Increasing walking among older people: A test of behaviour change techniques using factorial randomised *N*-of-1 trials

Samuel R. Nyman[a,b]* ⓘ, Kelly Goodwin[c], Dominika Kwasnicka[d] ⓘ and Andrew Callaway[c] ⓘ

[a]Faculty of Science and Technology, Department of Psychology, Bournemouth University, Dorset, UK; [b]Bournemouth University Dementia Institute, Bournemouth University, Dorset, UK; [c]Faculty of Management, Department of Sport and Physical Activity, Bournemouth University, Dorset, UK; [d]Faculty of Medical Sciences, Institute of Health and Society, Newcastle University, Fuse - UKCRC Centre for Translational Research in Public Health, Newcastle Upon Tyne, UK

Objective: Evaluations of techniques to promote physical activity usually adopt a randomised controlled trial (RCT). Such designs inform how a technique performs on average but cannot be used for treatment of individuals. Our objective was to conduct the first *N*-of-1 RCTs of behaviour change techniques with older people and test the effectiveness of the techniques for increasing walking within individuals.
Design: Eight adults aged 60–87 were randomised to a 2 (goal-setting vs. active control) × 2 (self-monitoring vs. active control) factorial RCT over 62 days. The time series data were analysed for each single case using linear regressions.
Main outcome measures: Walking was objectively measured using pedometers.
Results: Compared to control days, goal-setting increased walking in 4 out of 8 individuals and self-monitoring increased walking in 7 out of 8 individuals. While the probability for self-monitoring to be effective in 7 out of 8 participants was beyond chance ($p = .03$), no intervention effect was significant for individual participants. Two participants had a significant but small linear decrease in walking over time.
Conclusion: We demonstrate the utility of *N*-of-1 trials for advancing scientific enquiry of behaviour change and in practice for increasing older people's physical activity.

Keywords: behaviour change; *N*-of-1; older people; physical activity; self-regulation; walking

Introduction

The randomised controlled trial (RCT) is considered the gold standard design for evaluating the effectiveness of an intervention, including interventions to increase the health of older people (e.g. Barton, 2000). Conventional RCTs are used to estimate the average effect of an intervention in a population and usually require large sample sizes to achieve adequate statistical power. However, even if a significant positive effect for an

*Corresponding author. Email: snyman@bournemouth.ac.uk

intervention is demonstrated, the average effect often masks the variance within the intervention group where some individuals may not benefit or may even perform worse on the primary outcome (Guyatt et al., 1986; Johnston & Johnston, 2013). Therefore, conventional RCTs are limited in that they are unable to identify the mechanisms for change nor identify within- and between-participant variability in the intervention's effectiveness (Craig et al., 2008). N-of-1 trials offer an alternative design to overcome these limitations with the use of intensive data collection in small samples while still maintaining the internal validity of traditional trials such as experimental design and randomisation to conditions (Craig et al., 2008). In N-of-1 trials, instead of randomising groups of participants to experimental conditions (intervention or control), individual participants are randomised to conditions in a pre-determined order and time series fashion whereby each participant is exposed to both the intervention and control group on different days of the trial period.

The main function of N-of-1 trials is to evaluate the effectiveness of interventions in individuals and in so doing provide further data as to the heterogeneity of the effect, test theory and often inform the design of conventional trials (Craig et al., 2008). For example, the theory of planned behaviour was recently tested in two studies that used a series of N-of-1 trials and both found partial support for the theory. One study found the theory to predict at least one of three forms of physical activity in five of the six individuals tested, but overall the theory was of variable predictive utility within individuals and across physical activity behaviours (Hobbs, Dixon, Johnston, & Howie, 2013). Another study with individuals with chronic pain found the theory to predict walking activity, but it rarely predicted limitations in walking and when it did it was in the opposite direction than hypothesised (Quinn, Johnston, & Johnston, 2013). As well as a tool for scientific research, N-of-1 trials can inform best practice in patient care and in particular for older people (Price & Grimley Evans, 2002). For example, N-of-1 trials have been used to identify optimal treatment for patients (Scuffham et al., 2010) and improve access to expensive medications (Scuffham, Yelland, Nikles, Pietrzak, & Wilkinson, 2008).

However, despite a long history in other psychological disciplines such as clinical psychology and a more recent history in pharmaceutical medicine (Barlow, Knock, & Hersen, 2008; Gabler, Duan, Vohra, & Kravitz, 2011), there has been scant use of the N-of-1 design within health psychological science. This is to the detriment of current scientific understanding of behaviour and behaviour change, as the dominant theoretical models currently in use, such as the theory of planned behaviour, have been almost entirely evaluated with between-groups designs including conventional RCTs. While theories may be applicable on average and interventions may be successful on average, they may not necessarily be successful when applied to individuals (Johnston & Johnston, 2013; Molenaar & Campbell, 2009). It then follows that the lack of within-person research impedes health care as practitioners have difficulty in applying the evidence-base to their patients (Price & Grimley Evans, 2002).

In the current study, we report the first, to our knowledge, N-of-1 factorial RCT to test the effect of behaviour change techniques (BCTs) to increase physical activity among older people. We build upon the first N-of-1 factorial RCT of BCTs on physical activity that although recruited one participant aged 67, was primarily conducted with young and middle-aged adults with an average age of 36.9 years (Sniehotta, Presseau, Hobbs, & Araújo-Soares, 2012). We conducted a partial-replication of this earlier study

with two BCTs used in control theory (Carver & Scheier, 1982). Control theory proposes a model of self-regulation whereby individuals set themselves a target to achieve and then they monitor their behaviour against that target. If there is a discrepancy between the target and behaviour, then either efforts are made to reduce the discrepancy, or the individual withdraws from the ambition of reaching the target. If there is no discrepancy between the target and behaviour, then no more effort is undertaken (Carver & Scheier, 1982).

The techniques that were tested in the current study were goal-setting and self-monitoring. For goal-setting, an individual is asked to define something that they want to achieve, and if this is clear, specific and challenging, then they will generally perform better than simply being asked to try their best (Locke & Latham, 1990, 2002). For self-monitoring, an individual is asked to regularly monitor their behaviour and compare their progress against their goals. For example, pedometers are used to monitor walking step counts and have been found to significantly increase physical activity among the general adult population including older people (Baker et al., 2008; Bravata et al., 2007; Fitzsimons, Baker, Gray, Nimmo, & Mutrie, 2012). In systematic reviews, both self-regulatory techniques have been found to be effective, and in particular self-monitoring, for improving healthy eating and physical activity in adults from the general population (Bird et al., 2013; Michie, Abraham, Whittington, McAteer, & Gupta, 2009), who are chronically ill (Conn, Hafdahl, Brown, & Brown, 2008), and obese (Dombrowski et al., 2012). These self-regulation techniques are also included in UK national guidelines for improving public health through physical activity and BCTs (National Institute for Health and Clinical Excellence, 2006, 2007). In the previous *N*-of-1 factorial RCT of BCTs (Sniehotta et al., 2012), across the 10 participants overall, goal-setting approached significance and self-monitoring led to a significant increase in walking. Individually, goal-setting significantly increased walking in two of the 10 participants and self-monitoring significantly increased walking in 2 other participants. However, these findings with adults of a mean age of 36.9 years cannot be extrapolated to older populations and requires testing.

Age may influence how BCTS are applied and their relative efficacy in engaging individuals to adopt healthy behaviours (e.g. Renner, Spivak, Kwon, & Schwarzer, 2007). For example, a study indicated that the use of coping planning – a BCT to anticipate barriers and mentally simulate successfully overcoming them – mediated the success of interventions designed to increase physical activity (Scholz, Sniehotta, Burkert, & Schwarzer, 2007). However, older adults had a different pattern of using this technique compared with middle-aged and younger adults; older adults had high levels of coping planning at baseline and so did not increase in their use of this technique over time like their younger counterparts. Therefore, further *N*-of-1 factorial RCTs are required to test self-regulation techniques within older people's walking activity. Walking was the target health behaviour as it does not require any special skills or equipment, is convenient, self-regulated, inherently safe (Morris & Hardman, 1997) and has been a consistently popular activity among the general population including older people (Canada Fitness Survey, 1983; Nathan, Wood, & Giles-Corti, 2014; Scholes & Mindell, 2013; Skelton, Young, Walker, & Hoinville, 1999). But despite the established health benefits of physical activity (Department of Health, Physical Activity, Health Improvement, and Protection, 2011), the general population and older people in particular are insufficiently physically active (Department of Health, Physical Activity, Health

Improvement, and Protection, 2011), and the evidence for interventions to increase walking have only demonstrated efficacy rather than effectiveness (Ogilvie et al., 2007). Therefore, further trials are required to establish how best to promote walking among older people.

Methods

Design

Eight 2 (goal-setting vs. active control) × 2 (self-monitoring vs. active control) factorial randomised controlled *N*-of-1 trials were conducted in the south of England. Participants were randomised daily to either intervention or control conditions (days) over a period of 62 days independently for goal-setting and self-monitoring. With the factorial design, this meant that the interventions were randomised and analysed separately with approximately 31 days' data for each condition per participant (on some days participants did both interventions, on some days none and on other days only one of the two interventions). The randomisation sequence was computer generated by a blinded member of the research team. A study period of 62 days was used, similar to the previous *N*-of-1 RCT (Sniehotta et al., 2012), based on the rule of thumb that 30 participants per condition arm would provide at least 80% statistical power (Cohen, 1988).

Intervention

For goal-setting conditions, participants were requested to set themselves a realistic goal to achieve each day. The intervention required participants to set themselves a goal for how many steps they would walk that day and write it down. The active control required participants to set themselves a goal for how many fruit and vegetables they would consume that day and write it down. For self-monitoring conditions, participants were provided with two pedometers (Omron Walking Style III, HJ-203-EK) and asked to wear one of them each day. The intervention required participants to wear a pedometer with the step count visible to allow for self-monitoring (open condition). The active control required participants to wear a pedometer that was covered with tape to prevent participants from being able to see the number of steps walked that day (sealed condition).

Measurements

The primary outcome for the study was objectively measured walking behaviour recorded through daily pedometer step counts. The pedometer used in this study has been validated in previous studies against triaxial accelerometry and direct observation (Steeves et al., 2011; Sugden et al., 2008). In the current study, the pedometers were piloted to ensure equivalent measurement of step counts. Three pedometers were randomly selected and placed in the right pocket of a volunteer whose step counts were recorded over four days on all pedometers to test intra-reliability. The percentage error from the mean of the three devices was calculated (.68, −2.31, and 1.63%) along with Cronbach's α (.998), which indicated that the pedometers had a very high level of agreement in measurement of step counts between devices. Thus, any differences in step counts observed during the study within participants would be highly unlikely to result from measurement error.

Participants

Prior to recruitment, the study was approved by the Bournemouth University, School of Tourism Research Ethics Committee. Participants were 10 adults aged 60+ who were recruited from a weekly aerobic exercise class for the over 60s held at a church hall. To be included in the study, each participant had to be aged 60+, living in their own home (not in residential care) and able to commit to the study for the whole duration. After participants had consented to take part in the study, they completed a health screening questionnaire (Australian Sports Commission, 2000) to ensure they were fit to participate. They were then provided with a pedometer to wear for one week with no randomisation to conditions. This served to familiarise participants with the study requirements, get used to wearing the pedometer each day, and avoid future dropouts from the study. After the first week with no randomisation, one participant declined to continue with the study. In addition, after the study was completed, another participant was excluded from the analysis as they had rarely worn the pedometer and provided only minimal data over the study period (69% data missing). Therefore, a total of eight participants were included in the statistical analysis: five females and three males aged 60–87 ($M = 71.75$, SD = 9.63).

Procedure

The intention was to contact participants each morning via mobile phone text message as conducted previously with younger adults (Sniehotta et al., 2012). However, the participants reported that they did not frequently use mobile phones and did not wish to be notified of condition allocation by text message. Therefore, the study protocol was changed and instead, participants agreed to be instructed of allocation to experimental conditions using a series of ordered and sealed envelopes and opening one of them each morning. Each envelope contained a message indicating the type of goal to be set and pedometer to be worn for that day. Each week, a researcher met the participants at the community hall where they undertook their exercise class. During the class, the researcher recorded the pedometer step counts for the previous week and issued the next set of envelopes to be opened for the following week. At these meetings, participants were asked if they had tampered with the devices, such as peeling off the tape to view step counts (in active control condition vs. self-monitoring) or shaking the pedometers to artificially increase the step count. Participants volunteered that on occasion they had done either of these, and on these days the data were excluded due to violation of experiment condition (see Table 1).

Analysis

Analyses were conducted using IBM SPSS.20. Descriptive statistics were first calculated for the overall sample and each participant. As the data contained distributions that had some abnormality (mainly due to outliers), medians and interquartile ranges are reported. Thereafter, the data were analysed for each participant separately. The data were then prepared for inferential statistics by addressing missing cases and autocorrelation.

Table 1. Frequency and reasons for missing data by each participant.

| Participant | Number (%) of missing days | Reasons for missing data | | | | |
		Violation of experiment condition	Acute illness	Participant forgot to wear device	Hospital operation	Holiday
1	6 (9.7%)				6	
2	14 (22.6%)	6	7	1		
3	11 (17.7%)		1	5		5
4	4 (6.5%)	1	2	1		
6	20 (32.3%)	17	2	1		
8	6 (9.7%)	6				
9	13 (21.0%)	8		5		
10	15 (24.2%)		13	2		
Total	89 (17.9%)	38	25	15	6	5

Notes: We defined a day as missing if the step count was lower than 100 steps. Data were considered unacceptable due to violation of experiment condition for one of two reasons: (1) when step counts were more than four times higher than the average steps per day for a given participant, as this indicated that the pedometers had been influenced (shaken) by the participant, which was acknowledged when asked; (2) when participants removed the seal and viewed the step count on the pedometer device on days when they were allocated to the active control (versus self-monitoring) and had acknowledged that they did this.

Missing cases

Across all the participants, 89 cases (data for one day) out of 496 were missing, the majority of which had a record of zero ($n = 78$) due to either violation of experiment condition or the participant not wearing the pedometer on the study day due to illness / forgetting (see Table 1). A further 11 cases were deemed missing due to a record of below 100 steps that was deemed as an invalid entry as this would equate to less than 1 min's walking activity (Tudor-Locke et al., 2011). Missing cases were input using Amelia II software (AmeliaView 1.7.2, http://gking.harvard.edu/amelia) that uses a bootstrapping-based algorithm to create multiple imputations to fill the missing cases. This approach is sensitive to time-series designs and is considered superior to other approaches in dealing with missing data such as listwise deletion and mean substitution (Honaker & King, 2010). Amelia II generates a default five data-sets with each case imputed in accordance with the uncertainty of variation in imputed values. As recommended, rather than combining them into one data-set for analysis, the five data-sets were analysed separately and then the results were combined (Honaker & King, 2010). After imputation, there were five data-sets with 62 days data for each participant, the results from which were combined and represented 62 data points per participant. Across the participants, the combined analyses had 245 and 251 days, respectively, for the goal-setting vs. control comparison, and 246 and 250 days, respectively, for the self-monitoring vs. control comparison.

Autocorrelation

The 40 data-sets (5 per participant) were then analysed separately. Before conducting the inferential analyses, each data-set was first checked for autocorrelation. Autocorrelation is likely to be present in time series data-sets as each participant's daily step count could

be correlated with previous or later step counts recorded during the study period. Where significant autocorrelations were identified, we accounted for them using pre-whitening, a standard procedure described previously (Hobbs et al., 2013; Quinn et al., 2013). The pre-whitening procedure removes autocorrelation from the data to satisfy the assumption of parametric tests that each day's data should be independent (and not correlated) to allow for further analysis. Where required, this was conducted in each instance on the most significant lag identified up to a maximum of seven lags (one week), as conducted previously, because removal of the largest auto-correlation relationship is most likely to successfully pre-whiten the data (Hobbs et al., 2013). For example, if lag1 had the largest autocorrelation, this was pre-whitened to account for first-order correlations (that step counts are correlated with step counts recorded the day before). Pre-whitening was performed by creating a time series version of the step count dependent variable for the most significant lag (e.g. lag1). The new lagged variable was then regressed onto the original variable and the unstandardised residuals were saved as the new pre-whitened dependent variable to be used for the analysis. Where pre-whitening was conducted, the new pre-whitened variable was then re-checked for the presence of significant autocorrelation. Pre-whitening successfully removed autocorrelation for every data-set except for three imputations for participant 1 (see footnote, Table 3).

Linear regressions

For each time series in each participant, the outcome variable of step count (original or pre-whitened as applicable) was entered into a linear regression as a function of intervention condition using dummy coding (1 = intervention day and 0 = active control day). Analyses were conducted separately for goal-setting (vs. control) and self-monitoring (vs. control) as the study was not powered to detect interaction effects (as was the case in the previous study by Sniehotta et al., 2012). Each linear regression was checked for meeting parametric assumptions by examining the residuals. Specifically, Durbin-Watson tests were conducted and all the values were close to 2 (and not below 1 or above 3), indicating the residuals were free from autocorrelation. Histograms of the standardised residuals were inspected and they all indicated patterns representing normal distributions. While on occasion one or two outliers were present, the results of tests to identify influential cases indicated that each outlier had no significant influence on the results and so were retained in the analyses (with all Cook's distance values below 1, all leverage values close to the average and all Mahalanobis distances below 2). Once each analysis was performed, the results were combined as described previously (Hobbs et al., 2013; Quinn et al., 2013). The mean values for B and p are presented, and the average standard error is presented, which was multiplied by a factor that corrects for bias. Cohen's d effect sizes were calculated by comparing the mean difference in step counts divided by the pooled standard error. In addition, we conducted linear regressions to test for linear time trends that could indicate possible carryover effects of the interventions. For each time series in each participant, the original outcome variable of step count (not pre-whitened) was entered into a linear regression as a function of cumulative day (from 1 through to 62) and the results combined in the same way described above. While the majority of Durbin-Watson values were close to 2 (and not below 1 or above 3) indicating the residuals were free from autocorrelation, some values were close to 1 and so should be interpreted cautiously for participants 1, 2 and 6.

Examination of consistency of effect

Finally, we examined the consistency of the effect of the interventions both across all the participants and then for each individual. For examination across all the participants, for each intervention, we calculated the binomial probability of the number of participants who had a result in favour of the intervention compared with the control. For both goal-setting and self-monitoring interventions, this was calculated from the probability of success occurring (.5, i.e. result in favour or not in favour of the intervention), the number of successes and the number of trials in the experiment (8, i.e. the total number of participants) using a freely available calculator (Soper, 2015). We then examined the descriptive statistics for each participant to observe the direction of effects across the two interventions (i.e. if goal-setting led to an increase in step counts for a given participant, we examined whether self-monitoring also led to an increase in step counts for this participant).

Results

The median step count was 5409 (interquartile range = 6265), which equates to an average of approximately 54 min walking per day (Tudor-Locke et al., 2011). Descriptive statistics for the step counts are presented in Table 2 as a function of intervention condition.

Goal-setting

Overall, compared to the control, the intervention led to a median increase of 922 steps per day. Four of the eight participants had higher median step counts in the goal-setting intervention condition compared to the control days. One participant had a median difference of over 1000 step counts from the goal-setting intervention, with an increase of 1663 steps (participant 9), equivalent to over 15 min extra walking per day (Tudor-Locke et al., 2011). Three participants had moderate increases in walking from

Table 2. Descriptive statistics between and within participants.

		Median (IR) steps			
		Goal-setting condition		Self-monitoring condition	
Participant	Age, gender	Intervention	Control	Intervention	Control
Overall		5781 (6204)	4859 (6363)	5912 (6700)	4962 (5809)
1	87, F	3366 (1916)	2878 (1294)	3178 (1842)	2921 (1044)
2	82, M	2581 (1982)	2572 (5395)	2678 (2741)	2364 (3053)
3	79, F	2350 (1646)	2032 (3776)	2392 (2848)	2082 (2338)
4	66, M	6511 (7494)	6832 (5612)	6823 (8605)	6511 (5727)
6	68, M	9381 (2977)	9921 (2570)	9548 (3200)	9429 (2544)
8	68, F	12,305 (5193)	12,977 (7856)	11,631 (6576)	12,956 (6744)
9	64, F	6047 (4352)	4384 (4394)	5396 (5214)	4416 (4808)
10	60, F	7185 (2364)	6626 (3260)	7210 (2646)	6629 (3257)

Notes: IR = interquartile range; F = female; M = male. The descriptive statistics presented are from the data before imputation of missing cases.

goal-setting of 559 (participant 10), 488 (participant 1) and 318 steps (participant 3). However, compared to the control days, for one participant, goal-setting led to an increase of only 9 steps (participant 2), and three participants had lower step counts in the goal-setting condition, with median differences of −672 (participant 8), −540 (participant 6), and −321 steps (participant 4). However, none of the combined results from the linear regression analyses were significant (see Table 3). A non-significant and small effect was observed in favour of the intervention overall ($d = .08$, 95% CI = −.42 to .58), and non-significant and small to small-medium effects were found both in favour of the intervention and the active control group for the participants individually (see Table 3).

Self-monitoring

Overall, compared to the control, the intervention led to a median increase of 950 steps per day. Seven of the eight participants had higher median step counts in the self-monitoring intervention condition compared to the control days. One participant had a median difference of just under 1000 step counts from the self-monitoring intervention, with an increase of 980 steps (participant 9), equivalent to almost 10 min extra walking per day (Tudor-Locke et al., 2011). Six participants had moderate increases in walking from self-monitoring of 581 (participant 10), 314 (participant 2), 312 (participant 4), 310 (participant 3), 257 (participant 1) and 119 steps (participant 6). However, compared to the control days, one participant had a lower step count in the intervention condition, with a median difference of −1325 steps (participant 8). However, again, none of the combined results from the linear regression analyses were significant (see Table 3). A non-significant and small effect was observed in favour of the intervention overall ($d = .11$, 95% CI = −.39 to .61), and non-significant and small to small-medium effects were found both in favour of the intervention and the active control group for the participants individually (see Table 3).

Linear time trend

For six participants, the combined results from the linear regressions showed no significant time trend for an increase (participants 1, 3, 6 and 10) or decrease (participants 2 and 9) in steps over the course of the study period. However, for participants 4 and 8, the combined results from the linear regression analyses showed a significant but small decrease in steps over time (see Table 3).

Consistency of effect

The probability of goal-setting to lead to an increase in step counts in four out of eight participants was 1/3.66 and not significant ($p = .27$). However, the probability of self-monitoring to lead to an increase in step counts in seven out of eight participants was 1/32 and significant ($p = .03$). There was consistency in the effect of the interventions across five of the eight participants: participant 9 had the highest increase in walking from both the goal-setting and self-monitoring interventions at around 1000 steps or more. Participants 1, 3 and 10 had a moderate increase in walking from both interventions, and participant 8 had a lower step count from both interventions. However, for

Table 3. Main effects by each participant.

Participant	Goal-setting					Self-monitoring					Linear time trend			
	B	SE	p	95% CI	d (95% CI)	B	SE	p	95% CI	d (95% CI)	B	SE	p	95% CI
1	369	611	.53	−828, 1567	.16 (−.34, .66)	419	671	.51	−897, 1735	.17 (−.33, .67)	31	21	.08	−9, 71
2	−259	982	.66	−2183, 1666	−.32 (−.82, .18)	837	904	.35	−936, 2609	.43 (−.08, .93)	−10	30	.54	−69, 49
3	566	1051	.58	−1495, 2626	−.14 (−.64, .36)	758	940	.41	−1084, 2599	.47 (−.04, .97)	30	32	.31	−34, 93
4	683	1318	.61	−1900, 3265	.14 (−.36, .64)	1290	1333	.33	−1322, 3902	.36 (−.15, .86)	−86	36	.02	−156, −16
6	−263	1035	.46	−2291, 1766	.02 (−.47, .52)	574	937	.43	−1262, 2411	−.07 (−.57, .43)	4	24	.72	−44, 51
8	329	1179	.75	−1981, 2640	−.01 (−.50, .49)	126	1177	.78	−2181, 2432	−.32 (−.82, .18)	−82	32	.01	−144, −20
9	−362	823	.66	−1975, 1251	.34 (−.17, .84)	−191	1013	.61	−2176, 1794	.33 (−.17, .83)	−5	27	.61	−57, 48
10	768	788	.31	−777, 2312	.38 (−.13, .87)	1055	739	.15	−393, 2503	.32 (−.18, .82)	9	23	.54	−37, 55

Notes: Linear time trend was analysed using the original values with missing cases imputed because pre-whitening would remove the effect being tested, while goal-setting and self-monitoring were analysed using the pre-whitened values with missing cases imputed. For the goal-setting and self-monitoring analyses, pre-whitening was required twice for participant two (at lag1 for imputations 2 and 5), once for participant three (at lag1 for imputation 1), three times for participant 6 (at lag1 for imputations 3–5), once for participant 8 (at lag2 for imputation 4), once for participant 9 (at lag4 for imputation 1) and twice for participant 10 (at lag1 for imputation 2 and lag5 for imputation 3). For participant 1, pre-whitening was required in four instances (at lag1 for imputations 2–5). However, in three instances, pre-whitening did not successfully remove significant autocorrelation (imputations 2 and 4–5). Therefore, we conducted a sensitivity analysis whereby the combined results of including all five imputations (one with no autocorrelation and four pre-whitened at lag1) were compared with only including the two imputations where pre-whitening was successfully removed (one with no autocorrelation and one pre-whitened at lag1). In the table, we report the results with all five imputations combined. When only the two imputations that had auto-correlation successfully removed with pre-whitening were included, the combined results were weaker for both goal-setting (B = 156, SE = 568, p = .74, 95% CI = −958, 1270) and self-monitoring (B = 111, SE = 528, p = .78, 95% CI = −925, 1146). The effect sizes (Cohen's d) were calculated using the means and standard deviations of the original data before imputation of missing cases.

participants 2, 4 and 6, while they all had a moderate increase in walking from the self-monitoring intervention, they had either no increase in walking (participant 2) or lower step counts from the goal-setting intervention (participants 4 and 6).

Discussion

This study reported the first test of BCTs among older people with the use of *N*-of-1 factorial RCTs. Overall, goal-setting and self-monitoring both increased step counts by an average of over 900 steps per day, which equates to approximately over nine minutes extra daily walking. For goal-setting, there was a trend for an increase in step counts for four of the eight participants; however, the probability of this result occurring was not significant. In the single case analyses, step counts were on average over 1600 higher on goal-setting days in one participant; however, goal-setting had no effect for one participant and a negative effect on walking for three others. For self-monitoring, there was a trend for an increase in step counts for seven of the eight participants, of which the probability of this result occurring was significant. In the single case analyses, step counts were on average just under 1000 step counts higher on self-monitoring days in one participant; however, self-monitoring had a negative effect on walking for one participant. From the combined results of the linear regressions, no significant difference was found for either goal-setting or self-monitoring BCTs to increase walking among the eight individuals in this study. We also found a small significant decrease in walking over time in two participants.

In comparison with the previous factorial *N*-of-1 RCTs with almost exclusively young and middle-aged adults (Sniehotta et al., 2012), a similar pattern of results from the descriptive statistics were observed at the overall and individual levels in favour of both the goal-setting and self-monitoring interventions compared to control days. This would suggest that these two self-regulation BCTs may be of similar utility (with a small to small-medium effect) when employed with the general healthy adult population regardless of age. In addition, in the current study there was a trend for consistency of effect of both goal-setting and self-monitoring interventions across participants, with a stronger effect observed for self-monitoring. However, the trends observed in our study were not supported by significant single case analyses, and Sniehotta et al. (2012) did not find consistency in the effect of the BCTs in their study. While the use of different BCTs may produce interaction effects and demonstrate additive benefit of combining techniques, *N*-of-1 RCTs can identify inconsistencies in intervention effects within and between participants that is obscured in conventional RCT designs, and highlight that not every behaviour change technique works for every participant. It is also of note that in the Sniehotta et al.'s study four participants showed small significant linear time trend effects and we identified a similar effect in two participants. While the direction was positive in the previous study and negative in our study, these findings suggest that BCTs have potential for carryover effects across conditions (whether they lead to an increase or decrease in the target behaviour) and future research may adopt more sophisticated trial designs to better account for the strength of such effects (Sniehotta et al., 2012).

There were three points of divergence between the current and prior factorial *N*-of-1 RCT study (Sniehotta et al., 2012). First, participants in the current study did not wish to be notified of their daily experimental condition by text message, but instead agreed

to be informed by opening a sealed envelope each morning. This may be of surprise given that the independent regulator and competition authority for the UK communications industries has reported that 94% of adults in the UK own or use a mobile phone (Ofcom, 2013). However, the same regulator recently reported that 38% of adults aged 65+ were not using a mobile phone (Ofcom, 2011), and identified several barriers to the industry for better addressing the needs of older people (Freeman & Lessiter, 2009). While our sample indicated a lack of confidence in daily use of mobile phone technology (some did not own a mobile phone and others who did were not familiar with using text messages, and so interest in the study was low when mobile phone use was a requirement), many other older people are likely to have access to and be proficient in the use of technology such as smartphones that have other applications that could be used to automatically collect data for N-of-1 RCTs (e.g. accelerometers to measure physical activity). Future research could compare different mediums and frequency of contacts in which experimental conditions are communicated to participants, data are collected from them and different ways of communicating BCT instructions. For example, text messaging could be used as an alert or reminder service, which has been identified as a powerful behaviour change tool in isolation and when used to supplement website-based interventions (Orr & King, 2015; Webb, Joseph, Yardley, & Michie, 2010).

A second point of divergence with the previous N-of-1 RCT study (Sniehotta et al., 2012) was that participants in this study completed on average over 2800 step counts fewer compared to the younger participants assessed by Sniehotta et al. This would be expected given that the sample recruited for the current study was older by an average of 35 years. In our study, participants were requested to set their own step count goals, but this difference in step counts achieved highlights the need to tailor goal-setting to the abilities and aspirations of the individual should goals be co-created by both the participant and interventionist. Future studies could examine in more detail the physical activity of individuals and make more fine-grained comparisons between age groups. For example, accelerometers could be used to ascertain the intensity and duration of bouts of physical activity completed between intervention and control conditions across age groups. In addition, with the use of smart homes and global positioning system technology, physical activity data collected on smartphones could be wirelessly connected with rooms around the house and points in the local area to indicate where and when people of different ages are most physically active (e.g. Demiris & Thompson, 2012; Fay, 2014). This would afford dynamic integration of environmental factors that are known to predict physical activity, such as access to services in the local neighbourhood (Eisenstein et al., 2011).

A third point of divergence with the previous N-of-1 RCT study (Sniehotta et al., 2012) was that the current study did not obtain statistically significant differences for the eight participants despite having a similar number of data points per participant (our study had two extra days per participant). For the one participant aged 67 in Sniehotta et al.'s study, this participant had a higher number of step counts in the goal-setting (+297 steps) and self-monitoring conditions (+1453 steps), of which the latter finding was significant. However, we reported lower median step counts, and so the proportion of within-participant variance in relation to median step counts was far higher in our study. This was reflected in lower B values in relation to the SE values for all bar one of the participants (see Table 3). Despite some significant findings, Sniehotta et al. observed

a lack of statistical power in their study and the current study was similarly limited due to high levels of within-participant variance in walking per day. A further limitation of our study was a high level of missing data across participants, which may have reduced power to detect significant differences given we had to rely on imputed data to replace missing cases (whereas Sniehotta et al., 2012 were able to use complete data from each participant). Our results highlight that the effect size of interventions and resultant power varies between participants in *N*-of-1 trials. In our study, non-significant effect sizes ranged −.01 to .38 for goal-setting and −.07 to .47 for self-monitoring, which suggests that a larger volume of data points to increase power may have led to detection of significant differences in some participants. For other participants, the interventions appeared ineffective. While future studies could collect a larger volume of data per participant to increase statistical power, this will increase participant burden and potentially lead to greater attrition levels, and this strategy will only be effective for interventions with above small effect sizes.

Implications for practice

Based on the findings of the current study, we would recommend that interventionists make use of *N*-of-1 RCTs with older people to identify which BCTs might be most effective to increase individuals' physical activity. While further research is required, we found trends that would support the use of goal-setting and self-monitoring BCTs. Other techniques not tested in this study could also be used (Michie et al., 2013). *N*-of-1 RCTs could also be of use for explaining complex patterns of change in patient health behaviours, as recommended in a recent study that found several factors to predict maintenance of physical activity (Hekler et al., 2013). As for implementation of BCTs, the provision of training and instrumentation for staff has been highlighted to overcome barriers to the use of goal-setting in shared clinical decision-making (Schulman-Green, Naik, Bradley, McCorkle, & Bogardus, 2006).

One application of *N*-of-1 RCTs that may be of particular benefit in interventions with older people is facilitating the tailoring of BCTs (Cushing, Walters, & Hoffman, 2014). The aim of tailoring a health intervention is to make it more personally relevant to the individual and thereby making it more effective (Kreuter, Farrell, Olevitch, & Brennan, 2000). For example, a study demonstrated that a tailored version of website-based information was more persuasive in promoting physical activity among older people compared with a generic control version (Nyman & Yardley, 2009). The use of *N*-of-1 trials offers a whole new layer of possibility for tailoring interventions in that individuals could be better matched to BCTs if *N*-of-1 data were obtained alongside or prior to implementation of a behaviour change programme (Sniehotta et al., 2012). For example, a range of techniques could be used in *N*-of-1 RCTs to determine which technique(s) work particularly well for a particular individual.

Limitations of the current study and suggestions for future research

Given the small sample and short study period, it was not possible to identify variables that predict which participants would benefit from the interventions and which would not. Future studies could be conducted that combine within- and between-groups

approaches to explore the variables that may predict positive responses to interventions with the use of multilevel modelling (Curran & Bauer, 2011; Johnston & Johnston, 2013). For example, a previous study with young adults found that within-person but not between-person variance in daily physical activity was significantly associated with satisfaction with life (Maher et al., 2013). Another study with university students found that daily time in the pursuit of goals that conflicted with their regular physical activity negatively predicted objectively measured physical activity (Presseau, Tait, Johnston, Francis, & Sniehotta, 2013). In addition, with *N*-of-1 trials one cannot investigate beliefs that will have carryover effects. For example, it would not be feasible to manipulate outcome expectancy beliefs within an *N*-of-1 RCT, which have been shown to vary weekly in the first four weeks of initiation of physical activity in a sample of previously inactive adults (Loehr, Baldwin, Rosenfield, & Smits, 2014). However, *N*-of-1 designs are more suited to behavioural techniques such as those tested in the current study and could be combined with other interventions for example through stepped-wedge designs (Johnston & Johnston, 2013). Also, further research could screen larger samples to purposively select participants into *N*-of-1 RCTs to produce samples that are homogenous or stratified on key variables such as high body mass index to identify processes of behaviour change that are effective for specific patient groups.

Conclusion

This study demonstrated the utility of using an *N*-of-1 design to test the effect of BCTs among older people. While no combined linear regressions were significant, trends were observed that would suggest the self-regulation interventions were effective for some individuals but not all. Self-monitoring was also found to have a more consistent effect across participants. With advances in technology for monitoring and providing feedback to individuals based on their activity levels, there is scope for *N*-of-1 designs to play an important role in understanding how best to increase physical activity among older people with the use of BCTs.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Samuel R. Nyman http://orcid.org/0000-0003-1963-0814
Dominika Kwasnicka http://orcid.org/0000-0002-5961-837X
Andrew Callaway http://orcid.org/0000-0003-0987-9608

References

Australian Sports Commission. (2000). *Physiological tests for elite athletes* (1st ed.). Leeds: Human Kinetics.

Baker, G., Gray, S. R., Wright, A., Fitzsimons, C., Nimmo, M., Lowry, R., & Mutrie, N. (2008). The effect of a pedometer-based community walking intervention "Walking for Wellbeing in the West" on physical activity levels and health outcomes: A 12-week randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity, 5*, e44.

Barlow, D. H., Knock, M. K., & Hersen, M. (2008). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). London: Pearson.

Barton, S. (2000). Which clinical studies provide the best evidence? *British Medical Journal, 321*, 255–256.

Bird, E. L., Baker, G., Mutrie, N., Ogilvie, D., Sahlqvist, S., & Powell, J. (2013). Behavior change techniques used to promote walking and cycling: A systematic review. *Health Psychology, 32*, 829–838.

Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., ... Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health: A systematic review. *Journal of the American Medical Association, 298*, 2296–2304.

Canada Fitness Survey. (1983). *Fitness and lifestyle in Canada*. Ottawa, ON: Directorate of Fitness and Amateur Sport.

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality-social, clinical, and health psychology. *Psychological Bulletin, 92*, 111–135.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Conn, V. S., Hafdahl, A. R., Brown, S. A., & Brown, L. M. (2008). Meta-analysis of patient education interventions to increase physical activity among chronically ill adults. *Patient Education & Counseling, 70*, 157–172.

Craig, P., Dieppe, M., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *British Medical Journal, 337*. Published online 29 September, doi:10.1136/bmj.a1655

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583–619.

Cushing, C. C., Walters, R. W., & Hoffman, L. (2014). Aggregated N-of-1 randomized controlled trials: Modern data analytics applied to a clinically valid method of intervention effectiveness. *Journal of Pediatric Psychology, 39*, 138–150.

Demiris, G., & Thompson, H. J. (2012). Mobilizing older adults: Harnessing the potential of smart home technologies. Contribution of the IMIA working group on smart homes and ambient assisted living. *Yearbook of Medical Informatics, 7*, 94–99.

Department of Health, Physical Activity, Health Improvement, and Protection. (2011). *Start active, stay active: A report on physical activity for health from the four home countries' chief medical officers*. London: Author.

Dombrowski, S. U., Sniehotta, F. F., Avenell, A., Johnston, M., MacLennan, G., & Araújo-Soares, V. (2012). Identifying active ingredients in complex behavioural interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: A systematic review. *Health Psychology Review, 6*, 7–32.

Eisenstein, A. R., Prohaska, T. R., Kruger, J., Satariano, W. A., Hooker, S., Buchner, D., ... Hunter, R. H. (2011). Environmental correlates of overweight and obesity in community residing older adults. *Journal of Aging and Health, 23*, 994–1009.

Fay, D. (2014). A method of predicting arrival times for operating a piece of equipment. *UK Preliminary Patent Application number 1305264.2*. Cork: University College Cork.

Fitzsimons, C. F., Baker, G., Gray, S. R., Nimmo, M. A., & Mutrie, N. (2012). Does physical activity counselling enhance the effects of a pedometer-based intervention over the long-term: 12-month findings from the Walking for Wellbeing in the west study. *BMC Public Health, 12*, e206.

Freeman, J., & Lessiter, J. (2009). *Exploring how manufacturers, suppliers and retailers address the needs of older and disabled people: What are the barriers and drivers?* London: Ofcom.

Gabler, N. B., Duan, N., Vohra, S., & Kravitz, R. L. (2011). N-of-1 trials in the medical literature: A systematic review. *Medical Care, 49*, 761–768.

Guyatt, G., Sackett, D., Taylor, D. W., Chong, J., Roberts, R., & Pugsley, S. (1986). Determining optimal therapy – Randomized trials in individual patients. *New England Journal of Medicine, 314*, 889–892.

Hekler, E. B., Buman, M. P., Poothakandiyil, N., Rivera, D. E., Dzierzewski, J. M., Morgan, A. A., ... Giacobbi, P. R. J. (2013). Exploring behavioral markers of long-term physical activity maintenance: A case study of system identification modeling within a behavioral intervention. *Health Education & Behavior, 40*, 51S–62S.

Hobbs, N., Dixon, D., Johnston, M., & Howie, K. (2013). Can the theory of planned behaviour predict the physical activity behaviour of individuals? *Psychology & Health, 28*, 234–249.

Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science, 54*, 561–581.

Johnston, D. W., & Johnston, M. (2013). Useful theories should apply to individuals. *British Journal of Health Psychology, 18*, 469–473.

Kreuter, M. W., Farrell, D., Olevitch, L., & Brennan, L. (2000). *Tailoring health messages: Customizing communication with computer technology.* Mahwah, NJ: Lawrence Erlbaum Associates.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *American Psychologist, 57*, 705–717.

Loehr, V. G., Baldwin, A. S., Rosenfield, D., & Smits, J. A. J. (2014). Weekly variability in outcome expectations: Examining associations with related physical activity experiences during physical activity initiation. *Journal of Health Psychology, 19*, 1309–1319.

Maher, J. P., Doerksen, S. E., Elavsky, S., Hyde, A. L., Pincus, A. L., Ram, N., & Conroy, D. E. (2013). A daily analysis of physical activity and satisfaction with life in emerging adults. *Health Psychology, 32*, 647–656.

Michie, S., Abraham, C., Whittington, C., McAteer, J., & Gupta, S. (2009). Effective techniques in healthy eating and physical activity interventions: A meta-regression. *Health Psychology, 28*, 690–701.

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., ... Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine, 46*, 81–95.

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*, 112–117.

Morris, J. N., & Hardman, A. E. (1997). Walking to health. *Sports Medicine, 23*, 306–332.

Nathan, A., Wood, L., & Giles-Corti, B. (2014). Examining correlates of self-reported and objectively measured physical activity among retirement village residents. *Australasian Journal on Ageing, 33*, 250–256.

National Institute for Health and Clinical Excellence. (2006). *Four commonly used methods to increase physical activity: Brief interventions in primary care, exercise referral schemes, pedometers and community-based exercise programmes for walking and cycling.* London: Author.

National Institute for Health and Clinical Excellence. (2007). *Behaviour change at population, community and individual levels (NICE public health guidance 6)*. London: Author.

Nyman, S. R., & Yardley, L. (2009). Web-site-based tailored advice to promote strength and balance training: An experimental evaluation. *Journal of Aging and Physical Activity, 17*, 210–222.

Ofcom. (2011). *UK adults' media literacy*. London: Ofcom.

Ofcom. (2013). *Communications market report 2013*. London: Ofcom.

Ogilvie, D., Foster, C. E., Rothnie, H., Cavill, N., Hamilton, V., Fitzsimons, C. F., & Mutrie, N. (2007). Interventions to promote walking: Systematic review. *British Medical Jornal, 334*, e1204.

Orr, J. A., & King, R. J. (2015). Mobile phone SMS messages can enhance healthy behaviour: A meta-analysis of randomised controlled trials. *Health Psychology Review*. Published online 5 March, doi:10.1080/17437199.2015.1022847

Presseau, J., Tait, R. I., Johnston, D. W., Francis, J. J., & Sniehotta, F. F. (2013). Goal conflict and goal facilitation as predictors of daily accelerometer-assessed physical activity. *Health Psychology, 32*, 1179–1187.

Price, J. D., & Grimley Evans, J. (2002). N-of-1 randomized controlled trials ('N-of-1 trials'): Singularly useful in geriatric medicine. *Age and Ageing, 31*, 227–232.

Quinn, F., Johnston, M., & Johnston, D. W. (2013). Testing an integrated behavioural and biomedical model of disability in N-of-1 studies with chronic pain. *Psychology & Health, 28*, 1391–1406.

Renner, B., Spivak, Y., Kwon, S., & Schwarzer, R. (2007). Does age make a difference? Predicting physical activity of South Koreans. *Psychology and Aging, 22*, 482–493.

Scholes, S., & Mindell, J. (2013). Physical activity in adults. In R. Craig & J. Mindell (Eds.), *Health survey for England – 2012* (pp. 1–49). London: The Health and Social Care Information Centre.

Scholz, U., Sniehotta, F. F., Burkert, S., & Schwarzer, R. (2007). Increasing physical exercise levels: Age-specific benefits of planning. *Journal of Aging and Health, 19*, 851–866.

Schulman-Green, D. J., Naik, A. D., Bradley, E. H., McCorkle, R., & Bogardus, S. T. (2006). Goal setting as a shared decision making strategy among clinicians and their older patients. *Patient Education and Counseling, 63*, 145–151.

Scuffham, P. A., Nikles, J., Mitchell, G. K., Yelland, M. J., Vine, N., Poulos, C. J., ... Glasziou, P. (2010). Using N-of-1 trials to improve patient management and save costs. *Journal of General Internal Medicine, 25*, 906–913.

Scuffham, P. A., Yelland, M. J., Nikles, J., Pietrzak, E., & Wilkinson, D. (2008). Are N-of-1 trials an economically viable option to improve access to selected high cost medications? The Australian experience. *Value in Health, 11*, 97–109.

Skelton, D. A., Young, A., Walker, A., & Hoinville, E. (1999). *Physical activity in later life: Further analysis of the Allied Dunbar national fitness survey and the Health Education Authority national survey of activity and health*. London: Health Education Authority.

Sniehotta, F. F., Presseau, J., Hobbs, N., & Araújo-Soares, V. (2012). Testing self-regulation interventions to increase walking using factorial randomized N-of-1 trials. *Health Psychology, 31*, 733–737.

Soper, D. S. (2015). Binomial probability calculator [software]. Retrieved March 3, 2015, from http://www.danielsoper.com/statcalc

Steeves, J. A., Tyo, B. M., Connolly, C. P., Gregory, D. A., Stark, N. A., & Bassett, D. R. (2011). Validity and reliability of the Omron HJ-303 tri-axial accelerometer-based pedometer. *Journal of Physical Activity & Health, 8*, 1014–1020.

Sugden, J. A., Sniehotta, F. F., Donnan, P. T., Boyle, P., Johnston, D. W., & McMurdo, M. E. T. (2008). The feasibility of using pedometers and brief advice to increase activity in sedentary older women – A pilot study. *BMC Health Services Research, 8*, e169.

Tudor-Locke, C., Craig, C. L., Aoyagi, Y., Bell, R. C., Croteau, K. A., De Bourdeaudhuij, I., ... Blair, S. N. (2011). How many steps/day are enough? For older adults and special populations. *International Journal of Behavioral Nutrition and Physical Activity, 8*, e80.

Webb, T. L., Joseph, J., Yardley, L., & Michie, S. (2010). Using the Internet to promote health behavior change: A systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *Journal of Medical Internet Research, 12*, e4.