1   **Effectiveness of Structured Physical Activity (PA) Interventions**
2   **through the evaluation of PA levels, Adoption, Retention,**
3   **Maintenance and Adherence Rates – A Systematic Review and**
4   **Meta-Analysis**

5

6

7   Authors: Willinger, Nadja[1]; Steele, James[2,3]; Atkinson, Lou[4]; Liguori, Gary[5]; Jimenez, Alfonso[1,6];

8   Mann, Steve[7]; Horton, Elizabeth[1]

9   Affiliations: [1]Coventry University (UK); [2]ukactive Research Institute (London, UK); [3]Solent

10   University (Southampton, UK); [4]Aston University (Birmingham, UK); [5]University of Rhode Island,

11   Kingston, RI; [6]GO fit LAB, Ingesport (Madrid, Spain); [7]Places for People leisure (London, UK).

12

13   Email: Willinger, Nadja* (willingn@uni.coventry.ac.uk); Steele, James (james.steele@solent.ac.uk);

14   Atkinson, Lou (l.atkinson1@aston.ac.uk); Liguori, Gary (gliguori@uri.edu); Jimenez, Alfonso

15   (ab9935@coventry.ac.uk); Mann, Steve (stevemann@pfpleisure.org); Horton, Elizabeth

16   (apx242@coventry.ac.uk).

17   *Corresponding Author.

18

19

20   Keywords: Efficacy, Feasibility, Dropout, Structured Physical Activity Intervention, Behaviour

21   Change Techniques (BCTs).

22   Submission Date: May 2019

23   Manuscript type: Systematic review

24   Word Count (excluding title page and abstract): 8460

25   Abstract

26   Background: Structured physical activity (PA) interventions can potentially be implemented in a variety

27   of facilities, and therefore can reach a large proportion of the population. The effectiveness of

28   interventions is historically evaluated through examination of group differences in outcome measures.

29   Often the proportions of individuals meeting thresholds for PA outcomes related to intervention

30   implementation are not considered. Our aim was to summarise the effectiveness of structured

31   interventions both through reported group differences in outcomes, adoption and maintenance rates,

32   and adherence and retention rates, providing information on intervention feasibility.

33   Methods: Database screening resulted in the inclusion of 12 interventions.

34   Results: There was a tendency for structured programmes to result in a significantly greater increase in

35   PA levels than the control conditions in the short-term, with more varying results in the long-term. Only

36   3 studies published adoption and maintenance rates. On average $67\pm16\%$ of participants were reported

37   as adopting PA, with only $29\pm13\%$ maintaining this effect. A mean retention rate of $75\pm13\%$ was

38   observed, and $61\pm21\%$ of intervention sessions were attended as described through adherence rates.

39   Conclusion: Structured interventions were classified as overall effective in short-term on the basis of

40   group differences in PA levels; however, adoption and maintenance rates were rarely reported.

41

42   Abstract word count: 197 words

Background

43

44 Physical activity (PA) levels in the UK are low, with almost 50% of the population currently being

45 classified as inactive or insufficiently active, despite a multitude of interventions aiming to increase PA

46 levels[1]. Many of these can be classified as structured interventions as they provide a clear

47 recommendation on the frequency of attendance of pre-planned exercise sessions, amongst other

48 specified components, thus allowing for standardisation and replication[2]. Structured programmes, if

49 proven effective in increasing PA, can potentially be implemented in a large variety of facilities in the

50 private and public health sector, therefore contributing to the global target of reducing inactivity by

51 10% by 2025 as defined through the WHO[3]. Thus, it could be said that this population target is based

52 around concepts of adoption and maintenance of PA.

53 Adoption and maintenance are frequently used terms, but their definition and measurement in academic

54 literature vary widely[4-6]. Generally, adoption refers to an individual's uptake of PA at recommended

55 levels, whereas maintenance is described as long-term behaviour change[5,6]. Retention on the other hand

56 refers to the number of participants in a study available for follow-up assessments, while adherence

57 rates describe the degree to which participants take part in intervention sessions[7,8]. Both retention and

58 adherence are indicators for intervention feasibility. Many structured interventions to increase PA can

59 potentially be delivered through leisure facilities and as such a further outcome of importance relating

60 to these behaviours may be the retention of those participants at the leisure centres.

61 An intervention's effectiveness is commonly assessed via the analysis of mean differences in outcome

62 measures (i.e. PA levels) between treatment groups typically through the use of null hypothesis

63 significance testing. However, the extent of its effectiveness is often only expressed as whether a

64 difference between groups meets a threshold for statistical significance. Meaningfulness of this

65 difference is often not considered and arguably this difference better reflects the *efficacy* of an

66 intervention anyway[9]. *Effectiveness* instead must consider both the changes in outcome measures in

67 addition to the implementation of interventions within the real world. We propose by reporting the

68 proportion of participants able to achieve and maintain the respectively recommended activity levels,

69 one cannot only get insights on the effectiveness of the intervention but also, in combination with mean

70 changes or differences in outcomes, assess its clinical relevance. Being physically active for 75 to 150

71 minutes at vigorous or moderate intensity respectively is associated with a multitude of health benefits

72 and analysing the proportion of participants meeting this threshold can be used to infer likelihood that

73 on national and global levels behaviour change might occur and ultimately result in a reduction of

74 healthcare costs and burden of disease associated with inactivity[10-13].

75 Reporting adoption and maintenance rates of PA enables readers to compare treatment effects and

76 implementation characteristics between treatment groups and offers more depth than by merely

77 comparing means and measures of variability (i.e. standard deviations). Though of course these sample

78 group level effects may be representative of population effects where studies are adequately powered,

79 and indeed responder counts may merely reflect group mean differences[14], group level effects only

80 provide limited evidence on the overall effectiveness of an intervention's implementation, particularly

81 for a behavioural outcome such as PA whereby interventions are designed with the intention of

82 increasing PA to meet a particular threshold. Often the proportions of individuals meeting thresholds

83 for outcomes such as PA levels over time are not considered (i.e. adoption, maintenance, retention, or

84 adherence). As such the group level effects seen may not be accurate reflections of the interventions

85 efficacy when implemented in real world settings.

86 This review seeks to address the issue of underreporting of the effectiveness of interventions by

87 summarising the adoption, adherence, retention and maintenance rates of structured PA programmes

88 targeting non-clinical adults. Therefore, the primary aim is to assess, analyse and draw conclusions

89 about the effectiveness of structured PA interventions and programmes in promoting the adoption and

90 maintenance of PA in an insufficiently active adult population. We further aim to draw conclusions on

91 intervention feasibility by summarising adherence and retention rates of participants. The following

92 research questions will be addressed: 1) What proportion of articles on structured PA interventions have

93 evaluated and reported intervention effects for adoption and maintenance?; 2) What are the adoption

94 and maintenance rates achieved by structured exercise programmes aiming to promote an increase in

95 PA levels of insufficiently active adults?; 3) Are structured interventions effective in increasing PA

96 levels compared to a control group or another treatment group? ; and 4) What are the retention and

97 adherence rates of participants of structured interventions and which conclusions on intervention

98 feasibility can therefore be made?.

99 Methods

100 The protocol for this review can be found under the PROSPERO registration number

101 CRD42017061009, and therefore will only be outlined briefly in the following. We followed the

102 PRISMA and AMSTAR 2 guidelines for the conduction and description of this review[15,16].

103 Literature Search Strategy

104 A combined search of 5 EBSCO databases (MEDLINE, SPORTDiscus, PsycINFO, CINAHL,

105 Academic Search Complete) was carried out in addition to separate searches through Scopus and the

106 Cochrane library from the earliest available date until April 2019. The search strategy can be viewed in

107 full on PROSPERO. Terms related to 'adoption', 'intervention', 'physical activity', 'maintenance',

108 'retention' and 'adults' were combined to identify relevant articles. Terms related to 'nutrition',

109 'workplace', 'mass media' and 'children' were excluded from the search.

110 Reference lists of eligible articles were screened for relevant articles. In addition, a separate search of

111 grey literature in the form of evaluation reports of PA programmes in the UK was carried out.

112 Inclusion Criteria

113 The studies and reports had to meet the inclusion criteria detailed below. Only articles published in

114 English language were included. In order to widen the evidence, a pragmatic approach was taken, and

115 this review also includes study designs other than randomised controlled trials e.g. prospective cohort

116 studies. This is taken into account when interpreting interventions results through the assessment of

117 methodological quality of included articles through an evaluation of risk of bias according to Cochrane

118 Guidelines, as detailed in the following paragraphs. Inclusion criteria were as follows: 1) The mean age

119 of participants lay between 18-64 years; 2) Participants are characterised by an insufficiently active

120 lifestyle at baseline, defined through activity levels of less than 150 minutes of moderate to vigorous

121 intensity PA per week; 3) The population sampled was healthy adults independent of their weight status;

122 4) The design is described as a randomised control trial, quasi-experimental trial, or pre- and post-

123   intervention i.e. prospective cohort study; 5) The intervention group participated in a structured PA

124   programme, characterised through a recommendation of a defined amount of PA per week achieved

125   through the provision of exercise sessions; 6) Participants were observed for at least 6 months; 7) The

126   primary aim of the intervention(s) was to increase PA levels; 8) The intervention(s) aimed to change

127   PA behaviour only (single behaviour change); and 9) The intervention(s) did not use mass media and

128   were not described as a home-based or lifestyle intervention.

129   Data Extraction

130   Data were extracted via standardised forms and characteristics of included studies (e.g. type and

131   duration of the intervention(s), follow-up assessments, PA goal or recommendation, applied behaviour

132   change techniques) and study subjects (e.g. gender, age, body mass index (BMI)) were recorded.

133   Adoption and maintenance rates, and intervention effectiveness expressed as changes in PA levels were

134   defined as primary outcomes. Due to the lack of explicit definitions of the terms 'adoption' and

135   'maintenance', in this review they were defined as the proportion of participants reaching the respective

136   researcher's recommended PA levels yet with consideration of at what time point this was measured.

137   For the purpose of this review we considered maintenance as a long-term effect measured at least 6

138   months after adoption (i.e. the first follow-up assessment after the beginning of the intervention). In

139   addition to adoption and maintenance rates, we summarised intervention effectiveness in respect to the

140   reported difference in the magnitude of effects in influencing PA levels between the structured

141   intervention(s) and the control condition.

142   Within each of the studies the Behaviour Change techniques (BCTs) applied in each intervention were

143   extracted and coded. Two members of the research team independently coded the BCTs on the base of

144   the description of interventions published in each article using the taxonomy of Michie et al. (2013) and

145   discrepancies were resolved through discussions[17].

146   Risk of bias

147   The Cochrane risk of bias tool was used to detect biases in random sequence generation, allocation

148   concealment, blinding of outcome assessors, attrition and reporting[18]. We did not assess bias for

149   blinding study personal and assessors for which intervention a participant received, due to the nature of

150   PA interventions.

151   For each study, every item was graded into 'high risk' or 'low risk' of bias, or 'unclear' where

152   description of methodology was insufficient. Studies were subjectively graded into high, medium or

153   low risk of bias, considering the types of bias and their severity in addition to the perceived overall

154   methodological complexity. If one item was evaluated as high risk in combination with unclear and low

155   risks for the remaining 4 categories the overall study quality was assumed to be medium, whereas more

156   than 2 categories graded as high risk resulted in a low overall quality of the study. Trials for which the

157   risk of random sequence generation has been classified as high were also be given an overall high risk

158   for bias, as this is a potentially strong confounding factor[18]. An overall rating of low risk of bias and

159   therefore high methodological quality was given to studies for which at least 3 categories were classified

160   as low risk in combination with an unclear risk of bias for the remaining 2 items. Any study for which

161   the risk assessment of at least 3 items was not possible due to insufficiently reported methods was be

162   given a moderate overall rating. No studies were excluded due to a poor rating of methodological

163   quality.

164   Results

165   A total of 20,659 articles were identified through database searching. After duplicates were removed

166   17,008 articles were checked for eligibility through title and abstract screening, resulting in 159 articles

167   analysed through full-text screening, of which two described the same study at different time points[19,20].

168   After the exclusion of 147 articles which did not meet inclusion criteria the reference lists of the

169   remaining 12 included articles were screened. This search yielded the identification of 5 additional

170   articles, of which one was included in this review, resulting in a final inclusion of 13 papers reporting

171   12 studies[19-31]. The screening process is outlined in the flow chart in Figure 1.

172   The search of grey literature yielded 8 evaluation reports summarising a multitude of PA programmes

173   in the UK. None of those met all of the inclusion criteria, therefore this information was not included

174   in this review.

Study characteristics

The characteristics of the 12 included studies are summarised in Table 1. Studies were published between 1982 and 2017 and most commonly conducted in the USA (50.0%). The design of the studies was predominantly described as a randomised controlled trial (RCT) (41.2%), and less than half of included articles were of a high methodological quality (n=5, 41.67%). Supplementary file 1 provides more information on the risk of bias assessment.

A total of 2116 participants of an average age of 49 years were recruited in the studies included in this review. For 4 studies (33.3%) there was no age reported. However, these articles were still included in this review as in the methods it was specified participants had to be of an age of 65 or below. Participants in 4 studies (33.3%) had a predominantly white background, whereas 3 studies (16.6%) selectively included Latinos or African Americans (8.3%). Half of the studies included women only, and in the remaining 6 articles 58.7% of participants were female.

The interventions were of a mean duration of 10±5 months and participants were observed for 14±6 months. In 4 studies (33.3%) the intervention consisted of a combination of group sessions and self-directed PA, 2 interventions were held in a church-setting, a community-setting was chosen by 3 research groups (25.0%) and another 2 studies (16.6%) started in a group setting but were continued as self-directed PA programmes. Participants were advised to be active ranging from 60 to 270 minutes per week in 1 to 6 supervised and unsupervised sessions. Most interventions were compared to a home-based programme (n=3, 25.0%) or no treatment control groups (n=3, 25.0%).

Interventions applied a mean of 7±3 (median=7.5) behaviour change techniques (BCTs), ranging from 3 to 12, whereas the control condition consisted of 0 to 12 BCTs with a mean of 2±4 (median=0.0). In most studies (n=10) there were more BCTs applied in the intervention group, but in 2 the same BCTs were applied in the control condition and the structured programme. As all interventions provided exercise sessions, the most frequently used BCTs were 'Instruction on how to perform the behaviour' (4.1; n=12), 'Demonstration of the behaviour' (6.1, n=12) and 'Behavioural practise/rehearsal' (8.1; n=12), followed by 'Goal setting (behaviour)' (1.1, n=8) and 'Graded tasks' (8.7; n=8). The BCTs

201 'Action planning' (1.4), 'Feedback on outcomes of behaviour' (2.7), 'Social support (practical)' (3.2),

202 'Social Support (emotional)' (3.3), 'Prompts/Cues' (7.1) and 'Material incentive (behaviour)' (10.1)

203 were applied only once in different studies. Control conditions most frequently used 'Goal setting

204 (behaviour)' (1.1) in addition to the provision of group sessions (BCTs 4.1, 6.1, 8.1).

205 Change in PA levels

206 Effectiveness is commonly assessed through the comparison of interventions to control conditions with

207 respect to whether their effect on increasing PA outcome measures is statistically significantly greater.

208 To assess short-term effectiveness of structured interventions for increasing PA, results at the first

209 follow-up assessment were summarised. Three categories of methods used to define and evaluate PA

210 levels were identified: 1) Changes in objectively measured PA; 2) Changes in an intended outcome of

211 PA behaviour as a proxy (i.e. change in VO_{2max}); and 3) Changes in self-reported PA levels.

212 Changes in objectively measured PA levels were assessed by only one study using accelerometry.

213 Three studies examined changes in VO_{2max} through maximal treadmill test as a proxy for PA behaviour

214 (i.e. it was inferred that an increase in cardiorespiratory fitness was indicative of a change in PA levels).

215 Different questionnaires were used to determine self-reported PA levels, e.g. 7-day recalls, the Modified

216 Baecke Questionnaire or the Voorips questionnaire.

217 On average, the first follow-up assessment took place after 5.92 months (range: 2-12). A total of 7

218 structured interventions were shown to be more effective than the controls (Table 2). Two trials resulted

219 in an increase in PA outcomes in both treatment groups, indicating the effectiveness of both the control

220 condition and the structured intervention. Only two studies did not result in an effect in either group, of

221 which one was described as a pilot study and was underpowered for statistical analysis.

222 We did not identify differences in control conditions, PA recommendations, or study quality, between

223 short-term effective and non-effective interventions. All but one of the 6 interventions selectively

224 including women were effective. BCT coding revealed that of 7 effective structured interventions 4

225 included 'problem solving' (1.2) into their intervention sessions.

226   To assess long-term changes, we summarised follow-up assessments at least 6 months after short-term

227   effects were assessed. We further distinguished between long term-effects while the intervention was

228   ongoing and after its termination. Of 7 interventions resulting in short-term effects PA levels were

229   assessed in 3 studies while the intervention was ongoing, whereas 3 only evaluated effects after the

230   termination of the intervention. Only one study assessed for effectiveness at both time points. During

231   the intervention period, differences in effects were still significant in 2 out of 4 trials, assessed after an

232   average of 15 months (range: 12-18), whereas after their termination 3 structured programmes sustained

233   higher PA outcomes than the control (21 months; range: 12-24). The number of included studies was

234   too low to be able to identify trends in characteristics of intervention effectiveness in long-term.

235   Adoption and maintenance rates

236   Information on the proportion of participants reaching the PA recommendations from authors within

237   the studies (which notably varied between studies and often diverged from current guidelines; see table

238   1) was only provided for 3 of 12 articles. One trial published a graph on adoption and maintenance

239   rates, but no information on the exact percentages was provided.

240   In most cases (n=10), adoption and maintenance behaviour were assessed via the same methods as for

241   physical activity levels, with exemption of 2 studies using activity recalls for the analysis of PA levels,

242   but assessed the stages of change for the evaluation of adoption and maintenance rates. The Stages of

243   Change are also referred to as the 'Transtheoretical Model' and describe a theory aiming to explain

244   behaviour change or somebody's readiness for change. It is based on the assumption each individual

245   moves through 5 stages when aiming for sustainably change his or her behaviour, namely

246   precontemplation, contemplation, preparation, action, maintenance. Whereas stage 1 and 2 are

247   characterised by someone's intentions to become more active (stage 2) or the absence of such (stage 1),

248   individuals in the preparation stage will undertake first steps towards fulfilling their goal. Stage 4 and

249   5 describe active individuals who either have (stage 5) or have not (stage 4) been fulfilling their activity

250   goals for at least 6 months[32]. Therefore, a change from stage 1, 2 or 3 to stage 4 and from stage 4 to 5

251   describe adoption or maintenance of PA, respectively.

252   For structured interventions, a mean adoption rate was reported as 67±16% (n=3), whereas on average

253   45±31% participants of the control group were adopting (n=3) (Table 2). This was assessed after 6 and

254   12 months. Of those, long-term effects were evaluated in 2 cases after 12 or 24 months. The average

255   maintenance rate was at 29±13% (n=2) for the intervention group and 18±3% (n=2) for the control

256   condition. Due to the low number of studies we were unable to evaluate differences in adoption and

257   maintenance rates and underlying mechanisms.

258   Retention Rates and Dropout

259   Retention rates were published in 9 articles (Table 3). The participation in a structured intervention

260   resulted in a mean retention rate of 75±13% as opposed to 68±11% in the control group. For 8 studies

261   providing information on reasons for dropout, no differences were found between control conditions

262   and interventions. Six studies reported the frequency of dropout reasons for 136 participants in the

263   control groups and 156 participants in the intervention groups. For the control conditions, most

264   frequently reported reasons for the termination of the participation in the interventions were: Loss to

265   follow-up (n=41); Lack of time and/or motivation (n=54), and withdrawal, illnesses and/or injuries

266   (n=14), and non-interest in the study (n=11). Participants in the intervention groups most commonly

267   dropped out due to a lack of time and/or motivation (n=69); loss to follow-up (n=27); withdrawal

268   (n=20); illnesses and/or injuries (n=14); and other unspecified reasons (n=14). For both the intervention

269   and control groups rarely reported reasons included: relocation (n=3); death (of family member) (n=2);

270   unreliable responses (n=1); pregnancy (n=1). Other reasons where no frequencies were reported were

271   non-participation, lack of spousal support, domestic violence, and missing consent forms.

272   Adherence Rates

273   Adherence was assessed by evaluating the proportion of attended intervention sessions in relation to

274   the recommendation (Table 3). This was documented through self-reported exercise logs and registers

275   at the exercise sessions. In 8 articles adherence to the structured programme was reported, resulting in

276   a mean rate of 61±21%. In 4 cases, control conditions providing alternative sessions were assessed for

277   adherence, with a mean rate of 59±30%.

278   Adherence to the structured programmes was similar between included studies, with the exemption of

279   Young and Steward (2006), who reported an adherence rate of 18%[31]. For the control conditions Yang

280   et al. (2016) report an adherence rate differing majorly from the average, as participants adhered to

281   100% to the activity recommendation, in contrast to the intervention group with 67%[30].

282   Discussion

283   This review highlights how adoption and maintenance rates of structured PA interventions are rarely

284   reported, yet, alongside reporting of mean changes/differences at group level, are needed to draw more

285   comprehensive conclusions on an intervention's efficacy, effectiveness, and feasibility, and to interpret

286   the relevance of observed effects. Though of course these sample group level effects may be

287   representative of population effects, this method only provides limited evidence on the overall

288   effectiveness of an intervention's implementation. As such the group level effects in terms of efficacy

289   seen may not be accurate reflections of the interventions effectiveness when implemented in real world

290   settings due to these behavioural elements. The discrepancy between the assessment of intervention

291   feasibility by considering proportions of individuals adopting and maintaining behaviours, and overall

292   group differences in outcome measures, is highlighted when interpreting the results published by Dunn

293   et al. (1999), who showed that despite a maintenance rate of only 20%, a significant increase in PA

294   levels at follow-up compared to baseline levels was observed[19]. Dunn et al. (1999) applied different

295   questionnaires to assess individual and group effects, which might partly explain the variance of

296   findings, though this further emphasises the need for consistency in reporting. This example does

297   however serve to underline how group level changes may be found to be statistically significant even

298   when only a small proportion of individuals in the sample groups have considerable improvements.

299   Thus, merely considering whether a statistically significant change in a chosen outcome measure such

300   as PA levels occurs does not reflect the relevance and magnitude of effects and ultimately the

301   interventions effectiveness.

302   The mechanisms of why some PA interventions are effective, whereas others do not result in the

303   anticipated effect are not fully understood. We found a tendency for structured programmes to result in

304   a greater increase in PA levels than the control conditions in short-term, with more varying results in

305   long-term. A meta-analysis found PA interventions to result in higher PA levels than controls up to 15

306   months, providing evidence for their effectiveness in long-term[33]. Howlett et al (2019) recently

307   published a meta-analysis including both interventions aiming to increase activity levels and reduce

308   sedentary time in healthy adults, and showed that PA interventions were effective in short and long-

309   term[34]. They further conducted meta-regressions, analysing the associations of BCTs influencing

310   intervention effectiveness. A BCT is described as an 'active ingredient' of an intervention, detailing

311   how a targeted behaviour is intended to be changed[17]. Interventions usually consist of a combination of

312   different BCTs and the analysis of patterns in effective or non-effective interventions can therefore

313   contribute to understanding the mechanisms of each intervention[35]. As such, consideration of the BCTs

314   included within interventions may aid in the understanding of which are most effective for enhancing

315   adoption, retention, and maintenance.

316   Howlett et al. (2019) reported that the BCTs 'Biofeedback' (2.6), 'Demonstration of behaviour' (6.1),

317   'Behavioural practice/rehearsal' (8.1), and 'Graded tasks' (8.7) were found to be an indicator of

318   intervention effectiveness in short-term, whereas the inclusion of 'problem solving' (1.2), 'Review of

319   behavioural goals' (1.5), and 'Feedback on the behaviour' (2.2) had a detrimental effect[34]. In long-term

320   a larger effect was observed in studies applying the BCTs 'Action planning' (1.4), 'Instruction on how

321   to perform the behaviour'(4.1), 'Prompts/cues' (7.1), 'Behaviour practice/rehearsal' (8.1), 'Graded

322   tasks' (8.7), and 'Self-reward' (10.9), as opposed to the inclusion of 'Information about antecedents'

323   (4.2), which resulted in a smaller effect. Another review identified the BCTs 'goal setting (behaviour)'

324   (1.1) and 'self-monitoring of behaviour' (2.3) as effective for increasing PA levels of overweight and

325   obese adults in both short-and long-term, whereas 'goal setting (outcome) (1.3), 'feedback on outcome

326   of behaviour' (2.7), 'graded tasks' (8.7) and 'adding objects to the environment (12.5) were predictors

327   for long-term effectiveness only[36].

328   In our review the application of these BCTs resulted in a large heterogeneity of findings. Due to the low

329   number of included articles however we were unable to formally assess intervention characteristics

330   influencing effectiveness. Of the 8 studies including 'goal setting (behaviour)' (BCT 1.1)', 4 were found

331   to be effective in short-term, whereas 1 of 4 studies showed long-term effects. The BCT 'Graded tasks'

332   (BCT 8.7) was applied in 6 studies assessing long-term effectiveness, of which 3 showed long-term

333   effectiveness. In 6 studies both BCTs 1.1 and 8.7 were applied, but only one resulted in long-term

334   effectiveness. In regards of self-monitoring of behaviour (2.3), 3 out of 3 studies reported to be effective

335   in short-term and 2 out of 3 in long-term, whereas for self-monitoring of behavioural outcomes (2.4) 3

336   out of 4 were effective in short term and 2 out of 4 in long-term. The one study applying 'feedback on

337   outcome of behaviour' (BCT 2.7) was found the be effective in short-term, which is accordance to

338   Samdal et al. (2017). Clearly the application of BCTs is at present quite heterogeneous which makes it

339   difficult to draw specific conclusions on which are most effective. To draw clearer conclusions of the

340   effectiveness of specific BCTs applied in structured interventions to increase PA levels, more research

341   empirical research comparing specific approaches to facilitate meta-analysis is warranted.

342   More information and ultimately more research is needed to systematically summarise adoption and

343   maintenance rates of PA interventions. Of 12 interventions only 3 were assessed for adoption and or

344   maintenance rates. These 3 proved themselves to be successful in inducing adoption in 50-85% in

345   participants. However, where long-term changes were assessed after the termination of the programmes,

346   only 20-35% of participants were able to sustain effects and were therefore classified as maintainers.

347   The small number and heterogeneity of studies prevented us from conducting meta-analyses on these

348   rates as intended and we were unable to identify intervention components resulting in high adoption

349   and maintenance rates. The low maintenance rates observed in this review indicate the need for future

350   research to improve the effectiveness of interventions in order to increase the proportion of participants

351   meeting the desired behaviours.

352   The effectiveness of an intervention is further reflected through a participant's retention and adherence

353   rate. In this review, three quarters of participants were available for all follow-up assessments, which

354   in turn can be translated in a dropout rate of approximately 25% of those who begin the intervention.

355   In a meta-analysis of yoga interventions an average dropout rate of 11% was observed, rising to 15%

356   for interventions of a duration of 12 weeks or longer[37]. Similar results have been published by Stubbs

357   et al. (2016) in their meta-analysis, showing that 18% of participants with depression do not complete

358   the full course of a PA programme[38]. This lies in the recommended range of up to 20% dropout, as

359   specified by Cochrane guidelines[39]. Although our results show a slightly higher dropout, our findings

360   are in accordance with the dropout rates found in PA interventions with schizophrenia (27%) and HIV

361   (29%) patients[40,41]. However, as we only included interventions recruiting healthy individuals, the

362   comparability to these other reviews is limited.

363   In a review of adherence of cancer survivors to the attendance of exercise sessions, adherence was

364   between 62-78%, whereas for older people this proportion dropped to 58-77%[42,43]. We showed that

365   participants attended on average over 60% of the recommended exercise sessions. This might have

366   affected the fidelity of the interventions, and the heterogeneity observed in regards to the effectiveness

367   of structured interventions is perhaps likely to be caused by both different intervention designs in

368   addition to non-compliance of participants. Compared to the other included studies, Young and Steward

369   (2006), reported an adherence rate marginally lower than the average[31]. In fact, this trial was the only

370   one selectively including African American women, and one of two trials implementing a culturally-

371   adapted church-based programme. As the recommendation of being active for 60 minutes per week lies

372   at the lower end of the spectrum of recommendations in the included studies, this is unlikely to have

373   resulted in the low adherence rate. Historically, African Americans as a demographic are reported to be

374   less active than white adults[44,45]. Common barriers to PA among African American women are lack of

375   motivation, family obligations and lack of social support, and haircare maintenance and the preference

376   of a more voluminous body shape, which are less commonly found in any other ethnic group[46]. This

377   indicates African American Women need to be supported more in becoming and staying physically

378   active, and might explain why adherence in this study was lower compared to the other included articles.

379   For the control conditions Yang et al. (2016) report an adherence rate differing majorly from the

380   average, as participants adhered to 100% to the activity recommendation, in contrast to the intervention

381   group with 67%[30]. This suggests that in this case the same amount of PA was more easily integrated

382   into an individual's weekly routine than conversing a group-based programme into a home-based

383   programme after 2 months. However, the low sample size of this pilot trial does not allow clear

384   conclusions. In both groups the same BCTs were applied.

385   This review has several limitations. Firstly, we selectively included structured programmes, as they are

386   most commonly applied with respect to PA interventions and, due to their standardisation are more

387   readily replicated[2]. However, while searching for literature, the identification of an intervention as

388   structured was often difficult due to insufficient reporting of methodological design and thus there may

389   be studies that went unidentified and were thus excluded. This insufficient reporting also impaired the

390   risk of bias assessment and the coding for BCTs, although we perceive there to have been an

391   improvement in the descriptions of more recent publications. For each included treatment group, it is

392   possible more BCTs have been applied than we coded for, due to insufficient descriptions of

393   intervention design. We further included academic literature published since 1990, which is possible to

394   have confounded our results as advances in research methods might reduce the comparability between

395   older and more recent publications.

396   Our ability to draw clear conclusions from this review is impaired by the heterogeneity of structured

397   PA interventions, highlighting the second limitation of this review. Recommended PA levels, the

398   content and delivery of exercise sessions, assessment of PA levels, control conditions, intervention

399   duration and observation period are amongst the multitude of characteristics of studies differing

400   considerably between trials, resulting in a large number of factors potentially influencing intervention

401   effectiveness. Our review was unable to facilitate quantitative synthesis using meta-analysis or permit

402   meta-regression and subgroup analyses taking intervention characteristics into account, therefore we

403   are unable to explore the mechanisms associated with effective structured interventions.

404   We further highlight the issue of using surrogate measures of PA like maximal treadmill tests for the

405   assessment of effectiveness of PA interventions. Those measures only provide indirect insights on PA

406   levels where it is assumed that, where a PA intervention is delivered compared to a control, any

407   improvements in cardiorespiratory fitness will only be due to increased PA levels and thus can be used

408   as a surrogate marker of this behaviour. While this might appear reasonable and indeed improvements

409   in cardiorespiratory fitness are an often and intended outcome of performance as a result of PA

410   behaviour being linked to morbidity and mortality[47], many factors might influence changes in

411   cardiorespiratory fitness including genetics[48] as well as both the volume and intensity of effort of any

412   PA behaviours[49,50]. As such it is argued for future work that PA behaviours, and indeed the fidelity of

413   any intervention with respect to the PA behaviours (volume, intensity of effort, frequency etc.) should

414   be assessed in addition to the intended outcome of those behaviours (e.g. cardiorespiratory fitness,

415   strength, mental wellbeing etc.). This would permit greater understanding of both what impact

416   interventions have upon PA behaviours and subsequently the degree to which those behaviours might

417   mediate improved health, fitness, and wellbeing.

418   We originally intended to compare academic literature with grey literature in the form of evaluation

419   reports of structured public health interventions. However, we were unable to identify non-academic

420   literature outlining PA interventions in sufficient detail to be included in this review. We believe

421   evaluation of public health interventions can contribute to the improvement of current knowledge on

422   effective PA interventions, however, lack of control conditions and poor reporting has historically been

423   an issue within the sector (PP 1), though does seem to have improved (PP2) and thus this may be a

424   possibility in future reviews[51].

425   We perceive our strict inclusion criteria as a strength rather than a limitation, as this enabled us to draw

426   focused conclusions on the effectiveness of structured interventions by reducing heterogeneity of PA

427   programmes. However, we suggest for future reviews to include a broader spectrum of interventions to

428   be able to carry out a more comprehensive review and meta-analysis, and to conduct sub-analyses where

429   appropriate. A further strength of this review is our adherence to PRISMA and AMSTAR 2 guidelines

430   to produce a review to the recommended standard. Thirdly, we pre-registered the research protocol for

431   this review on PROSPERO prior to conducting the search, enabling researchers conducting similar

432   reviews to understand, reproduce or improve our approach. Moreover, our results can be used to inform

433   future evidence-based structured interventions. On the basis of this review we have developed and are

434   trialling a structured intervention aiming to increase adoption and maintenance rates of members of

435   leisure centres. As part of this research, we are further investigating which factors influence

436   implementation of PA programmes, potentially providing guidelines for academic trials in leisure

437   centres.

438 ## Conclusion

439 From the limited data available it can be concluded that structured interventions are effective in

440 influencing PA levels in the short-term and the adoption of PA behaviours, whereas for maintenance

441 no clear conclusions can be drawn. Implementation of interventions seems to be feasible in short- and

442 long-term, as expressed through moderate adherence and retention rates. However, to assess the

443 effectiveness of structured interventions more information on maintenance rates is needed.

444 To our knowledge this is the first review aiming to systematically summarise adoption, retention

445 maintenance, and adherence rates of long-term structured PA interventions in non-clinical adults. As

446 these rates are rarely reported, we propose a new point of view in regards of the evaluation of studies

447 considering these, and highly recommend future research to address this issue of underreporting by

448 publishing information on adoption and maintenance rates relative to the recommended amount of PA.

449 This will contribute to the improvement of our understanding of the feasibility of PA interventions, the

450 mechanism through which they are effective in changing PA behaviour, and therefore the design of

451 future PA interventions aiming to tackle global inactivity rates. A more comprehensive summary and

452 meta-analysis of literature is needed, including a wider range of PA interventions.

453 ## Acknowledgements

454 None.

455 ## Funding Sources

457 References

458 1. Sport England. Active people survey. https://www.sportengland.org/research/about-our-

459 research/active-people-survey/. Updated 2016. Accessed 04/21, 2017.

460 2. Slade S, Dionne C, Underwood M, et al. Consensus on exercise reporting template (CERT):

461 Modified delphi study. *Physical Therapy*. 2016;96(10):1514-1524.

462 3. World Health Organisation (WHO). Noncommunicable diseases and mental health; target 3:

463 Reduce prevalence of physical inactivity. http://www.who.int/nmh/ncd-tools/target3/en/. Updated

464 2018. Accessed 09/18, 2018.

465 4. Fjeldsoe B, Neuhaus M, Winkler E, Eakin E. Systematic review of maintenance of behaviour

466 change following physical activity and dietary interventions. *Health Psychol*. 2011;30(1):99-109.

467 5. Kahlert D. Maintenance of physical activity: Do we know what we are talking about? *Prev Med*

468 *Rep*. 2015;2:178-180.

469 6. Marcus B, Dubbert P, Forsyth L, et al. Physical activity behavior change: Issues in adoption and

470 maintenance. *Health Psychol*. 2000;19(1):32-41.

471 7. Page SJ, Persch AC. Recruitment, retention, and blinding in clinical trials. *Am J Occup Ther*.

472 2013;67(2):154-161.

473 8. Persch AC, Page SJ. Protocol development, treatment fidelity, adherence to treatment, and quality

474 control. *Am J Occup Ther*. 2013;67(2):146-153.

475 9. Beedie C, Mann S, Jimenez A, et al. Death by effectiveness: Exercise as medicine caught in the

476 efficacy trap! *Br J Sports Med*. 2016;50(6):323-324.

477 10. Friedenreich CM. Physical activity and breast cancer: Review of the epidemiologic evidence and

478 biologic mechanisms. *Recent Results Cancer Res*. 2011;188:125-139.

479 11. Löllgen H, Bockenhoff A, Knapp G. Physical activity and all-cause mortality: An updated meta-

480 analysis with different intensity categories. *Int J of Sports Med*. 2009;30(3):214-224.

481 12. Public Health England. Physical inactivity: Economic costs to NHS clinical commissioning

482 groups.

483 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524

484 234/Physical_inactivity_costs_to_CCGs.pdf. Updated 2016. Accessed 20/04, 2017.

485   13. Sattelmair J, Pertman J, Ding EL, Kohl III HW, Haskell W, Lee IM. Dose response between

486   physical activity and risk of coronary heart disease: A meta-analysis. *Circulation*. 2012;124(7):798-

487   795.

488   14. Atkinson G, Williamson PJ, Batterham A. Issues in the determination of "responders" and "non-

489   responders" in physiological research. *Exp Physiol*. 2019;In press.

490   15. Liberati A, Altman D, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews

491   and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration.

492   *BMJ*. 2009;339(b2700).

493   16. Shea B, Reeves B, Wells G, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that

494   include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*.

495   2017;358(j4008).

496   17. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93

497   hierarchically clustered techniques: Building an international consensus for the reporting of behavior

498   change interventions. *Ann Behav Med*. 2013;46(1):81-95.

499   18. Higgins J, Green S. *Cochrane handbook for systematic reviews of interventions version 5.1.0

500   [updated march 2011].* The Cochrane Collaboration; 2011.

501   19. Dunn A, Marcus B, Kampert J, Garcia M, Kohl H, Blair S. Comparison of lifestyle and structured

502   interventions to increase physical activity and cardiorespiratory fitness. *JAMA*. 1999;281(4):327-334.

503   20. Dunn AL, Marcus BH, Kampert JB, Garcia ME, Kohl HW,3rd, Blair SN. Reduction in

504   cardiovascular disease risk factors: 6-month results from project active. *Prev Med*. 1997;26(6):883-

505   892.

506   21. Arredondo EM, Elder JP, Haughton J, et al. Fe en accion: Promoting physical activity among

507   churchgoing latinas. *Am J Public Health*. 2017;107(7):1109-1115.

508   22. Cox KL, Burke V, Gorely TJ, Beilin LJ, Puddey IB. Controlled comparison of retention and

509   adherence in home- vs center-initiated exercise interventions in women ages 40-65 years: The

510   S.W.E.A.T. study (sedentary women exercise adherence trial). *Prev Med*. 2003;36(1):17-29.

511   23. De Jong J, Lemmink KAPM, Stevens M, et al. Six- month effects of the groningen active living

512   model (GALM) on physical activity, health and fitness outcomes in sedentary and underactive older

513   adults aged 55– 65. *Patient Educ Couns*. 2006;62(1):132-141.

514   24. Hertogh E, Vergouwe Y, Schuit A, Peeters P, Monninkhof E. Behavioral changes after a 1-yr

515   exercise program and predictors of maintenance. *Med Sci Sports Exerc*. 2010;42(5).

516   25. Hovell MF, Mulvihill MM, Buono MJ, et al. Culturally tailored aerobic exercise intervention for

517   low-income latinas. *Am J Health Promot*. 2008;22(3):155-163.

518   26. KETTUNEN O, VUORIMAA T, VASANKARI T. A 12-month exercise intervention decreased

519   stress symptoms and increased mental resources among working adults - results perceived after a 12-

520   month follow-up. *International Journal of Occupational Medicine & Environmental Health*.

521   2015;28(1):157-168.

522   27. King AC, Haskell WL, Young DR, Oka RK, Stefanick ML. Long-term effects of varying

523   intensities and formats of physical activity on participation rates, fitness, and lipoproteins in men and

524   women aged 50 to 65 years. *Circulation*. 1995;91:2596-2604.

525   28. Kukkonen K, Rauramaa R, Siitonen O, Hänninen O. Physical training of obese middle-aged

526   persons. *Annals Clin Res*. 1982;14(34):80-85.

527   29. Lee C, White SW. Controlled trial of a minimal-intervention exercise program for middle- aged

528   working women. *Psychol Health*. 1997;12(3):361-374.

529   30. Yang K, James KA. Yoga, as a transitional platform to more active lifestyle: A 6-month pilot

530   study in the USA. *Health Promot Int*. 2016;31(2):423-429.

531   31. Young DR, Stewart KJ. A church-based physical activity intervention for african american

532   women. *Fam Community Health*. 2006;29(2):103-117.

533   32. Marcus BH, Rakowski W, Rossi JS. Assessing motivational readiness and decision making for

534   exercise. *Health Psychol*. 1992;11(4):257-261.

535   33. Murray JM, Brennan SF, French DP, Patterson CC, Kee F, Hunter RF. Effectiveness of physical

536   activity interventions in achieving behaviour change maintenance in young and middle aged adults: A

537   systematic review and meta-analysis. *Soc Sci Med*. 2017;192:125-133.

538 34. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy

539 inactive adults effective in promoting behavior change and maintenance, and which behavior change

540 techniques are effective? A systematic review and meta-analysis. *Transl Behav Med*. 2019;9(1):147-

541 157.

542 35. Michie S, Abraham C, Eccles M, Francis J, Hardeman W, Johnston M. Strengthening evaluation

543 and implementation by specifying components of behaviour change interventions: A study protocol.

544 *Implement Sci*. 2011;6(10).

545 36. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for

546 physical activity and healthy eating in overweight and obese adults; systematic review and meta-

547 regression analyses. *Int J Behav Nutr Phys Act*. 2017;14(42).

548 37. Cramer H, Haller H, Dobos G, Lauche R. A systematic review and meta-analysis estimating the

549 expected dropout rates in randomized controlled trials on yoga interventions. *Evid Based Complement

550 Alternat Med*. 2016;2016:5859729.

551 38. Stubbs B, Vancampfort D, Rosenbaum S, et al. Dropout from exercise randomized controlled

552 trials among people with depression: A meta-analysis and meta regression. *J Affect Disord*.

553 2016;190:457-466.

554 39. Furlan AD, Pennick V, Bombardier C, van Tulder M, Editorial Board, Cochrane Back Review

555 Group. 2009 updated method guidelines for systematic reviews in the cochrane back review group.

556 *Spine (Phila Pa 1976)*. 2009;34(18):1929-1941.

557 40. Vancampfort D, Mugisha J, Richards J, et al. Dropout from physical activity interventions in

558 people living with HIV: A systematic review and meta-analysis. *AIDS Care*. 2017;29(5):636-643.

559 41. Vancampfort D, Rosenbaum S, Schuch FB, Ward PB, Probst M, Stubbs B. Prevalence and

560 predictors of treatment dropout from physical activity interventions in schizophrenia: A meta-

561 analysis. *Gen Hosp Psychiatry*. 2016;39:15-23.

562 42. Kampshoff C, van Mechelen W, Schep G, et al. Participation in and adherence to physical

563 exercise after completion of primary cancer treatment. *Int J Behav Nutr Phys Act*. 2016;13(100).

564   43. Picorelli A, Pereira L, Pereira D, Felício D, Sherrington C. Adherence to exercise programs for

565   older people is influenced by program characteristics and personal factors: A systematic review. *J*

566   *Physiother*. 2014;60(3):151-156.

567   44. American Heart Association. Heart Disease and Stroke Statistics—2018 Update: A report from

568   the American Heart Association. *Circulation*. 2018;137(e67):e492.

569   45. American Heart Association. Heart Disease and Stroke Statistics—2006 Update. *Circulation*.

570   2006;113(6):e85-e151.

571   46. Joseph RP, Ainsworth BE, Keller C, Dodgson JE. Barriers to physical activity among african

572   american women: An integrative review of the literature. *Women Health*. 2015;55(6):679-699.

573   47. Ross R, Blair SN, Arena R, et al. Importance of assessing cardiorespiratory fitness in clinical

574   practice: A case for fitness as a clinical vital sign: A scientific statement from the american heart

575   association. *Circulation*. 2016;134(24):e653-e699.

576   48. Peter I, Papandonatos GD, Belalcazar LM, et al. Genetic modifiers of cardiorespiratory fitness

577   response to lifestyle intervention. *Med Sci Sports Exerc*. 2014;46(2):302-311.

578   49. Montero D, Lundby C. Refuting the myth of non-response to exercise training: 'Non-responders'

579   do respond to higher dose of training. *J Physiol*. 2017;595(11):3377-3387.

580   50. Ross R, de Lannoy L, Stotz PJ. Separate effects of intensity and amount of exercise on

581   interindividual cardiorespiratory fitness response. *Mayo Clin Proc*. 2015;90(11):1506-1514.

582   51. Public Health England. Identifying what works for local physical inactivity interventions.

583   http://researchinstitute.ukactive.com/downloads/managed/Identifying_what_works.pdf. Updated

584   2014. Accessed 14/03, 2018.

585

586   Tables

Author (Year)	Country	Type of Trial	Study Quality	Sample size	Age (Years)	% Female	Ethnicity	Setting	Intervention Duration	Recommendation	Control	BCTs (Intervention)	BCTs (Control)
Arredondo et al. (2017)	USA	Cluster-RCT	high	I: 187 C: 183	I: (44.5±9.8 C: 44.4±9.4	100%	Latino	Culturally adapted church-based intervention	12 months	150 minutes MVPA, individual choice of the amount of sessions (Walking groups, cardio dance and strength training classes)	Cancer screening classes	1.1; 1.2 3.1 4.1 6.1 8.1 10.1; 10.2	NIL
Cox et al. (2003)	Australia	RCT	high	I: 64 C: 62	I: 48.0 C: 48.3	100%	predominantly white	Group-based programme	18 months	3 times or more per week 20 minutes or longer MVPA (walks, aerobics, circuits)	Home-based	1.1; 1.2 2.1; 2.2; 2.3; 2.4 3.1 4.1 5.1 6.1 8.1; 8.7	1.1; 1.2 2.1; 2.2; 2.3; 2.4 3.1 4.1 5.1 6.1 8.1; 8.7
De Jong et al. (2006)	Netherlands	Cluster-RCT	moderate	I: 79 C: 102	I: 59.6±2.4 C: 58.8±2.7	I: 54.4% C: 56.9%	NIL	Gym-based leisure-time programme, including the 15 most popular sports for older adults	6 months	One session of 60 min per week	Waiting list	4.1 6.1 8.1	NIL
Dunn et al. (1997; 1999)	USA	RCT	high	I: 121 C: 114	I: 46.2±6.5 years C: 45.9±6.8	I: 50.9% C:50%	predominantly white	Group-based programme, (after 6 months home-based)	6 months (24 months observation)	150 min MVPA, stage 4 or 5 of the TTM initially 3 sessions, then 5 session per week, duration of 20-60 min, 50-85% of maximal aerobic power	Home-based	1.1; 1.4 2.1 3.1 4.1 6.1 8.1; 8.7	1.1; 1.2 4.1 6.1 8.1 10.3

Study	Country	Design	Quality	N	Age	%	Ethnicity	Intervention	Duration	Details	Control	Codes	
Hertogh et al. (2010)	Nether-lands	RCT	moderate	I: 78 C: 68	I: 59.0±4.5 C: 58.7±4.0	100%	NIL	Group-based and individual sessions	12 months (24 months observation)	3 sessions per week, twice supervised (60 min, strength and endurance), once individual (30 min brisk walking or cycling)	No intervention	2.3; 2.4 4.1 6.1 8.1; 8.7	NIL
Hovell et al. (2008)	USA	RCT	high	I: 68 C: 66	NIL	100%	Latino	Community programme for low-income, monolingual Spanish speaking Latinas	6 months (12 months observation)	3 times 90 minutes per week (aerobic dance exercise)	Safety education	1.2 2.2; 2.6 3.2; 3.3 4.1 5.1 6.1 8.1; 8.7 10.2	NIL
Kettunen et al. (2015)	Finland	C-ITS	low	I: 338 C: 33	I: 45±8.8 C: 41±6.9	I: 62.72% C: 51.5%	NIL	Supervised and individual exercise sessions' (after 12 months unsupervised exercise)	12 months (24 months observation)	2 day training camps at assessments (1-2 supervised sessions per 5 months), 3-5 unsupervised sessions per week (walking, skiing, biking)	No intervention	2.3; 2.4 4.1 6.1 8.1	NIL
King et al. (1995)	USA	RCT	high	I: 74 C: 75	NIL	I: 45.9% C: 45.3%	predominantly white	Group-based intervention	12 months	3 sessions of 60 min per week (walking/jogging, treadmills and stationary cycles)	No intervention	1.1 4.1 6.1 8.1; 8.7	NIL
Kukkonen et al. (1982)	Finland	Cohort	Low	I: 169	NIL	57.4%	NIL	supervised and individual sessions, individual training programme for each participant	17 months	30-60 min 3-6 times a week (walking, skiing, jogging, swimming, cycling), one a week supervised (calisthenics, volleyball)	No control group	1.1 2.1; 2.4 4.1 6.1 8.1; 8.7	NIL

25

Lee et al. (1997)	Australia	Cross-over	moderate	I: 14 C: 11	NIL	100%	NIL	self-administered programme of low-impact aerobic exercise, weekly activity and education	3 months (12 months observation) *	weekly low intensity aerobic exercise session, (duration unknown, walking or exercise to music) 2-3 individual sessions	Waiting list	1.1; 1.2 2.6; 2.7 4.1 6.1 8.1; 8.6; 8.7	2.6; 2.7
Yang et al. (2016)	USA	Pilot RCT	low	I: 7 C: 7	I: 58.4±6.8 C: 58.7±4.1	I: 85.71% C: 85.71%	predominantly white	pilot study, group-based yoga programme (after 2 months home-based)	6 months	1 supervised session (90 min) and 2 unsupervised sessions at least twice a week	Home-based (Yoga DVD), 2 initial supervised sessions	1.1 4.1 6.1 8.1; 8.6	1.1 4.1 6.1 8.1; 8.6
Young and Steward (2006)	USA	Cluster RCT	high	I: 123 C: 73	I: 48.2±.24.4 C: 48.4±19.65	100%	African American	Culturally adapted church-based aerobic intervention	6 months	individualised activity plans, 60 min per week	alternating weekly low-intensity stretching classes and health lectures	1.1 3.2 4.1 6.1 7.1 8.1; 8.7 10.2	3.2

587   Table 1: Study Design and Intervention Characteristics.

588   *Due to the cross-over design of this study only data until 24 weeks were included in this review. BCTs were coded in accordance with the BCT taxonomy published by Michie et al. (2013).

Author (Year)	Adoption					Maintenance (during intervention)					Maintenance (after intervention)		
	Definition	PA Measure	Time point	Rate	Short-term increase in PA levels	Definition	PA Measure	Time point	Rate	Long-term increase in PA levels	Time point	Rate	Long-term increase in PA levels
Objectively measured physical activity													
Arredondo et al. (2017)	Proportion of participants reaching 150min MVPA per week	Accelerometer	12 months	I: 53% C: 40%	Intervention group			No follow-up			No follow-up		
Physical activity behaviour outcome as proxy													
Kettunen et al. (2015) [B]		Maximal Treadmill test (VO_{2max})	4 months	NIL	Intervention group		Maximal Treadmill test (VO_{2max})	12	NIL	Intervention group	24 months	NIL	Intervention group
King et al. (1995) [A, B]		Maximal Treadmill test (VO_{2max})	6 months	NIL	Intervention group		Maximal Treadmill test (VO_{2max})	12 months	NIL	Intervention group	No follow-up		
Kukkonen et al. (1982) [B]		Maximal Treadmill test (VO_{2max})	2 months	NIL	No difference		Maximal Treadmill test (VO_{2max})	17 months	NIL	No difference	No follow-up		
Stages of Change													
Cox et al. (2003) [C]	Stage 4 of the TTM (assessed via Stage of Change instrument Questionnaire	7-day Recall (energy expenditure)	6 months	unclear	Intervention group	Stage 5 of the TTM	Stage of Change instrument (stage 5)	18 months	unclear	No difference	No follow-up		
Dunn et al. (1997; 1999) [+]	Meeting ACSM recommendation (Stage 4)	7-day recall (6m), Maximal treadmill test (24m)	6 months	I: 85% C: 78%	No difference	stage 5 of the TTM 0	Proportion of participants meeting the ACSM recommendation (stage 5)	Not assessed			24 months	I: 20% C: 20%	No difference
Changes in self-reported physical activity behaviour													
De Jong et al. (2006)		Voorrips questionnaire (energy expenditure)	6 months	NIL	No difference			No follow-up			No follow-up		
Changes in self-reported physical activity behaviour (continued)													
Hertogh et al. (2010)		Modified Baecke Questionnaire (MET/h)	12 months	NIL	Intervention group		Modified Baecke Questionnaire (MET/h)	Not assessed			24 months	NIL	Intervention group

Study	Outcome	Measure	Time	Result	Favoured	Outcome	Measure	Result		Time	Result	Favoured
Hovell et al. (2008)	Proportion of participants reaching 150min MVPA per week	7-day recall	6 months	I: 63% C: 16%	Intervention group	150min MVPA per week	7-day recall	Not assessed		12 months	I: 38% C: 15%	Intervention group
Lee et al. (1997)		Physical activity recall (MVPA)	3 months	NIL	Intervention group			Not assessed		No follow-up		
Yang et al. (2016)		Modifiable Activity Questionnaire (MET/h)	2 months	NIL	No difference			No follow-up		No follow-up		
Young and Steward (2006)		Physical activity recall	6 months	NIL	No difference			No follow-up		No follow-up		

Table 2: Adoption and Maintenance of Physical Activity.

[A] only information of 2 groups included in this review. [B] Separate analysis for men and women. [C] No numbers published (only graph). *Data from 2 publications summarised. Physical activity (PA); Intervention (I); Control (C); Moderate to vigorous physical activity (MVPA); Metabolic equivalents of tasks per hour (MET/h); transtheoretical model (TTM).

| Author (Year)	Length of Follow-up (months)	Retention Rate	Adherence	
			Rate	Assessment tool
Arredondo et al. (2017)	12	I: 84.33% C:84.93%	Not assessed	
Cox et al. (2003) [C]	18	I: 81.25% C: 61.29%	I: 65.4% C: 50.8%	attendance records, self-report
De Jong et al. (2006)	6	I: 48% C: 67%	I: 80%	attendance records
Dunn et al. (1997; 1999) [A]	24	I: 78.26 % C: 81.97%	Not assessed	
Hertogh et al. (2010)	24	I: 81.25% C: 73.12%	Not assessed	
Hovell et al. (2008)	12	Not assessed	I: 65.8% C: 55.4%	attendance records
Kettunen et al. (2015)	24	I: 84.85% C: 52.66%	Not assessed	
King et al. (1995) [A]	12	Not assessed	I: 52.6 %	attendance records
Kukkonen et al. (1982)	17	I (men): 56.94% I (women): 55.67%	Not assessed	
Lee et al. (1997)	12	I: 73.6% C: 72%	I: 80%	attendance records
Yang et al. (2016)	6	I: 85.71% C: 57.14%	I: 67% C: 100%	exercise logs
Young and Steward (2006)	6	I: 60.98% C: 57.53%	I: 18.3% C: 28.6%	attendance records

592 Table 3: Retention and Adherence Rates.

593 [A] Only information of 2 groups included in this review. [B] No numbers published (only graphic visualisation). Physical activity (PA);

594 Intervention (I); Control (C).

595   Figures/Graphics
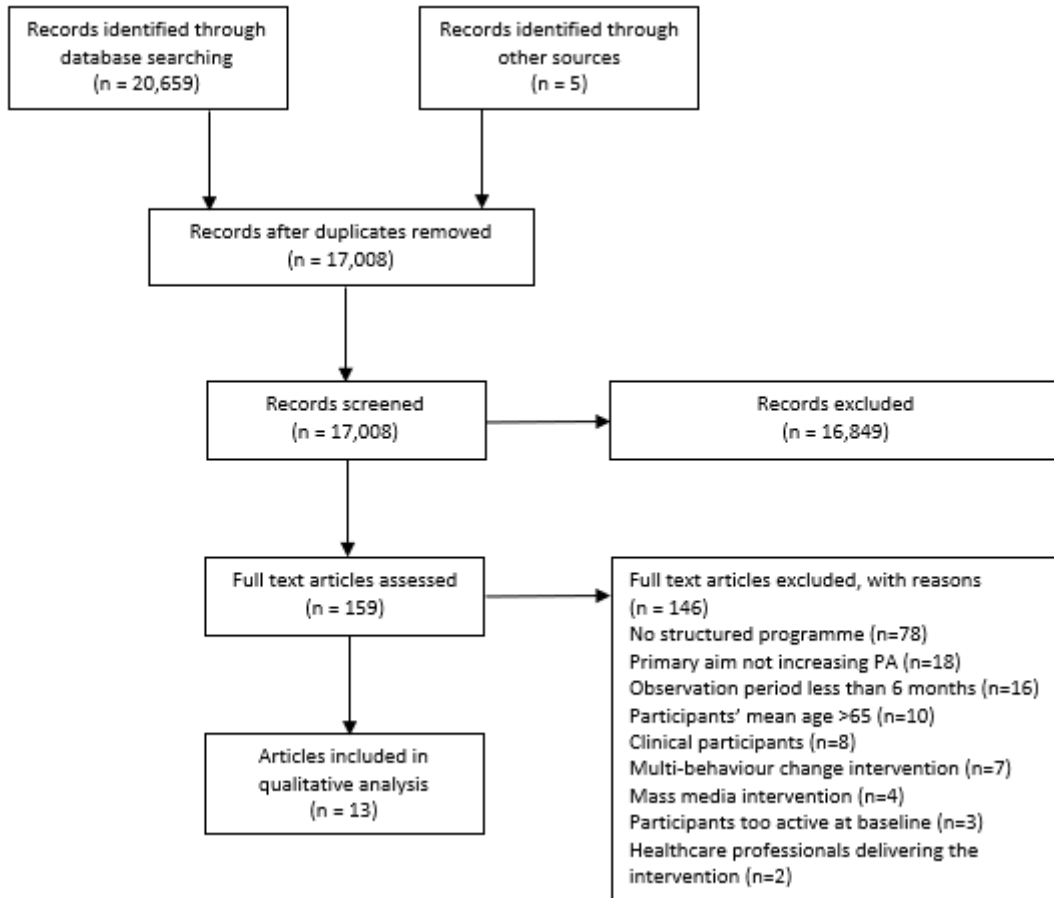
596

597   Figure 1: PRISMA Chart Screening Process