**Human Reproduction Update, Vol.19, No.2 pp. 124–135, 2013**
Advanced Access publication on November 23, 2012 doi:10.1093/humupd/dms045

human reproduction update

# Why we should talk about compliance with assisted reproductive technologies (ART): a systematic review and meta-analysis of ART compliance rates

**S. Gameiro[1,2,*], C.M. Verhaak[3], J.A.M. Kremer[4], and J. Boivin[2]**

[1]Faculty of Psychology and Educational Sciences, University of Coimbra, Rua do Colégio Novo, Apartado 6153, Coimbra 3001-802, Portugal [2]Cardiff Fertility Studies Research Group, School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff, Wales CF10 3AT, UK [3]Department of Medical Psychology, Radboud University Nijmegen, Medical Centre, Nijmegen, The Netherlands [4]Department of Obstetrics and Gynaecology, Radboud University Nijmegem, Medical Centre, Nijmegen, The Netherlands

Correspondence address. E-mail: GameiroS@cardiff.ac.uk

*Submitted on June 8, 2012; resubmitted on August 21, 2012; accepted on September 3, 2012*

## TABLE OF CONTENTS



**BACKGROUND:** The goal of this systematic review and meta-analysis was to estimate the rate of compliance with assisted reproductive technologies (ART) and examine its relationship with treatment success rates.

**METHODS:** Six databases were systematically searched from 1978 to December 2011. Studies were included if they reported data on patient progression through three consecutive standard ART cycles. Compliance was estimated for the first three ART cycles (typical ART Regimen Compliance, TARC) and after the first and the second failed cycles (CAF1, CAF2). Treatment success rates for all patients who started ART and for those who fully complied with the three ART cycles were estimated.

**RESULTS:** Ten studies with data for 14 810 patients were included. TARC was 78.2% [95% confidence interval (CI) 68.8–85.3%], CAF1 was 81.8% (73.3–88.1%) and CAF2 was 75.3% (68.2–81.2%). The overall success rate was 42.7% (32.6–53.6%) for all patients starting ART and 57.9% (49.4–65.9%) for those who complied with three ART cycles. Compliance rates did not vary according to study quality, but TARC

was higher for studies that reported data on doctor-censored patients versus those that did not (84.2% 95% CI 75.5–90.2 versus 70.6% 95% CI 58.3–80.5, $P = 0.043$). Analysis of funnel plots and the Egger test indicated publication bias for CAF1.

**CONCLUSIONS:** Findings from this meta-analysis should reassure clinics and patients that most patients are able to comply with three cycles of ART. Compliers could increase their chances of success by as much as 15%. A more detailed assessment of compliance requires monitoring long-term treatment trajectories through the creation of national registries.

**Key words:** assisted reproductive technologies / compliance rates / discontinuation / success rates

# Introduction

Most couples have life plans that include having children but 9–15% will have problems conceiving spontaneously (Boivin *et al*., 2007). Infertility is a significant impairment of function, which the first World Disability Survey ranks as 5th in the list of moderate to severe disabilities within the global population under the age of 60 (World Health Organization and The World Bank, 2011). Fortunately, the chances of achieving parenthood are high for couples undergoing fertility treatment. The world live birth rate with assisted reproductive technologies (ART, e.g. IVF) is 22% per single initiated cycle of treatment (de Mouzon *et al*., 2009) but can be 49% (Stern *et al*., 2010) or higher (Witsenburg *et al*., 2005; Verhagen *et al*., 2008) if people undergo the optimal number of cycles, typically three [National Institute for Clinical Excellence (NICE), 2004, p.5]. However, many couples do not undergo multiple cycles of ART, even when there is a favourable prognosis and ability to cover the costs of treatment (Domar 2004; Brandes *et al*., 2009). Indeed, discontinuation rates as high as 65% mainly due to psychological demands of treatment (Smeenk *et al*., 2004; Brandes *et al*., 2009) have been reported (Rajkhowa *et al*., 2006). Practice guidelines and national regulations emphasize the importance of discussing treatment success rates but not the rates of discontinuation [National Institute for Clinical Excellence (NICE), 2004; European Society of Human Reproduction and Embryology (ESHRE), 2008; The Practice Committee of the Society for Assisted Reproductive Technology and the Practice Committee of the American Society for Reproductive Medicine, 2008]. Recently, the UK National Institute for Clinical Excellence (NICE) recommended using compliance as a way of auditing treatment delivery at clinics [National Institute for Clinical Excellence (NICE), 2004, p.42], but to our knowledge, this has not been done. The World Health Organization (WHO) defines treatment compliance (or adherence) as '… the extent to which a person's behaviour follows medical advice or corresponds with agreed recommendations from a health care provider…' (WHO, 2003, p.3). In medical practice, in general, compliance means 'the degree of constancy and accuracy with which a patient follows a prescribed regimen' (http://medical-dictionary.thefreedictionary.com/compliance). Therefore, in ART, compliance would refer to the uptake of the ART cycles recommended by the doctor until pregnancy is achieved or until there is a recommendation to end treatment (as well as compliance with medication, which is not addressed in the present review.) Although the terminology is compatible with the concepts of shared and informed decision-making on the part of the patient, there has been a reluctance to conceptualize discontinuation in ART as a compliance issue or to influence patient decision-making about pursuing treatment. Reference to compliance is made implicitly, when clinicians mention cumulative pregnancy rates or offer financial packages that take into account better success rates with multiple cycles (Garrido *et al*., 2011); however, few patients recall having the opportunity to discuss the advantages (24%) or disadvantages (18%) of ending/continuing treatment (Peddie *et al*., 2004). The lack of emphasis on compliance in fertility treatment may be due to several factors. Unlike other disease contexts, people can opt out of fertility treatment without threatening their physical health and opting out can at times have beneficial consequences, for example on mental health (Peddie *et al*., 2005). Active intervention to encourage compliance could also be avoided because of popular conceptions of fertility doctors taking advantage of desperate infertile couples (Thompson, 2005). However, even if doctors want to discuss compliance with their patients, they lack precise information as its prevalence has not yet been systematically estimated from the available literature. Whatever the cause, providing explicit information about compliance at the start of treatment (e.g. compliance rate, consequence of ending/continuing treatment on success rate) is essential for informed consent; otherwise, patients begin treatment optimistic about success without fully realizing that the demands of treatment (e.g. physical, emotional and practical) may be such that they are unable to pursue the optimal number of cycles even when their prognosis is favourable and costs of treatment are covered (Domar, 2004; McDowell and Murray, 2011).

There is high variability in the discontinuation rate reported in primary research, ranging from 15% (Brandes *et al*., 2009) to 65% (Rajkhowa *et al*., 2006), which makes it difficult to be confident about compliance. Variability may, in large part, be explained by the lack of consensus on the definition and monitoring of compliance; for example, in many studies the non-complier group includes poor prognosis patients who discontinued treatment because they were advised to stop treatment (De Vries *et al*., 1999), some studies monitor patients for too short a follow-up period to accurately conclude on compliance (Land *et al*., 1997) and most studies do not control for patients who continue treatment at different clinics (Stolwijk *et al*., 1996; Verhagen, *et al*. 2008) or at a later time in their lives (Pearson *et al*., 2009). Other issues that contribute to variability in the compliance rate reported are treatment reimbursement policy, the type of population under study (e.g. previous experience with ART and parity), the type of ART treatment investigated and other methodological aspects (e.g. design, assessment of treatment initiation and success). Another important issue is that primary research has shown that ART success rates cannot be accurately estimated without considering discontinuation (Land *et al*., 1997), and therefore, the aforementioned issues would also impact on the reporting of success rates in ART. Further, the clinics' success rates

may also influence compliance as past research has shown that people move to clinics perceived to have higher pregnancy rates to improve their chances of success (Marcus et al., 2005). A systematic review taking into account these issues would help achieve greater clarity on compliance in ART and its association with treatment success rates.

The aims of the present systematic review and meta-analysis were 3-fold. The first goal was to provide the first estimate of compliance among typical infertile patients undergoing standard ART treatment. In order to promote future consensus on how to define, monitor and report compliance, the second goal was to examine conceptual and methodological causes of variability in compliance. Finally, the third goal was to assess how compliance is associated with treatment success rates.

# Methods

## Systematic search

The present work is part of a larger review that investigated reasons and predictors of discontinuation from fertility treatment (Gameiro et al., 2012). The Sure Support Unit for Research Evidence (Cardiff University) searched six databases (Medline, Medline In Progress, EMBASE, BNI, PsycINFO and The Cochrane Library) from 1978 to December 2011 (inclusive). A search strategy was created using terminology from the International Committee for Monitoring Assisted Reproductive Technology and the WHO-revised glossary of ART (Zegers-Hochschild et al., 2009) for fertility treatment (e.g. ART, IVF) AND discontinuation (e.g. dropout, compliance and discontinuation), which, with small adaptations, was used in all databases (see Supplementary data, Table SI). MeSH terms were used in PubMed. No restriction was made on the type (journal, conference paper or dissertation) or language of publication. The reference sections of all identified articles were examined by S.G. and a research specialist (Debbie Moss, see funding) to identify other relevant manuscripts.

## Inclusion and exclusion criteria

Studies were included if data were reported (or could be obtained from the corresponding author) on patient progression through a maximum of three consecutive standard ART (IVF or ICSI) cycles (i.e. number of patients starting, pregnant, discontinuing, continuing after failed treatment) or, if fewer, until pregnancy or until the clinician recommended the patient to end treatment (i.e. doctor censoring, where this information was provided). Three cycles were used because it is the typically recommended and/or subsidized number of cycles that patients face for an optimal chance of pregnancy in an ART programme [National Institute for Clinical Excellence (NICE), 2004]. Only studies that focused on patients with no previous experience of ART were included. Studies that solely investigated single groups (e.g. third-party reproduction, recurrent miscarriage) or specific ART treatment (e.g. modified natural IVF, transport IVF/ICSI) were also excluded to focus on the typical ART population. Duplicate or secondary publications on the same sample were excluded to avoid multiple-publication bias. In these cases, we prioritized the publication that focused on discontinuation from treatment and, if this criterion did not apply, the publication that reported data for the largest sample. Excluded studies were classified according to reason for exclusion (see Fig. 1).

## Data extraction

S.G. and a research specialist (D.B.) extracted data using a standardized protocol. Disagreement was resolved by discussion. Data were extracted or obtained from the corresponding author on characteristics of the study (e.g. country of origin, design), study population (e.g. average female age), clinical protocol (e.g. type of ART), health context (e.g. availability of subsidized/reimbursed treatment) and methodology (e.g. duration of follow-up period, inclusion or exclusion of cryopreserved IVF cycles in data reported). The data extracted to calculate the compliance rates were the numbers of patients who started treatment, who had successful or failed treatment, who were recommended to end treatment by their doctor (i.e. doctor censoring, where provided) and who discontinued or continued after a failed cycle. For those studies that reported on doctor censoring, data on its medical indication were also extracted.

## Quality assessment

S.G., J.B. and C.M.V. assessed study quality according to the Newcastle–Ottawa Quality (NOQ) assessment scale (Wells et al., 2010) adapted for the present study. The NOQ is used to appraise quality in terms of population representativeness, measurement of outcome (compliance), within-population comparability (compliers versus discontinuers) and adequacy of follow-up (completion rates). The specific criteria used for quality assessment were already described elsewhere (Gameiro et al., 2012). Low-, moderate- and high-quality labels were assigned to scores of 0–2, 3–5 and 6–7, respectively (see Supplementary data, Table SVI).

## Data analysis

Studies differed in terms of the number of subsidized treatment cycles and the number of cycles followed up. To control for this variability, we based our compliance calculations on the treatment uptake for the first three ART cycles. Uptake of the first cycle was 100% because studies only followed up patients who did a first cycle. We assumed that after failure on first or second cycles, patients would be expected to undertake a further cycle unless they were recommended to end treatment (i.e. doctor censoring).

Ideally, treatment success should be defined as achievement of a live birth. However, that is often not the case in primary research. Thus, treatment success (versus failure) was defined according to the success outcome reported in the primary study, which could be a β-hCG urine or blood test $\leq$21 days after embryo transfer, an ultrasonographic visualization of fetal heart activity or a live birth, as per standard definitions (Zegers-Hochschild et al., 2009).

Three compliance rates were calculated per study: Typical ART Regimen Compliance (TARC) and compliance after the first and the second failed cycles (Compliance After-Failure, CAF1, CAF2).

The TARC rate referred to patients who complied with all treatments recommended to them, that is, patients who continued with treatment for up to three cycles or until treatment success (as defined) or until advised to end treatment (i.e. doctor censoring). TARC was the sum of the number of patients who opted to undergo all three cycles when they failed on the first and the second cycles and of patients who stopped treatment either because it was successful or because they were censored by the doctor (where data on doctor censoring was reported), divided by the total number starting ART:

$$\text{TARC} = [\text{number of patients who underwent three cycles} + \text{number pregnant or with live birth} + \text{number doctor censored (if reported)}]/\text{number started}$$

The CAF rates provided an after-failure examination of compliance, that is, of patients who opted to undergo a further cycle after having had a failed cycle and therefore was the sum of the number of patients undergoing a further cycle divided by the number of patients with a failed cycle. Compliance after-failure was calculated for the first (CAF1) and the second (CAF2) failed

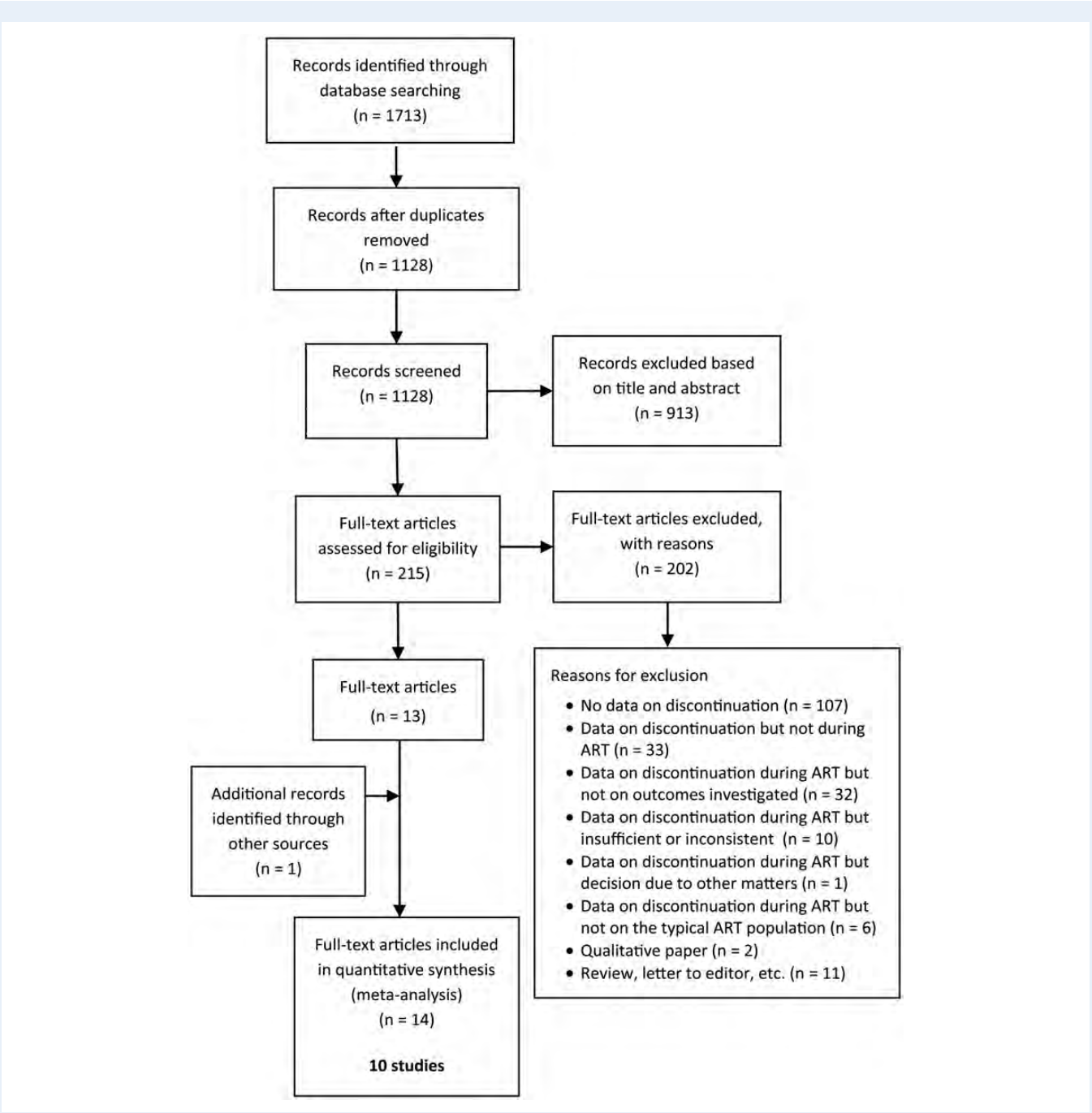


**Figure 1** Decision flowchart for identified studies.

cycles. Doctor-censored patients were excluded from the calculation of compliance after-failure rates (where such data were reported) because these patients would have been recommended to stop treatment, and therefore were not eligible for cycle uptake. The following formulas were used:

$$\text{CAF1} = (\text{number of patients who underwent second cycle})/[\text{number failed first cycle} - \text{number doctor censored after first cycle (if reported)}]$$

$$\text{CAF2} = (\text{number of patients who underwent third cycle})/[\text{number failed second cycle} - \text{number doctor censored after second cycle (if reported)}]$$

To examine whether the clinic's success rates per cycle were associated with compliance after those cycles, we computed treatment success rates per cycle (first and second cycle) for each study. As all but one study (Rufat

et al., 1994) (excluded from this analysis) were single centre, this was equivalent to providing first and second cycle success rates for each clinic. The rates per cycle (first and second cycle) were the number of patients with a successful outcome in the first or the second cycle divided by the number of patients who underwent the first or the second cycle.

To investigate how treatment success rates varied when compliance was taken into account, we calculated an overall success rate, which was the number of patients with a successful outcome in the first three ART cycles divided by the number of all patients who started the first ART cycle. We then calculated a separate typical regimen success rate that included only compliers (as defined in preceding TARC formula), that is, the number of patients with a successful outcome in the first three ART cycles divided by the number of compliers. Therefore, for each study, we had three types of success rates: clinic success rates per cycle [first and second cycles, excluding the study by Rufat et al. (1994)], overall success rate and success rate for compliers.

In order to correct for variations in study sample size, pooled estimates across studies were obtained by means of random-effects models, after log transformation. We chose a random-effects model because single-group meta-analysis produces substantial heterogeneity. The $I^2$ index was used to describe the proportion of total variation in study estimates that was due to heterogeneity (Higgins et al., 2003). Subgroup and meta-regression analyses based on the random-model were performed to identify causes of heterogeneity in compliance rates among studies. Causes were defined *a priori* and referred to characteristics related to the studies' clinical aspects [clinic geographic location, clinic's success rates per cycle (assessed only for CAF), number of embryo transfer policy and whether treatment was subsidized/reimbursed], patients (parity) and methodology (study design, handling of doctor censoring, length of follow-up, definition of start-of-cycle and success, handling of cryopreserved IVF cycles, quality rating, year of publication). The $\chi^2$ test was used to assess differences between the subgroups and the significance of the meta-regression coefficients were assessed with a $Z$-test. Publication bias was examined via visual inspection of the funnel plots (of the natural log of the rates against its standard error) and the Egger's test (Egger et al., 1997). Trim and fill was used to adjust the pooled rates for the presence of publication bias (Duval and Tweedie, 2000). We used the Comprehensive Meta Analysis software (Biostat Inc, 2011).

# Results

## Description of studies

The systematic search yielded 1128 non-duplicated records. Figure 1 presents the study decision flow chart. S.G. and D.B. agreed inclusion on all studies and agreed on reasons for exclusion for 91% of studies (see Table II of supplemental material for reasons for exclusion of full manuscripts screened). The authors of the 10 papers with missing or inconsistent data and of 5 other included papers were contacted to obtain missing data from the manuscripts. Four authors replied stating that the requested data were not available.

The 10 included studies sampled 14 810 patients from five countries. The population characteristics and design features of the studies are shown in Tables I and II. See Supplementary data, Tables SIII–V for treatment trajectory data. Critical appraisal of the studies is shown in Table III. NOQ ratings indicated no low-quality study, three average studies (30%) and seven high-quality studies (70%) with substantial inter-rater agreement (S.G. and C.M.V.: Cohen's $\kappa = 0.750$, $P = 0.007$; S.G. and J.B.: Cohen's $\kappa = 0.872$, $P < 0.001$). See Table SIV of supplementary data for details on critical appraisal of the studies.

## Meta-analysis

### *Compliance rates*

Figure 2 shows the pooled TARC rate for the random-effect model. One study, Brandes et al. (2009), did not report on data per cycle and was not included in the calculation of CAF rates. The meta-analysis showed that TARC was 78.2% [95% confidence interval (CI) 68.8–85.3%, $I^2 = 99.17$], CAF1 was 81.8% (73.3–88.1%, $I^2 = 98.66$) and CAF2 was 75.3% (68.2–81.2%, $I^2 = 95.64$).

### *Subgroup and meta-regression analyses*

Table IV presents the results of subgroup analysis performed. It was not possible to perform subgroup analysis on the basis of the geographical location of the clinic, parity or embryo transfer policy, whether treatment was subsidized/reimbursed, the definition of initiated cycle and handling of cryopreserved embryo transfers, because at least one of the subgroups had only one or no study. Variability among studies was explained only by how compliance was defined. More precisely, those studies that reported on the number of doctor-censored patients (and thus considered them to be compliers) presented higher TARC rates than studies that did not. The differences observed between subgroups related to the study design and population, length of follow-up and definition of treatment success were not significant. Finally, meta-regressions showed that publication year was not significantly related to compliance (TARC: Slope = 0.08, $Z = 1.716$, $P = 0.086$; CAF1: Slope = 0.06, $Z = 1.312$, $P = 0.190$; CAF2: Slope = 0.03, $Z = 0.813$, $P = 0.416$).

We excluded the study by Rufat et al. (1994) from the examination of associations between per cycle success rates of the clinics and subsequent compliance because, as already explained, Rufat pooled data from several fertility clinics. In addition, another study, Brandes et al. (2009), did not report data per cycle and could not be included. The clinic's first-cycle success rate was not significantly associated with compliance after that first cycle (CAF1: Slope = 0.49, $Z = 0.196$, $P = 0.845$) and the clinic's second-cycle success rate was not associated with compliance after that second cycle (CAF2: Slope = 0.34, $Z = 0.137$, $P = 0.891$).

### *Study quality and publication bias*

We performed subgroup analysis according to study quality (moderate or high) but the results of this analysis were not significant (see Table IV).

Egger's test indicated the presence of publication bias for TARC (intercept = 14.21, $t = 6.045$, $P < 0.001$), CAF1 (intercept = 10.54, $t = 5.31$, $P = 0.001$) and CAF2 (intercept = 6.22, $t = 4.70$, $P = 0.002$). Investigation of publication bias through visual inspection of the funnel plot (see Supplementary data, Figs. S1–III) was confirmed only for CAF1, where one study was found to the lower right of the pooled compliance rate and none to the left (Supplementary data, Fig. S2). The trim and fill method only identified one missing study for CAF1, estimating a new compliance rate of 80.5% (95% CI 72.0–86.9).

**Table I Sample characteristics reported in the 10 included studies.**

| Study | Country | Sample size | Selected population If yes, description | Age of women in years, mean ± SD (range) | Duration of infertility in years, mean ± SD | Parity (none or at least one child) |
|---|---|---|---|---|---|---|
| Brandes, *et al.* (2009, 2011) | The Netherlands | 373 | No | COMP:31.0 ± 4.1, DISC:33.3 ± 5.1[a] | COMP:1.31 ± 1.0, DISC:1.9 ± 1.68[a] | NR |
| De Vries, *et al.* (1998, 1999) | Belgium | 1169 | No | COMP:31 ± 4.3, DISC:32 ± 5.5[a] | NR | NR |
| Emery *et al.* (1997) and Slade *et al.* (1997) | UK | 130 | No | 32.21 ± 3.37 | 8.27 ± 2.97 | None |
| Land *et al.* (1997) | The Netherlands | 197 | No | NR | NR | NR |
| Pearson *et al.* (2009) | USA | 2245 | Excluded patients using donor gametes | 35.2 ± 4.3 (20–49) | NR | At least one child |
| Rufat *et al.* (1994) | France | 8362 | No | 33.1 ± 4.3 | NR | NR |
| Smeenk *et al.* (2004) | The Netherlands | 380 | No | 34.1 ± 3.9 (21–43) | 3.7 ± 2.2 (1–16) | NR |
| Stolwijk *et al.* (1996) | The Netherlands | 616 | Excluded patients using donor gametes | NR | NR | NR |
| Verhagen *et al.* (2008) | The Netherlands | 588 | Excluded patients starting IVF for preimplantation genetic diagnosis, surgical sperm aspiration or using donor gametes | COMP:32.9 ± 3.6, DISC:33.8 ± 4.1[a] | COMP:3.0 ± 2.2, DISC:3.5 ± 2.4[a] | NR |
| Witsenburg *et al.* (2005) | The Netherlands | 750 | No | 33.0 ± 4.0 | NR | NR |

IVF, *In vitro* fertilization; COMP, group of patients who complied with treatment; DISC, group of patients who discontinued; NR, not reported; USA, United States of America; UK, United Kingdom.
[a]Average age and duration of infertility for total sample not reported.

**Table II Design characteristics of the 10 included studies.**

| Study | Prospective design[a] (yes/no) | Data collection period | Data available on number of doctor-censored patients (yes/no, if yes, reason for censoring) | Definition of cycle start (started ovarian stimulation, had oocyte retrieval) | Definition of treatment success (positive test, positive scan, live birth[b]) | Number of embryo transfer policy | Follow-up period (<12 months, ≥12 months[c]) | IVF cycles exclude cryopreserved embryo transfers (yes/no) | Subsidized/ reimbursed treatment (yes/no) |
|---|---|---|---|---|---|---|---|---|---|
| Brandes *et al.* (2009) | No | 2002–2004 | Yes, 'poor prognosis (doctor's refusal)' | Ovarian stimulation | Positive scan | NR | ≥12 months | NR | Yes |
| De Vries *et al.* (1999) | No | 1993–1996 | No | Ovarian stimulation | Positive test | NR | ≥12 months | NR | NR |
| Emery *et al.* (1997) | Yes | 1 year | No | Ovarian stimulation | Positive test | NR | ≥12 months | NR | Yes |
| Land *et al.* (1997) | No | 1993–1994 | Yes, 'denied further treatment for medical reasons (poor response to hMG or poor fertilization)' | Ovarian stimulation | Positive scan | NR | <12 months | NR | Yes |
| Pearson *et al.* (2009) | No | 1994–1998 and 1999–2003 | No | Ovarian stimulation | Live birth | NR | NR | Yes | NR |
| Rufat *et al.* (1994) | No | 1988–1992 | No | Oocyte retrieval | Positive scan | NR | ≥12 months | NR | NR |
| Smeenk *et al.* (2004) | Yes | 1999–2000 | Yes, 'active censuring' | Ovarian stimulation | Positive scan | NR | ≥12 months | No[d] | Yes |
| Stolwijk *et al.* (1996) | No | 1988–1993 | Yes, 'a previous treatment with a fertilization rate of <10%, despite the presence of more than three large follicles (15 mm) on the day of HCG administration and the performance of oocyte aspiration, or three or less large follicles during two previous treatments' | Ovarian stimulation | Positive scan | NR | NR | No[e] | NR |
| Verhagen *et al.* (2008) | No | 2000–2003 | Yes, 'active censuring (poor response, poor fertilization, poor response with poor fertilization, overweight with BMI >30 kg/m$^2$, hypertension or improved semen quality not requiring ICSI any more)' | Ovarian stimulation | Positive test | NR | NR | No[e] | Yes |
| Witsenburg *et al.* (2005) | No | 1996–2000 | No | Ovarian stimulation | Live birth | Maximum of two when age <38, maximum of three when age ≥3 | <12 months | No[e] | Yes |

NR, not reported; hMG, human menopausal gonadotrophins; HCG, human chorionic gonadotropin; BMI, body mass index; ICSI, intra cytoplasmic sperm injection.
[a]Prospective studies are those where study design and data collection happened before any information on the outcome of interest was collected.
[b]Positive test: positive βhCG urine/blood test, positive scan: fetal heart activity at 6/7 weeks.
[c]or adequacy of follow period sufficiently justified by authors.
[d]No information was given about how cryopreserved embryo transfer cycles were considered.
[e]Transfers of cryopreserved embryos were considered to be part of the cycle from which the embryos resulted.

**Table III Quality ratings for the 10 included studies using an adapted Newcastle–Ottawa Quality assessment scale.**

| Study | Quality criterion | | | | Overall quality rating (0–7) |
|---|---|---|---|---|---|
| | Representative population[a] (0–1) | Ascertainment of treatment trajectory[b] (0–3) | Comparability[c] (0–2) | Follow-up[d] (0–1) | |
| Brandes *et al.* (2009) | 1 | 3 | 2 | 1 | 7 (high) |
| De Vries *et al.* (1999) | 1 | 2 | 2 | 1 | 6 (high) |
| Emery *et al.* (1997) | 1 | 2 | 2 | 1 | 6 (high) |
| Land *et al.* (1997) | 1 | 2 | 2 | 1 | 6 (high) |
| Pearson *et al.* (2009) | 1 | 1 | 1 | 1 | 4 (moderate) |
| Rufat *et al.* (1994) | 1 | 2 | 2 | 1 | 6 (high) |
| Smeenk *et al.* (2004) | 1 | 3 | 2 | 0 | 6 (high) |
| Stolwijk *et al.* (1996) | 1 | 2 | 1 | 1 | 5 (moderate) |
| Verhagen *et al.* (2008) | 1 | 2 | 2 | 1 | 6 (high) |
| Witsenburg *et al.* (2005) | 1 | 1 | 2 | 1 | 5 (moderate) |
| % of studies that meet criteria | 100% | 20% meet three criteria<br>60% meet two criteria<br>20% meet one criteria | 80% meet two criteria<br>20% meet one criteria | 90% | 70% (high)<br>30% (moderate)<br>0% (low) |

[a]The 'representativeness criterion' was met when >80% of eligible patients were invited and >80% agreed to participate, or when the study reported on all consecutive series of patients over a defined period of time, or when sample size was >300 (1 point).
[b]The 'ascertainment of treatment trajectory' criterion was met if the study provided enough data to ascertain that withdrawal from treatment was premature (before three cycles completed and not pregnant and not due to poor prognosis; 1 point), that withdrawal was either permanent (at least 12-monthperiod since last treatment cycle or permanence sufficiently justified by authors) or not only from the target clinic (patients did not go to other clinics) (1 point) and that withdrawal was ascertained from secure records (i.e. medical records, 1 point).
[c]The 'comparability criterion' was met if all participants did treatment during the same period (i.e. data collection period was <5 years) (1 point); and sample was homogeneous regarding access to treatment (i.e. insurance coverage or number of subsidized cycles was described) or poor prognosis factors (i.e. mean age for all sample <40 or no statistical significant difference in age between groups) or type of treatment (all patients received the same treatment protocol), or IVF cycles excluded cryopreserved embryo transfer excluded (1 point).
[d]The 'follow-up criterion' was met if all cases were accounted for or completion rate (number of patients with outcome at follow-up divided by the number of patients that initiated) was >80% or description of patients lost to follow-up showed lack of bias (1 point).
The overall quality rating was the sum of met criteria (maximum seven). Quality ratings were grouped into low (0–3), moderate (4–5) and high (6–7) quality studies.

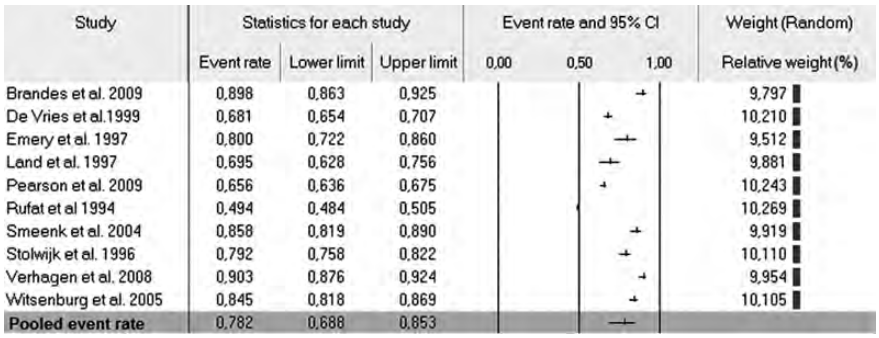


**Figure 2** Typical regimen compliance (event rate and 95% CIs) in ART treatment (TARC).

*Compliance and treatment success rates*

Two studies (Brandes *et al.*, 2009; Pearson *et al.*, 2009) did not report on the number of pregnancies achieved for the first three ART cycles and were not included in the calculation of the overall and typical regimen success rates. The overall success rate for the first three cycles, which included everyone who started treatment, was 42.7% (32.6–53.6%, $I^2 = 98{\cdot}8\%$). The typical regimen cycle success rate, which included only compliers (as defined in the TARC formula), was 57.9% (49.4–65.9%, $I^2 = 97.0\%$).

# Discussion

This meta-analysis shows that the vast majority of patients will comply with the typical ART regimen of three cycles, with about 2 of 10 patients discontinuing treatment earlier than would have been expected. Although many studies have pointed to alarmingly low compliance rates in ART (Malcolm and Cumming, 2004; Rajkhowa *et al.*, 2006), doctors can expect that 78% of patients will opt to undergo their ART regimen until they achieve pregnancy or are advised to

**Table IV** Compliance rates (typical and after the first or second failed cycle) according to subgroup analysis.

| Variables | | Typical ART regimen compliance (TARC) | | | | | Compliance after first failed cycle (CAF1) | | | | | Compliance after second failed cycle (CAF2) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *k* | Compliance rate | 95% CI LL | 95% CI UL | $\chi^2$ | *k* | Compliance rate | 95% CI LL | 95% CI UL | $\chi^2$ | | Compliance rate | 95% CI LL | 95% CI UL | $\chi^2$ |
| Clinical | | | | | | | | | | | | | | | |
| Population | | | | | 0.085 | | | | | 0.115 | | | | | 0.351 |
| General ART | 7 | 77.3 | 64.0 | 86.7 | | 6 | 80.9 | 69.5 | 88.7 | | | 73.7 | 62.7 | 82.3 | |
| Selected ART population | 3 | 80.2 | 60.2 | 91.6 | | 3 | 83.6 | 68.2 | 92.4 | | | 78.5 | 64.1 | 88.2 | |
| Geographic location | | | | | NA | | | | | NA | | | | | NA |
| Europe | 9 | 79.4 | 67.2 | 87.9 | | 8 | 82.6 | 72.0 | 89.8 | | | 76.6 | 67.1 | 84.1 | |
| USA | 1 | 65.6 | 22.6 | 92.6 | | 1 | 76.1 | 36.6 | 94.6 | | | 64.8 | 33.5 | 87.1 | |
| Patient | | | | | | | | | | | | | | | |
| Parity | | | | | NA | | | | | NA | | | | | NA |
| 0 | 1 | 80.0 | 72.2 | 86.0 | | 1 | 91.3 | 84.1 | 95.4 | | | 77.6 | 66.9 | 85.6 | |
| ≥1 child | 1 | 65.6 | 63.6 | 67.5 | | 1 | 76.1 | 73.9 | 78.1 | | | 64.8 | 61.6 | 67.9 | |
| Methodological | | | | | | | | | | | | | | | |
| Prospective design | | | | | 0.439 | | | | | 1.617 | | | | | 0.479 |
| Yes | 2 | 83.2 | 63.1 | 93.5 | | 2 | 89.2 | 74.3 | 95.9 | | | 79.4 | 64.5 | 89.1 | |
| No | 8 | 76.8 | 66.3 | 84.8 | | 7 | 79.4 | 69.3 | 86.8 | | | 74.1 | 66.3 | 80.7 | |
| Data available on number doctors-censored patients | | | | | **4.088*** | | | | | 0.642 | | | | | 3.341 |
| Yes | **5** | **84.2** | **75.5** | **90.2** | | 4 | 84.6 | 73.9 | 91.4 | | | 80.4 | 72.7 | 86.3 | |
| No | **5** | **70.6** | **58.3** | **80.5** | | 5 | 79.2 | 67.9 | 87.3 | | | 70.7 | 62.7 | 77.7 | |
| Length of follow-up | | | | | 0.007 | | | | | 0.267 | | | | | 0.651 |
| Twelve months or more | 5 | 77.0 | 61.6 | 87.4 | | 4 | 79.1 | 65.2 | 88.4 | | | 71.3 | 60.4 | 80.1 | |
| >12 months | 2 | 78.0 | 52.7 | 91.9 | | 2 | 83.9 | 66.0 | 933 | | | 77.9 | 63.6 | 87.6 | |
| Definition of initiated cycle | | | | | NA | | | | | NA | | | | | NA |
| Started hormonal stimulation | 9 | 80.5 | 73.7 | 85.9 | | 8 | 83.8 | 78.7 | 87.9 | | | 77.2 | 70.1 | 83.0 | |
| Had oocyte retrieval | 1 | 49.4 | 23.8 | 75.4 | | 1 | 57.7 | 35.8 | 76.9 | | | 59.2 | 35.4 | 79.4 | |
| Definition of treatment success | | | | | 0.141 | | | | | 0.905 | | | | | 0.148 |
| Live birth | 2 | 76.3 | 45.4 | 92.6 | | 2 | 83.8 | 61.7 | 94.3 | | | 75.4 | 53.8 | 88.9 | |
| Positive scan at 6/7 weeks | 5 | 77.1 | 58.8 | 88.9 | | 4 | 77.2 | 59.6 | 88.6 | | | 73.8 | 58.4 | 84.9 | |
| Positive βhCG urine/blood test | 3 | 81.1 | 58.4 | 92.9 | | 3 | 86.1 | 69.9 | 94.3 | | | 77.6 | 60.6 | 88.7 | |
| IVF cycles exclude cryopreserved embryo transfers | | | | | NA | | | | | NA | | | | | NA |
| No | 4 | 85.3 | 80.2 | 89.2 | | 4 | 87.4 | 83.4 | 90.6 | | | 83.2 | 80.7 | 85.4 | |
| Yes | 1 | 65.6 | 49.2 | 79.0 | | 1 | 76.1 | 73.9 | 78.1 | | | 64.8 | 61.6 | 67.9 | |
| Quality | | | | | 0.015 | | | | | 0.101 | | | | | 0.239 |
| High | 7 | 78.6 | 65.9 | 87.4 | | 6 | 81.0 | 70.3 | 88.4 | | | 74.0 | 63.5 | 82.3 | |
| Moderate | 3 | 77.3 | 56.4 | 90.0 | | 3 | 83.3 | 69.0 | 91.8 | | | 77.8 | 64.1 | 87.3 | |

*$P<$ 0.05, **$P<$ 0.01, *** $P<$ 0.001, *k* = number of studies, CI = confidence intervals, LL = lower limit, UL = upper limit, NA = not applicable because at least one of the subgroups only has one study, bold indicates $P<$ 0.05.

end treatment. Compliance is likely to decrease with ART failure, from 82% after the first failed cycle to 75% after the second failed cycle, but the decrease does not seem to be a function of the efficacy of the clinic. Compliance rates varied between 71 and 84% as a function of how compliance was defined (especially inclusion or exclusion of doctor-censored patients). Results suggest that a less rigorous definition of compliance may result in it being underestimated. To reach a definitive estimation of compliance in fertility treatment, researchers and practitioners need to reach consensus on the definition, monitoring and reporting of compliance. The chance of achieving a pregnancy for patients who initiated a typical three-cycle ART regimen was 43%, but 58% for those who complied. Patients need to be informed from the start of treatment of the possibility of facing a compliance decision (i.e. to continue treatment or not) and that chances of treatment success are optimal when people comply with recommendations.

The typical regimen ART compliance rate was 78% in a patient population who was expected to undergo treatment until they achieved pregnancy or were recommended to end treatment. It is reassuring for patients and clinics alike to realize that only about 2 of 10 patients do not comply with recommendations. Studies that reported on doctor censoring yielded even higher compliance rates. The reporting of active censoring is critical because it allows calculation of a compliance rate that takes into account whether the end of treatment was due to patient initiative or due to doctor recommendation. Including actively censored patients in the discontinuation group is misleading because these patients comply with medical recommendation. However, many studies do not consider this or other conceptual issues such as differentiation between permanent and temporary discontinuation or between definitive abandonment of treatment or of treatment at a given clinic only. This hinders research on compliance in ART, not only when assessing its prevalence but also when trying to understand its causes. This meta-analysis showed that when only the best available evidence is considered, compliance is 84%, supporting the idea that compliance in ART is indeed high.

The finding that compliance decreased with successive experience of unsuccessful cycles suggests that failure discourages couples from carrying on with treatment (Akyuz and Sever, 2009), maybe as a result of a subjective perception of poor prognosis or other factors such as cost. Although we could not do a subgroup analysis considering whether treatment was subsidized/reimbursed, the compliance rate when we considered only studies that clearly stated that treatment was subsidized/reimbursed was 84%. It may also be that ART is too demanding, an explanation consistent with patients' own stated reasons for discontinuation (Smeenk *et al*., 2004; Verhaak *et al*., 2007; Brandes *et al*., 2009; Boivin *et al*., 2012; Gameiro *et al*., 2012). As such, the compliance rate can also indicate that for 22% of couples the cost of treatment (financial, emotional) may be too high. It is relevant to note that the clinic's success rate per cycle (first and second cycles) was not associated with subsequent compliance, indicating that the clinic's efficacy does not dictate the compliance of their patient, despite strong beliefs within clinical communities that patients leave clinics with lower success rates (Marcus *et al*., 2005). It may be that patients disregard clinic success rates in favour of subjective perceptions of individual chances of success. It may also be that patients consider other outcomes beyond efficacy such as quality of care (van Empel *et al*., 2011) when considering uptake of further treatment.

Our results show that in every 100 typical couples starting ART treatment, 78 comply with three cycles and of these, 43 can expect to achieve pregnancy or live birth. However, if full compliance could be reached, 58 patients would achieve a pregnancy or live birth, which represents a 15% higher rate of success (if all other factors, including prognosis, are equal across three ART cycles). Therefore, addressing causes of non-compliance could help more people become parents, with a maximum estimated increase in success rates of 15%. In terms of number of treatment cycles, we would expect each clinic in Europe to carry out an additional 110 cycles per year if there was full compliance (based on European data: 402,039 cycles for 2007 in 1029 reporting clinics, excludes frozen embryo transfers, de Mouzon *et al*., 2012). Although a more precise knowledge of why patients discontinue treatment is still lacking, there are indications that to increase compliance clinics should focus on organizing treatments so that burden is diminished as much as possible and ensuring that patients receive support to meet the demands of treatment (see Gameiro *et al*., 2012 for reasons for discontinuation and Boivin *et al*., 2012 for an integrated model of fertility care). In addition, more explicit communication about compliance with patients and between health care providers is needed. Reports have shown that only 60% of women deciding to stop fertility treatment were satisfied with their decision (Peddie *et al*., 2004) and most felt they lacked the necessary information and counselling support (Peddie *et al*., 2005). Explicit information that ART success is likely to require multiple cycles and that treatment may entail emotional and physical side effects and disruptions to daily life, for example, would help address issues previously cited as causes of discontinuation (Rauprich *et al*., 2011; Boivin *et al*., 2012; Gameiro *et al*., 2012) and help patients have more realistic expectations of what a typical ART regimen entails for an optimal chance of pregnancy.

## Strengths and limitations

Considering the increasing debate surrounding the issue of compliance in fertility treatment, and in particular in ART, a meta-analysis on this literature was timely and appropriate. The strengths of this review are its systematic review of 30 years of research on discontinuation from seven databases, which yielded 10 studies from five countries, sampling the treatment trajectories of 14 810 patients. Data were independently extracted and quality evaluations made according to standard protocols for all studies. Compliance rates were calculated according to a clearly defined specification for the typical ART regimen and after-ART failure, which was consistent with ART practice and guidelines. Analytic methods included the overall meta-analysis and *a priori*-defined subgroup analyses according to relevant clinical, patient and methodological characteristics. Publication bias, including trim and fill, provided reliable estimates for 'missing' studies. Finally, although high heterogeneity in compliance rates was observed (above 95%), it was mainly due to statistical artefact and methodological issues. By statistical artefact, we mean that the majority of published meta-analyses report on effect sizes (e.g. risk ratios) from which it is statistically possible to remove the between-studies variance in base rates for the phenomenon under investigation (e.g. 1% difference in a base rate of 3 and 4% versus 80 and 81%). However, this is not the case in single-group studies and therefore meta-analyses of prevalence rates invariably produces high

heterogeneity (Borenstein et al., 2009, e.g. $I^2$ of 94% in a recent meta-analysis of the prevalence of depression in primary care, Mitchell and Sanjay Rao, 2009). All studies were published in peer-reviewed journals. They were of moderate to high quality and the quality of the studies was due to the fact that all used representative samples, and most studies could demonstrate homogeneity between compliers and non-compliers at the start of treatment and provided high completion rates for follow-up. The presence of publication bias for compliance after the first failed cycle (CAF1) did not markedly influence the magnitude of the rate reported (estimated to be 1.3% lower).

Despite these strengths, there were some limitations in primary research that were transmitted to the meta-analysis. In particular, the research does not provide a full account of patient progression through ART. Studies report on the proportion of patients who opted to undergo or stop treatment, but do not fully explain what then happened to patients registered as ending treatment at a particular clinic. These patients may have permanently ended treatment, as we assume, or they may have temporarily stopped or moved to another clinic. Analysis of the forest plot also revealed that one study (Rufat et al., 1994) presented a somewhat lower compliance rate with typical ART regimen than the other studies. This is one of the only two studies (Rufat et al., 1994; Stolwijk et al., 1996) that cover the pre-ICSI period when many causes of male infertility could not be addressed with treatment, which could explain the lower compliance reported. Studies focusing on groups of patients with poor prognosis or on specific treatments were excluded from analysis to control for clinical heterogeneity (i.e. use of specialist treatments, defined clinical subpopulations) so compliance in these groups is not known. Although these limitations need to be considered and addressed in the interpretation of the study findings and future research, the strengths of the systematic review and meta-analytic procedures adopted support the view that the compliance estimates reported are reliable and reflect current best available evidence.

### Conclusions and future research

Our results show that ~78% of patients undergo the cycles offered as part of the typical ART regimen, with uptake lower after ART failures but still high (82 and 75% after the first and the second failed cycles, respectively). These estimates are reassuring and should be transmitted to patients, who need to be informed from the start of treatment that, although ART is demanding, 8 out of every 10 patients comply with the typical regimen and that compliance with recommended cycles will offer the most optimal chance of success. Decision support should be developed to help people choose the best option (compliance, discontinuation) as ~22% will decide to end treatment for personal reasons and these patients need to be helped to reach equipoise about this decision. Future research should focus on trying to understand why patients discontinue treatment.

Despite these encouraging results, a definitive estimate of compliance may still be lacking because of primary research not providing a full account of patients' progression through the ART cycles. To progress compliance research, clinicians and researchers need to reach conceptual and methodological consensus on what is compliance and how to monitor it. An accurate assessment of compliance requires reporting the number of patients who undergo the typical ART regimen. While we studied three cycles, more or fewer ART cycles could be recommended depending on the patient population (e.g. poor responders) and ART protocol (e.g. minimal stimulation ART). In addition, patients who temporarily stop treatment, move on to another clinic or, as noted, are advised to end treatment should not be considered as non-compliers. In ART, there is no *a priori* time period in which the typical ART regimen should be completed. Most studies, therefore, set time limits for undergoing another cycle, typically 12 months, after which patients are considered to have abandoned treatment. These time limits should be evaluated for their representativeness of typical cycle uptake and, when used, reported. There is voluminous literature on success rates in ART yet few studies also report the number of patients opting not to undergo ART, which undermines the research base. Finally, it should be noted that the literature focuses exclusively on not undergoing the typical ART regimen (i.e. premature discontinuation). However, non-compliance can also occur when patients are advised to stop treatment but resist this idea (Boivin et al., 2005) and choose to continue ART at other clinics (i.e. over-persistence). This behaviour should also be monitored to reach an accurate estimation of the prevalence of 'over-persistence' and to obtain a better understanding of why couples are not able to follow recommendations to stop treatment. In summary, a precise assessment of compliance implies monitoring patients' long-term treatment trajectories. Such an endeavour requires the inclusion of compliance in national ART registers (e.g. in the UK the Human Fertilisation and Embryology Authority).

## Supplementary data

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

## Acknowledgements

We thank Laura Peronace, from Merck Serono SA, for her contribution to the conception and design of the larger review of compliance in fertility treatment. We thank Laura Peronace and Christos Venetis for providing feedback on the drafts of this manuscript.

## Authors' roles

S.G. did data extraction, critical appraisal, data analysis, data interpretation and writing of the report. J.B. and C.M.V. did critical appraisal, data interpretation and writing of the report. J.A.M.K. contributed to the writing of the report.

## Funding

S.G. holds a postdoctoral fellowship from the Portuguese Foundation for Science and Technology (FCT-SFRH/BPD/63063/2009). Merck Serono SA, Switzerland, funded the systematic literature search that was performed by Sure Support Unit for Research Evidence, Cardiff University and the second coding of articles that was performed by Debbie Moss and Caudex Medical. Merck Serono SA performed a scientific review of the publication but the views and opinions described in the publication are those of the authors.

# References


Akyuz A, Sever N. Reasons for infertile couples to discontinue in vitro fertilization (IVF) treatment. *J Reprod Infant Psychol* 2009;**27**:258–268.

Boivin J, Takefman J, Braverman A. Giving bad news: It's time to stop. In: Macklon N (ed). *Assisted Reproduction in the Complicated Patients*. London: Taylor & Francis, 2005.

Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum Reprod* 2007;**22**:1506–1512.

Boivin J, Domar AD, Shapiro DB, Wischmann T, Fauser BC, Verhaak CM. Tackling burden in ART: an integrated approach for medical staff. *Hum Reprod* 2012; **27**:941–950.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. West Sussex, UK: John Wiley & Sons, Ltd, 2009.

Brandes M, van der Steen JOM, Bokdam SB, Hamilton CJCM, de Bruin JP, Nelen WLDM, Kremer JAM. When and why do subfertile couples discontinue their fertility care? A longitudinal cohort study in a secondary care subfertility population. *Hum Reprod* 2009;**24**:3127–3134.

Brandes M, Hamilton CJCM, van der Steen JOM, De Bruin JP, Bots RSGM, Nelen WLDM, Kremer JAM. Unexplained infertility: overall ongoing pregnancy rate and mode of conception. *Hum Reprod* 2011;**26**:360–368.

de Mouzon J, Goossens V, Bhattacharya S, Castilla JA, Ferraretti AP, Korsak V. Assisted Reproductive Technology in Europe, 2007: results generated from European registers by ESHRE. *Human Reproduction* 2012;**27**:954–66.

de Mouzon J, Lancaster P, Nygren KC, Sullivan E, Zegers-Hochschild F, Mansour R, Ishihara O, Adamson D. World collaborative report on assisted reproductive technology, 2002. *Hum Reprod* 2009;**24**:2310–2320.

De Vries MJ, De Sutter P, Dhont M. Prognostic factors in patients continuing IVF treatment and dropouts. *Hum Reprod* 1998;**1998**:S54.

De Vries MJ, De Sutter P, Dhont M. Prognostic factors in patients continuing in vitro fertilization or intracytoplasmic sperm injection treatment and dropouts. *Fertil Steril* 1999;**72**:674–678.

Domar AD. Impact of psychological factors on dropout rates in insured infertility patients. *Fertil Steril* 2004;**81**:271–273.

Duval S, Tweedie R. Trim and fill: a simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;**56**:455–463.

Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997;**315**:629–634.

Emery JA, Slade P, Lieberman BA. Patterns of progression and nonprogression through in vitro fertilization treatment. *J Assist Reprod Genet* 1997;**14**: 600–602.

European Society of Human Reproduction and Embryology (ESHRE). *Good Clinical Treatment in Assisted Reproduction: An ESHRE Position Paper*. Beigem, Belgium: European Society of Human Reproduction and Embryology, 2008.

Gameiro S, Boivin J, Peronace LA, Verhaak CM. Why do patients discontinue fertility treatment? A systematic review of reasons and predictors of discontinuation in fertility treatment. *Hum Reprod Update* 2012:652–669.

Garrido N, Bellver J, Remohí J, Simón C, Pellicer A. Cumulative live-birth rates per total number of embryos needed to reach newborn in consecutive in vitro fertilization (IVF) cycles: a new approach to measuring the likelihood of IVF success. *Fertil Steril* 2011;**96**:40–46.

Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. *Br Med J* 2003;**327**:557–560.

Land JA, Courtar DA, Evers JL. Patients dropout in an assisted reproductive technology program: implications for pregnancy rates. *Fertil Steril* 1997; **68**:278–281.

Malcolm CE, Cumming DC. Follow-up of infertile couples who dropped out of a specialist fertility clinic. *Fertil Steril* 2004;**81**:269–270.

Marcus HJ, Marcus DM, Marcus SF. How do infertile couples choose their IVF centers? An internet-based survey. *Fertil Steril* 2005;**83**:779–781.

McDowell S, Murray A. Barriers to continuing in vitro fertilisation—why do patients exit fertility treatment? *Aus New Zealand J Obstet Gynaecol* 2011;**51**:84–90.

Mitchell A, Sanjay Rao AV. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009;**374**:609–619.

National Institute for Clinical Excellence (NICE). *Fertility: Assessment and Treatment for People with Fertility Problems*. London: NICE, 2004.

Pearson KR, Hauser R, Cramer DW, Missmer SA. Point of failure as a predictor of in vitro fertilization treatment discontinuation. *Fertil Steril* 2009;**91**:1483–85.

Peddie VL, van Teijlingen E, Bhattacharya S. Ending in-vitro fertilization: women's perception's of decision making. *Hum Fertil* 2004;**7**:31–37.

Peddie VL, van Teijlingen E, Bhattacharya S. A qualitative study of women's decision-making at the end of IVF treatment. *Hum Reprod* 2005;**20**:1944–1951.

Rajkhowa M, McConnell A, Thomas GE. Reasons for discontinuity of IVF treatment: a questionnaire study. *Hum Reprod* 2006;**21**:358–363.

Rauprich O, Berns E, Vollmann J. Information provision and decision-making in assisted reproduction treatment: results from a survey in Germany. *Hum Reprod* 2011;**26**:2382–2391.

Rufat P, Roulier R, Belaisch-Allart J, Bachelot A, de Mouzon J. Évolutions des critères pronostiques de fécondation in vitro selon le rang de la tentative. *Contracept Fertil Sex* 1994;**22**:282–286.

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, Stolwijk AM, Kremer JM, Braat DM. Reasons for dropout in an in vitro fertilization/intracytoplasmic sperm injection program. *Fertil Steril* 2004;**81**:262–268.

Stern JE, Brown MB, Luke B, Wantman E, Lederman A, Missmer SA, Hornstein MD. Calculating cumulative live-birth rates from linked cycles of assisted reproductive technology (ART): data from the Massachusetts SART CORS. *Fertil Steril* 2010; **94**:1334–1340.

Stolwijk AM, Hamilton CJCM, Hollanders JMG, Bastiaans LA, Zielhuis GA. A more realistic approach to cumulative pregnancy rate after in-vitro fertilization. *Hum Reprod* 1996;**11**:660–663.

The Practice Committee of the Society for Assisted Reproductive Technology and the Practice Committee of the American Society for Reproductive Medicine. Revised minimum standards for practices offering assisted reproductive technologies. *Fertil Steril* 2008;**90**:S165–S168.

Thompson C. *Making Parents: The Ontological Choreography of Reproductive Technologies*. Cambridge, MA: MIT Press, 2005.

van Empel IWH, Dancet EAF, Koolman XHE, Nelen WLDM, Stolk EA, Sermeus W, D'Hooghe TM, Kremer JAM. Physicians underestimate the importance of patient-centredness to patients: a discrite choice experiment in fertility care. *Hum Reprod* 2011;**26**:584–593.

Verhaak CM, Smeenk JM, Evers AWM, Kremer JM, Kraaimaat FW, Braat DM. Women's emotional adjustment to IVF: a systematic review of 25 years of research. *Hum Reprod Update* 2007;**13**:27–36.

Verhagen TEM, Dumoulin JCM, Evers JLH, Land JA. What is the most accurate estimate of pregnancy rates in IVF dropouts? *Hum Reprod* 2008;**23**:1793–1799.

Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. *The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomized Studies in Meta-analyses*. Ottawa, Canada: The Ottawa Health Research Institute, 2010.

WHO. *Adherence to Long-term Therapies: Evidence for Action*. Geneva, Switzerland: World Health Organization, 2003.

World Health Organization, The World Bank. *World Report on Disability*. Geneva: World Health Organization, 2011.

Witsenburg C, Dieben S, Van der Westerlaken L, BVerburg H, Naaktgeboren N. Cumulative live birth rates in cohorts of patients treated with in vitro fertilization or intracytoplasmic sperm injection. *Fertil Steril* 2005;**84**:99–107.

Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, Sullivan E, Vanderpoel S. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009. *Fertil Steril* 2009;**92**:1520–1524.