

Statistical flaws in design and analysis of fertility treatment studies on cryopreservation raise doubts on the conclusions

P. H. A. J. M. VAN GELDER¹, M. NIJS²

¹ PSCT B.V., Probabilistic Solutions Consult & Training, Nootdorpse Landingslaan 292, 2496TB Den Haag, the Netherlands.

² The Geertgen Foundation, Fertility Treatment Centre, Elsendorp, the Netherlands.

Correspondence at: Dr. Ir. P.H.A.J.M. van Gelder, PSCT. B.V., Nootdorpse Landingslaan 292, 2496TB Den Haag, the Netherlands.

E-mail: P.H.A.J.M.vanGelder@tudelft.nl; Tel.: 31.15.3618930; Fax : 31.15.2785124

Abstract

Decisions about pharmacotherapy are being taken by medical doctors and authorities based on comparative studies on the use of medications. In studies on fertility treatments in particular, the methodological quality is of utmost importance in the application of evidence-based medicine and systematic reviews. Nevertheless, flaws and omissions appear quite regularly in these types of studies. Current study aims to present an overview of some of the typical statistical flaws, illustrated by a number of example studies which have been published in peer reviewed journals. Based on an investigation of eleven studies at random selected on fertility treatments with cryopreservation, it appeared that the methodological quality of these studies often did not fulfil the required statistical criteria. The following statistical flaws were identified: flaws in study design, patient selection, and units of analysis or in the definition of the primary endpoints. Other errors could be found in p-value and power calculations or in critical p-value definitions. Proper interpretation of the results and/or use of these study results in a meta analysis should therefore be conducted with care.

Key words: Statistics in infertility studies, subfertility, methodological study requirements, cryopreservation.

Introduction

The Cochrane Collaboration (Cochrane, 1989) was set up to provide the pharmacological industry, health care specialists and patients, high quality and independent information on the impact of health care interventions, by means of systematic reviews of randomised studies. Two phases in the systematic evaluation, among which quality assessment of study methods and the statistical pooling of results so called 'meta-analysis', depend on the quality of the study reports. Concern about the quality of study reports and the risk of bias has lead to consolidated standards (CONSORT; Consolidated Standards of Reporting Trials (<http://www.consort-statement.org/>)) which have been adopted by a lot of journals, to which their publications must satisfy. Although

CONSORT offers a useful framework, studies on fertility treatment need additional requirements to the design of the study and to the analysis of the study results. This study tries to give an overview of typical incorrect statistical analyses in recent fertility treatment studies with cryopreservation. We investigate if statistical flaws of the following type were made: flaws in study design, patient selection, and units of analysis or in the definition of the primary endpoints, p-value errors, power calculations, minimal important differences or critical p-value definitions. Although previous studies (Barlow, 2003; Dickey, 2003; Salim Daya, 2003; Vail and Gardener, 2003; Arce *et al.*, 2005) already addressed such flaws, current study will focus on specific studies, namely those studies where frozen – thawed embryo transfers were included in the study objectives.

Materials and Methods

The random selection of publications was based on a literature search for peer reviewed publications in Thomson Reuters Web of Knowledge containing the following keywords: fertility, frozen embryo transfer, frozen embryo replacement, frozen embryo cycle, FET and cryopreservation and pregnancy during a 14 year period (1995 till 2009). Moreover, the authors hand searched conference abstracts of major proceedings (e.g. ESHRE and ASRM) as well as reference lists of selected papers. Based on titles and abstracts of the search results, a set of 11 studies could be selected by the authors for the current evaluation.

All selected studies were published in international scientific journals, including Human Reproduction, Fertility and Sterility, Molecular and Cellular Endocrinology. Each study was examined for the following typical statistical flaws:

1. Study design

Double blind studies

If 2 treatments have a possible different effect or route of administration, then the only correct study design is a double blind, a double placebo or a placebo controlled research. However, from a medical-ethical point of view, a double blind placebo controlled study often has a limited feasibility. Moreover, adapting the appearance of treatments is generally technically difficult and expensive and therefore less applicable. A blind research is then most apparent, but it means in this case no definite conclusions can be drawn. In practice, in this type of study it is nearly impossible to ensure that the researcher remains blind for all important parameters of the research because a patient can describe at any time which treatment she undergoes purely by the description and appearance of the used medication. Other types of design flaws, e.g. lack of superiority, equivalence and non-inferiority, avoidance of co-intervention or treatment bias (when some subjects are receiving other (unaccounted for) interventions at the same time as the study treatment), occur frequently as well and have been discussed in depth by Daya (2006). Other types, but less suitable studies for fertility treatment investigation are:

- prospective randomised study
- retrospective studies (looking back in time).

Crossover studies

In a crossover study the examined persons are divided into two groups. The first group receives

firstly treatment A followed by treatment B, whereas the second group is treated in reversed order. An advantage of this research set-up is that the minimum required number of test persons, which is necessary to detect an effect, can remain relatively small. Studies of fertility treatments are special, in the sense that the treatment stops, once success (a pregnancy) is being reached. A direct consequence of such extreme form of 'carry-over' is that cross-over studies are unsuitable for fertility treatments (Senn, 1993). In other fields, however, cross-over studies may well be suitable.

ITT principle

In the intention to treat - analysis (ITT) patients are analysed in the arm in which they have been classified initially and where drops outs are counted. If one does not count the persons who drop out, it leads, in general, to an over-estimate of the treatment effect. In this case one compares only those patients, who have continued the treatment (and who, for example, did not have too many side effects). One frequently uses for that a technique where the last measured value is counted as an endpoint (English: last observation carried forward).

2. Patient selection

The most used selection criteria in fertility treatment studies are age of the female patient, her ovulatory status and the body mass index (BMI). Also the number of previous IVF attempts is an important selection criterion. Since patients are seen in all kinds of stages of treatment, this can be accepted in medical practice. It is known, however, that the result of IVF decreases with the number of earlier attempts (Templeton *et al.*, 1996); patients with the most favourable forecast become pregnant more rapidly.

This can cause a possible bias of the results if the two groups are not divided equally by the number of previous IVF attempts. These type of important prognostic confounders must be taken into account in the study design, by means of stratification in assigning the treatment to the patient (by randomization of the medicines in subgroups with the same number of preceding IVF/ICSI attempts), or in the statistical analysis. Apart from achieving prognostic factor balance by stratification prior to randomization, it can also be achieved by the technique of minimization. At the point of assignment of each new patient to one of a number of treatments, minimization involves calculating for each treatment group the comparative degree of imbalance that would occur if the patient were assigned to that group (McEntegart, 2003).

3. 'Unit of analysis' flaws

Simple group comparing tests, such as the t-test or Mann - Whitney test for continuous data and χ^2 or Fisher's test for categorical data, require that the observations are statistically independent. At the allocation of test persons in several arms of the study, this will generally mean that only one observation per patient has been incorporated in such analysis. The hierarchical character of sub fertility data with, for example, several oocytes, embryos and several implants per treatment cycle, and several treatment cycles per woman, can lead rapidly to unit of analysis flaws. Use of several observations per woman leads to unforeseeable biases in the estimate of the treatment effect differences. It can lead to unjust narrow confidence intervals and low p-values.

4. Primary endpoints in the study

The primary outcome of sub fertility studies must be preferably live births, such as the baby take-home rate and the cumulative baby take-home rate. Side effects in fertility treatments can lead to flaws in the statistical analysis. Ovarian hyper-stimulation syndrome is a typical side effect, which can be considered as a treatment error. Other important side effects, such as ectopic pregnancy and aborted pregnancies only appear after a partially technical success to reaching a pregnancy. These events lead however to two methodological consequences. In the first place it is usual to report pregnancy-related side effects as a percentage of the pregnancies, instead of as a percentage of the randomised women. Such an approach loses however the advantages of randomised comparison. It can also be misleading, because it is possible to have a higher percentage miscarriage per pregnancy in the group with the lowest percentage miscarriage per woman.

5. p-Value calculation

The p-value or exceedance probability (of a given sample outcome) is the probability that the value of the test statistic is exceeded (left, right or two-tailed) given the distribution by the null hypothesis. The p-value indicates how far extreme the observed value is for the test statistic in the distribution of the null hypothesis: the smaller the p-value, the more extreme the outcome. In practice values of 5% and 1% are used as a border. P-values are mostly calculated in fertility treatment studies with the Cochran Haenszel test or Z-test. The Cochran-Mantel-Haenszel test is a test χ^2 that examines if an association between two variables after control for other variables is present. The test measures the strength

of this association. A Z-test is mostly applied on proportions where the test statistic is assumed as a normal distribution, which generally holds well for large sample sizes (Bland, 2000). Calculated p-values are smaller or larger than the significance level, but never equal to the significance level. Regularly, published studies show p-values exactly equal to 5%. Mostly, in such cases the calculated p-value has been rounded down.

6. Power calculations

Power calculations are calculations to determine the minimum sample size required for effective statistical significant differences between groups in a study. The sample size depends on the effect size which one expects, as well as on the probability which is required to find a result which is present in the population (the power). Frequently, power calculations in fertility treatment studies are, wrongfully, retrospectively carried out or are even absent.

7. Critical p-values in sequential studies

Sequential studies are usual practice in biomedical research as a result of ethical, administrative and economic reasons. Statistical hypotheses in such studies are repeatedly, at several times, tested, after a new group of new observations has been completed.

The analysis of the results takes place before the final number of experimental units has been reached. If this happens in an uncontrolled manner, the term peeking is used. A statistical fine must be applied in such case, because if enough interim analyses are carried out, and if the result of the statistical test lies on the border between significant and not significant, eventually one of the analyses, will result, wrongfully, in a $p < 0.05$. Sequential statistical methods include not only finding suitable methods for the provision of critical p-values on each interim test control, but also developing efficient inferential procedures for secondary analyses, such as parameter estimates, confidence interval calculations, etc. Techniques for sequential analysis, where data continues to accumulate, are available in literature. However, the practice is that such analyses are generally carried out without making arrangements for the necessary adaptations of the type I-errors. A simple method for correction of the critical p-value (although severe) is the use of the Sidak inequality (Sidak, 1967). This results in an adapted critical p-value which is given by the formula $1 - (1-p)^k$, where k the number of interim analyses and p the nominal critical p-value (generally 0.05). The reason for this is that the correction is arranged for several compar-

Table 1. — Critical p-values in sequential studies in fertility studies including frozen-thawed embryo transfers.

Number of interim analyses	Nominal critical p-value	Corrected critical p-value according to Armitage McPherson (1969)	Corrected critical p-value according to Sidak (1967)
1	0.05	0.05	0.05
2	0.05	0.025	0.03
5	0.05	0.01	0.016

isons when these are entirely independent. But successive interim analyses are mostly not completely independent of each other, but to a certain degree only. For this reason the Sidak adaptation gives a lower bound for the critical p-value. The Armitage-McPherson adaptation (Armitage *et al.*, 1969) is a less strict adaptation than the Sidak correction (Sidak, 1967) (Table 1).

8. Minimal important differences (MID)

One other methodological flaw is the absence of the definition of 'minimal important differences', being the smallest benefit of treatment that would result in clinicians recommending it to their patients. The MID is necessary to calculate sample size for randomized clinical trials, but its chosen value is often arbitrary. Power calculations can be performed to calculate a statistical difference given a defined alpha-error (incorrectly accepting that a difference exists between the two treatments) and beta-error (incorrectly accepting that no difference exists between the two treatments), but the question how big should the difference really be to be clinically relevant is frequently missing. This issue is very important to be defined prior to embarking on a trial and sophisticated statistics. Van Walraven *et al.* (1999) investigated the practicability of surveying physicians to elicit the MID for clinical trial sample-size calculation.

Results

Table 2 represents a detailed overview of the analysis for statistical flaws of 11 randomly selected fertility studies which include frozen-thawed embryo transfer analysis. Most of the 11 selected studies contain flaws in patient selection: patients were not randomly selected or they were incorrectly followed over the period of several cycles. In some studies, flaws in primary endpoints were observed. Other errors are reported as purely statistical flaws (unit of analysis errors, misapplication of cross-over design, technical errors in power or significance calculations), issues related to clinical preference (eligibility criteria, choices of primary outcome) and study

design issues (blinding, ITT, randomisation). All studies deal incorrectly with the units of analyses (namely cycles are used instead of patients). Also unsuitable retrospective designs are used in the majority of the studies. Furthermore, in only 1 study preceding power calculations had been carried out. In none of the studies investigated, minimal important differences have been presented, which is a major flaw and underestimated in its importance. Finally, adjustments of critical p-value in sequential studies were missing.

Discussion

Particularly in fertility treatment studies, the methodological quality is very important for the application in evidence-based medicine and systematic reviews. Nevertheless errors and omissions occur in these studies regularly. In this article an overview has been given of the most appearing statistical flaws. The seriousness and the impact of these flaws differ per study. A certain study even wrongly rejected the null hypothesis with respect to the most important reported parameter. The flaws that were identified put therefore doubts at the conclusions of these specific studies. It is of utmost importance that when studies are being registered and set up, the primary and secondary endpoints should be fixed. Also power calculations must be discussed. The correct application of medical statistics in reproduction studies is very important, but unfortunately in practice, it is not always conducted well. Although, it is not incorrect to publish a case-series or comparative cohort study of an intervention, even though the design is not as strong as a randomized controlled trial (RCT), the study cannot be interpreted with the same causal inference. RCT is the gold standard for evaluating the effectiveness/ efficacy of interventions. All other study designs where no random sequence generation is used are at risk of bias, leading us away from drawing correct conclusions from the studies.

Previous authors, such as Barlow (2003) Dickey (2003), Salim Daya (2003), Vail and Gardener (2003) and Arce *et al.* (2005), have addressed some of the above issues mentioned above. Current study has reconfirmed that statistical flaws still occur in

Table 2. — Overview of statistical flaws in 11 studies on fertility studies including transfer of frozen-thawed embryos. Studies were chosen at random. ^aMID: Minimal important differences.

Study number	Author	Subject of the study	Flaws in study design	Flaws in patient selection	Flaws in unit of analysis	Flaws in primary endpoints	MID ^a	Flaws in p-value and power calculations
1	Francsovits <i>et al.</i> (2009)	Use of human derived FSH versus recombinant FSH results in more embryos to be cryopreserved.	Less suitable prospective randomized assessor-blind study.	Patients were followed over several cryo embryo transfer cycles.	Data of cycles (not of women) were presented.	Clinical pregnancy rate was primary endpoint.	Not reported.	Statistical significance calculated with Mann-Whitney U-test and X2-analysis at end of study. Power calculations missing.
2	Gelbaya <i>et al.</i> (2006)	Cryopreserved-thawed embryo transfer in natural or down-regulated hormonally controlled cycles.	Unsuitable retrospective study.	Patients not randomly selected.	Data per woman presented.	Implantation rate, pregnancy rate, and number of live births per cycle en per embryo transfer (ET).	Not reported.	Statistical significance calculated with t-test and X2-analysis at end of study. Logistic regression analysis. Power calculations missing.
3	Kahn <i>et al.</i> (1999)	Study of either recombinant FSH (Puregon) or urinary FSH (Metrodin) in in vitro fertilization treatment.	Less suitable prospective randomized assessor-blind study.	Patients were followed over several cryo embryo transfer cycles.	Data of 3 IVF cycles per patient were studied.	Cumulative pregnancy rate was primary endpoint.	Not reported.	Statistical significance calculated with X2-analysis at end of study. Power calculations missing.
4	Oehninger <i>et al.</i> (2000)	Impact of different clinical variables on pregnancy outcome following embryo cryopreservation.	Unsuitable retrospective study.	Patients not randomly selected.	Data of cycles (not of women) were presented.	Cumulative pregnancy rate was primary endpoint.	Not reported.	Statistical significance calculated with Student's test, X2-analysis (with Yates' correction) and two-by-three contingency tables at end of study. Power calculations missing.
5	Out <i>et al.</i> (1995)	Study comparing recombinant and urinary follicle stimulating hormone (Puregon versus Metrodin) in in-vitro fertilization.	First part of study not double blind; second part of study was open.	Patients with a first IVF-attempt and patients with a maximum of three failed IVF-attempts were unjust included.	Use of multiple observations per woman.	Cumulative pregnancy rate (including returned frozen thawed embryo's) was not defined as primary or secondary endpoint, but still presented as primary endpoint.	Not reported.	Calculated p-Value exactly equal to 0.05 (according to Cochrane-Mantel-Haenszel test). Unjust statistical significance is concluded. p-Value should be smaller than 0.05 before null hypothesis can be rejected. Because of multiple interim-analyses, adjusted critical p-values of Sidak (1967) or Armitage - McPherson (1969) should be used. Power calculations missing.
6	Prades <i>et al.</i> (2009)	Analysis of cumulative pregnancy rates by freezing and thawing single embryos.	Unsuitable retrospective study.	Patients not randomly selected.	Data of cycles (not of women) were presented.	Implantation and pregnancy rate after fresh ETs and embryo survival - and pregnancy rate after FET.	Not reported.	Statistical significance calculated with nonparametric analysis of variance, Kruskal-Wallis tests, followed by pairwise comparisons with two-sample Wilcoxon tests, X2-tests and Fischer's exact tests. Univariate logistic regression analysis. Two stepwise multivariate logistic regression analysis. Wald's test. Power calculations missing.

7	Salumets <i>et al.</i> (2006)	Implications of clinical and embryological factors on the pregnancy outcome of frozen embryo transfers.	Unsuitable retrospective study.	Patients were followed over several cryo embryo transfer cycles.	Data on embryo transfers are not converted to the unit of woman.	Pregnancy - and clinical pregnancy rate were the primary endpoints.	Not reported.	Statistical significance calculated with Fisher's exact test, backward logistic regression and X2 at end of study. Power calculations missing.
8	Seelig <i>et al.</i> (2002)	Comparison of cryopreservation outcome with gonadotropin-releasing hormone agonists or antagonists in the collecting cycle.	Unsuitable retrospective study.	Patients not randomly selected.	Data of cycles (not of women) were presented.	Implantation-, pregnancy-, miscarriage rates.	Not reported.	Statistical significance calculated with Fisher's exact test, t-test and X2 at end of study. Power calculations missing.
9	El-Toukhy <i>et al.</i> (2004)	Pituitary suppression in ultrasound-monitored frozen embryo replacement cycles.	Less suitable prospective randomized assessor-blind study.	Patients were unjust followed over several cryo embryo transfer cycles.	Data of cycles (not of women) were presented.	Pregnancy - and clinical pregnancy rate were the primary endpoints.	Not reported.	Statistical significance calculated with Fisher's exact test, t-test and X2 at end of study. Power calculations have been presented in advance.
10	Wang <i>et al.</i> (2001)	Frozen-thawed embryo transfer: influence of clinical factors on implantation rate and risk of multiple conception.	Unsuitable retrospective study.	Patients not randomly selected.	Data of cycles (not of women) were presented.	Overall implantation rate.	Not reported.	Statistical significance calculated with Fisher's exact test and X2 at end of study. Power calculations missing.
11	Ziebe <i>et al.</i> (2007)	Influence of ovarian stimulation with HP-hMG or recombinant FSH on embryo quality parameters in patients undergoing IVF.	Correct randomized, assessor-blind, multinational trial.	Patients were unjust followed over several cryo embryo transfer cycles.	Data of cycles (not of women) were presented.	Ongoing pregnancy rate was the primary end-point.	Not reported.	Statistical significance calculated with logistic regressions and ANOVA at end of study. Power calculations missing.

recent studies. Methodological flaws in the study design compromise the internal and external validity of the results and conclusions. A plea is made to implement the publication of peer-reviewed study protocols before embarking on a trial to increase the quality of the studies on fertility treatment. In order for studies to be analysed correctly, researchers should receive post-academic courses on statistical theory and design and analysis of studies. Clinical trials are to be registered with one of the International Committees of Medical Journal Editors' recognised trial registers at the time of their inception. This registration process should include defin-

ing the type of statistical analysis to be used for that specific trial as well as the registration of the power analysis. Journals are advised to focus more on the statistical soundness of submitted papers, and should perhaps by default send each accepted manuscript to a specialised reviewer for analysis of the statistical soundness of that specific study. Reviewers should be trained to identify possible errors in the study design and statistical analysis in submitted manuscripts. A combination of such actions might help to reduce the occurrence of statistical flaws in research and will hence result in the publication of solid research papers.

A Call for action

As an 'exercise', we challenge the readers of this journal in identifying possible flaws in study design and statistical analysis of the paper of Zhang *et al.* (2006) that investigates the effect of traditional Chinese herbs combined with low dose human menopausal Gonadotropin applied in frozen thawed embryo transfer (Chin J Integr Med. 2006;12, 244-49). This paper has been randomly chosen and will fit the above purpose. In the next issue of this journal, the possible flaws associated to this 12th paper will be presented by the authors and can be compared with the readers' observations.

References

- Arce JC, Nyboe Andersen A, Collins J. Resolving methodological and clinical issues in the design of efficacy trials in assisted reproductive technologies: a mini-review. *Hum Reprod.* 2005;20:1757-71.
- Armitage P, McPherson CK, Rowe BC. Repeated significance tests on accumulating data. *J Royal Stat Soc A.* 1969;132:235-44.
- Barlow DH. The design, publication and interpretation of research in Sub fertility Medicine: uncomfortable issues and challenges to be faced, *Hum Reprod.* 2003;18:899-901.
- Bland M. *An Introduction to Medical Statistics*, Oxford University Press. 2000.
- Cochrane AL. *Effectiveness and Efficiency: Random Reflections of Health Services*, Oxford University Press. 1999.
- Daya S. Methodological issues in infertility research. *Best Practice & Research Clinical Obst Gyn.* 2006;20(6):779-97.
- Dickey RP. Clinical as well as statistical knowledge is needed when determining how subfertility trials are analysed. *Hum Reprod.* 2003;18,2495-8.
- El-Toukhy T, Taylor A, Khalaf Y *et al.* Pituitary suppression in ultrasound-monitored frozen embryo replacement cycles. A randomised study. *Hum Reprod.* 2004;19:874-9.
- Francsovits P, Tothne A, Murber A *et al.* Use of human derived FSH versus recombinant FSH results in more embryos suitable for cryopreservation. P-030, Proceedings of the 15th World Congress on IVF, International Society for in Vitro Fertilization, 2009.
- Gelbaya TA, Nardo LG, Hunter HR *et al.* Cryopreserved-thawed embryo transfer in natural or down-regulated hormonally controlled cycles: a retrospective study. *Fertil Steril.* 2006;85:603-9.
- Kahn JA, Sunde A, von Düring V *et al.* A prospective randomized comparative cohort study of either recombinant FSH (Puregon) or urinary FSH (Metrodin) in *in vitro* fertilization treatment. *Middle East Fertil Soc J.* 2009;4:206-14.
- McEntegart D.J. The Pursuit of Balance Using Stratified and Dynamic Randomization Techniques: An Overview. *Drug Infor J.* 2003;37:293-308.
- Oehninger S, Mayer J, Muasher S. Impact of different clinical variables on pregnancy outcome following embryo cryopreservation. *Molec Cell Endo.* 2000;169:73-7.
- Out HJ, Mannaerts BM, Driessen SG *et al.* A prospective, randomized, assessor-blind, multicentre study comparing recombinant and urinary follicle stimulating hormone (Puregon versus Metrodin) in *in-vitro* fertilization. *Hum Reprod.* 1995;10,2534-40.
- Prades M, Golmard JL, Vauthier D *et al.* Can cumulative pregnancy rates be increased by freezing and thawing single embryos? *Fertil Steril.* 2009;91:395-400.
- Daya S. Pitfalls in the design and analysis of efficacy trials in subfertility. *Hum Reprod.* 2003;18:1005-9.
- Salumets A, Suikkari AM, Mäkinen S *et al.* Frozen embryo transfers: implications of clinical and embryological factors on the pregnancy outcome. *Hum Reprod.* 2006;21:2368-74.
- Seelig AS, Al-Hasani S, Katalinic A *et al.* Comparison of cryopreservation outcome with gonadotropin-releasing hormone agonists or antagonists in the collecting cycle. *Fertil Steril.* 2002;77:472-5.
- Senn S. *Cross-over trials in clinical research.* Wiley and Sons. 1993.
- Sidak Z. Rectangular confidence regions for the means of multivariate normal distributions. *J Amer Stat Assoc.* 1967;62:626-33.

- Templeton A, Morris JK, Parslow W. Factors that affect outcome of *in-vitro* fertilisation treatment. *Lancet*. 1996;348:1402-6.
- Vail A, Gardener E. Common statistical errors in the design and analysis of subfertility trials. *Hum Reprod*. 2003;18:1000-4.
- van Walraven C, Mahon JL, Moher D *et al*. Surveying physicians to determine the minimal important difference: implications for sample-size calculation. *J Clin Epidemiol*. 1999; 52:717-23.
- Wang JX, Yap YY, Mathews CD. Frozen-thawed embryo transfer: influence of clinical factors on implantation rate and risk of multiple conception. *Hum Reprod*. 2001;16:2316-9.
- Ziebe S, Lundin K, Janssens R *et al*. Influence of ovarian stimulation with HP-hMG or recombinant FSH on embryo quality parameters in patients undergoing IVF. *Hum Reprod*. 2007;22:2404-13.
- Zhang HQ, Zhao HX, Gu DY *et al*. Effect of Traditional Chinese Herbs Combined with Low Dose Human Menopausal Gonadotropin Applied in Frozen thawed EmbryoTransfer. *Chin J Integr Med*. 2006;12:244-9.