of spontaneous recovery seen in baseline and unlikely to be an intervention effect) or accelerates (reflecting an intervention effect). A trend may be more or less obvious and it has been proposed that trend lines can be considered representative of the data within a phase only if 80% of the data fall into the trend line envelope [69], see Fig. 6c. The trend of a phase can by drawn by hand, using the split-middle trend line – a line joining the median (or alternatively the mean) of each half phase (see Appendix 1 in e-component).

Trend lines and level lines have been combined by Fisher et al. [70] in his "dual criterion method". In this method, one first calculates the baseline level (mean line based on baseline data) and then superimposes it over the subsequent data path. Next, a split-middle trend line is calculated based on baseline data and extended into the subsequent phase. An effect (i.e., a change in data across phases) is said to exist when a pre-specified number of data points have fallen above each of the lines according to a binomial equation. Fisher et al. demonstrated that the dual criterion method generally resulted in fewer Type I errors:

- variability bands, correspond to the spread of data points within a phase, and whatever the design, the smaller the variability
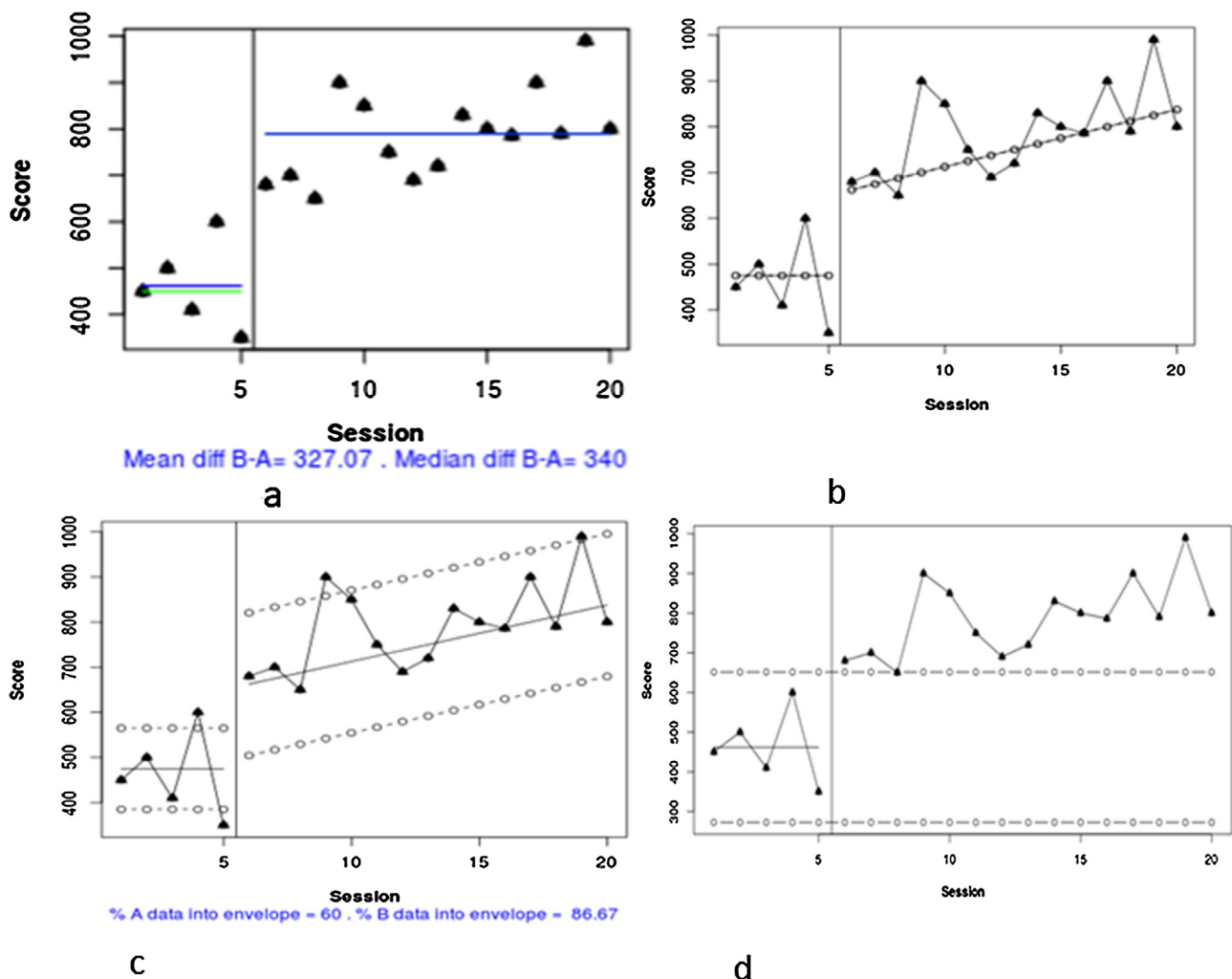

Fig. 6. Visual aids computed on a free single-case experimental designs (SCED)-analysis website: https://manolov.shinyapps.io/overlap. a: level; b: trend line; c: trend line stability envelope; d: two-standard deviation band.

within a phase, the easier to detect an intervention effect. The most used variability technique is the two-standard deviation band [17,71], drawn by calculating the mean of a phase and adding and subtracting 2 standard deviation (taking the standard deviation of the studied phase only) from it [17]; see Fig. 6d;

- overlap is the calculation of points overlapping between phases: the more overlap between two phases, the less likely an intervention effect exists. Conversely, if two phases show no overlap in the measures, the intervention is likely to have an effect of greater magnitude. Many overlap techniques exist, from simple hand calculation, often used in the history of SCEDs but highly criticised (percent of phase B data points which exceed the single highest phase A data point; percentage of phase B data points exceeding the median of the baseline phase [PEM]; percent of all data remaining after removing the minimum number of data points which would eliminate all data overlap between phases A and B; . . . [72]) to more complex calculations of overlap such as "non-overlap of all pairs – NAP" [72], Tau-U and baseline-corrected Tau-U [73] (that can be calculated from: http://ktarlow.com/stats/tau/). This last measure of overlap is probably the most useful in rehabilitation studies as it calculates overlap even in highly variable data, correcting for trend.

Most visual analysis aids can be drawn by hand (see tutorials on visual aids: What Works Clearinghouse [p. 17–21], Perdices and Tate [17], Lane and Gast [69]), but useful web-based resources allow computation by entering data in a text format (see especially the user-friendly website https://manolov.shinyapps.io/Overlap).

Systematic visual analysis has shown to increase the accuracy of interpretation of SCED data, especially in relation to the conservative dual criterion [70], as shown in several studies [74–76].

Some of the visual analysis criteria used in behavioral sciences [49], such as consistency across phase, immediacy of effect, coherence of measures within a phase are not always appropriate for SCEDs in rehabilitation because effects in rehabilitation are rarely immediate, and measures within a phase often variable (due to fatigue, pain, spontaneous recovery. . .).

11. Statistical interpretation of results

Many statistical procedures for interpreting SCED data exist and there is dynamic ongoing research on statistical analysis of SCED data. There is no consensus on which statistical procedure to choose to analyse SCED data [2,12,51,77]. Providing clinicians with statistical methods to interpret SCED data is beyond the scope of this paper. Interested readers are referred to the SCBIBE appendix [22] for the available statistics for SCED, to Manolov et al. [77] discussion on criteria for choosing a statistical procedure and to Manolov and Moeyaert's recommendations and practical statistical computing tutorials [78,79].

12. Specific issues in interpreting SCED data

When interpreting SCED results, the following should be kept in mind:

- visual analysis may be quite subjective, which can lead to increases in type 1 errors (estimated in several studies to be 0.24–0.28), especially when a trend exists and when analysis is done without an adequate visual analysis training [52];
- SCED data are usually autocorrelated [80] (because the behavior of a subject is rarely random), which violates the assumption of most statistical tests. Matyas and Greenwood estimated type

1 error (concluding that an intervention has an effect when it does not) rates at 0.24 on data when autocorrelated;
- reliability in measuring the outcome measure is crucial for SCED interpretation. Beside the problem of choosing an outcome that is easily measurable, the fact that the measure is repeated many times may make it less reliable because of:
  - possible retest effect,
  - observer drift (possibility that observers may change their observational definitions of the construct being measured over time, thereby not making scores comparable across phases of the experiment),
  - observational bias (possibility that observers may be influenced by a variety of factors associated with desired experimental outcomes),
  - reactivity (possibility that observational scores are higher as a result of the researcher monitoring the observers or observational process). Reporting the IRR and blinding the observer to the treatment phase are crucial to avoid those biases, especially if the outcome measure seems subjective (e.g. in Logan's study assessing "happiness" – a highly subjective construct-in children with multiple profound disabilities across different teaching conditions, happiness ratings were based on video-taped sessions visualized in random order by a blinded judge and rated on specific criteria predetermined and specific for each child [32]).

13. The issue of replication and generalizability of findings

The concept of applicability or generality is based on the assumption that inferences can be drawn from the condition in which an intervention effect was demonstrated, to other conditions based on known similarities and differences between these conditions [81].

A major criticism of SCEDs is that the findings cannot be generalized to other patients and that demonstrating an effect on three patients is not a proof of intervention effectiveness beyond those particular patients. To emphasise this, abstracts sometimes end with "These results cannot be generalized beyond the patients included in these trials." [82]

We argue however, as many other methodological papers, that this issue is not specific to SCEDs. RCTs report group means that mask the differential responses to treatments [1]. "Even in the most successful group design, there are individuals whose behavior remains unaffected, or is made worse, by the treatment" (p. 173) [59]. Besides RCTs often use very restrictive inclusion criteria that makes the results difficult to apply to "the messy complexities of clinical problems" [16, p. 2, 83] and their generalizability may therefore be poor. Further, many interventions, tested through group designs, have never been able to be replicated [84]. In rehabilitation, we are used to adapting rehabilitation techniques, devices and assistive technology, based on the patient's characteristics, personal and environmental factors [85]. In the age of personalized medicine, it seems almost unthinkable to apply blindly the result of an RCT to all patients.

Although, just like RCTs, SCED cannot prove that an intervention will work for all patients, there are ways to increase the generalizability of SCED findings:

- the first way is replication. As mentioned earlier, the highest SCED standards require that the experimental effect be replicated at least three times within the SCED study [2,54,59]. This type of replication increases the internal validity of the SCED design and is known as direct replication [9,22]. It refers to the replication of the experimental effect within the

design. Therefore, SCEDs using ABA designs, or MBD across 2 patients/settings/behaviours do not fulfil the replication criterion.

The second type of replication is replicating the whole SCED, with the same intervention but with other patients, settings, clinicians (also called systematic replication [9,22]). Each replication will add information regarding the generalizability of the findings. For example, Feeney and Ylvisaker [40,86,87] replicated across different students with TBI and across different schools their intervention for executive dysfunction first published by Feeney and Ylvisaker [88].

A number of studies meta-analysing SCEDs have been published [89–93] and contribute to establishing the evidence acquired from use of SCEDs. For example, Moss and Nicholas used data from 23 SCEDS, comprising a total of 57 subjects and show that for patients more than one year post onset of aphasia, time post onset is not related to response to treatment;

- the second way that generalisability is increased is by the precise description of patient's characteristics, history, associated deficits etc., to allow those reading the paper to apply the findings to similar patients.
- the third complementary way is the assessment of "social validity". Social validity reflects contextual aspects that will influence the replication and/or the use of the intervention in the real world and that might augment or inhibit the effectiveness of an intervention. These aspects include:
  ○ whether the treatment was easy or difficult to implement,
  ○ whether any unintended consequences developed in conjunction with the intervention, or,
  ○ whether the interventionist will continue or expand the use of the treatment and why. The reporting of social validity data is useful for identifying contextual trends.

When can we consider an intervention as effective using SCEDs? A number of criteria have been established. These criteria relate to two main issues:

- the quality of SCED research taken in consideration;
- the quantity of replication (an intervention being more likely effective, if it has been replicated many times).

Regarding the first criteria, standards exist that allow clinicians to evaluate the studies methodological qualities [2,54] (see paragraph on standards) and just like in group studies, only methodologically sound studies should be considered as evidence to implement an intervention in routine clinical practice. Romeiser Logan et al. [94] proposed a grading of levels of evidence for SCEDs, based on the SCED design type, the use of randomization, the number of direct replication. The second issue relates to the amount of replication needed to have confidence that the intervention effect will be present in everyday clinical practice. Although this is a 'rule of thumb', the 5-3-20 rule has been proposed by What works Clearinghouse: a minimum of five SCD research papers examining the intervention that meet quality standards, conducted by at least three different research teams in three different geographical locations with a combined number of 20 single-cases across the papers [49].

14. Challenges and prospects for SCEDs

Most papers using SCED appear to evaluate cognitive interventions (although this may be a search bias as traditionally SCED methodologies have been used in neuropsychology and education with clear references to SCED in the title/abstract, whereas papers from other domains do not always used "SCED" methodology key

words in pubmed). We argue that SCEDs are a useful tool in areas outside neuropsychological rehabilitation in PMR and would be helpful if applied to motor interventions, therapeutic education, adaptive devices and assistive technology choices etc.

If we are to increase the use of well-designed SCEDs in rehabilitation we suggest that it is necessary to:

- provide appropriate teaching on SCEDs to students of all rehabilitation fields;
- convince journal editors and funding institution that SCEDs are not case reports but high-quality research. For example, a SCED should not appear under "case report" section of journals, because they are fundamentally different from retrospectively reporting a case;
- SCED protocols should be entitled to be published as pre-registered trials (note this is beginning slowly, especially in N of 1 trails, see for example [20]);
- endorse reporting guidelines such as SCRIBE or CENT to standardize the reporting of SCEDs in PRM journals;
- encourage and support students to use SCED methodologies for their degree thesis, rather than poorly designed group studies, lacking power or descriptive case reports or retrospective data analysis.

The best opportunity for PRM research is perhaps not to increase the number of large scale studies, using group designs, but make the most of the small scale research done routinely in all rehabilitation centres, hospitals and private practices by rehabilitation students and professionals. Such "small" scale research often uses an inadequate methodology, while the same amount of work could yield a robust SCED. For example, students often use small groups tested pre-and post-intervention, whilst including just three patients but using a multiple baseline design could give stronger evidence of the intervention being tested. Another advantage is to have the student entirely responsible for his/her protocol, rather than joining ongoing research. This could hopefully result in acquiring better scientific reasoning than just performing a protocol. Another opportunity is to use multicentre SCEDs, for example creating collaborations between two PRM students for testing the same intervention in two different cities, using SCEDs: this type of design is easy to complete within an academic year (for example including two patients per student). Having pairs of students allows them to measure the IRR and procedural fidelity for 20% of sessions as recommended in standards.

15. Conclusion

SCEDs offer an opportunity to conduct high-quality research with a limited number of patients, to evaluate the effects of intervention for patients individually and to explore the effectiveness of novel approaches devised by local therapists outside the world of research.

SCEDs follow a series of rules for designing the experiment and interpreting data, but greater awareness of, and skills in, this methodology are needed. Following the steps described in this paper or other tutorials may be a first step for clinicians to use SCEDs.

Disclosure of interest

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.rehab.2017.12.002.

References

[1] Graham JE, Karmarkar AM, Ottenbacher KJ. Small sample research designs for evidence-based rehabilitation: issues and methods. Arch Phys Med Rehabil 2012;93:S111–6.

[2] Tate RL, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, et al. Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: the 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. Neuropsychol Rehabil 2013;23:619–38.

[3] Guyatt GH, Heyting A, Jaeschke R, Keller J, Adachi JD, Roberts RS. N of 1 randomized trials for investigating new drugs. Control Clin Trials 1990;11:88–100.

[4] Byiers BJ, Reichle J, Symons FJ. Single-subject experimental design for evidence-based practice. Am J Speech Lang Pathol 2012;21:397–414.

[5] Cohen LL, Feinstein A, Masuda A, Vowles KE. Single-case research design in pediatric psychology: considerations regarding data analysis. J Pediatr Psychol 2014;39:124–37.

[6] Lane JD, Ledford JR, Gast DL. Single-case experimental design: current standards and applications in occupational therapy. Am J Occup Ther 2017;71. 7102300010p1.

[7] Gorczynski P. The use of single-case experimental research to examine physical activity, exercise, and physical fitness interventions: a review. J Appl Sport Psychol 2013;25:148–56.

[8] Haegele JA, Hodge SR. The applied behavior analysis research paradigm and single-subject designs in adapted physical activity research. Adapt Phys Activ Q 2015;32:285–301.

[9] Onghena P, Edgington ES. Customization of pain treatments: single-case design and analysis. Clin J Pain 2005;21:56–68 [discussion 69–72].

[10] Dallery J, Cassidy RN, Raiff BR. Single-case experimental designs to evaluate novel technology-based health interventions. J Med Internet Res 2013;15:e22.

[11] Geist K, Hitchcock JH. Single-case design studies in music therapy: resurrecting experimental evidence in small group and individual music therapy clinical settings. J Music Ther 2014;51:293–309.

[12] Smith JD. Single-case experimental designs: a systematic review of published research and current standards. Psychol Meth [Internet] 2012;17 [Cited 2017 Jun 12. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3652808/].

[13] Evans JJ, Gast DL, Perdices M, Manolov R. Single-case experimental designs: introduction to a special issue of neuropsychological rehabilitation. Neuropsychol Rehabil 2014;24:305–14.

[14] Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013;14:365–76.

[15] Normand MP. Less is more: psychologists can learn more by studying fewer people. Front Psychol [Internet] 2016;7 [Cited 2017 Jun 19. Available from: http://journal.frontiersin.org/Article/10.3389/fpsyg.2016.00934/abstract].

[16] Romeiser-Logan L, Slaughter R, Hickman R. Single-subject research designs in pediatric rehabilitation: a valuable step towards knowledge translation. Dev Med Child Neurol [Internet] 2017 [Cited 2017 Mar 18. Available from: http://doi.wiley.com/10.1111/dmcn.13405].

[17] Perdices M, Tate RL. Single-subject designs as a tool for evidence-based clinical practice: are they unrecognised and undervalued? Neuropsychol Rehabil 2009;19:904–27.

[18] Johnston MV, Smith RO. Single-subject designs: current methodologies and future directions. Occup Participation Health 2010;30:4–10.

[19] Gharebaghy S, Rassafiani M, Cameron D. Effect of cognitive intervention on children with ADHD. Phys Occup Ther Pediatr 2015;35:13–23.

[20] Nikles CJ, McKinlay L, Mitchell GK, Carmont S-AS, Senior HE, Waugh M-CA. et al. Aggregated n-of-1 trials of central nervous system stimulants versus placebo for paediatric traumatic brain injury – a pilot study. Trials 2014;15:1.

[21] McKerracher G, Powell T, Oyebode J. A single-case experimental design comparing two memory notebook formats for a man with memory problems caused by traumatic brain injury. Neuropsychol Rehabil 2005;15:115–28.

[22] Tate RL, Perdices M, Rosenkoetter U, McDonald S, Togher L, Shadish W, et al. The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016: explanation and elaboration. Arch Sci Psychol 2016;4:10–31.

[23] Coker P, Lebkicher C, Harris L, Snape J. The effects of constraint-induced movement therapy for a child less than one year of age. Neurorehabil 2009;24:199–208.

[24] Mattacola CG, Lloyd JW. Effects of a 6-week strength and proprioception training program on measures of dynamic balance: a single-case design. J Athl Train 1997;32:127–35.

[25] Kirsch NL, Shenton M, Spirl E, Rowan J, Simpson R, Schreckenghost D, et al. Web-based assistive technology interventions for cognitive impairments after traumatic brain injury: a selective review and two case studies. Rehabil Psychol 2004;49:200–12.

[26] Blyth KM, McCabe P, Madill C, Ballard KJ. Ultrasound in dysphagia rehabilitation: a novel approach following partial glossectomy. Disabil Rehabil 2017;39:2215–27. http://dx.doi.org/10.1080/09638288.2016.1219400 [Epub 2016 Dec 28.].

[27] Raymer AM, Ciampitti M, Holliway B, Singletary F, Blonder LX, Ketterson T, et al. Semantic-phonologic treatment for noun and verb retrieval impairments in aphasia. Neuropsychol Rehabil 2007;17:244–70.

[28] Ten Berge SR, Boonstra AM, Dijkstra PU, Hadders-Algra M, Haga N, Maathuis CGB. A systematic evaluation of the effect of thumb opponens splints on hand function in children with unilateral spastic cerebral palsy. Clin Rehabil 2012;26:362–71.

[29] Ouellet M-C, Morin CM. Efficacy of cognitive-behavioral therapy for insomnia associated with traumatic brain injury: a single-case experimental design. Arch Phys Med Rehabil 2007;88:1581–92.

[30] Sunderland A, Walker CM, Walker MF. Action errors and dressing disability after stroke: an ecological approach to neuropsychological assessment and intervention. Neuropsychol Rehabil 2006;16:666–83.

[31] Baldwin VN, Powell T. Google calendar: a single-case experimental design study of a man with severe memory problems. Neuropsychol Rehabil 2015;25:617–36 [Epub 2014 Sep 29].

[32] Logan KR, Jacobs HA, Gast DL, Murray AS, Daino K, Skala C. The impact of typical peers on the perceived happiness of students with profound multiple disabilities. Research and practice for persons with severe disabilities 1998;23:309–18.

[33] Chang Y-J, Wang T-Y, Chen Y-R. A location-based prompting system to transition autonomously through vocational tasks for individuals with cognitive impairments. Res Dev Disabil 2011;32:2669–73.

[34] Hoogeboom TJ, Kwakkenbos L, Rietveld L, den Broeder AA, de Bie RA, van den Ende CHM. Feasibility and potential effectiveness of a non-pharmacological multidisciplinary care programme for persons with generalised osteoarthritis: a randomised, multiple baseline single-case study. BMJ Open 2012;2:e001161.

[35] Svoboda E, Richards B, Leach L, Mertens V. PDA and smartphone use by individuals with moderate-to-severe memory impairment: application of a theory-driven training programme. Neuropsychol Rehabil 2012;22:408–27.

[36] Ferland MB, Larente J, Rowland J, Davidson PSR. Errorless (re)learning of daily living routines by a woman with impaired memory and initiation: transferable to a new home? Brain Inj 2013;27:1461–9.

[37] Krasny-Pacini A, Limond J, Evans J, Hiebel J, Bendjelida K, Chevignard M. Context-sensitive goal management training for everyday executive dysfunction in children after severe traumatic brain injury. J Head Trauma Rehabil 2014;29:E49–64.

[38] Ducharme JM, Davidson A, Rushford N. Treatment of oppositional behavior in children of parents with brain injury and chronic pain. J Emotional Behav Disord 2002;10:241–8.

[39] Lui M, Falk TH, Chau T. Development and evaluation of a dual-output vocal cord vibration switch for persons with multiple disabilities. Disabil Rehabil Assistive Technol 2012;7:82–8.

[40] Feeney TJ. Structured flexibility: the use of context-sensitive self-regulatory scripts to support young persons with acquired brain injury and behavioral difficulties. J Head Trauma Rehabil 2010;25:416–25.

[41] Cosbey J, Muldoon D. EAT-UP™ Family-centered feeding intervention to promote food acceptance and decrease challenging behaviors: a single-case experimental design replicated across three families of children with autism spectrum disorder. J Autism Dev Disord 2017;47:564–78.

[42] Diamond MF, Ottenbacher KJ. Effect of a tone-inhibiting dynamic ankle-foot orthosis on stride characteristics of an adult with hemiparesis. Phys Ther 1990;70:423–30.

[43] Tunnard C, Wilson BA. Comparison of neuropsychological rehabilitation techniques for unilateral neglect: an ABACADAEAF single-case experimental design. Neuropsychol Rehabil 2014;24:382–99. http://dx.doi.org/10.1080/09602011.2013.872041 [Epub 2014 Jan 13.].

[44] Sumitani M, Rossetti Y, Shibata M, Matsuda Y, Sakaue G, Inoue T, et al. Prism adaptation to optical deviation alleviates pathologic pain. Neurology 2007;68:128–33.

[45] Nikles J, Mitchell G, McKinlay L, Waugh M-C, Epps A, Carmont S-A, et al. A series of n-of-1 trials of stimulants in brain injured children. NeuroRehabilitation 2007;40:11–21.

[46] Naganuma GM, Billingsley FF. Effect of hand splints on stereotypic hand behavior of three girls with Rett syndrome. Phys Ther 1988;68:664–71.

[47] Carey LM, Matyas TA. Training of somatosensory discrimination after stroke: facilitation of stimulus generalization. Am J Phys Med Rehabil 2005;84:428–42.

[48] Boman I-L, Bartfai A, Borell L, Tham K, Hemmingsson H. Support in everyday activities with a home-based electronic memory aid for persons with memory impairments. Disabil Rehabil Assistive Technol 2010;5:339–50.

[49] Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse

[50] Arco L. Neurobehavioural treatment for obsessive-compulsive disorder in an adult with traumatic brain injury. Neuropsychol Rehabil 2008;18:109–24.

[51] Tate RL, Perdices M, Rosenkoetter U, Shadish W, Vohra S, Barlow DH, et al. The Single-Case Reporting Guideline In BEhaviour Interventions (SCRIBE) 2016 statement. Neuropsychol Rehabil 2017;27:1–15.

[52] Ferron J, Jones PK. Tests for the visual analysis of response-guided multiple baseline data. J Exp Educ 2006;75:66–81.

[53] Tate RL, McDonald S, Perdices M, Togher L, Schultz R, Savage S. Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the Single-Case Experimental Design (SCED) Scale. Neuropsychol Rehabil 2008;18:385–401.

[54] Kratochwill TR, Hitchcock JH, Horner RH, Levin JR, Odom SL, Rindskopf DM, et al. Single-case intervention research design standards. Remedial Spec Educ 2013;34:26–38.

[55] Heyvaert M, Onghena P. Analysis of single-case data: randomisation tests for measures of effect size. Neuropsychol Rehabil 2014;24:507–27.

[56] Levin JR, Ferron JM, Gafurov BS. Comparison of randomization test procedures for single-case multiple baseline designs. Dev Neurorehabil 2016;1–21 [Published online: 01 Jul 2016 https://doi.org/10.1080/17518423.2016.1197708].

[57] Ferron J, Foster-Johnson L, Kromrey JD. The functioning of single-case randomization tests with and without random assignment. J Exp Educ 2003;71:267–88.

[58] Kratochwill TR, Levin JR. Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. Psychol Meth 2010;15:124–44.

[59] Horner R, Carr E, Halle J, McGee G, Odom SL, Wolery M. The use of single-subject research to identify evidence-based practice in special education. Exceptional Children 2005;71:165–79.

[60] Lane K, Wolery M, Reichow B, Rogers L. Describing baseline conditions: suggestions for study reports. J Behav Educ 2007;16:224–34.

[61] Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. J Public Health Dentistry 2011;71:S52.

[62] Borrelli B, Sepinwall D, Ernst D, Bellg AJ, Czajkowski S, Breger R, et al. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. J Consult Clin Psychol 2005;73:852–60.

[63] Hart T. Treatment definition in complex rehabilitation interventions. Neuropsychol Rehabil 2009;19:824–40.

[64] Ledford JR, Gast DL. Measuring procedural fidelity in behavioural research. Neuropsychol Rehabil 2014;24:332–48.

[65] Spillane V, Byrne MC, Byrne M, Leathem CS, O'Malley M, Cupples ME. Monitoring treatment fidelity in a randomized controlled trial of a complex intervention. J Adv Nurs 2007;60:343–52.

[66] Schulz FK, Altman DG, Consort DM. Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010. http://dx.doi.org/10.1136/bmj.c332340 [Published online 2010 Mar 23].

[67] Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 statement. BMJ 2015;350:h1738.

[68] Shamseer L, Sampson M, Bukutu C, Schmid CH, Nikles J, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015: explanation and elaboration. BMJ 2015;350:h1793.

[69] Lane JD, Gast DL. Visual analysis in single-case experimental design studies: brief review and guidelines. Neuropsychol Rehabil 2014;24:445–63.

[70] Fisher WW, Kelley ME, Lomas JE. Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. J Appl Behav Anal 2003;36:387–406.

[71] Pfadt A, Wheeler DJ. Using statistical process control to make data-based clinical decisions. J Appl Behav Anal 1995;28:349.

[72] Parker RI, Vannest K. An improved effect size for single-case research: nonoverlap of all pairs. Behav Ther 2009;40:357–67.

[73] Tarlow KR. An improved rank correlation effect size statistic for single-case designs: baseline-corrected Tau. Behav Modif 2017;41:427–67.

[74] Wolfe K, Slocum TA. A comparison of two approaches to training visual analysis of AB graphs. J Appl Behav Anal 2015;48:472–7.

[75] Lanovaz MJ, Huxley SC, Dufour M-M. Using the dual-criteria methods to supplement visual inspection: an analysis of nonsimulated data. J Appl Behav Anal 2017;50:662–7.

[76] Young ND, Daly EJ. An evaluation of prompting and reinforcement for training visual analysis skills. J Behav Educ 2016;25:95–119.

[77] Manolov R, Gast DL, Perdices M, Evans JJ. Single-case experimental designs: reflections on conduct and analysis. Neuropsychol Rehabil 2014;24:634–60.

[78] Manolov R, Moeyaert M. How can single-case data be analyzed? Software resources, tutorial, and reflections on analysis. Behav Modif 2017;41:179–228.

[79] Manolov R, Moeyaert M. Recommendations for choosing single-case data analytical techniques. Behav Ther 2017;48:97–114.

[80] Shadish WR, Sullivan KJ. Characteristics of single-case designs used to assess intervention effects in 2008. Behav Res Meth 2011;43:971–80.

[81] Gast DL, Ledford JR. Replication. In: Single-subject research methodology in behavioral sciences 1st ed., New York: Routledge; 2009. p. 110–28.

[82] Ferreira JJ, Mestre T, Guedes LC, Coelho M, Rosa MM, Santos AT, et al. Espresso coffee for the treatment of somnolence in Parkinson's disease: results of n-of-1 trials. Front Neurol [Internet] 2016;7 [Cited 2017 Jun 9. Available from: http://journal.frontiersin.org/Article/10.3389/fneur.2016.00027/abstract].

[83] Shelton JD. Evidence-based public health: not only whether it works, but how it can be made to work practically at scale. Global Health Sci Pract 2014;2:253–8.

[84] Open Science Collaboration. Estimating the reproducibility of psychological science. Science 2015;349:aac4716.

[85] World Health Organization (WHO). International classification of functioning, disability and health (ICF) [Internet]. Geneva: WHO; 2001 [Cited 2015 Dec 29. Available from: http://www.who.int/classifications/icf/en/].

[86] Feeney TJ, Ylvisaker M. Context-sensitive cognitive-behavioral supports for young children with TBI: a second replication study. J Positive Behav Interv 2008;10:115–28.

[87] Feeney T, Ylvisaker M. Context-sensitive cognitive-behavioural supports for young children with TBI: a replication study. Brain Inj 2006;20:629–45.

[88] Feeney TJ, Ylvisaker M. Context-sensitive behavioral supports for young children with TBI: short-term effects and long-term outcome. J Head Trauma Rehabil 2003;18:33–51.

[89] Moss A, Nicholas M. Language rehabilitation in chronic aphasia and time postonset: a review of single-subject data. Stroke 2006;37:3043–51.

[90] Wehmeyer ML, Palmer SB, Smith SJ, Parent W, Davies DK, Stock S. Technology use by people with intellectual and developmental disabilities to support employment activities: a single-subject design meta-analysis. J Vocational Rehabil 2006;24:81–6.

[91] Heinicke MR, Carr JE. Applied behavior analysis in acquired brain injury rehabilitation: a meta-analysis of single-case design intervention research: applied behavior analysis in ABI. Behav Interv 2014;29:77–105.

[92] Goh AE, Bambara LM. Individualized positive behavior support in school settings a meta-analysis. Remedial Spec Educ 2012;33:271–86.

[93] Maggin DM, Chafouleas SM, Goddard KM, Johnson AH. A systematic evaluation of token economies as a classroom management tool for students with challenging behavior. J School Psychol 2011;49:529–54.

[94] Romeiser-Logan L, Hickman RR, Harris SR, Heriza CB. Single-subject research design: recommendations for levels of evidence and quality rating. Dev Med Child Neurol 2008;50:99–103.