UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Outcome measures in reproductive medicine trials

van Rumste, M.M.E.

Publication date
2013

Link to publication

Citation for published version (APA):
van Rumste, M. M. E. (2013). *Outcome measures in reproductive medicine trials*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:18 Nov 2026

Chapter 2

Measuring outcomes in fertility trials - can we rely on clinical pregnancy rates?

Jane F. Clarke, Minouche M.E. van Rumste, Cindy M. Farquhar, Neil P. Johnson, Ben W.J. Mol, Peter Herbison

Fertility and Sterility 2010; 94:1647-51

ABSTRACT

OBJECTIVE To assess whether the estimates of treatment effect in randomized clinical trials (RCTs) in reproductive medicine differ when either clinical pregnancy or live birth is used as outcome measure.

DESIGN Meta-analysis.

SETTING We analyzed RCTs in reproductive medicine found in systematic reviews in the Cochrane Library that reported on both clinical pregnancy and live birth.

PATIENT(S) Subfertile couples.

INTERVENTION(S) For each individual RCT, data on clinical pregnancy and live birth were extracted.

MAIN OUTCOME MEASURE(S) We compared the outcome of each study by calculating a kappa-statistic (statistically significant treatment effective or not) and by comparing the odds ratio by calculating the ratio of the odds ratios (ROR).

RESULT(S) We found 67 systematic reviews, of which 42 reported on pregnancy and live birth. These 42 reviews included 654 RCTs, of which 143 (22%) reported both on pregnancy and live birth. The pregnancy loss rates in the treatment and control groups were comparable. Of the 143 RCTs, the conclusion based on pregnancy rate and live birth rate was comparable (kappa value of 0.81 (95% confidence interval [CI] .68 - .94)). The odds ratios estimating treatment effect from pregnancy and live birth were also comparable (ROR, 1.01, 95% CI 0.9 to 1.12).

CONCLUSION(S) Only a minority of randomized clinical trials in reproductive medicine report on live birth. Conclusions on the effectiveness of a treatment based on either clinical pregnancy or live birth as endpoints are comparable.

INTRODUCTION

The aim of reproductive medicine is to assist couples with an unfulfilled desire to become parents. Although the endpoint of a treatment for assisted reproduction is usually an ongoing pregnancy, such a pregnancy however does not guarantee the live birth of a baby (1). Recently, it has been proposed that singleton live birth rate should be the primary endpoint in subfertility trials (2). However, in some settings follow-up until live birth is an additional challenge for the investigator. When a third-line fertility treatment results in an ongoing pregnancy, the couple usually returns to a local gynecologist or midwife, thus limiting easy follow-up. In some countries, care for women with low-risk singleton pregnancies is provided in a primary care setting (3). Moreover, many trials are performed within limited budgets, thus hampering an additional follow-up time of 30 weeks, which is needed for a complete report on live birth. Therefore, a majority of subfertility studies report pregnancy instead of live birth as the endpoint.

The potential major drawback of using pregnancies as the endpoint of studies is the known pregnancy losses from the first trimester to live birth (4). First, a report of the absolute number of pregnancy rates would underestimate the number of couples that finally become parents. Second, differences in pregnancy loss rates between couples in two arms of an intervention trial would result in a difference in the estimates of treatment effect, because odds ratios calculated from pregnancy rates would be different from odds ratios calculated from live birth rates. If the latter were the case, odds ratios calculated from pregnancy rates will give false estimates of pregnancy rates.

Pregnancies from assisted reproduction technologies (ARTs) such as IVF and ovulation induction are known to have increased rates of multiple pregnancies. In clinics that do not have a single embryo transfer policy, the expected rates of multiple pregnancies following ART can be as high as 30–40% of pregnancies (5). Multiple pregnancies are associated with an increased risk of pregnancy loss, compared with singleton pregnancies. Therefore, it is of importance that subfertility trials report the birth of healthy children.

Unfortunately, the recent change from pregnancy to live birth as an endpoint is not based on clinical data. We hypothesized that pregnancy rates and live birth rates are not that different in subfertility studies. The aim of this study is to assess whether the use of either clinical pregnancy rates or live birth rates makes a difference in the statistical results and the conclusions of effectiveness of treatment.

MATERIALS AND METHODS

Search strategy

The Cochrane MDSG (Menstrual Disorders and Subfertility Group) clinical trial database contains over 4,300 subfertility trials. The integrity of Cochrane reviews is well recognised because they are also based on high-quality randomized controlled trials (RCTs), we decided to use their database for our search. We used the Web-based software that the Cochrane editorial groups use to support their editorial workflows (6). In January 2008, we conducted a search for all Cochrane systematic reviews in reproductive medicine that reported on both live birth and clinical pregnancy. Reviews and RCTs had to report odds ratios (ORs) with 95% confidence intervals (CIs) on both clinical pregnancy and live birth. The definition used for clinical pregnancy was an intrauterine gestational sac of at least 6 weeks' gestation. The definition of live birth was a pregnancy that resulted in the birth of at least one living baby. Two reviewers independently checked whether the reviews and RCTs reported these outcomes. In the case of disagreement, the judgment of a third reviewer was decisive. RCTs that combined ongoing pregnancies and live birth as one single outcome were not included. RCTs that were reported in more than one Cochrane review were included only once.

Statistical analysis

The systematic reviews were grouped into reviews of assisted reproductive technologies (ART) and non-assisted reproductivetechnologies (non-ART). The definition of ART was treatments and procedures involving the handling of human eggs and sperm or the use of fertility drugs for the purpose of establishing a pregnancy. In contrast, non-ART was defined as interventions that allow future natural conceptions, such as surgery or tubal flushing. Data extraction was done by two reviewers using the Cochrane software Review Manager 5 (Revman 5). We collected data on treatment allocation and on outcome (i.e. clinical pregnancy and live birth). First, we calculated the weighted mean rates of clinical pregnancy and live birth for treatment and control groups. This was done for all studies, as well as for ART and non-ART studies separately. The pooled treatment effect as estimated by studies from clinical pregnancy and live birth was expressed with an OR per study. Differences between estimates on treatment effect based on clinical pregnancy and live birth were expressed by calculating a ratio of the odds ratios (ROR) as previously described by Schulz (7). To calculate the ROR, we used the statistical program of Revman 5. The nearer the ROR is to 1, the smaller the differences in conclusion from studies reporting on clinical pregnancy and studies reporting on live birth as outcome.

Subsequently, the data were plotted in Bland-Altman graphs (8). In these graphs, the mean of the ORs of pregnancy and live birth is plotted on the x-axis, and the difference between the two ORs is plotted on the y-axis (live birth minus pregnancy). Points above the average

line have ORs for live birth that are higher than those for pregnancy. In the conventional Bland-Altman plot, each point has equal weight. We reported plotting symbols that are proportional to the size of the study to illustrate the weight of the included studies.

The presence or absence of concordance between outcomes in RCTs was also reported. RCTs were categorized as no evidence of a significant treatment effect if the 95% CI around the OR included 1, or as evidence of a significant treatment effect if the 95% CI around the OR did not include 1. After categorizing the RCTs in this way, conclusions based on clinical pregnancy and the conclusions based on live birth were compared in a two-by-two table. The agreement beyond chance between these conclusions was reported as a kappa statistic. A kappa value of 0 indicates no agreement beyond chance between the two ORs, whereas a kappa value of 1 indicates perfect agreement. Similarly as for RCTs, the concordance between Cochrane reviews reporting on clinical pregnancy and life birth was reported in the same manner as was done for RCTs—using kappa-statistics.

RESULTS

Search

From the Cochrane database, we identified 67 reviews of RCTs in reproductive medicine, of which 42 reported on both clinical pregnancy and live birth as an outcome. These 42 reviews contained a total of 654 RCTs. Four reviews containing 14 RCTs were excluded, because all the included RCTs combined ongoing pregnancy and live birth in one single outcome. Of the remaining 640 RCTs, 160 reported on live birth, of which 143 reported on both pregnancy and live birth outcomes (22%). Of these 143 RCTs, 111 reported on ART and 32 reported on non-ART.

Data

Table 1 shows the pooled data for clinical pregnancy and live birth as reported for treatment and control groups. For both treatment and control groups, a total of 1,219 pregnancies of the 6,512 clinical pregnancies (19%) were lost between diagnosis of clinical pregnancy and birth. There was a 5.4% difference between clinical pregnancy (30.3%) and live birth rates (24.9%) in the treatment groups, which represented a pregnancy loss rate of 18% (2780/3390). The difference between clinical pregnancy (27.9%) and live birth rates (22.4%) in the control groups was 5.5%, representing a pregnancy loss rate of 19.5% (2513/3122).

The Bland-Altman plots are shown for both ART (Fig. 1) and non-ART (Fig. 2). Both figures show that the majority of the included studies have a good correlation. The smaller studies with high ORs show less correlation between pregnancy and live birth. The mean difference in ORs was 0.1 for ART and –0.01 for non- ART studies. The ROR was 1.01 for

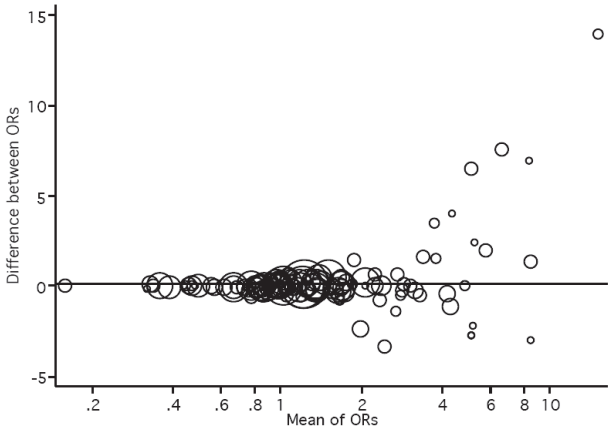

Figure 1 Bland-Altman plot of agreement between ORs for ART studies. The size of the circle is proportional to the size of the study. Points above the average line have ORs for live birth that are higher than those for pregnancy.

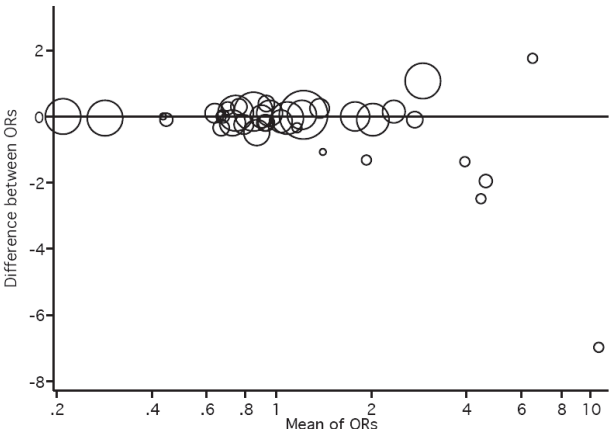

Figure 2 Bland-Altman plot of agreement between OR for non-ART studies. The size of the circle is proportional to the size of the study. Points above the average line have ORs for live birth that are higher than those for pregnancy.

all studies (95% CI, 0.9–1.12), 0.99 for ART studies only (95% CI, 0.87–1.13), and 1.03 for non-ART studies (95% CI, 0.86–1.23).

Of the 143 RCTs, 114 studies (80%) reported no evidence of significant treatment effect when based on either pregnancy or live birth rates. In 21 studies (15%), evidence of a significant treatment effect was reported. Five studies (3.5%) reported evidence of a significant treatment effect in pregnancy rate, but not in live birth rate. Three studies (2%) showed the opposite and reported evidence of significant treatment effect in live birth rate, but not clinical pregnancy rate. The kappa value indicating the agreement beyond chance was 0.81 (95% CI, 0.68–0.94).

Of the 63 subfertility Cochrane reviews, 41 (65%) reported no evidence of a significant treatment effect for pregnancy rates and live birth rates, and 19 (30%) reported evidence of significant treatment effect in both outcomes. Two reviews (3.2%) reported evidence of a significant treatment effect in pregnancy rate but not live birth rate, whereas one review (1.6%) reported evidence of significant treatmenteffect in live birth rate but not clinical pregnancy rate. The kappa value was 0.89 (95% CI, 0.77–1.00).

Table I. Treatment effect size for RCTs categorised as ART and non-ART

	ART / non-ART	No. of studies	Treatment group N/inclusion (%)	Control group N/inclusion (%)	OR for treatment effect (95% CI)
Clinical pregnancy rate	ART	111	2608/8277 (31.5%)	2256/7981 (28.3%)	1.2 (1.11-1.27)
	non-ART	32	782/2894 (27.0%)	866/3223 (26.9%)	1.0 (0.89-1.12)
	All	143	3390/11171 (30.3%)	3122/11204 (27.9%)	1.1 (1.07-1.20)
Live birth rate	ART	111	2207/8277 (26.7%)	1870/7981 (23.4%)	1.2 (1.13-1.31)
	non-ART	32	573/2894 (19.8%)	643/3223 (20.0%)	0.98 (0.87-1.12)
	All	143	2780/11171 (29.4%)	2513/11204 (22.4%)	1.2 (1.08 - 1.23)
			Difference between clinical pregnancy and live birth (%)	Difference between clinical pregnancy and live birth (%)	Ratio of the odds ratios (95% CI)
Difference in proportions for CPR and LBR	ART	111	4.8%	4.9%	0.99 (0.87-1.13)
	non-ART	32	7.2%	6.9%	1.03 (0.86-1,23)
	All	143	5.4%	5.5%	1.01 (0.9-1.12)

DISCUSSION

We found 67 systematic reviews, 42 of which reported on both outcomes. A minority of the RCTs reported both on clinical pregnancy and live birth (22%). Almost all included RCTs showed good correlation between the outcome pregnancy rates and live birth rates.

Strengths and Weaknesses

We extracted data from the Cochrane reviews and not the original RCTs. In this study, we relied on the authors and reviewers to have included good quality RCTs. However, the quality of RCTs in reproductive medicine has been found to be poor, notably in failing to report allocation concealment and live birth outcomes (9, 10). There is evidence of a progressive improvement in the quality of gynecological RCTs over the years (11). However, not all of the included Cochrane reviews have been updated, and some more recently published studies may have been overlooked. Less than 5% of all the available studies from the Cochrane MDSG register could be included. The definition of clinical pregnancy that we used was a gestational sac of at least 6 weeks' gestation. An outcome that may be a more reliable surrogate for live birth is a viable pregnancy (the presence of a fetal heart) or, even better, ongoing pregnancy (12 weeks' gestation with confirmation of a heartbeat). The definition of live birth was any birth after 20 weeks' gestation in which signs of life were present. A healthy child born at term would be a more preferable definition. We were reliant on the consistent use of definitions by the authors of the reviews.

Alignment with other findings

Couples undergoing fertility treatments want to know the likelihood of success, meaning at least one healthy, live infant (1). However, it is unusual to describe obstetric outcome in fertility studies. Most RCTs are powered to detect differences in pregnancy rates, but they are not powered to detect a significant difference in live birth rate or multiple pregnancy rates (12, 13). Although there is no debate that a multiple pregnancy increases the risk of many pregnancy complications, it is also important to consider that most multiple pregnancies end in the birth of two healthy babies. Consequently, as the studies are not powered to detect differences in twin rates, therefore they are not capable of detecting any differences in poor outcome relating to complications in pregnancy.

The application of surrogate outcomes in medical research has been previously debated (14). Surrogate outcomes are often used when observation of clinical outcomes requires long follow-up. A guiding principle for using surrogate outcomes is that there must be a valid reason to accept them. Usually, this is not believed to be the case in reproductive medicine, because it is not difficult to report live birth with adequate follow-up. The risk associated with making a link between any target outcome and its surrogate is that it is not valid. In this study, we have documented the extent of the link and reported a 5.4% difference between pregnancy rate and live birth rate. We want to stress that, in interventions that have interaction with miscarriage or pregnancy complications, it is important to consider differences in pregnancy rates and live birth rates carefully. For example, a treatment that increases multiple pregnancy rates, resulting in higher pregnancy risks, or a treatment that reduces pregnancy loss in the treatment group, such as metformin (15), will influence the ultimate outcome.

We found the pregnancy loss rate from clinical pregnancy to live birth rate to be almost 19%. A previous report on pregnancy loss after ART shows a total risk of pregnancy loss

Table 2. Discordant randomized control trials

Study	Pregnancy treatment events	Treatment total	Control events	Control total	OR M-H, Fixed (95%CI)
2.1 ART					
Gerris 199923	21	27	14	26	3.00 [0.91-9.87]
Levron 200225	8	46	20	44	0.25 [0.10-0.66]
Matorras 199621	31	47	17	41	2.74 [1.15-6.50]
Matorras 200226	56	255	47	260	1.28 [0.83-1.97]
Petersen 200528	10	40	3	40	4.11 [1.04-16.29]
2.2 NON ART					
Jinno 199722	12	32	3	31	5.60 [1.40-22.47]
Palomba 200427	31	55	39	54	0.50 [0.22-1.10]
Vandermolen 2001	6	12	1	15	14 [1.37-142.89]

of 29%, and 58% of all pregnancy losses occurred before 6 weeks' gestation (16). From 6 weeks'gestation onward, the risk of pregnancy loss ranged 10–45%, depending on female age and ART type (16). Pregnancy rates can therefore be used as primary outcome in fertility studies. The 19% pregnancy losses found in this study can be used to extrapolate pregnancy outcomes to live birth and help to counsel patients.

Disagreement with other findings

Pregnancy losses were greater in non-ART than in ART studies — 26.2% and 16.2%, respectively. This finding is not in agreement with the literature (17) and may be explained by the fact that the number of studies reporting non-ART interventions was limited. Another explanation might be that the embryos that survive the cleavage and blastocyst stages may be already self-selected for survival (18). Furthermore, there is evidence that twin pregnancies from ART result in relatively fewer pregnancy losses, compared with singleton pregnancies, than are usually reported (17, 19, 20).

We found discordant outcomes between pregnancy and live birth in eight studies (21–28), see Table 2. In these studies, the ORs of the treatment and control groups were in agreement with each other, but one of the ORs (treatment or control) was not statistically significant. Remarkably, in one study there was one more live birth than the number of pregnancies reported (28). We suspect that the authors of the review reporting this study did not correctly copy the data or that one twin birth was counted as two live births.

In conclusion, a minority of RCTs in reproductive medicine report on live birth. In RCTs, conclusions on treatment effect based on the endpoints (clinical pregnancy and live birth) are usually comparable, and the comparison between treatment groups will not be compromised. However, pregnancy losses should always be incorporated in clinical decision making, because live birth rate is the only outcome of real interest to the couple.

Live birth treatment events	Live birth total	Control events	Control total	OR M-H, Fixed (95%CI)
20	27	10	26	4.57 [1.42-14.71]
8	46	15	44	0.41 [0.15-1.09]
26	47	16	41	1.93 [0.83-4.53]
56	255	36	260	1.75 [1.11-2.77]
9	40	4	40	2.61 [0.73-9.32]
9	32	3	31	3.65 [0.88-15.08]
20	55	32	54	0.39 [0.18-0.85]
4	12	1	15	7.00 [0.66-73.93]

REFERENCES

1. Min JK, Breheny SA, MacLachlan V, Healy DL. What is the most relevant standard of success in assisted reproduction? The singleton, term gestation, live birth rate per cycle initiated: the BESST endpoint for assisted reproduction. Hum Reprod 2004;19:3-7.
2. Land JA, Evers JLH. Risks and complications in assisted reproduction techniques: Report of an ESHRE consensus meeting. Hum Reprod 2003;18:455-7.
3. Amelink-Verburg MP, Verloove-Vanhorick SP, Hakkenberg RM, Veldhuijzen IM, Bennebroek Gravenhorst J, Buitendijk SE. Evaluation of 280,000 cases in Dutch midwifery practices: a descriptive study. BJOG 2008;115:570-8.
4. Johnson NP. No more surrogate end-points in randomised trials - the PCOSMIC trial protocol for women with polycystic ovary syndrome using metformin for infertility with clomiphene. Aust N Z J Obstet Gynaecol 2006;46:141-5.
5. Human Fertilization and Embryology Authority. http://www.hfea.gov.uk/en/default.html
6. www.archie.cochrane.org
7. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408-12.
8. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
9. Helmerhorst FM, Belfield T, Kulier R, Maitra N, O'Brien P, Grimes DA. The Cochrane Fertility Regulation Group: synthesizing the best evidence about family planning. Contraception 2006;74:280-6.
10. Dias S, McNamee R, Vail A. Evidence of improving quality of reporting of randomized controlled trials in subfertility. Hum Reprod 2006;21:2617-2627.
11. Selman TJ, Johnson NP, Zamora J, Khan KS. Gynaecologic surgery from uncertainty to science: evolution of randomised controlled trials. Hum Reprod 2008;23:827-31.
12. Vail A, Gardener E. Common statistical errors in the design and analysis of subfertility trials. Hum Reprod 2003;18:1000-4.
13. Myers ER, McCrory DC, Mills AA, Price TM, Swamy GK, Tantibhedhyangkul J, et al. Effectiveness of assisted reproductive technology (ART). Evid Rep Technol Assess 2008;167:1-195.
14. Grimes DA and Schulz KF. Surrogate end points in clinical research: hazardous to your health. Obstet Gynecol 2005;105:1114-8.
15. Thatcher SS, Jackson EM. Pregnancy outcome in infertile patients with polycystic ovary syndrome who were treated with metformin. Fertil Steril 2006;85:1002-9.
16. Farr SL, Schieve LA, Jamieson DJ. Pregnancy loss among pregnancies conceived through assisted reproductive technology, United States, 1999-2002. Am J Epidemiol 2007;165:1380-8.
17. Tummers P, De Sutter P, Dhont M. Risk of spontaneous abortion in singleton and twin pregnancies after IVF/ICSI. Hum Reprod 2003;18:1720-3.
18. Los FJ, Van Opstal D, van den Berg C. The development of cytogenetically normal, abnormal and mosaic embryos: a theoretical model. Hum Reprod Update 2004;10:79-94.
19. La Sala GB, Nucera G, Gallinelli A, Nicoli A, Villani MT, Blickstein I. Lower embryonic loss rates among twin gestations following assisted reproduction. J Assist Reprod Genet 2005;22:181-4.
20. Lambers MJ, Mager E, Goutbeek J, McDonnell J, Homburg R, Schats R, et al. Factors determining early pregnancy loss in singleton and multiple implantations. Hum Reprod 2007;22:275-9.
21. Matorras R, Gorostiaga A, Diez J, Corcostegui B, Pijoan JI, Ramon O, Rodriguez-Escudero FJ. Intrauterine insemination with frozen sperm increases pregnancy rates in donor insemination cycles under gonadotropin stimulation. Fertil Steril 1996;65:620-5.
22. Jinno M, Katsumata Y, Hoshiai T, Nakamura Y, Matsumoto K, Yoshimura Y. A therapeutic role of prolactin supplementation in ovarian stimulation for in vitro fertilization: the bromocriptine-rebound method. J Clin Endocrinol Metab 1997;82:3603-11.

23. Gerris J, Neubourg D, Mangelschots K, Van Royen E, Vande Meerssche M, Valkenburg M. Prevention of twin pregnancy after in-vitro fertilization or intracytoplasmic sperm injection based on strict embryo criteria: a prospective randomized clinical trial. Hum Reprod 1999;14:2581-7.

24. Vandermolen DT, Ratts VS, Evans WS, Stovall DW, Kauma SW, Nestler JE. Metformin increases the ovulatory rate and pregnancy rate from clomiphene citrate in patients with polycystic ovary syndrome who are resistant to clomiphene citrate alone. Fertil Steril 2001;75:310-15.

25. Levron J. Shulman A, Bider D, Seidman D, Levin T, Dor J. A prospective randomized study comparing day 3 with blastocyst-stage embryo transfer. Fertil Steril 2002;77:1300-1.

26. Matorras R, Urquijo E, Mendoza R, Corcostegui B, Exposito A, Rodriguez-Escudero FJ. Ultrasound-guided embryo transfer improves pregnancy rates and increases the frequency of easy transfers. Hum Reprod 2002;17:1762-66.

27. Palomba S, Orio F Jr, Nardo LG, Falbo A, Russo T, Corea D, et al. Metformin administration versus laparoscopic ovarian diathermy in clomiphene citrate-resistant women with polycystic ovary syndrome: a prospective parallel randomized double-blind placebo-controlled trial. J Clin Endocrinol Metab 2004;89:4801-9.

28. Petersen CG, Mauri AL, Baruffi RL, Oliveira JBA, Massaro FC, Elder K, et al. Implantation failures: success of assisted hatching with quarter-laser zona thinning. Reprod Biomed Online 2005;10: 224-9.