Research article

Randomised controlled trial of a theoretically grounded tailored intervention to diffuse evidence-based public health practice [ISRCTN23257060]

Louise Forsetlund*[1], Peter Bradley[1], Lisa Forsen[2], Lena Nordheim[1], Gro Jamtvedt[1] and Arild Bjørndal[1]

Address: [1]Directorate for Health and Social Affairs, Postbox 8054 Dep, 0031 Oslo, Norway and [2]Norwegian Institute of Public Health, Postbox 4404 Nydalen, 0403 Oslo, Norway

Email: Louise Forsetlund* - louise.forsetlund@shdir.no; Peter Bradley - peter.michael.bradley@shdir.no; Lisa Forsen - lisa.forsen@folkehelsa.no; Lena Nordheim - lena.victoria.nordheim@shdir.no; Gro Jamtvedt - gro.jamtvedt@shdir.no; Arild Bjørndal - arild.bjorndal@shdir.no

* Corresponding author

Abstract

Background: Previous studies have shown that Norwegian public health physicians do not systematically and explicitly use scientific evidence in their practice. They work in an environment that does not encourage the integration of this information in decision-making. In this study we investigate whether a theoretically grounded tailored intervention to diffuse evidence-based public health practice increases the physicians' use of research information.

Methods: 148 self-selected public health physicians were randomised to an intervention group (n = 73) and a control group (n = 75). The intervention group received a multifaceted intervention while the control group received a letter declaring that they had access to library services. Baseline assessments before the intervention and post-testing immediately at the end of a 1.5-year intervention period were conducted. The intervention was theoretically based and consisted of a workshop in evidence-based public health, a newsletter, access to a specially designed information service, to relevant databases, and to an electronic discussion list. The main outcome measure was behaviour as measured by the use of research in different documents.

Results: The intervention did not demonstrate any evidence of effects on the objective behaviour outcomes. We found, however, a statistical significant difference between the two groups for both knowledge scores: Mean difference of 0.4 (95% CI: 0.2–0.6) in the score for knowledge about EBM-resources and mean difference of 0.2 (95% CI: 0.0–0.3) in the score for conceptual knowledge of importance for critical appraisal. There were no statistical significant differences in attitude-, self-efficacy-, decision-to-adopt- or job-satisfaction scales. There were no significant differences in Cochrane library searching after controlling for baseline values and characteristics.

Conclusion: Though demonstrating effect on knowledge the study failed to provide support for the hypothesis that a theory-based multifaceted intervention targeted at identified barriers will change professional behaviour.

Background

According to the evidence-based medicine paradigm the explicit utilisation of scientific information is an important tool to improve the quality of decision-making. Therefore, encouraging such practice is an important aim. It has been recommended that future trials of how to promote evidence-based practice should be embedded in a theoretical framework, identify barriers and facilitating factors within the target group and utilise evidence on effective strategies for behaviour change [1–3]. Such a framework for designing and evaluating complex interventions has subsequently been further elaborated by Campbell and colleagues [4].

The study described in this article is part of a larger project in which we was guided by the above-mentioned framework. The overall aim of the project was to encourage public health physicians in Norway to identify and use relevant scientific evidence in their decision-making and to promote understanding of such information through continuing professional development. The project investigated the extent that public health physicians used research information [5]; identified where public health physicians missed the opportunity to search for research information [6], and identified barriers to change [7]. A multifaceted intervention based on a theoretical model was planned during these stages.

The aim of this study was to evaluate whether a tailored theory-based and multifaceted intervention targeted at the whole process of evidence-based practice increased the explicit integration of research in public health physicians' decision-making. In turn, we wanted to find out whether municipalities were more likely to follow such evidence-based advice and whether this would influence the physicians' reported job-satisfaction.

Methods

Participants

All public health physicians working in municipalities in Norway with more than 3000 inhabitants (N = 332) were invited to participate in the project. The invitation letters explained that project participants would have free access to a library service. In return, they would be asked to return questionnaires and examples of written reports to be used for programme evaluation. We also stated that some participants would be asked to co-operate further during the project period.

Intervention components and theory

The intervention was carried out from April 1999 until the end of January 2001. The multifaceted intervention is illustrated graphically in Figure 1 and detailed in Table 1 (see Additional file 1). The barriers to the use of scientific evidence that we had identified were operationalised in

the intervention model as knowledge, attitudes, self-efficacy and physical access [7]. We also aimed to influence environmentally related barriers like organisational and social context. For example, by offering geographically spread physicians a communication network and establishing a dedicated project team as a point of contact and information service. Rogers' model of innovation diffusion [8] was used to guide the organisation of the different components of the intervention.
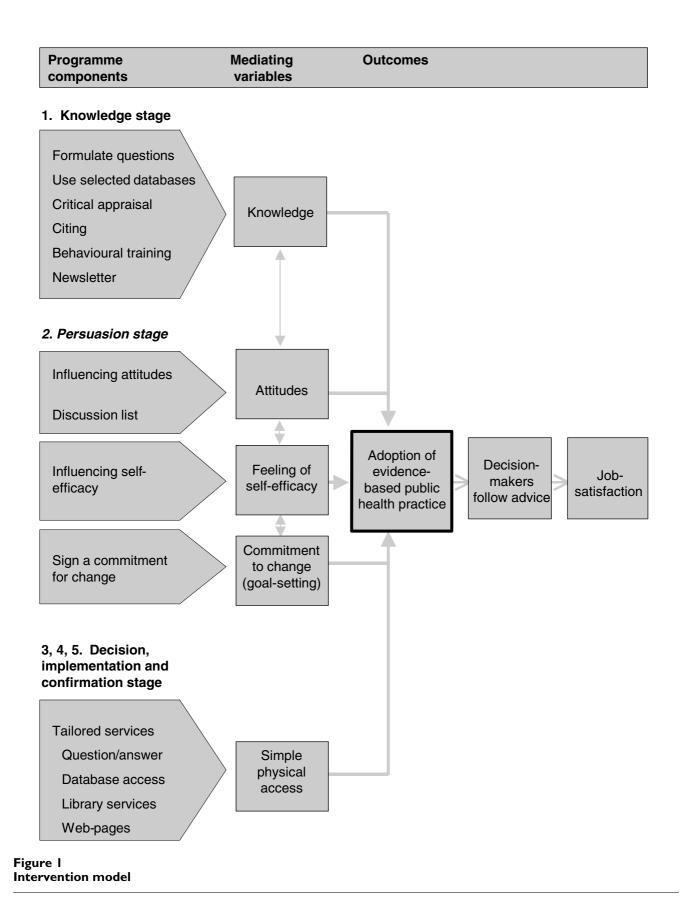
Some of the different strategies previously shown to be effective in changing professional behaviour in some settings were used [2]: multifaceted intervention as such, reminders and feedback (on a general level) and interactive educational meetings [9]. Thus, important components of the intervention were a workshop, an information service, a discussion list and access to several databases.

Rogers defines diffusion as "the process by which an innovation is communicated through certain channels over time among the members of a social system" [8]. In the innovation-diffusion process the individual will first gain knowledge of the innovation, then form an opinion on it, which will be used to adopt or reject it in the decisional stage. The individual's feeling of self-efficacy will also influence the eventual outcome. After the individual decides to adopt the innovation, implementation and confirmation of the decision follow. The intervention sequence was built to lead each participant through each of these five steps. To further influence future task performance, goal setting was used in the intervention as a motivational technique [10]. This involved participants signing a contract about what they would change in their practice. They were informed that they would be asked if they really had made the changes 6 months later.

In contrast, participants in the control group received a letter confirming free access to library services for one year. Because there are no organised library services in Norway for practitioners around the country, this represented a potentially useful service. However, knowing how difficult it is to achieve behaviour change we also assumed that this offer, made in a letter, would be equivalent to no intervention.

Outcome measures

Behaviour was considered the primary outcome and was measured by analysing the contents of local health service reports and of a hypothetical assignment, by a postal survey, a telephone survey and a questionnaire. The questionnaire was also used to measure the other, secondary outcomes: attitudes, knowledge of evidence-based practice information sources and concepts, task-related self-efficacy, decision-to-adopt and job-satisfaction.

Figure 1
Intervention model

Table 2: Internal consistency analysis

	Pilot	Stand. Alpha	Pretest	Post-test
Attitudes	0.83	9 items (out of 13)	0.82	0.83
Self-efficacy	0.84	6 items (out of 6)	0.72	0.73
Decision-to-adopt	0.87	2 items (out of 2)	0.87	0.90
Job satisfaction	0.83	6 items (out of 8)	0.80	0.84

Hypothetical assignment
Participants were asked to write a strategy for patients with serious psychiatric disorders in a medium-sized municipality with particular respect to how suggested measures might be supported. Five questions were added; e.g. how to identify initiatives, where to find relevant information and how to evaluate it. At post-test the topic was changed to accident prevention. The hypothetical assignment was developed through discussion with three experienced public health physicians. The assignment was enclosed with the questionnaire.

Postal survey
Participants were asked at the end of trial whether they had explicitly used research information in any of their written reports in the project period. Two examples were attached. Respondents responding affirmatively were asked to send in relevant documents. Reports on environmental health were excluded because they tended to have a very local focus.

Telephone survey
A report of the effectiveness of external hip protectors was distributed to all in the intervention group, accompanied by the following suggestion to the physician: "Inform the manager at your local nursing home and encourage them to take further action!" We called every nursing home in the appropriate municipalities, and enquired whether the local public health physician had contacted them regarding the use of external hip protectors.

Questionnaire
The questionnaire was based on previous literature [11–18]. In addition to questions on background variables it included items for measuring knowledge, attitude to the use of research information, task-related self-efficacy, decision-to-adopt, job satisfaction and on self-reported behaviour as mentioned above. Concepts from social cognitive theory considered equivalent to the concepts of attitudes, self-efficacy and decision-to-adopt from Rogers' theory of innovation diffusion were used to develop questionnaire items [13,14].

The questionnaire was pilot tested with 126 physicians working in municipalities with less than 3000 inhabitants of which 55 (43%) were returned. For most of the questionnaire, subjects were asked to rate each item on 7-point Likert-like scales ranging from "Strongly disagree" to "Strongly agree". Reversed items were converted for scoring. All the items that considered the concepts of attitude, decision-to-adopt, self-efficacy and job-satisfaction were summed and their means were computed. Thus, overall index measures of each concept with scores ranging from 1–7 were obtained. Some concepts had their number of items reduced. The analysis of internal consistency of scale items based on the 55 pilot test data yielded a Cronbach's alpha score ranging from 0.83 to 0.87, indicating a satisfactory level of agreement (Table 2).

The knowledge construct was divided into knowledge about terms of importance to critical appraisal (concept knowledge) and knowledge about information sources for evidence-based practice (source knowledge). Respondents were asked to grade self-perceived knowledge on scales ranging from 0 to 2 and from 0 to 3 respectively. An additional question was added to concept knowledge, scored as either 0 or 1. Scores were summed and means for individual overall scores for concept and source knowledge were computed.

Scoring
Frameworks for scoring the documents were developed: the planning documents, the hypothetical assignment and the additional question list. The criteria lists were pilot-tested with 10 cases for each document type, then discussed and revised. Then the lists were re-piloted with 10 more cases, after which some smaller changes were made.

Two assessors scored each document independently. The assessors gave a total score for the extent the document reflected the different evidence-based practice-elements that the intervention targeted, ranging from 1–5. Disagreement was resolved by a third party.

Sample size and randomisation

Using a table for sample size determination we specified a power of 80% to detect a medium-sized difference of 0.5 standardized effect size at a significance level of 5%. We found the required sample size to be 62 physicians in each group [19]. Public health physicians were enrolled by one of the authors (LForsetlund) upon receipt of the consenting letter. Enrolled physicians were subsequently randomised to one of two groups by an independent researcher using computer software.

Blinding

The registrar of the questionnaire data was blinded to group allocation. The researchers who scored the other study outcomes were blinded to the allocation of participants and whether the results were pre- or post-tests.

Analysis

The internal consistency for all indexes was estimated by using Cronbach's alpha. Interrater consistency was assessed by agreement in weighted Kappa score for the total document score (scale 1–5) for all three predefined criteria lists at pre- and post-test.

The discriminative validity of the instruments was examined by correlating the scores of each scale to the scores obtained in the others, using Spearman's non-parametric test.

The effect of the intervention was evaluated by t-tests for ordinal (scale) variables. Confidence intervals (95% CI's) were calculated. Binary variables were evaluated by means of Chi-squared analysis. Because of their skewness Mann-Whitney tests were used to compare quantitative discrete variables. The scores (1–5) for the hypothetical assignment and additional questions were also compared by means of the Mann-Whitney test, while the scores for reports were recoded and reported as 'used' or 'not used' research.

Data for all responding participants were analysed on an intention to treat basis, in the sense that even responders who had not received the intervention in full were included in the analysis. For those outcomes where an effect had been shown, sensitivity analyses were conducted by assigning the control group's lowest and average values in turn, to replace missing data in both groups.

Pre- and post-test analyses were planned because of potential threats of attrition and contamination. However, according to Vickers and Altman [20] analysing pre- and post-test change does not control for baseline imbalance because of regression to the mean. They suggest a type of multiple regression analysis (covariance analysis) to adjust each respondent's follow-up score with his or her baseline score. We expanded the model to also include baseline characteristics of possible prognostic strength.

Results

Participant flow

Overall, 148 physicians gave written consent to participate in the project. The randomisation process allocated 73 to the intervention group and 75 to the control group. See Figure 2 for a flow diagram. Six of 73 (8%) physicians from the experimental group withdrew from the project before answering any material. 50 of 73 (68%) physicians attended the workshop while 62 (85%) of them were members of the discussion list. A total of eight physicians had no Internet access and these were sent copies of the reports that were made by the team in the web-based question-and-answer service.

No control group participant explicitly withdrew from the project, but 7 physicians could not be contacted at follow-up because they had changed job or were on prolonged leave. One physician who had been randomised to the control group was in fact not a public health physician. He was treated as a non-responder.

Recruitment

Recruitment took place between January 1999 and January 2000. After randomisation, the participants were sent the baseline assessment forms. Follow-up measurements were started immediately at the end of the intervention.

Baseline data

Baseline characteristics revealed a possible imbalance for some variables (sex, number of years as a public health physician, specialist status, previous exposure to courses in critical appraisal and number of advisory reports written during the previous half year) (Table 3).

Numbers analysed

Response rates for all instruments varied from 59% to 83% at pre-test and from 57% to 100% at post-test, except for the response rate for reports which was 23% for the experiment group and 33% for the control group (Table 4). One questionnaire response in the intervention group at post-test was excluded because the majority of questions were not answered, so 58 were analysed. In the intervention group 49 (67%) and in the control group 53 (71%) answered the questionnaire at both pre- and post-test and were included in the regression analysis.

Outcomes and estimation

Analysis of internal consistency of scale items was repeated on the pre- and post-test material yielding an alpha between 0.73 and 0.90 at post-test (Table 2). The weighted Kappa scores for interrater agreement on use of research information for reports, hypothetical assignment
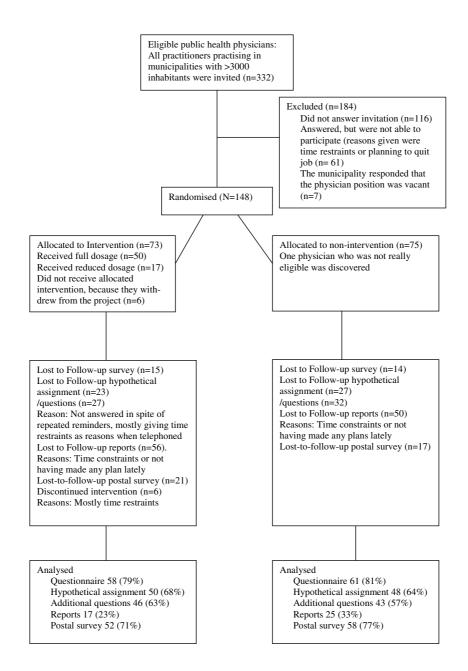
Eligible public health physicians:
All practitioners practising in
municipalities with >3000
inhabitants were invited (n=332)

Excluded (n=184)
    Did not answer invitation (n=116)
    Answered, but were not able to
    participate (reasons given were
    time restraints or planning to quit
    job (n= 61)
    The municipality responded that
    the physician position was vacant
    (n=7)

Randomised (N=148)

Allocated to Intervention (n=73)
Received full dosage (n=50)
Received reduced dosage (n=17)
Did not receive allocated
intervention, because they with-
drew from the project (n=6)

Allocated to non-intervention (n=75)
One physician who was not really
eligible was discovered

Lost to Follow-up survey (n=15)
Lost to Follow-up hypothetical
assignment (n=23)
/questions (n=27)
Reason: Not answered in spite of
repeated reminders, mostly giving time
restraints as reasons when telephoned
Lost to Follow-up reports (n=56).
Reasons: Time constraints or not
having made any plan lately
Lost-to-follow-up postal survey (n=21)
Discontinued intervention (n=6)
Reasons: Mostly time restraints

Lost to Follow-up survey (n=14)
Lost to Follow-up hypothetical
assignment (n=27)
/questions (n=32)
Lost to Follow-up reports (n=50)
Reasons: Time constraints or not
having made any plans lately
Lost-to-follow-up postal survey (n=17)

Analysed
    Questionnaire 58 (79%)
    Hypothetical assignment 50 (68%)
    Additional questions 46 (63%)
    Reports 17 (23%)
    Postal survey 52 (71%)

Analysed
    Questionnaire 61 (81%)
    Hypothetical assignment 48 (64%)
    Additional questions 43 (57%)
    Reports 25 (33%)
    Postal survey 58 (77%)

Figure 2
Flow chart

Table 3: Baseline demographic and other characteristics of control and intervention groups. Values are numbers (percentages of participants) and means (SD)

Baseline measure	Intervention group n = 59 (%)		Control group n = 62 (%)	
Demographic				
Women	8	(14)	17	(27)
Men	50	(86)	45	(73)
Specialist (yes/no)	37	(64)	30	(48)
Mean (SD) Size of municipality (no.inhabitants)	20137	(26421)	18494	(33391)
Mean (SD) age (years)	47	(6.4)	47	(7.9)
Mean (SD) Public health weekly working hours	16.4	(9.8)	17	(9.6)
Mean (SD) Experience (years as publ.health phys.)	12	(8.5)	9.5	(8.6)
Other characteristics				
Back ground variables				
Access to Internet (office/home)	52	(88)	52	(85)
Access to medical library	10	(18)	13	(22)
Access to Cochrane	5	(10)	5	(9)
Attended session(s) on searching (yes/no)	14	(24)	14	(23)
Attended session(s) in critical appraisal (yes/no)	24	(42)	18	(30)
Mean (SD) Data skill scale (1–7)	4.7	(1.9)	4.3	(1.9)
Mean (SD) Number of written reports	14.5	(15.1)	11.3	(10.8)

Table 4: Response rates at pre- and post-test for all instruments

Pre-test						
	Questionnaire	Hypothetical assignment	Additional questions	Reports		
	N (%)	N (%)	N (%)	N (%)		
Intervention group	59 (81)	49 (67)	45 (62)	43 (59)		
Control group	62 (83)	51 (68)	47 (63)	57 (76)		
Post-test						
	Questionnaire	Hypothetical assignment	Additional questions	Reports	Postal Survey	Telephone Survey
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Intervention group	59 (81)	50 (68)	46 (63)	17 (23)	52 (71)	73 (100)
Control group	61 (81)	48 (64)	43 (57)	25 (33)	58 (77)	75 (100)

and additional questions were 0.50, 0.91 and 0.87 at pre-test respectively and 0.89, 0.75 and 0.74 at post-test.

In the discriminative analysis the instrument for attitude demonstrated a small, though significant correlation to the self-efficacy, decision-to-adopt and job-satisfaction instruments (Table 5). No other correlations were demonstrated.

Primary outcomes
No evidence of differences for any of the objective behavioural variables could be observed at follow up, though a slight tendency for the intervention group to use research

to a somewhat greater extent could be observed although not to a level that was statistically significant (Tables 6 and 7).

The responses to the question ' number of times searching Cochrane' (or Medline) overestimated the number of searches compared to the search logs. It is presumably easier to remember having searched or not searched than how many times. The variables were therefore recoded to 'having made a search' or 'not having made a search' and analysed. There were statistically significant differences between groups for self-reported searching in the Cochrane database (χ^2 = 6.3, df = 1, p = 0.01) but not in

Table 5: Discriminant analysis using Spearman's correlation coefficient

	Attitudes	Self-efficacy	Decision-to-adopt	Job-satisfaction
Attitudes	-	0.3**	0.3**	0.2*
Self-efficacy	0.3**	-	0.1	0.2
Decision-to-adopt	0.3**	0.1	-	0.1
Job-satisfaction	0.2*	0.2	0.1	-

* Correlation is significant at the 0.05 level (2-tailed) ** Correlation is significant at the 0.01 level (2-tailed)

Table 6: Differences between groups for using research to some extent (tested by means of Mann-Whitney)

	Intervention			Control			
Behaviour	Number of respondents	Mean score	(SD)	Number of respondents	Mean score	(SD)	P
Hypothetical assignment	(50)	2.1	(1.3)	(48)	1.8	(1.2)	0.154
Additional questions	(46)	2.2	(1.4)	(43)	1.7	(1.0)	0.063

Table 7: Differences between groups for using research to some extent

Behaviour	Intervention (N = 73)			Control (N = 75)		
	(N) (= number of respondents)	n (= number using research to some degree)	(% of total = 73)	**(N)** (= number of respondents)	n (= number using research to some degree)	(% of total = 75)
Reports	(17)	0	(0)	(25)	1	(1)
Postal survey:						
Advice-giving documents	(52)	3	(4)	(58)	0	(0)
Telephone survey:						
Giving information on hip protectors to nursing homes	(73)	2	(3)	(75)	0	(0)

Medline ($\chi^2 = 0.1$, df = 1, p = 0.74) (Table 8). The sensitivity analysis (worst case scenario) rendered a narrowly significant result for searching the Cochrane database ($\chi^2 = 4.0$, p = 0.047). There was no evidence of differences in self-reporting of:

- number of articles ordered or critically appraised,

- number of problems identified as relevant for the use of research,

- number of instances when research was of help in decision-making,

- or number of cases where the physician experienced that the advice given was followed.

Secondary outcomes

Table 9 describes the effect of the intervention at post-test for the secondary outcomes. There were statistically significant differences between the groups for knowledge about information sources (mean diff = 0.4, 95% CI = 0.2 to 0.6) and knowledge about concepts (mean diff = 0.2, 95% CI = 0.0 to 0.3) but not for attitudes, self-efficacy, decision-to-adopt or job-satisfaction. Assigning the lowest value in the control group (0) to the missing values of the knowledge variables of both groups still rendered significant results for source knowledge (mean diff = 0.3, 95% CI = 0.1 to 0.5). The results for concept knowledge, however, became non-significant (mean diff = 0.1, 95% CI = -0.1 to 0.3). Assigning the mean value of the control group (1.1) to missing values of concept knowledge rendered a significant difference (mean diff = 0.2, 95% CI = 0.0 to 0.3).

Table 8: Differences at post-test between groups for self-reported searching of Cochrane and Medline. Chi square test

	Intervention	Control	DF	χ^2	P
	(N)	(N)			
Searched Cochrane	(55) 34	(60) 23	1	6.3	0.01
Searched Medline	(55) 31	(60) 32	1	0.1	0.74

Searched Cochrane: 'yes' (1), 'no' (0)
Searched Medline: 'yes' (1), 'no' (0)

Table 9: Student t test of differences between groups at post-test

	Intervention (N = 58) Mean (SD)	Control N = 61 unless otherwise stated) Mean (SD)	Mean diff	95% CI	t	DF	P
Source knowledge	1.1 (0.6)	0.7 (0.5)	0.4	0.2–0.6	4.3	111.5	0.00
Concept knowledge	1.3 (0.4)	1.1 (0.4)	0.2	0.0–0.3	2.6	115.3	0.01
Attitudes	5.4 (0.8) (n = 56)	5.2 (0.7)	0.1	-0.2–0.4	0.9	115	0.37
Decision-to-adopt	4.9 (1.2)	5.1 (0.9)	-0.2	-0.6–0.2	-0.9	97.8	0.35
Self-efficacy	4.0 (0.9)	3.9 (0.9)	0.1	-0.2–0.4	0.5	116.9	0.60
Job-satisfaction	4.3 (1.3)	4.0 (1.2)	0.3	-0.1–0.8	1.5	114.6	0.13

Knowledge of sources: Mean of additive score of 0 = 'unknown', 1 = 'known, but not used', 2 = 'read', 3 = 'used in a public health decision-making situation'. *Knowledge of concepts* : Mean of additive score of 0 = 'unknown', 1 ='known', 2 = 'so known that I can explain to others' + an extra point (1) if correctly answering "Method chapter" as to what is the most important chapter for deciding scientific quality of an article. *Attitudes:* Likert scale: 1 = 'totally disagree', 2 = 'disagree', 3 = 'partly disagree', 4 = 'neither agree nor disagree', 5 = 'partly agree', 6 = 'agree', 7 = 'totally disagree'. Decision-to-adopt: Likert scale: 1 = 'totally incorrect', 2 = 'incorrect', 3 = 'Somewhat incorrect' 4 = 'neither right nor wrong', 5 = 'somewhat correct', 6 = 'correct', 7 = 'totally correct'. *Job-satisfaction:* Same Likert scale as attitudes.

Ancillary analyses
The variables in the regression model were the group variable, baseline score and the variables demonstrating a potential important imbalance between the groups. The analysis changed the result for the self-reported variable 'searching Cochrane', which became non-significant. There was no substantial change for the other two significant results (data not shown).

Discussion
Interpretation
This study is of interest because it is the first empirically and theoretically based tailored multifaceted intervention for diffusing the whole process of evidence-based practice in a randomised-controlled design. The intervention had some effect on knowledge reported. This supports the conclusion from a recent systematic review [21] that teaching critical appraisal skills in health care settings has positive effects on participants' knowledge. However, even when combining teaching with an intervention encompassing the whole process of evidence-based practice (and not just critical appraisal) including supportive

elements like an information service, discussion list and newsletter, there was no evidence of impact on decision-making. Most importantly, this study does not support the hypothesis that a multifaceted intervention targeted at selected barriers changes professional behaviour [22].

According to diffusion-theory "the rate of awareness-knowledge for an innovation is more rapid than its rate of adoption" [8]. Innovations that can be tested and are simple and compatible with previous experience and practice have a shorter innovation-decision period. Measuring performance after a period of 1.5 year may still have been a too short time perspective. It appears that our intervention successfully led the participants through the stage of increasing knowledge, but did not reach the stage of persuasion. A change in knowledge is a necessary but insufficient criterion for changing practice, and, as it seems, also for changing attitudes and feeling of self-efficacy. The lack of evidence of effect on the variables 'advice followed' and 'job satisfaction' (Figure 1) is predictable from the lack of evidence of effect on practice. Although 43 out of 47 (3 missing) stated goals on leaving the workshop for how

they would adopt evidence-based practice, this did not seem to strengthen the change process. A meta-analysis by Wood et al. [23] reported that goal-setting effects are maximised for easy tasks. Since the majority of public health tasks are complex, one might anticipate a modest effect.

The adjustment analysis by multiple regression analysis did not change the interpretation of our results regarding the intermediate variables. The logistic regression analysis of the self-reported searching of Cochrane is more difficult to interpret, since the change in results may be due to a loss in power when including only those who have answered both pre- and post tests.

Limitations of the study
Statistical validity
Some relevant potential threats to the statistical conclusion validity of our study could be: low statistical power, unreliability of measures and unreliability of treatment implementation [24]. As for the first threat, the probability of making a faulty no-difference conclusion, i.e. a Type II error, increases when sample sizes are small. In our study the response rate for reports at post-test was especially low (Table 7). We could have made a greater effort to obtain more documents and thus increased the amount of data collected. However, we chose not to pursue this matter, since we received the same information through the postal survey: We are reasonably confident that the physicians would have reported being involved in writing either types of documents (reports and advice-giving documents). In addition, behaviour was also measured by the telephone survey.

The reliability of the instruments measuring the constructs; attitudes, self-efficacy, decision-to-adopt and job-satisfaction was tested for internal consistency and was satisfactory. Likewise, the weighted Kappa measure of inter-rater consistency for the use of criteria lists was of adequate size. The variables 'searching Cochrane/Medline' (Table 8) were checked against the search logs.

Several of the null hypotheses regarding outcome variables were not rejected (Table 9). Recalculating the power with the variances obtained in the study shows that the size of the study was big enough to detect 0.5 SD changes with more than 80% power, as intended. Though the changes in the non-significant results are less than 0.5 SD, the confidence intervals are so wide (Table 9) that we cannot accept the null hypothesis on the basis of the statistical analysis. On the other hand, the 'Users' guide to the medical literature' states that if the upper boundary of the confidence interval excludes any important benefit of the intervention, one may conclude that the trial is negative [25].

With this type of study there is always some difficulty of standardising the implementation of the intervention. According to Cook and Campbell [24] lack of standardisation will inflate error variance and decrease the chance of obtaining true differences. On the other hand, lack of standardisation is typical for pragmatic trials and reflects real situations [26]. There is a theoretical possibility that the intervention was never really adequately implemented, e.g. the quality of the educational part of the intervention may have been insufficient regarding both teaching methods and duration.

Internal validity
The risk of contamination between groups was felt to be limited since public health physicians in Norway are geographically scattered; one physician in each of the country's 435 municipalities. This initial assumption was supported by the fact that none of the physicians in the control group were recorded to use the library services offered. However, during the intervention period evidence-based practice was discussed in other public health settings. This may have influenced the general level of knowledge on the topic.

For those who provided post-test data, the response rates were fairly similar between the groups. Some physicians had changed jobs and some stated they did not have time, but there was no evidence of a differential attrition between the groups.

Construct validity
It is debatable how far the operalisations of the theoretical construct 'multifaceted intervention' on the input side, actually reflected this construct and whether the measurements of dependent variables really did measure what they were meant to measure. However, the theoretical foundation should to some extent account for face and content validity. Moreover, the discriminant validity of the instruments measuring attitudes, self-efficacy, decision-to-adopt and job-satisfaction was shown to be satisfactory by the low correlation between each of these indexes.

By using alternative measures of the primary outcome, with different means of recording responses (Tables 6,7), a potential threat from mono-method bias should have been met. The experiment group could, however, have guessed the hypothesis of the study to a greater extent than the control group. The differences we found in knowledge might reflect either this or the greater attention given to the experiment group.

Generalisability
The study sample contained highly motivated and interested physicians with some skills in data technology and

working experience in rural and urban settings. Considering that this group could be characterized with Rogers' terminology as 'innovators' or 'early adopters' the results are rather disappointing.

Conclusion

The multi-faceted intervention demonstrated effect on knowledge, but failed to demonstrate any other positive effects on the intermediate steps required to disseminate and implement (diffuse) new practice according to Roger's theoretical model. It is therefore not surprising that practitioners did not increase the use of evidence in practice.

Efforts to promote evidence-based practice could be strengthened by utilising networks and infrastructures that already exist. First and foremost, evidence-based methodology should become an integral part of undergraduate and continuing medical education. Central and local authorities, which support public health physicians, should use evidence-based methods to inform decision-making, for example in central strategy documents. We suspect, however, that this requires a culture shift regarding the perceived necessity for utilising research information on health issues.

The reasons underlying the program's failure to demonstrate any further effect cannot be illuminated by a randomised controlled design. As discussed by Wolff [27] and others [28,29] there may be some inherent problems in using the randomised trial design to evaluate social complex interventions. Moreover, effectiveness evaluations do not give much information on or understanding of the processes involved between program delivery and outcome [30]. A qualitative investigation of these processes may increase understanding and is, in this case, already in progress.

Competing interests
None declared.

Authors' contributions
LForsetlund participated in the conception and design of the trial, as well as analysing and interpreting data and writing the article. PB participated in the trial intervention, in the drafting, editing and critical revision of the article. LForsen participated in the analysis and interpretation of the data and in the critical revision of the article. LN participated in the trial intervention, collecting of data and in the critical revision of the article. GJ participated in the trial intervention and in the critical revision of the article. AB was responsible for the conception and design of the whole trial, and participated in the drafting and critical revision of the article. All authors read and approved the final manuscript.

Additional material

> ### Additional file 1
> Intervention goal and content
> *A description of the goal, timing and content of the intervention and which media we used.*
> Click here for file
> [http://www.biomedcentral.com/content/supplementary/1472-6920-3-2-S1.doc]

References
1. Grol R and Grimshaw J **Evidence-based implementation of evidence-based medicine.** *Jt Comm J Qual Improv* 1999, **25:**503-513
2. NHS Centre for Reviews and Dissemination **Getting evidence into practice.** *Effective Health Care* 1999, **5(1):**1-16
3. Moulding NT, Silagy CA and Weller DP **A framework for effective management of change in clinical practice: dissemination and implementation of clinical practice guidelines.** *Qual Health Care* 1999, **8:**177-183
4. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D and Tyrer P **Framework for design and evaluation of complex interventions to improve health.** *BMJ* 2000, **321:**694-696
5. Forsetlund L and Bjørndal A **Har samfunnsmedisinere tilfredsstillende tilgang til viktige informasjonskilder? [Do public health practitioners have satisfactory access to important information sources?]** *Tidsskrift for Den Norske Laegeforening* 1999, **119(17):**2456-2462
6. Forsetlund L and Bjørndal A **The potential for research-based information in public health: identifying unrecognised information needs.** *BMC Public Health* 2001, **1:**1[http://www.biomedcentral.com/1471-2458/1/1]
7. Forsetlund L and Bjørndal A **Identifying barriers to the use of research faced by public health physicians in Norway and developing an intervention to reduce them.** *J Health Serv Res and Pol* 2002, **7(1):**10-18
8. Rogers EM **Diffusion of innovation.** *New York, The Free Press* 1995,
9. Davis D, Thomson O'Brien MA, Freemantle N, Wolf FM, Mazmanian P and Taylor-Vaisey A **Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes?** *JAMA* 1999, **282:**867-874
10. Locke EA and Latham GP **A theory of goal setting & task performance.** *Englewood Cliffs, Prentice-hall* 1990,
11. Jacobs AM, Young DM and Dela Cruz FA **Evaluating prototype nursing continuing education programs.** *In: Measurement of nursing outcomes: Measuring nursing performance (Edited by: Strickland OL, Waltz FC) New York, Springer* 1988, **2:**349-363
12. McColl A, Smith H, White P and Field J **General practitioners' perceptions of the route to evidence based medicine: a questionnaire survey.** *BMJ* 1998, **316:**361-365
13. Schwarzer R and Fuchs R **Self-efficacy and health behaviours.** *In: Predicting health behaviour: research and practice with social cognition models (Edited by: Conner M, Norman P) Buckingham, Open University Press* 1996, 163-196
14. Conner M and Sparks P **The theory of planned behaviour and health behaviours.** *In: Predicting health behaviour: research and practice with social cognition models (Edited by: Conner M, Norman P) Buckingham, Open University Press* 1996, 121-155
15. Spector PE **Summated rating scale construction.** *London, Sage Publications 1992. Sage university papers series: Quantitative applications in the social sciences, no.82*

16.   Henerson ME, Morris LL and Fitz-Gibbon CT **How to measure attitudes.** *London, Sage* 1987,
17.   Sanders GL **MIS/DSS success measure: Systems objectives and solutions** 1984, **4:**29-34[http://wings.buffalo.edu/mgmt/courses/mgtsand/success/success.html] [cited 2001 Oct 18].
18.   Seashore SE, Lawler III EE, Mirvis PH, Camman C and editors **Assessing organizational change: A guide to methods, measures, and practices.** *New York, Wiley 1983. Wiley series on organizational assessment and change*
19.   Hinkle DW, Wiersma W and Jurs SG **Applied statistics for the behavioral sciences.** *Boston, Houghton Mifflin Company* 1988,
20.   Vickers AJ and Altman DG **Analysing controlled trials with baseline and follow up measurements.** *BMJ* 2001, **323:**1123-1124
21.   Parkes J, Hyde C, Deeks J and Milne R **Teaching critical appraisal skills in health care settings (Cochrane Review):** *In: The Cochrane Library, Issue 3, 2001. Oxford: Update Software*
22.   Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, Grilli R, Harvey E, Oxman A and O'Brien MA **Changing provider behavior: an overview of systematic reviews of interventions.** *Med Care* 2001, **39:**II2-45
23.   Wood RE, Mento AJ and Locke EA **Task complexity as a moderator of goal effects: A meta-analysis.** *J Appl Psychol* 1987, **72(3):**416-425
24.   Cook TD and Campbell DT **Quasi-experimentation: Design & analysis issues for field settings.** *London, Houghton Mifflin Company* 1979,
25.   Guyatt G and Rennie D **Users' guides to the medical literature: a manual for evidence-based clinical practice.** *Chicago, AMA Press* 2002,
26.   Roland M and Torgerson DJ **Understanding controlled trials: what are pragmatic trials?** *BMJ* 1998, **316:**285
27.   Wolff N **Randomised trials of socially complex interventions: promise or peril?** *J Health Serv Res Policy* 2001, **6:**123-126
28.   Norman GR and Schmidt H **Effectiveness of problem-based learning curricula: theory practice and paper darts.** *Med Ed* 2000, **34:**721-728
29.   Prideaux D **Researching outcomes of educational interventions: a matter of design.** *BMJ* 2002, **324:**126-127
30.   Lipsey MW **Theory as method: Small theories of treatments.** *New Directions for Program Evaluation* 1993, **Spring(57):**5-38

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1472-6920/3/2/prepub