Human Reproduction, Vol.26, No.10 pp. 2763–2776, 2011
Advanced Access publication on August 1, 2011 doi:10.1093/humrep/der246

human reproduction

META-ANALYSIS ***Psychology and counselling***

Stress, distress and outcome of assisted reproductive technology (ART): a meta-analysis

S.M.S. Matthiesen[1,*], Y. Frederiksen[1], H.J. Ingerslev[2], and R. Zachariae[3]

[1]Department of Psychology, University of Aarhus, Jens Chr. Skous vej 4, 8000 Aarhus, Denmark [2]The Fertility Clinic, Aarhus University Hospital, 8200 Aarhus, Denmark [3]Department of Oncology, Aarhus University Hospital, 8000 Aarhus, Denmark

*Correspondence address. E-mail: signese@psy.au.dk

Submitted on May 12, 2010; resubmitted on May 12, 2011; accepted on June 29, 2011

BACKGROUND: A number of studies have investigated the relationship between psychological factors such as stress and distress (measured as anxiety and depression) and outcomes of assisted reproductive technology (ART). The results, however, are inconsistent, and the strength of any associations remains to be clarified. We conducted a systematic review and meta-analysis of the results of studies reporting on the associations between stress, anxiety, and depression and ART outcomes.

METHODS: Prospective studies reporting data on associations between stress or distress in female patients and ART outcome were identified and evaluated by two independent researchers according to an *a priori* developed codebook. Authors were contacted in cases of insufficient data reporting. Stress was defined as perceived stress, work-related stress, minor life events or major life events, and distress was defined as anxiety or depression.

RESULTS: A total of 31 prospective studies were included. Small, statistically significant, pooled effect sizes were found for stress [ESr, effect size correlation) = −0.08; $P = 0.02$, 95% confidence interval (CI): −0.15, −0.01], trait anxiety (ESr = −0.14; $P = 0.02$, 95% CI: −0.25, −0.03) and state anxiety (ESr = −0.10, $P = 0.03$, 95% CI: −0.19, −0.01), indicating negative associations with clinical pregnancy rates. A non-significant trend (Esr = −0.11, $P = 0.06$) was found for an association between depression and clinical pregnancy. For serum pregnancy tests and live birth rates, associations between trait anxiety or state anxiety were not significant. The fail safe number did not exceed the suggested criterion in any analyses, between-study heterogeneity was considerable and the mean age, mean duration of infertility and percentage of first time ART attenders in the study samples were found to moderate several of the associations.

CONCLUSIONS: Small but significant associations were found between stress and distress and reduced pregnancy chances with ART. However, there were a limited number of studies and considerable between-study heterogeneity. Taken together, the influence of stress and distress on ART outcome may appear somewhat limited.

Key words: anxiety / depression / stress / infertility / ART

Introduction

Problems of infertility seem to be increasing in the western countries (Swan *et al.*, 2000; Skakkebaek *et al.*, 2006). Couples suffering fertility problems often turn to medically assisted reproduction (MAR), and the number of initiated treatments with assisted reproductive technology (ART) performed at public and private fertility clinics is rising. Studies conducted over the past two decades suggest that psychological stress and feelings of distress may reduce the chances of obtaining a viable pregnancy with IVF treatment (Demyttenaere *et al.*, 1992, 1994; Thiering *et al.*, 1993; Csemiczky *et al.*, 2000; Kee *et al.*, 2000; Smeenk *et al.*, 2001, 2005; Klonoff-Cohen, 2005). More specifically, several studies have shown that stress related to infertility, to participating in an ART treatment program, to relationship difficulties and to other environmental stressors is associated with reduced chances of achieving a positive ART treatment outcome (Stoleru *et al.*, 1997; Klonoff-Cohen *et al.*, 2001; Verhaak *et al.*, 2001; Klonoff-Cohen and Natarajan, 2004; Boivin and Schmidt, 2005; Barzilai-Pesach *et al.*, 2006; Ebbesen *et al.*, 2009). Other studies have explored the influence of anxiety and depression on ART outcome (Klonoff-Cohen, 2005), with several of these having found anxiety and depression to be associated with poorer outcomes of ART (Klonoff-Cohen, 2005). The

magnitude of a possible association between stress or distress and ART outcome, however, remains unclear. In contrast to the studies cited earlier, there are also several studies that have been unable to document such associations, thereby questioning whether there is a reliable effect of stress and distress on pregnancy chances (Milad et al., 1998; Lovely et al., 2003; Anderheim et al., 2005; de Klerk et al., 2008).

Studies of the possible impact of psychological factors on ART outcomes show considerable heterogeneity (Klonoff-Cohen, 2005). Inspection of the existing studies show variability in sample characteristics in terms of mean age (e.g. Demyttenaere et al., 1992, 1998; Klonoff-Cohen and Natarajan, 2004; Panagopoulou et al., 2006; Sohrabvand et al., 2008), in the proportions of ART inductees versus veterans (e.g. Merari et al., 1992; Anderheim et al., 2005; de Klerk et al., 2008), in the duration of infertility (e.g. Lancastle and Boivin, 2005; Panagopoulou et al., 2006; Karlidere et al., 2008; Sohrabvand et al., 2008; Ebbesen et al., 2009), in infertility etiology (e.g. Klonoff-Cohen et al., 2001; Smeenk et al., 2001; Barzilai-Pesach et al., 2006) and in the timing of stress or distress assessment, e.g. before ART treatment (Thiering et al., 1993; Visser et al., 1994; Boivin and Takefman, 1995; Anderheim et al., 2005), when enrolling in treatment (Sanders and Bruce, 1999; Klonoff-Cohen et al., 2001; de Klerk et al., 2008) or during treatment (Milad et al., 1998; Lovely et al., 2003). The first three variables are potentially interrelated, as they all may be associated with timing and therefore related to both psychological changes and changes in reproductive abilities. The mentioned between-study differences constitute a considerable source of heterogeneity, which may cloud possible associations between stress, distress and ART outcome.

Clarifications in this area of research are clearly needed, as both the associations between these psychological factors and ART outcomes as well as their magnitude are still unclear. This would be of potential importance both for decision-making policies for the provision of intervention programs for ART treatment-seeking couples and for the design of intervention studies with sufficient statistical power to assess the effects of stress reduction on ART outcome. In line with this view, it has been suggested that further research should focus on clarifying the existence and direction of a causal relationship between distress and pregnancy chances (Merari et al., 1992). A quantitative systematic review of the available prospective studies could provide a valuable test of the hypothesis by allowing the evaluation of the combined effects, as well as providing an estimate of the magnitude of any effects found and is therefore timely. The present study used this approach to address the question of whether stress and distress in women trying to conceive through ART treatment have a statistically significant impact on the overall chances of achieving a viable pregnancy with ART treatment. If such associations were found, a second aim was to provide estimates of the effect sizes. Additional moderator analyses were planned to address possible reasons for differences in effect sizes between studies in case of significant between-study heterogeneity.

Materials and Methods

Study eligibility criteria

To be included, studies had to be original prospective, empirical studies of female patients in ART treatment. ART refers to all treatments or procedures that include in vitro handling of both human oocytes and sperm (or embryos) for the purpose of establishing a pregnancy. This includes, but is not limited to, IVF and embryo transfer, gamete intrafallopian transfer, zygote intrafallopian transfer, tubal embryo transfer, gamete and embryo cryopreservation, oocyte and embryo donation and gestational surrogacy ART, but does not include assisted insemination (artificial insemination) using sperm from either a woman's partner or a sperm donor (Zegers-Hochschild et al., 2009). The studies had to focus on ART outcomes assessed as number of oocytes harvested, embryos transferred, fertilization rates, implantation rates, serum pregnancy test result, clinical pregnancy or live birth delivery. Studies had to report data on the association between stress or distress and ART outcome, with stress defined as perceived stress, infertility-related stress, work-related stress, minor life events (e.g. daily hassles with partner) or major life events (e.g. death of a family member) and distress defined as anxiety or depression. Results reported as composite measures were excluded due to the likelihood of confounding the latent variable under investigation, e.g. anxiety. A composite measure refers to a measure generated by collapsing two or more psychometric measures, e.g. deriving a 'negative affect measure' from combined scores of anxiety and other negative emotions. Studies had to be published in English language peer-reviewed journals accessible from databases such as PubMed, PsychINFO and Web of Science. Not included were articles reporting results of: non-prospective studies (cross-sectional studies and case–control studies), as this type of design is considered inappropriate to assess an impact of stress and distress on pregnancy chances; psychosocial intervention studies or studies confounded by such intervention programs; studies measuring effects of stress on male fertility indicators (e.g. sperm quality) and experimental studies measuring biological stress responses (e.g. heart rate). Letters to editors, dissertations, abstracts and conference papers were also excluded. There were no restrictions as to the publication date, but no studies were before 1978, as this was the year of the birth of the first baby born as result of ART.

Independent variables

Stress

The stress process, according to the theoretical framework proposed by Cohen et al. (1995), consists of three steps: (i) a stressor or environmental demand, e.g. a life event or a series of life events, followed by (ii) a set of appraisals and the subsequent perception of stress, which then may lead to (iii) affective, behavioral and/or biological stress responses, e.g. distress. The terms stress and distress are often confused and reported results for these psychological states are commonly referred to altogether as 'stress research'. Although part of the same process, stress and distress are different phenomena, which are distinguished by the presence of an environmental stressor. Environmental stressors and the related psychological stress may thus result in distress, e.g. anxiety or depression, but in contrast to feelings of distress, which may originate from several different sources, all stress perceptions and reactions have an external stressor in common. Stress measures in the studies to be included in this meta-analysis had to concur with the acknowledged definitions of stress (Cohen et al., 1997), and could include either measures of perceived stress, e.g. the Perceived Stress Scale (PSS-10) (Cohen and Williamson, 1988), or measures of environmental stressors generally believed to be able to induce stress, including scales measuring the number of stressful life events, e.g. the list of recent events (LRE) (Henderson et al., 1981), or specific stressors, e.g. occupational or marital stress.

Anxiety

Anxiety is an emotion described by a subjective feeling of tension, apprehension, nervousness and worry, and by activation or arousal by the

autonomic nervous system (Spielberger *et al.*, 1977). Anxiety may occur as a transitory state (i.e. state anxiety) or as a more stable, enduring disposition of anxiety-proneness (i.e. trait anxiety). Individuals high on trait anxiety are considered more prone to react to their environment with state anxiety (Spielberger *et al.*, 1977). To be included, studies had to report results for one or more measure of state and/or trait anxiety, e.g. the commonly used State-Trait Anxiety Inventory (STAI) (Spielberger *et al.*, 1977).

Depression

The core symptom of depression is sadness, with other emotional symptoms, including a negative view of the self, hopelessness and lack of motivation. Behavioral and somatic depressive symptoms include loss of appetite, weight loss, loss of sexual interest and sleep disturbances (Rosenhan and Seligman, 1995). To be included, studies had to report on either a continuous measure of the number and degree of depressive symptoms, e.g. Beck's Depression Inventory (Beck *et al.*, 1996), or as a categorical variable, e.g. using a cut off score on a continuous measure of depression. Other instruments (e.g. the Profile of Moods Questionnaire) aim to measure a more fluctuating depressed mood state, which does not translate into clinical depression versus non-depression.

Dependent variables

To be included, articles had to report results concerning one or more reproductive end-points, the most common being serum pregnancy tests, clinical pregnancy or live birth outcome but also included were the number and quality of oocytes harvested and the number and quality of embryos transferred, fertilization rates and implantation rates. A positive 'serum pregnancy test' is a pregnancy diagnosed only by the detection of HCG in serum or urine (commonly defined by a serum HCG-level of >20 IU 2 weeks following embryo transfer) that does not necessarily develop into a clinical pregnancy. A 'clinical pregnancy' refers to a pregnancy diagnosed by ultrasonographic visualization of one or more gestational sacs or definite clinical signs of pregnancy. It includes ectopic pregnancy. 'Live birth' is defined in the literature as a birth in which a fetus is delivered with signs of life after complete expulsion or extraction from its mother, beyond 20 completed weeks of gestational age (live births are counted as birth events, e.g. a twin or triplet live birth is counted as one birth event) (Zegers-Hochschild *et al.*, 2009). Where no information on pregnancy type was included, it was assumed that pregnancy result referred to clinical pregnancy.

Literature search

A comprehensive forward search for articles published from 1978 to present was conducted in spring/summer 2009 and replicated in January 2010, using the databases PsychINFO (search period: 22/04/09–26/01/10), PubMed (search period: 10/07/09–22/01/10) and Web of Science (search period: 04/06/09–25/01/10), and keywords were chosen on the basis of preliminary exploratory literature search and keywords listed below abstracts in published studies. The following keywords were used in combinations. Stress terms and synonyms included 'stress, occupational stress, stressful life events, major life events' and 'stressor'. Distress terms and synonyms were 'distress, anxiety and depression'. ART terms were 'IVF and ARTs'. General outcome terms included 'pregnancy, IVF outcome, live birth, miscarriage, spontaneous abortion', while the specific outcome terms used were 'oocyte, egg, embryo, fertilization and implantation'. In these literature searches, a limit was set to 700 hits, which was considered manageable. Literature searches exceeding this limit were specified with additional (general outcome) keywords to limit cases to an acceptable number for review (i.e.$<$700). Non-restricted literature searches using different combinations of one or more stress terms or one or more distress terms together with ART and general outcome terms resulted in a total of 859 hits (PubMed), 873 hits (Web of Science) and 138 hits (PsychINFO). Evaluation of titles and abstracts from literature searches led to a total of 35 articles being printed for further reading. A backward literature search on the basis of references from already reviewed studies resulted in additional 26 studies selected for evaluation (total $n = 61$). A number of control searches were conducted with fewer combined key words, which resulted in a large number of hits. This, however, did not result in additional target articles. A second literature search was conducted in order to identify studies investigating the effects of stress and distress on more specific reproductive end-points using the stress and distress terms, together with ART terms and specific outcome terms. No studies emerged from these literature searches that had not already been identified in the primary searches.

Coding procedures

A codebook was developed (which can be obtained upon request from the first author), and used by two independent researchers to review the identified articles according to the specified inclusion/exclusion criteria. Articles selected for review on the basis of titles and abstracts during the literature search were examined in full-text and evaluated according to the codebook. The two coders consisted of the first (S.M.S.M.) and second author (Y.F.). In cases of disagreement, the fourth author (R.Z.) was consulted and a final coding was agreed upon.

Calculating effect sizes

Effect sizes were computed for each study. In most articles, the results were presented as mean scores with standard deviations (or standard error of measurement) of stress or distress in the groups of pregnant and non-pregnant women. In some studies, stress or distress scores had been dichotomized (e.g. as depressed versus non-depressed women) and frequencies were reported for pregnant and non-pregnant women. We contacted, by e-mail, 10 authors who had reported insufficient data for effect size calculation, asking them to provide these data, and three authors who had reported only data from multivariate analyses, asking them to provide raw data. We received replies from six of the thirteen contacted authors, and three of these were able to provide the necessary information. For one set of authors who had reported insufficient data in their paper, we were unable to retrieve contact information. If the authors could not be contacted, did not respond or if this information was no longer available to them, we attempted to estimate the effect sizes using the available statistical information. In cases where data for one or more results were presented as 'non-significant' without further data, and the contacted authors did not respond or were unable to provide the necessary data, the effect size was conservatively set at 0.0. This was done in six cases. In the three cases, where data from multivariate analyses were reported and the contacted authors were unable to provide the basic statistics, effect sizes were estimated on the basis of adjusted data, e.g. relative risks, odds ratios or beta statistics from logistic or linear regressions.

The effect sizes calculated for each study represent the magnitude of the association between stress or distress and ART outcome. Calculations of effect sizes were done independently by the first and second author, and any differences in results were discussed with the fourth author until agreement was reached. Finally, effect sizes for each study were combined to represent a global effect size for the dependent variable in question. The effect size correlation coefficient (ESr) was used (Rosenthal and Rubin, 2003). The ESr can be interpreted as a correlation coefficient equivalent to Pearson's r with values between -1.00 and 1.00. A negative effect size refers to the hypothesized direction of the association tested, i.e. a negative association between stress or distress and ART outcome, e.g. a high level of stress associated with a reduction in the chance of

pregnancy. In the present review, a negative ESr thus indicates a result confirming the main hypothesis, i.e. an adverse effect of stress and distress on ART outcome.

A calculation of an overall effect size across all measured ART outcomes was not considered meaningful, as these measures represent different stages in the same overall outcome. Pooled effect sizes were therefore calculated for each of the reproductive end-points, for which we considered that there were sufficient studies to conduct meta-analysis (>2 independent results). The ART outcomes examined were serum pregnancy test rates, clinical pregnancy rates and live birth rates.

Independency of results

From articles reporting results for more than one outcome, e.g. both number of oocytes retrieved and number of clinical pregnancies, all results were used in the individual analyses (models) regarding the particular outcomes; however, only one result per study was allowed in each model. In studies providing data on more than one measure of the same independent variable, e.g. anxiety, data for the most construct-relevant measure were used, e.g. STAI anxiety measure over the Profile of Mood States (POMS). Where several types of the same stressor were investigated, the one thematically closest to the source of stress was chosen over the other (e.g. stress related to infertility or specific procedures of ART treatment rather than costs of treatment). For instance, for infertility-related stress, results for 'worry about fertility treatment' were chosen over 'worry about missing work'. In studies reporting multiple results for different stressors in the same sample, e.g. perceived stress and number of stressful life events, a weighted mean ES(r) was calculated and used in analyses.

Combining effect sizes

Effect sizes were combined to test for a statistically significant negative association between stress, anxiety and depression for each of the chosen ART outcomes using a fixed or random-model approach, depending on whether studies showed signs of heterogeneity or not (see later text). Three studies were considered a minimum for conducting a meta-analysis for a given association. If studies appeared heterogeneous, we investigated possible sources of between-study differences in effect sizes. Statistical analyses were conducted using Comprehensive Meta-analysis (Bohrenstein and Rothstein, 2006).

Assessing between-study heterogeneity

Since studies on the possible impact of psychological factors on ART outcomes show considerable heterogeneity (Klonoff-Cohen, 2005), a common effect size was considered less likely. Dispersion is likely to reflect true differences in effect sizes across studies (Borenstein *et al.*, 2009). Formal tests of heterogeneity were therefore conducted by calculating Q (a χ^2 statistic used to quantify levels of heterogeneity), and in cases of a significant Q-value, a random effects model was used when combining effect sizes from primary studies. Due to the low power of this test, the alpha level was set at 0.10, as previously suggested (Borenstein *et al.*, 2009).

Quality assessment

The use of assigning a quality score to each study and using this score to weight the results in meta-analysis is highly debated (Detsky *et al.*, 1992; Kunz and Oxman, 1998; Juni *et al.*, 1999; Greenland and O'Rourke, 2001), and we therefore chose not to use such a scoring procedure. Instead, we have addressed relevant quality factors related to study designs and sampling methods by investigating whether they moderate the associations found between stress, distress measures and ART outcomes.

Moderator analyses

In addition to testing the overall associations between the independent variables of anxiety, depression and stress and the various outcomes of ART, we planned to explore the following possible moderators of these associations, in case of a significant heterogeneity test: (i) mean age of the study sample, (ii) percentage of first time ART attendees, (iii) mean duration of infertility, (iv) percentage of female infertility factor (i.e. where the female part of the couple is the one who is infertile) and (v) timing of stress/distress measurement. Timing of assessment was coded as follows: 'Before treatment' refers to stress or distress being measured within a period of up to 3 months before enrolling in an ART treatment cycle. 'Enrollment' refers to stress or distress measured at the time of the patients' enrollment in ART treatment. If no information was given for the time of baseline measurement, this was coded as enrollment, as this is by far the most common time of measurement. 'Procedural' refers to stress or distress measured after the commencement of ART treatment, and this may be at the time of oocyte retrieval or embryo transfer or at any other time during a treatment cycle. If a study reported results based on procedural stress measured on several occasions after treatment start, a mean effect size was calculated. The possible role of continuous moderators, i.e. mean age, percentage first-time attenders, mean duration of infertility and percentage female infertility factor, was analyzed using meta-regression, while categorical moderators, i.e. timing of assessment, was analyzed with between-group meta-analysis of variances (ANOVAs).

Publication bias

All contacted authors were encouraged to come forth with any published or unpublished results regarding the research question of interest. Finally, for each analysis, we calculated the fail-safe number, which addresses the possibility of a 'file drawer problem' by referring to the minimum number of unpublished papers reporting null results that would lead to a different conclusion in the meta-analysis.

In order to ensure high quality reporting of our results, this paper adheres to the PRISMA statement (Liberati *et al.*, 2009), an updated version of the QUOROM statement from 1999 (Moher *et al.*, 1999).

Results

Study characteristics

For a description of the study selection procedure, see Fig. 1. A total of 34 articles were found eligible for further analysis, fulfilling all of the inclusion criteria. See Table I for an overview of the studies within the articles. A total of 25 effect sizes from 14 articles were then excluded from the data analysis due to an inability to establish the effect direction, the use of composite stress or distress measures or an insufficient number of independent studies (<3) regarding the specific outcome, e.g. number of oocytes retrieved (Smeenk *et al.*, 2001; Klonoff-Cohen *et al.*, 2001; Klonoff-Cohen and Natarajan, 2004; Lancastle and Boivin, 2005; Ebbesen *et al.*, 2009), implantation rate (Gallinelli *et al.*, 2001), fertilization rate (Stoleru *et al.*, 1997; Klonoff-Cohen *et al.*, 2001; Klonoff-Cohen and Natarajan, 2004) and transfer rates (Klonoff-Cohen *et al.*, 2001; Klonoff-Cohen and Natarajan, 2004). The excluded studies are listed in Supplementary data, Table SI. Because other studies in these articles were suitable for final inclusion, only three articles were excluded in this step (Fig. 1).

The final 31 articles found suitable for meta-analysis had investigated a total of 4902 participants with an average sample size of 158

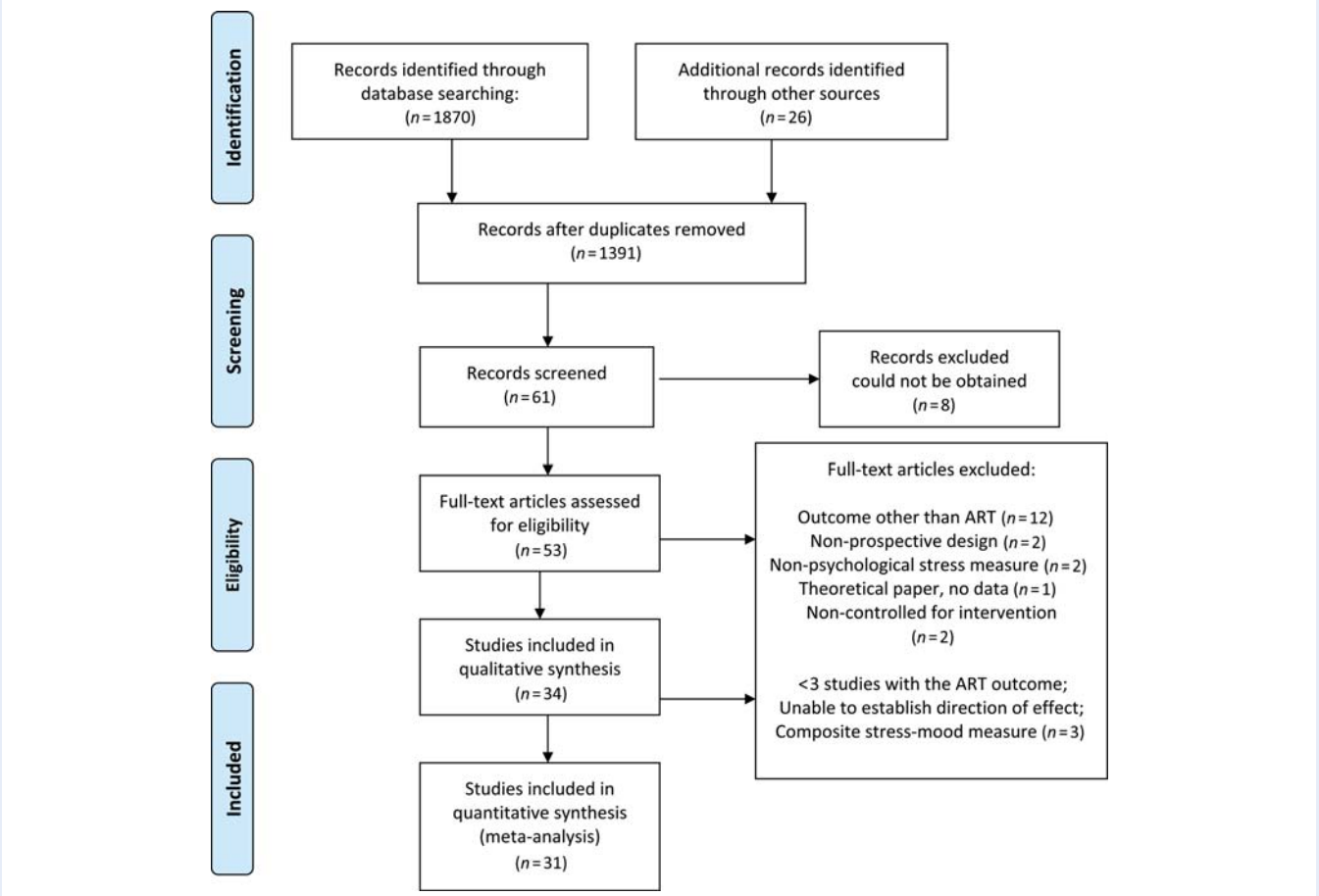

Figure 1 Flow diagram for study selection.

Table I Investigated associations between stress, distress and reproductive end-points—studies published between 1992 and 2009.

Outcomes	Independent variables				
	State anxiety	Trait anxiety	Depressive symptoms	Stress[a]	Total[b]
Rate of positive serum pregnancy test	5	3	2	1	11
Clinical pregnancy rate	17	9	16	6	48
No. of oocytes	1	1	1	3	6
Fertilization rate	2	1	1	1	5
Implantation rate	1	1			2
Transfer rate			1	1	2
Live birth rate	4	1	1	2	8
Total	30	16	22	14	82

[a]Stress includes: occupational stress, major life events, infertility related or perceived stress.
[b]The numbers exceed the total number of articles, as they may provide results for more than one independent and dependent variable.

participants [range: 22 (Csemiczky *et al.*, 2000) to 791 (Ebbesen *et al.*, 2009)]. The characteristics of each study can be seen in Table II (and Supplementary data, Table SII), with the numbers referring to the actual number of participants included in the various analyses.

The majority of studies concerned the investigation of an association between clinical pregnancy rates and depression or state anxiety, whereas pregnancy rates associated with trait anxiety and stress were less frequently examined. The dependent variable most

Table II Studies included in the analyses investigating associations between the psychological variables of stress and distress and outcomes of ART.

Authors (year)	n	Mean age (years)[a]	Mean infertility duration (years)	Outcome variable[b]	Independent variable	Measure[c]	Reported result (+, ns, −)[d]	Effect size (ESR)[e]
Demyttenare et al. (1992)	40	32.4	6.1	CPR	Depression	Zung	+	−0.25
Demyttenare et al. (1992)	40	32.4	6.1	CPR	State anxiety	STAI	+	−0.32
Merari et al. (1992)	85	NA	NA	SPR	Trait anxiety	STAI	Ns	−0.02
Merari et al. (1992)	85	NA	NA	SPR	State anxiety	STAI	Ns	+0.16
Thiering et al. (1993)	316	34.0	NA	CPR	State anxiety	STAI	Ns	−0.32
Thiering et al. (1993)	316	34.0	NA	CPR	Trait anxiety	STAI	Ns	−0.19
Visser et al. (1994)	65	NA	NA	CPR	State anxiety	STAI	Ns	−0.01
Visser et al. (1994)	65	NA	NA	CPR	Depression	HSC	Ns	−0.10
Schover et al. (1994)	120	31.0	NA	CPR	State anxiety	Brief symptom questionnaire	Ns	0.00
Schover et al. (1994)	120	31.0	NA	CPR	Depression	Brief symptom questionnaire	Ns	0.00
Schover et al. (1994)	120	31.0	NA	CPR	Stress (infertility-related)	Stress and infertility questionnaire	Ns	0.00
Boivin and Takefman (1995)	40	33.0	4.43	SPR	State anxiety	STAI	Ns	−0.18
Boivin and Takefman (1995)	40	33.0	4.43	SPR	Trait anxiety	STAI	Ns	−0.10
Harlow et al. (1996)	88	NA	NA	CPR	State anxiety	STAI	Ns	0.00
Harlow et al. (1996)	36	NA	NA	CPR	Trait anxiety	STAI	Ns	0.00
Facchinetti et al. (1997)	29	33.9	6.3	SPR	State anxiety	STAI	+	−0.37
Facchinetti et al.(1997)	29	33.9	6.3	SPR	Trait anxiety	STAI	Ns	−0.29
Slade et al. (1997)	144	32.21	8,27	CPR	Depression	BDI	Ns	−0.06
Slade et al. (1997)	144	32.21	8.27	CPR	State anxiety	STAI	Ns	0.00
Slade et al. (1997)	200	32.21	8.27	CPR	Trait anxiety	STAI	Ns	0.00
Demyttenaere et al. (1998)	98	29.7	4.1	CPR	Depression	Zung	Ns	+0.02
Milad et al. (1998)	40	33.25	3.2	LBR	State anxiety	STAI	Ns	−0.16
Sanders and Bruce (1999)	90	32.6	NA	CPR	Trait anxiety	STAI	+	−0.18
Sanders and Bruce (1999)	90	32.6	NA	CPR	State anxiety	STAI	Ns	−0.06
Sanders and Bruce (1999)	90	32.6	NA	CPR	Depression	POMS	Ns	−0.10
Csemiczky et al. (2000)	22	33.4	4.3	CPR	State anxiety	STAI	Ns	−0.40
Klonoff-Cohen et al. (2001)	123	36.8	4.06	LBR	State anxiety	POMS	+	−0.08

Continued

Table II *Continued*

Authors (year)	*n*	Mean age (years)[a]	Mean infertility duration (years)	Outcome variable[b]	Independent variable	Measure[c]	Reported result (+, ns, −)[d]	Effect size (ESR)[e]
Gallinelli *et al.*(2001)	40	NA	NA	LBR	Trait anxiety	STAI	Ns	−0.04
Gallinelli *et al.* (2001)	40	NA	NA	LBR	State anxiety	STAI	+	−0.34
Smeenk *et al* (2001)	237	33.4	3.7	SPR	State anxiety	STAI	+	−0.01
Smeenk *et al.* (2001)	237	33.4	3.7	SPR	Depression	BDI	+	−0.14
Verhaak *et al.* (2001)	207	33.4	3.7	CPR	State anxiety	STAI	Ns	−0.11
Verhaak *et al.* (2001)	207	33.4	3.7	CPR	Depression	BDI	+	−0.19
Merari *et al.* (2002)	113	33.9	NA	CPR	Trait anxiety	STAI	Ns	−0.02
Merari *et al.* (2002)	113	33.9	NA	CPR	State anxiety	STAI	Ns	+0.13
Merari *et al.* (2002)	113	33.9	NA	CPR	Depression	DACL	Ns	+0.13
Lovely *et al.* (2003)	42	31.2	NA	SPR	State anxiety	STAI	Ns	+0.19
Klonoff-Cohen *et al.* (2004)	132	36.81	4.06	CPR	Stress (infertility-related)	CART	+	−0.25
Boivin and Schmidt (2005)	818	31.5	4.09	CPR	Stress (infertility-related)	COMPI Fertility Problem Stress Scales	+	−0.07
Anderheim *et al.* (2005)	139	32.1	4.45	CPR	State anxiety	PGWB	Ns	−0.08
Anderheim *et al.* (2005)	139	32.1	4.45	CPR	Depression	PGWB	Ns	−0.02
Lancastle and Boivin (2005)	76	33.33	7.77	CPR	Trait anxiety	STAI	+	−0.07
Lancastle and Boivin (2005)	76	33.33	7.77	CPR	State anxiety	STAI	+	−0.16
Smeenk *et al.* (2005)	168	34.3	3.7	CPR	State anxiety	STAI	Ns	0.00
Smeenk *et al.* (2005)	168	34.3	3.7	CPR	Depression	BDI	Ns	0.00
Panagopoulou *et al.* (2006)	342	34.4	1.2	CPR	Stress (infertility-related)	COMPI Fertility Problem Stress Scales	Ns	−0.05
Barzalai-Pesach *et al.* (2006)	75	31.1	NA	CPR	Occupational stress	NA	+	−0.27
Karlidere *et al.* (2008)	104	30.23	8.55	CPR	Trait anxiety	STAI	+	−0.40
Karlidere *et al.* (2008)	104	30.23	8.55	CPR	State anxiety	STAI	+	−0.39
Karlidere *et al.* (2008)	104	30.23	8.55	CPR	Depression	BDI	+	−0.40
Sohrabvand *et al.* (2008)	106	29.65	7.79	CPR	Trait anxiety	Iranian Cattle anxiety questionnaire	+	−0.76

Continued

Table II *Continued*

Authors (year)	n	Mean age (years)[a]	Mean infertility duration (years)	Outcome variable[b]	Independent variable	Measure[c]	Reported result (+, ns, −)[d]	Effect size (ESR)[e]
Sohrabvand et al. (2008)	106	29.65	7.79	CPR	Depression	BDI	+	−0.72
de Klerk et al. (2008)	289	32.8	3.6	LBR	State anxiety	HADS	Ns	+0.01
Lintsen et al. (2009)	783	33.2	3.4	CPR	State anxiety	STAI	Ns	−0.01
Lintsen et al. (2009)	783	33.2	3.4	CPR	Depression	BDI	Ns	−0.04
Ebbesen et al. (2009)	781	31.2	2.63	CPR	Depression	BDI	Ns	0.00
Ebbesen et al. (2009)	791	31.2	2.63	CPR	Stress (life events)	LRE	+	−0.08
Ebbesen et al. (2009)	782	31.2	2.63	CPR	Stress (perceived)	PSS-10	Ns	+0.03

[a]NA, not available.
[b]ART outcome: SPR, serum pregnancy rate; CPR, clinical pregnancy rate; LBR, live birth rate.
[c]Measures: ZUNG, ZUNG Depression Scale; STAI, State-Trait Anxiety Scale; HSC, Hopkins Symptom Checklist; BDI, Becks Depression Inventory; POMS, Profile of Mood States; DACL, Lubin's Depression Adjective Checklist; CART, concern during assisted reproductive technologies; PGWB, psychological well-being index; HADS, Hospital Anxiety and Depression Scale; PSS-10, The Perceived Stress Scale; LRE, the list of recent events; SCL-90R, Symptom Checklist.
[d]Reported results: +, significant result in the expected directed; ns, non-significant result; −, significant result in the opposite of the expected direction.
[e]ESR, effect size correlation; negative correlation = association in the expected direction, i.e. high scores associated with reduced likelihood of pregnancy.

frequently analyzed was clinical pregnancy. Several types of stress were measured, the most commonly studied being infertility-related stress. Enrollment in ART treatment was the most frequent baseline measurement time, although some studies had baseline measurements before the start of treatment or during the course of an ART treatment cycle. All data and references can be seen in Table II. Regarding the follow-up time in studies, most studies investigated the possible influence of stress and distress on fertility-related outcomes in the window of one reproductive cycle, with follow-up time defined by the outcome measured with respect to the course of fertility or pregnancy-related events (e.g. oocyte retrieval, pregnancy test result, live birth rate etc.) (see Table II for outcome measures in the studies). Exceptions were four studies that assessed outcomes 1 year after the baseline measurement of stress (Thiering *et al.*, 1993; Strauss *et al.*, 1998; Sanders and Bruce, 1999; Boivin and Schmidt, 2005), one study evaluating pregnancy status 20 months after baseline psychological evaluation (Schover *et al.*, 1994) and one study measuring pregnancy status 6 months after termination of a program of three treatment attempts (Slade *et al.*, 1997). The percentage of first time ART attendees ranged from 0% (Merari *et al.*, 1992) to 100% (Anderheim *et al.*, 2005; Smeenk *et al.*, 2005; Karlidere *et al.*, 2008; Sohrabvand *et al.*, 2008; Ebbesen *et al.*, 2009; Lintsen *et al.*, 2009), but in the majority of studies, samples consisted mostly of first-time ART attendees (data not shown). Mean duration of infertility varied in the samples from 1.2 years (Ebbesen *et al.*, 2009) to 8.55 years (Karlidere *et al.*, 2008). Most articles presented results from bi- or univariate analyses without adjusting for other factors, yielding a total of 53 effect sizes to be analyzed. Four articles (Klonoff-Cohen *et al.*, 2001; Smeenk *et al.*, 2001; Klonoff-Cohen and Natarajan, 2004; Panagopoulou *et al.*, 2006) reported results from multivariate analyses yielding five effect sizes.

Associations between stress, distress and ART outcomes

Stress

A simple vote count showed that four out of six studies had found a statistically significant association between stress and clinical pregnancy rate in the expected direction (Table II). As seen in Table III, the heterogeneity test reached statistical significance ($P < 0.10$), and a random effects models was therefore used, showing a pooled effect size in the expected direction (ESr = −0.08, $P = 0.02$).

State anxiety

Vote counts revealed statistically significant ($P < 0.05$) results in the expected direction in 3 out of 5 studies with serum pregnancy test rate, 3 out of 15 with clinical pregnancy rate and 2 out of 4 with live births as ART outcome (Table II). As seen in Table III, only the pooled effect size (ESr = −0.10) for the association between state anxiety and clinical pregnancy reached statistical significance ($P = 0.03$). The heterogeneity tests were statistically significant ($P < 0.10$) for serum pregnancy rate, and clinical pregnancy rate and random effects models were therefore used for these models, whereas a non-significant heterogeneity test for live birth suggested the use of a fixed effects model.

Table III Results of meta-analyses of prospective studies of the influence of trait anxiety, state anxiety, depressive symptoms and stress on reproductive outcomes of ART.

Outcome[b]	Predictor	Sample size		Heterogeneity[a]			Global effect sizes			Failsafe N[e]	Criterion[f]
		K[c]	Total *n*	Q	Df	*P*	ESr[d]	95% CI	*P*		
CPR	Stress	6	1869	9.7	5	0.083	−0.08[h]	(−0.15 to −0.01)	0.02	12	40
LBR	State anxiety	4	492	5.0	3	0.171	−0.09[g]	(−0.22 to 0.04)	0.19	—	—
SPR	State anxiety	5	433	8.7	4	0.068	−0.01[h]	(−0.18 to 0.15)	0.89	—	—
CPR	State anxiety	15	2131	44.1	12	0.000	−0.10[h]	(−0.19 to −0.01)	0.03	44	75
SPR	Trait anxiety	3	125	1.5	2	0.464	−0.09[g]	(−0.25 to 0.07)	0.28	—	—
CPR	Trait anxiety	8	907	17.0	7	0.017	−0.14[h]	(−0.25 to −0.03)	0.02	26	50
CPR	Depression	13	2803	99.4	12	0.000	−0.11[h]	(−0.23 to 0.01)	0.06	—	—

[a]*P*-values of 0.1 or less were taken to suggest heterogeneity, Q: a Chi-square statistic assessing between-study differences in effect sizes; significance level = 10%.
[b]Outcome defined as SPR, serum pregnancy rate; CPR, clinical pregnancy rate; LBR, live birth rate.
[c]To maximize statistical power while ensuring independency of results, studies with multiple results were either combined or excluded in an outcome category. e.g. for stress and CPR, only one of two effect sizes from Ebbesen *et al.* could be entered in the analysis, and for depression and CPR, data from Verhaak *et al.* and Smeenk *et al.* were the same and analyzed only once.
[d]ESr, effect size correlation. A negative value indicating an effect size in the hypothesized direction, i.e. a negative association between anxiety or depression and chances of a pregnancy or live birth following ART.
[e]Failsafe *n* = number of non-significant studies that would bring the *P*-value to non-significant ($P > 0.05$).
[f]A Failsafe N exceeding the criterion (five × *k* + 10) indicates a robust result.
[g]Fixed effects model.
[h]Random effects model.

Trait anxiety

All three studies investigating the association between serum pregnancy test results and trait anxiety reported non-significant results, while four out of eight studies with clinical pregnancy as the ART outcome showed significant results in the expected direction (Table II). The pooled effect size for the association between trait anxiety and clinical pregnancy rate reached statistical significance (ESr = −0.14, $P = 0.02$), while the result for serum pregnancy tests did not reach statistical significance. The heterogeneity test was statistically significant ($P < 0.10$) for clinical pregnancy, but not for serum pregnancy tests, and a fixed model was therefore used for the latter outcome.

Depression

Vote counts showed that 4 out of 14 studies investigating the association between depression and clinical pregnancy found a statistically significant result in the hypothesized direction (see Table II). The pooled effect size showed a non-significant trend in the expected direction (ESr = −0.11, $P = 0.06$).

As seen in Table III, for all the pooled effect sizes that reached statistical significance, none of the fail-safe numbers exceeded the suggested criterion of five x the number of studies + 10, indicating less than robust results.

Moderating influences

Mean age of the study sample

The meta-regression revealed a significant influence of age on the association between trait anxiety and clinical pregnancy, with the analysis showing a positive slope (seven studies; slope: + 0.14; $Q = 44.2$; $P < 0.0001$), indicating that the association between trait anxiety and reduced chance of clinical pregnancy was stronger in younger patients. A similar result was found regarding the association between depression and clinical pregnancy (11 studies; slope: + 0.07; $Q = 18.7$; $P < 0.0001$), indicating that the association between depression and reduced chance of clinical pregnancy is stronger in younger patients. No moderating influence of age was found on the association between state anxiety and clinical pregnancy (nine studies; slope: + 0.04; $Q = 2.14$; $P = 0.14$) or on the association between stress and clinical pregnancy (six studies; slope: −0.02; $Q = 2.8$; $P = 0.09$).

Percentage of first time IVF/ICSI patients

A statistically significant positive slope was found for the influence of the percentage of first-time attendees on the association between state anxiety and clinical pregnancy (10 studies; slope + 0.002; $Q = 5.23$; $P = 0.02$), indicating a more pronounced association between state anxiety and reduced chance of clinical pregnancy in first-time attendees. In contrast, for the association between trait anxiety and clinical pregnancy, a significant negative slope was found (seven studies; slope: −0.004; $Q = 18.8$; $P < 0.0001$), indicating that the association between trait anxiety and reduced chance of clinical pregnancy was less pronounced in first-time attendees. Results did not reach statistical significance for the influence of the percentage of first-time attendees on the association between clinical pregnancy and depression (11 studies; slope: −0.00; $Q = 3.28$; $P = 0.07$). The number of studies was insufficient to calculate moderating effects of percent first-time attendees on the association between stress and clinical pregnancy (i.e. <3).

Mean duration of infertility

A statistically significant influence of mean infertility duration was found for the association between state anxiety and clinical pregnancy, as indicated by a significant negative slope (six studies; slope −0.04; $Q = 7.5$; $P = 0.006$), i.e. the association between state anxiety and

reduced chance of clinical pregnancy was more pronounced in patients with longer infertility duration. For the association between depression and clinical pregnancy, a significant negative slope was also seen (eight studies; slope: −0.06 $Q = 37.2$; $P < 0.0001$), indicating that the association between depression and reduced chance of clinical pregnancy was more pronounced in patients with longer infertility duration. The results for the association between trait anxiety (four studies; slope:+0.07; $Q = 1.67$; $P = 0.20$) or stress (three studies; slope: −0.02; $Q = 1.26$; $P = 0.26$) and clinical pregnancy rate did not reach statistical significance.

Assessment time

For the associations between pregnancy and state anxiety, trait anxiety or depression, none of the differences in effect sizes between studies assessing the independent variable before and at the time of enrollment reached statistical significance (Q: 0.41–2.8, P: 0.20–0.52) (data not shown). For stress, the number of studies to conduct meta-ANOVA analyses to test a moderating role by assessment time was insufficient (<3).

Infertility etiology

The number of studies was insufficient (<3) to estimate the influence of percentage of participants with female factor infertility on associations between the independent variables state anxiety, trait anxiety, depression or stress on the dependent variable clinical pregnancy.

Discussion

The aims of this systematic review and meta-analysis were to examine the evidence for associations between the psychological variables of stress and distress and ART outcome, to estimate the effect sizes, and to identify possible moderators. For the 31 independent studies included, vote counts showed inconsistent results, and the individual study effect sizes varied from small to large. Meta-analysis revealed statistically significant, but small (Cohen, 1988) negative associations between stress and clinical pregnancy, and between state or trait anxiety and clinical pregnancy. No significant association was found between depression and clinical pregnancy. Furthermore, the associations between state or trait anxiety and serum pregnancy test results, and between state anxiety and live births did not reach statistical significance. The fail-safe number did not exceed the criterion in any analyses, indicating that the results of the currently available studies should be considered less robust.

As seen in Table II, several studies have reported statistically significant associations in the hypothesized direction between stress, depression or anxiety and a number of pregnancy-related measures, with no studies reporting results in the opposite of the hypothesized direction. Due to few studies available, it is too early to state any final conclusions without considerable risk of type one or two error. For several initial measures related to pregnancy outcome, e.g. the number of oocytes and whether the embryo implants, all we can say for now is that partial, preliminary support is found for a negative impact of anxiety, depression and stress, and that the results warrant further replication. It is regrettable that we were unable to conduct meta-analyses for these measures, since the very early phase of IVF is critical for the outcome.

Our finding of significant negative associations between most psychological parameters and clinical pregnancy rates, but no associations between frequency of positive serum pregnancy tests or live birth rate, is not biologically plausible, and the latter analyses may have been underpowered. The number of studies in several analyses was small (<10), and analyses regarding associations between state anxiety and serum pregnancy test result or live birth, as well as associations between trait anxiety and serum pregnancy test result were based on total sample sizes below 500 and showed non-significant results.

A high level of heterogeneity across studies may generally have contributed to the less robust results. In contrast to a fixed effects model, which assumes one true effect behind the studies, the random effects model allows for the true effect to vary between study samples, e.g. the effect could be more pronounced in younger as opposed to older participants etc. If an effect varies across studies, the result is less robust as indicated by the failure to meet the fail-safe number criterion, suggesting the need for further explorative analyses of factors that may moderate the magnitude of the effects. Our results indicated several possible moderators of the associations found between stress or distress and ART outcome. The moderators found (i.e. mean age, proportions of ART inductees versus veterans, and duration of infertility) are possibly inter-related with respect to the time perspective of treatment, and the variations observed may be associated both with changes in reproductive abilities and with psychological changes. The period from oocyte pick-up to the delivery of the child is a 'black box' of less well-defined factors rarely accounted for in sufficient detail, e.g. culture conditions, embryo selection criteria and embryo transfer techniques. Emotional reactions of couples seeking ART treatment seem to change over time. An acute stress reaction to infertility is typically seen in the initial phase of diagnosing infertility and treatment enrollment, followed later by a more prolonged period of distress when no pregnancy occurs despite several ART attempts (Lalos *et al.*, 1986; Berg and Wilson, 1991; Beaurepaire *et al.*, 1994; Slade *et al.*, 1997; Boivin *et al.*, 1998; Hammarberg *et al.*, 2001; Verhaak *et al.*, 2005, 2007).There is some evidence to suggest that emotional reactions may interfere with several important steps in ART treatment procedures (i.e. number of oocytes and fertilization), or may induce menstruation cycle disturbance (Chrousos and Gold, 1992; Klonoff-Cohen *et al.*, 2001; Klonoff-Cohen and Natarajan, 2004; Ebbesen *et al.*, 2009). The severity of the stressor may determine the magnitude of cycle disturbances, with acute stress reactions inhibiting the reproductive system through activation of the HPA axis (Chrousos and Gold, 1992). On the other hand, the more prolonged stress reactions are generally associated with increased cortisol levels (Bloom and Lazerson, 1988), which may cause estradiol inhibition as indicated by impaired granulosa cell function and possibly compromised follicular maturation and lowered number of oocytes to be harvested (Lancastle and Boivin, 2005). Both stress reactions have been suggested to be associated with cycle disturbances and reproductive failure (Chrousos and Gold, 1992; Demyttenaere *et al.*, 1992; Lancastle and Boivin, 2005). This may provide some explanation for the observed between-population differences in that distress may more easily exceed a certain threshold for distress-mediated impact on reproductive abilities in a younger ART population, in the less ART experienced and in the long-term infertile group. Another possibility lies in age-group differences in the biological sensitivity to impacts of

stress and distress. Previous research has documented reduced physiological reactivity (e.g. lower reactivity for systolic blood pressure and heart rate) and encoding (stress mediated adverse biological alterations) following negative emotional experiences (Levenson *et al.*, 1994; Mather *et al.*, 2004) or stressful encounters (Uchino *et al.*, 2010) in older adults compared with young adults (and reduced physiological reactivity has been seen in ART pregnant women compared with non-pregnant women (Facchinetti *et al.*, 1997)), providing another possible explanation for the moderating effects found for age and previous ART experience on ART outcome. However, more research is needed before we are able to fully understand the complex interaction patterns between psychological states and ART outcome within a time frame that stretches over several years.

Assessment timing did not appear to influence the results, which can be interpreted as a sign of valid measurements of the targeted latent variables. Depression and trait anxiety should be stable phenomena and relatively independent of smaller events occurring during the course of an ART treatment attempt, and it should be expected that these latent variables are relatively independent of the timing of their measurement. This was confirmed in our findings. State anxiety, however, is expected to be more sensitive to fluctuations in anxiety over time and thus more easily influenced by ART-related events. The lack of influence of assessment timing within the time frame of the ART treatment program is therefore somewhat surprising and could therefore potentially be interpreted as lack of sufficient measurement sensitivity in the state anxiety measures used.

Duration of infertility or sample mean age did not moderate the associations between stress and ART outcome. While it is possible that these factors do not influence the associations between stress and ART outcome, insufficient statistical power may be a more likely explanation, insofar as the facts that there were only few primary stress studies available to us and that meta-analytic moderator tests generally have low statistical power (Hedges and Pigott, 2001).

We chose not to use quality assessment scales in this meta-analysis or adjust the results according to quality scores as the reliability and validity of such scales have been disputed. There appears to be considerable disagreement between authors of different scales as to what should be regarded as good study quality (Juni *et al.*, 1999), and furthermore, many quality assessment scales pertain to the quality of the reporting in the study rather than to the study quality per se (e.g. Vandenbroucke *et al.*, 2007) for the evaluation of observational studies) (Juni *et al.*, 1999). The use of quality scores as weights has therefore been discouraged by several authors due to the lack of statistical or empirical justification (Detsky *et al.*, 1992; Juni *et al.*, 1999). Quality scales have been suggested as useful instruments to distinguish good studies from poor (Moher *et al.*, 1995), but in their replication of a previously conducted meta-analysis using 25 different quality assessment scales suggested by Moher *et al.* (1995), Juni *et al.* (1999) found reason to question their usefulness. When reanalyzing the data, Juni *et al.* (1999) found that the original findings were confirmed only for some quality rating scales, whereas others led to opposite conclusions. Instead of using quality scores quantitatively in reviews, it has been suggested to define relevant methodological aspects prior to study conduction. We have attempted to accommodate this through the use of strict inclusion criteria in our meta-analysis and through founding the selection of studies on an *a priori* developed protocol, listing selection criteria perceived as relevant in this area of research. The primary reasons for the use of quality assessment in reviews are to avoid bias and enhance precision (Detsky *et al.*, 1992), but quality scores have been suggested to lack sensitivity for bias and may therefore constitute a less preferable tool in systematic reviews (Kunz and Oxman, 1998). Instead, we have used more informal quality assessments when addressing relevant quality factors related to study designs and sampling methods and have chosen to explore whether they moderate the associations found between stress, distress measures and ART outcomes. We have thereby sought to take possible bias into account without reducing the existing variation. Unfortunately, we have not been able to address all possible sources of bias, as we rely on the available data in the primary studies. Within stress-infertility research in ART, lifestyle factors, and perhaps clinical factors such as infertility etiology, may influence both the levels of psychological stress and the markers of fertility. However, as the majority of studies in the field have failed to control for these factors in their analyses or have not controlled for the same factors in their analyses, we were unable to address these possible sources of bias.

When considering the relatively small effect sizes found for the evaluated associations between stress, distress and ART outcome, our results may be seen as encouraging for ART patients and health professionals, who may be worried about the 'double punishment' of both experiencing infertility-related stress or distress and being at risk of further reducing the chances of obtaining a pregnancy through ART. Whereas the relevance of trying to alleviate infertility-related stress through counseling and psychosocial intervention seems evident (Boivin, 2003; de Liz and Strauss, 2005), it may appear less likely that psychotherapy should make a marked difference, at least for the group of ART patients taken as a whole. It is, however, possible that subgroups of patients with more pronounced physiological effects of stress and distress may benefit, e.g. younger patients, long-term infertility patients or ART inductees.

Based on our results, we find that an attempt to establish a clinically relevant effect size is premature at this stage. Being granted a baby and family life is highly important for ART patients and of vital importance to their life satisfaction, and therefore even small increases in the chances of obtaining a pregnancy may justify various steps taken to succeed. Furthermore, we are currently unable to fully understand how stress and distress may interfere with the process taking place from harvested oocytes to live births, let alone how to best intervene with the aim of reducing stress- or distress-mediated risks of negative ART outcomes.

More well-designed studies in the area will allow for a more robust approximation of the population effects in future meta-analyses, and future studies investigating possible psychoneuroendocrine and behavioral mediators will aid us in learning more about the processes at work in the case of stress or distress influencing ART outcome. This knowledge could also help us gain sufficient information on the basis of which to determine clinically relevant effect sizes and, if warranted, provide a basis for future intervention procedures.

Strengths and limitations

We have attempted to address the question of whether reliable associations exist between ART outcome and the psychological variables of anxiety, depression and stress. We have done so using

strict inclusion criteria and through adherence to an *a priori* developed protocol and existing guidelines for good quality reporting of results in meta-analyses. In spite of these strengths, there are also a number of potential weaknesses that should be considered when interpreting the results of our meta-analysis. First, our choice to use non-adjusted data could have resulted in overestimation of effect sizes. Only three studies presented results from multivariate analyses, and the limited number of studies did not allow us to statistically compare effect sizes with those from studies reporting unadjusted data. Second, most analyses in our meta-analysis are based on relatively few studies, and between-study heterogeneity was high, making the results less robust. Third, we chose only to include English language papers. It has been suggested that results are more likely to be published in English-language journals if the results are statistically significant, and language selection may thus constitute a source of bias. However, when contacting thirteen first authors, we encouraged them to provide any relevant data they might be in possession of, regardless of the nature of the results. However, none of the authors responded to this request. Finally, it should be noted that patients receiving intrauterine insemination were excluded from this study, as the vast majority of MAR studies focus on ART. Due to the shared problem of infertility that IUI patients have in common with ART patients, the impact of infertility and treatment for this group may be underestimated and future meta-analyses could preferably include this group also.

Conclusion and perspectives

Significant, but small, effects were found between the psychological variables of stress and distress and ART outcome, i.e. clinical pregnancy rate. Results were non-significant for associations between trait anxiety and serum pregnancy test result and for state anxiety and serum pregnancy test result and live birth rates. The results of the available prospective studies appeared less robust, and between-study heterogeneity was high. Sample mean age, mean duration of infertility and percentage of first time ART attendees in the studies appeared to moderate several of the reported associations between distress (but not stress) and ART outcome. Future meta-analyses including more studies as they become available are needed to replicate our findings. Based on the small effect sizes found in this study, a general recommendation for psychological interventions with the aim of enhancing pregnancy cannot be made, although it is possible that subgroups of more stress-susceptible patients may benefit from intervention. Overall, the results could be regarded as encouraging for the general population of ART patients so far as the evidence from this meta-analysis suggests that there may be only a relatively limited influence of psychological stress and distress on chances of achieving a pregnancy through ART treatment.

Supplementary data

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

Authors' roles

B.Z. played a role in substantial contributions to conception and design, analysis and interpretation of data, revising the manuscript for important intellectual content and the final approval of the version to be published. Y.F. took part in acquisition of data, revision of the article critically for important intellectual content and the final approval of the version to be published. H.J.I. made substantial contributions to conception and design, and interpretation of data, revising the article critically for important intellectual content and the final approval of the version to be published. S.M.S.M. was involved in conception and design, and acquisition and interpretation of data, drafting the article and revising it critically for important intellectual content, and the final approval of the version to be published.

Acknowledgements

The authors thank Rita Jensen, Librarian and Head of the Institute Library, Institute of Psychology, Aarhus University for providing valuable assistance in the initial phases of the literature search.

Funding

The study has not received funding but is part of a common university based phd-fellowship.

References

Anderheim L, Holter H, Bergh C, Moller A. Does psychological stress affect the outcome of *in vitro* fertilization? *Hum Reprod* 2005; **20**:2969–2975.

Barzilai-Pesach V, Sheiner EK, Sheiner E, Potashnik G, Shoham-Vardi I. The effect of women's occupational psychological stress on outcome of fertility treatments. *J Occup Environ Med* 2006;**48**:56–62.

Beaurepaire J, Jones M, Thiering P, Saunders D, Tennant C. Psychosocial adjustment to infertility and its treatment—male and female responses at different stages of IVF et treatment. *J Psychosom Res* 1994; **38**:229–240.

Beck A, Steer R, Brown G. *Manual: Beck Depression Inventory*, 2nd edn. San Antonio, TX: The Psychological Corp. Hartcourt & Brace, 1996.

Berg BJ, Wilson JF. Psychological functioning across stages of treatment for infertility. *J Behav Med* 1991;**14**:11–26.

Bloom FE, Lazerson A. *Brain, Mind and Behavior*, 2nd edn. New York: W.H. Freeman and Company Worth Publishers, 1988

Bohrenstein M, Rothstein H. *Comprehensive Metaanalysis*. Eaglewood, NJ: Biostat, 2006

Boivin J. A review of psychosocial interventions in infertility. *Soc Sci Med* 2003;**57**:2325–2341.

Boivin J, Schmidt L. Infertility-related stress in men and women predicts treatment outcome 1 year later. *Fertil Steril* 2005;**83**:1745–1752.

Boivin J, Takefman JE. Stress level across stages of *in vitro* fertilization in subsequently pregnant and nonpregnant women. *Fertil Steril* 1995; **64**:802–810.

Boivin J, Andersson L, Skoog-Svanberg A, Hjelmstedt A, Collins A, Bergh T. Psychological reactions during *in-vitro* fertilization: similar response pattern in husbands and wives. *Hum Reprod* 1998; **13**:3262–3267.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-analysis*, 1st edn. West Sussex, UK: John Wiley and Sons Ltd., 2009.

Chrousos GP, Gold PW. The concepts of stress and stress system disorders. Overview of physical and behavioral homeostasis. *JAMA* 1992;**267**:1244–1252.

Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

Cohen S, Williamson G. Perceived stress in a probability sample of the United States. In: Spaapan S, Oskamps M (eds). *The Social Psychology of Health*. Californa: Sage Publications Inc., 1988.

Cohen S, Kessler RC, Gordon LU. Conceptualizing stress and its relation to disease. In: Cohen S, Kessler RC, Gordon LU (eds). *Measuring Stress—a Guide for Health and Social Scientists*. New York: Oxford University Press, 1995, 3–26.

Cohen S, Kessler RC, Gordon LU. *Measuring Stress. A Guide for Health and Social Scientists*, 1st edn. New York: Oxford University Press, 1997.

Csemiczky G, Landgren BM, Collins A. The influence of stress and state anxiety on the outcome of IVF-treatment: psychological and endocrinological assessment of Swedish women entering IVF-treatment. *Acta Obstet Gynecol Scand* 2000;**79**:113–118.

de Klerk C, Hunfeld JAM, Heijnen EMEW, Eijkemans MJC, Fauser BC, Passchier J, Macklon NS. Low negative affect prior to treatment is associated with a decreased chance of live birth from a first IVF cycle. *Hum. Reprod* 2008;**23**:112–116.

de Liz TM, Strauss B. Differential efficacy of group and individual/couple psychotherapy with infertile patients. *Hum Reprod* 2005; **20**:1324–1332.

Demyttenaere K, Nijs P, Evers-Kiebooms G, Koninckx PR. Coping and the ineffectiveness of coping influence the outcome of *in vitro* fertilization through stress responses. *Psychoneuroendocrinology* 1992;**17**:655–665.

Demyttenaere K, Nijs P, Evers-Kiebooms G, Koninckx PR. Personality characteristics, psychoendocrinological stress and outcome of IVF depend upon the etiology of infertility. *Gynecol Endocrinol* 1994; **8**:233–240.

Demyttenaere K, Bonte L, Gheldof M, Vervaeke M, Meuleman C, Vanderschuerem D, D'Hooghe T. Coping style and depression level influence outcome in *in vitro* fertilization. *Fertil Steril* 1998; **69**:1026–1033.

Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992;**45**:255–265.

Ebbesen SM, Zachariae R, Mehlsen MY, Thomsen D, Hojgaard A, Ottosen L, Petersen T, Ingerslev HJ. Stressful life events are associated with a poor in-vitro fertilization (IVF) outcome: a prospective study. *Hum Reprod* 2009;**24**:2173–2182.

Facchinetti F, Matteo ML, Artini GP, Volpe A, Genazzani AR. An increased vulnerability to stress is associated with a poor outcome of *in vitro* fertilization-embryo transfer treatment. *Fertil Steril* 1997;**67**:309–314.

Gallinelli A, Roncaglia R, Matteo ML, Ciaccio I, Volpe A, Facchinetti F. Immunological changes and stress are associated with different implantation rates in patients undergoing *in vitro* fertilization-embryo transfer. *Fertil Steril* 2001;**76**:85–91.

Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics* 2001;**2**:463–471.

Hammarberg K, Astbury J, Baker HWG. Women's experience of IVF: a follow-up study. *Hum Reprod* 2001;**16**:374–385.

Harlow CR, Fahy UM, Talbot WM, Wardle PG, Hull MG. Stress and stress-related hormones during *in-vitro* fertilization treatment. *Hum Reprod* 1996;**11**:274–279.

Hedges L, Pigott TD. The power of statistical tests for moderators in meta-analyses. *Psychol Methods* 2001;**9**:426–445.

Henderson S, Byrne DG, Duncan-Jones P. *Neurosis and the Social Environment*. Sydney: Academic Press, 1981.

Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;**282**:1054–1060.

Karlidere T, Bozkurt A, Ozmenler KN, Ozsahin A, Kucuk T, Yetkin S. The influence of emotional distress on the outcome of *in-vitro* fertilization (IVF) and/or intracytoplasmic sperm injection (ICSI) treatment among infertile Turkish women. *Isr J Psychiatry Relat Sci* 2008;**45**:55–64.

Kee BS, Jung BJ, Lee SH. A study on psychological strain in IVF patients. *J Assist Reprod Genet* 2000;**17**:445–448.

Klonoff-Cohen H. Female and male lifestyle habits and IVF: what is known and unknown. *Hum Reprod Update* 2005;**11**:179–203.

Klonoff-Cohen H, Natarajan L. The concerns during assisted reproductive technologies (CART) scale and pregnancy outcomes. *Fertil Steril* 2004; **81**:982–988.

Klonoff-Cohen H, Chu E, Natarajan L, Sieber W. A prospective study of stress among women undergoing *in vitro* fertilization or gamete intrafallopian transfer. *Fertil Steril* 2001;**76**:675–687.

Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;**317**:1185–1190.

Lalos A, Lalos O, Jacobsson L, Von Schoultz B. Depression, guilt and isolation among infertile women and their partners. *J Psychosom Obstet Gynaecol* 1986;**5**:197–206.

Lancastle D, Boivin J. Dispositional optimism, trait anxiety, and coping: unique or shared effects on biological response to fertility treatment? *Health Psychol* 2005;**24**:171–178.

Levenson RW, Carstensen LL, Gottman JM. The influence of age and gender on affect, physiology, and their interrelations: a study of long-term marriages. *J Pers Soc Psychol* 1994;**67**:56–68.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;**6**: e1000100.

Lintsen AM, Verhaak CM, Eijkemans MJ, Smeenk JM, Braat DD. Anxiety and depression have no influence on the cancellation and pregnancy rates of a first IVF or ICSI treatment. *Hum Reprod* 2009;**24**: 1092–1098.

Lovely LP, Meyer WR, Ekstrom RD, Golden RN. Effect of stress on pregnancy outcome among women undergoing assisted reproduction procedures. *South Med J* 2003;**96**:548–551.

Mather M, Canli T, English T, Whitfield S, Wais P, Ochsner K, Gabrieli JD, Carstensen LL. Amygdala responses to emotionally valenced stimuli in older and younger adults. *Psychol Sci* 2004;**15**:259–263.

Merari D, Feldberg D, Elizur A, Goldman J, Modan B. Psychological and hormonal changes in the course of *in vitro* fertilization. *J Assist Reprod Genet* 1992;**9**:161–169.

Merari D, Chetrit A, Modan B. Emotional reactions and attitudes prior to *in vitro* fertilization: an interspouse study. *Psychol Health* 2002; **17**:629–640.

Milad MP, Klock SC, Moses S, Chatterton R. Stress and anxiety do not result in pregnancy wastage. *Hum Reprod* 1998;**13**:2296–2300.

Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;**16**:62–73.

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. *Lancet* 1999;**354**:1896–1900.

Panagopoulou E, Vedhara K, Gaintarzti C, Tarlatzis B. Emotionally expressive coping reduces pregnancy rates in patients undergoing *in vitro* fertilization. *Fertil Steril* 2006;**86**:672–677.

Rosenhan DL, Seligman MEP. *Abnormal Psychology*, 3rd edn. New York: Norton & Company, Inc., 1995.

Rosenthal R, Rubin DB. r equivalent: a simple effect size indicator. *Psychol Methods* 2003;**8**:492–496.

Sanders KA, Bruce NW. Psychosocial stress and treatment outcome following assisted reproductive technology. *Hum Reprod* 1999; **14**:1656–1662.

Schover LR, Greenhalgh LF, Richards SI, Collins RL. Psychological screening and the success of donor insemination. *Hum Reprod* 1994; **9**:176–178.

Skakkebaek NE, Jorgensen N, Main KM, Rajpert-De Meyts E, Leffers H, Andersson AM, Juul A, Carlsen E, Mortensen GK, Jensen TK *et al*. Is human fecundity declining? *Int J Androl* 2006;**29**:2–11.

Slade P, Emery J, Lieberman BA. A prospective, longitudinal study of emotions and relationships in *in-vitro* fertilization treatment. *Hum Reprod* 1997;**12**:183–190.

Smeenk JM, Verhaak CM, Eugster A, van Minnen A, Zielhuis GA, Braat DD. The effect of anxiety and depression on the outcome of *in-vitro* fertilization. *Hum Reprod* 2001;**16**:1420–1423.

Smeenk JM, Verhaak CM, Vingerhoets AJ, Sweep CG, Merkus JM, Willemsen SJ, van Minnen A, Straatman H, Braat DD. Stress and outcome success in IVF: the role of self-reports and endocrine variables. *Hum Reprod* 2005;**20**:991–996.

Sohrabvand F, Abedinia N, Pirjani R, Jafarabadi M. Effect of anxiety and depression on ART outcome. *Ir J Reprod Med* 2008;**6**:89–94.

Spielberger C, Gorsuch R, Lushene R. *Manual for the State—Trait Anxiety Inventory*, Palo Alto, CA: Consulting Psychologists Press, 1977.

Stoleru S, Cornet D, Vaugeois P, Fermanian J, Magnin F, Zerah S, Spira A. The influence of psychological factors on the outcome of the fertilization step of *in vitro* fertilization. *J Psychosom Obstet Gynaecol* 1997; **18**:189–202.

Strauss B, Hepp U, Staeding G, Mettler L. Psychological characteristics of infertile couples: can they predict pregnancy and treatment persistence? *J Community Appl Soc Psychol* 1998;**8**:289–301.

Swan SH, Elkin EP, Fenster L. The question of declining sperm density revisited: an analysis of 101 studies published 1934–1996. *Environ Health Perspect* 2000;**108**:961–966.

Thiering P, Beaurepaire J, Jones M, Saunders D, Tennant C. Mood state as a predictor of treatment outcome after *in vitro* fertilization/embryo transfer technology (IVF/ET). *J Psychosom Res* 1993;**37**:481–491.

Uchino BN, Birmingham W, Berg CA. Are older adults less or more physiologically reactive? A meta-analysis of age-related differences in cardiovascular reactivity to laboratory tasks. *J Gerontol B Psychol Sci Soc Sci* 2010;**65B**:154–162.

Vandenbroucke JP, Von EE, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007;**147**:W163–W194.

Verhaak CM, Smeenk JM, Eugster A, van Minnen A, Kremer JA, Kraaimaat FW. Stress and marital satisfaction among women before and after their first cycle of *in vitro* fertilization and intracytoplasmic sperm injection. *Fertil Steril* 2001;**76**:525–531.

Verhaak CM, Smeenk JM, van Minnen A, Kremer JA, Kraaimaat FW. A longitudinal, prospective study on emotional adjustment before, during and after consecutive fertility treatment cycles. *Hum Reprod* 2005; **20**:2253–2260.

Verhaak CM, Smeenk JM, Evers AWM, Kremer JAM, 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.

Visser AP, Haan G, Zalmstra H, Wouters I. Psychosocial aspects of *in vitro* fertilization. *J Psychosom Obstet Gynaecol* 1994;**15**:35–43.

Zegers-Hochschild F, Adamson GD, de MJ, Ishihara O, Mansour R, Nygren K, Sullivan E, van der PS. 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.