Journal of Integrative Medicine 19 (2021) 203–210

Contents lists available at ScienceDirect

Journal of Integrative Medicine

journal homepage: www.jcimjournal.com/jim
www.journals.elsevier.com/journal-of-integrative-medicine

Methodology

Is the N-of-1 method applicable in bodywork research? Lessons learned using a trial as a methodological pilot

Stergios Tsiormpatzis [1]

Open University, University of Helsinki, Helsinki 00014, Finland

ARTICLE INFO

Article history:
Received 21 July 2020
Accepted 8 January 2021
Available online 2 February 2021

Keywords:
N-of-1 trials
Bodywork
Pilot
Cross-over studies
Bias control
Methods

ABSTRACT

N-of-1 trial designs have rarely been used in bodywork research. Using a recent trial as a methodological pilot, critical issues related to the applicability of N-of-1 trials to bodywork are discussed. These include the issues of carry-over effects, bias-controlling approaches and statistical analysis. The discussion highlights the importance of mixed methods and draws some suggestions for a future research program. N-of-1 trials could be used to provide insights about some essential elements of bodywork modalities and their effectiveness.

Please cite this article as: Tsiormpatzis S. Is the N-of-1 method applicable in bodywork research? Lessons learned using a trial as a methodological pilot. *J Integr Med.* 2021; 19(3): 203–210.

1. Introduction

Recently the author presented the results of a mixed method N-of-1 trial within a whole systems research case study, aiming to assess the effect of *shiatsu* on the health-related quality of life of a person with secondary progressive multiple sclerosis (SPMS) [1] (henceforth called "our study"). In our study, one patient with no previous experience of *shiatsu* was recruited by a letter of invitation sent to former patients of the practitioner for sharing. The first responder was screened and satisfied our study's eligibility criteria [1].

Our study was a six-period counterbalanced design comparing periods of 2-week usual care with periods of 2-week twice-weekly *shiatsu* sessions (each session comprised of a *shiatsu* treatment of 60–90 minutes. During the trial, 12 sessions were provided in total) in addition to usual care. The processes of the study are summarized in Fig. 1.

The *shiatsu* sessions were offered at no cost, during visits to the house of the patient. In order to assess the effect of the treatment, the short version of the Multiple Sclerosis Quality of Life Inventory (MSQLI) [2] was administered at the end of each period. At the end of the trial, a semi-structured interview with the patient explored her treatment experience, the influence of the trial on her life, and any possible adverse events. After each *shiatsu* session, case notes were kept and used as evidence during the analysis and synthesis of the case. The results of the MSQLI questionnaires, as well as the verbatim transcript of the interview and the case notes, are available in the original publication of our study [1]. For ethical reasons, our study did not include enough data collection points that could permit statistical analysis of the questionnaires (a single administration at the end of each period). The collected data were analysed and synthesised as a descriptive case study. Possible improvements in spasticity, bowel function, fatigue, pain, sleep, and relaxation were found, without adverse events.

The rationale for writing this present methodological commentary arose while considering the applicability challenges of designing and conducting our study and the lack of relevant literature that could guide N-of-1 trials for bodywork.

N-of-1 trials are prospective (experimental or observational) studies aimed to determine the effect of an intervention on a single participant, ideally in a blinded and randomized manner. Its most common design is the repeated challenge-withdrawal "in which multiple crossovers between treatment(s) and control (placebo, standard care, or alternate treatment) are continued for a pre-specified amount of time or until treatment effectiveness is determined" [3]. Due to the role of the patient as its own control in the multiple crossovers, there is the opportunity for the treatment and its evaluation to be tailored to the needs of the specific patient [4] and a user-centred design is feasible [5]. It has been suggested that N-of-1 trials can offer benefits for patients and clinicians as well as for the healthcare system overall [6]. Yet until recently, it was understood that their potential has remained mostly unfulfilled [7,8]. Following the 2014 publication of a

[1] Tuohilammentie 388, Otalampi 03300, Finland
E-mail address: stergios.tsiormpatzis@helsinki.fi

https://doi.org/10.1016/j.joim.2021.01.010
2095-4964/

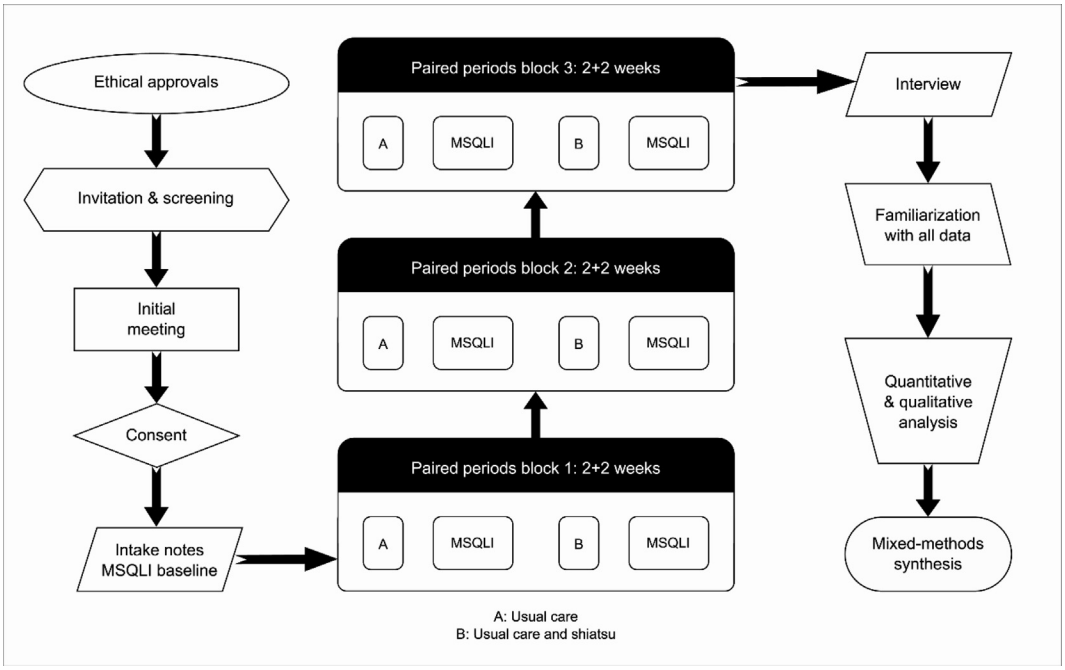


Fig. 1. Flow diagram of our study. MSQLI: Multiple Sclerosis Quality of Life Inventory.

detailed user's guide [9], the 2015 essential guide book [10], the publication of the 2015 CONSORT extension for reporting N-of-1 trials (CENT) [11,12] as well as the 2016 series "N-of-1 trials to enhance patient outcomes" [4,13], a new interest in N-of-1 trials was generated. Until that time, N-of-1 clinical trials had been mostly used to assess pharmaceutical interventions [14]. Since then, the growing interest is in new areas of investigation.

That trend continues despite the recently disappointing but vibrant results of the PREEMPT study [15–17], probably the most comprehensive N-of-1 trial ever attempted, considering the number of participants and interventions. Its aim was to examine whether patients with chronic musculoskeletal pain "would experience less pain and improved health, adherence, satisfaction and shared decision making" [16] when randomly allocated in an N-of-1 trial and supported by a mobile health application compared to those allocated to usual care. Despite the improvement observed in both groups, the results failed to show any benefit of the N-of-1 allocation in the reported outcomes. The publication of the PREEMPT study generated an interesting discussion, beginning with an invited commentary, which concluded that after PREEMPT, there is no evidence in support of the usage of N-of-1 trials for improvement of clinical outcomes, "a beautiful idea being vanquished by cruel and ugly evidence" [18]. The discussion that followed [19–24] highlighted issues with the study itself and with N-of-1 trials overall, some of which are relevant to our study. Thus it was indicated that while the study showed no difference in treatment effect for the majority of the study participants, this is not an argument about N-of-1 trials but about treatment efficacy [19,23]. Instead, the results could be interpreted as supportive of the idea that N-of-1 trials can be used not only to evaluate the benefit of a treatment to a specific patient, but also for comparison in terms of costs or harms between treatments of comparable efficacy, as well as a tool to improve the therapeutic relationship and promote scientific literacy and self-efficacy of the patients. This could possibly lead to better-informed compliance with a treatment regimen [20,24]. The researchers revealed their goal was to balance the experimental rigour of their study with the choice and convenience of the patients [24]. As well, it was observed that the documented clinically important results of the study (including the improvement in shared decision making regarding medications) were ignored in favour of looking for statistical significance [22,24]. Yet, it was argued that the length of the intervention periods were too short to show an effect for the included non-pharmacological treatments since the "onset" and "half-life" of their effects (if they existed) could only be hypothesized [23,24]. Moreover, a later publication taking account of the medical record data on analgesic prescribing at the baseline and 6-month follow-up for participants of PREEMPT found a statistically significant reduction in nonsteroidal anti-inflammatory drug usage for the patient allocated in the N-of-1 trial, as well as a clinically meaningful but not statistically significant reduction in opioid usage [25]. The authors of the later publication attribute the results to four possible factors: patient's improvement of health literacy, improved therapeutic relationship, improved self-awareness leading to a different way of experiencing pain, and changes to the treatment regime by the patients due to the usage of the mobile application [25], something indicated by the qualitative part of the PREEMPT study as well [17]. It is worth mentioning that a capsule commentary that followed highlighted previously discussed design issues that could explain the lack of statistical significance on opioids, i.e., short length of the trial compared to the many months needed for reducing opioid usage, and was underpowered among patients using opioids [26].

2. Bodywork, N-of-1 trials, and methodological pilot studies

Bodywork, defined as "therapeutic touching or manipulation of the body by using specialized techniques" [27], refers to many different therapeutic modalities [28]. Its individualized application according to the assessment of the condition by the practitioner at each session [29] and the constant development of the practice of each practitioner are two key challenging factors in researching bodywork modalities [30]. In addition, bias control (such as blinding, control, and randomization) is always difficult to achieve due to its touch-based nature [31].

Critical elements to consider when designing an N-of-1 trial are: (a) the condition should be relatively stable or chronic (even if their applicability in conditions with expected natural recovery

is under investigation [32]); (b) there should be substantial uncertainty regarding the comparative effectiveness of the examined treatments with a heterogeneity of their effects to be expected; and (c) the treatments should have rapid onset and washout [33–35].

SPMS is a chronic, relatively stable progressive condition, without expectation of improvement [36]. *Shiatsu*, like most bodywork modalities, is practised in a heterogeneous, highly personalised way but lacks robust data on its efficacy. The effect onset and washout of *shiatsu* are unknown and probably not rapid. To the author's knowledge, before our study there have been no N-of-1 trials for bodywork modalities other than physiotherapy, possibly due to the fact that bodywork cannot usually satisfy the requirement of rapid onset and short washout.

This issue should not prohibit the exploration of ways that could remedy the problem, making N-of-1 trials applicable in bodywork research. Extrapolation is used in science when dealing with a new phenomenon: trying to explain the unknown with the help of what is already known. Reaching the applicability limits of a method can show what those limits are, either by its successful application in new areas of investigation or by its failure that could guide to a new approach [37]. Our study extrapolates knowledge coming from different areas of investigation rarely used in N-of-1 trials.

In the present commentary, our study is considered as a methodological pilot that could highlight the limits of applicability for N-of-1 trials in bodywork. A pilot study is "a small-scale test of the methods and procedures to be used on a larger scale if the pilot study demonstrates that these methods and procedures can work" [38]. In the context of N-of-1 trials, pilot studies have been used to examine their feasibility in terms of actual recruitment of patients [39], or to examine the feasibility of methodologically innovative designs in settings not usually considered well-fitted for N-of-1 investigation [32]. They can also help to prepare a series of N-of-1 with multiple participants or group studies [40], or to show the feasibility of them in clinical practice [41]. Yet, during peer-debriefing of our study, it appeared that the concept of an N-of-1 trial as a methodological pilot was poorly understood. Comments such as "I am not convinced that one can/should 'pilot' an N-of-1 approach" and "if this were to be scaled up to a full-size study (the anticipated trajectory of a 'pilot') it would still have only a single participant. This is a definitional/logic issue that flaws the whole piece" has made clear to the author the need to elaborate further on the approach used in our study.

Our study highlighted the fact that the results "are not generalizable but refer to the specific participant in the specific setting" [1], a limitation often considered embedded to N-of-1 trials overall [42]. Yet it has been supported that a series of patients taking place in identical N-of-1 trials can provide a basis for "making assertions about overall treatment effectiveness within the group" [43], even though generalizing the results to other populations should be done with caution. Thus, it is suggested that our study could take the role of a pilot for a larger N-of-1 study with several participants suffering from SPMS. Moreover, by considering our study as a methodological pilot, it is not implied that the time-related and pharmacology-compatible characteristics of the bodywork modality examined in our study (*shiatsu*) are quantitatively directly transferable to other settings and populations. Yet, it is suggested that due to the current lack of relevant evidence it is reasonable to rely on the common scientific rationale that "similar" outcomes should occur in "similar" situations in the future [44]. Thus the approach taken in our study and as further specified below could be used in future studies of *shiatsu* or other bodywork modalities in order to provide quantitative estimates of the time-related and pharmacology-compatible characteristics of the modality, that are necessary in order to design a robust N-of-1 trial [23,24].

Some critical issues related to the applicability of N-of-1 trials to bodywork are discussed below in light of the experience of our study. These include (1) carry-over effects, (2) bias-controlling approaches, and (3) statistical analysis.

The discussion also highlights the importance of mixed methods [45,46] and makes some suggestions for a future research program. Its scope covers the exploratory investigation of time-related and pharmacology-compatible bodywork characteristics (investigation of the process of change that is connected with the practice of the modality [47]) as well as effectiveness research.

3. Considering onset, half-life, washout, and dosage as applicable concepts in bodywork: Dealing with the carryover effects

The main challenge for bodywork N-of-1 trials is carryover effects, which are difficult to detect and interfere with the results since there is no previous knowledge about their amplitude [48]. They are related to the quantitative concepts of onset, half-life, washout, and dosage as used in pharmacological research. By knowing the values of each substance used, pharmacological N-of-1 trials can be adjusted to accommodate the design to the specific timeframes needed for the analysis of the effect of the specific substances. N-of-1 is well fitted to inform optimal dosage, onset and washout [7], but are those time-related and pharmacology-compatible concepts applicable to bodywork?

The non-specificity of mechanisms of action [49] and other characteristics of bodywork poses a challenge for research either at a basic laboratory or clinical effectiveness level [50–52]. This is true of *shiatsu* as well, with a fuzziness of definitions and different styles of practice sharing only a few common characteristics [1,53,54]. By acknowledging the lack of previous knowledge regarding those time-related and pharmacology-compatible concepts about *shiatsu*, our study attempted to extrapolate the approach of "physiological effect models" as used in pharmacodynamics [55]. Since the effect of *shiatsu* in the organism is quantitatively unknown, it was aimed to collect data as rich as possible, in order to facilitate the usage of a "looking at the data" [55] approach. Thus, mixed methods were used in the hope of partly substituting the quantitative unknown with the qualitative data.

Even if triangulation was used in our study with all the collected data in order to reach credible results (despite the lack of statistical power), a specific example might highlight the role that mixed methods can have in terms of the carryover-related issues.

The interview that followed the clinical part of our study revealed the effect of the treatment on spasticity as the most significant for the patient, despite the fact that it was not included in the MSQLI [1]. The case notes kept after each session provide evidence that spasticity was the primary complaint during the first three sessions with *shiatsu* (period B1) but not afterwards during the last sessions of the B1 period and the following B2 period. Spasticity reappeared during the usual care period and became again the major complaint during the first three sessions of the final B3 period when *shiatsu* was reintroduced [1]. The inclusion of the interview at the end of our study, as well as the decision to include case notes, and not only rely on the quantitative questionnaire was crucial in documenting an effect not anticipated before-hand. Additionally, it was possible to infer a timeframe of its onset until the peak (three sessions in one and a half weeks) and washout (two sequential usual care periods—4 weeks in total—to reach a level that required another three sessions during one and a half weeks to repeat the effect). Those results, even if not documented using quantitative methods, could be used in conjunction with other, presumably more quantifiable effects as documented in the MSQLI, reinforcing (or not) an idea of a common

timeframe for the occurrence of different effects that could provide further justification for quantifying the time-related and pharmacology-compatible characteristics of *shiatsu*.

3.1. Onset and peak

The onset and peak of the effect of most forms of bodywork cannot be determined based on biochemical characteristics as happens in pharmacology, at least not based on the current knowledge about physiological effects of bodywork at the micro-level and the ability to infer from them effects at the macro-level [50–52]. In our study, attempts were made to determine the onset and peak by observing the changes in the effect that occurred after each session. It was found that the onset occurs during one treatment period, but eight sessions over four weeks (the trial's maximum treatment period) were necessary for the effect to reach the peak for some of the examined domains (Fig. 2). In pharmacology, that would be considered a very long onset and peak time, prohibiting N-of-1 methodologies. In the bodywork context, N-of-1 designs with even longer, possibly adaptive [17,56] treatment periods might be exploratory, aiming to establish the onset and peak of different bodywork modalities for different conditions.

3.2. Half-life and washout

Similar challenges exist for the determination of the half-life for effects of bodywork. Despite the limited data collection points in our study, it was found that the maximum effect on bowel control that was achieved in eight sessions over four weeks, declined steadily for every two weeks without treatment (Fig. 2). It is suggested that by using exploratory N-of-1 trials for specific modalities, it could be possible to reach preliminary estimations for an equivalent of half-life, i.e., at what rate an observed effect is fading away. Such trials should permit robust statistical analysis, ideally including varying/adaptive periods of treatment and control. By using half-life estimates determined by pilot studies, a satisfying washout and follow-up period could be inferred and used for larger studies.

Recently, it has been proposed that "multiple baseline design" (MBD) could be used when long-lasting effects are expected [57], as is often the case for bodywork. Yet MBDs originated from behavioural research [58], and are very unlikely to be feasible in bodywork. In bodywork, MBDs will require the ability to locally restrict effects and change their location over time [59]. Bodywork modalities aiming to bring a systemic effect in the functioning of the organism as a whole are weak candidates for MBD's designs. On the other hand, as it has been shown for physiotherapy [60–62], strictly localised manipulation aiming to have a particular effect does not need an MBD.

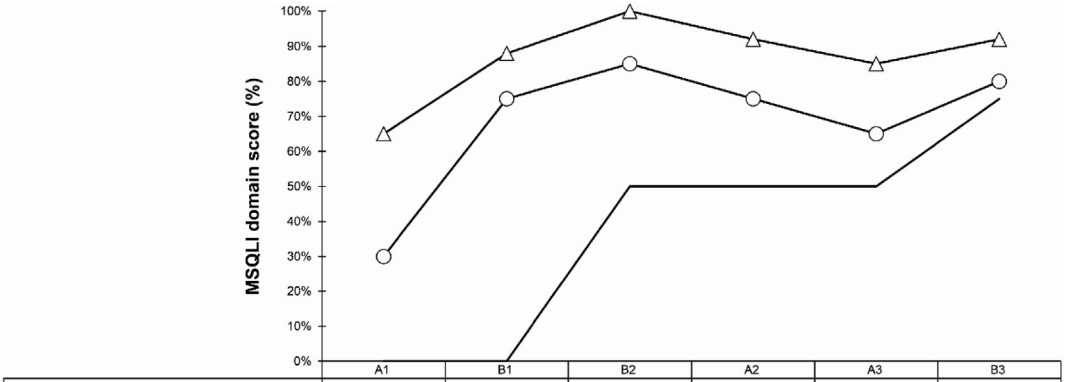
3.3. Dosage

The concept of dosage is most commonly understood in the manner used in pharmacotherapy. There is inadequate knowledge regarding which modality might be more beneficial for a specific condition and its required dosage [63]. Dosage in the context of bodywork might mean how long each session is (amount), how often sessions are given (frequency), and how long a treatment schedule lasts (length). Given this knowledge gap, N-of-1 trials for bodywork run the risk of under-treatment. In our study, it was explained that the dosage regime of the trial was not following the common practice of *shiatsu*. The amount of treatment was adjusted according to the assessment of the condition by the practitioner at each session, mostly exceeding the common practice of 45–60 min per session. The frequency was based on the clinical experience of the practitioner, and was twice as often as in standard practice which suggests once-weekly sessions. The length of our study was limited due to ethical and logistical reasons, a limitation that from a methodological perspective was partly corrected by the counterbalanced crossover that provides two consecutive *shiatsu* periods. The results of our study are supportive of the idea that the concept of dosage is applicable in bodywork. Pilot N-of-1 trials with adaptive periods and varying dosage characteristics could provide more specific details about an estimate of the actual dosage requirements which in turn could be used to design more robust bodywork N-of-1 or group trials.

4. Controlling and placebo

The only control used in our study was a period of standard care. Yet, as an anonymous peer-reviewer of our study indicated, this is not a good enough way to control for "doing nothing." This is an issue that might apply to all kinds of treatment *vs* standard care trials. On reflecting from a methodological angle, the author believes that while in a single N-of-1 bodywork trial it might be challenging to account for the therapist effect (the "doing nothing"), controlling is not impossible.

Controlling in bodywork is a problematic issue [31]. There are three ways to control with different aims: no-treatment, alternative treatment, and placebo treatment periods [64]. The used



	A1	B1	B2	A2	A3	B3
Role-Physical Scale (RP)	0%	0%	50%	50%	50%	75%
Modified Fatigue Impact Scale - 5 Item Version	30%	75%	85%	75%	65%	80%
Bowel Control Scale (BWCS)	65%	88%	100%	92%	85%	92%

Fig. 2. Selected MSQLI scores. MSQLI: Multiple Sclerosis Quality of Life Inventory.

standard care could be equivalent to using a no-treatment period. While it might be relevant to clinical practice, it is not possible to differentiate specific bodywork from placebo effects [65].

When using the placebo effect as a methodological instrument to separate the specific and non-specific effects of the treatment, it is assumed to be "treatments theorized not to be effective for a condition or symptoms by virtue of their intrinsic properties" which should "mimic the appearance of the verum treatment except for its particular, hypothesized, remedial factor" [66]. Physical placebo interventions (including surgeries) have been found to be associated with strong effects, and the more invasive the intervention the greater the effect [67]. In bodywork research, there have been attempts to use sham or placebo control [68]. Yet their efficiency remains questionable [69] as long as they are not objectively verified by monitoring physiological effects [70,71]. Potential sham methods include active components and have been found to be not credible [64], while placebo effects are an essential active mechanism of bodywork [72,73]. A heterogeneous mixture of approaches has been used [74] including hands-on techniques that might belong to other bodywork modalities and thus produce a bodywork specific effect [75], and physical modalities requiring mechanical equipment. The most successful attempts so far are related to simulation of specific, localised, brief manipulations [76], an approach not generalizable to modalities that aim to have a systemic effect.

A more "ontological" (the term is used as per Blease & Annoni [77] with distinction between "ontological-motivated" and "methodologically-motivated" definitions of placebo concepts) consideration of the factors influencing placebo effects might provide a useful alternative when attempting to control bodywork. Contextual factors that mediate the placebo effects are related to "physical, psychological and social elements that characterize the therapeutic encounter" [78], including (but not limited to) the expectations of the patients, elements of the therapeutic relationship such as verbal and non-verbal communication, behaviour and "branding" of the practitioner, and therapeutic performance [77,79]. These are essential elements in the clinical setting of bodywork [78,80] and can be explored by using mixed methods. While quantification of their role in different modalities is not available, it is accepted that through them, the placebo effects are part of the clinical bodywork practice [72,73].

An attempt to include control periods that are planned to elicit those factors without a physical placebo might provide some partial solution in an effort to differentiate the specific and non-specific effects of the tested modality. For example, in our study, instead of using controlled usual care periods, control periods could have included the usual care with the addition of visits to the patient without any physical intervention. Those visits would be as frequent as the *shiatsu* sessions with the aim of an equal presence of the therapeutic relationship and other contextual elements of placebo during treatment and control periods. One aim of a pilot N-of-1 trial in bodywork that attempts such a kind of control could be to evaluate the success of the placebo control achieved, by including semi-structured interviews with the patient and the practitioner and structured diaries [81–83] in a mixed methods manner.

5. Statistical analysis

Due to concerns regarding the possible burden to a vulnerable patient, our study included only one data collection point per period after the baseline (six in total)—a single administration of MSQLI at the end of each period. That approach has been used before to address carryover effects in a secondary analysis [84], yet it was a critical weakness that made impossible the statistical analysis of our findings. By looking at the results from a methodological point of view, a linear serial correlation between observations is expected during the onset until reaching the peak as well as during the washout. In our study with the counterbalancing design, a linear serial correlation could run not only inside a period but between periods inside blocks and between blocks as well, making statistical analysis without previous knowledge about the onset and washout extremely challenging.

For the sake of simplicity, the simulated scores that are used below in the example of statistical analysis, are considered as coming from a bodywork N-of-1 trial with balanced crossover and five blocks (ABABABABAB), including ten data collection points per period. The simulated values (Table 1) are in a scale 0–100 with 0 = worst, 100 = best, with A indicating control and B bodywork periods. For the statistical calculations of the example R (v.4.0.0) with R Studio (v.1.2.5042) was used, using the sample code provided in the publication by Tang and Landes [85].

While in the literature there are complicated statistical approaches available for the analysis of N-of-1 trials, the recent work by Tang and Landes [85] provides detailed guidance on a statistical approach precisely prepared for single N-of-1 trials and is simple enough to be used by investigators with only basic statistical skills. Their suggestion is to use "serial *t*-tests for rate-change" when a linear dependency is assumed or "serial *t*-tests for level-change" when linear independence is assumed. That approach has weaknesses too, since it does not consider carryover effects. In the following example, we combine those serial *t*-tests with a modification of a method proposed in another paper by Wang and Schork [86] (supposing long periods with multiple data collection points that will minimise the false-positive chances [85]): instead of contrasting the mean score of the initial treatment period with the mean of the aggregated treatment periods that follow, the rate-change of the beginning part (of which the length in a mixed methods study could be estimated by using the case notes of the practitioners and by "looking at the data" of the questionnaires) of each treatment period can be compared with the rate-change of the rest of each treatment period, to estimate the onset. Similarly, with the control periods the comparison of the rate-change could provide details about the washout of the active treatment.

Step 1: By plotting the scores and triangulating with case notes, we can identify a trend at the first two to three scores of the B-periods, possibly indicating the onset till peak time. Similarly, for the A-periods we can identify a trend at the first three scores that follow B-periods, possibly indicating the washout.

Step 2: We can test the observation of step 1 by using the 2-sample serial *t*-test for rate-change in each period with the results indicating small or relatively small *P*-values for most of the periods (Table 2). For the sake of the example, we consider that the tests confirm the observation, meaning that for the examined bodywork modality, the first two or three scores will be considered as the onset period as well as the first three scores of the control periods being considered as the washout period.

Step 3: By excluding the scores belonging to the onset and washout periods (step 2), the corrected means of all periods can be assumed to be linear independent and can be tested by using the serial *t*-test for level-change. The results of the tests are in Table 3 (considered with two and three scores as the onset period).

It should be noted that the simulated data and example hypothesis or decisions are intentionally simple in order to describe the process of analysis, while in real life, more complex relationships and decisions are expected. An example from our study might highlight the usefulness of the mixed methods in generating useful clinical questions even when the quantitative data remain inconsistent (Fig. 3).

Table 1
Simulated scores.

Block	Period	Data collection point									
		1	2	3	4	5	6	7	8	9	10
Block 1	A1	30	30	30	35	35	20	30	30	30	30
	B1	30	35	45	60	75	80	85	80	80	80
Block 2	A2	70	60	55	40	45	40	35	40	30	40
	B2	50	65	70	85	85	80	75	80	85	85
Block 3	A3	70	55	35	40	40	30	35	30	30	30
	B3	55	65	70	80	80	75	85	90	85	85
Block 4	A4	60	55	45	40	35	40	45	35	30	35
	B4	60	70	80	75	85	90	95	90	95	95
Block 5	A5	60	40	40	30	30	35	30	40	35	35
	B5	60	65	75	80	80	75	85	90	80	80

A: control period; B: bodywork period. Each data collection point refers to a score completed at different points in time during each period.

Table 2
Two-sample serial *t*-tests for rate-change.

Compared scores	Standard deviation	Standard error	*t*-statistic	2-sided *P*-value
A2[1:3]/A2[3:10]	4.2375	2.1174	−3.0361	0.0084
A3[1:3]/A3[3:10]	2.9039	1.5980	−9.8338	< 0.0001
A4[1:3]/A3[3:10]	4.3757	3.1957	−2.0117	0.0906
A5[1:3]/A3[3:10]	4.4766	2.1812	−5.0759	0.0001
B1[1:3]/B1[3:10]	6.0831	4.8752	1.0256	0.3562
B2[1:3]/B1[3:10]	4.2725	3.8466	2.5997	0.0692
B3[1:3]/B1[3:10]	3.8704	2.7686	2.1930	0.0679
B4[1:3]/B1[3:10]	3.9340	2.9547	2.4174	0.0553
B5[1:3]/B1[3:10]	4.6877	3.3951	2.0513	0.0847

X[a:b]/X[c:d]: scores a to b of period X vs. scores c to d of period X.

Table 3
Paired serial *t*-test for level change.

Compared periods	Standard deviation	Standard error	*t*-statistic	2-sided *P*-value
A[4:10]/B[3:10]	3.6204	3.1816	−14.5141	0.0401
A[4:10]/B[4:10]	3.1298	1.9808	−24.2322	0.0017

A[a:b]/B[c:d]: scores a to b of A periods *vs.* scores c to d of B periods. Two different score sets [c:d] are used, to show the potential of the described process for the statistical identification of the washout period.

As discussed in our study [1], the two pain-related scores that are included in the MSQLI follow opposing trends, with the Bodily Pain Scale doing worst during the B1–B2 periods that included *shiatsu*. However, the patient considers the relief from her pains as an essential element of the *shiatsu* treatments she received. While the case notes indicate the appearance of temporary local pains during the sessions, these were never mentioned in the post-treatment adverse effect enquiries. The suggestion of our study, considering all forms of data, was that the pain documented by the Bodily Pain Scale did not correlate with the effects of the pain assessed by Medical Outcomes Study Pain Effects Scale but corresponds to theory and experientially consistent "transitional effect" that is known to occur in *shiatsu* practice [87]. The contradiction between the scores was an issue in that without using mixed methods, it would probably not be possible to explain in a clinically relevant way even if the MSQLI was completed enough times to permit statistical analysis.

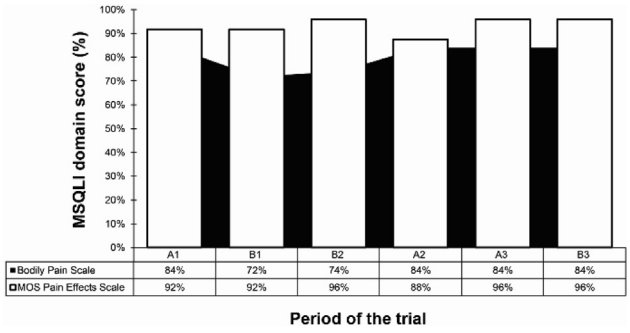


Fig. 3. Contradicting pain-related MSQLI scores. MSQLI: Multiple Sclerosis Quality of Life Inventory; MOS: medical outcomes study.

6. Conclusion and suggestions for future studies

N-of-1 trials with mixed methods as an exploratory design could be used as pilots, which aim to provide details about the time-related and pharmacology-compatible characteristics of the tested modalities (onset and peak, washout, dosage) [88,89]. Those details could feed simulations [90] so that more extensive (N-of-1 or group) studies could be optimally designed and the appropriate analytical approach chosen, making them more appropriate and specific to each tested modality. In a feedback-loop manner, better accuracy to the estimators of the time-related characteristics could be achieved, using methods specifically developed for the estimation of the effect size [91] and Bayesian analysis [92].

Considering the high probability that bodywork N-of-1 trials will continue for the near future to be very small and limited, possibly having to face similar ethical concerns as our study, it should be highlighted that with the methods discussed in the present commentary, a minimum of five periods per examined treatment

should be included in order for a statistical analysis at the single level to be possible [85]. A higher number of observations during each period could promote more accuracy in the onset and wash-out estimations. Aggregation of multiple N-of-1 trials, aiming for inferences of overall treatment effectiveness [93], should follow the most appropriate approaches indicated in the literature, probably using Bayesian linear mixed models [94].

When moving from the single N-of-1 trial to larger trials with multiple participants and practitioners, power analysis could be considered following recent methodological suggestions [5,86,95]. In such trials instead of the standard care control periods, it might be worth considering partial controlling by using a kind of open-label placebo [96], in the form of the simple presence of the therapist without touch interaction, for the same time as the sessions of the active treatment periods. This might be more feasible for bodywork modalities with a short intervention time. Elements of randomization and blinding (for example on which practitioner treats whom, on treatment or control allocation per session, and during analysis) as well as routine physiological measurements beside the chosen patient-reported questionnaires, could be utilized.

Besides providing further insights about critical elements of bodywork modalities and their effectiveness, those studies could permit a comparative analysis between different bodywork modalities and styles, providing a way to explore the "Dodo Bird Verdict" [97], translated as the absolute *vs* relative efficacy controversy [98] for bodywork, which, as long as the specificity of time-related and pharmacology-compatible bodywork characteristics is not improving, remains an open issue [99,100].

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author would like to thank Sarah Budd for offering her valuable time to discuss English and Greek language issues, Daniel Tsiormpatzis for his last minute help with the graphics and to the anonymous reviewers for their constructive comments.

References

[1] Tsiormpatzis S. Effects of shiatsu on the health-related quality of life of a person with secondary progressive multiple sclerosis: a mixed methods N-of-1 trial within a whole systems research case study. Eur J Integr Med 2019;32:101006.
[2] The Consortium of Multiple Sclerosis Centers Health Services Research Subcommittee. Multiple Sclerosis Quality of Life Inventory: a user's manual. (1997) [2020-10-27]. http://www.nationalmssociety.org/NationalMSSociety/media/MSNationalFiles/Brochures/MSQLI_-A-User-s-Manual.pdf.
[3] 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.
[4] Vohra S. N-of-1 trials to enhance patient outcomes: identifying effective therapies and reducing harms, one patient at a time. J Clin Epidemiol 2016;76:6–8.
[5] Kwasnicka D, Inauen J, Nieuwenboom W, Nurmi J, Schneider A, Short CE, et al. Challenges and solutions for N-of-1 design studies in health psychology. Health Psychol Rev 2019;13(2):163–78.
[6] Nikles J. What are N-of-1 trials? In: Nikles J, Mitchell G. The essential guide to N-of-1 trials in health. Dordrecht: Springer Netherlands; 2015:9–17.
[7] Guyatt G. N-of-1 randomized trials: a commentary. J Clin Epidemiol 2016;76:4–5.
[8] Kravitz RL, Paterniti DA, Hay MC, Subramanian S, Dean DE, Weisner T, et al. Marketing therapeutic precision: potential facilitators and barriers to adoption of N-of-1 trials. Contemp Clin Trials 2009;30(5):436–45.
[9] Kravitz RL, Duan N. Design and implementation of N-of-1 trials: a user's guide. Rockville: AHRQ Publication; 2014.
[10] Nikles J, Mitchell G. The essential guide to N-of-1 trials in health. Dordrecht: Springer, Netherlands; 2015.
[11] 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. J Clin Epidemiol 2016;76:9–17.
[12] 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. J Clin Epidemiol 2016;76:18–46.
[13] Knottnerus JA, Tugwell P, Tricco AC. Individual patients are the primary source and the target of clinical research. J Clin Epidemiol 2016;76:1–3.
[14] Li J, Gao W, Punja S, Ma B, Vohra S, Duan N, et al. Reporting quality of N-of-1 trials published between 1985 and 2013: a systematic review. J Clin Epidemiol 2016;76:57–64.
[15] Barr C, Marois M, Sim I, Schmid CH, Wilsey B, Ward D, et al. The PREEMPT study—evaluating smartphone-assisted n-of-1 trials in patients with chronic pain: study protocol for a randomized controlled trial. Trials 2015;16:67.
[16] Kravitz RL, Schmid CH, Marois M, Wilsey B, Ward D, Hays RD, et al. Effect of mobile device-supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain. JAMA Intern Med 2018;178(10):1368–77.
[17] Whitney RL, Ward DH, Marois MT, Schmid CH, Sim I, Kravitz RL. Patient perceptions of their own data in mHealth Technology-enabled N-of-1 trials for chronic pain: qualitative study. JMIR Mhealth Uhealth 2018;6(10):e10291.
[18] Mirza RD, Guyatt GH. A randomized clinical trial of n-of-1 trials—tribulations of a trial. JAMA Intern Med 2018;178(10):1378–9.
[19] Vohra S, Punja S. A case for n-of-1 trials. JAMA Intern Med 2019;179(3):452.
[20] Smith AK. A case for n-of-1 trials. JAMA Intern Med 2019;179(3):452–3.
[21] Kravitz RL, Sim I, Duan N. A case for n-of-1 trials—reply. JAMA Intern Med 2019;179(3):453.
[22] Chapple AG, Blackston JW. Finding benefit in n-of-1 trials. JAMA Intern Med 2019;179(3):453–4.
[23] McDonald S, McGree J, Bazzano L. Finding benefit in n-of-1 trials. JAMA Intern Med 2019;179(3):454–5.
[24] Kravitz RL, Schmid CH, Sim I. Finding Benefit in n-of-1 trials—reply. JAMA Intern Med 2019;179(3):455.
[25] Odineal DD, Marois MT, Ward D, Schmid CH, Cabrera R, Sim I, et al. Effect of mobile device-assisted N-of-1 trial participation on analgesic prescribing for chronic pain: randomized controlled trial. J Gen Intern Med 2020;35(1):102–11.
[26] Keller MS. Capsule commentary on Odineal et al., effect of mobile device-assisted N-of-1 trial participation on analgesic prescribing for chronic pain: randomized controlled trial. J Gen Intern Med 2020;35(1):404.
[27] Merriam-Webster. Definition of bodywork. [2020-10-27]. https://www.merriam-webster.com/dictionary/bodywork.
[28] Porcino AJ, Boon HS, Page SA, Verhoef MJ. Meaning and challenges in the practice of multiple therapeutic massage modalities: a combined methods study. BMC Complement Altern Med 2011;11:75.
[29] Milanese S. The use of RCT's in manual therapy—are we trying to fit a round peg into a square hole?. Man Ther 2011;16(4):403–5.
[30] Porcino AJ, Boon HS, Page SA, Verhoef MJ. Exploring the nature of therapeutic massage bodywork practice. Int J Ther Massage Bodywork 2013;6(1):15–24.
[31] Mehling WE, DiBlasi Z, Hecht F. Bias control in trials of bodywork: a review of methodological issues. J Altern Complement Med 2005;11(2):333–42.
[32] Nikles J, Tate RL, Mitchell G, Perdices M, McGree JM, Freeman C, et al. Personalised treatments for acute whiplash injuries: a pilot study of nested N-of-1 trials in a multiple baseline single-case experimental design. Contemp Clin Trials Commun 2019;16:100480.
[33] Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, et al. A clinician's guide for conducting randomized trials in individual patients. CMAJ 1988;139(6):497–503.
[34] Kravitz RL, Duan N, Vohra S, Li J; The DEcIDE Methods Center N-of-1 Guidance Panel. Introduction to N-of-1 trials: indications and barriers. In: Kravitz R, Duan N; The DEcIDE Methods Center N-of-1 Guidance Panel. Design and implementation of N-of-1 trials: a user's guide. Rockville: Agency for Healthcare Research and Quality; 2014:1–12.
[35] Davidson KW, Cheung YK, McGinn T, Wang YC. Expanding the role of N-of-1 trials in the precision medicine era: action priorities and practical considerations. (2018) [2020-10-27]. https://doi.org/10.31478/201812d.
[36] Lublin FD. New multiple sclerosis phenotypic classification. Eur Neurol 2014;72(Suppl 1):1–5.
[37] Vazulin VA. To the question of the "mechanisms" of the development of theoretical knowledge. (1964) [2020-10-27]. https://ilhs-org.net/ru/stat2.pdf [Russian].
[38] Porta M. Pilot investigation, study. In: A dictionary of epidemiology. Oxford: Oxford University Press; 2014. p. 215.
[39] Clough AJ, Hilmer SN, Naismith SL, Gnjidic D. The feasibility of using N-of-1 trials to investigate deprescribing in older adults with dementia: a pilot study. Healthcare (Basel) 2019;7(4):161.

[40] Hardy JR, Carmont SAS, O'Shea A, Vora R, Schluter P, Nikles CJ, et al. Pilot study to determine the optimal dose of methylphenidate for an N-of-1 trial for fatigue in patients with cancer. J Palliat Med 2010;13(10):1193–7.
[41] Kronish IM, Cheung YK, Shimbo D, Julian J, Gallagher B, Parsons F, et al. Increasing the precision of hypertension treatment through personalized trials: a pilot study. J Gen Intern Med 2019;34(6):839–45.
[42] Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The N-of-1 clinical trial: the ultimate strategy for individualizing medicine?. Per Med 2011;8 (2):161–73.
[43] Nikles J, Mitchell GK, Schluter P, Good P, Hardy J, Rowett D, et al. Aggregating single patient (N-of-1) trials in populations where recruitment and retention was difficult: the case of palliative care. J Clin Epidemiol 2011;64(5): 471–80.
[44] Edgington ES. Randomized single-subject experimental designs. Behav Res Ther 1996;34(7):567–74.
[45] Van Ness PH, Murphy TE, Ali A. Attention to individuals: mixed methods for N-of-1 health care interventions. J Mix Methods Res 2017;11(3):342–54.
[46] Aalbers S, Spreen M, Bosveld-van Haandel L, Bogaerts S. Evaluation of client progress in music therapy: an illustration of an N-of-1 design in individual short-term improvisational music therapy with clients with depression. Nord J Music Ther 2017;26(3):256–71.
[47] Janosky JE. Use of the single subject design for practice based primary care research. Postgrad Med J 2005;81(959):549–51.
[48] Hills M, Armitage P. The two-period cross-over clinical trial. Br J Clin Pharmacol 1979;8(1):7–20.
[49] Bialosky JE, George SZ, Bishop MD. How spinal manipulative therapy works: why ask why?. J Orthop Sports Phys Ther 2008;38(6):293–5.
[50] Standley PR. My personal journey that led to the crossroads of interdisciplinary manual medicine research: serendipitous opportunities afforded a basic scientist. J Bodyw Mov Ther 2013;17(1):79–82.
[51] Chaitow L. Can we describe what we do?. J Bodyw Mov Ther 2014;18 (3):315–6.
[52] Standley PR. Towards a Rosetta Stone of manual therapeutic methodology. J Bodyw Mov Ther 2014;18(4):586–7.
[53] Tsiormpatzis S. Safety and risks of shiatsu: protocol for a systematic review. Eur J Integr Med 2019;28:20–6.
[54] Cabo F, Aguaristi I. Similarities and differences in East Asian massage and bodywork therapies: a critical review. OBM Integr Complement Med 2020;5 (1):17.
[55] Holford NH, Sheiner LB. Understanding the dose-effect relationship: clinical application of pharmacokinetic-pharmacodynamic models. Clin Pharmacokinet 1981;6:429–53.
[56] Duan N, Kravitz RL, Schmid CH. Single-patient (N-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. J Clin Epidemiol 2013;66(8):S21–8.
[57] Bradbury J, Avila C, Grace S. Practice-based research in complementary medicine: could N-of-1 trials become the new gold standard?. Healthcare 2020;8(1):15.
[58] Tate RL, Perdices M. N-of-1 trials in the behavioral sciences. In: Nikles J, Mitchell G. The essential guide to N-of-1 trials in health. Dordrecht: Springer Netherlands; 2015: 19–41.
[59] Farrell JC. Transfer ability in a patient with spinal cord compression: a multiple baseline design. Physiother Theory Pract 1991;7(1):39–43.
[60] Gursli S, Sandvik L, Bakkeheim E, Skrede B, Stuge B. Evaluation of a novel technique in airway clearance therapy—specific cough technique (SCT) in cystic fibrosis: a pilot study of a series of N-of-1 randomised controlled trials. SAGE Open Med 2017;5:205031211769750.
[61] Beattie AJM. The effectiveness of spinal mobilisation in the treatment of low back pain: a single case study. Physiother Theory Pract 1991;7(1):57–62.
[62] Riddoch J, Lennon S. Evaluation of practice: the single case study approach. Physiother Theory Pract 1991;7(1):3–11.
[63] Riddoch J, Lennon S. Single subject experimental design: one way forward?. Physiotherapy 1994;80(4):215–8.
[64] Hancock MJ, Maher CG, Latimer J, McAuley JH. Selecting an appropriate placebo for a trial of spinal manipulative therapy. Aust J Physiother 2006;52 (2):135–8.
[65] Vickers AJ, De Craen AJM. Why use placebos in clinical trials? A narrative review of the methodological literature. J Clin Epidemiol 2000;53 (2):157–61.
[66] Blease CR. Psychotherapy and placebos: manifesto for conceptual clarity. Front Psychiatry 2018;9:379.
[67] Brim RL, Miller FG. The potential benefit of the placebo effect in sham-controlled trials: implications for risk-benefit assessments and informed consent. J Med Ethics 2013;39:703–7.
[68] Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. Cochrane Database Syst Rev 2010;1:CD003974.
[69] Lougee H, Johnston RG, Thomson OP. The suitability of sham treatments for use as placebo controls in trials of spinal manipulative therapy: a pilot study. J Bodyw Mov Ther 2013;17(1):59–68.
[70] Henley CE, Wilson TE. Use of beat-to-beat cardiovascular variability data to determine the validity of sham therapy as the placebo control in osteopathic manipulative medicine research. J Am Osteopath Assoc 2014;114(11):860–6.
[71] Manzotti A, Cerritelli F, Lombardi E, La Rocca S, Chiera M, Galli M, et al. Effects of osteopathic treatment versus static touch on heart rate and oxygen saturation in premature babies: a randomized controlled trial. Complement Ther Clin Pract 2020;39:101116.
[72] Bialosky JE, Bishop MD, Penza CW. Placebo mechanisms of manual therapy: A sheep in Wolf's clothing?. J Orthop Sports Phys Ther 2017;47(5):301–4.
[73] Karas S, Mintken P, Brismée JM. We need to debate the value of manipulative therapy and recognize that we do not always understand from what to attribute our success. J Man Manip Ther 2018;26(1):1–2.
[74] Cerritelli F, Verzella M, Cicchitti L, D'Alessandro G, Vanacore N. The paradox of sham therapy and placebo effect in osteopathy: a systematic review. Medicine (Baltimore) 2016;95(35):e4728.
[75] Rosner A. Fables or foibles: inherent problems with RCTs. J Manipulative Physiol Ther 2003;26(7):460–7.
[76] Chaibi A, Benth JŠ, Russell MB. Validation of placebo in a manual therapy randomized controlled trial. Sci Rep 2015;5:11774.
[77] Rossettini G, Carlino E, Testa M. Clinical relevance of contextual factors as triggers of placebo and nocebo effects in musculoskeletal pain. BMC Musculoskelet Disord 2018;19(1):27.
[78] Blease C, Annoni M. Overcoming disagreement: a roadmap for placebo studies. Biol Philos 2019;34:18.
[79] Rossettini G, Testa M. Manual therapy RCTs: should we control placebo in placebo control?. Eur J Phys Rehabil Med 2018;54(3):500–1.
[80] Latey P. Placebo: a study of persuasion and rapport. J Bodyw Mov Ther 2000;4 (2):123–36.
[81] Stone DA, Kerr CE, Jacobson E, Conboy ScD LA, Kaptchuk TJ. Patient expectations in placebo-controlled randomized clinical trials. J Eval Clin Pract 2005;11(1):77–84.
[82] White P, Bishop FL, Prescott P, Scott C, Little P, Lewith G. Practice, practitioner, or placebo? A multifactorial, mixed-methods randomized controlled trial of acupuncture. Pain 2012;153(2):455–62.
[83] Snowden M. Use of diaries in research. Nursing Standard 2015;29(44):36–41.
[84] Zucker DR, Ruthazer R, Schmid CH, Feuer JM, Fischer PA, Kieval RI, et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. J Rheumatol 2006;33(10):2069–77.
[85] Tang J, Landes RD. Some *t*-tests for N-of-1 trials with serial correlation. PLoS ONE 2020;15(2):e0228077.
[86] Wang Y, Schork NJ. Power and design issues in crossover-based N-of-1 clinical trials with fixed data collection periods. Healthcare 2019;7(3):84.
[87] Long AF, Esmonde L, Connolly S. A typology of negative responses: a case study of shiatsu. Complement Ther Med 2009;17(3):168–75.
[88] Huang H, Yang P, Xue J, Tang J, Ding L, Ma Y, et al. Evaluating the individualized treatment of traditional Chinese medicine: a pilot study of N-of-1 trials. Evid Based Complement Alternat Med 2014;2014:148730.
[89] Huang H, Yang P, Wang J, Wu Y, Zi S, Tang J, et al. Investigation into the individualized treatment of traditional Chinese medicine through a series of N-of-1 trials. Evid Based Complement Alternat Med 2018;2018:5813767.
[90] Percha B, Baskerville EB, Johnson M, Dudley JT, Zimmerman N. Designing robust N-of-1 studies for precision medicine: simulation study and design recommendations. J Med Internet Res 2019;21(4):e12641.
[91] Hedges LV, Pustejovsky JE, Shadish WR. A standardized mean difference effect size for single case designs. Res Synth Methods 2012;3(3):224–39.
[92] Barnard-Brak L, Richman DM, Watkins L. Treatment burst data points and single case design studies: a Bayesian N-of-1 analysis for estimating treatment effect size. Perspect Behav Sci 2020;43(2):285–301.
[93] Punja S, Xu D, Schmid CH, Hartling L, Urichuk L, Nikles CJ, et al. N-of-1 trials can be aggregated to generate group mean treatment effects: a systematic review and meta-analysis. J Clin Epidemiol 2016;76:65–75.
[94] Krone T, Boessen R, Bijlsma S, van Stokkum R, Clabbers NDS, Pasman WJ. The possibilities of the use of N-of-1 and do-it-yourself trials in nutritional research. PLoS ONE 2020;15(5):e0232680.
[95] Senn S. Sample size considerations for N-of-1 trials. Stat Methods Med Res 2019;28(2):372–83.
[96] Kaptchuk TJ, Miller FG. Open label placebo: can honestly prescribed placebos evoke meaningful therapeutic benefits?. BMJ 2018;363:k3889.
[97] Budd R, Hughes I. The Dodo Bird Verdict—controversial, inevitable and important: a commentary on 30 years of meta-analyses. Clin Psychol Psychother 2009;16(6):510–22.
[98] Staines GL. The relative efficacy of psychotherapy: reassessing the methods-based paradigm. Rev Gen Psychol 2008;12(4):330–43.
[99] Cleland JA, Childs JD. Does the manual therapy technique matter? (2005-09) [2020-10-27]. https://web.archive.org/web/20170829143457/http://www.sigurdmikkelsen.no/admin/dokumenter/Cleland%20&%20Childs%20-%202008%20-%20Does%20the%20manual%20therapy%20technique%20matter.pdf.
[100] Huijbregts P. Clinical prediction rules: Time to sacrifice the holy cow of specificity? J Man Manip Ther 2007;15(1):5–7.