A multi-faceted strategy to improve the use of national fertility guidelines; a cluster-randomized controlled trial

Selma M. Mourad1,2, Rosella P.M.G. Hermens2, Janine Liefers2, Reinier P. Akkermans2, G.A. Zielhuis3, E. Adang3, Richard P.T.M. Grol2, Willianne L.D.M. Nelen1, and Jan A.M. Kremer1,*

1Department of Obstetrics and Gynecology, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands 2Scientific Institute for Quality of Healthcare (IQ Healthcare), Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands 3Department of Epidemiology, Biostatistics and Health Technology Assessment, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands

Submitted on February 2, 2010; resubmitted on August 9, 2010; accepted on September 30, 2010

BACKGROUND: Proper use of clinical practice guidelines can decrease variation in care between settings. However, actual use of fertility guidelines is suboptimal and in need of improvement. Hence, a cluster-randomized controlled trial was designed to study the effects of two strategies to implement national Dutch guidelines on comprehensive fertility care.

METHODS: Sixteen fertility clinics participated in the trial. A minimal, professional-oriented implementation strategy of audit and feedback was tested versus a maximal multi-faceted strategy that was both professional and patient oriented. The extent of adherence to guideline recommendations, reflected in quality indicator scores, was the primary outcome measure. To gain an insight into unwanted side effects, patient anxiety and depression scores were gathered as secondary outcomes. Data collection encompassed medical record search, patient and professional questionnaires.

RESULTS: A total of 1499 couples were included at baseline and 1396 at the after-measurement. No overall significant improvement in indicator scores was found for either strategy [odds ratios ranging from 0.23 (95% confidence interval (CI): 0.06–0.95) to 6.66 (95% CI: 0.33–132.8)]. Secondary outcomes did not differ significantly for both groups, although selected anxiety scores appeared lower in the maximal intervention group. Process evaluation of the trial revealed positive patient experiences with the intervention material [e.g. an increased understanding of their doctor’s treatment policy (61%), an increased ability to ask questions about the treatment (61%)]. Professionals’ appreciation of intervention elements varied, and execution of the multi-faceted strategy appeared incomplete.

DISCUSSION: Absence of an intervention effect may be due to the nature of the strategies, incomplete execution or flaws in study design. Process evaluation data raise the question of whether professionals should be the only stakeholder responsible for guideline implementation. This study therefore contributes to an increased understanding of fertility guideline implementation in general, and the role of patients in particular.

Clinical Trials.gov: NCT00119925.

Key words: quality indicators / fertility / guidelines / quality of care

Introduction

The burden of infertility weighs heavily on ~80 million couples worldwide (Vayena et al., 2002; Nachtigall, 2006). For western Europe, this affects 10–15% of couples of reproductive age (Evers, 2002; Boivin et al., 2007; Oakley et al., 2008; Bhattacharya et al., 2009), of which an estimated 50% seek medical assistance (Boivin et al., 2007). Providing those infertile couples with best available healthcare, i.e. consensus- and evidence-based diagnostic tests and treatment options, is the common aim of several national and international societies of fertility professionals. For this purpose, they have developed clinical practice guidelines to guide healthcare providers during daily practice, while at the same time decreasing variation in care between settings. However, the mere existence of such guidelines does not automatically imply that they are widespread or commonly used. Dissemination of new guidelines should therefore ideally be
followed by robust implementation efforts. However, research on the implementation of fertility guidelines, in particular, is scarce. This is remarkable, as the diagnostics and treatment commonly used for infertility lead to an extensive use of healthcare resources and are associated with substantial physical and psychological consequences for the patients involved (Greil, 1997; Cousineau and Domar, 2007; Schieve et al., 2007).

It is commonly known in guideline implementation research that there is no ‘magic bullet’ for successful implementation of every clinical problem or in every practice setting. For example, the literature is still inconclusive regarding the effects of multi-faceted versus single interventions for guideline implementation (Grol and Grimshaw, 2003; Grimshaw et al., 2004, 2006). The most frequently studied interventions encompass audit and feedback on current practice, dissemination of educational materials, reminders and the organization of educational meetings or outreach visits, which all seem to have only small to moderate effects on the improvement of professional performance (Jamtvedt et al., 2006; O’Brien et al., 2007; Farmer et al., 2008). However, mainly individual professionals are targeted in these interventions. The role of patients in guideline implementation is, surprisingly, still rather unexplored and evidence is therefore scarce (Griffin et al., 2004; Grol et al., 2005; Wetzels et al., 2007). As fertility patients are generally young, critical of their care providers and thus sometimes regarded by professionals as one of many barriers to guideline implementation (Haagen et al., 2005), we hypothesize patients could be just the stakeholders we need for successful implementation of fertility guidelines.

Hence, we designed a cluster-randomized controlled trial (c-RCT) to study the effects of two different strategies to improve the use of recommendations in national Dutch fertility guidelines on comprehensive clinical fertility care. The literature was reviewed to identify barriers for the implementation of clinical practice guidelines in reproductive medicine, and fertility care in particular (Cabana et al., 1999; Cardenas et al., 2002; Farquhar et al., 2002; Foy et al., 2002, 2005; Graham et al., 2004a; Haagen et al., 2005; Say and Foy, 2005; Chan et al., 2008), as a prospective identification of barriers is still assumed to lead to better adaptation of interventions (Shaw et al., 2005). Based on this literature, barriers were expected to be found in the contexts of the guideline itself, organizational aspects of care, patient characteristics and professional characteristics. Our strategies were subsequently tailored to these barriers. We hypothesized that a maximal implementation strategy consisting of a multi-faceted intervention, tailored to barriers from the literature and directed at both professionals and patient couples, would prove to be more effective than a minimal implementation strategy consisting of a single, professional-oriented intervention of audit and feedback. We assume these effects can be seen at the cluster level. To increase our understanding of factors influencing the impact of the implementation strategies, a process evaluation of the trial was also performed.

**Materials and Methods**

**Study design**

We performed a c-RCT in 16 fertility clinics from the Fertility Network East, a network of fertility clinics in the eastern region of The Netherlands exchanging knowledge and clinical experiences. The study was approved for all clinics by ‘the Regional Review Board for Research with Human Subjects (CMO) Region Amhem-Nijmegen’ (CMO no. 2004/193), as well as two local research ethics committees. The trial was registered with ClinicalTrials.gov (ID NCT00119925) and results are reported according to the CONSORT statement for cluster-randomized trials (Campbell et al., 2004).

**Randomization**

Fertility clinics instead of couples were the unit of randomization to avoid crossover contamination of both intervention strategies; clinics were assigned to either a minimal or a maximal implementation strategy. Prior to randomization to one of the implementation strategies, participating clinics were stratified according to clinic size (small, medium and large) and treatment facilities (with or without IVF/ICSI facilities; Fig. 1). Two independent research associates performed the randomization procedure by drawing blinded envelopes. Allocation sequence was concealed until the interventions were assigned. Usual care measured at baseline served as a control for both groups.

**Blinding**

All patients and professionals were blinded to group assignment and remained unaware of intervention contents. The research group, but not the trained data extraction personnel, was aware of group assignment.

**Participants**

**Clinics**

In total, 16 clinics of the Fertility East Network were invited and agreed to participate. There was one academic and one tertiary care clinic, both running a licensed IVF laboratory, of which there are only 13 in total in The Netherlands. Seven of the 16 clinics offered secondary care, as an autonomous satellite (i.e., performing an IVF/ICSI cycle up to the ovarian stimulation phase) or transport IVF/ICSI clinic satellite (i.e., performing an IVF/ICSI cycle up to the oocyte pick-up phase). These clinics are also teaching clinics and of large or intermediate size; the other seven clinics are smaller, non-teaching facilities. In total, 15 clinics are in the national health system, and one of the smaller secondary care clinics is a private clinic. The different types of clinics were chosen to reflect the average national fertility care.

**Patients**

To include a representative patient group, potential participating couples were retrospectively selected in each clinic by means of the clinics’ financial registration database (Diagnosis Treatment Combination code). In this nationwide registration system, patients undergoing diagnostics or treatment for infertility are identified with a specific Fertility-code (F-code). A baseline measurement was performed which included a random sample of 1499 couples that had an active F-code anytime between the 1st of April 2005 and July 1st 2005 and concerned the fertility care these couples received in the period between 1 January and 1 July 2005. For the after-measurement, couples were eligible if they had an active F-code anytime between 1 December 2007 and 1 March 2008. The after-measurement focused on the care these couples received in the period between 1 September 2007 and 1 March 2008. From each clinic, a random sample of eligible couples (n = 50, 150 or 500, stratified according to clinic size) was invited to participate in data collection for the study (postal survey). Patients in each study arm were included according to the ‘intention to treat’ principle, i.e. independent from diagnosis or treatment type. This procedure was followed for both the baseline and after-measurement. Eligible couples were sent a questionnaire which
included an informed consent form concerning the use of medical data from both patient records and the questionnaire.

**Study interventions**

**Minimal strategy: audit and feedback**

The minimal implementation strategy was entirely directed at professionals, and in June 2007 introduced in each of the eight clinics in the minimal intervention arm. This strategy consisted of professional audit and feedback. Audit encompassed results of the baseline assessment of the clinics’ scores to previously developed quality indicators (Mourad *et al.*, 2007), regarding care provided in the period January–July 2005. Based on these indicator scores, clinic-specific feedback reports regarding current care were formulated by the study group. Each of these minimal strategy clinics was sent a sufficient number of these feedback reports for dissemination among its fertility professionals (i.e. gynaecologists, fertility doctors, nurses and physician assistants). An instruction letter for interpretation was enclosed with each report. Per quality indicator, feedback on current care was given by means of a bar chart showing the total range of performance on a scale from 0 to 100% including the median adherence of all participating clinics. Feedback was kept anonymously and each individual clinic received a feedback report with a clear marking of their own performance in the bar chart. After sending these feedback reports, no further contact was established with the minimal intervention clinics until the data collection for the after-measurement was started.

![Flow chart of the clinics and the included patients.](image-url)
Maximal strategy: multi-faceted intervention

The maximal implementation strategy was multi-faceted, and tailored to barriers known from the literature as well as to baseline performance; the strategy was directed at both professionals and patient couples.

The strategy included the following five professional-oriented elements:

(i) Audit and feedback discussions

Similar to the minimal strategy, clinic-specific feedback reports on current care were developed and sent to the eight maximal intervention clinics in May 2007. After ~3 weeks, a multi-disciplinary (e.g. addressing gynaecologists, fertility doctors, residents, fertility nurses and quality officers) 1 h meeting was organized in each clinic, where the feedback report was presented and commented on by one of the authors (S.M.M.). During each meeting, the clinic’s performance was discussed in the light of the other 15 clinics’ performances, and possibilities for quality improvement were highlighted.

Moreover, professionals in each of the maximal strategy clinics were provided with the following implementation tools:

(ii) List of suggested tools for implementation of the guidelines at local level

Per indicator, a specific tool or action was provided to improve performance. For example, the proposed action to improve ‘history-taking of the couple’ consisted of ‘the development and introduction in the outpatient record of a systematic history-taking form’.

Besides these professional-oriented interventions, a patient-oriented intervention was introduced in these maximal strategy clinics:

(iii) Leaflet on shared decision making (SDM) in the fertility consultation

The leaflet on SDM explained the general principles of SDM and contained a suggested literature list. It also provided practice examples for fertility care in which patient preferences do not always match guideline contents, and how SDM could be helpful to reach bilateral agreement in policy (e.g. concerning single versus double embryo transfer or the need for an expectant period before start of assisted reproduction).

(iv) Patient information checklists

Checklists for the provision of patient information were disseminated for professionals to use in the consultation room. These checklists consisted of laminated sheets that contained patient information items to be discussed with patients and recommended in the guidelines (e.g. the chances of pregnancy after transfer of one versus two embryos, the side effects of medication for ovulation induction).

(v) Educational patient leaflets

Based on the fertility guidelines and accompanying quality indicators, three educational patient leaflets were developed concerning: (a) the initial assessment of fertility, male infertility, endometriosis, tubal pathology and idiopathic infertility; (b) ovulation induction, intrauterine insemination and ovarian hyperstimulation syndrome; and (c) IVF/ICSI treatment and ovarian hyperstimulation syndrome. These leaflets explained the background and contents of the professional guidelines in lay language. The text moreover aimed to encourage patients to start a dialogue with their doctor regarding diagnostic testing, treatment plan and treatment itself, while promoting the concept of SDM. The leaflets contained a prompt sheet for questions during consultation and also referred to the national fertility guidelines published on the website of the Dutch Society for Obstetrics and Gynaecology. Professionals were asked to distribute the relevant leaflets among the infertility patients consecutively visiting their (outpatient) clinic in the implementation period (June 2007–January 2008).

Outcomes

The effectiveness of both implementation strategies was reflected in the scores on a pre-developed set of guideline-based quality indicators (Mourad et al., 2007). The studied guidelines included nine fertility guidelines of the Dutch Society for Obstetrics and Gynaecology (NVvOG): initial assessment of fertility, anovulation, male infertility, tubal pathology, endometriosis, premature ovarian failure, IUl, indications for IVF treatment and ovarian hyperstimulation syndrome. Moreover, the model protocol of the Dutch embryo Act was considered, which includes clinical statements on the provision of IVF treatment. The indicators were systematically developed structure and process indicators for comprehensive clinical fertility care. They cover topics like ‘indications for treatment’, ‘diagnostic procedures’, ‘treatment procedures’ and ‘patient information’. Indicators are scored dichotomously, in which the value ‘1’ reflects ‘adherence’ and the value ‘0’ reflects ‘non-adherence’ to a defined guideline recommendation. All indicators were tested during the baseline measurement for several quality criteria (i.e. measurability, reliability, applicability, improvement potential, discriminatory capacity, complexity and case-mix stability), thus exploring their value as instruments for monitoring and improving clinical performance. After the baseline measurement, indicators with high applicability were considered adequate primary outcome measures to reflect the degree of adherence to guideline recommendations for the c-RCT (Mourad et al., 2008). Indicators which also have high improvement potential are suitable to assess changes in adherence, whereas indicators with low improvement potential are only suitable to monitor adherence in time. To gain an insight into any unwanted side effects of the applied strategies, we included at patient level ‘anxiety’ and ‘depression’ as secondary outcome measures.

Data collection

Data collection to calculate quality indicator scores was performed from either medical records (e.g. process indicators concerning treatment aspects), a professional questionnaire (e.g. structure indicators concerning clinic’s structure) or a patient questionnaire (e.g. process indicators concerning patient information provision) (Mourad et al., 2009). Medical record extraction was performed by three trained data collectors who entered data in digital forms, specifically designed to enhance systematic and complete data collection by using computerized algorithms for data entry. Data collectors followed an intensive 1-month training and reliably assessed a series of 30 test records before starting official data collection. During data collection, two independent reviewers abstracted a random sample of 10% (n = 32) of medical records from two participating clinics. The extent of agreement between these data reviewers on the level of process indicator scores, corrected for chance, was calculated using Cohen’s kappa coefficient (Landis and Koch, 1977). Reliability among the data collectors was substantial, reflected in average Cohen’s kappa coefficients of 0.86 (baseline) and 0.82 (after-measurement) (range 0.48–1.0).

Data for the secondary outcome measures were gathered by the patient questionnaire. Anxiety was measured by a 10-item short version of the State Trait Anxiety Index (STAI) (Spielberger et al., 1970; Van Der Ploeg et al., 1980) and 12 additional infertility-related anxiety items, e.g. ‘anxiety for treatment outcome, both on a 4-point scale (‘almost never’, ‘sometimes’, ‘frequently’ and ‘almost always’). Depression was measured on a 4-point scale by the Beck Depression Index for Primary Care (BDI-PC) (Beck et al., 1997).

Process evaluation of intervention elements

We applied the process-evaluation framework described by Hulscher et al. (2003) to evaluate both ‘exposure to’ and each stakeholder’s ‘experience
with the several elements of the interventions. This was done by means of a professional questionnaire and an addendum to the patient questionnaire in the after-measurement.

Sample size

The study was designed to provide at least 80% power in order to detect a difference of 15% in guideline adherence between the two study arms (70% for the minimal strategy and 85% for the maximal strategy) at the 0.05 two-sided significance level. In a non-clustered design this would mean an average of 241 couples. However, the design effect of using a clustered design had to be taken into account. The median intracluster correlation coefficient (ICC) for all indicators was 0.12 (range 0.00–0.51), but to assure inclusion of enough patients, we chose to use an ICC of 0.15 for the further sample size calculation. The design effect was calculated at 5.35 using the formula of Thorsen and Makela (1999) [design effect = 1 + (n − 1) ICC]. Considering an ICC of 0.15 and assuming an average 30 infertile couples seen per professional and an average 2–3 professionals per clinic (43 professionals in total), 1290 couples were needed for both the baseline measurement and the after-measurement.

Analysis

To analyse the effectiveness of the two implementation strategies, we assessed the proportion of patients that were treated in accordance with the guidelines, which is reflected in quality indicators scores before and after the implementation period in both intervention arms. The difference in indicator scores was analysed with adjustment for clustering of patients within clinics. Therefore, for each indicator, multi-level logistic regression analyses were performed in which ‘intervention arm’ acted as the independent variable and ‘indicator score’ as the dependent variable. Analyses were based on the modified intent to treat principle, meaning all participants were included in the arm to which they were originally assigned, regardless of whether they completed the intervention given to the arm. Differences at baseline were corrected for by taking baseline scores as a covariate in the final multi-level model composed to assess differences between the two intervention arms. The statistical program SAS for Windows V8.2 was used to compose a multi-level model correcting for clustering in the different clinics.

Results

Participant flow

At baseline, 726 couples were included for the maximal implementation strategy and 773 for the minimal implementation strategy; for the after-measurement these numbers were 697 and 696, respectively (Fig. 1).

Comparability of data at baseline

Demographic characteristics of both intervention arms at both measurement periods are shown in Table I. For most characteristics, the arms were comparable. There were significant differences between the minimal and maximal intervention arms at baseline, for the variables ‘type of infertility’ (68 versus 77% primary infertility; \( P = 0.000 \)) and ‘duration of infertility’ (35 versus 41 months; \( P = 0.000 \)).

Outcomes and estimation

Primary outcomes

Table II shows the indicator scores of both the baseline and the after-measurement for the guideline ‘initial assessment of fertility’. For the other guidelines, these scores can be found in supplementary data, Table SII. The indicators with low improvement potential at baseline showed in the after-measurement the same high adherence percentages, meaning care remained of high standard. This is, for example, true for the indicator ‘Couple’s history-taking should cover at least: age of both partners, duration of infertility, type of couple’s infertility’.

The remaining indicators with high improvement potential at baseline, showed mixed effects in the after-measurement (Table II and supplementary data, Table SII).

Only for two indicators, concerning ‘the aim of the initial assessment of fertility to result in a diagnosis and prognosis’ and ‘lifestyle advice concerning smoking’, was there a significant surplus value of the maximal intervention compared with the minimal intervention arm. For one indicator, concerning the monitoring of ovarian response by transvaginal ultrasound in the case of IUI in the stimulated cycle, was there a significant ‘lower decrease’ in adherence to the guideline when comparing minimal and maximal intervention arms.

Secondary outcomes

Both at baseline and in the after-measurement, there were no significant differences in BDI-PC and STAI scores for the minimal and maximal intervention arms (data not shown). Regarding the infertility-related anxiety items at baseline, women in the maximal arm were more fearful of definite childlessness (\( P = 0.033 \)). This difference did not exist in the after-measurement. At after-measurement, women in the minimal arm had significantly higher anxiety scores regarding the effect of stress on the physical relationship with their partner (\( P = 0.013 \)) as well as higher anxiety scores for the presence of a twin pregnancy (\( P = 0.006 \)).

Process evaluation of study interventions

Patients

Of the 696 patients included in the maximal intervention arm in the after-measurement, 260 (37%) unique patients reported having received one or more of the patient leaflets. The majority of those patients appeared highly satisfied with the leaflets (97%). They reported an increased knowledge of potential causes (71%), treatment procedures (90%) and guidelines (51%), an increased understanding of their doctor’s treatment policy (61%), and an increased ability to ask questions about the treatment (61%). The scores for an improved communication with the doctor (22%), as well as perceived increased empowerment for decision making during consultations (22%) were lower. In total, 83% would want to receive comparable leaflets in the future and 97% would recommend the leaflets to peers.

Professionals

Professionals from both intervention arms rated the feedback report as easily accessible and comprehensible, and reported that the report actually contributed to the implementation of the guidelines in their clinics. In the maximal intervention arm, the feedback meeting was rated of equal value. In the maximal intervention clinics, the ‘List of suggested tools for implementation of the guidelines
at local level’ was moreover highly appreciated, whereas professionals were indifferent to the leaflet on SDM and the patient information checklists. One remarkable reason that came up frequently for not using the offered intervention material for fertility guideline implementation was that it was considered to be ‘not my job responsibility’ to initiate practice changes.

**Table 1** Demographic characteristics of the participating couples.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Baseline measurement (n = 1499)</th>
<th>After-measurement (n = 1393)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age in years (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>32.48 (SD 4.1)</td>
<td>33.06 (SD 4.5)</td>
</tr>
<tr>
<td>Male</td>
<td>35.11 (SD 4.9)</td>
<td>35.53 (SD 5.1)</td>
</tr>
<tr>
<td>Ethnic background (%)¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>98.4</td>
<td>98.5</td>
</tr>
<tr>
<td>Non-Dutch</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td>Gross monthly family income in € (%)²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1100</td>
<td>1.3</td>
<td>0.8</td>
</tr>
<tr>
<td>1100–1760</td>
<td>4.4</td>
<td>2.3</td>
</tr>
<tr>
<td>1760–2750</td>
<td>20.9</td>
<td>14.8</td>
</tr>
<tr>
<td>&gt; 2750</td>
<td>73.3</td>
<td>82.1</td>
</tr>
<tr>
<td>Education level per couple (%)³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>7.1</td>
<td>3.1</td>
</tr>
<tr>
<td>Intermediate</td>
<td>41.1</td>
<td>36.6</td>
</tr>
<tr>
<td>High</td>
<td>51.9</td>
<td>60.3</td>
</tr>
<tr>
<td>Type of infertility (%)⁴</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>68.4</td>
<td>67.8</td>
</tr>
<tr>
<td>Secondary</td>
<td>31.6</td>
<td>32.2</td>
</tr>
<tr>
<td>Mean duration of infertility in months (SD)⁵</td>
<td>34.7 (SD 20.4)</td>
<td>35.8 (SD 23.7)</td>
</tr>
</tbody>
</table>

¹Ethnic background of the couples was determined by the origin of both partners: Dutch = one or both partners are of Dutch origin; non-Dutch = both partners are not of Dutch origin.

²Gross monthly family income was categorized according to social security standards in 2005 and modal income in Euros: < 1100 = less than Dutch modal income; 1100–1760 = Dutch modal income; 1760–2750 = up to 1.5 times Dutch modal income; > 2750 = more than twice Dutch modal income.

³Education level of the couples was determined by the highest education level of both partners: low = primary or lower vocational education; intermediate = secondary or intermediate vocational education; high = higher professional education or university.

⁴Type of infertility was determined for the couple.

⁵Duration of infertility was defined as the period between the start of regular unprotected sexual intercourse and the beginning of the study period: respectively 1 January 2005 (for the baseline measurement) or 1 September 2007 (for the after-measurement).

*Significant difference at the P = 0.05 level.

**Discussion**

To our knowledge, we report the first trial to evaluate the implementation of a set of fertility guidelines, and one of few trials in general that seeks to influence professional behaviour by a patient-oriented intervention. However, both the minimal professional-directed strategy and the maximal tailored, multi-faceted patient- and professional-oriented strategy did not improve the ‘overall’ use of recommendations in our national guidelines on comprehensive fertility care. Wide variation between clinics might be partly due to differences in local protocols. As some local protocols contain guideline recommendations, professionals are more prompted to adhere, whereas professionals in other clinics might not specifically value these recommendations, explaining the wide differences in adherence. This underlines the need for a thorough process evaluation to detect the rationale behind such differences in variation.

For only two indicators were significant effects seen in favour of the maximal intervention, compared with the minimal intervention. Not having found a sustainable effect of either strategy for the entire guideline programme raises the question of whether these specific strategies are ineffective in themselves, or whether ineffectiveness is largely caused by flaws in study design or incomplete execution of the intervention strategies. We will further elaborate on these possibilities.

**Rationale of the tested interventions**

From the reviews of the Cochrane Effective Practice and Organisation of Care Group, we learned that the impact of different interventions to promote healthcare interventions is widely variable. Both ‘audit and feedback’ and ‘educational outreach visits’ are reported to have a small to modest effect on improving professional practice, and the
effect of ‘printed educational materials’ seems to be only beneficial when compared with ‘no intervention’ (Farmer et al., 2008). In the field of reproductive healthcare, a recent review of evidence-based strategies for implementing guidelines in obstetrics (Chaillet et al., 2006) showed especially positive effects for interventions including audit and feedback, reminders and multi-faceted strategies. It also clearly demonstrated that the prospective identification of barriers to change leads to better adaptation of interventions. Recent studies on barriers to implementation in reproductive medicine identified mainly organizational barriers (Cardenas et al., 2002; Farquhar et al., 2002; Graham et al., 2004b; Haagen et al., 2005; Say and Foy, 2005). As the professionals involved in obstetric, gynaecological and fertility care are largely similar, and evidence for implementation within fertility care is very scarce, an extrapolation of these results may be assumed to be reasonable. However, we have to keep in mind that the major pitfall in implementing change to practice lies in the fact that previously successful interventions may prove worthless in other settings. This assumption to extrapolate and rely on previously performed barrier research and interventions might be one explanation for why the tested strategies in our c-RCT did not prove to be successful.

**Study design**

We chose to perform a c-RCT to evaluate the effect of the tested interventions, as this is considered the ‘gold standard’ in implementation research (Ovretveit and Gustafson, 2002). Randomization was performed at clinic rather than professional or patient level to avoid any risk of contamination of both study arms. However, such a clustered design requires more patients to achieve an appropriate study power, which can be done by increasing either the number of clusters or the number of patients per cluster. However, as the patient sample was randomly selected from electronic databases that are maintained for financial purposes, in which treatment or diagnosis type was not indicated, sufficient inclusion per indicator could not be guaranteed beforehand. Incomplete inclusion unfortunately made it extremely difficult to achieve statistical power for some of the less frequently applicable indicators. Proving changes in effect of the interventions was moreover complicated because effect sizes in implementation research are in general only small to moderate (5–15%) (Bero et al., 1998; Grol and Grimshaw, 1999; Grimshaw et al., 2004).

Unfortunately, non-responder data were not available due to privacy restrictions. Due to the nature of the patient selection (using a financial registration system), it remains unclear to what extent non-responders were those who did not consent to participate, or were those non-eligible for participation.

Choosing a complete set of guidelines with a large number of concrete recommendations for practice as a study subject was an ambitious and challenging task. The same professionals are involved in providing diagnosis and treatment of the various types of infertility, improving the use of national fertility guidelines

<table>
<thead>
<tr>
<th>Table II</th>
<th>Indicator scores at baseline and after-measurement for the minimal and maximal intervention strategies: guideline ‘initial assessment of fertility’.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indicator (Initial assessment of fertility)</strong></td>
<td><strong>Minimal intervention</strong></td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
</tr>
<tr>
<td></td>
<td>% (n)</td>
</tr>
<tr>
<td>The initial fertility assessment should result in both a diagnosis and a prognosis</td>
<td>45.2 (231)</td>
</tr>
<tr>
<td>The initial fertility assessment should consist of three parts: semen analysis, tubal occlusion and cycle analysis</td>
<td>32.3 (100)</td>
</tr>
<tr>
<td>Couple’s history-taking should cover at least: age of both partners, duration of infertility, type of couple’s infertility (primary or secondary)</td>
<td>97.0 (209)</td>
</tr>
<tr>
<td>Woman’s physical examination should include assessment of the body mass index</td>
<td>74.1 (193)</td>
</tr>
<tr>
<td>Life style advice should be part of the counselling regarding pregnancy probabilities</td>
<td>50.8 (65)</td>
</tr>
<tr>
<td>Advice concerning smoking</td>
<td></td>
</tr>
<tr>
<td>Advice concerning alcohol</td>
<td>45.2 (115)</td>
</tr>
<tr>
<td>Advice concerning bodyweight</td>
<td>48.9 (90)</td>
</tr>
</tbody>
</table>

Some patient numbers or percentages shown in the table regarding the baseline measurement may slightly differ from those reported in Mourad et al. (2008). Increased insight during the study period led to the researchers deciding to make minor adjustments in the complex algorithms behind some indicators. Indicators of both baseline and after-measurement in the table were calculated using exactly the same indicator algorithms, and are thus comparable over time.

NS, non-significant.

<sup>a</sup>Corrected for baseline differences.

<sup>★</sup>Significant difference in favour of the maximal intervention.
so interventions were aimed at these professionals irrespective of background characteristics of their patients (i.e. diagnosis or treatment type). Had the interventions had an overall positive effect on guideline adherence, a large group of patients could have benefitted from an increase in health gain and overall efficacy of care. This in return, would have had a substantial beneficial impact on cost-effectiveness of the strategies. However, taking the lack of observed effects for the comprehensive set of recommendations into account, and the difficulties of including enough patients of some of the indicators, some further remarks should be made for future studies. We recommend to carry out a future c-RCT with a more strictly specified subgroup of patients, for example with the shared diagnosis ‘male infertility and a more limited set of recommendations’. This would have several advantages. First, the patient sample could be more easily identified by either clinical or laboratory registrations. Secondly, when enrolling the intervention, professionals only have to focus on a distinct patient group and limited clinical care pathways, thus probably increasing their awareness of the provided interventions.

**Incomplete execution of the intervention**

The patient leaflet was received by only 37% and read by 30% of intended patients from the maximal intervention clinics. The vast majority of leaflet recipients read the material (87, 92 and 87%, respectively, for each of the leaflets) and highly appreciated it. We learn from these results that patients on the one hand are amenable to this innovative type of implementation strategy, whereas the professionals on the other hand either appeared to be quite indifferent to the execution of this intervention or did not consider it their job to do so. Subgroup analysis of patients that actually received the patient information leaflets was unfortunately not possible because of low numbers per indicator. Obviously the success of a maximal intervention as described in this paper depends on optimal dissemination of intervention materials, for example due to the presence of local opinion leaders who commit themselves to the project. We moreover recommend searching for alternative ways of implementing such interventions, for example direct automated mailing of material through identification of patients from, for example, electronic patient files, systematic dissemination on the floor through specialized nurses or, perhaps quite unorthodox, even by bypassing the already busy professionals by entrusting the patients associations alone with this task.

As the feedback report and meeting were generally rated as the most valuable intervention elements by the professionals, it could very well be possible that the surplus value of the maximal implementation strategy was limited. This is in accordance with the literature, which recently questioned the previously reported advantages of multi-faceted over single interventions (Grimshaw et al., 2006). Change needs time, certainly when it concerns effecting a change in the communication and decision-making process by patient education and empowerment. It could be a slow process requiring a redefinition of traditional roles and a paradigm shift in doctor–patient interaction. As stated above, the challenge of implementing a comprehensive set of guidelines by addressing all dimensions of fertility care at the same time is a huge one. A longer implementation period (e.g. 12 or even 24 months) might have been more justified to uncover intervention effects, especially as the interventions were introduced just before the regional summer holiday season.

**Patient involvement**

So far, most evidence from the literature on guideline implementation strategies concerns professional-oriented and to a lesser extent organizational interventions. The patient is seldom directly involved in introducing change and improvement. However, individual preferences of the patient can indeed have major impact on decisions about which healthcare is delivered (Wensing and Elwyn, 2003; Elwyn et al., 2005). Although this c-RCT did not prove any beneficial effect of either of the tested strategies, we think that the role of patients in influencing and improving healthcare provision should be a continued focus of attention in future research on fertility care. Negative effects of the interventions, such as an increase of anxiety or depression scores, were not observed. Patients from the maximal intervention arm were even found to be less anxious about twin pregnancy or the effect on the physical relationship with their partner, which could be caused by more open communication with their care providers. Regarding the infertility-related anxiety items, women in the maximal arm were more fearful of definite childlessness ($P = 0.033$) at baseline than at the after-measurement. This could be caused by positive ascertainment from their doctors or from increased understanding of their problem due to the patient leaflets; however, it might also be explained by longer duration of infertility and higher frequency of primary infertility at baseline. However, the large sample size in this study could have caused significant, but not necessarily clinically important, differences for the fertility-related anxiety items. As these secondary outcome measures were included in the study to get insight into any unwanted ‘side-effects’ of the used interventions, we can conclude that these did not exist, as differences were in favour of the maximal intervention. Using these items in future research to establish validated cut-off points for clinical use would be helpful.

Summarizing, the results of our study, although not evidently effective, can contribute to an increased understanding of the potential role of patients in clinical guideline implementation, as the process evaluation data on the patient-oriented intervention showed promising results; patients did feel empowered to act as a partner in the diagnostic and treatment process and experienced an improvement in communication. Whether professionals are also prepared to accept patients as equal partners in clinical decision making, remains another challenging focus for further research (Jung et al., 2002). Similarly, the roles of leadership, team climate and local ‘culture for change’ could be interesting subjects for future research, (Davies et al., 2000; Bower et al., 2003), particularly because some professionals indeed felt and reported that ‘initiating practice changes’ was not their responsibility.

In conclusion, in this c-RCT we did not find an overall sustainable effect of the studied professional- or patient-oriented interventions, which aimed to improve the use of a set of national guidelines for comprehensive fertility care. However, an evaluation of the patient-oriented intervention showed generally positive experiences with patient empowerment and doctor–patient communication. Professionals appeared quite indifferent to the disseminated implementation materials, raising the question of whether they should be the only stakeholders responsible for guideline implementation. This provides a promising basis for further studies on the effects of patient-oriented implementation of professional guidelines.
Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Acknowledgements

We would like to acknowledge all patient couples for their participation, the 16 fertility clinics and the coordinating gynaecologists for their cooperation in data sampling: P.A. van Dop (St Catharina Ziekenhuis, Eindhoven), B.W. Mol (Maxima Medisch Centrum, Veldhoven), A.M.H.W. Franssen (Canisius Wilhelmina Ziekenhuis, Nijmegen), M. Van Rozendaal (Ziekenhuis Bernhoven, Veghel), J.H.A. Vollebergh (Ziekenhuis Bernhoven, Oss), A.G. Minkhorst (Maasziekenhuis Pantein, Boxmeer), C.J.C.M. Hamilton (Jeroen Bosch Ziekenhuis, den Bosch), H. Ruis (behandelcentrum stichting Geertrgen, de Mortel), G.J.M. Muijser (Rivierenland Ziekenhuis, Tiel), J.H. Schagen van Leeuwen (St Antonius Ziekenhuis, Nieuwegein), L.J. van Dam (Gelre Ziekenhuis, Apeldoorn), E. Scheenjes (Ziekenhuis Gelderse Vallei, Ede), A.P.E. Schmoutziguer and L.F. Bancsi (Ziekenhuis Rijnstate, Arnhem), V.M. Blom (Streekziekenhuis Zevenaar, Zevenaar), E.M. Tepe (Slingeland Ziekenhuis, Doetinchem). Furthermore, special thanks to Anke Borgers, Petra Frenken, Monica van Zanten and Jasper van den Berg for their help with data collection. Evelyn Verheijden and Janine Liefers for their help with data management and analysis.

Conflict of interest: none declared.

Funding

The Netherlands Organization for Health Research and Development (ZonMw) funded this project (grant no. 945-14-116).

References
