INTRODUCTION

Clinical trials are nowadays widely accepted as the paradigm for evaluating the effects of therapeutic, diagnostic, and preventive strategies [1]. In health-care research, however, for methodological and practical reasons, performing a randomized controlled trial is often considered impossible. Alternatives are quasi-experimental and observational designs [2-4].

Most trials are performed under conditions created especially for that trial and, in general, adaptations in health care are required. However, when a health-care activity itself is studied, the design is often adapted to fit within health care. This is especially the case in evaluating the effectiveness of procedures already applied in routine care. Special experimental conditions may be inappropriate in such cases. It can be preferable to perform a study in a routine health-care context rather than in an optimized artificial situation that does not reflect actual care. Furthermore, studying an activity or procedure that is already implemented in routine care can offer methodological opportunities. One of these opportunities is the possibility of refraining from informing the participants about the specific trial conditions, if such information would prevent a valid evaluation. Considering this approach is relevant if interventions can not be blinded and if study subjects can be influenced by it if they know
that their behavior is observed. This phenomenon is named after the Hawthorne plant of the Western Electric Company, where in the 1920s, a study was conducted to increase productivity by improving illumination. More light was accompanied by increased productivity. However, productivity also increased with less light. Later it was found that effects were not caused by better illumination but by the perception of employees that extra attention was given to working conditions [5].

We present an example of how the effects of activities applied in routine health care were studied in a secretly performed trial. In this trial we studied the effects of routine feedback on test ordering by family physicians. The aim of the trial was to confirm a causal relation between feedback and a change in test ordering behavior.

When feedback is given, a Hawthorne effect is likely to occur [6]. The fact that doctors know that performance is reviewed may change behavior, apart from the specific changes caused by the intervention itself. In assessing the real effects of feedback in a trial, a Hawthorne effect should be prevented or controlled for. When this effect is unequally present in the comparison groups, bias of the trial result can occur.

In practice, a Hawthorne effect is not necessarily undesirable. Regarding feedback, it may cause a general alertness that makes the physician request fewer tests. However, although reducing the number of inappropriate test requests is a goal of feedback, the reduction through a Hawthorne effect is probably unspecific and therefore appropriateness of test ordering behavior may not improve accordingly. Refraining from test requests should be done selectively in specific situations where requests would be inappropriate. Otherwise, underutilization of tests cannot be ruled out.

Furthermore, in any intervention based on transfer of knowledge, there is a risk of information leakage from the intervention group to the control group. Such leakage leads to contamination of the trial arms, which is almost impossible to take into account. Information leakage can be prevented or minimized by strict separation.

We deviated from accepted rules for conducting a trial by performing the trial in secret, without informing study subjects (physicians) about the trial. The pros and cons, requirements and limitations of this approach are discussed in this paper.

**METHODOLOGY**

**Background and Objectives**

The Diagnostic Coordinating Center Maastricht (DCC) coordinates all diagnostic test requests of the family physicians (±85) in the Maastricht region. To realize a more appropriate test use, the DCC gives personal feedback on test ordering behavior to every individual family physician affiliated to the center. The feedback is given twice a year and focuses on both the number and the rationality of test requests. Request forms submitted by each family physician in the course of one recent month are discussed. Discussing rationality of requests is feasible since the request forms contain routine clinical data (history, physical examination, possible diagnosis, reason for request, etc.). Feedback is given as written comments by a respected expert peer. This feedback procedure started in 1985. Before the feedback was initiated in 1985, there was a continuous increase in the number of requests per year. Soon after the onset of feedback request numbers decreased sharply, as demonstrated in an observational study [7]. It was not certain that this represented a cause–effect relation. Therefore, before advising implementation of feedback elsewhere, a causal relation had to be demonstrated, and a randomized trial was developed to that end. However, at the start of the trial in 1989 the feedback had become a routine activity, integrated in daily health care [8].

**The Trial Design**

We assessed the effects of routine individual feedback on the rationality and volume of test requests by family physicians. For the purpose of the trial we provided feedback on tests not discussed before. Since feedback started several years earlier, specific effects on tests previously discussed could already be achieved before the trial was started. While before the experiment feedback was given predominantly on hematological, serological, and clinical chemistry tests, after four years several tests remained that were not, or only on rare occasions, discussed in feedback. Thus, the feedback evaluated in the trial focused on these tests (electrocardiography, endoscopy, cervical smears, allergy tests, and a variety of radiologic or ultrasonographic tests).

Family physicians were assigned at random to two groups (1 and 2). Using an incomplete balanced block design, both groups of physicians received feedback on one of two clusters of tests (A or B), while serving as a control for the other test cluster on which they did not receive feedback (Table 1). The individual tests discussed in the "experimental" feedback during the trial were not assigned at random since several tests were closely related to each other.

Two factors, a Hawthorne effect and contamination of the trial arms through information leakage could reduce the accuracy of our trial results.

**Hawthorne Effect**

The Hawthorne effect, described in the introduction, was dealt with as follows: First, we performed the trial without obtaining informed consent. The experimental (new) feedback was appended to the routine feedback as provided since 1985. For the family physicians, the trial was therefore nothing but a small extension to the usual feedback, now also focusing on tests not discussed before. Such extensions were not unusual. The amount of feedback to each group of
TABLE 1. Trial design: Tests discussed in the trial in relation to the compared groups of family physicians (FP)

<table>
<thead>
<tr>
<th>Tests</th>
<th>FP group 1</th>
<th>FP group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test-group A</td>
<td>Intervention</td>
<td>Control</td>
</tr>
<tr>
<td>Electrocardiography</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endoscopy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervical smears</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allergy tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test-group B</td>
<td>Control</td>
<td>Intervention</td>
</tr>
<tr>
<td>X-rays:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervical spine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thoracic spine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lumbar spine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvis/hip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knee joint(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ankle joint(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sinuses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ultrasound:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver/biliary tract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidneys</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

family physicians was similar, since each physician received experimental feedback on only one of both clusters of new tests.

Second, the incomplete balanced block design would minimize bias from a Hawthorne effect: a Hawthorne effect would occur in both groups of physicians since both groups received feedback during the trial. Therefore, a Hawthorne effect could not be responsible for differences between both groups [9].

**Information Leakage**

It is possible that two doctors, each assigned to another group, collaborate in daily practice. Consequently, a family physician belonging to a control group for a test cluster could discuss these tests or the feedback with a colleague belonging to the intervention group involved. This would imply a risk of information leakage, possibly resulting in an underestimation of differences between the intervention and control group, due to contamination of the trial arms.

Such a contamination due to information leakage was minimized by two closely related features of our design. Basically, information leakage was avoided by the fact that family physicians were unaware of the trial. This was achieved by adding feedback on new topics (tests not discussed before) to the existing routine feedback. The usual feedback was only extended and not substantially changed. Closely related to this was the fact that informed consent was not obtained for this study. After informed consent, the trial would no longer be secret and contamination of the trial arms was very likely to occur, thereby preventing an accurate evaluation of results. Altogether, in some routine situations (like the described one) one might prefer not to obtain informed consent, so offering the best possibility to assess the effects of an intervention.

**EVALUATION**

The above design was followed by the DCC during 2.5 years, from October 1989 until May 1992. After randomization, two groups were composed that appeared to have similar characteristics (Table 2). In addition, it was indeed feasible to add new information to the existing feedback without interfering with the routine procedure (feedback given twice a year since 1985).

The feedback caused effects on the rationality of test requests (agreement with guidelines) and on test numbers. Overall, non-rational test requests were seen less frequently ($p = 0.009$). Rationality mainly improved for test cluster B, with fewer non-rational requests seen in the group receiving feedback on this test cluster ($p = 0.04$). At test level rationality improved, especially for lumbar spine x-rays [10].

Overall, test numbers decreased by the feedback ($p = 0.036$). For test cluster A these decreased for the intervention group (-7%), while they increased for the control group (+13%) ($p = 0.04$). Cervical smears were predominantly responsible for this difference.

Although effects of the feedback could be demonstrated, these were smaller than expected. It is possible that earlier feedback also influenced test ordering behavior for tests not discussed, probably due to a general learning effect [7]. Therefore, part of the improvement already occurred before the trial. At the end of the trial, due to the positive results, we continued the feedback introduced within the scope of the trial.

We had no sign or question whatever from the family physicians indicating that the trial was surmised. At the end
of the study, all family physicians were informed about the study and its results. Even then, after stopping the trial, no such sign reached us in retrospect. The relations between the DCC and the family physicians are still good and based on mutual confidence.

DISCUSSION

According to the Helsinki declaration, investigators “should obtain the subject’s freely given informed consent” from each patient unless “the doctor considers it essential not to obtain informed consent; the reasons for doing so should be stated in the study protocol.” There are situations in which obtaining full informed consent would lead to a methodologically invalid and therefore unacceptable design. For our study, the combined effects of a Hawthorne phenomenon and information leakage, while evaluating an unblindable procedure, would represent such a situation. Full informed consent was not obtained on the following grounds: First, the feedback was started in 1985 at the insistence of our physicians. Since feedback is not possible without insight into test requests, implicitly they gave general consent to monitoring test requests and test ordering behavior. Second, the feedback procedure was the study object, not the tests discussed. The procedure remained virtually unchanged, only the contents of the feedback were extended for the trial. Third, at least usual care was provided; no normally available procedure was withheld in the trial setting. Finally, it is clear that the studied intervention is not harmful.

By not obtaining informed consent both a Hawthorne effect and contamination of trial arms due to information leakage can be avoided. In health-care research, a Hawthorne effect is an important threat to an accurate evaluation of the effect of an intervention on physicians’ behavior. It can lead to behavioral changes which could erroneously be attributed to that intervention [6]. This risk is virtually absent when the physicians enrolled are unaware of the trial, as in our study. In addition, an incomplete balanced block design is a way to minimize bias resulting from a Hawthorne effect, since in such a design physicians in both groups are equally influenced.

A second threat to a health-care trial, potentially responsible for bias, is contamination of the trial arms due to information leakage stemming from communication between doctors in the intervention group and the corresponding controls. This affects internal validity and differences between both groups may be underestimated. Such an information leakage can be minimized by performing the trial without informed consent, especially when it is not customary for family physicians to discuss the experienced interventions with colleagues.

Some might find it unethical not to obtain informed consent. However, if this is the only way to evaluate an activity already in use, it is more unethical not to perform the study. In that case, effectiveness cannot be estimated and ineffective routine activities are continued. Clearly, also an invalid and misleading study would be highly unethical.

Monitoring behavior is necessary for feedback. If no informed consent at all is obtained, one might fear a deterioration of the relations between physicians and those who monitor behavior. Therefore, a general consent should be obtained in advance. By asking collaborating doctors to agree that their behavior is monitored and could be evaluated without study-specific informed consent, the need for health-care evaluation and ethical principles can be brought into harmony. Of course, it should be warranted that the monitoring data are confidentially dealt with.

We conclude that a trial under routine health-care conditions may reflect the actual daily situation better than a trial under artificially optimized conditions. The fact that the procedure under study is desired by physicians and used in routine care, enables a trial without study-specific informed consent provided that a general consent is obtained at the start of routine procedures. This consent could include possible research (under specific conditions) in the future. If the health-care procedure under study is in routine use but still needs to be evaluated since it may have great impact on (the quality of) care, if it is not harmful, if full informed consent would prevent proper evaluation due to contamination of the trial arms or a Hawthorne effect and if at least usual care is offered to the intervention group and the controls, one should consider the possibility of not obtaining full informed consent. In such cases, a general informed consent should suffice. An independent ethical committee should evaluate study proposals as to whether these criteria are met.

For the development of health-care research, an open debate on the possibilities and conditions for not obtaining informed consent is important. Especially the medical profession, whose performance has great impact on the quality of care, should not be reluctant to allow a proper evaluation of its achievements.

References

7. Winkens RAG, Pop P, Grol RPTM, Kester ADM, Knottnerus

