Human Reproduction Vol.23, No.1 pp. 187–192, 2008 doi:10.1093/humrep/dem352
Advance Access publication on October 31, 2007

Estimating the success of an *in vitro* fertilization programme using multiple imputation

N. Soullier[1,2,3,6], J. Bouyer[1,2,3], J.L. Pouly[4], J. Guibert[5] and E. de La Rochebrochard[1,2,3]

[1]INSERM, U822, 82 rue du Général Leclerc, F-94276 Le Kremlin-Bicêtre Cedex, France; [2]INED, 82 rue du Général Leclerc, F-94276 Le Kremlin-Bicêtre Cedex, France; [3]Faculté de Médecine Paris-Sud, Université Paris-Sud 11, F-94276 Le Kremlin-Bicêtre Cedex, France; [4]Unité de FIV, CHU de Clermont-Ferrand, F-63003 Clermont-Ferrand, France; [5]Unité de Médecine de la Reproduction, Service de Gynécologie-Obstétrique II, Hôpital Cochin AP-HP, F-75014 Paris, France

[6]Correspondence address. Tel: +33-1-45-21-23-43; E-mail: soullier@vjf.inserm.fr

BACKGROUND: The outcome of *in vitro* fertilization (IVF) has been widely investigated over the last 30 years, but evaluation was mostly based on pregnancy rate per oocyte retrieval. Our objective was to estimate the cumulative live birth rate after four IVF aspirations, using multiple imputation that takes into account treatment interruptions. METHODS: We analysed data from 3037 couples beginning IVF treatment between 1998 and 2002 in two French IVF units. Multiple imputations were used at each aspiration to impute the IVF outcome (delivery or not) for couples who interrupted treatment. The global success rate after four aspirations was then computed. RESULTS: At the first aspiration, 21% of couples obtained a live birth and 24% discontinued treatment. The multiple imputation method provided an estimated cumulative live birth rate at each aspiration as if no couple discontinued treatment: 35% at the second aspiration and 41% at the third. The cumulative success rate after four aspirations was estimated at 46% (95% CI: 44–48%). CONCLUSIONS: Multiple imputation is a promising method for estimating the cumulative success rate of IVF. It could provide new insight on IVF evaluation and should be tested in further studies.

Keywords: IVF; cumulative success rate; multiple imputation

Introduction

About 9–14% of couples have difficulty in conceiving a child (Thonneau *et al*., 1991; Boivin *et al*., 2007). To address these subfertility problems, assisted reproductive technology (ART), *in vitro* fertilization (IVF) in particular, has developed. Following Min Chueh Chang's application of IVF to animals in 1959, the technique was developed for humans by Patrick Steptoe and Robert Edwards in the UK (Steptoe and Edwards, 1978). The first "test-tube baby", Louise Brown, was born in Oldham, UK, on 25 July 1978. Nowadays, >41 000 IVF cycles are performed annually in France, 46 500 in Germany, 24 500 in the UK and 63 500 in the USA (Adamson *et al*., 2006). Averaging data from 49 countries, the world collaborative report estimated that pregnancy rate was 27% and delivery rate was 19% (Adamson *et al*., 2006). In France, these percentages were 24% for pregnancies and 18% for deliveries (Adamson *et al*., 2006).

To date, there is no single simple answer as to how success in assisted reproduction should be measured. An important debate took place on this issue in Human Reproduction in 2004. The original article (Min *et al*., 2004) encouraging programmes to report the Birth Emphasizing a Successful Singleton at Term (BESST) gave rise to much discussion. Furthermore, most reported success rates are based on the number of pregnancies or deliveries per aspiration cycle, excluding results from frozen embryo transfers (FETs).

Instead of considering each aspiration, a different approach would be to consider the whole run from the first to the last aspiration in the assisted conception unit. Using this measure, we would be able to inform couples starting IVF treatment of their chances of having a child. This approach requires taking into account the fact that some couples abandon treatment without a successful pregnancy. Classically, life table analysis is used to estimate a cumulative success rate. It assumes that couples discontinuing treatment have the same chance of having a child as couples who continue. Sharma *et al*. (2002) retrospectively analysed data from 2056 patients undergoing 2708 IVF cycles from April 1992 to March 1999 in the assisted conception unit of a university hospital in the UK. After four aspirations, the estimated cumulative success rate obtained with a life table analysis reached 75% for clinical pregnancies and 66% for live births.

Using questionnaires sent to 1510 couples who had undergone IVF treatment at Ninewells Hospital and Medical School in Dundee (UK) between January 1995 and December 2001, Rajkhowa *et al*. (2006) estimated that 25% of couples not

pregnant after the first aspiration discontinued treatment. This proportion could be similar in countries where state funding exists. In a retrospective cohort study of 202 couples applying for their first ART treatment cycle between July 1993 and December 1994 in a university hospital-based tertiary care fertility clinic in The Netherlands (where costs are covered by health cost insurance), Land *et al*. (1997) estimated that after one aspiration without achieving pregnancy, 26% of couples interrupt the treatment. Sharma *et al*. (2002) estimated that this proportion reaches 64% in the UK where state funding is low. Moreover, they noted that patients who discontinued treatment after the first aspiration differed from those who had a second aspiration in age, number of oocytes retrieved and number of embryos transferred: women who discontinued treatment were more likely to be older than 35 years, to have five or fewer oocytes retrieved and to have two or fewer embryos transferred. Thus, it seems likely that couples discontinuing treatment have fewer chances of having a child than couples continuing treatment, and we cannot disregard the large number of couples discontinuing treatment before they in fact conceive. Such differences raise doubts about the basic assumption of life table estimates, which is that women discontinuing treatment would have the same chance of having a child as those who continue.

The aim of this work is to apply an imputation method that takes into account women interrupting treatment in order to estimate the probability that a woman entering an IVF programme would conceive if she did not interrupt treatment.

Materials and Methods

Materials

Data were provided by two French IVF units (Cochin in Paris and Clermont-Ferrand, a medium-sized city in the centre of France). All women having their first aspiration in the unit between 1998 and 2002 were included. All women included thus had at least one aspiration. Information was collected from medical records and concerned all aspirations (both conventional IVF and intracytoplasmic sperm injection, ICSI) undergone by the patient in the IVF unit, as well as data on FET, up to 2005. The data collected were the woman's date of birth, date of aspiration, number of oocytes retrieved, IVF technique used, number of embryos transferred, number of embryos frozen, pregnancy and delivery.

Data collection was discontinued after the woman's fourth aspiration, four being the number of aspirations which is refunded by the French social security system. Therefore, for these first four cycles, the economic factor did not intervene.

Elective single embryo transfer was not applied and treatment-independent pregnancies were not included.

The standard stimulation protocol associated pituitary desensitization, with a daily agonist started in the luteal phase, and stimulation with gonadotrophins (hMG or recombinant FSH).

No woman aged over 42 was included unless she had a normal ovarian reserve and a specific IVF indication. The ovarian reserve was used as an inclusion criterion for couples aged over 37.

Definitions

An attempt was defined as an aspiration, i.e. an oocyte retrieval. Success was defined as a delivery resulting from fresh or frozen embryo transfer. Only the first delivery was considered here. Consequently, each rate was calculated among women who underwent aspiration and who did not conceive at one of the preceding transfers. Interruption of treatment was defined as no event (aspiration, embryo transfer, FET, pregnancy or delivery) recorded in the IVF unit for at least two years at the time of data acquisition.

Methods

Three methods were applied to estimate the IVF success rate. The first two were basic statistical methods, while the third was a new method tested here.

(i) *Kaplan–Meier life table analysis*: As already stated, it assumes that women discontinuing treatment would have the same chance of conceiving as women continuing treatment.

(ii) *Observed percentage*: This method consists of counting the total number of successes and dividing it by the number of women who underwent a first aspiration. This observed percentage is taken as the success rate. It assumes that no woman discontinuing treatment would have had a child if she had continued.

(iii) *Multiple imputation* (Schafer, 1997; Little and Rubin, 2002): As there were no subsequent IVF aspirations after interruption, the corresponding data were treated as missing. Their values were estimated using the known characteristics of the woman and of her previous aspirations, according to imputation methodology. Simple imputation replaces each missing value by a single value. This does not take into account the uncertainty about the correct value to impute. Multiple imputation therefore replaces each missing value with a set of plausible values. Each complete data set is then analysed using standard statistical methods for complete data and results are combined.

Thus, multiple imputation inference involves three phases.

(i) The missing data are filled in m times to generate m complete data sets.

(ii) The m complete data sets are analysed using standard statistical methods.

(iii) The results from the m complete data sets are combined to produce inferential results. The final estimate is the average of the m estimates.

Multiple imputation was carried out using the SAS MI procedure (SAS Institute Inc. 2004. SAS/STAT® 9.1 User's Guide. Cary, NC: SAS Institute Inc.).

All women included underwent at least one aspiration, so they all had a response (success or failure) at the first aspiration and no imputation was necessary. At the second aspiration, we did not impute data for women whose first aspiration was successful (only the first delivery was of interest), nor for women who had a second aspiration (their response was not missing). Thus, at the second aspiration, we imputed data for women whose first aspiration was not successful and who discontinued treatment. Then, at the third aspiration, we imputed data for women who had no success (observed or imputed) at the preceding aspirations and who did not undergo a third aspiration. Similarly, at the fourth aspiration, we imputed data for women who had no success (observed or imputed) at the preceding aspirations and who discontinued before the fourth aspiration. Following Rubin's recommendations, we made five imputations by aspiration (Rubin, 1996). For an aspiration, each imputation gave a success rate, which was the number of successes (observed or imputed) at this aspiration divided by the number of women. These estimates were combined to result in one estimate per aspiration. For the success rate, the combination was the mean of the estimates, i.e. the sum of the estimates for the aspiration (one per imputation) divided by the number of imputations. For the cumulative success rate, the same approach was used,

the estimate being the number of women having had at least one success at the aspiration or at the preceding ones divided by the number of women.

The variables included in the multiple imputation model to estimate the outcome (delivery or not) were the IVF unit, the woman's age at the aspiration date, the number of oocytes retrieved and the total number of embryos (defined as the sum of the number of embryos transferred and the number of embryos frozen). Consequently, the probability of success was predicted according to these covariates, which are frequently cited in the literature as factors affecting success for IVF. In the imputation process, linear regression was used for the characteristics (woman's age, number of oocytes retrieved and total number of embryos), and the discriminant function method was used for the outcome, which was a binary variable (success/failure).

Results

The data concerned 3037 women. Table I describes the study population by IVF unit. The population was almost equally distributed between the two IVF units (52% in Cochin and 48% in Clermont-Ferrand) and between the two IVF techniques (conventional IVF and ICSI) at the first aspiration. On average, women were 33 years old and 10 oocytes were retrieved at the first aspiration. From these oocytes, an average of five embryos was obtained, two were transferred and two were frozen. In 58% of these first aspirations, no embryo was frozen. Within the 3–5 embroys transferred category (Table 1), 91% of the transfers were transfers of three embryos. At the first aspiration, only 1% of the transfers in Cochin and 2% of the transfers in Clermont were transfers of four or five embryos.

Women underwent an average of two aspirations in the unit (SD = 1.2), the range being 1–8. Forty-one percent of the women underwent only one aspiration and 69% underwent one or two aspirations. For 64% of the women, the time elapsing between the first aspiration and the last event recorded in the unit was $\leq$24 months.

Birth rate decreased as the aspiration rank increased, i.e. as the number of aspirations increased (Table II). The interruption rate per aspiration cycle was high and increased as the aspiration rank increased (Table II).

Table III compares, at the first aspiration, women interrupting treatment without having a child, women continuing and undergoing another aspiration without previously having a child and women having a child. Women interrupting treatment were significantly more likely to be older than 35 years, to have five or fewer oocytes retrieved and to have more than two embryos.

Table IV presents the results of multiple imputation according to aspiration rank. The observed success rate was 21% at the first aspiration (no imputation performed), and the estimated rates with imputation were 18% at the second, 9% at the third and 8% at the fourth aspiration. The cumulative success rates were 35% at the second, 41% at the third and 46% at the fourth aspiration.

The proportion of imputed data were 30% at the second aspiration (30% of women failing at the first aspiration interrupted treatment), 55% at the third aspiration (55% of women failing at the first and at the second aspirations—observed or imputed failures—interrupted treatment) and 71% at the fourth aspiration (71% of women failing at the first, second and third aspirations—observed or imputed failures—interrupted treatment).

Table V presents the success rates after four aspirations as estimated with multiple imputation and the two basic methods. The success rate after four aspirations estimated with multiple imputation was 46%. Life table analysis (Kaplan–Meier) gave a 52% success rate after four aspirations. The observed percentage was 37%.

Table I. Characteristics at the first aspiration of patients at the two IVF units

	IVF unit			
	Clermont-Ferrand		Cochin	
	n	%	*n*	%
Patient age (years)				
17–34	1028	71	983	62
35–39	338	23	469	30
40–46	87	6	132	8
Technique				
IVF	888	61	762	49
ICSI	565	39	808	51
Oocytes retrieved				
0–6	537	37	546	35
7–15	679	47	786	50
16–60	237	16	247	16
Embryos obtained				
0–1	295	20	305	19
2–3	334	23	401	26
4–33	824	57	864	55
Embryos transferred				
0–1	374	26	369	24
2	765	53	1014	65
3–5	314	22	181	12
Embryos frozen				
0	1009	69	756	48
1–2	185	13	325	21
3–21	259	18	484	31

Discussion

According to our data, 37% of the couples entering an IVF unit leave with a child. This percentage results from both the efficacy of the IVF technique and the acceptability (physical, psychological and social factors) of the IVF programme. To estimate a success rate precluding acceptability, we applied to our data a multiple imputation method that takes into account women interrupting treatment. The estimated

Table II. Outcomes according to aspiration rank

Aspiration rank	*n*	Women succeeding (%)	Women failing and continuing (%)	Women failing and interrupting (%)
1	3037	21	55	24
2	1624	19	52	29
3	813	13	55	31
4	397	13	48	39

Table III. Women's characteristics according to the outcome of the first aspiration

Characteristics	Women succeeding		Women failing and continuing		Women failing and interrupting		Chi-square test*
	n	%	*n*	%	*n*	%	*P*-value
Patient age (years)							
≤35	542	83	1219	73	458	63	<0.0001
>35	108	17	440	27	270	37	
Number of oocytes retrieved							
≤5	105	16	478	29	253	35	0.004
>5	545	84	1176	71	475	65	
Number of embryos transferred							
≤2	528	81	1376	84	618	85	0.56
>2	122	19	263	16	110	15	
Total number of embryos**							
≤2	165	25	904	55	360	50	0.01
>2	484	75	734	45	365	50	

*For 'women failing and continuing' versus 'women failing and interrupting'.
**Sum of the number of embryos transferred and the number of embryos frozen.

cumulative IVF success rate (i.e. delivery rate) after four aspiration cycles was 46%.

This study developed an innovative approach to the estimation of success in IVF. The objective was to consider the IVF programme in its entirety by linking all the aspiration cycles undergone by a woman. Success was considered as the result of one or more aspirations, including FETs. These represented a sizeable proportion of all transfers (30%) and successful pregnancies were in fact achieved: 8% of FETs resulted in deliveries.

In comparison with women having a second aspiration, those dropping out after the first aspiration were older (more women were aged over 35 years), were more likely to have had five or fewer oocytes retrieved and were more likely to have had a total number of embryos (transferred and frozen) greater than two. No difference was found in the number of embryos transferred, which in most cases, was two. All in all, patients with poor prognostic factors were more likely to drop out, even if poor prognosis was not *per se* the only reason for abandoning treatment. Consequently, women interrupting treatment had rather lower chances of having a child. Except with regard to embryos, these results are in line with those of Sharma *et al*. (2002) but contrast with those of Roest *et al*. (1998). These authors showed no difference in fertilization rate or in predictors of poor treatment outcome (oocyte yield ≤2 and replacement of <2 embryos) in the first and second cycles between patients who continued treatment and those who dropped out, in 1211 patients of the IVF programme at the Zuiderziekenhuis in Rotterdam (The Netherlands) who had their first attempt between January 1989 and June 1994 (Roest *et al*., 1998).

Few studies have attempted to take into account successive aspirations in women entering an IVF programme. Using life table analysis, Sharma *et al*. (2002) estimated that the cumulative live birth rate after four aspirations was 66%. With the same method, Stolwijk *et al*. (2000) estimated the success rate at 56%, using data from 1315 couples who underwent their first IVF treatment from March 1991 to December 1997 at the University Hospital Nijmegen in The Netherlands. Finally, Olivius *et al*. (2002) found a success rate of 69% in a retrospective, observational study of 974 couples (1985 stimulated cycles initiated) who started their first conventional IVF or ICSI cycle between January 1996 and December 1997 in a university hospital in Sweden.

As women discontinuing treatment have lower chances of having a child than women continuing treatment, life table estimations are optimistic and overestimate the success rate. On the contrary, cumulative success rates after four aspirations estimated by the observed percentage are pessimistic and underestimate the success rate. Stolwijk *et al*. (2000) and Olivius *et al*. (2002) found 42 and 55%, respectively. These authors proposed more realistic estimates, between the pessimistic and the optimistic estimates.

Stolwijk *et al*. (2000) proposed that women interrupting treatment after failing should be divided in two groups according to the reason for discontinuation. Women interrupting because of a medical reason were supposed to have a probability of having a child in the future equal to zero (observed percentage method). Women interrupting because of a reason

Table IV. Multiple imputation estimation of success according to aspiration rank

Aspiration rank	Estimated success rate (%)	95% CI	Cumulative success rate (%)	95% CI
1	21*	20–23	21	20–23
2	18	16–20	35	33–37
3	9	7–11	41	39–43
4	8	6–11	46	44–48

*Observed percentage (all women underwent at least one aspiration).

Table V. Estimated success rates after four aspirations

Method	Estimated success rate (%)	95% CI
Kaplan–Meier	52	49–55
Observed percentage	37	35–38
Multiple imputation	46	44–48

other than medical were supposed to have the same probability of delivery as women continuing after a failed attempt (Kaplan–Meier method). Thus, they estimated the success rate at 52%. Olivius *et al*. (2002) used the same idea, dividing women canceling treatment according to their prognosis (poor/not poor). In this case, the estimated success rate was 65%. These methods are attractive. We generalized the method, estimating the probability of success not by prognosis group only but for each woman. Thus, our multiple imputation method works as if each group consisted of only one woman and then her probability of delivery is determined according to her characteristics.

The cumulative success rate given by multiple imputation after four aspirations was 46%. This is a pertinent estimation of the probability that a couple entering an IVF unit will have a child if they do not interrupt the treatment until four aspirations or a delivery.

Our estimations [pessimistic 37% (observed percentage), realistic 46% (multiple imputation) and optimistic 52% (Kaplan–Meier)] were lower than those of Stolwijk *et al*. and Olivius *et al*. This is consistent with the world collaborative report (Adamson *et al*., 2006) which observed that France had lower pregnancy and delivery rates (24 and 18%) than Sweden (30 and 23%) and a lower pregnancy rate than The Netherlands (30%, delivery rate not available).

Compared with the methods of Stolwijk and Olivius, multiple imputation has the advantage of providing completed data sets, which include both observed and imputed data. The total number of successes (observed percentage) is thus available and it is no longer necessary to use life table analysis to estimate the success rate. It is then possible to use these data sets to consider aspects other than success rate estimation, and particularly to analyse factors associated with IVF outcome. Our method is thus promising, even if some points require closer investigation. First, multiple imputation worked well with two IVF units. The results did not differ if we entered the IVF unit as a covariate in the imputation model and then examined the results by IVF unit, or if we carried out imputation separately for each IVF unit. This demonstrates that the unit effect was well dealt with by the multiple imputation model. However, we have to examine the results with more than two IVF units. Then multiple imputation until the fourth aspiration cycle implied that a large proportion of the data were simulated at the last aspiration cycle. Despite that, the confidence intervals are small. But simulations should be done to determine how reasonable it is to impute such a large proportion of missing data. Furthermore, multiple imputation assumes that missing data are missing at random (MAR), which means that given the observed data, the missingness mechanism does not depend on the unseen data. The MAR assumption can never be tested but the main predictors were included in the imputation model, so that we were as close as possible to the MAR assumption. Finally, the covariates included here (woman's age, number of oocytes retrieved and total number of embryos) are factors affecting success that are often found in the literature. Moreover, no other sociodemographic data on women were available at the time of analysis. Consequently, more research should be done to decide which covariates should be included in the imputation model, to better estimate the probability of success of each woman.

In conclusion, multiple imputation is a promising method. It can provide an estimated cumulative success rate for IVF that takes into account the chances of delivery of women interrupting treatment, and it should be more thoroughly explored. IVF results would gain in transparency if they presented not only the result aggregating the efficacy and the acceptability of the technique (result given by the observed percentage), but also separate results for efficacy (e.g. with multiple imputation) and for acceptability (one of its aspects being given by the interruption rate). It would be very interesting to use multiple imputation in another context, such as in other countries where interruption rates and reasons for interrupting treatment are different, and to compare the results. Couples starting an IVF programme ask for the fullest information possible, and the success rate after four aspirations is concrete information.

Funding

Région Ile-de-France (doctoral grant, N.S.).

Author Roles

Data analysis and interpretation, drafting the article and approval of the final manuscript—N.S.

Data interpretation, critical revision and approval of the final manuscript—J.B.

J.L. Pouly: data acquisition, critical revision and approval of the final manuscript.

Data acquisition, critical revision and approval of the final manuscript—J.G.

Conception and design, data acquisition, data interpretation, critical revision and approval of the final manuscript—E. de L.R.

References

Adamson GD, de Mouzon J, Lancaster P, Nygren K-G, Sullivan E, Zegers-Hochschild F. World collaborative report on in vitro fertilization, 2000. *Fertil Steril* 2006;**85**:1586–1622.

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.

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

Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons Inc., 2002.

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.

Olivius K, Friden B, Lundin K, Bergh C. Cumulative probability of live birth after three in vitro fertilization/intracytoplasmic sperm injection cycles. *Fertil Steril* 2002;**77**:505–510.

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

Roest J, van Heusden AM, Zeilmaker GH, Verhoeff A. Cumulative pregnancy rates and selective drop-out of patients in in-vitro fertilization treatment. *Hum Reprod* 1998;**13**:339–341.

Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc* 1996; **91**:473–489.

Schafer JL. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall/CRC, 1997, 430.

Sharma V, Allgar V, Rajkhowa M. Factors influencing the cumulative conception rate and discontinuation of in vitro fertilization treatment for infertility. *Fertil Steril* 2002;**78**:40–46.

Steptoe P, Edwards R. Birth after the reimplantation of a human embryo. *Lancet* 1978;**2**:366.

Stolwijk AM, Wetzels AM, Braat DD. Cumulative probability of achieving an ongoing pregnancy after in-vitro fertilization and intracytoplasmic sperm injection according to a woman's age, subfertility diagnosis and primary or secondary subfertility. *Hum Reprod* 2000;**15**:203–209.

Thonneau P, Marchand S, Tallec A, Ferial ML, Ducot B, Lansac J, Lopes P, Tabaste JM, Spira A. Incidence and main causes of infertility in a resident population (1,850,000) of three French regions 1988–1989. *Hum Reprod* 1991;**6**:811–816.

Submitted on July 16, 2007; resubmitted on October 5, 2007; accepted on October 9, 2007