Process evaluation of a cluster randomized trial of tailored interventions to implement guidelines in primary care—why is it so hard to change practice?

Signe Flottorp, Kari Håvelsrud and Andrew D Oxman


**Background.** A cluster randomized trial of tailored interventions to support the implementation of guidelines for sore throat and urinary tract infection found little or no change in the main outcomes, which were antibiotic prescriptions, use of laboratory tests and use of telephone consultations. There was great variation between the practices in the change in these outcomes.

**Objectives.** Our aim was to evaluate how the interventions were received and to understand why practices did or did not change.

**Methods.** The trial was conducted in general practices in Norway. Data for this process evaluation were collected from the 120 practices that completed the trial. Multiple methods were used: observations, semi-structured telephone interviews, a postal survey and data extracted from electronic medical records. We investigated factors that might explain a lack of change, including: agreement with the guidelines; communication within each practice; degree of participation in the project; taking time to discuss the guidelines and their implementation; use of the components of the interventions; and routines for telephone consultations. Possible explanatory factors were explored in relation to variation in change and the overall extent of change in rates of use of antibiotics, laboratory tests and telephone consultations.

**Results.** Sixty-three per cent of practices agreed with the guidelines. Only 35% reported having regular meetings, and 33% discussed the project before its start, although 75% reported agreement about participating within the practice. Only 33% reported meeting to discuss the guidelines. Use of the components of the interventions ranged from 11% for the increased fee for telephone consultations to 48% for the computerized decision support. Forty-four per cent reported problems with telephone routines. No single factor explained the observed variation in the extent of change across practices.

**Conclusions.** Inadequate time, resources and support were the most salient factors that might explain a lack of change. Problems with internal communication and telephone routines were important contributing factors in many practices.

**Keywords.** Evaluation studies, guideline adherence, guidelines, primary health care, professional practice.

**Introduction**

We conducted a cluster randomized trial with 142 general practices in Norway to evaluate the effects of tailored interventions to support the implementation of guideline lines for the management of urinary tract infections and sore throat.1 Randomly selected practices that were invited consented to participate in the trial. The interventions were tailored to overcome identified barriers to change.2 The recommendations and the interventions, which had little effect on the main outcomes, are described in Box 1.

Rigorous methods are needed to evaluate the effect of interventions to change practice.3 However, implementation research should not only focus upon the effects on the main outcomes. Supplementary information is needed to describe what really happens in a trial, and to
create or explore hypotheses about why and how interventions are effective or ineffective.\textsuperscript{4–7} Process evaluations have been used in studies of health promotion and health education, in the social sciences and education research, but they have not been utilized commonly in implementation research. Both qualitative and quantitative methods are potentially useful to open ‘the black box’. We report here a process evaluation that was planned and conducted parallel to the trial, in which we have investigated factors that might explain the changes that were observed for the main outcomes of the trial. These factors included: agreement with the guidelines; communication within each practice; degree of participation in the project; taking time to discuss the guidelines and their implementation; use of the components of the interventions; and routines for telephone consultations. Other possible explanations for the lack of the effect that were explored include the need for more active interventions, a longer follow-up period and better economic compensation.

Methods

Our background and preconceptions
We have worked in general practice for several years. Our experience from daily practice made it easy to sympathize with busy practitioners not always responding to the project’s requirements. Although the practices were volunteers, we expected them to give priority to treating patients.

Data sources
Data were collected from four sources.

Observation. Written notes were taken from nearly all contacts between members of the project team and the participating practices. Technical information from every practice to follow-up the installation of software and the returning of the disks with data was collected routinely in an Excel file. The project assistant (KH) had close contact with all the practices via telephone and fax throughout the project. These contacts were discussed continuously in the project team, both to adjust the interventions and to understand what happened. Based on our informal discussions and an analysis of the records from the contacts with the practices, we established some factors that might explain variations in the main outcomes or a lack of change. These factors were formulated as five questions:

(i) Did the practice seem to be engaged in the project?
(ii) Was the internal communication in the practice good?
(iii) Did the GP assistants seem to be involved in the project?
(iv) Was the project implemented without problems?
(v) Did the software work technically ok?

The answers were scored on 5-point scales by KH for each practice.

Telephone interviews. A questionnaire with both open-ended and closed questions was mailed to each practice during the intervention phase of the project in September 2000. The interventions were initiated in May 2000 and ended in January 2001. Participants were asked to discuss the questionnaire and to try to reach agreement on their responses within each practice. KH collected the answers by telephone interview between October and December 2000. The telephone interviews were audio taped. SF listened to the tapes and transcribed verbatim the answers to the open-ended questions.

Postal survey. A second questionnaire with both open-ended and closed questions was developed after we knew the results of the trial. We asked participants to rate the importance of possible explanations for a lack of effect...
on 5-point Likert scales. Questionnaires were distributed by mail in June 2001. We sent one questionnaire to each GP and one for all the GP assistants in each practice that completed the trial. Non-responders were reminded twice by telephone and by fax to return the questionnaire.

Electronic medical records. Data for the main outcomes of the trial were extracted from electronic medical records.1

Analyses
SPSS 10.0 and SAS 8.2 were used to analyse the data. Means for responses in the two intervention groups were compared using the Mann–Whitney test. Results were analysed separately for each of the five main outcomes (Box 1). We calculated regression coefficients for each of the explanatory variables. For the ordinal variables, normality was assumed, whereas non-ordered categorical variables were treated as such.

Combinations of explanatory variables were tested in multiple weighted regression analyses. No formal adjustments were made for multiple comparisons, but we took account of this and of the limited power of the analyses in interpreting the results.

Results

Response rate
We conducted telephone interviews with 112 of the 120 practices (93%). The respondents had discussed the questionnaire in 69% of the practices before the telephone interview. At least one respondent from 99 practices (83%) returned the second questionnaire; the GP assistants in 72 of the 120 practices (60%) responded, and at least one of the physicians responded from 89 of the practices (74%). Overall, 136 of 265 questionnaires (51%) sent to participating GPs were returned.

Participation
Although our intention was that all the GPs and assistants in each practice should participate in the project, this was not an inclusion criterion. Everyone participated in 67% and all GPs participated in 89% of the included practices (Table 1). There was disagreement about participation in 8% of the practices (Table 2). Based on our observations, GP assistants did not seem to be involved in the project in 45% of the practices, and 18% of the practices did not seem engaged in the project (Table 2).

Discussion of the guidelines and engagement in the project
Half of the practices had hardly met to discuss the guidelines and a change in their routines when the first survey was performed (Table 2). Twenty-nine practices had held one or two meetings, while 16 practices had organized 3–5 meetings. Ten per cent reported that they had spent no time at all discussing the guidelines, and 52% spent <1 h. In the second survey, 33% agreed that time constraints might explain the lack of effect of the interventions.

Most of the practices were satisfied with their participation in the project and thought it had been useful to try to improve practice for a common clinical problem (Table 2). In the second survey, 52% of the respondents agreed that the interventions might have been more effective if we had been able to visit all the practices (Table 3). Thirty-eight per cent reported that they had started a process to change practice, but they needed more time. Thirty-nine per cent stated that they did not need to change their practice. However, self-reported need to change practice was not correlated with agreement with the guidelines, baseline rates for the main outcomes or change in practice.

Agreement with the guidelines
Although the guidelines differed from traditional care in Norwegian general practice, most of the practices agreed with the guidelines according to both surveys. In the first survey, 4% disagreed and in the second 8%, while 63 and 68% stated that they agreed (Table 2). Eighty-seven per cent reported that they tried to follow the guidelines.

Organizational factors
Many of the practices needed reminders to install the software, return the disks with data and to give responses to the questionnaires. It was difficult to get through to many of the practices by phone. Not everyone received the project newsletters, although we sent copies for everyone in each practice.

Thirteen per cent of the practices had serious problems with internal communication within the practice based on our observations, and it was difficult to run the project in 25% of the practices (Table 2).
Twenty per cent of the practices met weekly and 16% quite often, while 37% seldom or never met to discuss their routines in clinical problems. Assistants participated in these meetings in 67% of the practices. Clinical problems were rarely or never discussed in a common forum in 20% of the practices (Table 2).

Difficult access by telephone might have hindered an increased use of telephone consultations according to 35% of the respondents to the second questionnaire (Table 3). Eighteen per cent thought that economic disincentives were an impediment for reducing the use of laboratory tests.

Use of the interventions
Eleven of the 61 practices in the sore throat group (18%) and 20 of the 59 practices in the urinary tract infection group (34%) participated in the courses that we offered. Thirty-four per cent did not use the patient information (Table 2). The computer-based code to receive the increased fee for telephone consultations was only used by 11% of the fee-for-service practices; although only 25% of respondents indicated that the increased fee was not enough to stimulate increased use of telephone consultations (Table 3).

---

### Table 2 Factors possibly related to variations in change in the main outcomes between the practices

<table>
<thead>
<tr>
<th>Question with responses in numbers and percentages</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Do not know</th>
<th>Mean (median) score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participation in the project</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did you discuss the project in the practice before the start?</td>
<td>23 (21)</td>
<td>13 (12)</td>
<td>50 (45)</td>
<td>15 (14)</td>
<td>9 (8)</td>
<td>1 (1)</td>
<td>2.8 (3)</td>
</tr>
<tr>
<td>Was there agreement that the practice should participate?</td>
<td>57 (52)</td>
<td>25 (23)</td>
<td>18 (16)</td>
<td>5 (5)</td>
<td>4 (4)</td>
<td>1 (1)</td>
<td>1.8 (1)</td>
</tr>
<tr>
<td>Were the GP assistants informed about the project?</td>
<td>63 (56)</td>
<td>22 (20)</td>
<td>20 (18)</td>
<td>4 (4)</td>
<td>2 (2)</td>
<td>1 (1)</td>
<td>1.7 (1)</td>
</tr>
<tr>
<td>Agreement with the guidelines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did you study the article about the guideline thoroughly?</td>
<td>65 (32)</td>
<td>67 (33)</td>
<td>40 (19)</td>
<td>15 (7)</td>
<td>14 (7)</td>
<td>5 (2)</td>
<td>2.2 (2)</td>
</tr>
<tr>
<td>Did you agree with the recommendations in the guideline?</td>
<td>59 (29)</td>
<td>79 (39)</td>
<td>40 (20)</td>
<td>11 (5)</td>
<td>6 (3)</td>
<td>7 (4)</td>
<td>2.0 (2)</td>
</tr>
<tr>
<td>Do you try to follow the new recommendations?</td>
<td>47 (42)</td>
<td>28 (25)</td>
<td>22 (20)</td>
<td>3 (3)</td>
<td>3 (3)</td>
<td>9 (8)</td>
<td>1.9 (2)</td>
</tr>
<tr>
<td>Internal communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you have regular meetings in your practice?</td>
<td>21 (19)</td>
<td>18 (16)</td>
<td>24 (22)</td>
<td>20 (18)</td>
<td>27 (25)</td>
<td>3 (3)</td>
<td>3.1 (3)</td>
</tr>
<tr>
<td>Was the internal communication in the practice good?</td>
<td>12 (10)</td>
<td>67 (56)</td>
<td>22 (18)</td>
<td>15 (13)</td>
<td>1 (1)</td>
<td>3 (3)</td>
<td>2.4 (2)</td>
</tr>
<tr>
<td>Have you met to discuss the guideline in the practice?</td>
<td>21 (20)</td>
<td>12 (11)</td>
<td>20 (19)</td>
<td>9 (9)</td>
<td>43 (41)</td>
<td>0</td>
<td>3.4 (3)</td>
</tr>
<tr>
<td>Use of components of the interventions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did you use the patient information leaflet?</td>
<td>38 (19)</td>
<td>41 (20)</td>
<td>53 (26)</td>
<td>35 (17)</td>
<td>35 (17)</td>
<td>2 (1)</td>
<td>2.9 (3)</td>
</tr>
<tr>
<td>Do you use the increased fee for telephone consultations?</td>
<td>10 (10)</td>
<td>1 (1)</td>
<td>6 (6)</td>
<td>5 (5)</td>
<td>64 (62)</td>
<td>16 (15)</td>
<td>4.6 (5)</td>
</tr>
<tr>
<td>Was the computer-based advice helpful?</td>
<td>26 (13)</td>
<td>41 (20)</td>
<td>57 (28)</td>
<td>34 (17)</td>
<td>32 (16)</td>
<td>15 (7)</td>
<td>2.8 (3)</td>
</tr>
<tr>
<td>Did the practice seem to be engaged in the project?</td>
<td>15 (13)</td>
<td>29 (24)</td>
<td>55 (46)</td>
<td>20 (17)</td>
<td>1 (1)</td>
<td>0</td>
<td>2.7 (3)</td>
</tr>
</tbody>
</table>

---

### Table 3 Agreement with possible explanations of a lack of effect for the interventions

<table>
<thead>
<tr>
<th>Question with responses in numbers and percentages</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Do not know</th>
<th>Mean (median) score</th>
</tr>
</thead>
<tbody>
<tr>
<td>The project did not last long enough</td>
<td>48 (24)</td>
<td>52 (26)</td>
<td>46 (23)</td>
<td>24 (12)</td>
<td>10 (5)</td>
<td>22 (11)</td>
<td>2.8 (3)</td>
</tr>
<tr>
<td>We have started to implement the guideline, but the change is a process that needs more time</td>
<td>28 (14)</td>
<td>38 (19)</td>
<td>47 (23)</td>
<td>56 (28)</td>
<td>21 (10)</td>
<td>11 (6)</td>
<td>3.2 (3)</td>
</tr>
<tr>
<td>We had no reason to change practice</td>
<td>23 (11)</td>
<td>41 (20)</td>
<td>55 (27)</td>
<td>52 (26)</td>
<td>26 (13)</td>
<td>5 (3)</td>
<td>3.2 (3)</td>
</tr>
<tr>
<td>It might have been easier to change if we had been visited by the project team, with information and support</td>
<td>31 (15)</td>
<td>27 (13)</td>
<td>35 (17)</td>
<td>44 (22)</td>
<td>63 (31)</td>
<td>4 (2)</td>
<td>3.5 (4)</td>
</tr>
<tr>
<td>The increased fee for telephone consultation is not enough to stimulate increased use of telephone consultations</td>
<td>45 (22)</td>
<td>37 (18)</td>
<td>40 (20)</td>
<td>22 (11)</td>
<td>28 (14)</td>
<td>29 (14)</td>
<td>3.2 (3)</td>
</tr>
<tr>
<td>It is difficult to increase the use of telephone consultations because of problems with access by telephone</td>
<td>56 (27)</td>
<td>35 (17)</td>
<td>40 (20)</td>
<td>44 (22)</td>
<td>28 (14)</td>
<td>2 (1)</td>
<td>2.8 (3)</td>
</tr>
</tbody>
</table>

---

a Responses to the first questionnaire (n = 112), one mailed questionnaire to each practice which responded by telephone interviews.
b Responses to the second questionnaire (n = 208), one mailed questionnaire for the assistants in each practice and one for every GP.
c Variables scored by project assistant based on records taken during the project (n = 200).
d Practices not working on fee-for-service excluded.
installation, and in 12% it was difficult to use the software (Table 4). The majority did not report important problems with the software, and only 7% thought it was difficult to fill in the questionnaire that popped up during relevant consultations. Because of technical and other reasons, the program was not installed on all the computers in every practice. The mean level of installation was 75% of the computers in each practice. However, the software was usually not installed for all the intervention period, and the questionnaire that popped up was sometimes cancelled. We estimate that the decision support software was available in 2418 of 5031 sore throat consultations (48%), and 703 of 2522 urinary tract infection consultations (28%).

Regression analyses
The explanatory variables in Tables 2 and 4 and reported above in the text were able to explain little of the variation in the main outcomes across practices, and few of the results of these analyses were statistically significant in either the univariate or multivariate analyses. Degree of participation, discussion of the guidelines, agreement with the guidelines, organizational factors and use of the components of the interventions were not significantly associated with changes in practice.

Transcripts of telephone interviews
The transcripts of the responses to the open-ended questions in the telephone interviews supported the qualitative information from the surveys. Many of the respondents said that the project was relevant in daily practice and that they felt it gave inspiration to discuss common problems based on research from general practice. The assistants appreciated an attempt to agree on common routines in the practice. Several of the practices had intentions to start a change process, but had not yet managed to organize themselves. A common comment was that the project could have been more useful if they had spent more time on it, but, as one of the respondents asked: “The question is: ‘how do you do that?’” Some also reflected on how difficult it is to change, even when you have decided to do so: “It is thought provoking how difficult it is to change routines that are so ingrown”. Although many of the respondents were enthusiastic and positive, and the majority felt the project had been useful, others were perhaps more realistic: “If we had made a change it would have been useful, but we have not. There are so many topics that are interesting, but we end up doing things as we have been used to doing them. General practitioners are difficult to change”.

Discussion
The components of the interventions (Box 1) were implemented to various degrees in the participating practices. There is not a single explanation for the variation in change in practice or for the overall lack of change. A combination of organizational problems and lack of time and engagement is the most viable explanation for the lack of effect.

Although much of the data used in this assessment are self-reported, triangulation of methods, using different data sources and multiple respondents from each practice, strengthen the conclusions that can be drawn. Because the project team was responsible for the process evaluation, participants might have tried to please us by giving optimistic or uncritical responses, but variation in the responses fitted our impression of the activity in the participating practices quite well. Many of the respondents openly admitted that their practices had not devoted time to discuss the guidelines or their implementation.

The specific behaviours reflected in the main outcomes for the trial (Box 1) were difficult to change, particularly increasing the use of telephone consultations. Over 60% of the practices that were invited to participate in the trial were interested, and most had positive attitudes about receiving support to implement evidence-based guidelines. Despite a large degree of interest, a large response rate to the first questionnaire (93%) and relatively little disagreement with the guidelines, some of the participants clearly lacked interest in the subject matter or the project.

<table>
<thead>
<tr>
<th>Question</th>
<th>1 No problems</th>
<th>2 Very difficult</th>
<th>3 Very difficult</th>
<th>4 Very difficult</th>
<th>5 Very difficult</th>
<th>Do not know</th>
<th>Mean (median) score</th>
</tr>
</thead>
<tbody>
<tr>
<td>How did you experience the installation of the software?</td>
<td>39 (35)</td>
<td>20 (18)</td>
<td>26 (23)</td>
<td>10 (9)</td>
<td>14 (13)</td>
<td>3 (3)</td>
<td>2.5 (2)</td>
</tr>
<tr>
<td>How do you experience using the software?</td>
<td>44 (39)</td>
<td>32 (29)</td>
<td>20 (18)</td>
<td>9 (8)</td>
<td>4 (4)</td>
<td>3 (3)</td>
<td>2.1 (2)</td>
</tr>
<tr>
<td>How is it to fill out the questionnaires that ‘pop up’?</td>
<td>47 (42)</td>
<td>34 (30)</td>
<td>21 (19)</td>
<td>6 (5)</td>
<td>2 (2)</td>
<td>2 (2)</td>
<td>1.9 (2)</td>
</tr>
<tr>
<td>Did the software work technically ok?</td>
<td>38 (32)</td>
<td>43 (36)</td>
<td>25 (21)</td>
<td>13 (11)</td>
<td>1 (1)</td>
<td>0</td>
<td>2.1 (2)</td>
</tr>
</tbody>
</table>

* Responses to the first questionnaire (n = 112), one mailed questionnaire to each practice who responded by telephone interviews.

* Variables scored by project assistant based on records taken during the project (n = 200).
The project might have felt to some practices like a ‘top-down’ rather than a ‘bottom-up’ approach to developing and implementing guidelines. We did not ask the participants if they felt the project was top-down, but none of the respondents reported this in their answers to the open-ended questions in the surveys and none of the participants commented on this during the project. Nonetheless, for many practices, the project may have been alien to their normal way of making decisions. Many practices lacked routines for discussing guidelines and managing change for common problems, making it difficult to incorporate a project like this into daily practice.

Lack of time and lack of engagement are not necessarily independent factors. In a busy practice, it might be difficult to engage in discussions about implementing clinical guidelines, but on the other hand lack of engagement could lead to low prioritization and a view that there was insufficient time.

In Box 2 we have summarized what we would have done differently, if we were to repeat this project.

In future trials, other approaches to process evaluations are worth considering. Qualitative methods, for instance focus groups with participating GPs and assistants, in-depth interviews or case studies with direct observation in some practices might give information that is not revealed by a questionnaire.

Conclusions

Time, resources and adequate support are necessary to achieve desired changes in practice when implementing guidelines in general practice. Organizational factors, such as inadequate internal communication and routines that make it difficult to use telephone consultations, exacerbate the challenge of achieving changes in practice when change is indicated.

Acknowledgements

We are especially grateful to all the practices that devoted time and energy and generously shared their experiences from the project with us. Colleagues at our institute, partners in the BIOMED project ‘Changing professional practice’ and the advisory committee provided valuable input to the planning of the process evaluation. Doris Tove Kristoffersen (statistician) assisted in the analyses of the data. The Quality Assurance Fund of the Norwegian Medical Association and the Norwegian National Institute of Public Health financed this project.

Box 2  What would we now do differently to implement these guidelines?

• Use more active interventions, for instance outreach visit.
• Focus more on organizational issues to change routines that involve not only physicians.
• Use a longer time period for intervention and follow-up.
• In the future, safe on-line access might make it possible to communicate with the practices, install software and collect data online.

This would perhaps make it easier and potentially more cost-effective to participate in and run quality improvement and guideline implementation projects in primary care.

What is already known on this subject?
• Interventions to support the implementation of guidelines commonly have little or no effect on objectively measured professional practice when evaluated in randomized trials.
• A randomized trial of tailored interventions to support the implementation of guidelines for sore throat and urinary tract infections found little or no change in antibiotic prescriptions, use of laboratory tests or use of telephone consultations.
• Process evaluations using both qualitative and quantitative methods can help to understand why practices did or did not change.

What does this study add?
• No single factor explains why some practices changed and others did not or the overall lack of change.
• Inadequate time, engagement, resources and support are the most likely explanations for the overall lack of change that we observed.
• Organizational factors, particularly problems with internal communication and telephone routines, were important contributing factors in many practices.

References


