



The interpretation of clinical relevance in randomised clinical trials in patients with chronic low back pain: protocol for a meta-research study



Tiziano Innocenti^{a,b,*}, Tim Schleimer^c, Stefano Salvioli^{b,d}, Silvia Giagio^{e,f},
Raymond Ostelo^{a,g}, Alessandro Chiarotto^{a,h}

^a Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, the Netherlands

^b GIMBE Foundation, Italy

^c Department of Applied Health Sciences, Division of Physiotherapy, Hochschule für Gesundheit, Bochum, North Rhine-Westphalia, Germany

^d Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Genoa, Italy

^e Division of Occupational Medicine, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy

^f Department of Biomedical and Neuromotor Sciences (DIBINEM), Alma Mater Studiorum, University of Bologna, Bologna, Italy

^g Department of Epidemiology and Data Science, Amsterdam UMC, Location Vrije Universiteit, Amsterdam Movement Sciences research institute, the Netherlands

^h Department of General Practice, Erasmus MC, University Medical Center, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:

Meta-epidemiology
Clinical relevance
Low back pain

ABSTRACT

Chronic low back pain (cLBP) is one of the leading worldwide causes of disability. The smallest worthwhile effect (SWE) parameter has been proposed to find a threshold of clinical relevance. Specific values of the SWE have been calculated in patients with cLBP for pain intensity, physical functioning and time to recovery for physiotherapy compared with no intervention. Our objectives are 1) To evaluate how authors have interpreted the clinical relevance of the effect of physiotherapy compared to no-intervention on pain, physical functioning and time to recovery; 2) To reinterpret the clinical relevance of these between-group differences based on the available SWE estimates; 3) To evaluate, for descriptive purposes, whether the studies are adequately powered or underpowered considering the published SWE values and a power threshold of 80%. A systematic search in Medline, PEDro, Embase and Cochrane CENTRAL will be conducted. We will search for RCT investigating the effectiveness of physiotherapy as compared to no interventions in people with cLBP. We will compare the authors' interpretation of results for clinical relevance with their results to determine if they meet their a-priori definitions. Then, we will perform a re-interpretation of the between-group differences based on SWE values published for cLBP.

* Corresponding author.

E-mail address: t.innocenti@vu.nl (T. Innocenti).

Social media:  (T. Innocenti)

<https://doi.org/10.1016/j.mex.2023.102239>

Received 11 March 2023; Accepted 31 May 2023

Available online 1 June 2023

2215-0161/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications table

Subject area:	Medicine and Dentistry
More specific subject area:	Meta-research
Name of your protocol:	Not applicable
Reagents/tools:	Not applicable
Experimental design:	Meta-research study
Trial registration:	Not applicable
Ethics:	This study will not involve humans
Value of the Protocol:	<ul style="list-style-type: none"> • Several stakeholders (e.g. researchers, policymakers) should make a judgment about the statistical significance and clinical relevance of the results • To find new reliable estimates for clinical relevance is crucial for clinical decision making

Description of protocol

Introduction

Chronic low back pain (cLBP) is defined as pain, muscle tension, or stiffness lasting longer than 12 weeks [1] and it is one of the worldwide leading causes of disability, which accounts for large costs for healthcare systems and work productivity [2,3]. Around 65% with LBP still report pain after one year [4], and it affects physical, emotional and social functioning [3].

Many treatment options are available for patients with cLBP [5]. Randomised controlled trials (RCT) provide the best study design yielding accurate estimates of the effectiveness of an intervention [6]. To determine the effectiveness of an intervention, various stakeholders (e.g. researchers, policy makers) should make a judgment about the statistical significance and clinical relevance of the results [7].

Clinical relevance indicates whether the intervention should be used in clinical practice to improve healthcare outcomes, and it involves the interpretability of between-group differences [8]. Between-group differences represent the effect between two or more treatments to determine if the effect of the treatment is clinically relevant [8]. Many parameters have been proposed to find a threshold of for interpreting the clinical relevance of between-group differences [9,10]. Several terms have been used to identify a threshold for clinical relevance over the years [10,11]. It was first described as “Minimal Clinical Important Difference – MCID” in 1989 by Jaeschke et al. [9]. In 2005 the “C” was removed because it focused on the clinic rather than the patients’ experience [12]. Since its first definition, many other studies [13–16] have described similar entities with slightly different names, but it is not always clear what construct these parameters are intended to measure [11]. Now, the minimal important difference (MID) is defined as “the smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important, either beneficial or harmful, and which would lead the patient or clinician to consider a change in the management” [12].

Good estimates of between-group clinical relevance interpretation should satisfy two conditions, as argued by Barrett et al. [10]. The first is that they should be intervention-specific and must involve weighing the benefits of the intervention against its costs, risks, and side effects, as compared to a control intervention. The second is that the judgments about whether the benefits of intervention should be based on the perspective of patients who receive the intervention. In addition, these estimates should be expressed in terms of a between-group difference effect rather than a within-group change on a specific outcome measure, especially if they want to be used to inform the design and interpretation of clinical trials [8,17].

The MID does not fully satisfy these aforementioned requirements: costs, risks, and other harms are not generally taken into account within the existing MID framework [10], it is not intervention-specific, and it should be used to interpret a within-group change on an outcome measure rather than a between-group difference. To overcome these issues, in 2005, Barrett et al. [10,18] described a methodological approach, the “benefit harm trade-off method,” to estimate the smallest worthwhile effect (SWE) of health interventions. The SWE approach captures the judgments of recipients of care; allows patients to weigh the benefits of treatment against the risks, costs, and side effects; and potentially provides estimates based on an intervention-control comparison. It is expressed as a percentage of the smallest amount of patient-valued benefit that an intervention would require to justify associated costs, risks, and other inconveniences [10].

Two studies [19,20] calculated SWE values for pain intensity, physical functioning and time to recovery for physiotherapy interventions (exercise and manual therapy) compared with no intervention for patients with cLBP. These studies consistently found that patients need to experience at least a 20% additional improvement in pain and physical functioning, and to speed up their recovery by 10 days to consider that the effect of physiotherapy is worthwhile given its costs, potential side effects, and inconveniences, as compared to receiving no intervention. In other terms, 20% value corresponds to the minimum difference that people expect to experience with physiotherapy compared to no intervention departing from their current level (i.e. baseline score in a clinical trial). In a clinical trial, this corresponds to the between-group difference at the follow-up considered. Exploring how the authors of published RCTs on physiotherapy for cLBP have interpreted their results and reinterpreting the clinical relevance in light of the SWE values for between-group differences could be relevant to future research (e.g., to drive the sample size calculation in RCTs) and clinical practice (e.g., to make clinicians and patients able to better judge the clinical relevance of physiotherapy interventions in cLBP).

Therefore, the objective of our meta-research study are the following:

Primary objectives:

- To evaluate if authors have interpreted the clinical relevance of the effect of physiotherapy compared to no-intervention on pain, physical functioning and time to recovery following their a-priori definition;

- To re-interpret the clinical relevance of the between-group differences of the published RCTs based on the available SWE estimates for physiotherapy compared to no intervention in patients with cLBP [19,20].

Secondary objectives:

- To evaluate, for descriptive purposes, whether the studies are adequately powered or underpowered considering the published SWE values and a power threshold of 80%.

Material and methods

We will follow and adapt the PRISMA 2020 checklist [21] for the reporting of this manuscript (or a specific reporting checklist for meta-research studies, if available at the time of reporting [22]).

Data source and search strategies

For retrieving eligible articles, a literature search on the following electronic databases will be conducted from inception to November 30th, 2022: Medline (through the PubMed interface), PEDro, Embase (Embase.com), and Cochrane CENTRAL (www.cochranelibrary.com). In addition to the electronic database search, other potentially relevant studies will be searched in the reference lists of the included articles and recent systematic reviews on the topic [1], in all relevant gray literature sources such as conference proceedings, OpenGray, and Google Scholar. The search will be re-run before the final analyses to retrieve any additional records.

Medical subjects' headings (MeSH) and all relevant free text words will be combined using the boolean operators (e.g. AND, OR): (1) condition (chronic low back pain); (2) intervention (exercise and manual therapy); (3) study design (Randomized Controlled Trial, Controlled Clinical Trial). The Cochrane MEDLINE sensitivity-maximizing RCT filter will be used to ensure the maximum sensitivity to our search strategy since it has very high sensitivity and a slightly better precision relative to more sensitive filters [23].

No year restrictions will be applied. We will include studies written in English, French, Italian, German, Dutch, Spanish and Portuguese. If we find eligible articles in other languages (e.g. Chinese or Japanese), we will try to translate them using online web services or look for native-speaker readers to perform their assessment. Then, if we cannot reach a suitable translation or find a native speaker, we will exclude them from our sample.

A detailed search strategy will be created for each electronic database (see Appendix 1).

Eligibility criteria

To be consistent with the criteria used in the validation studies for the SWE [19,20], the following inclusion criteria were applied to the retrieved literature:

- (1) conducted in adult patients (>18 years old) with non-specific cLBP (defined as LBP lasting for more than 12 weeks)
- (2) The intervention consisted of a combination of the following physiotherapy treatment components: any exercise therapy modality (e.g. but not limited to resistance exercise, strengthening exercise, aerobic exercise) AND/OR any form of manual therapy (e.g. but not limited to muscle energy techniques, massage therapy, spinal manipulative therapy). Other treatments (e.g. education, psychologically informed physiotherapy, electrotherapy or acupuncture) can be included if they represent only a minor part of the intervention (i.e. $\leq 25\%$ of the treatment time/sessions).
- (3) The control group is 'no intervention' (e.g., waiting list or total absence of interventions). Regardless of the name given to the comparison by the original authors, we will include it if no active treatment is given to the patients. If, for example, authors describe usual care as "participants were instructed to follow their normal schedule of medications and physical activity", we will include it. If only instructions were given without "an active role of the health care provider," we will include it.
- (4) at least one of the outcomes is pain intensity, physical functioning or time to recovery
- (5) RCT design and cross-over design (only the data before the crossover will be included), (6) the follow-up time point is short-term, defined as ≤ 12 weeks (since the study by Ferreira [20] and Christiansen [19] focused on 2 and 6 weeks, respectively).

The following exclusion criteria will be used:

- (1) patients undergoing back surgeries in the last year
- (2) patients with severe psychiatric co-morbidities that impede participation
- (3) other treatments (e.g., pain medication, injections) combined with physiotherapy. RCTs including a "mixed" population of patients with non-specific LBP and other musculoskeletal disorders (e.g., specific LBP, neck pain, acute LBP) will be included only if at least 75% of the patients displayed chronic LBP.

Data selection

First, the titles and abstracts of the retrieved studies will be divided into two sets. Each set will be screened independently by two pairs of reviewers (TI, TS and SS, SG) to evaluate whether they meet the eligibility criteria. Second, the full texts of the potentially eligible studies will be downloaded and assessed for eligibility by the same two pairs. If the full text cannot be retrieved, we will

contact the authors with a maximum of three emails. A consensus meeting will be held to determine agreement on the selection; in case of disagreement, a third reviewer (AC) will decide on inclusion. The reasons for the full-text exclusions will be recorded. The study selection process will be summarised through the PRISMA flowchart [21]. EndNote [24] will be used to remove duplicates and manage the bibliography. Rayyan QCRI systematic review software [25] will be used to carry out the selection process.

Data extraction

The data will be extracted by two independent reviewers (TI and TS) and double-checked by a third reviewer (SS) to ensure the precision and quality of the retrieved information. The following data will be extracted from the included studies through a standardised extraction Excel sheet: design, study setting (author, year of publication, country, and language), in- and exclusion criteria, sample size, patient characteristics (e.g. age, sex, pain, physical functioning and time to recovery at baseline), type of intervention (including intervention characteristics such as frequency, and duration), outcomes measures used (for physical functioning, pain and time to recovery), effect sizes, cut-off used for statistical significance and clinical relevance, the description of the framework used (e.g. MID, SWE), follow-up time point.

Regarding statistical significance, the authors' a-priori definition (e.g. two side p-value < 0.05) will be searched in the text, and the alpha value will be used for the sample size calculation. For clinical relevance, it will be checked how the authors a-priori defined the MID or SWE for the between-group differences; these parameters will be searched in the article text and the sample size calculation (either in the study protocol when available or the full-text), looking at the difference they used to calculate it.

The following predefined decision rules will be used to select data from trials to prevent selective inclusion: (a) where trialists reported both unadjusted and adjusted values for the same outcome; adjusted values will be extracted; (b) where trialists reported data analysed; based on the intention-to-treat (ITT) sample and another sample (e.g. per-protocol, as-treated), ITT-analyzed data will be extracted.

Missing data or additional details will be gathered by contacting the corresponding author of included studies with a maximum of three emails. For Cross-over trials, only the data before the crossover will be extracted and included in the analyses.

Data analysis

Descriptive data of the included studies will be reported in tables as mean (standard deviations) or median (interquartile range) values. The following descriptive analyses will be performed:

Primary analyses

- We will compare the authors' interpretation of results for statistical significance and clinical relevance with their results to determine if they meet their a-priori definitions. If the authors do not specify any clinical relevance or statistical significance threshold, we will use the difference hypothesised for sample size calculation. If the authors do not report any threshold or the sample size calculation, we will report the lack of information.
- We will determine if the author's a priori-definition of clinical relevance meets the SWE threshold.
- We will re-interpret the clinical relevance, comparing the between-group differences for every individual trial based and the published SWE values [19,20]. For pain intensity and physical functioning, the results will be considered clinically relevant in case of a 20% additional improvement in the intervention group (physiotherapy) compared to the control (no intervention), starting from the baseline values of the respective outcomes. For time to recovery, in case of 10 days or more. To retrieve the percentage of between-group differences, we will calculate it by comparing the between-group difference with the mean baseline value of the study for each outcome.

Secondary analysis

- We will calculate the power for each included study considering the alpha error used in the study, sample size and a between-group difference of 20% (the SWE values already published) on the primary outcome. In this way, each study will be classified as the following:
 - Underpowered if the power achieved is less than 80%.
 - Well-powered if the difference is 80% or above.

Pilot trials evaluating the feasibility of a future large RCT will be excluded from this analysis.

SPSS software (IBM SPSS Statistics for Macintosh, Version 28.0. Armonk, NY: IBM Corp) will be used for the analyses. G*Power software will be used for a posteriori power calculation [26].

Ethics and dissemination

This study does not require an ethics review as we will not collect personal data; it will summarize information from publicly available studies.

A manuscript will be prepared and submitted for publication in an appropriate peer-reviewed journal upon study completion. The study findings will be disseminated at a relevant (inter)national conference. All results of this meta-research study will also be announced at (inter)national scientific events in musculoskeletal rehabilitation and research methods. The results of this investigation will be relevant to researchers paying more attention to the synthesis of the evidence to translate clinical implications to key stakeholders (healthcare providers and patients).

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.

CRediT authorship contribution statement

Tiziano Innocenti: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Tim Schleimer:** Data curation, Investigation, Methodology, Writing – review & editing. **Stefano Salvioli:** Data curation, Investigation, Methodology, Writing – review & editing. **Silvia Giagio:** Data curation, Investigation, Methodology, Writing – review & editing. **Raymond Ostelo:** Data curation, Investigation, Methodology, Supervision, Writing – review & editing. **Alessandro Chiarotto:** Data curation, Investigation, Methodology, Supervision, Writing – review & editing.

Data availability

No data was used for the research described in the article.

Acknowledgments

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

References

- [1] J.A. Hayden, J. Ellis, R. O'gilvie, A. Malmivaara, M.W. van Tulder, Exercise therapy for chronic low back pain, *Cochrane Database Syst. Rev.* (9) (2021), doi:[10.1002/14651858.CD009790.pub2](https://doi.org/10.1002/14651858.CD009790.pub2).
- [2] J. Hartvigsen, M.J. Hancock, A. Kongsted, et al., What low back pain is and why we need to pay attention, *Lancet* 391 (10137) (2018) 2356–2367 Jun 9, doi:[10.1016/S0140-6736\(18\)30480-X](https://doi.org/10.1016/S0140-6736(18)30480-X).
- [3] A. Cieza, K. Causey, K. Kamenov, S.W. Hanson, S. Chatterji, T. Vos, Global estimates of the need for rehabilitation based on the global burden of disease study 2019: a systematic analysis for the global burden of disease study 2019, *Lancet* 396 (10267) (2021) 2006–2017 Dec 19, doi:[10.1016/S0140-6736\(20\)32340-0](https://doi.org/10.1016/S0140-6736(20)32340-0).
- [4] C.J. Itz, J.W. Geurts, M. van Kleef, P. Nelemans, Clinical course of non-specific low back pain: a systematic review of prospective cohort studies set in primary care, *Eur. J. Pain.* 17 (1) (2013) 5–15 Jan, doi:[10.1002/j.1532-2149.2012.00170.x](https://doi.org/10.1002/j.1532-2149.2012.00170.x).
- [5] M. Artus, D.A. van der Windt, K.P. Jordan, E.M. Hay, Low back pain symptoms show a similar pattern of improvement following a wide range of primary care treatments: a systematic review of randomized clinical trials, *Rheumatology* 49 (12) (2010) 2346–2356 Dec, doi:[10.1093/rheumatology/keq245](https://doi.org/10.1093/rheumatology/keq245).
- [6] J.M. Kendall, Designing a research project: randomised controlled trials and their principles, *Emerg. Med. J.* 20 (2) (2003) 164–168, doi:[10.1136/emj.20.2.164](https://doi.org/10.1136/emj.20.2.164).
- [7] W.-C. Leung, Balancing statistical and clinical significance in evaluating treatment effects, *Postgrad. Med. J.* 77 (905) (2001) 201–204, doi:[10.1136/pmj.77.905.201](https://doi.org/10.1136/pmj.77.905.201).
- [8] R.H. Dworkin, D.C. Turk, M.P. McDermott, et al., Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations, *Pain.* 146 (3) (2009) 238–244 Dec, doi:[10.1016/j.pain.2009.08.019](https://doi.org/10.1016/j.pain.2009.08.019).
- [9] R. Jaeschke, J. Singer, G.H. Guyatt, Measurement of health status. Ascertaining the minimal clinically important difference, *Control Clin. Trials* 10 (4) (1989) 407–415 Dec, doi:[10.1016/0197-2456\(89\)90005-6](https://doi.org/10.1016/0197-2456(89)90005-6).
- [10] B. Barrett, D. Brown, M. Mundt, R. Brown, Sufficiently important difference: expanding the framework of clinical significance, *Med. Decis. Making* 25 (3) (2005) 250–261 May-Jun, doi:[10.1177/0272989X05276863](https://doi.org/10.1177/0272989X05276863).
- [11] M.L. Ferreira, R.D. Herbert, P.H. Ferreira, et al., A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain, *J. Clin. Epidemiol.* 65 (3) (2012) 253–261 Mar, doi:[10.1016/j.jclinepi.2011.06.018](https://doi.org/10.1016/j.jclinepi.2011.06.018).
- [12] H.J. Schunemann, G.H. Guyatt, Commentary—goodbye M(C)ID! Hello MID, where do you come from? *Health Serv. Res.* 40 (2) (2005) 593–597 Apr, doi:[10.1111/j.1475-6773.2005.00374.x](https://doi.org/10.1111/j.1475-6773.2005.00374.x).
- [13] A. Beurskens, H.C.W. de Vet, A.J.A. Koke, Responsiveness of functional status in low back pain: a comparison of different instruments, *Pain* 65 (1) (1996) 71–76 Apr, doi:[10.1016/0304-3959\(95\)00149-2](https://doi.org/10.1016/0304-3959(95)00149-2).
- [14] D. Cella, M. Bullinger, C. Scott, I. Barofsky, Clinical Significance Consensus Meeting G. Group vs individual approaches to understanding the clinical significance of differences or changes in quality of life, *Mayo Clin. Proc.* 77 (4) (2002) 384–392 Apr, doi:[10.4065/77.4.384](https://doi.org/10.4065/77.4.384).
- [15] M.J. Yelland, P.J. Schluter, Defining worthwhile and desired responses to treatment of chronic low back pain, *Pain Med.* (Malden, Mass) 7 (1) (2006) 38–45 Jan-Feb, doi:[10.1111/j.1526-4637.2006.00087.x](https://doi.org/10.1111/j.1526-4637.2006.00087.x).
- [16] C. Demoulin, R. Ostelo, J.A. Knottnerus, R.J. Smeets, Quebec back pain disability scale was responsive and showed reasonable interpretability after a multidisciplinary treatment, *J. Clin. Epidemiol.* 63 (11) (2010) 1249–1255 Nov, doi:[10.1016/j.jclinepi.2009.08.029](https://doi.org/10.1016/j.jclinepi.2009.08.029).
- [17] M.L. Ferreira, R.D. Herbert, What does 'clinically important' really mean? *Aust. J. Physiother.* 54 (4) (2008) 229–230, doi:[10.1016/s0004-9514\(08\)70001-1](https://doi.org/10.1016/s0004-9514(08)70001-1).
- [18] B. Barrett, R. Brown, M. Mundt, et al., Using benefit harm tradeoffs to estimate sufficiently important difference: the case of the common cold, *Med. Decis. Making* 25 (1) (2005) 47–55 Jan-Feb, doi:[10.1177/0272989X04273147](https://doi.org/10.1177/0272989X04273147).
- [19] D.H. Christiansen, N.B. de Vos Andersen, P.H. Poulsen, R.W. Ostelo, The smallest worthwhile effect of primary care physiotherapy did not differ across musculoskeletal pain sites, *J. Clin. Epidemiol.* 101 (2018) 44–52 Sep, doi:[10.1016/j.jclinepi.2018.05.019](https://doi.org/10.1016/j.jclinepi.2018.05.019).
- [20] M.L. Ferreira, R.D. Herbert, P.H. Ferreira, et al., The smallest worthwhile effect of nonsteroidal anti-inflammatory drugs and physiotherapy for chronic low back pain: a benefit-harm trade-off study, *J. Clin. Epidemiol.* 66 (12) (2013) 1397–1404 Dec, doi:[10.1016/j.jclinepi.2013.02.018](https://doi.org/10.1016/j.jclinepi.2013.02.018).
- [21] M.J. Page, J.E. McKenzie, P.M. Bossuyt, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, *Bmj* 372 (2021) n71, doi:[10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71).
- [22] D.O. Lawson, L. Puljak, D. Pieper, et al., Reporting of methodological studies in health research: a protocol for the development of the Methodological Study reporting Checklist (MISTIC), *BMJ Open* 10 (12) (2020) e040478, doi:[10.1136/bmjopen-2020-040478](https://doi.org/10.1136/bmjopen-2020-040478).
- [23] J. Glanville, E. Kotas, R. Featherstone, G. Dooley, Which are the most sensitive search filters to identify randomized controlled trials in MEDLINE? *J. Med. Libr. Assoc.* 108 (4) (2020) 556–563 Oct 1, doi:[10.5195/jmla.2020.912](https://doi.org/10.5195/jmla.2020.912).
- [24] [EndNoteVersion EndNote 20, Clarivate, 2013](https://www.endnote.com/).
- [25] M. Ouzzani, H. Hammady, Z. Fedorowicz, A. Elmagarmid, Rayyan-a web and mobile app for systematic reviews, *Syst. Rev.* 5 (1) (2016) 210 Dec 5, doi:[10.1186/s13643-016-0384-4](https://doi.org/10.1186/s13643-016-0384-4).
- [26] F. Faul, E. Erdfelder, A.G. Lang, A. Buchner, G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences, *Behav. Res. Methods* 39 (2) (2007) 175–191 May, doi:[10.3758/bf03193146](https://doi.org/10.3758/bf03193146).