

# Quality indicators for all dimensions of infertility care quality: consensus between professionals and patients

E.A.F. Dancet<sup>1,2,3,\*†</sup>, T.M. D’Hooghe<sup>1,†</sup>, C. Spiessens<sup>1</sup>, W. Sermeus<sup>3</sup>,  
D. De Neubourg<sup>1</sup>, N. Karel<sup>3</sup>, J.A.M. Kremer<sup>4</sup>, and W.L.D.M. Nelen<sup>4</sup>

<sup>1</sup>Leuven University Hospital, Leuven University Fertility Centre, Herestraat 49, Leuven 3000, Belgium <sup>2</sup>Center for Reproductive Medicine, Women’s and Children’s Hospital, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands <sup>3</sup>Centre for Health Services & Nursing Research, Catholic University Leuven, Kapucijnenvoer 35, 3000 Leuven, Belgium <sup>4</sup>Department of Obstetrics and Gynecology, Radboud University Nijmegen Medical Centre, Geert Grooteplein Zuid 10, Nijmegen 6525 GA, The Netherlands

\*Correspondence address. E-mail: eline.dancet@uzleuven.be

Submitted on July 29, 2012; resubmitted on January 17, 2013; accepted on February 13, 2013

**STUDY QUESTION:** What is the relative importance of the six dimensions of quality of care according to different stakeholders and can a quality indicator set address all six quality dimensions and incorporate the views from professionals working in different disciplines and from patients?

**SUMMARY ANSWER:** Safety, effectiveness and patient centeredness were the most important quality dimensions. All six quality dimensions can be assessed with a set of 24 quality indicators, which is face valid and acceptable according to both professionals from different disciplines and patients.

**WHAT IS KNOWN ALREADY:** To our knowledge, no study has weighted the relative importance of all quality dimensions to infertility care. Additionally, there are very few infertility care-specific quality indicators and no quality indicator set covers all six quality dimensions and incorporated the views of professionals and patients.

**STUDY DESIGN, SIZE AND DURATION:** A three-round iterative Delphi survey including patients and professionals from four different fields, conducted in two European countries over the course of 2011 and 2012.

**PARTICIPANTS/MATERIALS, SETTINGS AND METHODS:** Dutch and Belgian gynaecologists, embryologists, counsellors, nurses/midwives and patients took part ( $n = 43$  in round 1 and finally 30 in round 3). Respondents ranked the six quality dimensions twice for importance and their agreement was evaluated. Furthermore, in round 1, respondents gave suggestions, which were subsequently uniformly formulated as quality indicators. In rounds 2 and 3, respondents rated the quality indicators for preparedness to measure and for importance (relation to quality and prioritization for benchmarking). Providing feedback allowed selecting indicators based on consensus between stakeholder groups. Measurable indicators, important to all stakeholder groups, were selected for each quality dimension.

**MAIN RESULTS:** All stakeholder groups and most individuals agreed that safety, effectiveness and patient centeredness were the most important quality dimensions. A total of 498 suggestions led to the development of 298 indicators. Professionals were sufficiently prepared to measure 204 of these indicators. Based on importance, 52 (7–15 per dimension; round 2) and finally 24 (4 per dimension; round 3) quality indicators were selected.

**LIMITATIONS, REASONS FOR CAUTION:** The final quality indicator set does not cover the entire care process, but rather takes a ‘sample’ of each quality dimension. Although the quality indicators are face valid and acceptable, their psychometric characteristics need to be tested by further research.

**WIDER IMPLICATIONS OF THE FINDINGS:** Quality management should focus on safety, effectiveness and patient centeredness of care. Clinics can use the quality indicator set to assess all quality dimensions of their care.

**Key words:** infertility / quality control / assisted reproduction

† These authors contributed equally as first authors to this paper.

## Introduction

Quality management is important to infertility care, as new artificial reproductive technologies (ARTs), are frequently introduced without sufficient prior safety and effectiveness studies (Schatten, 2002; Winston and Hardy, 2002) in vulnerable patients burdened emotionally, socially and sometimes also physically (Verhaak *et al.*, 2007; Hinton *et al.*, 2010).

Quality management starts with measuring quality. Quality indicators are measurable elements of health care for which there is evidence or consensus that they assess quality of care, and which are useful to quality management (Donabedian, 1988; Mainz, 2003a). The vast majority (89.3%) of medical quality indicators addresses safety and effectiveness of care (Copnell *et al.*, 2009). However, indicator sets should fully represent health-care quality and, therefore, cover the six dimensions of quality of care, including effectiveness, safety, efficiency, timeliness, equity and patient centeredness (Mainz, 2003a).

Infertility care has very few speciality specific quality indicators compared with other fields (Copnell *et al.*, 2009) and experts have called for the development and selection of quality indicators (Temmerman *et al.*, 2006; Nelen *et al.*, 2007; Gerris, 2011). The previous European 'Reprostat' indicator set addressed reproductive *health status* and not *health care* ([http://ec.europa.eu/health/ph\\_projects/2001/monitoring/fp\\_monitoring\\_2001\\_a1\\_frep\\_02\\_en.pdf](http://ec.europa.eu/health/ph_projects/2001/monitoring/fp_monitoring_2001_a1_frep_02_en.pdf)). Global ART monitoring solely focuses on effectiveness and safety (Nygren *et al.*, 2011). Two previous guideline-based indicator sets developed by our research group have addressed four of the six quality dimensions (i.e. effectiveness, safety, efficiency and timeliness; Mourad *et al.*, 2007; Haagen *et al.*, 2010).

So far, studies developing and selecting quality indicators for infertility care have mainly involved physicians (Mourad *et al.*, 2007; Haagen *et al.*, 2010), and have not included the perspective from other health professionals involved in infertility care, or the view from patients who receive this care. In contrast, patients have been involved in the selection of quality indicators through focus groups in the field of end-of-life cancer care (Grunfeld *et al.*, 2008) and through Delphi questionnaires in the fields of primary mental health care (Shield *et al.*, 2003; Campbell *et al.*, 2004), lung cancer care (Hermens *et al.*, 2006) and general practice care (Jeacocke *et al.*, 2002).

Additionally, to our knowledge, no study has weighted the relative importance of all quality dimensions to infertility care. Discrete choice experiments with physicians and patients weighted the importance of effectiveness and some, but not all, other quality dimensions (Ryan, 1999; Musters *et al.*, 2011; Palumbo *et al.*, 2011; van Empel *et al.*, 2011). These studies demonstrated the significant value of effectiveness, but also the willingness to trade it off for other quality dimensions, like patient centeredness (Ryan, 1999; Palumbo *et al.*, 2011; van Empel *et al.*, 2011).

The aim of this study was 2-fold. First, we aimed to evaluate the relative importance of the six dimensions of quality of care to different stakeholders. Secondly, we aimed to develop a quality indicator set addressing all six quality dimensions and incorporating the views from professionals working in different disciplines and from patients.

## Methods

In a systematic iterative Delphi-consensus study, we used a coded questionnaire and gave feedback on group responses (Jones and Hunter, 1995; Campbell *et al.*, 2002) in order to (i) rank the dimensions of quality of care, (ii) generate quality indicators and (iii) select a comprehensive set of face-valid (i.e. they look as though they are measuring the appropriate construct; Polit and Beck, 2008), acceptable quality indicators. The perspectives of different stakeholders were incorporated, by questioning gynaecologists, embryologists, counsellors, nurses/midwives and patients from both Belgium and the Netherlands. Ethical approval was granted in the first stage without further revision. Participants were provided with written study information and could contact the researchers with questions.

### Selection of experts

A purposive (two clinics were hand picked based on their University character and based on their location in different countries), stratified (including stakeholders from the following strata: gynaecologists, embryologists, counsellors and nurses/midwives) sample of professionals was recruited (Polit and Beck, 2008). Fifteen staff members of the Leuven University fertility clinic (Belgium) took part as a mandatory element of the clinics' quality management system (Van den Broeck *et al.*, 2012; Willemen *et al.*, 2012; i.e. to prevent self-selection bias; Sica, 2006). Fourteen staff members from the Nijmegen University fertility clinic (the Netherlands) were asked to take part on a voluntary basis. In both clinics, all gynaecologists, senior embryologists and counsellors were addressed (i.e. to prevent sample bias; Sica, 2006) and four nurses/midwives were selected based on their clinical experience and commitment to quality management.

A convenience sample (Polit and Beck, 2008) of patients was recruited via an announcement posted during 2 weeks on the websites of the national patient associations ([www.deverdwaaaldeoievaar.be](http://www.deverdwaaaldeoievaar.be); [www.freya.nl](http://www.freya.nl)). Fourteen Dutch and 10 Belgian patients contacted the researchers and were all given the opportunity to take part.

### Three Delphi rounds

In Delphi round 1, respondents ranked the six dimensions of quality of care for importance and proposed 0–3 quality indicators for each quality dimension based on the definitions of quality indicators (Lawrence and Olesen, 1997; Campbell *et al.*, 2002) and quality dimensions (Corrigan *et al.*, 2001; Supplementary data, Table S1). Patients additionally answered demographic and medical questions. Analysis of the data collected in Delphi round 1 (described below) led to the development of the Delphi-round-2 questionnaire.

In Delphi round 2, respondents re-ranked the dimensions of quality of care based on the feedback on their own ranking and their stakeholder group's ranking in Delphi round 1 (i.e. group ranking and ranges of individual ranking per dimension). Furthermore, respondents rated the indicators (developed based on the input in Delphi round 1 from all respondents) in Delphi round 2 for both their preparedness to measure a specific indicator (not by patients) and for the strength of the relation between a specific indicator and quality of care on nine-point scales (1 = not at all prepared or no relation; 9 = most prepared or strongest relation; Hermens *et al.*, 2006). Next, respondents selected for each dimension their top-three indicators for the benchmarking of clinics. A glossary based on a Dutch translation of the International Committee for Monitoring Assisted Reproductive Technology glossary was added to the questionnaire (Zegers-Hochschild *et al.*, 2009; De Neubourg *et al.*, 2012). Analysis of the data collected in Delphi round 2

(described below) led to the development of the Delphi-round-3 questionnaire.

In Delphi round 3, respondents re-rated a selection of the developed indicators (selected based on the ratings from Delphi round 2) for their relation to quality of care and re-ranked them for the benchmarking of clinics. In order to carry out this task, each respondent received feedback on their own Delphi-round-2 ratings and on the following Delphi-round-2 ratings from others: medians for all respondents, number of top-three priorities across all respondents and specification of the stakeholder groups contributing to indicator selection. This last Delphi round aimed to increase the level of consensus among the stakeholder groups and to further limit the number of finally selected indicators.

## Ranking the dimensions of quality of care

The final rankings of the six quality dimensions were based on the results of Delphi round 2, which were, in turn, influenced by feedback on the results of Delphi round 1.

Rankings of respondents of the same group were combined by adding up the following scores per dimension: six points per respondent that ranked the dimension first, five points per respondent that ranked the dimension second and so on. The dimension with the highest score was considered the groups' first priority and so on.

The dimensions' rankings were considered at eight levels, including five stakeholder group rankings, two country rankings and the final ranking (giving equal weight to the five stakeholder group rankings). Additionally, the number of individual respondents agreeing on a rank was considered.

The level of intra-group and inter-group agreement (Polit and Beck, 2008) on the dimensions' rankings in Delphi round 2, at all nine levels, was described with Kendall's *W* coefficients of concordance (ranging from 0: no agreement to 1: complete agreement).

The evolution of the rankings and agreement on rankings between Delphi rounds 1 and 2 was appraised by comparing the final rankings and rankings per stakeholder group after both rounds.

## Development of quality indicators

Quality indicators were developed per dimension based on suggestions given by all respondents in Delphi round 1. First, the suggestions were grouped for content and relocated to the best fitted dimensions, if necessary. Secondly, suggestions were transformed to a uniform format for indicators (Mainz, 2003a) and internationally accepted terminology for reproductive medicine was used (Zegers-Hochschild et al., 2009; De Neubourg et al., 2012). 'Process indicators' describing what is actually done in giving and receiving care and 'outcome indicators' describing the effects of care on patients' health status, knowledge and behaviour (Mainz, 2003a) were formulated as 'rate-based indicators'. Rate-based indicators include a numerator, a denominator (specifying the population at risk for an event limiting the need for case-mix adjustments) and a period of time over which the event may take place (Mainz, 2003a). 'Structure indicators' describing the attributes of the settings in which care occurs (Mainz, 2003a) were provided with the timestamp 'at a certain moment in time'.

## Selecting quality indicators

*Selection for Delphi round 3 based on the results of Delphi round 2*

A limited set of indicators (aimed at maximal 60) covering all quality dimensions, and including indicators that professionals were prepared to measure and that were important to respondents (i.e. most related to quality of care and/or most often a top-priority), was selected for Delphi round 3. Therefore, for each indicator respondents' ratings on 'preparedness to measure', 'relation to quality of care' and their 'top-three

priorities' were considered at the following three levels: 'per group', 'all respondents' and 'across groups' (Table I). Dimension-specific cut-off values for the selection of quality indicators were the dimension-specific medians, 25th and 75th percentiles (Jones and Hunter, 1995; based on ratings of the entire group of respondents) and the proportions of top-three priorities (Table I).

*Selection for the final indicator set based on the results of Delphi round 3*

Finally, an even more limited set of indicators (aimed at four per dimension), covering all six quality dimensions and supported by all stakeholder groups (based on ratings for relation to quality of care and/or benchmarking priorities), was selected based on the results of Delphi round 3.

Ratings for relation to quality of care and benchmarking priorities were considered at stakeholder group level. Indicators were selected if at least two of the following four inclusion criteria were fulfilled for all five stakeholder groups: (i) at least one stakeholder group member allocated its' highest score for relation to quality of care for that dimension; (ii) half or more of the stakeholder group members allocated a score for relation to quality of care higher than the stakeholder group's median score for relation to quality of care for the respective dimension; (iii) at least one stakeholder group member gave a top-three priority and (iv) half or more of the stakeholder group members gave a top-five priority.

## Results

### Respondents

Of the 53 respondents invited for Delphi round 1, 30 (57%) took part in all rounds (Fig. 1). Each country and all stakeholder groups were represented in every round, but response rates were highest for, respectively, Belgium (84% instead of 32% in the Netherlands) and gynaecologists (Fig. 1). Reasons for not responding were mainly time constraints among professionals; three patients said the questionnaires were too difficult. In the course of the three Delphi rounds, the researchers only needed to provide additional clarification to three respondents (an embryologist, a midwife and a patient). In all three cases, the difference between rating indicators on a nine-point scale for importance and ranking the indicators per dimension for priority (Delphi round 3) had to be explained.

In Delphi round 1, all patients were women with a (University) college degree (average age = 32.6 years). Almost half of them (43.8%) filled the questionnaire out together with their male partner. A small minority ( $n = 3$ ; 18.8%) had at least one child, mostly ( $n = 2/3$ ) conceived through assisted reproduction and two patients were pregnant. Patients had experienced IVF/ICSI treatment (69%), IUI (69%) and/or timed intercourse with ovulation induction (31%).

### Rankings of the dimensions

The final ranking of the quality dimensions (Delphi round 2) in order of importance was: 'safety', 'effectiveness' and 'patient centeredness' (shared second and third places), 'efficiency', 'timeliness' and 'equity' (Table II). The level of agreement among the five stakeholder groups on this final ranking was good ( $W = 0.87$ ; Delphi round 2) and increased after Delphi round 1 ( $W = 0.81$ ). Final agreement was considerable at a country level ( $W = 0.64-0.65$ ) and was lowest for patients ( $W = 0.52$ ) and highest for nurses/midwives ( $W = 0.93$ ; Table II). An inter-stakeholder-group agreement was high, ranging

**Table 1** Exclusion and inclusion criteria for indicator selection based on Delphi round 2<sup>a,b</sup>.

	Measurement level	Per group	All respondents	Across groups
Exclusion criteria <sup>c</sup>	Preparedness to measure	At least one group has a group median for preparedness to measure < the dimension-specific P25 for preparedness to measure	The entire group median for preparedness to measure < the dimension-specific median for preparedness to measure	At least three groups (= all but one group judging preparedness to measure) have a group median < the dimension-specific median for preparedness to measure
Inclusion criteria <sup>d</sup>	Relation to quality of care	At least one group has a group median for relation to quality of care > the dimension-specific P75 for relation to quality of care	Importance The entire group median for relation to quality of care > the dimension-specific median for relation to quality of care	At least four groups (= all but one group judging relation to quality) have a group median > the dimension-specific median for relation to quality of care
	Top-three priority	For at least one group, more than one-third of the group members have selected the indicator as a top-three priority	At least one-fourth of the entire group has selected the indicator as a top-three priority	For at least four groups (= all but one group selecting top-three priorities) minimal one group member has selected the indicator as a top-three priority

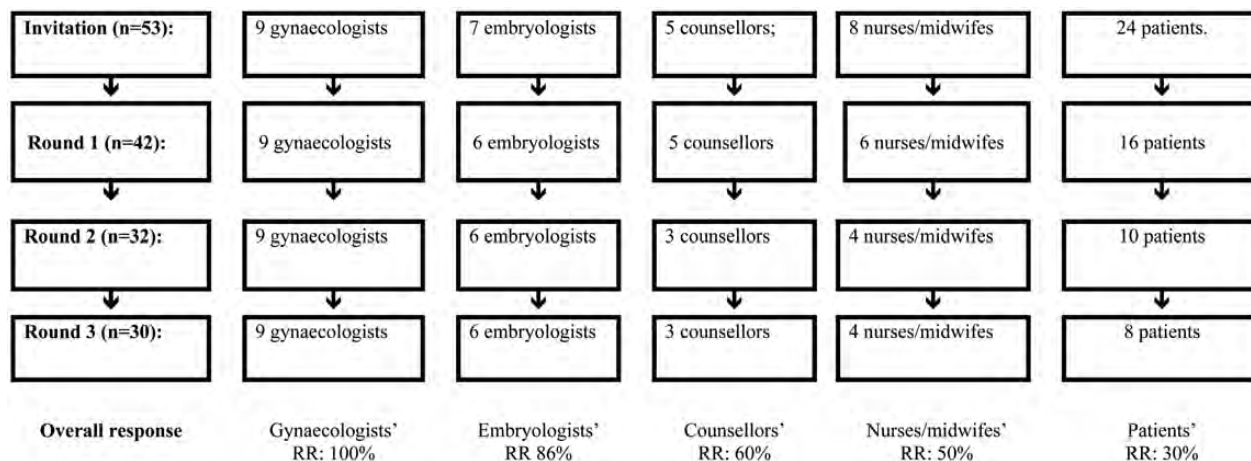
P25, the 25th percentile; P75, the 75th percentile.

<sup>a</sup>Cut-off values were defined based on the ratings of all respondents (i.e. including all stakeholder groups).

<sup>b</sup>In order to be selected, indicators could not meet the exclusion criteria and should meet the inclusion criteria.

<sup>c</sup>The exclusion criteria needed to be met at minimal one of the three levels (i.e. per group or all respondents or across groups).

<sup>d</sup>The inclusion criteria needed to be met at minimal one of the three levels (i.e. per group or all respondents or across groups).



**Figure 1** Number of respondents per Delphi round, per group and overall response rates per group.

from 0.84 between counsellors and nurses/midwives to 0.99 between counsellors and patients.

When the rankings at all eight levels and the individual rankings of most respondents ( $n = 21/32$ ; 66%) were considered, the three most important quality dimensions were 'safety', 'effectiveness' and 'patient centeredness'. Of these three dimensions, 'safety' was most important according to stakeholder group rankings of gynaecologists, embryologists and nurses/midwives and according to the individual rankings of the majority of respondents ( $n = 17/32$ ). 'Effectiveness' was more important than 'patient centeredness' according to gynaecologists and embryologists and the individual rankings from most respondents ( $n = 19/32$ ). Patient centeredness was most important to the patients' and counsellors' stakeholder groups (shared first place with effectiveness for counsellors).

Of the three least important dimensions, 'efficiency' seemed most and equity seemed least trivial as agreed on at all but one of the eight ranking levels considered (not the stakeholder group' rankings of 'counsellors') and by individual rankings from half of the respondents (16/32).

### Suggestions for quality indicators

In total, 36 respondents (86%) gave 498 suggestions for indicators in Delphi round 1. Four patients (25%) and three professionals (12%; a gynaecologist, a counsellor and a nurse) did not give any suggestions.

A minority of the suggestions (18%) needed to be re-classified in the six dimensions of quality of care.

Grouping similar suggestions limited the 498 suggestions to 298. More specifically, the number of proposed indicators per dimension



**Table II** Delphi-round-2 rankings of the dimensions of quality of care, the number of group members agreeing on the top-three and Kendall's coefficient of concordance at all the nine levels (including: five stakeholder group, two country, the overall ranking and the final ranking).

Levels at which the rankings were considered	The dimensions of quality of care <sup>a</sup>						Number of group members agreeing on top-three	Kendall's coefficient of concordance (W; P-value)
	Safety	Effectiveness	Patient centeredness	Efficiency	Timeliness	Equity		
Gynaecologists (n = 9)	1 (51)	2 (46)	3 (34)	4 (28)	5 (16)	6 (14)	7/9	0.82 (P < 0.0001)
Embryologists (n = 6)	1 (35)	2 (31)	3 (18)	4 (17)	5 (15)	6 (6)	4/6	0.91 (P < 0.0001)
Counsellors (n = 3)	3 (11)	1–2 (15)	1–2 (15)	4–5 (7)	4–5 (7)	6 (4)	1/3	0.61 (P = 0.105)
Nurses (n = 4)	1 (24)	3 (16)	2 (19)	4 (16)	6 (5)	5 (7)	3/4	0.93 (P = 0.002)
Patients (n = 10)	3 (44)	2 (46)	1 (51)	4 (25)	5 (24)	6 (20)	6/10	0.52 (P < 0.0001)
Belgian respondents (n = 21)	1 (106)	2 (104)	3 (92)	4 (62)	5 (44)	6 (33)	16/21	0.64 (P < 0.0001)
Dutch respondents (n = 11)	1 (59)	2 (50)	3 (49)	4 (42)	5 (23)	6 (18)	5/11	0.65 (P < 0.0001)
Overall ranking (n = 32; all respondents equal weight)	1 (165)	2 (154)	3 (137)	4 (93)	5 (67)	6 (29)	21/32	0.64; (P < 0.0001)
Final ranking (n = 5; equal weight for five stakeholder groups)	1 (26)	2–3 (24,5)	2–3 (24,5)	4 (14,5)	5 (9,5)	6 (5)	21/32	0.87 (P = 0.004)

<sup>a</sup>Rankings per group and total allocated scores within brackets.

were 72 for 'patient centeredness', 63 for 'safety', 59 for 'efficiency', 40 for 'effectiveness', 35 for 'timeliness' and 29 for 'equity'.

All suggestions needed to be transformed into the format of rate-based process or outcome indicators and of structure indicators.

## Quality indicator selection for Delphi round 2

Of the 298 indicators presented in Delphi round 2, 52 indicators (17%), including 6 ('equity') to 15 ('patient centeredness') per dimension, were selected as professionals were prepared to measure them and as they were judged important (Supplementary data, Tables SII and SIV).

The difference in dimension-specific cut-off values based on ratings on the nine-point scale in Delphi round 2 showed that respondents were more prepared to measure indicators for 'effectiveness' compared with other dimensions and that scores for relation to quality of care were comparable across all dimensions (Supplementary data, Table SII).

One-fifth of the developed indicators (n = 60/298) met the inclusion criteria for importance. One-third of the developed indicators (n = 94/298) met the exclusion criteria for insufficient preparedness to measure. A minority of the developed indicators (n = 8/298; 3%) met both the inclusion and exclusion criteria and were not selected; all but one of these indicators belonged to the dimension safety (Supplementary data, Table SII).

All stakeholder groups contributed to the inclusion of indicators based on importance (between 17 of 60 for counsellors and 30 of 60 for gynaecologists) and the exclusion based on preparedness to measure (between 3 of 94 for nurses and 72 of 94 for counsellors).

## Selection of the final indicator set (Delphi round 3)

Finally, 24 quality indicators were selected, which assessed the six dimensions of quality of care and which were strongly related to quality of care and/or highly prioritized for benchmarking according to gynaecologists, embryologists, counsellors, midwives/nurses and patients (Table III).

The dimension and group-specific cut-off values for selection differed. Each of the four inclusion criteria led to the identification of 15–21 indicators and included indicators from all dimensions (1–5 per dimension; Supplementary data, Table SIII). Whereas 17 of 24 indicators were selected based on both their strong relation to quality of care and their high benchmarking priority in each group, 7 of 24 indicators were selected based on one of both.

The final indicator set included process (n = 10), structural (n = 8) and outcome indicators (n = 6; Table III).

Suggestions given in Delphi round 1 from all groups contributed to the finally selected indicators. Each group had a unique contribution as at least one of the finally selected indicators was developed based on suggestions of only that specific group (Table III).

**Table III** Indicators selected for the final indicator set, specification of their type and the support from the different stakeholder groups in Delphi round 2 and support from the entire group of respondents in Delphi round 3.

Indicators per dimension	Type of indicator (Mainz, 2003a)	Stakeholder groups giving suggestion(s) leading to indicator development (Delphi round 1)	Support from stakeholder groups in Delphi round 2		Measures representing Delphi-round-3 ratings from all respondents	
			Professional stakeholder groups prepared to measure <sup>a</sup>	Stakeholder groups directly contributing to inclusion based on importance considered at group and across group level <sup>b</sup>	Number of respondents considering the indicator as top-three priority for benchmarking	Respondents' median score for relation to quality of care
<i>Safety</i>						
The number of fresh ART cycles with severe complications (OHSS, bleeding, infection, complaints of serious pain) resulting from the fertility treatment, which require hospitalization relative to the total number of fresh ART cycles during a certain time period	Outcome indicator	All	All	All but counsellors	26/30	8
The number of fresh ART cycles with complications (OHSS, haemorrhage, infection) as a result of MAR relative to the total number of fresh ART cycles during a certain time period	Outcome indicator	All	All	All but counsellors	10/30	7
The number of MAR cycles in which gametes or embryos get lost as a result of an accident, human error or mistake relative to the total number of MAR cycles during a certain time period	Process indicator	Embryologists	All but counsellors	Gynaecologists, patients	10/30	8
The number of reported mistakes or incidents caused by all care providers relative to the number of treatment cycles during a certain time period	Process indicator	Embryologists	All	Embryologists, nurses	13/30	8
<i>Effectiveness</i>						
The number of treated patients who go home with a live born baby relative to the total number of treated patients during a certain time period	Outcome indicator	All but nurses	All	all but counsellors	14/30	9

Continued

**Table III** *Continued*

Indicators per dimension	Type of indicator (Mainz, 2003a)	Stakeholder groups giving suggestion(s) leading to indicator development (Delphi round 1)	Support from stakeholder groups in Delphi round 2		Measures representing Delphi-round-3 ratings from all respondents	
			Professional stakeholder groups prepared to measure <sup>a</sup>	Stakeholder groups directly contributing to inclusion based on importance considered at group and across group level <sup>b</sup>	Number of respondents considering the indicator as top-three priority for benchmarking	Respondents' median score for relation to quality of care
The number of patients who after a maximum of three <i>fresh</i> ART cycles (oocyte aspiration actually performed) had a live birth (the expulsion or extraction of minimally one fetus showing evidence of life) relative to the total number of patients starting an ART cycle during a certain time period	Outcome indicator	Gynaecologists	All	Gynaecologists	21/30	7 <sup>c</sup>
The number of live births (the complete expulsion or extraction of a product of fertilization that shows evidence of life) after a <i>fresh</i> ART cycle with embryo transfer relative to the total number of <i>fresh</i> ART cycles with embryo transfer during a certain time period	Outcome indicator	Gynaecologists	All	None	9/30	7.5
The number of pregnancies in women younger than 36 years old as a result of a <i>fresh</i> ART cycle relative to the total amount of <i>fresh</i> ART cycles in women younger than 36 years old during a certain time period	Outcome indicator	Gynaecologists, embryologists	All	All but counsellors	13/30	8
<i>Patient centeredness</i>						
The number of patients of a fertility clinic to whom psychosocial counselling was offered relative to the total number of patients of that fertility clinic during a certain time period	Process indicator	All	All	All but nurses	17/30	7
The regular organization of a multidisciplinary meeting of the fertility clinic in which the psychosocial context of the patient can be discussed if necessary during a certain time period	Structural indicator	All	All	Gynaecologists, nurses	10/30	8
The provision of the offer to patients of psychosocial counselling at a certain moment in time	Structural indicator	All	All but nurses	Patients	7/30	8

The number of patients who opinionated that their personal experiences and wishes were actually heard relative to the total number of interrogated patients during a certain time period	Process indicator	Embryologists	All but counsellors	None	4/30	8
<i>Efficiency</i>						
The number of patients undergoing a very thorough diagnostic phase and reaching a diagnosis prior to starting MAR relative to the total number of patients starting MAR during a certain time period	Process indicator	Embryologists, nurses, patients	All	All but nurses	24/30	8
The existence of a website of the fertility clinic containing all the basic information, contracts and information about studies and FAQs at a certain moment in time	Structural indicator	Gynaecologists	All	All but counsellors	16/30	8
The provision of the use of an electronic patient record containing all relevant clinical information and allowing the extraction of letters and reports at a certain moment in time	Structural indicator	Gynaecologists	All	All but nurses	22/30	8
The total number of FTE care providers relative to the total number of treated patients per type of care provider during a certain time period	Structural indicator	Patients	All	All but patients	9/30	7
<i>Timeliness</i>						
The average duration of the waiting time per new patient between the asking and the getting of the first appointment during a certain time period	Process indicator	Gynaecologists, embryologists, patients	All but counsellors	All but embryologists	16/30	7
The average duration of the waiting time during MAR per patient between having the need for and attending an urgent consultation in case of unexpected negative results (e.g. fertilization failure) during a certain time period	Process indicator	Gynaecologists, patients	All	All but embryologists	22/30	8
The average duration of the waiting time per patient between the first appointment and the start of the first treatment cycle during a certain time period	Process indicator	All but embryologists	All but counsellors	None	25/30	7
The average duration of the waiting time in the waiting room per patient between the agreed time to start a consultation and actual starting time of the consultation during a certain time period	Process indicator	All	All	All but counsellors	21/30	7.5

Continued



**Table III** *Continued*

Indicators per dimension	Type of indicator (Mainz, 2003a)	Stakeholder groups giving suggestion(s) leading to indicator development (Delphi round 1)	Support from stakeholder groups in Delphi round 2		Measures representing Delphi-round-3 ratings from all respondents	
			Professional stakeholder groups prepared to measure <sup>a</sup>	Stakeholder groups directly contributing to inclusion based on importance considered at group and across group level <sup>b</sup>	Number of respondents considering the indicator as top-three priority for benchmarking	Respondents' median score for relation to quality of care
<i>Equity</i>						
The number of patients who opinionated that she/he is being respected by her/his physician relative to the total number of interrogated patients during a certain time period	Process indicator	Gynaecologists, counsellors	All	All but embryologists	14/30	7.5
The provision of clearly described in- and exclusion criteria for MAR in the fertility clinic (among others taking into account the national legislation) at a certain moment in time	Structural indicator	Gynaecologists, counsellors, nurses	All	All	24/30	7.5
The provision of a clearly explained vision of the fertility clinic concerning ethical limitations (e.g. no surrogacy) of which at no time nor for no reason (e.g. power, money) can be deviated at a certain moment in time	Structural indicator	Nurses	All but counsellors	All but nurses	13/30	8
The provision of protocols that are in accordance with international guidelines/ recommendations of care concerning equity and taking account of the universal needs at a certain moment in time	Structural indicator	Counsellors	All	All but nurses	16/30	7

ART, artificial reproductive technology; OHSS, ovarian hyperstimulation syndrome; MAR, medically assisted reproduction; FTE, full-time equivalent.

<sup>a</sup>A stakeholder group was judged prepared to measure the indicator if the group median for preparedness to measure  $\geq$  the dimension-specific median.

<sup>b</sup>At a group level, stakeholder groups were considered to contribute directly to inclusion of indicators based on importance if their group gave exceptionally high ratings at the group level (group median for relation to quality of care > the dimension-specific P75; > 1/3 of the group members selected the indicator as a top-three priority). At across group level, stakeholder groups were considered to contribute directly to inclusion of indicators based on importance or if their group and three other groups gave high ratings at across group level ( $\geq 4/5$  groups have a group median for relation to quality of care > the dimension-specific median ( $\geq 4/5$  groups had minimal one group member selecting the indicator as a top-three priority). Besides this direct contribution to inclusion at group and across group levels, indicators could also be included at entire group level.

<sup>c</sup>The only finally selected indicator that did not have a median score (among all respondents) for relation to quality of care higher or equal to the dimension-specific median (i.e. eight for effectiveness and seven for the other dimensions).

All professional stakeholder groups were prepared to measure most ( $n = 18/24$ ) of the finally selected indicators (Table III). All but one professional stakeholder group (counsellors or nurses) were prepared to measure the remaining indicators ( $n = 2$ ; Table III).

All, but one, of the selected indicators finally had a median score (among all respondents) for relation to quality of care higher or equal to the dimension-specific median (i.e. eight for effectiveness and seven for the other dimensions; Table III). Additionally, all selected indicators were selected as top-three indicators for benchmarking in Delphi round 3 by 4–26 respondents (Table III). The inclusion of most ( $n = 16/24$ ) of the 24 indicators was already directly supported by the 'importance and ratings' (i.e. ratings for the strength of the relation to quality of care and/or top-three priority rankings) of four in five stakeholder groups in Delphi round 2 (Table III).

### Patients' perspective and contribution

Patients agreed with professionals that safety, effectiveness and patient centeredness were the most important quality dimensions (Table II). Nevertheless, other than all professionals but counsellors (who gave top-priority to safety), the patients' stakeholder group gave their top-priority to patient centeredness (Table II). All professional stakeholder groups had a high level of agreement with patients on the ranking of the quality dimensions ( $W \geq 0.84$ ) and agreement was highest between the patient group and the counsellor group ( $W = 0.99$ ).

Twenty-six of the 298 indicators presented in Delphi round 2 came exclusively from patients' suggestions. Of these, nine indicators were excluded after Delphi round 2 based on professionals' insufficient preparedness to measure them, and 15 indicators were not included after Delphi round 2 based on insufficient importance. Two of these 26 indicators were selected for Delphi round 3 (thanks to benchmarking priorities of nurses and counsellors, not patients), and one was selected for the final indicator set.

## Discussion

### Novelty

This is the first study in the field of infertility care to rank the six dimensions of quality of care for importance and to develop a quality indicator set that covers all six quality dimensions. Additionally, including the input from professionals from four different disciplines as well as from patients is novel.

### The ranking of the six quality dimensions for infertility care

The top-priority of safety, effectiveness and patient centeredness is not surprising given the historical focus on effectiveness and safety (Nygren *et al.*, 2011) and the recent popularity of patient centeredness in infertility care (Dancet *et al.*, 2010, 2011a,b, 2012; Van Empel *et al.*, 2010; Aarts *et al.*, 2012a).

The fact that safety was chosen as top-priority over effectiveness was, on the one hand, surprising as successful treatment outcome (e.g. pregnancy rates) is often considered the golden standard in the evaluations of fertility treatments (Haan, 1991; Neumann *et al.*, 1994; Ryan, 1999; Min *et al.*, 2004; Chambers *et al.*, 2009; Griffiths *et al.*, 2010; Musters *et al.*, 2011; Palumbo *et al.*, 2011; van Empel

*et al.*, 2011). On the other hand, the group of gynaecologists, embryologists and nurses rating safety as top priority possibly feel responsible for treatment safety.

The fact that patients ranked safety less highly than effectiveness has been reported before in female infertility patients (Scotland *et al.*, 2007; Palumbo *et al.*, 2011). It is difficult to decide whether the patients' or the professionals' perspective should be given more weight. We believe that patient autonomy cannot be interpreted as an absolute ethical principle (Chervenak and McCullough, 1990). For example, it may be more effective (increase pregnancy rate) for a 30-year-old infertile woman to have four or more embryos transferred when compared with single embryo transfer. However, this choice is also associated with potentially serious obstetric and perinatal complications for her and her offspring if a high-order multiple pregnancy is established. It remains the ethical responsibility of the physicians to prevent potential harm in the first place. Therefore, the conclusion that safety is a top priority together with effectiveness and patient centeredness, on which both patients and professionals agree, is most valid.

The high priority given to patient centeredness might be specific to the field, as gynaecologists are known to have a more patient-centred attitude than surgeons (Chan and Ahmad, 2012).

The importance of the three lowest ranked quality dimensions should not be forgotten as fertility patients are willing to trade-off effectiveness of treatment for equity (i.e. costs) and timeliness (Ryan, 1999; Musters *et al.*, 2011; Palumbo *et al.*, 2011).

### The multidimensional quality indicator set for infertility care

This study is the first to show that all quality dimensions of infertility care can be assessed with quality indicators.

The final set is face valid and acceptable as it is underpinned by consensus among five different stakeholder groups, including those being assessed and those benefitting from the assessment (Campbell *et al.*, 2002). The set meets the need for process and structure indicators, besides outcome indicators (Nelen *et al.*, 2007).

The final quality indicator set does not cover the entire care process, but rather takes a 'sample' of each quality dimension, as many different suggestions for indicators were given for each dimension and as understanding quality requires many different measures (Mainz, 2003a). We defined that we wanted to select only 24 quality indicators, based on the highest agreement among stakeholder groups. In retrospect, many more interesting quality indicators were developed and rated as rather face valid and acceptable. These additional indicators (Supplementary data, Table SIV) could also be used for quality management and for further research.

As the indicators were not based on the literature or guidelines but on the creativity of experts and patients in order to generate innovative indicators covering all six quality dimensions, the following paragraphs compare the final indicator set to the literature.

With respect to safety, not focusing on multiple pregnancies is unexpected (Land and Evers, 2004; Nygren *et al.*, 2011). The inclusion of bleeding, infection and complaints of serious pain, besides ovarian hyperstimulation syndrome, as severe complications of ARTs in a safety indicator is novel (Mourad *et al.*, 2007; Nygren *et al.*, 2011). The identification of lost gametes and embryos as an aspect of

medically assisted reproduction (MAR) safety as a specific mistake or incident is innovative.

Concerning efficiency, the inclusion of an indicator on thorough diagnosis was not surprising (Mourad et al., 2007; Haagen et al., 2010). However, it is unclear what is meant by a thorough diagnosis as for example, the clinical rationale (Meuleman et al., 2009) and cost-effectiveness of diagnostic laparoscopy have been debated (Mol et al., 2001; Moayeri et al., 2009). To label the use of an electronic patient record and staff productivity as indicators of efficiency is in line with, respectively, the finding that electronic health records relate to efficient nursing care for hospitalized patients (Kutney-Lee and Kelly 2011) and that the economic recess requires improving staff productivity (Holland and McIntosh, 2012). Regarding the indicators on the clinic's website, educating patients via the internet is gaining popularity in infertility care (Aarts et al., 2012b); however, its (cost-)efficiency has not been examined.

With respect to effectiveness, referring to baby-take-home rates (Haan et al., 1991) and live birth delivery rates (Zegers-Hochschild, 2009) is not unexpected. However, using the number of treated patients during a certain time period besides the number of initiated cycles, aspirations or embryo transfers as a denominator (Nygren et al., 2011) are novel and serves to include all MAR treatments provided by a clinic in one measure.

Regarding patient centeredness, it is surprising that half of the final indicators relate to emotional support provided by counsellors as these concerns only 1 out of 10 dimensions of patient-centred infertility care (Dancet et al., 2011b, 2012). However, the need for all staff members to address patients from a psychosocial besides a medical perspective during routine care (Boivin et al., 2001; Dancet et al., 2011b) was addressed by the other two indicators.

With respect to timeliness, the final indicators remarkably focus on urgent consultations and waiting times in waiting rooms and for new patients rather than on daily availability of the clinic for treatments (Haagen et al., 2010).

Regarding equity of infertility care, the finally selected indicators focus on the general need for explicit vision and inclusion and exclusion criteria for MAR, sticking to equity guidelines (Kennedy et al., 2006) and respect for patients but do not specifically focus on equity of access for lesbians, single and poor women (Peterson, 2005) or on financial determinants of access to MAR (ESHRE—European Society for Human Reproduction and Embryology—Task Force on Ethics and Law, 2008).

## Critical appraisal of the study methodology

The used methodologies for ranking the quality dimensions and generating and selecting quality indicators had three overall strengths.

First, the international perspective was more adequately represented by the inclusion of respondents from two different European countries with their respective different educational, health care and reimbursement systems (Mainz, 2003b).

Secondly, different but complementary perspectives were aggregated by involving other professionals than only physicians (Mainz, 2003b), who had been previously questioned on these issues (Mourad et al., 2007; Haagen et al., 2010; van Empel et al., 2011). For example, all professional groups, except nurses, exclusively suggested at least one of the finally selected indicators. The fact that

nurses did not exclusively suggest an indicator seems due to the limited size of their stakeholder group and to the less innovative nature of their suggestions. Furthermore, feedback from other stakeholder groups influenced respondents' ratings (Campbell et al., 1999).

Thirdly, this study effectively involved patients in ranking the six quality dimensions and was the first across all health-care fields to prove the feasibility of involving patients in indicator development. Suggestions from patients and suggestions from professionals required similar efforts for transformation into correctly formulated indicators. Suggestions from patients contributed to half of the finally suggested indicators and like all other groups, patients had a unique contribution as they were the only group suggesting one of the finally selected indicators. We hypothesize that highly educated patients are best placed to represent the patients' perspective in Delphi studies as taking part was even challenging for some patients with a University (College) degree. However, the fact that only patients with a University degree were part of the patient group may have introduced some bias. Participation in focus groups, instead of answering Delphi questionnaires, might be a better method to get the patient's perspective from a patient group with a more realistic mix of different educational levels (Grunfeld et al., 2008).

Furthermore, four critical remarks need to be made regarding this study's methodology.

First, although the Delphi-method was successful in reaching consensus on the dimensions' rankings, it may not be as appropriate as discrete choice experiments to incorporate multidimensional influences on health decisions (Wensing and Elwyn, 2002).

Secondly, our development of quality indicators can be criticized because we did not start the Delphi rounds with indicators extracted from the literature (Mainz, 2003b). This was not possible as only few indicators for infertility care have been reported in the literature (Copnell et al., 2009). Furthermore, recommendations from guidelines do not cover all six quality dimensions (Mourad et al., 2007; Haagen et al., 2010). Instead, we used an open and creative process to develop quality indicators directly based on respondents' suggestions, which allowed the identification of many different aspects of quality of care (Campbell et al., 2002). It was not surprising that respondents' suggestions for indicators were not formulated as rate-based indicators (Mainz, 2003a,b), as the literature on infertility quality indicators reports on guideline recommendations instead of indicators (Campbell et al., 2002; Mourad et al., 2007; Haagen et al., 2010).

Thirdly, not all respondents' top priorities for benchmarking could be selected as this would have resulted in the selection of too many indicators. Our aim to reach interpersonal agreement among stakeholders is in line with the aim of consensus techniques (Campbell et al., 2002), but probably does not identify the most novel or original indicators (Bains, 2009). Additionally, maximal support from members from all groups was obtained by choosing the indicators that were consistently best rated across stakeholder groups and reporting cut-offs transparently, as the lack of a golden standard did not allow defining cut-off levels for quality indicators beforehand.

Fourthly, although the sampling strategy was designed to prevent sample bias and self-selection bias (Sica, 2006), the threat of response bias was present as not all respondents took part in all three rounds despite receiving reminders (Sica, 2006). Therefore, the final set might reflect the perspective of the respondents most motivated for quality management (Sica, 2006). However, biased responses were limited by

using coded questionnaires and providing the opportunity to reconsider personal ratings in subsequent rounds.

### Implications for clinical practice

Clinics can use the presented indicator set to assess their multidimensional quality of care in a manner supported by several important stakeholders of infertility care. The indicator set can additionally be used for benchmarking institutions, for targeting and evaluating improvement projects, for supporting accountability, regulations and accreditation and for assisting consumers' choice of providers (Mainz, 2003a,b).

Clinics are advised to involve professionals from different groups and patients in their quality management as this study demonstrated their contribution.

### Implications for research

The development and selection of an indicator set drawing its credibility from expert consensus demonstrated that indicators can be generated for all quality dimensions of infertility care, and that different stakeholder groups can be involved.

Real-life applications of the presented face valid and acceptable quality indicators should be studied by further research (Campbell *et al.*, 2002; Mainz, 2003a; de Bruin-Kooistra *et al.*, 2012). Each indicator should be examined for the following criteria: reliability of data collection, applicability to a clinic's specific patient population, improvement potential, discriminatory capacity, complexity and case-mix stability (Mourad *et al.*, 2008). Regarding the indicator 'the number of live births after a fresh ART cycle with embryo transfer relative to the total number of fresh ART cycles with embryo transfer during a certain time period', the 10th and 90th percentile of the results of 2010 of the Belgian clinics were, respectively, 10.1 and 24.8 (median = 18.0; De Neubourg *et al.*, 2013, submitted for publication). For the other finally selected indicators, ranges of results of Belgian and/or Dutch clinics were not reported (De Neubourg *et al.*, 2013, submitted for publication; [http://www.nvog.nl/Sites/Files/0000002744\\_IVF%20cijfers%20\(centra\)%202010.pdf](http://www.nvog.nl/Sites/Files/0000002744_IVF%20cijfers%20(centra)%202010.pdf)).

This study showed that patients can be involved in deciding how to evaluate health care besides involving them in the evaluation of their health status, which is currently promoted (Barry, 2011; Washington and Lipstein, 2011). Involving additional stakeholder groups in future studies, like insurance companies and society, could be interesting.

In line with the recent studies conceptualizing patient centeredness of infertility care from the patients' perspective (Dancet *et al.*, 2010, 2011a, 2012), it would be interesting to examine the nearly unexplored fertility patients' perspective on the five other quality dimensions, especially effectiveness and safety.

### Conclusion

Fertility clinics are advised to primarily focus their quality management on improving the safety, effectiveness and patient centeredness of their care. The final quality indicator set assesses all quality dimensions of infertility care, covers several novel care aspects compared with the literature, is face valid and acceptable according to five different stakeholder groups, and should be tested in real-life applications.

## Supplementary data

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

## Acknowledgements

This is a Leuven University funded study. All patient associations, patients and professionals taking part in this study are acknowledged for their contribution.

## Authors' roles

E.A.F.D., T.M.D., W.L.D.M.N., J.A.M.K., C.S. and W.S. designed the study. E.A.F.D. and N.K. collected the data. Data were analysed by E.A.F.D., N.K., W.L.D.M.N., D.D. and C.S. and interpreted by all authors. E.A.F.D. drafted the article. All authors critically revised the article and approved the version to be published.

## Funding

This study received funding from KU. Leuven (mecenate grant).

## Conflict of interest

The authors did not have conflicts of interests.

## References

- Aarts JW, Huppelschoten AG, van Empel IW, Boivin J, Verhaak CM, Kremer JA, Nelen WL. How patient-centred care relates to patients' quality of life and distress: a study in 427 women experiencing infertility. *Hum Reprod* 2012a;**27**:488–495.
- Aarts JW, van den Haak P, Nelen WL, Tuil WS, Faber MJ, Kremer JA. Patient-focused internet interventions in reproductive medicine: a scoping review. *Hum Reprod Update* 2012b;**18**:211–227.
- Bains W. Leadership and innovation. How consensus management blocks genuine innovation. *Biosci Hypotheses* 2009;**2**:277–281.
- Barry MJ. Helping patients make better personal health decisions: the promise of patient-centered outcomes research. *JAMA* 2011;**306**:1258–1259.
- Boivin J, Appleton TC, Baetens P, Baron J, Bitzer J, Corrigan E, Daniels KR, Darwish J, Guerra-Diaz D, Hammar M *et al.*; European Society of Human Reproduction and Embryology. Guidelines for counselling in infertility: outline version. *Hum Reprod* 2001;**16**:1301–1304.
- Campbell SM, Hann M, Roland MO, Quayle JA, Shekelle PG. The effect of panel membership and feedback on ratings in a two-round Delphi survey: results of a randomized controlled trial. *Med Care* 1999;**37**:964–968.
- Campbell SM, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002;**11**:358–364.
- Campbell SM, Shield T, Rogers A, Gask L. How do stakeholder groups vary in a Delphi technique about primary mental health care and what factors influence their ratings? *Qual Saf Health Care* 2004;**13**:428–434.
- Chambers GM, Sullivan EA, Ishihara O, Chapman MG, Adamson GD. The economic impact of assisted reproductive technology: a review of selected developed countries. *Fertil Steril* 2009;**91**:2281–2294.



- Chan CM, Ahmad WA. Differences in physician attitudes towards patient-centredness: across four medical specialties. *Int J Clin Pract* 2012;**66**:16–20.
- Chervenak FA, McCullough LB. Clinical guides to preventing ethical conflicts between pregnant women and their physicians. *Am J Obstet Gynecol* 1990;**162**:303–307.
- Copnell B, Hagger V, Wilson SG, Evans SM, Sprivilis PC, Cameron PA. Measuring the quality of hospital care: an inventory of indicators. *Intern Med J* 2009;**39**:352–360.
- Corrigan JM, Donaldson MS, Kohn LT, Maguire SK, Pike KC. *Crossing the Quality Chasm. A New Health System for the 21st Century*. Washington: Institute of Medicine, National Academy of Sciences, National Academy Press, 2001.
- Dancet EA, Nelen WL, Sermeus W, De Leeuw L, Kremer JA, D'Hooghe TM. The patients' perspective on fertility care: a systematic review. *Hum Reprod Update* 2010;**16**:467–487.
- Dancet EA, Ameye L, Sermeus W, Welkenhuysen M, Nelen WL, Tully L, De Bie B, Veit J, Vedsted-Hansen H, Zondervan KT et al. The ENDOCARE questionnaire (ECQ): a valid and reliable instrument to measure the patient-centeredness of endometriosis care in Europe. *Hum Reprod* 2011a;**26**:2988–2999.
- Dancet EA, Van Empel IW, Rober P, Nelen WL, Kremer JA, D'Hooghe TM. Patient-centred infertility care: a qualitative study to listen to the patient's voice. *Hum Reprod* 2011b;**26**:827–833.
- Dancet EA, D'Hooghe TM, Sermeus W, van Empel I, Strohmer H, Wyns C, Santa-Cruz D, Nardo LG, Kovatchki D, Vanlangenakker L et al. Patients from across Europe have similar views on patient-centred care: an international multilingual qualitative study in infertility care. *Hum Reprod* 2012;**27**:1702–1711.
- de Bruin-Kooistra M, Amelink-Verburg MP, Buitendijk SE, Westert GP. Finding the right indicators for assessing quality midwifery care. *Int J Qual Health Care* 2012;**24**:301–310.
- De Neubourg D, van Duijnhoven NT, Nelen WL, D'Hooghe TM. Dutch translation of the ICMART-WHO revised glossary on ART terminology. *Gynecol Obstet Invest* 2012;**74**:233–248.
- Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988;**260**:1743–1748.
- ESHRE Task Force on Ethics and Law; Pennings G, de Wert G, Shenfield F, Cohen J, Tarlatzis B, Devroey P. ESHRE Task Force on Ethics and Law 14: equity of access to assisted reproductive technology. *Hum Reprod* 2008;**23**:772–774.
- Gerris J. Invited lecture: 'Evidence based safety in ART: principles, paradigms and practice'. In: *ESHRE Annual Meeting*, Stockholm, 6 July 2011.
- Griffiths A, Dyer SM, Lord SJ, Pardy C, Fraser IS, Eckermann S. A cost-effectiveness analysis of in-vitro fertilization by maternal age and number of treatment attempts. *Hum Reprod* 2010;**25**:924–931.
- Grunfeld E, Urquhart R, Mykhalovskiy E, Folkes A, Johnston G, Burge FI, Earle CC, Dent S. Toward population-based indicators of quality end-of-life care: testing stakeholder agreement. *Cancer* 2008;**112**:2301–2308.
- Haagen EC, Nelen WL, Grol RP, Braat DD, Hermens RP, Kremer JA. Variation in guideline adherence in intrauterine insemination care. *Reprod Biomed Online* 2010;**20**:533–542.
- Haan G. Effects and costs of in-vitro fertilization. Again, let's be honest. *Int J Technol Assess Health Care* 1991;**7**:585–593.
- Haan G, Bernardus RE, Hollanders JM, Leerentveld RA, Prak FM, Naaktgeboren N. Results of IVF from a prospective multicentre study. *Hum Reprod* 1991;**6**:805–810.
- Hermens RP, Ouwens MM, Vonk-Okhuijsen SY, van der Wel Y, Tjan-Heijnen VC, van den Broek LD, Ho VK, Janssen-Heijnen ML, Groen HJ, Grol RP et al. Development of quality indicators for diagnosis and treatment of patients with non-small cell lung cancer: a first step toward implementing a multidisciplinary, evidence-based guideline. *Lung Cancer* 2006;**54**:117–124.
- Hinton L, Kurinczuk JJ, Ziebland S. Infertility; isolation and the Internet: a qualitative interview study. *Patient Educ Couns* 2010;**81**:436–441.
- Holland A, McIntosh B. Optimising productivity, quality and efficiency in community nursing. *Br J Community Nurs* 2012;**17**:390–392.
- Jeacocke D, Heller R, Smith J, Anthony D, Williams JS, Dugdale A. Combining quantitative and qualitative research to engage stakeholders in developing quality indicators in general practice. *Aust Health Rev* 2002;**25**:12–18.
- Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;**311**:376–380.
- Kennedy R, Kingsland C, Rutherford A, Hamilton M, Ledger WW; British Fertility Society. Implementation of the NICE guideline—recommendations from the British Fertility Society for national criteria for NHS funding of assisted conception. *Hum Fertil (Camb)* 2006;**9**:181–189.
- Kutney-Lee A, Kelly D. The effect of hospital electronic health record adoption on nurse-assessed quality of care and patient safety. *J Nurs Adm* 2011;**41**:466–472.
- Land JA, Evers JL. What is the most relevant standard of success in assisted reproduction? Defining outcome in ART: a Gordian knot of safety, efficacy and quality. *Hum Reprod* 2004;**19**:1046–1048.
- Lawrence M, Olesen F. Indicators of quality of health care. *Eur J Gen Pract* 1997;**3**:103–108.
- Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care* 2003a;**15**:523–530.
- Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. *Int J Qual Health Care* 2003b;**15** (Suppl. 1):i5–i11.
- Meuleman C, Vandenabeele B, Fieuws S, Spiessens C, Timmerman D, D'Hooghe T. High prevalence of endometriosis in infertile women with normal ovulation and normospermic partners. *Fertil Steril* 2009;**92**:68–74.
- Min JK, Breheny SA, MacLachlan V, Healy DL. What is the most relevant standard of success in assisted reproduction? The singleton, term gestation, live birth rate per cycle initiated: the BESST endpoint for assisted reproduction. *Hum Reprod* 2004;**19**:3–7.
- Moayeri SE, Lee HC, Lathi RB, Westphal LM, Milki AA, Garber AM. Laparoscopy in women with unexplained infertility: a cost-effectiveness analysis. *Fertil Steril* 2009;**92**:471–480 (Epub 2008 Aug 22).
- Mol BW, Collins JA, Van Der Veen F, Bossuyt PM. Cost-effectiveness of hysterosalpingography, laparoscopy, and Chlamydia antibody testing in subfertile couples. *Fertil Steril* 2001;**75**:571–580.
- Mourad SM, Hermens RP, Nelen WL, Braat DD, Grol RP, Kremer JA. Guideline-based development of quality indicators for subfertility care. *Hum Reprod* 2007;**22**:2665–2672.
- Mourad SM, Nelen WL, Hermens RP, Bancsi LF, Braat DD, Zielhuis GA, Grol RP, Kremer JA. Variation in subfertility care measured by guideline-based performance indicators. *Hum Reprod* 2008;**23**:2493–2500.
- Musters AM, de Bekker-Grob EW, Mochtar MH, van der Veen F, van Mello NM. Women's perspectives regarding subcutaneous injections, costs and live birth rates in IVF. *Hum Reprod* 2011;**26**:2425–2431.
- Nelen WL, Hermens RP, Mourad SM, Haagen EC, Grol RP, Kremer JA. Monitoring reproductive health in Europe: what are the best indicators of reproductive health? A need for evidence-based quality indicators of reproductive health care. *Hum Reprod* 2007;**22**:916–918.
- Neumann PJ, Gharib SD, Weinstein MC. The cost of a successful delivery with in vitro fertilization. *N Engl J Med* 1994;**331**:239–243.
- Nygren KG, Sullivan E, Zegers-Hochschild F, Mansour R, Ishihara O, Adamson GD, de Mouzon J. International Committee for Monitoring Assisted Reproductive Technology (ICMART) world report: assisted reproductive technology 2003. *Fertil Steril* 2011;**95**:2209–2222, 2222.e1–17.

- Palumbo A, De La Fuente P, Rodríguez M, Sánchez F, Martínez-Salazar J, Muñoz M, Marqueta J, Hernández J, Espallardo O, Polanco C *et al*. Willingness to pay and conjoint analysis to determine women's preferences for ovarian stimulating hormones in the treatment of infertility in Spain. *Hum Reprod* 2011;**26**:1790–1798.
- Peterson MM. Assisted reproductive technologies and equity of access issues. *J Med Ethics* 2005;**31**:280–285.
- Polit DF, Beck CT. *Nursing Research. Generating and Assessing Evidence for Nursing Practice*, 8th edn. Philadelphia: Lippincott Williams & Wilkins, 2008, pp. 758.
- Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Soc Sci Med* 1999;**48**:535–546.
- Schatten GP. Safeguarding ART. *Nature Cell Biol* 2002;**4**:19–22.
- Scotland GS, McNamee P, Peddie VL, Bhattacharya S. Safety versus success in elective single embryo transfer: women's preferences for outcomes of in vitro fertilisation. *BJOG* 2007;**114**:977–983.
- Shield T, Campbell S, Rogers A, Worrall A, Chew-Graham C, Gask L. Quality indicators for primary care mental health services. *Qual Saf Health Care* 2003;**12**:100–106.
- Sica GT. Bias in research studies. *Radiology* 2006;**238**:780–789.
- Temmerman M, Foster LB, Hannaford P, Cattaneo A, Olsen J, Bloemenkamp KW, Jahn A, da Silva MO. Reproductive health indicators in the European Union: The REPROSTAT project. *Eur J Obstet Gynecol Reprod Biol* 2006;**126**:3–10.
- Van den Broeck U, Spiessens C, Dancet E, Bakelants E, Vrancken A, Demyttenaere K, Enzlin P, D'Hooghe T. Patient evaluation of infertility management in an ISO 9001:2008-certified centre for reproductive medicine. *Reprod Biomed Online* 2012;**24**:293–300.
- van Empel IW, Aarts JW, Cohlen BJ, Huppelschoten DA, Laven JS, Nelen WL, Kremer JA. Measuring patient-centredness, the neglected outcome in fertility care: a random multicentre validation study. *Hum Reprod* 2010;**25**:2516–2526.
- van Empel IW, Dancet EA, Koolman XH, Nelen WL, Stolk EA, Sermeus W, D'Hooghe TM, Kremer JA. Physicians underestimate the importance of patient-centredness to patients: a discrete choice experiment in fertility care. *Hum Reprod* 2011;**26**:584–593.
- Verhaak CM, Smeenk JM, Evers AW, Kremer JA, Kraaimaat FW, Braat DD. Women's emotional adjustment to IVF: a systematic review of 25 years of research. *Hum Reprod Update* 2007;**13**:27–36.
- Washington AE, Lipstein SH. The patient-centered outcomes research institute—promoting better information, decisions, and health. *N Engl J Med* 2011;**365**:e31.
- Wensing M, Elwyn G. Research on patients' views in the evaluation and improvement of quality of care. *Qual Saf Health Care* 2002;**11**:153–157.
- Willemens D, D'Hooghe T, Knoop I, De Neubourg D, Spiessens C. Does the European Union tissues and cells directive improve quality in the in vitro fertilization laboratory? A case study in a tertiary referral center. *Semin Reprod Med* 2012;**30**:191–198.
- Winston RM, Hardy K. Are we ignoring potential dangers of IVF and related treatments? *Nature Cell Biol* 2002;**4**:14–18.
- Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, Sullivan E, van der Poel S; International Committee for Monitoring Assisted Reproductive Technology; World Health Organization. The International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) Revised Glossary on ART Terminology, 2009. *Hum Reprod* 2009;**24**:2683–2687.