Intervention fidelity and effectiveness of a UK worksite physical activity intervention funded by the Bupa Foundation, UK

REBECCA LAWTON1*, ROSIE MCEACHAN2, CATH JACKSON3, ROBERT WEST4 and MARK CONNER5

1Institute of Psychological Sciences, University of Leeds, Leeds, UK, 2Born in Bradford, Bradford Institute for Health Research, Bradford, UK, 3Health Sciences, University of York, York, UK, 4Health Sciences, University of Leeds, Leeds, UK and 5Psychological Sciences, University of Leeds, Leeds, UK

*Corresponding author. E-mail: r.j.lawton@leeds.ac.uk

SUMMARY

The main aim of this study was to test whether the effectiveness of a worksite physical activity intervention delivered in five work organizations varied as a function of intervention fidelity. We conducted a fidelity analysis as part of a large matched-pair cluster randomized controlled trial of a worksite physical activity intervention (AME for Activity). Participants (N = 1260) were employees from five organizations in the UK. The primary trial outcome was physical activity at 9 months post intervention. Adherence, exposure, quality of delivery and participant responsiveness/engagement were measured to assess fidelity. Qualitative data about the context in which the intervention was delivered were collected via focus groups, interviews and field notes. Multi-level modelling was used to provide a comparison of the effect of the intervention on increases in physical activity for worksites where intervention fidelity was good, compared with those where intervention fidelity was poor or moderate. Intervention fidelity was poor in two organizations, moderate in two organizations and good in one organization (local council). Re-analysis of the trial data comparing employees in the local council (N = 443) with employees in all other worksites (N = 611) revealed a significant effect of the intervention on physical activity levels among council employees only. These findings suggest that the measurement of fidelity and the testing of the effects of intervention fidelity on outcomes, as part of the evaluation of complex interventions, are essential to understand the context and conditions in which interventions are most effective.

Key words: fidelity; physical activity intervention; cluster randomized controlled trial; worksite

INTRODUCTION

There is now convincing evidence that people who are physically active live longer and have lower morbidity (Allender et al., 2007; Lee et al., 2012). However, more than half the population of the UK, USA and Canada (Centers for Disease Control and Prevention, 2008; Statistics Canada, 2009; The Information Centre for Health and Social Care, 2009) do not achieve the recommended 30 min of moderate intensity activity, on five or more days of the week (O’Donovan et al., 2010).

Most working adults spend half of their waking hours at work, many in sedentary occupations, making the workplace an excellent setting for promoting health (Black, 2008; Department for Work and Pensions, Department of Health, 2008; Centers for Disease Control and Prevention, 2012). In addition to improving employee health, there is a strong business case for promoting physical activity (Kreis and Bodeker, 2004; Black, 2008). Many of the economic benefits are related to better health (e.g. reduced absenteeism and reduced back pain), but less obvious benefits of
physical activity have also been demonstrated (e.g. increased productivity, increased stress tolerance and improved decision-making; Kreis and Bodeker, 2004). A number of physical activity interventions have therefore focused on the workplace as a place where interventions could be delivered; however, evidence as to their effectiveness has been mixed. In a quantitative review, Dishman and colleagues (Dishman et al., 1998) found no convincing evidence for the effectiveness of workplace interventions with a meta-analysis of 45 effects yielding a heterogeneous and small effect size of $r = 0.11$ (95% CI −0.20 to 0.40). Two more recent systematic reviews have reported more optimistic findings (Dugdill et al., 2008; Conn et al., 2009); however, they highlight the generally poor methodological quality of studies in this area, e.g. failure to follow principles of randomized controlled trials. Furthermore, Abraham and Graham-Rowe (Abraham and Graham-Rowe, 2009) report greater effect sizes for the outcome of increased fitness among worksite interventions that focused specifically on physical activity ($d = 0.29$), rather than on healthy lifestyle more generally ($d = 0.08$). Thus, there is a need for more methodologically robust studies that take into account key principles of randomized controlled trials such as randomization, blinding and compliance and which focus specifically on promoting physical activity.

In addition, there is a growing argument that interventions must also be theoretically grounded to promote efficacy and the advancement of understanding in this field (Davidson et al., 2003; Des Jarlais et al., 2004). Despite the importance of theory, a recent review of workplace physical activity intervention studies found that only 13 of 22 explicitly mentioned use of a theoretical framework, and furthermore, only four were clear about how the framework was applied (Taylor et al., 2012). This finding is supported by a recent review of physical activity interventions across a broader range of settings (Rhodes and Nasuti, 2011).

To address the lack of robust cluster RCTs and the lack of clarity about the theoretical premise of many interventions, McEachan et al. (McEachan et al., 2008) conducted a cluster RCT to test the effectiveness of a worksite physical activity intervention (AME for Activity) based on the Theory of Planned Behaviour. The main findings of this effectiveness trial are reported by McEachan et al. (McEachan et al., 2011) and demonstrate non-significant effects on the primary outcome which was self-reported physical activity over 12 months. However, employees in the intervention worksites did show significantly greater decreases between baseline and 12-month follow-up in systolic blood pressure and resting heart rate than employees in the control worksites.

It is now widely acknowledged that when testing complex interventions via randomized controlled trials, it is important to collect data about how the intervention is delivered in practice (fidelity) and whether this varies according to the context (Bellg et al., 2004; Oakley et al., 2006; Craig et al., 2008). This evaluation of the process of intervention delivery is deemed to be particularly important in multi-centre trials where the ‘same intervention may be implemented and received in different ways’ [(Oakley et al., 2006), p. 413].

In any complex intervention treatment, fidelity is a key challenge (Bellg et al., 2004; Dane and Schneider, 1998). For example, in a large-scale community-based physical activity intervention study, Hardeman et al. (Hardeman et al., 2008) found that facilitators delivered only around 44% of the specified intervention techniques across four key sessions. If an intervention is not implemented as directed and no effect is found, then one cannot be sure whether this is due to lack of efficacy of the intervention or simply that it has not been implemented correctly (Hasson, 2010). It is recommended, therefore, that evaluations of intervention fidelity become an integral part of the conduct and evaluation of all health behaviour intervention research (Bellg et al., 2004). Indeed, a recent study, similar to that reported here found that the degree of physical activity change achieved in a workplace physical activity intervention was directly related to the level of implementation with greater increases found in worksites who implemented more of the intervention (Wilson et al., 2010).

The AME for Activity intervention, which is the focus of this article, was a detailed intervention tool kit, delivered by employees trained in the intervention who then implemented it within their own workplace over a 3-month period. These trained employees are referred to as facilitators. The potential, therefore, for the fidelity of the intervention to vary considerably across worksites and organizations was great. Furthermore, the intervention was designed to be flexible to meet the needs of different organizations (bus company, local council, university, hospital, government organization) and the varying needs of employees themselves. The objective
of the intervention was to increase the levels of at least moderate intensity activity with the aim of achieving the current recommended levels of at least 30 min on at least 5 days of the week (Haskell et al., 2007; O’Donovan et al., 2010). Moderate intensity activity was defined as any activity that ‘makes you feel warmer, breathe harder and make your heart beat faster than usual, but not so much you can’t have a conversation whilst being active’ (British Heart Foundation, 2014); an increase in moderate intensity activity was defined as any increase lasting at least 10 min, as this 10 min period was posited, at the time, to be the minimum length required for health benefits (US Department of Health and Human Services, 1996). Precise standardization of the delivery of the intervention was neither desirable nor feasible and any employee who met the Physical Activity Readiness criteria (PAR-Q; Thomas et al., 1992) could volunteer to participate, irrespective of current level of activity. We predicted that there would be variation in both the delivery of the intervention (by facilitators) and the response to the intervention (by employees) and therefore the fidelity of the intervention. The main aim of the study reported here was to test whether the effectiveness of a complex worksite physical activity intervention delivered in five different work organizations varied as a function of intervention fidelity. We addressed three research questions (RQs):

(1) Does the quality of delivery of the intervention (fidelity) vary across work organizations?
(2) What contextual factors (particularly features of organizational and social context of work) affect the delivery of the intervention?
(3) Does the intervention significantly increase the levels of physical activity among employees in those work organizations who deliver the intervention as intended?

METHOD

A full description of the trial design can be found in McEachan et al. (McEachan et al., 2011).

Design of the trial

A matched pairs cluster randomized controlled trial was used. Pairs of worksites, matched according to the function and size, were randomly assigned to be in either the intervention or control group. Participants completed measures of physical activity at baseline (T1) and again immediately after the intervention (T2), 3 months after the intervention (T3) and 9 months after the intervention (T4: 12 months after baseline measures). Thus, the design had one repeated measure—time—with four levels (baseline and three follow-ups) and one between subjects factor—group—with two levels (intervention and control). Based on previous research and experience, we set out to detect a standardized effect size (d) of 0.2 on physical activity levels, between the intervention and control groups using a two-sided significance level of 0.05 and minimum power of 0.80. This yielded a total sample size of 902 (451 in each group).

Participants

Eligible worksites were those which could be matched with a similar worksite (in the same organization) in a different geographical location (e.g. two schools, two bus depots). All organizations approached agreed to take part. From these organizations, 4370 participants from 44 worksites were invited to take part. Of these 1225 booked and attended a health check at baseline, 1164 completed a questionnaire at T1 and 1124 did both. Eighty-six respondents were excluded on health grounds (as advised by the occupational health physician linked to the project). Participants were recruited from five work organizations (bus company, local council, hospital, government organization, university). A total of 1260 individuals from 44 worksites provided a measurement of at least one time point and were included in the analysis.

Intervention

The nine key components of the intervention called ‘AME (Awareness, Motivation and Environment) for activity’ were: a launch week, interactive leaflets, posters, a knowledge quiz, team challenges, reminders, letters of management support, newsletters and fridge magnets to allow self-monitoring of physical activity. The intervention employed 18 of the 26 behaviour change techniques identified in a widely used taxonomy of behaviour change techniques (Abraham and Michie, 2008), but most frequently provided information about health and the consequences of physical activity, prompted intention formation, encouraged consideration of barriers, encouraged specific goal setting, provided rewards and environmental prompts,
provided social support and social comparison and encouraged self-monitoring and feedback. The intervention was delivered by an employee (the facilitator) from each worksite who received 7 h training from the research team and a step by step guide to the delivery in the form of an intervention manual. All materials (e.g. leaflets, posters, keeping track magnet) were delivered to facilitators before the intended launch of the intervention. As a guide, if implemented according to instruction to groups of up to 50 individuals, the intervention could be delivered in as little as 15 h by one facilitator over a 3-month period. The majority of the intervention was delivered to individuals or involved local prompts (e.g. posters), with the exception of the team challenges which involved groups of employees attempting to jointly achieve a goal (e.g. walking up a number of stairs equivalent to a mountain, e.g. Snowdon). This component was included to facilitate social support and social comparison.

Measures

Primary outcome
The primary outcome measure for this intervention and therefore for addressing RQ3 was self-reported moderate to vigorous physical activity. Total MET minutes of moderate and vigorous physical activity a week were calculated using the validated short form of the IPAQ (International Physical Activity Questionnaire; Hagstromer et al., 2006).

Fidelity measures (Table 1)
Intervention fidelity was assessed following the guidance of Dane and Schneider (Dane and Schneider, 1998). Information was gathered at both the individual level and worksite level. A detailed description of these measures is available from the corresponding author. Adherence measured the extent to which facilitators had delivered each of the nine intervention components (absent or present).

Quality of delivery of the intervention was assessed by asking facilitators to respond to the following question on a 10-point scale from 1 (not at all well) to 10 (extremely well): thinking about the 3-month intervention that you were asked to deliver, how well do you think this was actually delivered.

Facilitators were also asked to report the number of hours they spent implementing the intervention. This provided us with an indication of commitment to the intervention delivery and was based on the estimate (pre-specified in the training manual for facilitators) that a minimum of 5 h per month were required to fully implement the intervention.

Exposure assessed the extent to which recipients indicated that they had received each of nine intervention components (yes or no response in questionnaire).

Participant responsiveness measured perceptions of usefulness of each of the components of the intervention on a scale from 1 (not very useful) to 7 (very useful).

Participant engagement asked participants to indicate whether they had taken part in the team challenges (the active component of the intervention).

An organization with a consistently high score across the measures of fidelity was concluded to have delivered the intervention as intended. This allowed us to address RQ1.

Contextual factors
Three sources of information were used to develop an understanding of the organizational context in which the intervention was delivered. These were: (i) focus groups (see McEachan et al., 2008 for more detail); (ii) field notes made by RL and RM during worksite visits during the recruitment of organizations and at subsequent meetings prior to the delivery of the intervention; and (iii) monthly telephone calls (by C.J.) to facilitators in which they were asked a specific open-ended questions about the barriers to implementation and general progress with the delivery of the intervention. These data were coded by two independent researchers (R.L. and R.M.) and used to explain the quantitative data collected as part of the fidelity measurement. This provided us with data to address RQ2.

Procedure for data collection
Ethical approval for this study was granted by Institute of Psychological Sciences University of Leeds and Sheffield East NHS local research ethics committee. Worksites were recruited between October 2007 and May 2008 and invitations were sent to all employees within these worksites. Eligible employees (those meeting inclusion criteria) were then sent a T1 questionnaire and a health check appointment was arranged (see
Table 1: Measurement of fidelity (based on Dane and Schneider, 1998)

<table>
<thead>
<tr>
<th>Fidelity measure</th>
<th>Description</th>
<th>Level</th>
<th>How collected</th>
<th>How analysed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adherence</td>
<td>Index of how many components delivered</td>
<td>Worksite</td>
<td>Structured log completed by facilitators</td>
<td>Two members of the research team (R.M. and R.L.) used these sources to code the presence or absence of each of the nine intervention components to create an intervention adherence score. Where a worksite had more than one facilitator, scores were averaged. The mean Kappa score of agreement across the nine components was 0.83 (ranging from 0.57—team challenges, 1.00, management support letters, knowledge quiz). Disagreements were discussed and resolved</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Monthly progress telephone interviews with facilitators.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Researcher (C.J.) took field notes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure</td>
<td>Extent to which the recipients of the intervention report receiving</td>
<td>Individual</td>
<td>Dichotomous items in post-intervention postal questionnaire at T2</td>
<td>Items were summed to create a total exposure score for each participant. These were then aggregated to create a worksite exposure score for each worksite</td>
</tr>
<tr>
<td></td>
<td>intervention components</td>
<td></td>
<td>‘Tick if you noticed’ [component]:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Components: launch/quiz/leaflets/posters/keeping track magnet/letters or emails from management/letters or emails from staff members/newsletters/team challenges</td>
<td></td>
</tr>
<tr>
<td>Quality of delivery</td>
<td>Overall quality of delivery of the intervention (how 'well' were the</td>
<td>Worksite</td>
<td>Final monthly progress telephone interview with facilitators</td>
<td>Scores from the facilitator and the researcher were averaged to give an overall index of quality of delivery</td>
</tr>
<tr>
<td></td>
<td>intervention components delivered?)</td>
<td></td>
<td>‘Thinking about the 3-month intervention that you were asked to deliver, how well do you think this was actually delivered?’ (1 not at all well to 10 extremely well). Both facilitator and researcher (C.J.) provided a score</td>
<td></td>
</tr>
<tr>
<td>Participant</td>
<td>How useful participants felt the intervention components were</td>
<td>Individual</td>
<td>Items in post-intervention postal questionnaire at T2</td>
<td>A mean of these items was created for each component which were aggregated to provide a mean responsiveness score for each worksite</td>
</tr>
<tr>
<td>responsiveness</td>
<td></td>
<td></td>
<td>How useful did you find [each intervention component listed separately]? (1—not very useful to 7—very useful)</td>
<td></td>
</tr>
<tr>
<td>Participant</td>
<td>The number of participants reporting taking part in the team challenges</td>
<td>Individual</td>
<td>Tick if you used/participated in: team challenges</td>
<td>% was recorded for each worksite</td>
</tr>
<tr>
<td>engagement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
below for further details) irrespective of group. At the time of the health check, control participants received a brief leaflet describing ways of improving health through diet and activity.

On completion of health checks at a particular worksite, all intervention materials were dispatched and facilitators set a launch date for the intervention within 2 weeks. The second questionnaire (T2) was sent out post intervention (3 months after the intervention start date) for each matched pair of worksites, T3 questionnaires were sent out 3 months post intervention and T4 questionnaires were sent 9 months post intervention (total follow-up of 12 months). The post intervention questionnaire included the measures of exposure and responsiveness/engagement described above.

One of the researchers (C.J.) telephoned each facilitator on a monthly basis after the start of the intervention for a total of 3 months. The second questionnaire (T2) was sent out post intervention (3 months after the intervention start date) for each matched pair of worksites, T3 questionnaires were sent out 3 months post intervention and T4 questionnaires were sent 9 months post intervention (total follow-up of 12 months). The post intervention questionnaire included the measures of exposure and responsiveness/engagement described above.

One of the researchers (C.J.) telephoned each facilitator on a monthly basis after the start of the intervention for a total of 3 months. C.J. was blind to other fidelity information and was not part of the research team collecting data from the intervention and control sites. Facilitators were asked what components of the intervention had been delivered in the previous months and their views on how well the intervention was going. Responses were recorded in written form. Adherence, quality of delivery and commitment of the facilitator were each assessed in this way. If the researcher was unable to contact the facilitator, this was also recorded.

### Analysis

To address RQ1, we report descriptive statistics for the measures of fidelity across each of the five work organizations. We then explain these differences with reference to qualitative data elicited in telephone interviews with facilitators and information about the organizational context derived from field notes of site visits and focus groups (RQ2). These qualitative data were analysed using deductive content analysis (Elo and Kynga, 2008). A data categorization matrix was created and data were independently coded by RM and RL as barriers or facilitators to implementation for each worksite.

The analysis strategy for RQ3 was based on the multi-level modelling of the entire cohort reported in McEachan et al. (McEachan et al., 2011). In line with consort guidelines (Campbell et al., 2004), data were analysed using multi-level modelling to control for the fact that time points were clustered within participants who were clustered within worksites, allowing for variation at both an individual (different time points) and worksite level. When building our model, we retained those variables identified as important predictors of physical activity in previous analyses (e.g. gender, socio-economic status, season of measurement, health status, time point) (McEachan et al., 2011). We included an intervention variable (coded as 0—control, and 1—intervention) and an intervention * time interaction variable in recognition of our prediction that the effect of the intervention would be strongest immediately post intervention and decrease over time. Separate multi-level models were run for (1) organizations demonstrating ‘good’ fidelity and (2) organizations demonstrating ‘moderate’ or ‘poor’ fidelity. For an organization to be classed as having good intervention fidelity, scores on each of the criteria (adherence, exposure, quality and participant responsiveness) were required to be above the midpoint on the scale, together with a reported minimum of 15 h of intervention delivery time (as per protocol). A rating of ‘moderate’ fidelity was based on at least

<table>
<thead>
<tr>
<th>Table 1: Continued</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fidelity measure</strong></td>
</tr>
<tr>
<td><strong>Facilitator commitment</strong></td>
</tr>
<tr>
<td><strong>Context (background information)</strong></td>
</tr>
</tbody>
</table>
three of the scores being above the midpoint and at least half of the specified hours being devoted to the delivery of the intervention. ‘Poor’ fidelity rating was based on less than three of the scores being above the midpoint and less than half of the specified hours being devoted to the delivery of the intervention.

RESULTS

Research question 1
Table 2 shows the scores for each of our measures of fidelity for each organization. Across the measures of fidelity, the bus company scored poorly, as did the university worksites. Facilitators in these organizations delivered fewer of the intervention components (2.5 and 5 out of 9, respectively) and this is reflected in the lower exposure scores (among participants) and the poorer quality rating. Facilitators in the local council, hospital and government organizations delivered a higher number of components (approximately 7 of the 9 components) and this again is reflected in the higher exposure scores of participants in these organizations. However, the quality of the intervention delivery and the participant responsiveness and engagement scores were poor for the government organization compared with the local council and hospital, being highest among the former. Finally, none of the organizations, with the exception of the local council (hours = 23), spent the minimum recommended number of hours (15 h or 5 h per month) delivering the intervention. Thus, the bus company and University were classed as demonstrating ‘poor’, the government organization and the hospital as ‘moderate’ and the local council as demonstrating ‘good’ intervention fidelity.

Research question 2
Telephone interviews with facilitators and focus groups conducted with employees supplemented the quantitative data and provided important information about the context for, and process of, intervention delivery. Content analysis identified two main factors that represented barriers or levers for the fidelity of intervention delivery; ‘nature of the job’ and ‘commitment and engagement’. Both are described below.

Nature of the job
In some organizations, the intervention was much more difficult to deliver, practically, than in others. The facilitators in the bus company reported that bus drivers are rarely at the worksite. This made delivering the intervention, particularly the team challenges, very difficult and exposure to posters and reminders was not optimized. This insight is supported by the low exposure rating among participants in this organization which indicates that few of the drivers

| Table 2: Fidelity scores for each organization (Means, range and n) |
|----------------------------------|------------------|------------------|------------------|------------------|------------------|
|                                  | Local council    | Hospital         | Bus company      | Government organization | University |
| Adherence (0–9)                  | 7.28             | 7.57             | 2.50             | 7.6               | 5.00         |
| Range                            | 4–9              | 3–9              | 2–3              | 0                 | 4–6          |
| N (worksites)                    | 10               | 7                | 2                | 1                 | 2            |
| Exposure (0–9)                   | 7.05             | 6.85             | 2.63             | 6.99              | 3.0          |
| SD                               | 2.04             | 2.13             | 2.19             | 2.08              | 1.81         |
| N (participants)                 | 153              | 66               | 27               | 67                | 34           |
| Quality (1–10)                   | 5.86             | 5.89             | 0.50             | 4.30              | 2.50         |
| Range                            | 0–9              | 2–7.5            | 0–1              | 0                 | 1–4          |
| N (worksites)                    | 10               | 7                | 2                | 1                 | 2            |
| Participant responsiveness (1–7) | 4.54             | 4.21             | 2.67             | 3.91              | 3.69         |
| SD                               | 1.39             | 1.33             | 1.61             | 1.16              | 1.57         |
| N (participants)                 | 121              | 67               | 16               | 49                | 14           |
| Participant engagement (team challenges) | 42.1% (69) | 36.4% (37) | 3% (2) | 18.8% (13) | 0 |
| N (participants)                 | 164              | 102              | 66               | 69                | 0            |
| Commitment (number of hours)     | 23.38            | 7.88             | 0a               | 11.67             | 0a           |
| N (facilitators)                 | 11               | 9                | 3                | 4                 | 3            |

*aWe were unable to contact facilitators in these organizations to determine number of hours spent delivering the intervention. Where there are more facilitators than worksites, this is because some facilitators worked in pairs.*
received more than two of the nine key components. Staff at the university, even when employed within the same department, had a variety of different roles and were scattered across a large campus again hindering delivery of the intervention. Within the remaining organizations (hospital, government organizations, local council), delivering the intervention was more straightforward as employees did come together as teams, at least for some points during the day (e.g. in the staff room at lunch time), or worked in the same physical location for large parts of the day (e.g. a social services department).

**Commitment and engagement**

It was clear from the telephone interviews with facilitators that both the commitment of facilitators and the responsiveness/engagement of employees varied markedly. For example, the facilitators in the government organizations struggled to engage staff, the majority of whom (95%) were in managerial or professional roles and who reported finding some elements of the intervention (for example, the cartoons used in the leaflets) simplistic and childish.

In the bus company although the senior management were strongly supportive of the initiative at the outset, local facilitators reported that this did not manifest in practical support (time and resources) for the facilitators beyond release to attend the training day. Despite local commitment from the hospital facilitators, they received little managerial support and they reported difficulties in finding the time to deliver certain elements of the intervention (for example, the cartoons used in the leaflets) simplistic and childish.

Most facilitators in this group felt that their hard work was recognized by members of their team as well as their managers. Clear differences in commitment of facilitators and engagement of staff were apparent in the council organization. The differences in how well the intervention was delivered corresponded with the number of hours facilitators spent implementing the intervention (23 h in the local council organization, compared with 12 h for the Government organization and 8 h for the hospital).

Based on the local council organization being the only organization to score consistently well across the measures of fidelity, including the number of hours spent delivering the intervention, a finding supported and explained by the qualitative investigation, the analysis below tested the effect of the intervention among local council employees (who received the intervention as intended) compared with employees in all other work organizations (who did not receive the intervention as intended).

**Research question 3**

To explore the differential effect of the intervention for those employees who received the intervention as intended versus those who did not we ran two identical multi-level models (local council employees model 1a; all other organizations model 1b). The results of these models can be found in Table 3.

It can be seen from Table 3 that, as predicted, among local council employees the intervention was successful at increasing moderate–vigorous MET minutes of activity ($B = 846.92, p = 0.007, z = 2.69, 95\%CI: 228.96–1464.88$). The significant intervention * time interaction ($B = -220.56, p = 0.027, z = -2.22, 95\%CI: -415.45 to -25.66$) indicates this level falls off significantly over time. Immediately post intervention (T3), those in the local council intervention group are performing an additional 626 moderate–vigorous MET minutes of activity compared with the local council controls. By 9 months post intervention (T4), they are performing just 185 more moderate–vigorous MET minutes of activity more than the council controls. The analysis was repeated for all other worksites (Table 3, right-hand column). There was no effect of the intervention or the intervention * time interaction for these worksites.
DISCUSSION

The findings of the trial of the AME for activity worksite physical activity intervention found that it did not significantly increase self-reported levels of physical activity (McEachan et al., 2011), although it did produce changes in some indicators of health (systolic blood pressure and resting heart rate). In this article, we offer a more sophisticated and nuanced interpretation of these data based on an evaluation of the fidelity and context of intervention delivery. The local council, an organization which demonstrated high levels of commitment to the intervention, was the only organization to show good intervention fidelity. The fidelity of the intervention was either poor (bus company and university) or moderate (hospital and government organization) in the remaining work organizations. When data were reanalysed so as to compare the effectiveness of the intervention in the local council worksites (N = 20) and all other worksites (N = 24) separately, we identified a significant effect of the intervention on our primary outcome within the local council worksites only. This finding suggests that when the AME for Activity intervention is delivered as intended, it has the potential to lead to significant increases in activity among employees. Our results showed an increase in the equivalent of 156 min of moderate intensity activity per week (or 78 min vigorous intensity activity) immediately post-intervention: a level allowing individuals to meet the physical activity recommendations of at least 30 min moderate intensity activity on at least 5 days a week. By 9 months post intervention, this reduced to a level analogous with an additional 46 min moderate intensity activity per week. Two recent meta-analyses have identified that health benefits of physical activity show a dose response, with benefits greatest for those moving from no activity to moderate amounts of activity. For example, Sattelmair et al. (Sattelmair et al., 2011) found that individuals who engaged in the equivalent of 150 min per week of moderate-intensity leisure-time physical activity (minimum amount, 2008 US federal guidelines) had a 14% lower coronary heart disease risk (relative risk, 0.86; 95% CI: 0.77 to 0.96) compared with those reporting no leisure-time physical activity. Samitz et al. (Samitz et al., 2011), in a large meta-analysis of cohort studies, found that an additional 30 min of moderate-intensity activity per day was associated with a 5% lower risk in all-cause mortality. They also conclude that 1 h per week of physical activity (compared with 0 h) offers nearly two-thirds of the mortality reduction associated with 10 h per week. These findings suggest that even relatively small increases in the amount of physical activity being undertaken, particularly among those individuals who are largely sedentary, can have substantial health benefits.

The findings above suggest that this worksite physical activity intervention which was evaluated in a range of worksite organizations can be effective, but only given a specific set of ideal circumstances. Based on the fidelity ratings and contextual information we gathered, these ideal conditions include facilitator commitment (in part a function of support from management) which was predicted to approximate to 15 h...
across the 3-month intervention, employee receptiveness and engagement and the delivery of the intervention to a team within a physical space where they all met. These results support the idea that a one size fits all approach to complex interventions is not appropriate (Hawe et al., 2004). In this intervention, we built in some flexibility. For example, a range of possible team challenges were specified that all had the same function: to encourage social comparison and support, but varied in form (step challenges, stair climbing challenges etc.). This approach is recommended by Hawe and colleagues (Hawe et al., 2004). There was also some flexibility in the way particular components might be delivered, e.g. the launch event. However, this might not have been sufficient to support implementation in these very different work environments. Based on the ideas of Hawe et al. (Hawe et al., 2004), an intervention that specified the function of the component parts (e.g. team competition and social comparison and allowed the worksites/facilitators to devise an appropriate form of delivery) might have served to allow better tailoring of the intervention to the local context. A counter argument, of course, is that this places a greater burden on the local worksites to develop their own techniques for intervention that may not be informed to the same extent by theories of change.

Although some might argue that this type of post hoc fidelity analysis to help understand what the ‘ideal’ circumstances are for intervention delivery is counter-scientific, others argue that when a complex intervention is being tested there is a much greater blurring of the boundaries between evaluations of efficacy and effectiveness (Glasgow et al., 2003) and it is entirely appropriate to measure fidelity and to use this information to explain variations in effectiveness (Craig et al., 2008). In other words, although the trial reported here was a study of effectiveness of an intervention in real world conditions, it remains important to understand what circumstances the intervention is most likely to be effective. This would allow for more informed decision making about the commissioning and roll out of the intervention in other settings. The results of the present article demonstrate, as in other trials of complex interventions, that fidelity of intervention delivery can be poor (Hardeman et al., 2008) and that this may impact on the outcomes of the intervention (Hasson, 2010). Here, we have attempted, through our assessment of the organizational context in which the interventions were delivered, to understand the factors that explain variations in fidelity. This information is valuable because it can be used to make decisions about organizations and employees who are most, and least, likely to benefit from an intervention of this kind.

The finding here, and in other studies, that there is considerable variation in the quality of the delivery of interventions has other implications. It suggests that we may need to be more creative in the delivery of behaviour change techniques and rely less on people to deliver the components of the intervention and more on remote intervention delivery formats. For example, automatic text messages and on-line motivational interventions are increasingly being used to successfully promote health behaviour change (Webb et al., 2010). Another interpretation of the findings is that small environmental and social marketing changes in a worksite are insufficient to change physical activity and it is necessary to include intervention components with the appropriate level of organizational support to achieve change. For example, the team challenges were the least passive component of this intervention, but engagement in this component was low. While 42% of the participants in the government organization engaged in these challenges, an average of only 15% of participants in the other four organizations took part.

The measurement of fidelity is a challenging task. Our data suggest that reliance on simple quantitative metrics [e.g. how many components delivered (facilitators), how many components received (participants) may not be helpful in teasing apart the differences in the quality of delivery across the intervention sites or the response of the staff to these interventions]. For example, in our evaluation, making leaflets available to participants and putting up posters fulfilled the criteria for delivery of these two components. However, in one instance, a facilitator left leaflets in mail boxes, in another she personally handed out the leaflets and explained how they should be used. Quantitative measures of fidelity can struggle to pick up on these nuances. Qualitative and contextual analyses offer a more promising method of identifying how well the intervention was delivered and received in practice; however, the time and resources required may prohibit this as part of a large trial. Moreover, the nature of this data precludes entry into traditional analytic strategies.
(e.g. by including a ‘fidelity’ variable) and means that researchers need to find alternative ways of grouping participants into those who do receive the intervention successfully and those who do not. To reduce participant burden and based on the findings here, we would recommend that the amount of time spent delivering the intervention is a useful indicator of the quality of delivery and might also be a good proxy measure for organizational support/barriers.

The last 5 years have seen a growing interest in the measurement of fidelity as part of the process evaluation of complex interventions. More sophisticated conceptual models of fidelity measurement have been developed (Carroll et al., 2007) and are being tested (Hasson, 2010). The findings reported here support such advances. In fact, we would argue that without the measurement of fidelity and the understanding of contextual features that might impact fidelity, complex interventions might be erroneously dismissed as ineffective. Moreover, these analyses can help intervention developers to refine interventions so that they better fit the delivery context.

FUNDING

This research was funded by the BUPA Foundation medical charity (ref: 33a.05Lawton). The funders had no role in the conduct or reporting of the study.

REFERENCES


