ELSEVIER

Social Science & Medicine 60 (2005) 2135–2147

SOCIAL SCIENCE & MEDICINE

www.elsevier.com/locate/socscimed

Guiding the design and selection of interventions to influence the implementation of evidence-based practice: an experimental simulation of a complex intervention trial

Debbie Bonetti[a,*], Martin Eccles[b], Marie Johnston[c], Nick Steen[b], Jeremy Grimshaw[d], Rachel Baker[b], Anne Walker[e], Nigel Pitts[a]

[a]*Dental Health Services Research Unit, University of Dundee, UK*
[b]*School of Population and Health Sciences, Centre for Health Services Research, University of Newcastle upon Tyne, UK*
[c]*School of Psychology, University of Aberdeen, UK*
[d]*Clinical Epidemiology Programme, Ottawa Health Research Institute, Canada*
[e]*Health Services Research Unit, University of Aberdeen, UK*

Available online 8 December 2004

Abstract

A consistent finding in health services research is the report of uneven uptake of research findings. Implementation trials have a variable record of success in effectively influencing clinicians' behaviour. A more systematic approach may be to conduct Intervention Modelling Experiments before service-level trials, examining intervention effects on 'interim endpoints' representing clinical behaviour, derived from empirically supported psychological theories. The objectives were to: (1) Design Intervention Modelling Experiments by backward engineering a 'real-world' randomised controlled trial (NEXUS); (2) examine the applicability of psychological theories to clinical decision-making; (3) explore whether psychological theories can illuminate how interventions achieve their effects.

A 2×2 factorial randomised controlled trial was designed with pre- and post-intervention data collection by postal questionnaire surveys. The first survey was used to generate feedback data and the interventions were delivered in the second survey. General medical practitioners (GPs) in England and Scotland participated. First survey respondents were randomised twice to receive or not audit and feedback and educational reminder messages.

The main outcome measures included behavioural intention (general plan to refer for lumbar X-rays) and simulated behaviour (specific, scenario-based, decisions to refer for lumbar X-ray). Predictors were attitude, subjective norm, perceived behavioural control (theory of planned behaviour), self-efficacy (social cognitive theory) and decision difficulty.

Both interventions significantly influenced simulated behaviour, but neither influenced behavioural intention. There were no interaction effects. All theoretically derived cognitions significantly predicted simulated behaviour. Only subjective norm was not predictive of behavioural intention. The effect of audit and feedback on simulated behaviour was mediated through perceived behavioural control.

The results of this study suggest that Intervention Modelling Experiments, using psychological models to help isolate mediators of clinical decision-making, may be a means of developing more potent interventions, and selecting

*Corresponding author. DHSRU, The Mackenzie Building, Kirsty Semple Way, Dundee, Scotland, DD2 4BF. Tel.: +44 1382 420154; fax: +44 1382 420051.

E-mail address: d.bonetti@chs.dundee.ac.uk (D. Bonetti).

0277-9536/$ - see front matter
doi:10.1016/j.socscimed.2004.08.072

implementation interventions with a greater likelihood of success in a service-level randomised controlled trial.

Keywords: Implementation modelling experiments; Randomised controlled trials; The theory of planned behaviour; Social cognitive theory; Evidence-based practice

Introduction

Clinical and health services research is continually producing new findings that may contribute to effective and efficient patient care. However, despite the considerable resources devoted to this area, a consistent finding is that the transfer of research findings into practice is unpredictable and can be a slow and haphazard process. Studies in the UK, USA and the Netherlands suggest that about 30–40% of the patients do not receive care according to current scientific evidence and about 20–25% of care provided is not needed or potentially harmful (e.g. Ketley & Woods, 1993; Grol, 2001; Schuster, McGlynn, & Brook, 1998).

Extensive resources have been devoted to reducing this variation, and hence inappropriate care, in trials of interventions to encourage health professionals to implement evidence-based practice. Over the last decade a large number of implementation intervention studies have been published and reviewed (Bero et al., 1998; Oxman, Thomson, Davis, & Haynes, 1995; Grimshaw et al., 2001). Overall, there has been variable success in effectively influencing health care professionals' behaviour, and little progress has been made in understanding why any particular implementation intervention succeeds or fails. The substantial heterogeneity within interventions, targeted behaviours, and study settings, makes generalising results of implementation trials problematic; a position aggravated by not drawing on theoretical models specifically directed at behaviour change. While the imperative of reducing variation in clinical practice by encouraging evidence-based practice remains, research conducted to date provides little guidance on the design or selection of implementation interventions for further evaluation in trials.

A possible solution to this problem is to have more systematic methods of developing implementation interventions, optimising service level trials and enhancing the generalisability of research findings. This approach accords with the framework for evaluating complex interventions suggested by the Medical Research Council (MRC, 2000). The MRC framework includes 5 ordered stages: Theory, Modelling, Exploratory Trial, Definitive randomised controlled trial, and Long-term Implementation. To date, implementation research has focused on the definitive randomised controlled trial stage. By drawing on the 3 prior phases, it should be possible to enhance an understanding of how the interventions operate and to concentrate resources on interventions with greater likelihood of success of definitive randomised controlled trials than is presently the case.

One approach to conducting the Exploratory Trial stages of the MRC framework is to conduct Intervention Modelling Experiments. In such experiments, key elements of an intervention are manipulated in a manner that simulates the 'real-world' as much as possible, but the measured outcome is an interim endpoint—a proxy for the clinical behaviour that will be the definitive randomised controlled trial outcome. Using interim endpoints, which are easier and cheaper to measure than actual behaviour, means that an Intervention Modelling Experiment functions as a resource-efficient tool to enable intervention design modification, while providing a scientific rationale for intervention selection. An intervention should successfully change the experiment's proxy outcome before being considered eligible for testing on actual behaviour in a definitive randomised controlled trial. Of course an intervention that successfully influences a proxy outcome in a modelling experiment is not guaranteed to successfully influence behaviour—the experiment may not manage to replicate the exact ecological conditions of real life or may investigate only short-term effects. Nevertheless, an intervention that influences a modelling experiment's proxy outcome is *more* likely to influence the represented behaviour in a service-level trial than an intervention that does not.

Possible interim endpoints can be identified from the large body of research utilising psychological theories to predict and change health behaviour and long-term health outcomes. For example, behaviour change is unlikely to occur if motivation is lacking (Fishbein et al., 2001). Using empirically supported psychological theories to identify interim endpoints should also further an understanding of the interventions themselves. If the interventions achieve their effects by influencing predictive cognitions derived from these models, then this information can guide the future design and development of implementation interventions, thereby increasing the chance of successful definitive randomised controlled trials.

However, the validity of using psychological models in relation to clinician behaviour has yet to be established. It is possible that the uptake of evidence on effective clinical practice may differ from the health-related behaviours and outcomes generally the subject of studies using psychological models. If so, then cognitions drawn from these models could not be used as modelling experiment outcomes. It also may not be possible to capture 'real-world' effects of interventions on clinical decision-making at an experimental level. Although modelling experiments in other fields have demonstrated that short-term laboratory effects have been replicated in 'real-world' studies of long-term health outcomes (Paul, 1966; Bandura, 1969; Mathews, Gelder, & Johnston, 1981), the validity of using modelling experiments in relation to clinician's implementation behaviour has yet to be established.

Whilst the intended use of a modelling experiment is to optimise the subsequent design of an intervention, we chose, at this exploratory stage, to examine the feasibility of the method by replicating a recently completed trial, the North–East X-ray Utilisation Study (NEXUS: Eccles et al., 2001). The idea was to 'backward engineer' a definitive randomised controlled trial that tested two commonly used implementation interventions which have a variable record of success. The NEXUS trial evaluated the relative effectiveness of audit & feedback and educational reminder messages (reminders) in changing general practitioners' (GP) radiology ordering behaviour for lumber spine and knee X-rays within a 2×2 factorial cluster randomised controlled trial design. The educational reminder messages were based on the Royal College of Radiologists' guidelines and were provided on the report of every X-ray ordered over a 12-month period. The audit and feedback (which was comparative and of practice-level referrals over 6 months) was delivered to individual GPs at the start of the intervention period and again 6 months later. The study found that reminders led to a significant reduction of approximately 20% in X-ray requests, whilst audit/feedback led to a non-significant reduction of around 1% in X-ray requests (Fig. 1).

The first aim of this study was to examine whether it was possible to design a modelling experiment to reflect a definitive randomised controlled trial. It was acknowledged that reproducing the NEXUS results would not, in itself, test the validity of the methodology, given that the purpose of a modelling experiment is to determine an intervention's eligibility for taking to the definitive randomised controlled trial stage, not guaranteeing its success at that stage. Nevertheless, being able to

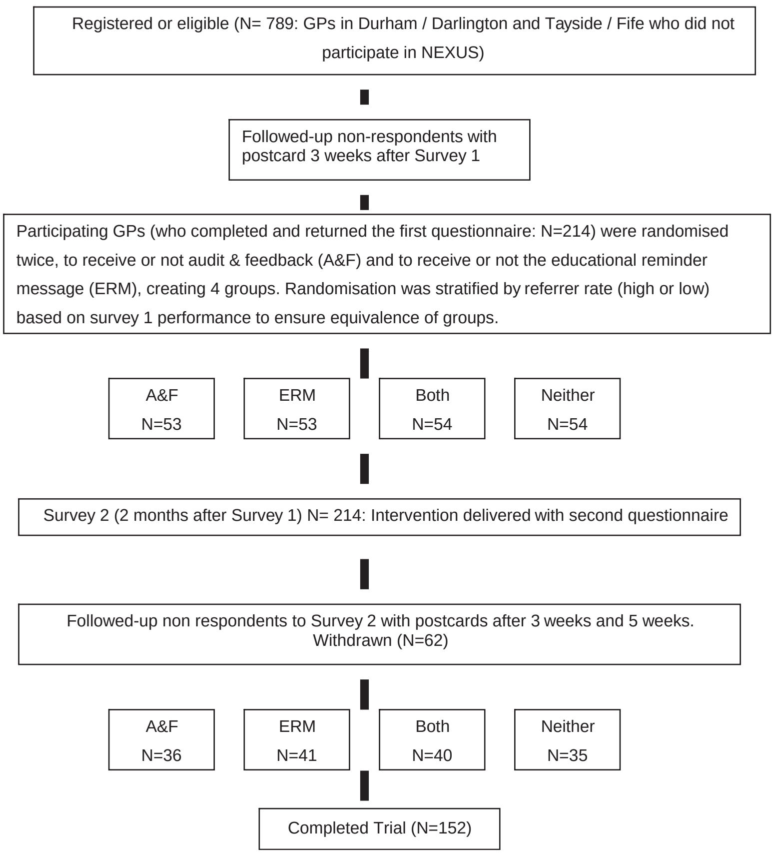

Fig. 1. NIME trial profile.

reproduce both types of interventions at this experimental level would be crucial for the validity of this methodology. Another critical issue for the validity of the modelling experiment method is being able to identify interim endpoints to proxy clinical behaviour. Therefore, the second aim of this study was to examine the applicability of psychological models to identify interim endpoints to clinical decision-making.

The theory of planned behaviour (Ajzen, 1991) and social cognitive theory (Bandura, 1997, 2000) are psychological models which have been successfully used to predict variation in many different behaviours in many different populations (e.g. Norman & Conner, 1993; Conner & Sparks 1996; Godin & Kok 1996; Cox et al., 1998; Albarracin, Johnson, Fishbein, & Muellerleile, 2001; Armitage & Conner, 2001; Hardeman et al., 2002).

The theory of planned behaviour proposes that the strength of an individual's intention to engage in behaviour, and the degree of control they feel that they have over the behaviour (perceived behavioural control) are the proximal determinants of engaging in it (see Fig. 2). Intention is defined in terms of a conscious plan to exert effort to perform the behaviour. Intention strength is posited as determined by attitudes towards the behaviour, subjective norms and perceived behavioural control, all of which, in turn, are based upon salient beliefs about the behaviour. Attitudes towards the behaviour are proposed to arise from a combination of beliefs about its consequences (behavioural beliefs) and evaluations of those consequences (outcome evaluations). Subjective norms are based on perceptions of the views about the behaviour of other individuals or groups (normative beliefs), and the strength of the individual's desire to gain approval of these groups (motivation to comply). Perceived behavioural control is a function of beliefs about factors likely to facilitate or inhibit the behaviour, including factors such as patient preferences or resource constraints on health professional practice.

In social cognitive theory, it is proposed that behaviour is determined by self-efficacy cognitions, outcome expectations, impediments and proximal goals (Fig. 2). It is proposed that behaviour change and maintenance are directly linked with the modification of self-efficacy cognitions, which is usually the only variable employed when utilising this theory. Self-efficacy cognitions are beliefs about capabilities to organise and execute courses of action required to produce given attainments and are a function of learning and experience. Self-efficacy is expected to vary depending on the task and context, affecting the acquisition, inhibition and disinhibition of behaviour by influencing affective states, choice of tasks and situations, persistence, and effort investment (Bandura, 1997).

The third aim of this study was to explore whether it was possible to use psychological models to further our

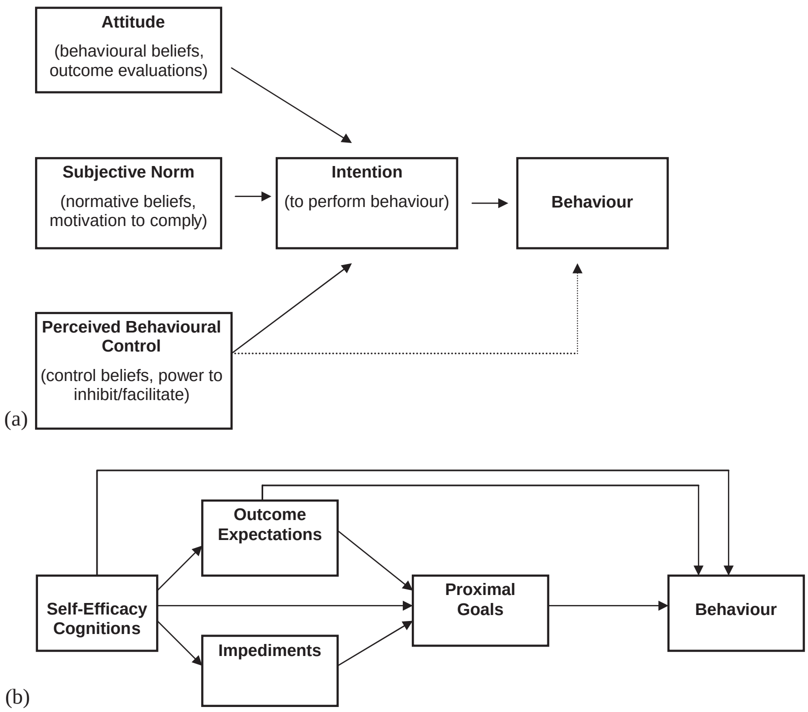

Fig. 2. Psychological models of behaviour. (a) The theory of planned behaviour (adapted from Ajzen, 1991). (b) Social cognitive theory (adapted from Bandura, 2000).

understanding of the interventions themselves—of how they achieve their effects. To explore the specific cognitions influenced by the interventions, measures of decision difficulty, and variables derived from the theory of planned behaviour and social cognitive theory were employed as predictors.

In view of the importance of the interim endpoint in the modelling experiment methodology, this study used two different measures to proxy GPs' referral behaviour. One was a theoretically derived measure, *behavioural intention*. This measure was chosen because there is considerable evidence supporting intention as the single best predictor of subsequent health-related behaviour (Ajzen, 1991; Randall & Wolff, 1994; Conner & Norman, 1996). The second outcome measure, *behavioural simulation*, was designed to simulate GPs' referral behaviour by asking GPs to make decisions about who they would refer from a set of scenarios depicting patients with back pain.

To achieve the objectives of this study, four questions were explored in the context of an intervention modelling experiment relating to the referral of patients for lumbar spine X-ray:

1. Do audit and feedback or educational reminder messages influence GPs' behavioural intention?
2. Do audit and feedback or educational reminder messages influence GPs' simulated behaviour?
3. Do audit and feedback or educational reminder messages influence GPs' cognitions (attitude, subjective norm, perceived behavioural control, self-efficacy and decision difficulty)?
4. Can cognitive variables derived from psychological models predict GPs' behavioural intention or simulated behaviour?

Method

Design

This study replicated the design of the NEXUS trial, and was a 2×2 factorial randomised control trial (Fig. 1). Data collection was by two postal questionnaire surveys, two months apart at baseline and post intervention. The baseline survey was used to recruit trial participants and to generate data for the interventions (see below).

Following the baseline survey, respondents were randomised twice—to receive or not to receive audit and feedback and reminders. This produced the four groups of audit/feedback alone, reminders alone, both and neither. Randomisation was stratified by 'referral rate', based on the median of the number of X-ray results requested in the first survey (see behavioural simulation, below). 'High referrers' were GPs who made more than three requests for X-rays and 'low referrers' were GPs who made three or less. The referrer groups underwent the randomisation process separately to ensure that each intervention group had a similar mix of high and low referrers—an element that might have masked intervention effects.

The second survey delivered the interventions and collected post-intervention data.

Participants

The study participants were 789 GPs randomly identified from practitioner lists (supplied by Primary Care Groups and Primary Care Trusts in England and Scotland) in the districts of Durham/Darlington and Tayside/Fife. None of these GPs participated in the NEXUS Trial.

Interventions

All participants received the X-ray reports they requested from the first survey scenarios. These reports had the same format as those used by radiology departments in their region.

To replicate the NEXUS audit and feedback intervention, GPs randomised to receive this intervention were sent a graph indicating their 'referral rate' (the number of decisions they made to refer from the first survey scenarios) on a histogram of the rates of all participants in the study.

In NEXUS, the educational reminder message intervention involved adding text on X-ray reports, which read: 'In either acute (less than 6 weeks) or chronic back pain, without adverse features, X-ray is not routinely indicated'. GPs randomised to receive this intervention had this message attached to each of the simulated X-ray reports they requested.

Outcome measures

These were measured at baseline and again following the intervention.

Behavioural intention

This measure comprised of 3 items: 'For every 10 patients, approximately how many would you refer for an X-ray: (a) presenting with acute back pain for the first time? (b) Presenting with acute back pain for the second time? (c) Presenting with acute back pain for the third time?' Scores from the 3 items were summed to create a total score (from 0 to 30) showing overall strength of intention to refer.

Simulated behaviour

There were 20 scenarios in total (10 per survey) describing patients presenting in primary care with back pain. Key clinical, GP, and patient elements, which might influence GPs' decisions to refer for a lumbar X-ray, were varied in each scenario. The elements were derived from the radiology guidelines (Royal College of Radiologists Working Party, 1995), the literature (e.g. Kerry et al., 2000; Kendrick et al., 2001), and a qualitative study on GPs' management of back pain (Bond, 1999). The clinical elements were: duration of episode, age of the patient, indicators of concern (e.g. steroid use, recent infection, pain worse with rest, previous carcinoma, arthritis (spondylitis), weight loss, neurological deficit, motor loss), general health of the patient, pre-existing/coexisting medical problems, time off work, previous episodes (number, duration), number of visits to the GP for this episode, previous X-ray, medication, change in the pattern of symptoms, and use of alternative services (physiotherapy; chiropractic; osteopathy; aromatherapy/massage; acupuncture). The GP element was knowledge of the patient (their patient or their partner's patient). The patient elements were: gender, general tendency to visit the GP, pain behaviour (wincing, slow movement, groaning), anxiety, distress, expectation that the GP should do something, desire for an X-ray, hostility toward GP (talk of neglect/problem not taken seriously), and occupation.

In NEXUS, the feedback in the audit and feedback intervention related to the number of referrals made by participating practices and not whether the referrals were appropriate according to the guidelines. Nevertheless, it was decided to control for 'appropriateness', since this factor was felt to be likely to influence referral decision-making. Age, duration of episode, and the presence of a red flag were elements cited in the guidelines as relevant to the recommendation for a lumbo-sacral spine X-ray. Since there was no information in the guidelines as to their relative importance for referral, nor to a required combination, expert opinion guided the criterion for an appropriate referral decision (the presence of at least 2 referral elements) as well as the proportion of 'appropriate' scenarios. Two out of ten scenarios in each questionnaire described patients who 'should' be referred for an X-ray. All scenarios included 5–8 elements, which allowed variation while keeping the content relatively consistent and clinically plausible. The scenarios were stratified by 'appropriateness' and patient's age (<16 years old, >74 years old, all others), and then randomly assigned to appear in the pre or post-intervention questionnaire.

GPs were given the following instructions at the beginning of the scenario sets: 'The following 10 scenarios differ slightly in various elements which may influence your decision to refer a patient with back pain for an X-ray, including whether or not the patient is "your own" or normally cared for by a partner. We appreciate how unlikely it is that you would see a surgery composed solely of patients with back pain, as this set is arranged. We are also aware that the scenario format means that skills you may normally draw on, such as evaluating non-verbal clues from the patient, cannot be a factor in your assessment. Nevertheless, given this understanding, we hope that you address *each* scenario and make a referral decision and on the accompanying scale ring the number that best indicates how difficult this decision was. We have left space for you to comment on your referral decisions, if you so choose.' Decisions in favour of referring the depicted patients for a lumbar X-ray were summed to create a total score out of a possible 10.

Predictive measures

The theoretically derived measures follow the operationalisation protocols of Ajzen (1991), Bandura (1997, 2000) and Conner and Sparks (1996).

Attitude

Three items that assessed behavioural beliefs that referring for a lumbar X-ray would result in a particular consequence ('reassure the patient', 'allay my uncertainty', 'make me more confident about managing the patient's symptoms') were rated on a 7-point scale from 'likely' to 'unlikely'. There were also three corresponding evaluations of the consequences (e.g. 'Reassuring the patient is…') rated on a 7-point scale from 'important' to 'unimportant'. Higher summed scores represented a stronger attitude in support of referring patients for lumbar X-rays.

Subjective norm

Three items assessed GPs' motivation to comply with persons or groups likely to influence GPs in their decision to refer a patient for a lumbar X-ray ('doing what my colleagues think I should do is'; 'doing what the NHS thinks I should do is'; 'doing what patients think I should do is') rated on a 7-point scale from 'unimportant' to 'important'). Higher summed scores reflected greater perceived social pressure to refer a patient for a lumbar X-ray.

Perceived behavioural control

Four items assessed control beliefs in regard to making a diagnosis without an X-ray, and making the decision to refer a patient for an X-ray, rated on a 7-point scale from 'very difficult' to 'not at all difficult' and the perceived power of each control belief to inhibit or facilitate referring for a lumbar X-ray (e.g. *My making a diagnosis without an X-ray is*) rated on a 7-point scale from 'less likely' to 'more likely'. Higher summed scores represented higher perceived behavioural control over practicing without an X-ray and so lower dependence on X-rays (i.e. behaviour control should, theoretically, be negatively related to referring behaviour).

Self-efficacy

Four items assessed self-efficacy regarding referral ('How confident are you in your ability: to diagnose back problems without an X-ray', 'to treat back problems without an X-ray', 'to present a diagnosis to a difficult or anxious patient without an X-ray', 'to follow guideline recommendations for X-ray referral') rated on a 7-point scale 'extremely confident' to 'not at all confident'. Higher summed scores represent greater self-efficacy in practicing without an X-ray and lower dependence on X-rays.

Decision difficulty

GPs rated how difficult it was for them to make the referral decision for each scenario, on a 10-point scale from 'not at all difficult' to 'extremely difficult'. Higher summed scores represented greater perceived difficulty in making referral decisions in the scenario set.

Analysis

Univariate descriptive statistics were examined for accuracy of data entry. If variables had missing values, these were replaced with the group mean, providing 90% of the measure was completed. The data was examined for univariate outliers using z scores >3.29 ($p = 0.001$). All variables were examined for their approximation to a normal distribution using skewness and kurtosis statistics with cut-off greater than ± 1. An item analysis was performed to examine variables for internal reliability and multidimensionality. Baseline differences and the effects of the interventions were investigated using χ^2, ANOVA (GLM) and ANCOVA. Relationships between predictors and outcome variables were examined using Pearson correlations and stepwise regression analyses. Mediation was tested using the methodology of Sobel (1982) and Baron and Kenny (1986). The analyses for research question 1 were performed using the following parameters:

(i) *Explanatory analyses*: The main purpose of a modelling experiment is to develop, refine and compare interventions. It was acknowledged that GPs in the reminder group who did not order an X-ray in the first round could not receive the educational message. Therefore, to focus on the effectiveness of the interventions whilst removing the possibility of introducing bias by selective deletion of GPs, analyses investigating intervention effects included only GPs who ordered at least one X-ray from the first scenario set, regardless of group.

(ii) *Without the baseline outcome measure as a covariate*: This analysis was performed to explore the effect of only collecting follow up data (i.e. excluding baseline outcome data) in the design of future modelling experiments.

Results

The descriptive statistics for all measures are reported in Table 1.

Response rate and non-response analyses

Two hundred and fourteen GPs completed the baseline (T1) questionnaire. These GPs were sent the post-intervention (T2) questionnaire, which was completed by 152 GPs (71%). There were no significant differences ($p>0.05$) between the 62 GPs who only completed a baseline questionnaire and the 152 who completed the study in: gender ($\chi^2(1,210) = 0.00$, ns); area ($\chi^2(1,213) = 0.49$, ns); trainer status ($\chi^2(1,205) = 0.08$, ns); year since qualified ($t(1,205) = 0.22$, ns); or intervention group ($\chi^2(3,211) = 2.54$, ns). The 152 GPs who completed the study had been qualified an average of 20 years (SD = 7.9 years, ranging from 5 to 37 years), 100 (66%) were male, and 84 (45%) were from Tayside/Fife.

Power

A preliminary power calculation had suggested 100 GPs in total, 25 per group, were required to detect intervention effects. Post hoc power analysis showed that this study ($N = 152$: 2×2 factorial design) had power = 0.99 to detect a large effect size (0.40) and power = 0.86 to detect a medium effect size (0.25) (F-test on means in ANOVA) at alpha = 0.05 (Faul & Erdfelder's (1992) Gpower program).

Outcome measures: item analysis

The Cronbach alphas for T1 and T2 behavioural intention were moderate to high (0.66 and 0.82, respectively). Scores on the individual items followed a similar pattern both pre- and post-intervention, with GPs increasing their referral intention as the number of return visits by patients increased.

The Cronbach alphas for T1 and T2 simulated behaviour were relatively low (0.51 and 0.54, respectively) and were not improved by dropping scenarios least related to the total score. Nevertheless, T1 and T2 simulated behaviour were normally distributed around the mean (3.67, SD = 1.89), with number of decisions to refer ranging from 0 to 9 (T1) and 0 to 10 (T2).

Table 1
Descriptive statistics of measures by group ($N = 214$)

	Audit and feedback				Educational reminder message				Total		
	No A&F		Yes A&F		No ERM		Yes ERM		All participants		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Alpha
T1											
Attitude	26.24	4.36	26.09	4.65	26.32	4.49	26.02	4.51	26.17	4.49	0.53
Subjective norm	8.64	2.63	9.09	2.39	8.74	2.54	8.99	2.50	8.86	2.51	0.55
PBC	18.86	3.36	18.98	3.36	18.86	3.46	18.97	3.25	18.92	3.35	0.63
Self-efficacy	19.31	3.02	19.23	3.57	19.19	3.44	19.35	3.17	19.27	3.29	0.69
Decision difficulty	40.09	12.33	39.53	13.21	40.82	13.05	38.77	12.41	39.81	12.75	0.81
Behavioural intention	5.83	3.62	5.78	3.19	6.05	3.48	5.55	3.33	5.50	2.97	0.71
Simulated behaviour	3.59	2.10	3.70	1.96	3.75	2.04	3.55	2.01	3.64	2.01	0.59
T2											
Attitude	25.66	5.26	25.68	4.87	25.65	5.27	25.69	4.85	25.69	4.85	0.65
Subjective norm	8.90	2.33	8.95	2.29	8.93	2.22	8.92	2.39	8.92	2.30	0.41
PBC	18.54	3.22	19.37	3.16	18.93	3.34	18.98	18.93	19.01	2.95	0.59
Self-efficacy	19.22	3.23	19.42	3.37	19.35	3.29	19.30	19.35	19.35	3.21	0.74
Decision difficulty	41.16	12.11	38.61	13.39	40.31	13.14	39.46	40.31	40.29	12.18	0.78
Behavioural intention	5.39	4.16	4.87	3.66	5.43	3.92	4.82	3.89	4.79	3.82	0.82
Simulated behaviour	3.47	2.07	3.14	2.04	3.60	2.00	3.01	2.08	3.16	1.89	0.54

T1 = Pre-intervention (baseline) measures; T2 = Post-intervention measures; PBC = perceived behavioural control.

An exploratory factor analyses (principal components, varimax rotation) was performed to examine the possibility that the low Cronbach alphas for the simulated behaviour measure was due to the scenarios being multidimensional in nature. Three factors were identified. Factor 1 (patient consequences) contained elements which mainly related to the patients' perception of impact (previous episodes, days off work, distress); Factor 2 (patient demands) contained elements which related mainly to patient demands on the GP (number of previous visits, patient's expectation of an X-ray, patient's desire for an X-ray, patient's hostility toward the GP), and Factor 3 (patient symptoms) had items weighted with clinical symptoms (duration of pain, weakness, red flags, general health). Elements which influenced decisions in favour of referral (derived from the 2 scenarios most often referred in each set) were: age, number of previous visits, general health, medication (not relieving pain), patient's anxiety, patient's expectation of being referred, hostility, patient's desire for an X-ray, and patient distress.

Two scenarios in each set depicted patients who 'should' have been referred based on clinical indices outlined by the guidelines. At baseline, one GP (0.7%) referred only these patients. After the intervention, six GPs (3.9%) referred only these patients.

Equivalence of groups

χ^2 and ANOVA analyses revealed no significant (ns) differences ($p > 0.05$) between the intervention groups for any background variables (gender: $\chi^2(3, 210) = 1.51$, ns; years qualified: $F(3, 203) = 0.30$, ns; % GP Trainers: χ^2 (3, 205) = 1.62, ns; Area: χ^2 (3, 207) = 0.80, ns), pre-intervention cognitions (attitude: $F(3, 210) = 0.10$, ns; subjective norm: $F(3, 209) = 1.04$, ns; perceived behavioural control: $F(3, 209) = 0.04$, ns; self-efficacy: $F(3, 209) = 0.05$, ns; decision difficulty: $F(3, 179) = 1.29$, ns) or outcome variables (behavioural intention: $F(3, 203) = 0.37$, ns; simulated behaviour: $F(3, 207) = 0.26$, ns).

1. Do audit and feedback or educational reminder messages influence GPs' behavioural intention?

There were no significant main or interaction effects of the interventions on behavioural intention in the explanatory analysis. The results remained non-significant when the baseline measure was omitted as a covariate (Table 2).

2. Do audit and feedback or educational reminder messages influence GPs' simulated behaviour?

Both the audit and feedback and educational reminder message interventions significantly influenced simulated behaviour. The results were non-significant when

Table 2
Main treatment effects on outcome and predictive measures: parameter estimates with 95% confidence intervals

Dependent variable	Audit & feedback		Educational reminder message	
	Parameter (*B*)	95% CI	Parameter (*B*)	95% CI
Behavioural intention				
Explanatory analysis [a]	−0.86	−1.87 to 0.15	−0.16	−1.18 to 0.85
Without baseline as a covariate	−1.49	−3.25 to 0.28	−1.16	−2.99 to 0.68
Simulated behaviour				
Explanatory analysis	−0.69	**−1.21 to −0.18**	−0.59	**−1.11 to −0.08**
Without baseline as a covariate	−0.55	−1.33 to 0.24	−0.52	−1.34 to 0.30
Attitude				
Explanatory analysis	0.01	−1.29 to 1.48	0.33	−1.06 to 1.71
Without baseline as a covariate	0.17	−2.18 to 2.52	0.11	−2.34 to 2.56
Subjective norm				
Explanatory analysis	−0.29	−0.82 to 0.24	−0.17	−0.70 to 0.36
Without baseline as a covariate	−0.19	−0.79 to 1.18	−0.32	−0.71 to 1.34
Perceived behavioural control				
Explanatory analysis	1.27	**0.52 to 2.03**	−0.01	−0.85 to 0.66
Without baseline as a covariate	1.18	−0.12 to 2.47	−0.68	−2.03 to 0.68
Self-efficacy				
Explanatory analysis	0.63	−0.17 to 1.44	0.03	−0.78 to 0.83
Without baseline as a covariate	0.81	−1.65 to 2.27	0.11	−1.41 to 1.63
Decision difficulty				
Explanatory analysis	−3.75	**−7.31 to −0.19**	−0.11	−3.68 to 3.47
Without baseline as a covariate	−5.19	−10.85 to 0.46	−2.69	−8.47 to 3.09

[a] $N = 143$: T1 simulated behaviour >0.

the baseline measure was omitted as a covariate (Table 2). There were no significant interaction effects between audit and feedback and educational reminder message in any analysis.

3. Do audit and feedback or educational reminder messages influence GPs' cognitions (attitude, subjective norm, perceived behavioural control, self-efficacy and decision difficulty)?

There was a significant main effect of audit and feedback on GPs' perceptions of control (Table 2). GPs who experienced audit and feedback increased their perceived behavioural control whereas GPs who did not decreased their perceived behavioural control. An increase in perceived behavioural control is the desired direction; higher perceived behavioural control reflects lower dependence on X-rays. There was also a significant main effect of audit and feedback on decision difficulty (Table 2). GPs who experienced audit and feedback found it less difficult to make a referral decision on the post-intervention scenario set compared to GPs who did not experience audit and feedback. However, there were no significant main effects or interaction effects of the interventions on attitude, subjective norm, or self-efficacy.

Further investigation examined whether the effect of the audit and feedback intervention on simulated behaviour was mediated through perceived behavioural control or decision difficulty. Analyses showed perceived behavioural control did significantly mediate the relationship between the audit and feedback intervention and simulated behaviour ($z = 2.35$, $> \pm 1.96$), but not decision difficulty ($z = -1.66$, $< \pm 1.96$).

4. Can cognitive variables derived from psychological models predict GPs' behavioural intention or simulated behaviour?

Attitude, perceived behavioural control, and self-efficacy significantly predicted behavioural intention at the $p<0.05$ level. Decision difficulty predicted behavioural intention at the $p<0.10$ level (Table 3). When attitude, subjective norm, perceived behavioural control, self-efficacy and decision difficulty were entered into a stepwise regression model, only theory of planned behaviour variables entered the regression equation, together predicting 30% of the variance in behavioural intention (Table 4).

With the exception of subjective norm, all of the predictive cognitions derived from psychological theories significantly predicted simulated behaviour, as did

Table 3
Pearson correlations between post-intervention (T2) predictive and outcome variables

Outcomes	Predictive cognitions				
	Attitude	S. Norm	PBC	Self-efficacy	D. Difficulty
Behavioural intention	0.35***	−0.14 ns +	−0.53***	−0.44***	0.18*
Simulated behaviour	0.28***	−0.19*	−0.31***	−0.31***	0.22**

S. Norm = subjective norm; PBC = perceived behavioural control; D. Difficulty = decision difficulty; ns + significant at $p<0.10$; *$p<0.05$; **$p<0.01$; ***$p<0.001$.

Table 4
Results of linear (stepwise) regression analyses predicting GPs' behavioural intention and simulated behaviour[a]

Dependent variable	IV in model	B [b]	95% CI (B)	Beta [c]	Sr^2 [d]	R^2 (Adj.)	df	F
Behavioural intention	PBC	−0.57	−0.76 to −0.38	−0.45***	0.18			
	Attitude	0.15	0.05 to 0.26	0.21**	0.04	0.31 (0.30)	2,142	30.88***
Simulated behaviour	PBC	−0.14	−0.25 to −0.03	−0.22**	0.04			
	S. Norm	−0.18	−0.31 to −0.04	−0.20**	0.04			
	Attitude	0.07	0.01 to 0.13	0.19*	0.03	0.16 (0.14)	3,139	8.78***

[a]All variables are post-intervention (T2). Independent variables (IV) in both analyses were entered in the following order: attitude, subjective norm (S. Norm), perceived behavioural control (PBC), self-efficacy, decision difficulty.
[b]Unstandardized regression coefficients.
[c]Standardized regression coefficients.
[d]Sr^2 = squared semi partial correlations. These represent the unique contribution of each IV to R^2; Shared Variability = R^2 − sum of Sr^2 (Tabachnik & Fidell, 1996); *$p<0.05$; **$p<0.01$; ***$p<0.001$.

decision difficulty (Table 3). When attitude, subjective norm, perceived behavioural control, self-efficacy and decision difficulty were entered into a stepwise regression model, only the theory of planned behaviour variables entered the regression equation, together predicting 14% of the variance in simulated behaviour (Table 4).

Discussion

This study examined whether it was possible to design an intervention modelling experiment to reflect a real-world, definitive randomised controlled trial, testing the effect of audit and feedback and educational message reminders on GPs' X-ray ordering. The presentation format of the NEXUS interventions was reproduced in this study. In NEXUS, the interventions were individually tailored, based on participants' referrals during a set period. In this study, GPs' referral decisions from the scenarios in the first survey provided the platform to enable a similarly tailored 'dosage' in the second survey. Thus, it was found possible to simulate the overall NEXUS design at the experimental level.

However, the NEXUS results were not exactly replicated. In NEXUS, educational reminder messages significantly influenced referral behaviour and audit and feedback did not. In our experiment, both interventions significantly influenced simulated referral behaviour. However, replication of the NEXUS results is not necessarily important for the validation of the intervention modelling experiment methodology. The intended purpose of a modelling experiment is to gauge an intervention's appropriateness for subsequent testing in a service-level randomised controlled trial. Had this modelling experiment been conducted, it would have suggested that audit and feedback and educational reminder message were both reasonable interventions to include in a subsequent service level trial. Why audit and feedback was effective in this experiment and not in NEXUS is an interesting issue. One possible explanation may be an unavoidable methodological difference in the two studies. In the modelling experiment, the time between the delivery of the intervention and completing the outcome measure may have ranged from a few minutes to 2 months, whereas in NEXUS it may have been up to 6 months. Intervention modelling experiments would be an ideal forum to test the hypothesis that variations in the time lag between audit and feedback in implementation trials account for the variable success of this form of intervention.

Another aim of this study was to examine the applicability of psychological models to clinical decision-making. Cognitions relating to X-ray referral performed just as predicted from the theory of planned behaviour and social cognitive theory. Positive attitudes towards referring for an X-ray were significantly related to a greater number of decisions in favour of referring in the behavioural simulation scenarios, as well as a greater general intention to refer—the proxy outcomes for referral behaviour. Lower dependence on X-rays, defined in terms of higher perceived control and higher self-efficacy in diagnosing and treatment management without an X-ray, was significantly related to decisions not to refer along with lower intention to refer. Also, participants with higher motivation to follow the advice of groups supporting fewer referrals (NHS and colleagues) were significantly less likely to refer patients in the scenarios and had lower general intention to refer. It appears that psychological models can indeed be used to effectively predict cognitions relating to a specific clinical behaviour. Since psychological models generally come with methods to both measure and change their component variables—identifying predictors of clinical practice from these models may better inform the design of interventions to influence practice.

When the relationships between each outcome and all of the predictors were further examined using stepwise regression analyses, variables from the theory of planned behaviour predicted 30% of the variance in behavioural intention and 14% of the variance in simulated behaviour. Caution is warranted in making generalisations about the effect on actual practice, in this and other studies, because the correlational and cross-sectional nature of much of the research precludes conclusions about cause and effect. A further aim of this study was to explore whether it was possible to use variables derived from psychological models to understand how the interventions achieve their effects. There was evidence that audit and feedback achieved its effects through increasing GPs' perceptions of control over referring a patient for an X-ray. NEXUS did not investigate cognitions relating to X-ray referral, and so this result cannot be compared. Nevertheless, it is reasonable to posit that audit and feedback failed in the definitive randomised controlled trial because it failed to influence perceived behavioural control, i.e. perceptions of barriers and facilitators of referring. If this experiment had been conducted prospectively, then the interventions could have been modified to specifically address the relevant predictive cognitions, to further increase the chance of a successful definitive randomised controlled trial. We have illustrated this in another intervention modelling experiment, when we showed that dentists' cognitions regarding third molar extraction can be influenced by systematic recall of alternative management of the symptoms (Bonetti et al., 2003).

In terms of the design, interpretation and role of modelling experiments, this intervention modelling experiment raises a number of important issues. The first issue concerns how the proxy outcome is measured. In this intervention modelling experiment, two different measures of outcome were used to capture applied and theoretical aspects of referral intention. Simulated behaviour was designed to mimic GPs' referral behaviour by asking GPs to make decisions about which patients they would refer from a set of scenarios. More in line with how intention is usually operationalised in studies using psychological models, behavioural intention was designed to measure GPs' beliefs about their general intention to refer patients for lumbar spine X-rays. The operationalisation of the outcome did influence the results of this study. The measure of intention derived from GPs' simulated decision-making was significantly influenced by the interventions, whilst the intention belief measure showed no effect of the interventions. It may be that the intention belief measure needs redesigning because it was not sufficiently sensitive. On the other hand, the measure of simulated behaviour might simply be a better proxy for clinical behaviour. The scenario responses required actual decisions, albeit for a simulated patient, whereas the intention belief responses were self-reports of beliefs. Also, despite evidence showing that intention to perform a behaviour is the best predictor of that behaviour, it is widely recognised that an 'intention-behaviour gap' exists. This is the variance in behaviour that is not explained in the theory of planned behaviour, using the simple intention report. Gollwitzer has proposed an additional variable, implementation intention, to bridge this gap (Gollwitzer & Brandstatter, 1997; Orbell, Hodgkins, & Sheeran, 1997). Evidence of implementation intention requires that the behavioural content (i.e. when, where, what) be specified in detail. This would more closely resemble the simulated behaviour measure than that of behavioural intention. So, while the results are not consistent with the simple theory of planned behaviour formulation, they may be consistent with the elaborated implementation intention model that would be expected to produce results closer to the actual behaviour i.e. the NEXUS results. Whatever the reason, it appears that a more sophisticated measure may be required to better proxy clinicians' behaviour than the more usual format when measuring intention.

Further analysis of the simulated behaviour measure revealed that scenario elements associated with high referral rates have little to do with clinical indices and more to do with how GPs deal with patients' demands on them and also with patients' perception of the impact of their symptoms. Two scenarios in each set depicted patients who 'should' have been referred based on clinical indices outlined by the guidelines. While they were ranked among the more frequently referred scenarios, less than 4% of GPs referred only these scenarios. These results suggest that guidelines may need to encompass 'non-clinical' factors in order to increase their relevance to clinical practice.

How data are analysed may also be an issue for intervention modelling experiments. The main objective of this study was to test the potential of the intervention modelling experiment methodology for identifying and understanding intervention effects, so an explanatory analysis was used. In this analysis, which only included GPs who experienced the interventions by ordering at least one X-ray from the first scenario set, both interventions significantly reduced GPs' simulated referrals. When the analyses were performed without the baseline outcome measures as covariates, there were no significant intervention effects. It appears that including a baseline outcome measure enhanced the power of the experiment to detect an effect. The results of this study suggest that, whenever possible, such data should be included in definitive randomised controlled trials, otherwise intervention effects may be underestimated.

The final issue concerns the low initial consent rate for this study (27%), which may bias or affect the generalisability of the results. However, there was no significant difference in the consent rate between the Intervention and Control groups, so the relative estimate of the effect is unlikely to be biased. Also, evidence of the generalisability and ecological validity of the results of this study is that the effect of the educational reminder message intervention did reflect the order and magnitude of the real-world trial on which this experiment was based, showing a reduction of approximately 1 X-ray request per GP.

Conclusion

The results of this study show that it is possible to reproduce implementation interventions at an experimental level, although care needs to taken when addressing delivery differences in the 'real-world'. The results also suggest that psychological theories may help isolate mediators of clinical decision-making, thereby enhancing the potential of intervention modelling experiment methodology. The application of psychological models can help pinpoint the further refinements which may be necessary to develop more potent interventions, with a greater likelihood of success in a service-level randomised controlled trial.

Considerable resources are currently being devoted to encouraging clinicians to implement evidence-based practice using interventions with erratic success records, or no known applicability to a specific clinical behaviour, selected mainly by means of researchers' intuition or optimism. Implementation modelling experiments may not be an infallible filtering system. An intervention that successfully influences a proxy outcome in an intervention modelling experiment may fail to influence behaviour in a definitive randomised controlled trial, as in this study. Nevertheless, the overall results suggest that intervention modelling experiments may have a role as a practical and systematic method of enhancing intervention design and selecting implementation interventions to trial.

Acknowledgements

We are grateful to Tayside Research Network (TayReN) for their support and assistance in this study.

References

Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes*, *50*, 179–211.

Albarracin, D., Johnson, B. T., Fishbein, M., & Muellerleile, P. A. (2001). Theories of reasoned action and planned behavior as models of condom use: a meta-analysis. *Psychological Bulletin*, *127*, 142–161.

Armitage, C. J., & Conner, M. (2001). Social cognition models and health behaviour: a structured review. *Psychology & Health*, *15*, 173–189.

Bandura, A. (1969). *Principles of behaviour modification*. London: Hope, Rinehard & Winston.

Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: Freeman.

Bandura, A. (2000). Health promotion from the perspective of Social Cognitive Theory. In P. Norman, C. Abraham, & M. Conner (Eds.), *Understanding and changing health behaviour from health beliefs to self-regulation* (pp. 299–339). Amsterdam: Harwood Academic Publishers.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Bero, L., Grilli, R., Grimshaw, J. M., Harvey, E., Oxman, A. D., & Thomson, M. A. (1998). Closing the gap between research and practice: an overview of systematic reviews of interventions to promote implementation of research findings by health care professionals. *British Medical Journal*, *317*, 465–468.

Bond, S. (1999). The reassurance value to general practitioners and patients of lumbar spine radiography for low back pain. Unpublished study.

Bonetti, D., Johnston, M., Pitts, N., Deery, C., Ricketts, I., Bahrami, M., Ramsay, C., & Johnston, J. (2003). Can psychological models bridge the gap between clinical guidelines and clinicians' behaviour? A randomised controlled trial of an intervention to influence dentists' intention to implement evidence-based practice. *British Dental Journal*, *195*, 403–408.

Conner, M., & Norman, P. (Eds.). (1996). *Predicting health behaviour*. Buckingham: Open University Press.

Conner, M., & Sparks, P. (1996). The theory of planned behaviour and health behaviours. In M. Conner, & P. Norman (Eds.), *Predicting health behaviour. Research and practice with social cognition models* (pp. 121–162). Buckingham: Open University Press.

Cox, D. N., Anderson, A. S., Reynolds, J., McKellar, S., Lean, M. E. J., & Mela, D. J. (1998). Take five, a nutrition education intervention to increase fruit and vegetable intakes: impact on consumer choice and nutrient intakes. *British Journal of Nutrition*, *80*, 123–131.

Eccles, M., Steen, N., Grimshaw, G., Thomas, L., McNamee, P., Soutter, J., Wilsdon, J., Matowe, L., Needham, G., Gilbert, F., & Bond, S. (2001). Effect of Audit & Feedback and educational reminder messages on primary care radiology referrals. *Lancet*, *357*, 1406–1409.

Faul, F., & Erdfelder, E. (1992). *GPOWER: a priori, post-hoc, and compromise power analyses for MS-DOS (Computer program)*. Bonn, FRG: Bonn University, Department of Psychology.

Fishbein, M., Triandis, H., Kanfer, F., Becker, M., Middlestadt, S., & Eichler, A. (2001). Factors influencing behaviour and behaviour change. In A. Baum, T. Revenson, & J. Singer (Eds.), *Handbook of health psychology* (pp. 3–19). London: Lawrence Erlbaum Associates.

Godin, G., & Kok, G. J. (1996). The theory of planned behavior: a review of its applications to health-related behaviors. *American Journal of Health Promotion*, *11*, 87–98.

Gollwitzer, P. M., & Brandstatter, V. (1997). Implementation intentions and effective goal pursuit. *Journal of Personality and Social Psychology*, *73*, 186–199.

Grimshaw, J. M., Shirran, L., Thomas, R., Mowatt, G., Fraser, C., Bero, L., Grilli, R., Harvey, E., Oxman, A., & O'Brien, M. A. (2001). Changing provider behaviour—an overview of systematic reviews of interventions. *Medical Care*, *39*(8), II2–II45.

Grol, R. (2001). Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Medical Care*, *39*(1146), 1154.

Hardeman, W., Johnston, M., Johnston, D., Bonetti, D., Wareham, N., & Kinmouth, A. (2002). Application of the theory of planned behaviour in behaviour change interventions: a systematic review. *Psychology & Health*, *17*(2), 123–158.

Kendrick, D., Fielding, K., Bentley, E., Miller, P., Kerslake, R., & Pringle, M. (2001). The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomized (unblended) controlled trial. *Health Technology Assessment, 5*(30).

Kerry, S., Hilton, S., Patel, S., Dundas, D., Rink, E., & Lord, J. (2000). Routine referral for radiography of patients presenting with low back pain: is patients' outcome influenced by GPs' referral for plain radiography? *Health Technology Assessment, 4*(20).

Ketley, D., & Woods, K. L. (1993). Impact of clinical trials on clinical practice: example of thrombolysis for acute myocardial infarction. *Lancet*, *342*, 891–894.

Mathews, A. M., Gelder, N. G., & Johnston, D. W. (1981). *Agoraphobia—nature and treatment*. London: Tavistock.

Medical Research Council (MRC) (2000). A framework for development and evaluation of RCTs for complex interventions to improve health.

Norman, P., & Conner, M. (1993). The role of social cognition models in predicting attendance at health checks. *Psychology & Health*, *8*, 447–462.

Orbell, S., Hodgkins, S., & Sheeran, P. (1997). Implementation intentions and the theory of planned behaviour. *Personality and Social Psychology Bulletin*, *23*, 945–954.

Oxman, A. D., Thomson, M. A., Davis, D. A., & Haynes, B. (1995). No magic bullets: a systematic review of 102 trails of interventions to improve professional practice. *Canadian Medical Association Journal*, *153*(10), 1423–1431.

Paul, G. L. (1966). *Insights vs. desensitisation in psychotherapy*. Stamford: Stamford University Press.

Randall, D. M., & Wolff, J. A. (1994). The time-interval in the intention-behaviour relationship: meta-analysis. *British Journal of Social Psychology*, *33*, 405–418.

Royal College of Radiologists Working Party (1995). *Making the best use of a Department of Radiology: Guidelines for doctors* (3rd Ed.). London: The Royal College of Radiologists.

Schuster, M., McGlynn, E., & Brook, R. H. (1998). How good is the quality of health care in the United States? *Milbank Quarterly*, *76*(517), 563.

Sobel, M. E. (1982). Asymptomatic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). Washington, DC: American Sociological Association.

Tabachnik, B. G., & Fidell, L. S. (1996). *Using multivariate statistics*. New York: Harper & Row.