

# BMJ Open Evaluation of the fidelity of an interactive face-to-face educational intervention to improve general practitioner management of back pain

Simon D French,<sup>1,2</sup> Sally E Green,<sup>1</sup> Jill J Francis,<sup>3</sup> Rachelle Buchbinder,<sup>1,4</sup> Denise A O'Connor,<sup>1</sup> Jeremy M Grimshaw,<sup>5,6</sup> Susan Michie<sup>7</sup>

**To cite:** French SD, Green SE, Francis JJ, *et al*. Evaluation of the fidelity of an interactive face-to-face educational intervention to improve general practitioner management of back pain. *BMJ Open* 2015;**5**:e007886. doi:10.1136/bmjopen-2015-007886

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-007886>).

Received 9 February 2015  
Revised 29 May 2015  
Accepted 15 June 2015



CrossMark

For numbered affiliations see end of article.

## Correspondence to

Dr Simon D French;  
[simon.french@queensu.ca](mailto:simon.french@queensu.ca)

## ABSTRACT

**Objectives:** Implementation intervention effects can only be fully realised and understood if they are faithfully delivered. However the evaluation of implementation intervention fidelity is not commonly undertaken. The IMPLEMENT intervention was designed to improve the management of low back pain by general medical practitioners. It consisted of a two-session interactive workshop, including didactic presentations and small group discussions by trained facilitators. This study aimed to evaluate the fidelity of the IMPLEMENT intervention by assessing: (1) observed facilitator adherence to planned behaviour change techniques (BCTs); (2) comparison of observed and self-reported adherence to planned BCTs and (3) variation across different facilitators and different BCTs.

**Design:** The study compared planned and actual, and observed versus self-assessed delivery of BCTs during the IMPLEMENT workshops.

**Method:** Workshop sessions were audiorecorded and transcribed verbatim. Observed adherence of facilitators to the planned intervention was assessed by analysing the workshop transcripts in terms of BCTs delivered. Self-reported adherence was measured using a checklist completed at the end of each workshop session and was compared with the 'gold standard' of observed adherence using sensitivity and specificity analyses.

**Results:** The overall observed adherence to planned BCTs was 79%, representing moderate-to-high intervention fidelity. There was no significant difference in adherence to BCTs between the facilitators. Sensitivity of self-reported adherence was 95% (95% CI 88 to 98) and specificity was 30% (95% CI 11 to 60).

**Conclusions:** The findings suggest that the IMPLEMENT intervention was delivered with high levels of adherence to the planned intervention protocol.

**Trial registration number:** The IMPLEMENT trial was registered in the Australian New Zealand Clinical Trials Registry, ACTRN012606000098538 ([http://www.anzctr.org.au/trial\\_view.aspx?ID=1162](http://www.anzctr.org.au/trial_view.aspx?ID=1162)).

## BACKGROUND

Many people do not receive effective and safe healthcare.<sup>1–3</sup> To improve the uptake of

## Strengths and limitations of this study

- This study has demonstrated that it is feasible to undertake a rigorous fidelity assessment of the delivery of an implementation intervention by coding it according to specific behaviour change techniques (BCTs).
- The reliability of the study findings is limited due to a lack of data about the coverage of planned BCTs during the intervention workshop small group discussions.
- This study evaluated only one aspect of intervention fidelity, the implementation intervention delivery, with actual general medical practitioner behaviour not being measured.

evidence into clinical practice, with the aim of improving patient outcomes, a wide variety of complex implementation interventions have been designed. However, implementation interventions have had limited and varied effects.<sup>4</sup> There is a growing body of research demonstrating that interventions to change health professional practice behaviour in line with research findings should be targeted at specific health professional clinical practice behaviours and be underpinned by appropriate theory.<sup>5</sup> The evaluation of intervention fidelity complements this approach by identifying whether interventions delivered were faithful to the planned intervention specified in the protocol; high fidelity increases the likelihood that evaluations produce valid findings and that effective interventions achieve their aims.

Intervention fidelity refers to both the methodological strategies used to monitor and enhance the reliability and validity of delivery of interventions,<sup>6</sup> and the extent to which an intervention as delivered is faithful to the intervention as planned.<sup>7</sup> Fidelity of interventions influences the internal and the

external validity of trials evaluating intervention effects. Regarding internal validity, if an intervention was effective but fidelity was low, the effect may be due to unknown factors that were unintentionally added to, or omitted from, the intervention. If fidelity was low and the intervention was found to be ineffective, it could not be known if this was due to an ineffective intervention or to poor delivery of an effective intervention. External validity is compromised by low fidelity by not being able to draw conclusions about the generalisability of factors found to be responsible for any effects. Hence, without information about intervention fidelity, potentially effective interventions may be prematurely discarded or incorrectly applied, and ineffective interventions may be inappropriately implemented.<sup>8</sup>

The need to examine intervention fidelity has been highlighted in the CONSORT statement for reporting complex, non-pharmacological interventions in controlled trials.<sup>9</sup> However, intervention fidelity in implementation research has not been extensively studied or reported.<sup>10</sup> While there have been a number of previous studies of implementation fidelity in behaviour change research, these studies have been mostly undertaken to assess intervention fidelity of interventions to change the behaviour of people with health conditions<sup>6 11 12</sup> or those at risk of ill-health.<sup>13–15</sup> Only a few studies have assessed intervention fidelity of interventions aiming to change healthcare practitioner clinical behaviour<sup>16</sup> and when an assessment of intervention fidelity has been made, it has often not been undertaken systematically.<sup>17</sup>

Multiple intervention fidelity frameworks exist;<sup>18</sup> however, there is currently no standard method for assessing intervention fidelity in implementation research. Bellg *et al*<sup>7</sup> developed a fidelity framework for use in behaviour change research, and a set of guidelines and recommendations for best practice that cover the categories of: Design, Training, Delivery, Receipt, and Enactment. ‘Design’ refers to the theoretical foundations of the development of the intervention and the dose of the delivery. ‘Training’ refers to training of the providers who deliver the intervention, and monitoring and sustaining their skills. ‘Delivery’ refers to the processes to monitor and improve delivery of the intervention. ‘Receipt’ includes the processes to monitor and improve the ability of intervention recipients to understand and perform skills and cognitive strategies delivered during the intervention. ‘Enactment’ refers to the processes to monitor and improve the ability of intervention recipients to perform intervention-related behavioural skills and cognitive strategies in relevant real-life settings.

As well as the need to establish valid fidelity frameworks, there is also a need to establish routine procedures for monitoring the fidelity with which behaviour change interventions, and other complex interventions, are delivered.<sup>15 19</sup> One area of research where fidelity of delivery has been assessed rigorously, using a reliable method of specifying intervention content in terms of behaviour change techniques (BCTs), is behavioural

support for smoking cessation<sup>14 15</sup> and an intervention to increase physical activity among sedentary adults.<sup>13</sup> These fidelity studies evaluated face-to-face or telephone-delivered behavioural support, comparing actual BCT delivery to that specified in treatment manuals. These studies demonstrated that fidelity can be reliably assessed using transcripts of audiotaped sessions and underline the need to establish routine procedures for monitoring the fidelity with which behaviour change interventions are delivered in clinical practice.

This study assessed the fidelity of an implementation intervention that was delivered as part of the IMPLEMENT cluster randomised trial, and addressed the Design and Delivery aspects of the Bellg *et al*<sup>7</sup> intervention fidelity framework. The IMPLEMENT trial evaluated an implementation intervention designed to improve the uptake of evidence-based recommendations for the management of acute low back pain in a general medical practice setting.<sup>20 21</sup> The IMPLEMENT intervention was designed to improve the uptake of two evidence-based general practitioner (GP) target clinical behaviours—avoiding X-ray ordering and giving advice to stay active. The intervention was designed to address the barriers to, and enablers of, the uptake of the target behaviours in clinical practice. Details of the development process of the IMPLEMENT intervention are published elsewhere,<sup>22</sup> as is the trial protocol and the trial results.<sup>20 21 23</sup>

In the IMPLEMENT trial, 47 practices (53 GPs) were randomised to the control and 45 practices (59 GPs) to the intervention. For one of the main trial outcomes, simulation of clinical behaviour of X-ray referral, the IMPLEMENT intervention group GPs were more likely to adhere to guideline recommendations about X-ray (OR 1.76, 95% CI 1.01 to 3.05) and were more likely to give advice to stay active (OR 4.49, 95% CI 1.90 to 10.60). However, actual imaging referral was not statistically significantly different between groups, with rate ratio 0.87 (95% CI 0.68 to 1.10) for X-ray or CT scan. Overall, the trial findings demonstrated that the IMPLEMENT intervention led to small changes in GP intention towards being more evidence based, but the intervention did not result in statistically significant changes in actual behaviour.

The principle aims of this study were to investigate the fidelity of this implementation intervention and to document an approach for assessing the fidelity of implementation interventions. The specific objectives were to assess: (1) observed facilitator adherence to planned BCTs; (2) variation in BCT adherence across different facilitators, sessions and BCTs and (3) agreement between observed and self-reported BCT adherence to sections of the workshop.

## METHOD

### Description of the IMPLEMENT intervention

The IMPLEMENT intervention was delivered by trained facilitators in a two-session workshop, with each session

lasting 3 hours. Thirty-six GPs (75% of GPs assigned to the intervention group) attended both sessions of the IMPLEMENT intervention workshop. The first session focused on the clinical behaviour of not referring a patient with acute low back pain for an X-ray (table 1), and the second session on the clinical behaviour of giving advice to stay active (table 2). Each session was run in different locations to give participants maximum opportunity to attend; the first was held on five occasions and the second, on six occasions. Each session had between two or three facilitators present, from a pool of six. Four of the facilitators were GPs with education and facilitator experience. Two of the facilitators were investigators of the study (SDF and SEG), with both having allied health clinical backgrounds and teaching experience.

Each of the two workshop sessions was made up of a number of sections, with each section consisting of the delivery of a number of BCTs. Fifteen distinct BCTs were planned for session I (table 1), and 19 BCTs were planned for session II (table 2). Across all workshop sessions, there were 189 instances when BCTs were planned to be delivered, but one of these sessions was not captured due to audiorecorder failure.

### Design and participants

The study was a comparison between planned and actual, and observed versus self-assessed delivery of BCTs during the IMPLEMENT workshops. The participants were the facilitators who delivered the workshops.

### Procedure

Each workshop session was audiorecorded and transcribed verbatim. A researcher not involved in any other components of the IMPLEMENT study checked each of the transcripts against the original audio files for accuracy. A coding frame and guidelines were developed (available as an online supplementary file), describing each of the planned BCTs and the criteria required for a BCT to be coded as delivered.

### Measures

Observed adherence of facilitators to the planned BCTs was assessed by coding the workshop transcripts for the presence of BCTs. We assessed adherence to planned BCT delivery and variation of delivery of the BCTs across different facilitators and sessions. Facilitator self-reported adherence to delivery of the IMPLEMENT intervention was measured using a checklist completed at the end of each workshop session. The facilitator indicated whether or not each section of a workshop was delivered, but did not record whether individual BCTs were delivered.

### Method of coding

To assess observed facilitator adherence, a coder recorded whether the facilitator delivered the specified BCTs as 0 (not applied) or 1 (applied). In some instances, some sections of the transcripts were double

coded because a facilitator could potentially deliver two BCTs at the same time. For example, a facilitator could discuss the identification of barriers to performing one of the target behaviours and planning ways of overcoming these, and at the same time use persuasive communication techniques. In this instance, the transcript lines were coded as both *barrier identification* and *persuasive communication*.

### Reliability of coding

To establish reliability of transcript coding, one transcript of a complete workshop was coded by two researchers independently. One of these researchers (SDF) was involved in the development and delivery of the intervention, and one was an independent researcher who had not been involved in these activities. Coding results were then discussed, and the coding frame and guidelines were modified until agreement of at least 80% was established on the occurrence of BCTs and the relevant text for each, as recommended by Lombard *et al.*<sup>24</sup> After reliable coding was established, SDF coded the remaining transcripts. A random check of 10% of the remaining coding was undertaken by the independent researcher and at least 80% agreement on the occurrence of BCTs was confirmed.

### Workshop content not coded

No coding was undertaken for the planned BCTs *prompt practice (rehearsal)* and *role play* because these BCTs were only delivered in small group discussions that were not able to be captured by the audiorecording device.

### Analysis

All data from the workshop transcripts were recorded in the coding frame and then entered into Microsoft Office Excel (V.2003) spreadsheets. Analyses were undertaken in Excel and Stata (V.9.0).<sup>25</sup> Simple summary statistics were used to assess observed adherence to planned BCTs and also to compare self-reported adherence to observed adherence. Observed facilitator adherence to BCTs was expressed as the number of BCTs recorded as delivered divided by the number of BCTs that were planned. The difference in adherence to planned BCT delivery between facilitators was assessed by the Pearson  $\chi^2$  test. Using observed adherence as the 'gold standard', sensitivity, specificity and predictive values of self-reported adherence and their respective 95% CIs were reported.

All participants gave their informed consent.

## RESULTS

### Observed adherence to protocol specified BCTs

The fidelity of each BCT type across all workshops is shown in table 3. The observed adherence to planned BCTs across all workshops was 79% overall, ranging from 33% to 100% per session. The BCT *provide information on*

**Table 1** Detailed planned content for session I of the IMPLEMENT intervention

<b>Session I: Confidence in diagnosis</b>			
<b>Section</b>	<b>Section title</b>	<b>Content</b>	<b>Planned behaviour change techniques</b>
I-1	Welcome and Introductions	Group introductions; Set boundaries; Agenda and content for session Viewing of Victorian WorkCover Authority (VWA) videos*	None Information provision
I-2	Small group work No.1: Discussion of pre-session activity about X-ray	Discussion in small groups (3–4) of pre-session activity  Feedback small group discussion to larger group. Facilitator to note-take barriers and enablers on whiteboard and revisit throughout session	► Barrier identification ► Provide opportunities for social comparison ► Barrier identification ► Provide information on consequences ► Persuasive communication ► Provide opportunities for social comparison
I-3	Guideline recommendations	Didactic presentation (facilitator led with group discussion): introduction to acute non-specific low back pain; guideline development and stakeholders; guideline key messages and evidence underpinning them	► Information provision ► Persuasive communication
I-4	Small group No.2: Making recommendations specific	In small groups re-write original guideline key messages: by who, applying to who, what, where, when	Barrier identification
I-5	Revisit small group discussions No.1 and No.2	Group discussion: challenge negative beliefs using persuasive communication and reinforce relevance of key message to GPs' role	Persuasive communication
I-6	Plain film X-ray for acute low back pain	Didactic presentation from radiologist Outlining harms and utility of X-ray	Information provision ► Provide information on consequences ► Persuasive communication
I-7	Red flag screening	Peer expert: presentation and demonstration (with simulated patient) of identifying red flags	Model/demonstrate the behaviour
I-8	Small group No.3: Red flag screening practical	GPs take history of trained simulated patients who are demanding an X-ray  Group discussion including feedback from simulated patients to the GPs about their consultations	► Prompt practice (rehearsal) ► Role play Provide information on consequences
I-9	Summary	Group discussion: reflect on barriers and enablers on whiteboard; questions; outstanding issues	► Barrier identification ► Persuasive communication ► Provide opportunities for social comparison

\*VWA videos were produced as part of a public health media campaign in the late 1990's. Details of the campaign are provided in Buchbinder *et al.*<sup>29</sup>  
GP, general practitioner.

**Table 2** Detailed planned content for session II of the IMPLEMENT intervention**Session II: Move it or lose it**

Section	Title	Content	Planned behaviour change techniques
II-1	Welcome and Introductions	Agenda and content for session Viewing of Victorian WorkCover Authority (VWA) videos* Recap on last session—opportunity to discuss any change in behaviour since last session (if appropriate, ie, session II not immediately following session I)	None Information provision  ▶ Barrier identification ▶ Persuasive communication ▶ Provide opportunities for social comparison
II-2	Small group work No.4: Discussion of pre-session activity about advice to stay active	Discussion in small groups (3–4) and feedback to larger group  Feedback small group discussion to larger group. Facilitator to note-take barriers and enablers on whiteboard and revisit throughout session	▶ Barrier identification ▶ Provide opportunities for social comparison ▶ Persuasive communication ▶ Provide information on consequences ▶ Provide opportunities for social comparison
II-3	Guideline recommendations about interventions	Didactic presentation about guideline recommendations for treatment and evidence supporting them	▶ Information provision ▶ Provide instruction ▶ Persuasive communication
II-4	Making recommendations behaviourally specific	In whole group, re-write original guideline key messages: by who, applying to who, what, where, when	None
II-5	10 steps of clinical management as a framework for low back pain management	Didactic presentation with group discussion including use of: activity log as a replacement for X-ray referral slip; patient education handout	▶ Information provision ▶ Provide instruction ▶ Persuasive communication ▶ Time management
II-6	Small group No.5: talking with patients: Putting recommendations into practice	Using pre-prepared outlines of patient scenarios, participants practice with a partner and create scripts for themselves about the 2 key messages that are workable, time efficient and reinforces patient education	▶ Prompt practice ▶ Role play ▶ Time management
II-7	Summary	Feedback small group discussion to larger group. Reflect on barriers on whiteboard; questions; outstanding issues	▶ Barrier identification ▶ Provide information on consequences ▶ Provide opportunities for social comparison ▶ Persuasive communication
II-8	Action planning	Participants to choose relevant issues they wish to change in their practice, using 'if-then' planning for goal achievement	▶ Provide instruction ▶ Prompt specific goal setting (action planning)

\*VWA videos were produced as part of a public health media campaign in the late 1990's. Details of the campaign are provided in Buchbinder *et al.*<sup>29</sup>

*consequences* had the lowest fidelity (70%), and *information provision* had the highest (97%).

### Variation in intervention fidelity across different facilitators, sessions and BCTs

The percentage of BCTs delivered across all sessions, for each session, and for each facilitator is shown in [figure 1](#). For session I, at least 80% of the planned BCTs were

delivered, and in three of the session I workshops, 100% of the planned BCTs were delivered. For session II, on no occasion were all of the planned BCTs delivered, with the percentage of planned BCTs delivered in session II ranging from 53% to 84%. There were significantly less BCTs delivered as planned in session II compared with session I ( $\chi^2$  test (df=1)=16.6,  $p<0.001$ ). Facilitator adherence to the planned BCT delivery varied from 86% to



**Table 3** Fidelity of each behavioural change technique (BCT) type across all workshops from content analysis of transcripts

BCT	Adherence n/N (%) <sup>*</sup>		
	Session I	Session II	Both sessions
Persuasive communication	21/25 (84)	18/30 (60)	39/55 (71)
Information provision†	14/15 (93)	18/18 (100)	32/33 (97)
Provide information on consequences	15/15 (100)	4/12 (33)	19/27 (70)
Provide opportunities for social comparison	10/10 (100)	11/18 (61)	21/28 (75)
Barrier identification	10/10 (100)	10/12 (83)	20/22 (91)
Provide instruction	0 (-)	14/18 (78)	14/18 (78)
Time management	0 (-)	5/6 (83)	5/6 (83)
Total	70/75 (93)	80/114 (70)	150/189 (79)

<sup>\*</sup>Instances of BCTs delivered versus instances planned.

†Missing data for one instance due to audiorecording device failure.

100% for session 1 and 53% to 79% for session II, but any difference between facilitators was not statistically significant ( $\chi^2$  test (df=5)=9.7, p=0.08).

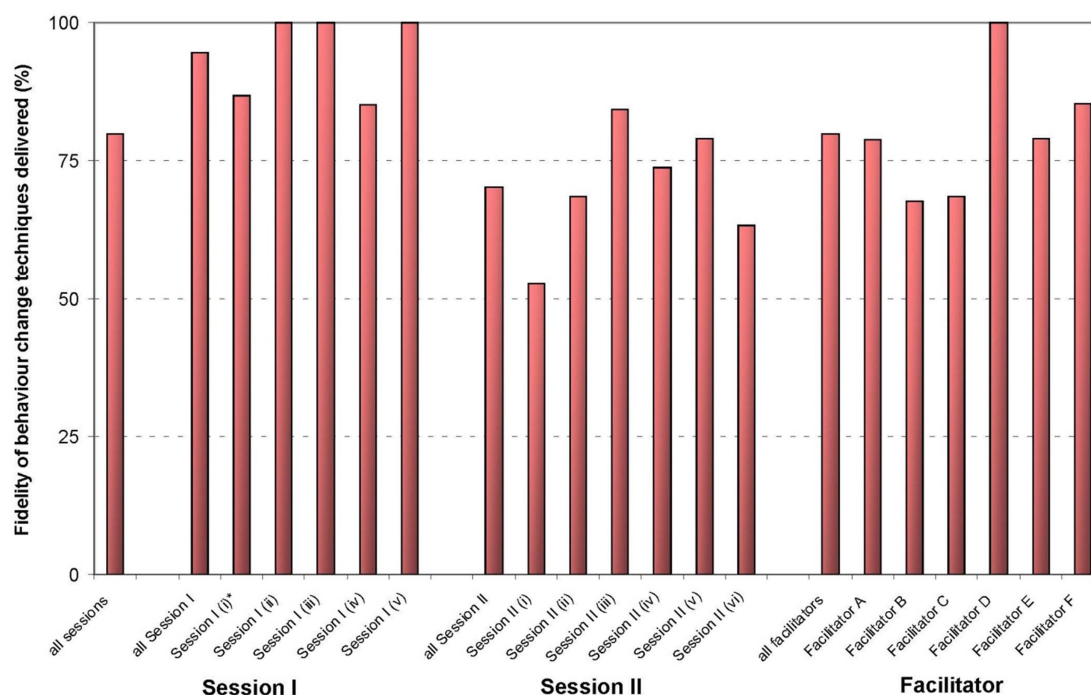
### Observed versus self-reported adherence

The sensitivity of self-reported adherence, that is, facilitators correctly identifying when a section of a workshop did occur according to the 'gold standard' of observed adherence, was 95% (95% CI 88 to 98). The specificity, that is, the ability of the facilitators to correctly identify when a section of a workshop did not occur according to the observed adherence, was 30% (95% CI 11 to 60).

The positive predictive value of self-reported adherence was 92% (95% CI 84 to 96), and the negative predictive value was 43% (95% CI 16 to 75). These results indicate that the facilitators could accurately determine if a section of the workshop was delivered, but were less able to accurately report when a section was not delivered.

### DISCUSSION

The overall observed adherence to BCTs in the IMPLEMENT workshops was 79%. It has been argued that 80–100% adherence to an intervention protocol



\* Roman numerals indicate each individual session that was delivered, e.g. Session I (ii) represents the second time we ran Session I, Session II (iv) the fourth time we ran Session II, etc

**Figure 1** Observed fidelity of behaviour change techniques delivered across all sessions, for each individual session and for each facilitator.

represents 'high' fidelity of delivery, whereas 50% adherence or less represents 'low fidelity'.<sup>8</sup> We use this convention to describe the IMPLEMENT intervention as being delivered with high levels of adherence to the planned intervention protocol. However, we recognise that this is a convention and further research needs to be conducted to ascertain the implications of different levels of adherence to different kinds of interventions in different contexts.

Although there was some variation in facilitator adherence to planned delivery of BCTs, there was no significant difference between the different facilitators who were involved in the workshops. This could be the result of the training programme we employed for facilitators, including a facilitator manual and telephone discussion of the details of the delivery of the intervention and because two of the study investigators were present at all workshops and provided input to all discussions. Our study design and analysis was unable to take into account the interactions between different facilitators and investigators.

Facilitators accurately reported the delivery of a section of a workshop, but did not accurately report its non-delivery. In this study, facilitators were only required to self-report delivery at the workshop section level, rather than the delivery of individual BCTs. If the facilitators were to self-report adherence in more detail to the level of specific BCTs rather than just record whether certain sections were delivered or not, it is likely that self-reported adherence would be less accurate. The findings suggest that observed adherence should be measured in fidelity studies, rather than by relying on self-reported adherence of those who deliver the intervention; this is consistent with findings in other fidelity studies of BCTs.<sup>13–15</sup>

Overall fidelity was lower in the second workshop session (70%) than the first workshop session (93%). The reason for this difference is unclear because our study was designed only to evaluate fidelity, not to determine why fidelity was low or high. Multiple reasons may explain this difference, including order effects, the nature of the target behaviours at the different sessions, or the different content and the nature of the BCTs themselves. The first session had a clinical focus about X-ray ordering, and the second was focused on giving advice to stay active. The trial results demonstrated large variation across general practices in the management of patients for giving advice to stay active, indicating that this behaviour was a more challenging behaviour for the GPs in the IMPLEMENT trial.<sup>21</sup> Further, the first session focused on stopping a behaviour (ordering of X-rays), but the second session focused on increasing a behaviour (giving advice to stay active), and this may account for the difference in fidelity between the sessions. Finally, the difference may have been due to differences between the actual BCTs that were planned to be delivered in the two sessions, with some more difficult to deliver than others. Our study was not designed to

determine why fidelity might vary across sessions, BCTs and facilitators; future fidelity evaluations should assess not only fidelity itself, but reasons for why fidelity may vary.

The main rater of observed adherence (SDF) was not an independent assessor, and was involved in the development and the delivery of the intervention, which is a study limitation. Also, we did not undertake double coding of all transcripts. We, therefore, cannot be certain whether coding reflects actual performance rather than unknown factors related to knowledge of the facilitators by the assessor. However, we did establish reliability of the coding frame by comparing the results of coding of the main rater to an independent assessor for one full workshop (two sessions), and a random check of 10% of the remaining coding by the independent researcher confirmed agreement. Coding was a time-intensive procedure and double coding of all transcripts was not feasible with our available resources. If feasible, future work in this area should employ an independent rater to undertake all coding.

Our scoring criteria for evaluating the delivery of individual BCTs of 'applied' and 'not applied' did not necessarily detect instances where facilitators potentially delivered a BCT, but did not provide correct information according to the guideline recommendations. For example, for the red flag screening demonstration (table 1), the planned BCT was *model/demonstrate the behaviour*. Facilitators could have delivered this BCT correctly, could have delivered the BCT but given incorrect information about red flags, or may not have delivered this at all. Our coding system was based on the assumption that if the BCTs were delivered, facilitators provided correct information. However, future fidelity studies should consider using a more sophisticated coding scheme to ensure that correct information is accurately coded. For instance, the coding instructions could be 'applied' and 'not applied or incorrectly applied'.

In this study, we addressed the Design, Training and Delivery aspects of the Bellg *et al*<sup>7</sup> intervention fidelity framework,<sup>8</sup> but did not address the receipt and enactment aspects. The receipt and enactment aspect of the Bellg *et al* framework as applied to the IMPLEMENT intervention are addressed in another paper reporting the results of the IMPLEMENT trial.<sup>21</sup> We were unable to analyse the responses of GP participants to the facilitator-led discussions in the workshops because the audiorecording device primarily captured the voice of the facilitator, and was not able to clearly capture the voices of the workshop participants. Nor could sound recordings capture the subtleties of facilitator delivery involving non-verbal behaviours, as well as individual and group interactions. We also did not measure actual GP clinical behaviour in their practices with patients with low back pain. Therefore, we have no data on how the GPs responded at the time the facilitators delivered the BCTs, and can draw no conclusions about how they received the BCTs at the time of delivery nor how they enacted the

learned behaviours with patients when they returned to practice. Also, some sections of the workshop were small group activities and were not captured by the recording device. Therefore, we have no data on the coverage of planned BCTs during the small group discussions, including the techniques of *prompt practice (rehearsal)* and *role play*. It is possible that some of the influences on GP behaviour change were when GPs had the opportunity to discuss their own behaviour with their peers and also role play with simulated patients. Future studies should attempt to capture all aspects of intervention delivery, receipt and enactment, if possible.

The assessment of the extent to which an intervention has adequately adhered to the planned intervention protocol is complex. The literature describes two distinct views about the fidelity of interventions in effectiveness research. One view is that fidelity should be strictly maintained with adherence to all aspects of the intervention protocol, whereas another view is that interventions should be flexible to change when implemented in different settings, but not so flexible that validity is compromised.<sup>26</sup> With the latter view, it may be recognised that essential components of the intervention should be strictly adhered to and an intervention may be considered successfully delivered if the 'essential' components of the intervention are delivered as planned.<sup>27</sup> However, standardisation of an intervention across settings does not necessarily ensure an effective intervention on the overall.<sup>28</sup> Adequate adherence to an intervention protocol may not require every single component of an intervention to be delivered, especially in the case of a complex intervention with multiple components.<sup>18 28</sup> However, for many interventions it may be difficult to identify what is essential and what is optional for a particular intervention in a particular context.<sup>26</sup> Owing to limitations in our fidelity evaluation, particularly that the small group activities were not coded, we are unable to confidently determine which components of the IMPLEMENT intervention are primarily responsible for its effects. Investigators of future fidelity evaluations should consider whether it is possible to a priori determine which components of the intervention are essential and should be given priority in fidelity evaluation.

The IMPLEMENT trial results demonstrated improvement in outcomes for both targeted behaviours (X-ray referral and giving advice to stay active). This was despite there being lower fidelity of the delivery of the intervention for one of the target behaviours, with the workshop on advice to stay active having fidelity of delivery of BCTs of 70%, whereas the X-ray workshop session had fidelity of delivery of BCTs of 93%. Considering we did not detect an association between fidelity of delivery and intervention effectiveness, we recommend that future fidelity studies explicitly investigate the link between fidelity and effectiveness to further examine the causal model of change.

The resources required to undertake this intervention fidelity evaluation were modest. Resources included

preparation of facilitator training materials, facilitator training, investigator time to plan the fidelity study, and time to undertake transcript coding and analysis. Owing to the relatively small number of workshop sessions, we were able to code all of the intervention delivered to all the trial participants. This is higher than has been recommended for treatment fidelity studies: code a minimum sample of 20–40% of sessions, balanced across the period of intervention delivery and conditions.<sup>27</sup> It may be that we overcoded our intervention and expended unnecessary resources. When the trial was originally planned and funding was sought, we did not propose to evaluate intervention fidelity and no resources were allocated to cover the costs associated with this. Future fidelity work should be incorporated into proposed study budgets at the funding application stage and a full plan for fidelity should be costed. This will require that funding bodies recognise the importance of this step in the evaluation of implementation interventions.

## CONCLUSIONS

The overall fidelity of delivery of the IMPLEMENT intervention was 79%. We are confident that the IMPLEMENT intervention was delivered with high levels of adherence to the planned intervention protocol and therefore, we anticipate that the trial outcome data were not the result of poor intervention delivery. Intervention fidelity is important in implementation research in order to understand the effects of an intervention. This study has demonstrated that it is feasible to undertake a rigorous fidelity assessment of a complex intervention by coding it according to specific BCTs.

### Author affiliations

<sup>1</sup>School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

<sup>2</sup>School of Rehabilitation Therapy, Queen's University, Kingston, Ontario, Canada

<sup>3</sup>Division of Health Services Research & Management, School of Health Sciences, City University London, London, UK

<sup>4</sup>Monash Department of Clinical Epidemiology, Cabrini Hospital, Melbourne, Victoria, Australia

<sup>5</sup>Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada

<sup>6</sup>Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada

<sup>7</sup>Department of Clinical, Educational and Health Psychology, Centre for Behaviour Change, University College London, London, UK

**Acknowledgements** The authors thank Dr Tari Turner for double coding of the transcripts.

**Contributors** SDF completed this work as part of his PhD project; SEG and RB were his supervisors. SDF, SEG, JJF, RB, DAO, JMG and SM conceived and designed the study. SDF and SEG were involved in the delivery of the intervention. SDF was involved in the data acquisition. SDF was involved in the data analysis, with input from SM, SEG and RB. SDF wrote the first draft of the manuscript. SM contributed to detailed revisions of the manuscript. All the authors contributed to the revisions of the manuscript and approved the final version.

**Funding** The IMPLEMENT trial was funded by the Australian National Health and Medical Research Council (NHMRC) by way of a Primary Health Care Project Grant (334060). SDF was partly supported by a NHMRC Primary



Health Care Fellowship (567071). RB is partially supported by an NHMRC Practitioner Fellowship. DAO is supported by a NHMRC Public Health Fellowship (606726). JMG holds a Canada Research Chair in Health Knowledge Transfer and Uptake.

**Competing interests** None declared.

**Ethics approval** Monash University Standing Committee on Ethics in Research involving Humans (2006/047).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

- Grol R, Berwick DM, Wensing M. On the trail of quality and safety in health care. *BMJ* 2008;336:74–6.
- Runciman WB, Hunt TD, Hannaford NA, *et al.* CareTrack: assessing the appropriateness of health care delivery in Australia. *Med J Aust* 2012;197:100–5.
- McGlynn EA, Asch SM, Adams J, *et al.* The quality of health care delivered to adults in the United States. *N Engl J Med* 2003;348:2635–45.
- Grimshaw JM, Thomas RE, MacLennan G, *et al.* Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;8:iii–iv, 1–72.
- Michie S, Johnston M, Abraham C, *et al.* Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care* 2005;14:26–33.
- Borrelli B, Sepinwall D, Ernst D, *et al.* A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *J Consult Clin Psychol* 2005;73:852–60.
- Bellg AJ, Borrelli B, Resnick B, *et al.* Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychol* 2004;23:443–51.
- Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *J Public Health Dent* 2011;71(Suppl 1):S52–63.
- Boutron I, Moher D, Altman DG, *et al.* Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148:295–309.
- Dusenbury L, Brannigan R, Falco M, *et al.* A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Educ Res* 2003;18:237–56.
- Forgatch MS, Patterson GR, Degarmo DS. Evaluating fidelity: predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behav Ther* 2005;36:3–13.
- Resnick B, Bellg AJ, Borrelli B, *et al.* Examples of implementation and evaluation of treatment fidelity in the BCC studies: where we are and where we need to go. *Ann Behav Med* 2005;29 (Suppl):46–54.
- Hardeman W, Michie S, Fanshawe T, *et al.* Fidelity of delivery of a physical activity intervention: predictors and consequences. *Psychol Health* 2008;23:11–24.
- Lorenzatto F, West R, Bruguera C, *et al.* A method for assessing fidelity of delivery of telephone behavioral support for smoking cessation. *J Consult Clin Psychol* 2014;82:482–91.
- Lorenzatto F, West R, Christopherson C, *et al.* Assessing fidelity of delivery of smoking cessation behavioural support in practice. *Implement Sci* 2013;8:40.
- McAteer J, Stone S, Fuller C, *et al.* Using psychological theory to understand the challenges facing staff delivering a ward-led intervention to increase hand hygiene behavior: a qualitative study. *Am J Infect Control* 2014;42:495–9.
- Michie S, Fixsen D, Grimshaw JM, *et al.* Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implement Sci* 2009;4:40.
- Carroll C, Patterson M, Wood S, *et al.* A conceptual framework for implementation fidelity. *Implement Sci* 2007;2:40.
- Mars T, Ellard D, Carnes D, *et al.* Fidelity in complex behaviour change interventions: a standardised approach to evaluate intervention integrity. *BMJ Open* 2013;3:e003555.
- McKenzie JE, French SD, O'Connor DA, *et al.* IMPLementing a clinical practice guideline for acute low back pain evidence-based management in general practice (IMPLEMENT): cluster randomised controlled trial study protocol. *Implement Sci* 2008;3:11.
- French SD, McKenzie JE, O'Connor DA, *et al.* Evaluation of a theory-informed implementation intervention for the management of acute low back pain in general medical practice: the IMPLEMENT cluster randomised trial. *PLoS ONE* 2013;8:e65471.
- French SD, Green SE, O'Connor DA, *et al.* Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. *Implement Sci* 2012;7:38.
- Mortimer D, French SD, McKenzie JE, *et al.* Economic evaluation of active implementation versus guideline dissemination for evidence-based care of acute low-back pain in a general practice setting. *PLoS ONE* 2013;8:e75647.
- Lombard M, Snyder-Duch J, Bracken CC. Content analysis in mass communication: assessment and reporting of intercoder reliability. *Hum Commun Res* 2002;28:587–604.
- Stata Statistical Software: Release 12 [program]. College Station, TX: StataCorp LP, 2011.
- Cohen DJ, Crabtree BF, Etz RS, *et al.* Fidelity versus flexibility: translating evidence-based research into practice. *Am J Prev Med* 2008;35(5 Suppl):S381–9.
- Schlosser R. On the importance of being earnest about treatment integrity. *Augment Altern Commun* 2002;18:36–44.
- Hawe P, Shiell A, Riley T. Complex interventions: how “out of control” can a randomised controlled trial be? *BMJ* 2004;328:1561–3.
- Buchbinder R, Jolley D, Wyatt M. Effects of a media campaign on back pain beliefs and its potential influence on management of low back pain in general practice. *Spine* 2001;26:2535–42.