

1 **Title:** Challenges and solutions for N-of-1 design studies in health psychology.

2

3 Dominika Kwasnicka, Jennifer Inauen, Wim Nieuwenboom, Johanna Nurmi, Annegret

4 Schneider, Camille E Short, Tessa Dekkers, A. Jess Williams, Walter Bierbauer, Ari

5 Haukkala, Federica Picariello, Felix Naughton

6

7 Author Note

8 Dominika Kwasnicka, School of Psychology, Curtin University, Kent Street, Perth, 6102,

9 Western Australia, +61 8 9266 2559, dominika.kwasnicka@curtin.edu.au; Jennifer Inauen,

10 Department of Environmental Social Sciences, Environmental and Health Psychology, Eawag

11 – Swiss Federal Institute of Aquatic Science & Technology, Ueberlandstrasse 133, 8600

12 Duebendorf, Switzerland, +41 58 765 5780, jennifer.inauen@eawag.ch; Wim Nieuwenboom,

13 University of Applied Sciences Northwestern Switzerland, School of Social Work, Institute for

14 Social Work and Health, Riggengbachstrasse 16, CH-4600 Olten, +41 62 957 21 09,

15 wim.nieuwenboom@fhnw.ch; Johanna Nurmi, (1) Faculty of Social Sciences, University of

16 Helsinki, Finland; (2) Behavioural Science Group, Institute of Public Health, Forvie Site,

17 Cambridge CB2 0SR, University of Cambridge, United Kingdom, +358 504484297,

18 johanna.nurmi@helsinki.fi; Annegret Schneider, Department of Clinical, Educational and

19 Health Psychology, University College London, Gower St, London, WC1E 6BT, United

20 Kingdom, +44 (0) 20 7679 25438, a.schneider@ucl.ac.uk; Camille E Short, The Freemasons

21 Foundation Centre for Men's Health, School of Medicine, University of Adelaide, Adelaide,

22 South Australia, Australia, +61 8 83130532, camille.short@adelaide.edu.au; Tessa Dekkers,

23 Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15,

24 2628 CE Delft, the Netherlands, +31 (0) 15 27 84 640, t.dekkers@tudelft.nl; A. Jess Williams,

25 Institute for Mental Health, School of Psychology, University of Birmingham, Edgbaston,

26 Birmingham B15 2TT, UK., AXW847@student.bham.ac.uk; Walter Bierbauer, Department of

1 Psychology, Applied Social and Health Psychology, University of Zurich, Binzmuehlestrasse
2 14, 8050 Zurich, Switzerland, +41 44 635 72 58; walter.bierbauer@psychologie.uzh.ch; Ari
3 Haukkala, Faculty of Social Sciences, University of Helsinki, Finland, +358 (0) 504480951,
4 ari.haukkala@helsinki.fi; Federica Picariello, Health Psychology Section, Psychology
5 Department, Institute of Psychiatry, Psychology, and Neuroscience, King's College London,
6 5th floor Bermondsey Wing, Guy's Campus, London, SE1 9RT, United Kingdom, +44 (0) 20
7 7188 9324, federica.picariello@kcl.ac.uk; Felix Naughton, School of Health Sciences,
8 University of East Anglia, Norwich NR4 7UL, United Kingdom, + 44 (0) 1603 593459,
9 f.naughton@uea.ac.uk.

10 This conceptual review was written as a result of an N-of-1 design workshop prior to 31st
11 European Health Psychology Society (EHPS) Conference in Padua, Italy, 2017. We would like
12 to acknowledge the EHPS for the opportunity to meet and discuss N-of-1 design. We would
13 like to thank all meeting participants who contributed to the discussions that instigated this
14 paper. The authors would also like to acknowledge the helpful statistical advice from Dr
15 George Savva and Dr Kathryn Richardson.

16 Correspondence concerning this article should be addressed to Dominika Kwasnicka, Faculty
17 of Health Sciences, School of Psychology, Curtin University, Bentley Campus, Perth,
18 Western Australia, 6845. E-mail: dominika.kwasnicka@curtin.edu.au.

19

20

21

1 **Abstract**

2 Theories of behaviour change and health behaviour change interventions are most often
3 evaluated in between-person designs. However, behaviour change theories apply to
4 individuals not groups and behavioural interventions ultimately aim to achieve within-person
5 rather than between-group change. Within-person methodology, such as N-of-1 (also known
6 as single case design), can circumvent this issue, though has multiple design-specific
7 challenges. This paper provides a conceptual review of the challenges and potential solutions
8 for undertaking N-of-1 studies in health psychology. Key challenges identified include
9 participant adherence to within-person protocols, carry-over and slow onset effects,
10 suitability of behaviour change techniques for evaluation in N-of-1 experimental studies,
11 optimal allocation sequencing and blinding, calculating power/sample size, and choosing the
12 most suitable analysis approach. Key solutions include involving users in study design,
13 employing recent technologies for unobtrusive data collection and problem solving by
14 design. Within-person designs share common methodological requirements with
15 conventional between-person designs but require specific methodological considerations. N-
16 of-1 evaluation designs are appropriate for many though not all types of interventions. A
17 greater understanding of patterns of behaviours and factors influencing behaviour change at
18 the within-person level is required to progress health psychology into a precision science.

19

20 See Supplementary Material 1 for video abstract.

21

22 *Keywords:* N-of-1, single case study, within-person design, idiographic design

23

24

1 **Introduction**

2 N-of-1 studies test predictions within a single object of investigation based on repeated
3 measurement of variables over time. N most often refers to an individual but it can also be
4 defined as one cluster or unit, e.g., a family, hospital, organisation. For N-of-1 designs the
5 power of the study is determined by the number of repeated observations. Therefore, it is
6 possible to satisfy study objectives with just one individual or unit of investigation (Duan,
7 Kravitz, & Schmid, 2013).

8 The N-of-1 design is a recommended method for testing behavioural theory within
9 individuals through repeated measures (Craig et al., 2008) and is the optimal design for
10 examining within-person variability in cognitions and behavioural outcomes (Johnston &
11 Johnston, 2013). N-of-1 studies have been successfully used in various settings to study a
12 range of health behaviours, including treatment adherence, physical activity, drug/alcohol
13 use, sleep, smoking and eating behaviour (McDonald et al., 2017). For example, Smith et al.
14 (2017) found in six N-of-1 investigations that no individuals had the same pattern of
15 associations between Social Cognitive Theory constructs and their physical activity, and the
16 theory derived constructs demonstrated bi-directional relationships with activity. Within-
17 person assessments of cognitions underlying health-related behaviours are increasingly
18 common in behavioural science and due to recent technology development (Versluis et al.,
19 2016) changes in predictor variables and associated outcome variables can be tested in real
20 time or near real time and in the context of daily life. For an extensive summary of examples
21 of how N-of-1 designs have been applied to health psychology research questions, see
22 McDonald et al.'s (2017) review.

23 N-of-1 approaches allow health psychologists to develop and conduct personalised
24 behaviour change studies. Different conditions or health interventions can be delivered in a
25 fixed or random order for each participant to investigate which intervention is the most

1 efficacious for each participant. The ‘traditional’ between-group randomised control trial
2 (RCT) only ever provides an estimate of effect at the group level, neglecting intra-individual
3 differences. Importantly, effects observed at the intra-individual level can differ from those
4 found at the between-participant level (Inauen, Shrout, Bolger, Stadler, & Scholz, 2016).
5 Using an N-of-1 design overcomes the issue of effects heterogeneity in nomothetic (between-
6 individual) designs. Furthermore, N-of-1 can more easily enable the modelling of temporal
7 changes. In terms of making individual treatment decisions, N-of-1 studies, namely N-of-1
8 RCTs, are regarded at the top of the evidence hierarchy (Guyatt et al., 2000) over and above
9 systematic reviews of RCTs.

10 Unlike health psychology, other disciplines have a relatively long tradition of using
11 within-person designs, providing indications of the type of questions this approach could help
12 to answer. In education research, repeated measures on the level of schools, classrooms or
13 individual learners contributed to the development of learning theories and individualised
14 learning support and have been recommended to document evidence-based best practice in
15 teaching, in particular in the field of special education (Horner et al., 2005; Kennedy, 2005;
16 Moeller, Dattilo, & Rusch, 2015). Other areas such as experimental economics, investigating
17 individuals’ responses to diverse choices or their willingness to pay or accept payments in
18 different scenarios, often used within-individual designs as such findings reflect more closely
19 real-world scenarios where people encounter a series of changing conditions over time
20 (Charness, Gneezy, & Kuhn, 2012; Hogarth, 2005).

21 Similarly, in medical research within-individual designs resemble clinical practice
22 where physicians treat their patients based on within-individual considerations (Davidson,
23 Peacock, Kronish, & Edmondson, 2014; Janosky, 2005). N-of-1 studies have also helped
24 advance other psychology sub-disciplines such as clinical, neuro or educational psychology
25 and broad behaviour change research, investigating how a person changes with changing

1 circumstances and exposure to different interventions (Barlow & Hersen, 1984; Hogarth,
2 2005; Sidman, 1960; Tate, Perdices, McDonald, Togher, & Rosenkoetter, 2014). Those
3 diverse applications of within-individual research designs illustrate a methodological
4 diversity, requiring further discipline-specific definition of key constructs and methods.

5 To help expand the use of N-of-1 methods in health psychology, the field would
6 benefit from a common terminology. Bolger and Laurenceau (2013) offer *intensive*
7 *longitudinal methods* as an umbrella term for methods involving sequences of repeated
8 measurements sufficiently frequent to allow characterising a separate change process for each
9 unit of assessment. N-of-1 studies, which are also known as *single-participant, within-person*
10 and *single-case study design*, fall under this umbrella term and include observational and
11 experimental multiple cross-over studies comparing two or more treatments within
12 individuals (Duan et al., 2013). N-of-1 is arguably the most commonly used term in health
13 psychology to describe this type of study (McDonald et al., 2017) and so we use it here,
14 understanding it as interchangeable with the aforementioned terms. While a single N-of-1
15 study seeks to understand idiographic within person changes, researchers might also be
16 interested in aggregating those processes at the between-person level to reach generalisable
17 conclusions. In this case the term *aggregated N-of-1 study* can be used, or also *cumulative N-*
18 *of-1s*, referring to the same principle.

19 In health psychology, there is currently no established tradition of N-of-1 studies
20 (Davidson et al., 2014; McDonald et al., 2017), meaning the design has been underused and
21 is often misunderstood in the field. While this idiographic design offers many advantages
22 over more traditional nomothetic approaches, it comes with its own challenges, some of
23 which are particularly pertinent to health psychology investigations. The purpose of this
24 paper is to review the key challenges for undertaking health psychology related N-of-1
25 research and provide potential solutions for resolving or minimising these and, in doing so,

1 encourage greater confidence in using this design. The key challenges were identified during
2 an N-of-1 design workshop prior to the 31st European Health Psychology Society (EHPS)
3 Conference in Padua, Italy, 2017.

4 **General challenges and solutions across N-of-1 designs**

5 There are two main types of N-of-1 studies, observational and experimental. The
6 principles of N-of-1 observational studies are usually more basic than principles of
7 experimental N-of-1 studies. The N-of-1 observational study involves repeated measures of
8 behavioural predictors and outcomes over time within an individual with no manipulation on
9 observed variables. The aim of the observational N-of-1 is to describe the relationship
10 between predictor and outcome over time often examining a temporal pattern, and time lags
11 between predictor and outcome, e.g., an individual reports higher energy levels immediately
12 after coffee than 2 hours later. N-of-1 observational designs can also address questions
13 regarding temporal association between variables (direction of association), e.g., does stress
14 result in you exercising less?; or does exercising result in you being less stressed?; or is it
15 both? Burg et al. (2017) demonstrated that the relationship between stress and exercise can be
16 uni- or bi-directional in either direction and varies from person to person. Observational N-
17 of-1s allow the testing of psychological theory within individuals, often demonstrating that
18 psychological variables predicting behaviours within individuals differ compared to between
19 individuals (Kwasnicka, Dombrowski, White, & Sniehotta, 2017). In addition, such studies
20 often find that the pattern and magnitude of outcome variance accounted for by predictor
21 variables differ between participants. Given the repetitive nature of data collection in N-of-1
22 studies, a high risk of missing data is a pertinent challenge of this study design.

23 **Challenge 1: Non-adherence to data collection and missing data**

24 Ecological Momentary Assessment (EMA; Stone & Shiffman, 1994), a form of
25 Ambulatory Assessment (AA; Trull & Ebner-Priemer, 2014), is one of the most common

1 methods of collecting data in N-of-1 studies. The repetition of EMA provides valuable
2 insights into an individual's behaviours and subjective states. However, it also poses a burden
3 on participants and may result in low adherence. In this context, adherence refers to the
4 extent to which a person's behaviour corresponds with the agreed terms of usage or agreed
5 recommendations (Sieverink, Kelders, & van Gemert-Pijnen, 2017). Interestingly, despite
6 increased burden on users, overall, adherence to EMA protocols reported in reviews is
7 relatively high, e.g., in older adults (Cain, Depp, & Jeste, 2009) and in youth (Wen,
8 Schneider, Stone, & Spruijt-Metz, 2017). In older adults only 4 out of 27 studies reported
9 adherence rates under 80% and in youth average adherence was 78% (N = 36). Cain et al.
10 (2009)'s review also found that among clinical populations adherence was greater in studies
11 with higher daily sampling frequencies (6+ times versus 2-3 or 4-5 times). The inverse was
12 true among nonclinical populations, with studies with a low sampling rate (2-3 times per day)
13 demonstrating the highest adherence. Adherence to event-contingent reporting (e.g.,
14 completing an EMA every time a participant smokes) was found in one study to be similar to
15 adherence to signal-based (random prompts) reporting (Schüz, Walters, Frandsen, Bower, &
16 Ferguson, 2013). Even with relatively high adherence rates, missing data still poses a
17 challenge for researchers undertaking N-of-1 studies. Two broad approaches to address
18 missing data in N-of-1 studies are imputation to manage missing data or, better still, user-
19 centred study design to reduce or avoid it in the first place.

20 Regarding the first potential solution – imputation – it is important to determine if
21 non-adherence is problematic by examining the pattern of missing assessments. Similar to
22 other methodologies, missing data in N-of-1 studies can be distinguished in three patterns,
23 missing completely at random (MCAR), missing at random (MAR) and missing not at
24 random (MNAR) (Rubin, 1976). Generally, we assume sporadic missing data to be MAR.
25 This means that the probability of a missing response is independent of both observed and

1 unobserved variables. MAR data can be imputed without having an impact on the causal
2 inference. Therefore (multiple) data imputation has been advocated as a suitable approach to
3 missing data in general, although this is only sometimes practiced in the context of N-of-1
4 studies (McDonald et al., 2017).

5 While multiple data imputation is common practice in cross-sectional and non-
6 intensive longitudinal designs, there are reasons to question this approach in N-of-1 studies.
7 Namely, since data in N-of-1 studies is repeatedly collected from individuals and potentially
8 auto-correlated, the assumption of independence between variables is likely violated. In fact,
9 longitudinal within-person studies often produce missing data that is a mixture of MAR and
10 MNAR (Feng, Cong, & Silverstein, 2012; Graham, 2009). For example, data may be missing
11 due to longitudinal attrition (e.g., lower response rates near the end of longitudinal data
12 collection) and increased respondent burden (Deeg, van Tilburg, Smit, & de Leeuw, 2002;
13 Twisk & de Vente, 2002); in these cases, multiple data imputation is unlikely to be a suitable
14 solution. The approaches that are designed specifically to be used to deal with missing data in
15 N-of-1 studies include using Amelia II software (www.gking.harvard.edu/amelia; Honaker &
16 King, 2010) that imputes missing data in a single cross-section from a time series or from a
17 time-series-cross-sectional data set implementing a bootstrapping-based algorithm. For a full
18 comparison of missing data methods and software to fit incomplete data regression models
19 see Horton and Kleinman (2007).

20 To mitigate non-adherence in N-of-1 studies User-Centred Design (UCD), also
21 known as participatory design or co-design, can be used. UCD describes design processes in
22 which users influence how a design takes shape (Abrams, Maloney-Krichmar, & Preece, 2004).
23 This includes inviting users to participate in feasibility and usability studies early and often
24 (Stappers & Giaccardi, 2017) and can also include users nominating their own predictor
25 variables of measurement based on their experiences. This design is often advocated as a

1 method to increase acceptance and uptake of the final study/intervention, which results from
2 the engaging users involved in a project design (e.g., Kent & Bush, 2018). A further feasible
3 option is using objectively measured data from unobtrusive measurement using technologies
4 such as smartphones and wearables. While measuring an individual's physical activity using
5 accelerometers is well established and practiced, new approaches, such as gesture recognition
6 to identify instances of smoking (Skinner, Stone, Doughty, & Munafò, 2018) offer
7 opportunities for high rates of behavioural measurement.

8 **Challenge 2: Calculating power/sample size**

9 Power analysis has been argued to be the most important statistical procedure when
10 planning a study (Bolger & Laurenceau, 2013). For N-of-1 studies, conducting power
11 analyses can be complex for several reasons. First, power has to be estimated for all levels of
12 analysis, i.e., within participants, and, where of interest, between participants (and potentially
13 further levels, e.g., schools). Second, conducting a power analysis for an N-of-1 study
14 requires in-depth knowledge of the statistical procedures to analyse N-of-1 data. Fortunately,
15 there are resources available that provide hands-on explanations of how to conduct power
16 analyses for N-of-1 studies. Bolger, Stadler, and Laurenceau (2012) and Bolger and
17 Laurenceau (2013), for example, offer a step-by-step approach to conduct power analysis
18 using simulations in Mplus. Third, and perhaps most critically, N-of-1 power analyses require
19 the assumptions about many more parameters than simpler observational and experimental
20 studies, such as effect heterogeneity. Information about these parameters is often not
21 available in the literature as N-of-1 studies in health psychology are still rare, and sometimes
22 researchers fail to report all model parameters. Informed guesses about estimates are
23 therefore often necessary. Some authors like Chen and Chen (2014) conclude from the results
24 of their simulation study that individual design of N-of-1 studies should not be considered
25 unless the effect size is sufficiently large. If using analysis techniques such as Autoregressive

1 Integrated Moving Average (ARIMA) models, some have recommended that at least 50
2 observations are required (Yaffee, 2012). However, such rules of thumb are highly dependent
3 on multiple factors such as effect size, anticipated variance in measures etc and so should be
4 considered with caution. In sum, and as with between-participant studies, those planning an
5 N-of-1 study should consider conducting a power analysis. It prepares the researcher for the
6 later analyses, and sensitises for the importance of detailed reporting of all model parameters
7 in later publications.

8 **Challenge 3: Autocorrelation**

9 A distinct feature of time series data, produced in N-of-1 studies, is that of
10 autocorrelation or serial dependency. This is where sequential data points for a given
11 measure, particularly when there is a short time interval between them, may be associated
12 with each other. For example, a person's stress levels today may be associated with their
13 stress yesterday, which would be an example of a 1st order autocorrelative relationship.
14 Another pattern sometimes observed when examining autocorrelation is day of the week,
15 e.g., lower stress on a Sundays, which would be a 7th order autocorrelative relationship (for a
16 graphical example of the autocorrelation of stress and a definition of autocorrelation see
17 Naughton and Johnston, 2014). Autocorrelation can provide valuable insight into the
18 influence of the past on present and future measurements and therefore modelling, rather than
19 eliminating, autocorrelation is preferable from an analysis perspective (Borckardt, Nash, &
20 Balliet, 2011). Not adjusting for autocorrelation can lead to inaccurate estimates of statistical
21 significance; a positive autocorrelation can increase the risk of a type I error (false positive)
22 and a negative autocorrelation, though less common, can increase the risk of a type II error
23 (false negative) (Vieira, McDonald, Araújo-Soares, Sniehotta, & Henderson, 2017).

24 Two main ways to statistically manage autocorrelation are an autoregressive model or
25 a dynamic model (Kravitz et al., 2014). Most methods using either of these broad approaches

1 enable adjustment for autocorrelation, e.g., ARIMA modelling, dynamic regression-etc.
2 Autocorrelation can also be accounted for when aggregating N-of-1s in a combined analysis,
3 such as through multi-level modelling. While historically the majority of N-of-1 studies have
4 not used appropriate statistical techniques that account for autocorrelation (McDonald et al.,
5 2017), there are multiple techniques such as those listed above that can account for
6 autocorrelation, though innovative statistical approaches are still needed (Davidson &
7 Cheung, 2017). Naughton and Johnston (2014) provide a guide to a simple method for
8 transforming an outcome or predictor variable that takes autocorrelation into account referred
9 to as ‘prewhitening,’ for use when analysing N-of-1s separately. Though care should be taken
10 with this approach if complex autocorrelation is expected beyond a simple 1st and/or 7th
11 autocorrelative relationship in case some of the effect being investigated is removed through
12 the transformation process (Vieira et al., 2017). However, in cases where there is insufficient
13 information to identify and accommodate autocorrelation patterns, often a comprehensive
14 descriptive analysis is preferable to using simpler statistical methods that cannot take
15 autocorrelation into account. While the challenges and solutions presented so far can apply to
16 all N-of-1 studies, there are a number of challenges specific to experimental N-of-1 studies.

17 **Challenges and solutions for experimental N-of-1 designs**

18 Experimental N-of-1 studies (also sometimes referred to as within-person experiments
19 and micro-randomised trials) are most often cross-over trials conducted with one participant
20 acting as their own control. As a result, most individual-level confounders are held constant
21 across different treatment periods; thus, controlling for their potential influence on the
22 outcome of interest. Although treatment periods or blocks can be ordered according to a non-
23 random schedule, it is preferable and more common for blocks to be randomly allocated.
24 Therefore, from here on we will refer to the N-of-1 RCT when referring to experimental N-
25 of-1 designs. N-of-1 RCTs are designed to include a sufficient number of treatment cross-

1 over points in order to minimise the influence of confounding and provide enough data to
2 establish the impact of a given treatment on the outcome of interest. In this section, we will
3 highlight advantages and disadvantages of these designs over between-participant designs
4 and describe how advantages can be optimised and disadvantages mitigated.

5 **Challenge 4: When is an N-of-1 RCT preferable to a traditional between-person RCT**

6 Currently, intervention evaluation design for health psychology-related research is
7 dominated by between-person designs. In most cases, we would estimate that alternative
8 designs, such as within-person methodology, are not considered at the design stage.
9 Understanding why and when an N-of-1 design might be preferable is an important step.
10 Firstly, an N-of-1 RCT could be advantageous when it is assumed that intra-individual effects
11 might differ from those found in between-participant studies. In other words, when
12 individuals have different change trajectories or different types or responses to an
13 intervention. For example, Brannon et al. (2017) demonstrate how a simple text message
14 intervention providing goal attainment feedback for a physical activity-based goal only
15 increased activity among three out of ten adolescents, with different responses depending
16 upon the source of the feedback (parent, peer or behavioural specialist). Furthermore, with N-
17 of-1, we may study whether and how the amount of exposure influences each participant: is
18 one dose enough, does efficacy increase gradually, or is there a saturation point when there
19 has been enough intervention exposure for a change to happen? If participants need different
20 exposure, group comparisons may conceal the effect. N-of-1 RCTs can also be used to
21 identify mechanisms of effect of interventions through assessing the temporal relationships
22 along a mediation pathway, e.g., does a momentary change in self-efficacy precede a change
23 in behaviour as a result of an intervention or is it a change in behaviour that precedes a
24 change in subsequent self-efficacy.

1 From a practical perspective, there are two clear advantages that increase the
2 feasibility of N-of-1 RCTs over group based RCTs, namely time and costs. In fast developing
3 fields, such as mHealth, the implementation of a full-scale RCT may be too slow for practical
4 purposes and technological solutions may be outdated before study results are published. In
5 addition, N-of-1 studies are a feasible platform for tailoring intervention delivery and data
6 collection, possibly increasing engagement and adherence with the intervention elements
7 (Yoon et al., 2018). For instance, participants may be interviewed to identify their preferred
8 physical activity, and the intervention and data collection can be tailored to that specific
9 activity. N-of-1 analyses can then examine which specific components in interventions
10 (elements of intervention, modes of delivery) were suitable and effective for each participant.
11 This type of tailored approach is encouraged in personalised medicine, which is driving
12 tailored health care solutions for individuals (Hood & Friend, 2011).

13 In some cases, for instance in the case of rare diseases, the participant number is
14 limited and there may not be enough statistical power to run a full-scale RCT (e.g., the
15 example of Xeroderma Pigmentosum, systematic intervention development in rare and
16 unstudied skin condition: Sainsbury, Walburn, Araujo-Soares, & Weinman, 2017). Moreover,
17 strict inclusion criteria in group-delivered RCTs aiming for high internal validity could limit
18 the intake of participants so much that external validity of the results would be diminished.
19 An N-of-1 design could provide more flexibility in these situations. The limited number of
20 participants and the high number of observations in N-of-1 studies may also be an advantage
21 for a mixed methods approach to exploring intervention effects, such as explaining
22 quantitative outcomes with patient interviews as done by Daughters, Magidson, Schuster, and
23 Safren (2010). N-of-1 RCTs are usually preferable to traditional RCTs when the intervention
24 can be delivered in a way that avoids carry-over effects and conditions can be randomised
25 within a person.

1 **Challenge 5: Carry-over and slow onset effects**

2 The impact of many types of health psychology-relevant interventions often does not
3 end abruptly after withdrawal of the treatment. For example, changing someone's attitude or
4 enhancing a person's self-efficacy might not be easily reversible, at least not in the short-
5 term. In fact, long-lasting effects are usually the goal of health psychology interventions. In
6 the context of an N-of-1 trial, the treatment effect may, therefore, carry-over into a period
7 when the treatment is removed and may influence participants' responses during this period
8 (Elbourne et al., 2002). In N-of-1 RCTs, health psychologists need to carefully consider
9 carry-over effects of the behaviour change techniques (BCTs) of interest.

10 Where carry-over effects are moderate, additional periods in-between intervention
11 and control periods should be considered to "washout" treatment effects. Washout by design
12 (i.e., purposefully built into the design of the evaluation) is the ideal approach for dealing
13 with carry-over effects, though it is possible to use analytical washout, where, in its simplest
14 form, observations immediately after a treatment has been stopped are excluded. More
15 advanced analytical approaches to address carry-over effects are described elsewhere (e.g.,
16 Senn, 2002). Where there are likely to be enduring carry-over effects from specific
17 interventions or BCTs, these interventions are unlikely to be appropriate for N-of-1 RCT trial
18 evaluations. Instead, more basic cross-over N-of-1 designs could be considered, e.g., AB
19 designs with long baseline and post-intervention data collection periods or 'traditional'
20 nomothetic between group approaches.

21 If such long-term carry-over effects are present, in typical health psychology N-of-1
22 studies this would almost always result in attenuation of any observed treatment effect and so
23 have an overall conservative effect on the study outcomes through inflation of the type II
24 error rate. Less well understood and recognised slow-onset effects refer to a situation in
25 which the full effect of an intervention may not occur immediately (Duan et al., 2013). For

1 example, the effect of daily self-monitoring may be more powerful after several days
2 compared to just the first day. Designing to account for “slow onset effects” can be done in a
3 similar way to designing for carry-over effects.

4 In preparation for N-of-1 trials, health psychologists therefore need to consult the
5 literature or directly investigate (e.g., pilot studies) which interventions or BCTs exhibit
6 carry-over and slow onset effects and the extent of these, and which do not. For this purpose,
7 experimental N-of-1s investigating the temporal dynamics of the effects of behaviour change
8 interventions are particularly valuable. Inauen et al. (2017), for example, investigated the
9 temporal effectiveness of smartphone-based support groups on healthy eating and found that
10 the support effects ended immediately after the end of the support groups. Hence, for this
11 particular behaviour change intervention, assignment or alternation of treatment sequences
12 would be acceptable.

13 A simple approach for identifying short-term carry-over effects is to compare all
14 observations in the period after any active treatments have stopped to those same post-
15 treatment periods but where the first or first few observations after active treatments have
16 stopped are omitted (Duan et al., 2013). If there are no differences then there are unlikely to
17 be short-term carry-over effects, although longer lasting carry-over effects are unlikely to be
18 identifiable using this approach. The same approach can be used to identify slow onset
19 effects, though where the active treatment periods rather than the periods after the treatment
20 period are compared with and without the first observation(s). While statistical techniques
21 can be used to identify carry-over effects, the most common approaches are controversial and
22 are not generally recommended (Duan et al., 2013; Senn, 2002). Therefore, designing, rather
23 than modelling, is the optimal way of managing such effects.

24 **Challenge 6: Identifying appropriate interventions for N-of-1 RCTs**

1 In light of the issue of carry-over effects identified above, only some behavioural
2 interventions or BCTs are likely to be appropriate or feasible for evaluation using N-of-1
3 RCTs, though this is likely to vary according to the behaviour or phenomenon under
4 investigation. As part of this review, we undertook a scoping assessment to explore which
5 BCTs (Michie et al., 2013) are most/least likely to generate carry-over effects. BCTs that are
6 time-specific were identified as being particularly suited to N-of-1 RCTs due to lesser
7 likelihood of carry-over effects. These included BCT domains such as *scheduled*
8 *consequences* (e.g., 14.1 *behaviour cost*) and BCTs such as 12.4 *distraction* and 7.1
9 *prompts/cues* that have temporal boundaries i.e. are most prominent ‘in the moment’. Other
10 BCTs, such as many of those within the *goals* and *planning* domain, could have potentially
11 only small carry-over effects when they were time restricted, e.g., *goal setting* for each day
12 separately rather than for a longer period of time. Goal setting has been successfully applied
13 in factorial N-of-1 RCTs testing and separating effects of goal setting and self-monitoring
14 used to increase physical activity in the general population (Sniehotta, Preece, Hobbs, &
15 Araújo-Soares, 2012) and older individuals (Nyman, Goodwin, Kwasnicka, & Callaway,
16 2016).

17 Less appropriate BCTs include those associated with *learning*, as learning has a high
18 (intended) likelihood of carry-over effects. BCTs which influenced *identity changes* (e.g.,
19 13.1 *identification of self as a role-model*), *attitudes* (e.g., 5.1 *information about health*
20 *consequences*) or *knowledge/skills* of how to conduct a behaviour (e.g., 4.1 *instruction of how*
21 *to perform behaviour*) were also identified as being less appropriate for N-of-1 evaluation.
22 Once this type of information is delivered and processed, it is relatively unlikely to be
23 ‘reversible’, i.e., unlearned, though this is dependent on the information, the individual and
24 the delivery. BCTs including gradual temporal progression, such as 8.6 *graded task*, are also
25 unlikely to be suitable due to the discrepancy between the graduation of tasks and the abrupt

1 end of an intervention between treatment blocks required in the typical N-of-1 RCT design.
2 Therefore, the most suitable BCTs are the ones that have time specific boundaries and the
3 least suitable BCTs are the ones that have intended long-term effects.

4 **Challenge 7: Optimal allocation sequencing and blinding**

5 A complexity experienced when undertaking N-of-1 RCTs, not encountered with
6 most between-person RCTs, is the need to allocate one or more cross-over periods
7 (treatments) to the same individual, as individuals act as their own control. Several authors
8 proposed randomisation of treatment sequences as the gold standard for N-of-1 trials (e.g.,
9 Edgington, 1996; Guyatt et al., 1990; Guyatt et al., 1988; Sackett, 1997). When the effects of
10 a treatment cease immediately after treatment withdrawal, randomisation should lead to the
11 strongest conclusions about the causal nature of the treatment (Tate et al., 2014). The
12 simplest approach for creating an allocation sequence for the simplest design (a single-
13 factorial study) is to randomly assign an equal number of treatment and control blocks during
14 the experimental period (e.g., 7 experimental and 7 control days for a 14-day period). Urn
15 randomisation (e.g., removing intervention and control ‘balls’ one at a time from an urn),
16 which is part of adaptive biased-coin designs, is ideal for an N-of-1 RCTs as it enables a
17 specified (usually equal) number of blocks for each treatment and increases the chances of
18 cross-over periods compared to standard permuted-block randomisation (Naughton &
19 Johnston, 2014). The latter is achieved using urn randomisation as the probability of
20 assignment to each block is influenced by the blocks already allocated – each time a control
21 ‘ball’ is taken from the urn, the chances of an intervention ‘ball’ being selected next
22 increases. Furthermore, when a random sequence is generated, when aggregating N-of-1s,
23 then as each participant will receive a different sequence, any ordering effects should be
24 cancelled out.

1 To help with understanding carry-over effects, particularly when this is not well
2 understood for an intervention of interest, block length can be adapted. For example, different
3 blocks might be created with 1, 2, or 3 control days after an intervention day, to observe the
4 length of possible carry-over effects during these ‘washout’ days. Alternatively, different
5 blocks with 1, 2, or 3 consecutive treatment days may be created to investigate whether
6 multiple treatment days increase treatment effectiveness compared to single days. Blocks
7 with varying lengths can again be randomly allocated to create the full allocation sequence.
8 Blocks with varying numbers of treatment or non-treatment days can also reduce the chances
9 of predicting the change from a treatment to non-treatment period or vice versa when
10 participants are blinded to allocation.

11 When undertaking a blinded N-of-1 RCT, it can be as important to blind participants
12 from cross-over times as it is for treatment allocation for each day. An example of an
13 allocation sequence generated using urn randomisation with different block lengths for a
14 double blind N-of-1 RCT can be found in Naughton and Johnston (2014). When the number
15 of participants is high (i.e., aggregated N-of-1s), different allocation sequences can be
16 assigned to participants to test the effect of the number of treatment days on study outcomes.
17 The same principles of allocation apply if testing multiple interventions simultaneously, e.g.,
18 with a 2x2 factorial design. An allocation sequence can be generated that ensures two or
19 more independent treatments are switched on and off in a random fashion such that the effect
20 of each can be efficiently assessed.

21 **Challenge 8: Type I error risk in N-of-1 designs**

22 A Type I error occurs if the null hypothesis is falsely rejected, i.e., when it is true.
23 Several [factors](#) may contribute to a Type I error in the context of an N-of-1 design. These
24 factors can be specific to N-of-1 designs or common to quantitative research designs in
25 general and just encountered in N-of-1 research. In terms of N-of-1-specific sources, one

1 factor that increases the risk of making a Type I error common to many N-of-1 studies is
2 through a failure to account for a positive autocorrelation, as described in the autocorrelation
3 section. A further factor is that multiple testing can be an issue when undertaking N-of-1
4 studies in at least two scenarios. Firstly, if analysing multiple N-of-1 datasets and evidence of
5 a significant association between variables of interest for any one of them would be
6 considered evidence of an association generalisable to a larger population, then this could be
7 considered a form of multiple testing. Though when undertaking N-of-1 research, inferences
8 about the findings are commonly considered to be limited to those individuals participating,
9 particularly in terms of informing treatment decisions (Guyatt et al., 2000), and so multiple
10 testing would not be relevant. The second scenario is if a series of N-of-1 datasets are
11 analysed individually and subsequently aggregated together in further analyses (e.g., multi-
12 level modelling) where the same general hypothesis is tested, perhaps in multiple subgroups.
13 In principle, this could be considered multiple testing, although a single aggregated analysis
14 would represent only one additional hypothesis test and could be considered only a minor
15 infringement of multiple testing conventions. For both scenarios, if deemed necessary,
16 multiple testing could be managed by a Bonferonni correction or, for a less conservative
17 approach, the false discovery rate could be controlled for using the Benjamini-Hochberg
18 procedure (Benjamini & Hochberg, 1995). Contrary to some erroneous views, analysing
19 repeated measures collected from an individual does not constitute a form of multiple testing
20 as each observation is collected at a separate time point.

21 Other factors inflating a Type I error risk that are not specific to N-of-1 but can be
22 encountered include studies where multiple dependent variables are used ~~measured~~ or where
23 dependent measures are broken down into sub-measures without a priori planning. As with
24 multiple testing, in principle the Type I error risk can be mitigated if required by using
25 correction approaches as described above, although some of these approaches can inflate

1 Type II error risk. A priori specification, such as in an openly accessible or published
2 protocol and statistical analysis plan, can avoid the need for corrections including p value
3 adjustment, which can be overly conservative.

4 **Discussion**

5 Despite a long tradition of regard for between-person RCTs as the design that generates the
6 best evidence, increasingly N-of-1 research is becoming highly valued to promote
7 individualised behavioural medicine and behaviour change interventions (Gabler, Duan,
8 Vohra, & Kravitz, 2011; Lillie et al., 2011). One factor that is likely to be inhibiting the
9 greater use of within-participant methods in health psychology is a lack of training and
10 expertise in these methods. As a consequence, poor management of the challenges identified
11 in this paper can either lead to poorly conducted research or discourage people from using N-
12 of-1 methods altogether, even when these methods would better address the research
13 questions. Challenges relevant to N-of-1 designs include participants' non-adherence to data
14 collection protocols and consequently missing data that is often inevitable with high
15 frequency assessments, resulting in high participant burden. Calculating power in an N-of-1
16 study is also often more of a challenge than in between-person studies due to a greater
17 number of potential parameters, such as the frequency of measurement, the variability in
18 measures over time within and between individuals and the feasible frequency for providing
19 and withdrawing an intervention (for N-of-1 RCTs).

20 The challenges that are specifically relevant to N-of-1 RCTs include assessing when
21 within-person designs are preferable to a traditional between-person RCT – deciding if a high
22 intensity within person assessment is necessary to evaluate predictors, outcomes and
23 alternating interventions. A key question is whether an intervention of interest is likely to
24 demonstrate carry-over or slow onset effects such that it cannot be mitigated through design
25 or statistical counter-measures. Many health psychology interventions aim to rapidly create

1 long lasting effects that are not easily reversible – such interventions are generally not well
2 suited to typical N-of-1 RCT studies with regular (e.g., daily) measurement. The researchers
3 designing N-of-1 RCTs also face a challenge of not achieving optimal allocation sequencing
4 and blinding and increased probability of type I error (as compared to conventional RCT) due
5 to multiple testing on the same participant. As well as identifying some of the key challenges,
6 in this conceptual review we present specific solutions for each of the challenges
7 (summarised in Table 1) and suggest future directions for overcoming them.

8 Please insert Table 1 here

9 **Future directions: overcoming N-of-1 study challenges**

10 As N-of-1 studies focus on single or a relatively small number of individuals assessed
11 with high frequency, they are an ideal method for not only detecting but also explaining the
12 detected behavioural patterns. Collecting and analysing multisource, multimethod data has
13 become easier with modern technology, e.g., N-of-1 RCTs have been conducted with
14 pedometers (Nyman et al., 2016; Sniehotta et al., 2012) and activity bracelets (Nurmi et al.,
15 2015) to study daily steps. Multidisciplinary collaboration is needed to create advanced data
16 management tools that enable integrating and interpreting, e.g., smartphone usage data, i.e.,
17 information from the calendar, social media, and location, to model behaviour more
18 accurately. Adjusting for automatically tracked life events helps to control factors that may
19 bias intervention outcomes. With consent from participating individuals, future studies could
20 utilise the huge pool of personal data collected actively and passively by social networks,
21 internet searches, digital calendars and smartphones, possibly through collaboration with
22 organisations that own the services (Onnela & Rauch, 2016). Another option for data
23 collection are keyboard apps that track all the text typed with a smartphone; the text input
24 could be then analysed with automatic text analysis methods, e.g., Linguistic Inquiry and
25 Word Count (Pennebaker, Francis, & Booth, 2001).

1 Combining multiple sources of data could give a multi-angle picture of the
2 determinants of behaviour (Munafò & Smith, 2018). One possibility is to let participants
3 explore their intervention data using think-aloud methods, helping them to provide possible
4 reasons for their actions (Kwasnicka, Dombrowski, White, & Sniehotta, 2015). Novel
5 technologies also enable self-adapting interventions that learn features that precede desired
6 outcomes and uses those more often in the future (Dallery, Kurti, & Erb, 2015). This type of
7 feature could further enhance smartphone apps that learn about antecedents of health
8 behaviours and then delivers behavioural support tailored to those antecedents when real time
9 sensor monitoring indicates a need or opportunity (Naughton et al., 2016). Evaluation of
10 these novel technologies could draw from engineering models, such as sequential multiphase
11 optimisation treatment (SMART) in which participant responses or characteristics influence
12 the interventions (Collins, Murphy, & Strecher, 2007).

13 N-of-1 research should be participatory as it focusses on building evidence for
14 individuals. Providing insight in data that carries such personal relevance may encourage
15 participation and provide individuals with an opportunity to be actively involved in their
16 health (Dunton, 2018). Furthermore, an invitation to participate in the development of the
17 study may help participants to be more adherent to data collection as explored in the clinical
18 health domain (e.g., Garcia, 2014). Participatory approaches of this kind may include inviting
19 participants to develop the research protocol and to decide on the number and timing of
20 measurements (Orlowski et al., 2015). However, there can be resource implications for this,
21 primarily researcher time. While a participatory approach to the development of N-of-1
22 research may improve adherence, it does introduce new difficulties. A key factor is that it
23 requires researchers to be flexible and open-minded towards participant-initiated changes to
24 research protocols. For example, participants may not all complete the same measures,
25 although a core set of measures can still be specified. Planning an N-of-1 requires taking into

1 account the aforementioned challenges; however, when appropriate, N-of-1 design can
2 provide higher accuracy data on temporal relationships and better intervention adherence than
3 conventional RCT design.

4 **Conceptual review strengths and limitations**

5 N-of-1 design is widely used in other disciplines including behavioural economics
6 and medicine. In health psychology, this design is underutilised due to several design-specific
7 challenges. The key strength of this conceptual review is explanation of the most topical
8 challenges that may often prevent health psychologists and behavioural scientists from
9 undertaking N-of-1 studies. We explored and explained each challenge and provided an
10 actionable solution for how to best overcome it in order to design systematic and high-quality
11 N-of-1 studies. However, the review of N-of-1 challenges is not exhaustive and there are
12 additional practical and methodological challenges that can be explored. Through this review,
13 we would like to encourage a conversation among health psychologists and behavioural
14 scientists about any additional prominent challenges and solutions in order to support and
15 promote the use of N-of-1 design.

16 **Conclusions**

17 The mismatch between the idiographic basis of most theories and intervention causal models
18 and the nomothetic approach of the methodologies typically used to evaluate them is
19 fundamentally problematic. Recent technological and statistical advances facilitate the use of
20 high measurement methodologies such as N-of-1 to test theoretical predictions and response
21 to intervention at the intra-individual level and sidestep this scientific incompatibility.
22 However, there are a mixture of challenges in carrying out N-of-1 studies that are likely to be
23 inhibiting its use in our field. Some of these challenges are unique to within-person
24 methodology, such as statistically addressing autocorrelation and identifying which types of
25 interventions are suitable to be evaluated using experimental N-of-1 design. Other challenges

1 are essentially the same as those encountered with between-person approaches, such as the
2 need to undertake power calculations, managing missing data and multiple testing issues.
3 However, the solutions to these challenges are largely different for within compared to
4 between-person designs. Looking ahead, due to recent technology development we are now
5 able to design behavioural studies and interventions which can be tailored to each individual;
6 this includes through unobtrusive data capture such as wearables and smartphone sensors.
7 These data combined with self-report EMA data can be used to create individual models of
8 behaviour using N-of-1 approaches to develop a truly personalised intervention for each
9 individual. We are therefore at an opportune time to expand our use of within-person designs
10 to better understand health behaviour and to deliver precision behaviour change
11 interventions.
12

References

- 1
2 Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. In W. Bainbridge
3 (Ed.), *Berkshire encyclopedia of Human-Computer Interaction* (pp. 763-768). Great
4 Barrington, MA: Berkshire.
- 5 Barlow, D. H. D. H., & Hersen, M. (1984). *Single case experimental designs; strategies for*
6 *studying behavior change*. Oxford: Pergamon Press.
- 7 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and
8 powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series*
9 *B (Methodological)*, 289-300.
- 10 Bolger, N., & Laurenceau, J. (2013). *Intensive longitudinal methods: An introduction to diary*
11 *and experience sampling research*. New York: Guilford Press.
- 12 Bolger, N., Stadler, G., & Laurenceau, J. P. (2012). Power analysis for intensive longitudinal
13 studies. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for*
14 *studying daily life* (pp. 285-301). New York: Guilford Press.
- 15 Borckardt, J. J., Nash, M. R., Balliet W. (2011). Statistical Analysis of Single-Case Data. In:
16 G.J. Madden, T. Hackenberg, K.A. Lattal (Eds.). *APA Handbook of Behavior Analysis*.
17 Washington, DC: APA.
- 18 Brannon, E. E., Cushing, C. C., Walters, R. W., Crick, C., Noser, A. E., & Mullins, L. L.
19 (2018). Goal feedback from whom? A physical activity intervention using an N-of-1
20 RCT. *Psychology & Health*, 33(6), 701-712. doi:10.1080/08870446.2017.1385783
- 21 Burg, M. M., Schwartz, J. E., Kronish, I. M., Diaz, K. M., Alcantara, C., Duer-Hefele, J., &
22 Davidson, K. W. (2017). Does stress result in you exercising less? Or does exercising
23 result in you being less stressed? Or is it both? Testing the bi-directional stress-exercise
24 association at the group and person (N of 1) level. *Annals of Behavioral Medicine*,
25 51(6), 799-809. doi:10.1007/s12160-017-9902-4
- 26 Cain, A. E., Depp, C. A., & Jeste, D. V. (2009). Ecological momentary assessment in aging
27 research: a critical review. *Journal of Psychiatric Research*, 43(11), 987-996.
28 doi:10.1016/j.jpsychires.2009.01.014
- 29 Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and
30 within-subject design. *Journal of Economic Behavior & Organization*, 81(1), 1-8.
31 doi:10.1016/j.jebo.2011.08.009
- 32 Chen, X., & Chen, P. (2014). A comparison of four methods for the analysis of N-of-1 trials.
33 *PLoS ONE*, 9(2), e87752. doi:10.1371/journal.pone.0087752

- 1 Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy
2 (MOST) and the sequential multiple assignment randomized trial (SMART): new
3 methods for more potent eHealth interventions. *American Journal of Preventive*
4 *Medicine*, 32(5), S112-S118. doi:10.1016/j.amepre.2007.01.022
- 5 Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008).
6 Developing and evaluating complex interventions: the new Medical Research Council
7 guidance. *BMJ*, 337, a1655. doi:10.1136/bmj.a1655
- 8 Dallery, J., Kurti, A., & Erb, P. (2015). A new frontier: Integrating behavioral and digital
9 technology to promote health behavior. *The Behavior Analyst*, 38(1), 19-49.
10 doi:10.1007/s40614-014-0017-y
- 11 Daughters, S. B., Magidson, J. F., Schuster, R. M., & Safren, S. A. (2010). ACT HEALTHY:
12 A combined cognitive-behavioral depression and medication adherence treatment for
13 HIV-infected substance users. *Cognitive and Behavioral Practice*, 17(3), 309-321.
14 doi:10.1016/j.cbpra.2009.12.003
- 15 Davidson, K. W., & Cheung, Y. K. (2017). Envisioning a future for precision health
16 psychology: innovative applied statistical approaches to N-of-1 studies. *Health*
17 *Psychology Review*, 11(3), 292-294. doi: 10.1080/17437199.2017.1347514
- 18 Davidson, K. W., Peacock, J., Kronish, I. M., & Edmondson, D. (2014). Personalizing
19 behavioral interventions through single-patient (N-of-1) trials. *Social and Personality*
20 *Psychology Compass*, 8(8), 408-421. doi: 10.1111/spc3.12121
- 21 Deeg, D. J., van Tilburg, T., Smit, J. H., & de Leeuw, E. D. (2002). Attrition in the Longitudinal
22 Aging Study Amsterdam: the effect of differential inclusion in side studies. *Journal of*
23 *Clinical Epidemiology*, 55(4), 319-328. doi:10.1016/S0895-4356(01)00475-9
- 24 Duan, N., Kravitz, R. L., & Schmid, C. H. (2013). Single-patient (n-of-1) trials: a pragmatic
25 clinical decision methodology for patient-centered comparative effectiveness research.
26 *Journal of Clinical Epidemiology*, 66(8), S21-S28. doi: 10.1016/j.jclinepi.2013.04.006
- 27 Dunton, G. F. (2018). Sustaining health-protective behaviors such as physical activity and
28 healthy eating. *JAMA*. doi:10.1001/jama.2018.6621
- 29 Edgington, E. S. (1996). Randomized single-subject experimental designs. *Behaviour*
30 *Research and Therapy*, 34(7), 567-574. doi:10.1016/0005-7967(96)00012-5
- 31 Elbourne, D. R., Altman, D. G., Higgins, J. P., Curtin, F., Worthington, H. V., & Vail, A.
32 (2002). Meta-analyses involving cross-over trials: methodological issues. *International*
33 *Journal of Epidemiology*, 31(1), 140-149. doi:10.1093/ije/31.1.140

- 1 Feng, D., Cong, Z., & Silverstein, M. (2012). Missing data and attrition. In J. T. Newsom, R.
2 N. Jones, & S. M. Hofer (Eds.). *Longitudinal data analysis: A practical guide for*
3 *researchers in aging, health, and social sciences* (71-96). New York: Routledge.
- 4 Gabler, N. B., Duan, N., Vohra, S., & Kravitz, R. L. (2011). N-of-1 trials in the medical
5 literature: a systematic review. *Medical Care, 49*(8), 761-768.
- 6 Garcia, J. C. (2014). *Beyond the numbers: A user-centered design approach for personal*
7 *reflective healthcare technologies* (Doctoral dissertation). Retrieved from
8 <https://research.utwente.nl/files/6088735/Thesis-JuanJimenez.pdf>
- 9 Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review*
10 *of Psychology, 60*, 549-576. doi:10.1146/annurev.psych.58.110405.085530
- 11 Guyatt, G., Keller, J. L., Jaeschke, R., Rosenbloom, D., Adachi, J. D., & Newhouse, M. T.
12 (1990). The n-of-1 randomized controlled trial: clinical usefulness. Our three-year
13 experience. *Annals of Internal Medicine, 112*(4), 293-299. doi: 10.7326/0003-4819-
14 112-4-293
- 15 Guyatt, G., McGinn, T. G., Wyer, P. C., Naylor, C. D., Stiell, I. G., Richardson, W. S., &
16 Group, E.-B. M. W. (2000). Users' guides to the medical literature: XXII: how to use
17 articles about clinical decision rules. Evidence-Based Medicine Working Group.
18 *JAMA, 284*(1), 79-84. doi:10.1001/jama.284.1.79
- 19 Guyatt, G., Sackett, D., Adachi, J., Roberts, R., Chong, J., Rosenbloom, D., & Keller, J. (1988).
20 A clinician's guide for conducting randomized trials in individual patients. *Canadian*
21 *Medical Association Journal, 139*(6), 497.
- 22 Hogarth, R. M. (2005). The challenge of representative design in psychology and economics.
23 *Journal of Economic Methodology, 12*(2), 253-263. doi:10.1080/13501780500086172
- 24 Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section
25 data. *American Journal of Political Science, 54*(2), 561-581. doi:10.1111/j.1540-
26 5907.2010.00447.x
- 27 Hood, L., & Friend, S. H. (2011). Predictive, personalized, preventive, participatory (P4)
28 cancer medicine. *Nature Reviews Clinical Oncology, 8*(3), 184. doi:
29 10.1038/nrclinonc.2010.227
- 30 Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of
31 single-subject research to identify evidence-based practice in special education.
32 *Exceptional Children, 71*(2), 165-179. doi:10.1177/001440290507100203

- 1 Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing: A comparison of missing
2 data methods and software to fit incomplete data regression models. *The American*
3 *Statistician*, *61*(1), 79-90. doi:10.1198/000313007X172556
- 4 Inauen, J., Bolger, N., Shrout, P. E., Stadler, G., Amrein, M., Rackow, P., & Scholz, U. (2017).
5 Using smartphone-based support groups to promote healthy eating in daily life: A
6 randomised trial. *Applied Psychology: Health and Well-Being*, *9*(3), 303-323.
7 doi:10.1111/aphw.12093
- 8 Inauen, J., Shrout, P. E., Bolger, N., Stadler, G., & Scholz, U. (2016). Mind the gap? An
9 intensive longitudinal study of between-person and within-person intention-behavior
10 relations. *Annals of Behavioral Medicine*, *50*(4), 516-522. doi:10.1007/s12160-016-
11 9776-x
- 12 Janosky, J. (2005). Use of the single subject design for practice based primary care research.
13 *Postgraduate Medical Journal*, *81*(959), 549-551. doi:10.1136/pgmj.2004.031005
- 14 Johnston, D. W., & Johnston, M. (2013). Useful theories should apply to individuals. *British*
15 *Journal of Health Psychology*, *18*(3), 469-473. doi:10.1111/bjhp.12049
- 16 Kennedy, C. H. (2005). *Single-case designs for educational research*. London: Pearson.
- 17 Kent, A., & Bush, P. (2018). Co-designing for care: Craft and wearable wellbeing. In R. Lynch
18 & C. Farrington (Eds.), *Quantified lives and vital data: Exploring health and*
19 *technology through personal medical devices* (pp. 251-280). Basingstoke, UK:
20 Palgrave Macmillan.
- 21 Kravitz, R., Duan, N., Duan, N., Eslick, I., Gabler, N., Kaplan, H., . . . Schmid, C. (2014).
22 Design and implementation of N-of-1 trials: a user's guide. *Agency for healthcare*
23 *research and quality*. Rockville, MD: Agency for Healthcare Research and Quality.
- 24 Kwasnicka, D., Dombrowski, S. U., White, M., & Sniehotta, F. F. (2015). Data-prompted
25 interviews: Using individual ecological data to stimulate narratives and explore
26 meanings. *Health Psychology*, *34*(12), 1191-1194. doi:10.1037/hea0000234
- 27 Kwasnicka, D., Dombrowski, S. U., White, M., & Sniehotta, F. F. (2017). N-of-1 study of
28 weight loss maintenance assessing predictors of physical activity, adherence to weight
29 loss plan and weight change. *Psychology & Health*, *32*(6), 686-708.
30 doi:10.1080/08870446.2017.1293057
- 31 Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., & Schork, N. J. (2011). The n-of-1
32 clinical trial: the ultimate strategy for individualizing medicine? *Personalized*
33 *Medicine*, *8*(2), 161-173. doi:10.2217/pme.11.7

- 1 McDonald, S., Quinn, F., Vieira, R., O'Brien, N., White, M., Johnston, D. W., & Sniehotta, F.
2 F. (2017). The state of the art and future opportunities for using longitudinal n-of-1
3 methods in health behaviour research: a systematic literature overview. *Health*
4 *Psychology Review, 11*(4), 307-323. doi:10.1080/17437199.2017.1316672
- 5 Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., . . . Wood,
6 C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically
7 clustered techniques: building an international consensus for the reporting of behavior
8 change interventions. *Annals of Behavioral Medicine, 46*(1), 81-95.
9 doi:10.1007/s12160-013-9486-6
- 10 Moeller, J. D., Dattilo, J., & Rusch, F. (2015). Applying quality indicators to single-case
11 research designs used in special education: A systematic review. *Psychology in the*
12 *Schools, 52*(2), 139-153. doi:10.1002/pits.21801
- 13 Munafò, M. R., & Smith, G. D. (2018). Robust research needs many lines of evidence. *Nature,*
14 *553,* 399-401. doi: 10.1038/d41586-018-01023-3
- 15 Naughton, F., Hopewell, S., Lathia, N., Schalbroeck, R., Brown, C., Mascolo, C., . . . Sutton,
16 S. (2016). A context-sensing mobile phone app (Q sense) for smoking cessation: a
17 mixed-methods study. *JMIR mHealth and uHealth, 4*(3). doi: 10.2196/mhealth.5787
- 18 Naughton, F., & Johnston, D. (2014). A starter kit for undertaking n-of-1 trials. *European*
19 *Health Psychologist, 16*(5), 196-205.
- 20 Nurmi, J., Knittle, K., Helf, C., Zwickl, P., Palacios, P. L., Tejedor, C. C., . . . Haukkala, A.
21 (2015). A personalised, sensor-based smart phone intervention for physical activity and
22 Diet-PRECIOUS N-of-1 trial. *Frontiers in Public Health, 4*(1), 1-1. doi:
23 10.3389/conf.FPUBH.2016.01.00098
- 24 Nyman, S. R., Goodwin, K., Kwasnicka, D., & Callaway, A. (2016). Increasing walking among
25 older people: A test of behaviour change techniques using factorial randomised N-of-1
26 trials. *Psychology & Health, 31*(3), 313-330. doi:10.1080/08870446.2015.1088014
- 27 Onghena, P., & Edgington, E. S. (2005). Customization of pain treatments. Single case design
28 and analysis. *The Clinical Journal of Pain, 21*(1), 56-68.
- 29 Onnela, J.-P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to
30 enhance behavioral and mental health. *Neuropsychopharmacology, 41*(7), 1691.
31 doi:10.1038/npp.2016.7
- 32 Orłowski, S. K., Lawn, S., Venning, A., Winsall, M., Jones, G. M., Wyld, K., . . . Smith, D.
33 (2015). Participatory research as one piece of the puzzle: a systematic review of

- 1 consumer involvement in design of technology-based youth mental health and well-
2 being interventions. *JMIR human factors*, 2(2). doi:10.2196/humanfactors.4361
- 3 Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count:
4 LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001), 2001.
- 5 Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.
6 doi:10.2307/2335739
- 7 Sackett, D. L. (1997). *Evidence-based medicine how to practice and teach EBM*. London:
8 Churchill Livingstone.
- 9 Sainsbury, K., Walburn, J., Araujo-Soares, V., & Weinman, J. (2017). Challenges and
10 proposed solutions for formative research to inform systematic intervention
11 development in rare and unstudied conditions: The case example of Xeroderma
12 Pigmentosum. *British Journal of Health Psychology*, 23(2), 229-237.
13 doi:10.1111/bjhp.12287
- 14 Schüz, N., Walters, J. A., Frandsen, M., Bower, J., & Ferguson, S. G. (2013). Compliance with
15 an EMA monitoring protocol and its relationship with participant and smoking
16 characteristics. *Nicotine & Tobacco Research*, 16(Suppl_2), S88-S92.
17 doi:10.1093/ntr/ntt142
- 18 Senn, S. (2002). *Cross-over trials in clinical research* (Vol. 5): John Wiley & Sons.
- 19 Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*.
20 New York: Basic Books.
- 21 Sieverink, F., Kelders, S. M., & van Gemert-Pijnen, J. E. (2017). Clarifying the concept of
22 adherence to eHealth technology: Systematic review on when usage becomes
23 adherence. *Journal of Medical Internet Research*, 19(12). doi:10.2196/jmir.8578
- 24 Skinner, A. L., Stone, C. J., Doughty, H., & Munafò, M. R. (2018). StopWatch: The
25 preliminary evaluation of a smartwatch-based system for passive detection of cigarette
26 smoking. *Nicotine & Tobacco Research*, 0(0), 1-5. doi:10.1093/ntr/nty008
- 27 Smith, G., Williams, L., O'Donnell, C., & McKechnie, J. (2018). A series of n-of-1 studies
28 examining the interrelationships between social cognitive theory constructs and
29 physical activity behaviour within individuals. *Psychology & Health*, 1-16. doi:
30 10.1080/08870446.2018.1500576
- 31 Sniehotta, F. F., Pesseau, J., Hobbs, N., & Araújo-Soares, V. (2012). Testing self-regulation
32 interventions to increase walking using factorial randomized N-of-1 trials. *Health*
33 *Psychology*, 31(6), 733-737. doi: 10.1037/a0027337

- 1 Stappers, P., & Giaccardi, E. (2017). Research through Design. In M. Soegaard & R. Friis-
2 Dam (Eds.) *The encyclopedia of human-computer interaction*, (pp. 43). Aarhus N,
3 Denmark: The Interaction Design Foundation.
- 4 Tate, R., Perdices, M., McDonald, S., Togher, L., & Rosenkoetter, U. (2014). The design,
5 conduct and report of single-case research: Resources to improve the quality of the
6 neurorehabilitation literature. *Neuropsychological Rehabilitation*, *24*(3-4), 315-331.
7 doi:10.1080/09602010802009201
- 8 Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological
9 science. *Current Directions in Psychological Science*, *23*(6), 466-470.
10 doi:10.1177/0963721414550706
- 11 Twisk, J., & de Vente, W. (2002). Attrition in longitudinal studies: how to deal with missing
12 data. *Journal of Clinical Epidemiology*, *55*(4), 329-337. doi:10.1016/S0895-
13 4356(01)00476-0
- 14 Versluis, A., Verkuil, B., Spinhoven, P., van der Ploeg, M. M., & Brosschot, J. F. (2016).
15 Changing mental health and positive psychological well-being using ecological
16 momentary interventions: a systematic review and meta-analysis. *Journal of Medical
17 Internet Research*, *18*(6), e152. doi: 10.2196/jmir.5642
- 18 Vieira, R., McDonald, S., Araújo-Soares, V., Sniehotta, F. F., & Henderson, R. (2017).
19 Dynamic modelling of n-of-1 data: powerful and flexible data analytics applied to
20 individualised studies. *Health Psychology Review*, *11*(3), 222-234.
21 doi:10.1080/17437199.2017.1343680
- 22 Wen, C. K. F., Schneider, S., Stone, A. A., & Spruijt-Metz, D. (2017). Compliance with mobile
23 ecological momentary assessment protocols in children and adolescents: A systematic
24 review and meta-analysis. *Journal of Medical Internet Research*, *19*(4).
25 doi:10.2196/jmir.6641
- 26 Yaffee, R. (2012). *An Introduction to Forecasting Time Series with Stata*. New York: Taylor
27 & Francis.
- 28 Yoon, S., Schwartz, J. E., Burg, M. M., Kronish, I. M., Alcantara, C., Julian, J., . . . Diaz, K.
29 M. (2018). Using behavioral analytics to increase exercise: A randomized N-of-1 study.
30 *American Journal of Preventive Medicine*, *54*(4), 559-567.
31 doi:10.1016/j.amepre.2017.12.011
- 32
33

Table 1

Challenges and solutions of N-of-1 studies in Health Psychology

Challenge type/name	Challenge defined	Proposed solution
<i>Challenge 1: Non-adherence to data collection and missing data</i>	N-of-1 study requires repetitive measurements performed on the same person, non-adherence to study protocol often leads to missing data; patterns of missing data vary.	Imputation under certain circumstances; employing user centred design; using non-obtrusive data collection techniques (e.g., sensors, GPS data) and novel technologies to collect data on the participant.
<i>Challenge 2: Calculating power/sample size</i>	Power of an N-of-1 study is relevant to the number of observations (not to the number of participants). As compared to conventional RCT designs, additional parameters need to be considered when calculating power/sample size.	Employing resources available to conduct power analyses for N-of-1 studies, for example a step-by-step approach using simulations in Mplus by Bolger, Stadler, and Laurenceau (2012) and Bolger and Laurenceau (2013). Sample size heuristics (e.g. a minimum of 50 observations to run an ARIMA model; Yaffee, 2012) are suggested by some, though these should be used with caution.
<i>Challenge 3: Autocorrelation</i>	Sequential data points, particularly when there is a short time interval between them, may be associated with each other. Not taking autocorrelation into account can lead to inaccurate estimates of statistical significance.	Using one of the many statistical techniques that enables the modelling or adjustment of autocorrelation, e.g., prewhitening, ARIMA modelling, dynamic regression, multi-level modelling.
<i>Challenge 4: When is an N-of-1 RCT preferable to a traditional between-person RCT?</i>	Understanding the circumstances when an N-of-1 RCT would be preferable to a traditional between-person RCT.	N-of-1 RCTs are preferable when: intra-individual effects differ from those found in between-participant studies; to study whether and how the amount of exposure influences each participant; to explore causal temporal relationships within the participant; to save time, cost, and to tailor

		<p>interventions effectively; to investigate rare conditions and behavioural treatments for rare conditions.</p>
<i>Challenge 5: Carry-over and slow onset effects</i>	<p>The impact of health psychology-relevant interventions sometimes does not start abruptly after a treatment is initiated (slow onset effect) and often does not end abruptly after withdrawal of the treatment (carry-over effect).</p>	<p>Using wash-out periods (e.g., no intervention present), or analytical wash-out, i.e., omitting certain intervention periods in analysis. Piloting interventions to explore if the effects do carry over time or have slow onset. Using single cross-over designs (e.g. AB designs) with long data collection periods.</p>
<i>Challenge 6: Identifying appropriate interventions for N-of-1 RCTs?</i>	<p>Not all interventions or Behaviour Change Techniques (BCTs) are suitable to be used in N-of-1 experimental studies due mainly to excessive carry over effects.</p>	<p>Only using N-of-1 methodology when testing appropriate interventions or BCTs, i.e., those that are time specific and not designed to cause enduring changes in participants. BCTs likely to be least suitable for N-of-1 RCTs are those that include learning, changing identity, and gaining knowledge or skills.</p>
<i>Challenge 7: Optimal allocation sequencing and blinding</i>	<p>The need to maximise the number of cross-over periods and, where appropriate, minimise the ability of participants to predict cross-over points in N-of-1 experimental studies.</p>	<p>Randomisation of treatment sequences has been proposed as the gold standard for N-of-1 trials, i.e., to randomly assign an equal number of treatment and control blocks during the experimental period. Urn randomisation is ideal as it enables a specified (usually equal) number of blocks for each treatment and increases the chances of cross-over periods compared to standard permuted-block randomisation. Differential (e.g., random) treatment block size can reduce predictability.</p>
<i>Challenge 8: Type I error risk in N-of-1 designs</i>	<p>It is possible that when using N-of-1 designs in certain ways the risk of a Type I error is inflated, such as when generalising from the findings of any identified effects across N-of-1 investigations, when analysing the same data more than once (e.g., as separate cases and in aggregated</p>	<p>Multiple testing could be managed by a Bonferonni correction or, for a less conservative approach, the false discovery rate could be controlled for using the Benjamini-Hochberg procedure (1995). A priori specification, such as in an openly accessible or published protocol and statistical analysis plan, can avoid the need for adjustment.</p>

analyses) and when more than one dependent variable is analysed.
