Full Title:
Reviewing evidence on complex social interventions: appraising implementation in systematic reviews of the health effects of organisational-level workplace interventions.

Short Title:
Reviewing evidence on complex social interventions

Matt Egan¹ Clare Bambra,² Mark Petticrew³, Margaret Whitehead⁴
¹ Medical Research Council Social and Public Health Sciences Unit, University of Glasgow, UK.
² Department of Geography, Wolfson Research Institute, Durham University, UK.
³ Public and Environmental Health Research Unit, London School of Hygiene and Tropical Medicine, UK.
⁴ Division of Public Health, University of Liverpool, UK.

Corresponding author (including requests for reprints): Matt Egan, Medical Research Council Social and Public Health Sciences Unit, University of Glasgow, 4 Lilybank Gardens, Glasgow, G128RZ, UK. Email: M.Egan@sphsu.mrc.ac.uk and Matt.Egan@health.scot.nhs.uk. Tel. 0141 354 2967, Fax. 0141 354 2901.

Sources of Support: Department of Health Policy Research Programme (Public Health Research Consortium), Economic and Social Research Council and the Chief Scientist Office of the Scottish Executive Health Department.

Keywords
1. Systematic review
2. Public Health
3. Implementation

Word count: 3066 (text only)

License
The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in JECH editions and any other BMJPL products to exploit all subsidiary rights, as set out in our licence (http://jech.bmjjournals.com/ifora/licence.pdf)
Abstract

Objective: We appraised the reporting of intervention implementation in studies included in systematic reviews of organisational-level workplace interventions. Implementation is taken to include such factors as intervention setting, resources, planning, collaborations, delivery and macro-level socio-economic contexts. Understanding how implementation affects intervention outcomes may help prevent erroneous conclusions and misleading assumptions about generalisability, but implementation must be adequately reported if it is to be taken into account.

Design: Data on implementation was obtained from four systematic reviews of complex interventions in workplace settings. Implementation was appraised using a specially-developed checklist, and by means of an unstructured reading of the text.

Results: We identified and appraised 103 studies evaluating 4 types of organisational-level workplace intervention (employee participation, changing job tasks, shift changes and compressed working weeks). Many studies referred to implementation but reporting was generally poor and anecdotal in form. This poor quality of reporting did not vary greatly by type or date of publication. A minority of studies described how implementation may have influenced outcomes. These descriptions were more usefully explored through an unstructured reading of the text, rather than by means of the checklist.

Conclusions: Evaluations of complex interventions should include more detailed reporting of implementation and consider how to measure quality of implementation. Our checklist helped us explore the poor reporting of implementation in a more systematic fashion. In terms of interpreting study findings and their transferability, the more qualitative appraisals however appeared to offer greater potential for exploring how implementation may influence the findings of specific evaluations. Implementation appraisal techniques for systematic reviews of complex interventions require further development and testing.

Abstract word count: 263
BACKGROUND
The case has been made for providing policy-makers with synthesized, detailed and robust accounts of the implementation of effective interventions in order to make better progress in tackling population morbidities and inequalities.[1] Advocates of a staged approach to the development and evaluation of complex interventions have also stressed the importance of accurately defining interventions and promoting effective implementation.[2] Implementation refers to the design and delivery of interventions.[3-6]. The way an intervention is implemented may influence its outcomes, and evaluations that do not take this into account risk (for example) misinterpreting negative outcomes which result from poor implementation as evidence that interventions are inherently ineffective.[7,8]. We developed a tool to appraise the quality of reporting of implementation and applied this tool to four systematic reviews of complex intervention evaluations affecting the workplace.

Implementation and complex interventions
Researchers and policy-makers have called for evidence from systematic reviews of social interventions affecting so-called ‘upstream’ health determinants such as employment, housing, transport etc.[9-10] Such interventions are often complex and difficult to evaluate.[11-12] They may involve multiple, context-specific interventions and an unstandardised approach to implementation.[13]

In our experience of conducting systematic reviews of ‘upstream’ interventions, it is often difficult from the reporting of a complex intervention evaluation to determine a) what exactly the intervention entailed, b) whether the intervention was implemented fully or adhered to good practice guidelines; and c) whether there were confounding factors in the wider social context that would affect the outcome of the intervention.[14-21] This contrasts with reports of less complex intervention in which a) the intervention is clear (e.g. a specific drug), b) intervention delivery was prescribed through a detailed protocol, and c) at least some attempt was made from the planning stage onwards to identify and reduce bias associated with key confounders.

Implementation appraisal
Implementation appraisal is not a new concern.[22-29] Some systematic reviews have considered whether interventions were delivered as prescribed by the study protocol (‘treatment integrity’ or ‘program adherence’).[30] However, appraisal tools used by systematic reviewers usually focus on the methodological characteristics of primary studies rather than implementation issues.[30,31] Such tools often take the form of checklists, although the practice of using checklist scores to appraise studies is problematic, leading some to advocate alternative approaches.[31,32]

Systematic reviews that attempt to rigorously appraise the implementation of complex interventions are the exception rather than the rule. A recent review of community-based injury-prevention initiatives which included appraisals of evidence on implementation found that reporting of implementation was poor.[33]

We developed and incorporated an appraisal checklist into four systematic reviews of organisational-level workplace interventions, along with a less structured exploration of textual accounts of implementation in the included studies.[17-20] The checklist covered reporting of intervention design (including whether or not interventions were specifically designed to affect employee health), target population, delivery, psychosocial factors and the characteristics of population sub-groups differentially affected by the interventions. Our primary aim was to appraise the reporting of implementation in primary studies, our study also considered whether or not there was evidence to suggest that higher standards of reporting were an indication of greater methodological rigour.[34]

METHODS
The four reviews that incorporated our appraisal tool synthesized evidence on the health effects of (i) workplace interventions to increase employee control and participation in decision-making[17]; (ii) changes to team structures and work allocation affecting employees’ day to day tasks[18]; (iii) the health effects of instigating compressed working weeks[19] and
(iv) shift work interventions[20]. Table 1 summarises the intervention types in these reviews. Their methods and outcomes have been described elsewhere.[17-20]

Our original checklist contained 28 criteria. These criteria were adapted from a number of sources, particularly Rychetnik et al, whose work had prompted our initial interest in implementation.[3,4,27,35-38] Two reviewers (ME and CB) piloted this checklist independently using 12 studies (taken from the participation and task restructuring reviews). On comparing their pilot appraisals, the reviewers agreed that the checklist had been difficult to interpret and apply consistently, and they criticised both its content and face validity. The reviewers ascribed these problems to the checklist being unclear (often because criteria had been adapted from other contexts). The pilot checklist also coped poorly with ambiguities in reports of implementation (often, the answers to specific checklist criteria were implied rather than explicitly stated in the brief reports of implementation we identified— and it was often difficult to agree on the point at which reviewers could distinguish mere implication from reported fact). We decided it would be preferable to work with a smaller number of broader criteria, and hence we shortened the checklist.

The final checklist included 10 criteria (response: yes/no). Studies were categorised by an implementation appraisal score (out of 10 – one point for the presence of each criteria), distinguishing the ‘lowest’, ‘intermediary’ and ‘higher’ scoring studies. The checklist is presented in table 2.
<table>
<thead>
<tr>
<th>Interventions</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Employee participation review [15]</strong></td>
<td></td>
</tr>
<tr>
<td>Employee participation committees</td>
<td>Employee representative committees (described variously: eg. action teams, problem solving committees, etc). These often focus on identifying and suggesting ways of overcoming workplace stressors. Some committees are led by external facilitators and/or have managerial representation.</td>
</tr>
<tr>
<td>Flexible working hours</td>
<td>Employees are given more control over choosing their working hours</td>
</tr>
<tr>
<td>Participation and individual level interventions</td>
<td>Employee committees combined with individual-level health promotion, education and behaviour programmes: such as anti-smoking or physical activity interventions, and training in relaxation techniques, stress reduction, and communication skills.</td>
</tr>
<tr>
<td>Participation and ergonomic interventions</td>
<td>Participatory committees combined with ergonomic interventions: i.e. attempts to reduce physical discomfort and workplace injuries by modifying physical environments (including technological improvements) and advising on posture and lifting.</td>
</tr>
<tr>
<td><strong>Task restructuring review [16]</strong></td>
<td></td>
</tr>
<tr>
<td>Production line</td>
<td>Production line interventions that increase the variety of tasks performed by a worker, increase the skills utilised and place more responsibility on individual workers.</td>
</tr>
<tr>
<td>Primary nursing</td>
<td>Increasing the skills utilised by workers by increasing the variety of work tasks. Primary nursing and personal caregiving are patient orientated care systems in which each patient is assigned to an individual nurse/carer, the nurse/carer takes 24 hour responsibility for the care of that patient including the planning and quality of the care provided.</td>
</tr>
<tr>
<td>Team working</td>
<td>Workers are given more collective responsibility and decision making power within the team but responsibility is not shared and supervisory structures remain in place.</td>
</tr>
<tr>
<td>Lean Production</td>
<td>Employee workloads are maximized, wasted time is reduced, tasks are distributed within the team, and work standards are determined by the employees themselves rather than solely by management.</td>
</tr>
<tr>
<td>‘Just in time’</td>
<td>‘Just in time’ requires that products are made ‘just in time’ to be sold — no stockpiling of products. Work groups, not individuals, are given autonomy and responsibility for specific tasks in order to achieve the required production flow.</td>
</tr>
<tr>
<td>Autonomous work groups</td>
<td>Autonomous workgroups are characterised by employee self-determination and involvement in the management of day to day work (including control over pace, task distribution, and training and recruitment).</td>
</tr>
<tr>
<td><strong>Compressed working week review [17]</strong></td>
<td></td>
</tr>
<tr>
<td>Compressed working week (CWW)</td>
<td>Hours worked per day are increased whilst the days worked are decreased in order to work the standard number of weekly hours in less than five days. Eg. The 12 hour CWW involves four 12 hour shifts (day, night) over four days with three or four days off. Under a 10 hour CWW, four 10 hour shifts are worked followed by three days off. The Ottawa system consists of three or four 10 hour morning or afternoon shifts for four days and two days off. Then a block of seven 8 hour nights and six days off.</td>
</tr>
<tr>
<td><strong>Shift work review [18]</strong></td>
<td></td>
</tr>
<tr>
<td>Changes affecting shift rotation</td>
<td>Eg. 1. Changing from slow to fast rotation: a change from 6 or 7 consecutive shifts of the same type to a maximum of 3 or 4. Eg. 2. Changing from backward (night, afternoon, morning) to forward (morning, afternoon, night) rotation or vice versa.</td>
</tr>
<tr>
<td>Changes affecting night work</td>
<td>Eg. 3. Changing from a rotating shift system to a permanent shift system.</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Later start and finish times</td>
<td>Starting and finishing shifts one hour (in the studies we identified) later.</td>
</tr>
<tr>
<td>Weekend shift changes.</td>
<td>Continuous (weekends on) to discontinuous shift system (weekends off) or vice versa</td>
</tr>
<tr>
<td>Decreased hours</td>
<td>Decrease in shift length (e.g. from 8 hours to 6 hours).</td>
</tr>
<tr>
<td>Self-scheduling</td>
<td>Self-scheduling enables individual shift workers to have some control over which shifts they work, their start times, or when their rest days occur.</td>
</tr>
<tr>
<td>Combined shift interventions</td>
<td>Combinations of the above, most usually shift rotation combinations e.g. change to fast and forward rotation</td>
</tr>
<tr>
<td>Theme</td>
<td>Checklist Question for Workplace Reviews</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1. Motivation</td>
<td>Does the study describe why the management decided to subject the employee population to the organisational change?</td>
</tr>
<tr>
<td>2. Theory of Change</td>
<td>Was the intervention design influenced by a theory of change describing the proposed pathway from implementation to health outcome?</td>
</tr>
<tr>
<td>3. Implementation Context</td>
<td>Does study provide any useful contextual information relevant to implementation of the intervention (eg. political, economic or managerial factors)?</td>
</tr>
<tr>
<td>4. Experience</td>
<td>Does study establish whether those implementing the intervention had appropriate experience (eg. Had the implementers conducted similar interventions before; or if managers/employees were involved, were they appropriately trained for new roles)?</td>
</tr>
<tr>
<td>5. Planning consultations</td>
<td>Is there a report of consultation/collaboration processes between managers, employees and any other relevant parties during the planning stage?</td>
</tr>
<tr>
<td>6. Delivery collaborations</td>
<td>Is there a report of consultation/collaboration processes between managers, employees and any other relevant parties during the delivery stage?</td>
</tr>
<tr>
<td>7. Manager support</td>
<td>Were on-site managers/supervisors supportive of the intervention (eg. Do authors comment on manager’s views of intervention)?</td>
</tr>
<tr>
<td>8. Employee support</td>
<td>Were employees supportive of the intervention (eg. do authors comment on employee’s views of intervention)?</td>
</tr>
<tr>
<td>9. Resources</td>
<td>Does study give information about the resources required in implementing the intervention (eg. time, money, people, equipment)?</td>
</tr>
<tr>
<td>10. Differential effects and population characteristics*</td>
<td>Does the study provide information on the characteristics of people for whom the intervention was beneficial, and the characteristics of those for whom it was harmful or ineffective?</td>
</tr>
</tbody>
</table>

*Note that whilst consideration of differential effects involves analysis of outcomes, it also provides contextual information on the characteristics of population sub-groups. This is of interest when exploring the transferability of research findings and the mechanisms by which some interventions affect different types of people in different ways. Hence we regard this issue as relevant to explorations of implementation and context (as well as to outcome analysis).
Two reviewers (ME and CB, or CB and MP) independently applied the checklist to all the studies included in the four systematic reviews. Differences were resolved through consultation. We then used crosstabulations to explore relationships between quality appraisal scores from our checklist and data on evaluation study designs, and with psychosocial and health outcomes (previous studies have suggested that more rigorous evaluations may be less likely to report positive outcomes).[34] We also explored whether reporting of implementation differed by date of publication (i.e. whether or not reporting has improved in recent years), and type of publication (i.e. whether reporting is better or worse in peer reviewed journals compared to other forms of publication).

Reported text that described implementation processes were also extracted from each study by one reviewer and checked by another to aid a less structured analysis of reporting of implementation for each review. We considered relevant data first on a case-by-case basis and explored the interactions between reported planning and implementation characteristics, contexts and outcomes. We discussed patterns and idiosyncrasies across different studies and synthesised key findings using a narrative approach. From this less structured process we gained some insights into how a minority of authors explained outcomes in terms of implementation characteristics.

RESULTS
Implementation appraisals were conducted on a total of 103 studies (references can be obtained from the original reviews[17-20]). Twenty-one studies were identified in the task restructuring review, 18 studies in the employee participation review, 40 studies in the compressed working week review and 26 studies in the shift work review.[17-20] Two studies appeared in two reviews. In Table 3, the numerical implementation scores are summarised for all studies and in Table 4 examples of summaries of implementation appraisals are presented for the higher scoring studies from each of the four reviews.

Summary of Implementation Appraisals
Most studies achieved low scores (see Table 3). The median score was 3 out of 10 (range = 0 to 7; lower and upper quartiles = 1 and 4). This varied slightly between reviews (from 2 to 4). The median score was 3 for studies published between 1996-2000 and 2 for studies published between 1991-5 and before 1991.

As few studies achieved a high implementation score, we have categorised the studies as follows: fourteen ‘higher’ scoring studies (scoring ≥5 in our implementation appraisal), Thirty-eight ‘intermediary’ scoring studies (scoring 3 or 4 in our appraisal), and fifty-one ‘lowest’ scoring studies (scoring <3).

The most commonly reported implementation themes were ‘motivation for intervention’ (Table 2, criteria 1 - appearing in 76% of included studies) and employee support of the intervention (criteria 8: appearing in 54% of the studies). All the other themes were reported in less than a third of the total studies. Criteria 10 (differential effects/population characteristics) was only reported in 8% of the studies, whilst no study described resourcing, costs or cost-benefits of interventions (criteria 9).

Table 3: Numerical summary of the results of the implementation appraisal checklist

<table>
<thead>
<tr>
<th></th>
<th>All Reviews (Excluding Duplicates)</th>
<th>Task Variety</th>
<th>Employment Participation</th>
<th>Compressed Working Week</th>
<th>Shift Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of studies</td>
<td>103</td>
<td>21</td>
<td>18</td>
<td>40</td>
<td>26</td>
</tr>
<tr>
<td>Mean implementation score</td>
<td>2.6</td>
<td>2.6</td>
<td>3.4</td>
<td>2.3</td>
<td>2.4</td>
</tr>
<tr>
<td>Median implementation score</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>Studies with lowest implementation score (0,1,2)</td>
<td>51</td>
<td>9</td>
<td>4</td>
<td>26</td>
<td>13</td>
</tr>
<tr>
<td>Citation</td>
<td>Intervention, study design &amp; population details</td>
<td>Implementation Details*</td>
<td>Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------</td>
<td>-------------------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Employee participation review</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Park et al (2004).[39]</td>
<td>Participatory committee to improve team communication and cohesiveness, work scheduling, conflict resolution, and employee rewards. Prospective repeat-cross sectional study. All employees, retail store, USA.</td>
<td>(1) Researcher initiated to act as a buffer against the adverse effects of recession and uncertainty. (2) Explicitly inspired by theories of psychosocial work reorganisation (3) Implementation took place during a period of recession and uncertainty (5) Professional facilitator assisted with delivery (6) Employee representative liaised with management and employees (10) Psychosocial improvements for black and Hispanic, but not white, employees</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mikkelsen and Saksvik (1999).[40]</td>
<td>Conference on working conditions followed by supervisor and employee workgroups meeting two hours a week, 9 times: intervention was moderated by consultants Prospective cohort study with comparison group. Manual and clerical workers, Post office depot, Norway.</td>
<td>(1) To improve workplace health (2) Explicitly inspired by theories of psychosocial work reorganisation (3) Company undergoing downsizing for financial reasons. (5) Researchers, managers and Union representatives helped design the intervention. (7) Management supported the intervention (8) Union representatives supported the intervention but authors report that in one department, the intervention was neither successfully implemented, nor effective, because steering group members lost interest, and personnel were relocated or made redundant.</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Task restructuring review</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wall et al (1990) [41]</td>
<td>Increased operator control on production line. Prospective cohort Manual workers, Factory floor, UK.</td>
<td>(1) Introduced to increase staff performance (2) Explicitly inspired by theories of psychosocial work reorganisation (4) Training was provided (6) Representative of employees of all grades and the researchers were involved in a working party overseeing the implementation of the intervention. (8) Some employees were resistant to the intervention</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wall et al (1986) [42]</td>
<td>Autonomous work groups Prospective cohort Manual and shop floor supervisors, Factory floor, UK.</td>
<td>(1) Intervention occurred in a purpose built factory which was designed with increasing factory floor responsibility and job redesign in mind. (2) Underpinned by theory about job redesign (4) Training on intervention was provided (6) Researchers were not involved in the design or implementation of the intervention. (8) Employee support for the intervention was mixed.</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Compressed working week review</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams, (1992).[43]</td>
<td>Six/seven 8hr shifts, two/four days off to Three/four 12hr shifts, two to seven days off Prospective cohort Operators, Chemical Plant, USA.</td>
<td>(1) Intervention initiated by staff to improve their work/life balance. 90% of staff were dissatisfied with the old system and other local factories had started using 12hr shifts. (3) Pressure from staff led to a management review of different shift schedules with the most popular schedule adopted. 83% voted for the implemented system. (4) Managers went on ‘fact finding’ visits to 12hr factories to learn about safety implications and how best to</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author(s)</td>
<td>Description</td>
<td>Changes to shift work</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>----------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wootten, (2000a, 2000b)[44, 45]</td>
<td>7.5hr to 12hr shifts. Retrospective cohort Nurses, Hospital, UK.</td>
<td>(1) Introduced to improve staff health and well being. (2) Explicitly inspired by theories of psychosocial work reorganisation (4) Colleagues who had implemented similar changes elsewhere were consulted. (5) Staff were consulted over the change (6) Implemented via collaborations between staff, supervisors and unions. (7) Managers were supportive of the intervention. (8) Union was supportive of the intervention.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brinton, (1983)[46]</td>
<td>Five 8hr shifts, two days off to Four 12hr shifts, Three/Four days off. Retrospective repeat cross section Wood yard workers, Paper Mill, USA.</td>
<td>(1) Workers' idea (2) Explicitly inspired by theories of psychosocial work reorganisation (3) Flexibility needed by both Union and management to get the new system implemented. (5) New system designed and agreed with the Union. (6) Committee set up between Union and managers to monitor safety on the new system. (7) Supported by supervisors. Company management agreed that they would implement the change if a majority of the workforce supported it. (8) Supported by the Union</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gauderer &amp; Knauth, (2004)[47]</td>
<td>Self-scheduling of shifts Prospective cohort with comparison group Bus drivers, Public transport depot, Germany.</td>
<td>(1) Introduced to improve the ergonomic design of shifts by involving drivers in their own scheduling. (2) Explicitly inspired by theories of psychosocial work reorganisation (4) Those involved in implementation collected information from other companies that had experienced a similar intervention. Workers attended training workshops to learn how to design their own schedules. (5) Staff, managers and researchers involved in designing the system. (8) Workers' council voted in favour of the change and at end of the 1 year trial period workers voted to keep the new system.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kandolin &amp; Huida, (1996)[48]</td>
<td>Slow to fast rotation; Backward to forward rotation; Self-scheduling of shifts. Prospective cohort with comparison group Midwives, Hospital, Finland.</td>
<td>(1) Introduced to reduce fatigue by decreasing the number of ‘quick returns’ and changing to a forward rotation. To increase role of midwives in their own scheduling. (3) Only a third of the midwives said that they had actually experienced a change to forward rotation, but more experienced less quick returns on the new system. A higher proportion of staff now participated in their own scheduling. (4) Managers had previous experience. (5) Managers carried out the rescheduling (6) Managers carried out the rescheduling (8) 55% said they preferred the old system due to the longer continuous free time.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Type of Publication**

Forty-nine included studies were published in peer review health journals, 41 in other peer review journals (mainly social science, occupational and managerial studies journals) and 13 in edited books or theses. Twelve per cent of articles from health journals received higher implementation scores compared to 15% of studies from both other journals and from books or theses. Forty-seven percent of articles from health journals received lower implementation scores compared to 51% of studies from other journals and 54% from books or theses.
Implementation and Study Design
Implementation appraisal scores were not useful predictors of robust study designs. We identified 32 prospective cohort studies with appropriate controls and have classed these as the most robust study designs: 36% of the studies with ‘higher’ implementation scores were “most robust”, compared to 45% of studies with intermediary scores, and 20% of studies with low scores.

Implementation and health effects
All 103 studies included in the reviews evaluated at least one health outcome.[17-20] We have categorised the studies as follows: (1) those which reported at least one positive health outcome and no negative outcomes (n = 47); (2) those which reported at least one negative health outcome and no positive outcomes (n = 14); and (3) those which report conflicting health outcomes (positive and negative) or reported little/no change in all the health outcomes measured (n = 42).

We found no conclusive evidence that better reporting of implementation might be associated with positive health outcomes. There was a similar range of implementation scores for both the 47 studies with positive outcomes (47% scored <3, 40% scored 3 or 4, and 13% scored ≥5) and the 42 studies with conflicting/little change in outcomes (45% scored <3, 40% scored 3 or 4, and 14% scored ≥5). Fourteen studies reported negative outcomes of which 84% scored <3, 15% scored 3 or 4, and none scored ≥5 on the implementation checklists.

Unstructured appraisals of implementation
We extracted textual data on implementation from all the included studies for less structured, more qualitative appraisals. However, we focus on the 14 studies with negative health outcomes.

Implementation reporting tended to be brief and anecdotal. It was often unclear how authors had obtained their information about implementation and whether they had taken steps to avoid bias or error. These (important) objections aside, our more qualitative approach to implementation appraisal did appear to uncover potential explanations for how implementation characteristics of some studies may have contributed to negative outcomes.

In the participation review we found that the only two studies with negative health outcomes evaluated participatory interventions that had been implemented in workplaces undergoing organisational downsizing.[17] We found that in the “task variety” review, negative health outcomes were more likely to result from interventions that were motivated for business reasons (managerial efficiency, productivity, cost etc) rather than by employee health concerns.[18] However the studies identified for the compressed working week and the shift work reviews provide evidence of both positive, negative or ‘little change’ outcomes resulting from interventions regardless of whether they were motivated by business concerns, health concerns or pressure from employees.[19,20]

Discussion
Promoting effective implementation is regarded as a key stage in the design and evaluation of complex interventions, and syntheses of evidence from such evaluations should incorporate data on implementation.[1,2] We incorporated implementation data into four systematic reviews of workplace interventions, using both a specially-developed checklist for measuring reporting of intervention design and implementation, and a more qualitative approach to assessing such reports. We found that reporting of implementation was generally poor. Our experience led us to reflect upon whether a checklist is the best tool for appraising implementation, particularly as our qualitative approach was easier to conduct and, we conclude, more useful than the checklist-based approach.

Quality of Reporting
In most cases, authors of included studies presented brief and anecdotal reports of implementation. We identified few descriptions of how authors obtained information about implementation, whether any prior code of good practice existed against which the quality of
implementation could be measured, and whether any attempts were made to prevent biased reporting of implementation. Roen and colleagues’ recently published details of their attempts to appraise the implementation of injury prevention interventions, which identified similarly poor standards of reporting.[33] However, they found that studies with methodologically stronger designs tended to provide poorer descriptions of implementation. We found no clear evidence of this relationship in our reviews.

Our checklist-based appraisals did find that most included studies provided some information about what motivated the implementers to deliver the intervention, and whether employees supported them. However, data on cost-effectiveness and differential effects on population subgroups were rarely reported, despite the widely-stated view that research to inform public health policy and practice should provide evidence on these issues[11,12]. We also found that reporting of implementation varied little by year or type of publication.

We also took a less structured (and less score-focused) approach to identifying reported data on implementation appraisal. This did identify some potential explanations for how implementation may have affected psychosocial and health outcomes: e.g., organisational downsizing, lack of management support, and the aim of increasing individual productivity without regard to employee wellbeing were all offered as explanations for negative results. These issues were usually described anecdotally within the studies, yet they often provided the most plausible explanations for negative outcomes available to reviewers.

We note that other systematic reviewers have employed more qualitative approaches to implementation appraisal.[29] Our own experience now leads us to advocate variations on this approach, perhaps as an adjunct to use of the use of implementation checklists.

**Limitations**

More methodological work is required to develop our approach (and alternative approaches[33]) to implementation appraisal: in particular to test inter-rater reliability and validity (the lack of such tests is a limitation to this study). We would focus our efforts on developing and testing qualitative implementation appraisal methods as we believe these may potentially provide greater insights than a checklist-based approach.

We do not rule out the possibility that a systematic review checklist could be developed to assist with implementation appraisals, but in our experience this approach was problematic. The checklist we developed assessed reporting of implementation - this is not the same as appraising the quality of implementation, but good reporting is one pre-requisite for such an appraisal.[1] Our checklist therefore helped to demonstrate the urgent need for improved reporting but did not help us to understand how implementation affected outcomes.

It should also be remembered that this paper only examines reviews of employment interventions. The generalisability of these findings depends on the degree to which employment researchers tend to report implementation differently or similarly to researchers working in other fields.

We also note that our implementation checklist analysis explored psychosocial and health outcomes. Whilst it is legitimate for public health researchers to be particularly interested in outcomes relevant to their field, we recognise that complex interventions such as those included in our reviews often have other outcomes (e.g. financial, managerial) of equal or greater importance to the implementers than health outcomes.

**Quality of Implementation**

As stated above, our checklist was not designed to appraise quality of implementation. Even if we could have directly appraised the quality of implementation, the checklist scores would still have been problematic. Summary appraisal scores reveal the number and variety, but not the importance of reported implementation characteristics. It may only take one flaw in the implementation to cause an intervention to fail, so a high intervention score is no guarantee of effectiveness.[32]
Furthermore, the development of a detailed checklist for measuring quality of implementation requires an *a priori* knowledge of the criteria that will distinguish well implemented interventions from poorly implemented interventions.[1] This may be feasible in some areas of research, when there is a strong consensus regarding standards of best practice, but that consensus does not always exist for every type of intervention. We attempted to develop such a list but quickly realised that the included interventions were too varied and there was often no clear way of prescribing in detail what constituted good or bad practice. We suspect that this uncertainty over best practice may increase with the complexity of an intervention, particularly if the intervention is flexible in design and context specific.

For example, what resources are sufficient to adequately resource an intervention? Is collaboration with employees always desirable, or can interventions achieve similar or better results if they are imposed by managers taking a ‘strong leader’ approach? It may be desirable for people managing implementation processes to have appropriate experience, but ‘appropriate’ needs to be defined: must managers have prior experience of delivering specific interventions, or is their general role in management to be regarded as appropriate enough?

The answers to all these questions seem to us to depend on the intervention and specific circumstances.

**Conclusion**

Guidance on improving reporting of implementation has been published elsewhere along with the recommendation that “adding simple criteria to reporting standards will significantly improve the quality and usefulness of published evidence and increase its impact on public health program planning.”[1] Such guidance may need to be adapted to suit specific interventions and our own checklist includes criteria that may be useful when reporting workplace interventions. However, we would advise caution against assuming that appraising the implementation of complex interventions is a simple matter. The appraisal tool we developed – like other appraisal tools – could only assess how well the implementation process was reported, rather than the quality of the process itself, and in most cases, reporting was poor. We also lacked detailed criteria on what constitutes well-implemented workplace interventions that could safeguard or improve employee health.

Nonetheless information on implementation and context is crucial for a nuanced assessment of the impact of complex interventions. Improvements in the reporting and appraisal of such information are overdue.
ACKNOWLEDGEMENTS
This work was partly funded by ESRC grant no: H141251011 (as part of the ESRC Centre for Evidence-based Public Health Policy), and as part of the Department of Health Policy Research Programme’s Public Health Research Consortium. (www.york.ac.uk/phrc). Mark Petticrew was funded by the Chief Scientist Office of the Scottish Executive Department of Health while most of this work was carried out. The views expressed are those of the authors not the funders.

COMPETING INTERESTS
The authors declare no competing interests.

CONTRIBUTORS
ME planned the study, collected and analysed data, is lead author and guarantor. CB, MP, MW and HT assisted in all aspects of the study including writing-up.
What is known on this subject

- Systematic reviews have been advocated as a means of identifying and appraising evidence on the health effects of complex social interventions.
- Implementation should be an important feature of these types of systematic review.
- However, to date, reviewers have often placed more emphasis on appraising the methodological characteristics of evaluations rather than the intervention itself and how it is implemented.
- Implementation appraisal tools have therefore remained relatively under-developed in the systematic review literature, especially as regards more complex social interventions.

What this study adds

- Aside from highlighting the lack of reporting of implementation issues in primary studies, this study also reveals which aspects of implementation were most commonly reported.
- We do not recommend our checklist as a means of appraising how implementation influences the outcomes of interventions. Implementation appraisal may be best achieved through less structured and more qualitative approaches.
- Future evaluations of implementation need to incorporate more contextual, qualitative information.

Policy Implications

- Primary studies and systematic reviews that are intended to influence policy and practice risk making erroneous recommendations if the quality of intervention implementation is not more robustly appraised.
References


