Analysis of multiple-cycle data from couples undergoing *in vitro* fertilization: Methodologic issues and statistical approaches

Stacey A. Missmer[1,2,3], **Kimberly R. Pearson**[4], **Louise M. Ryan**[4], **John D. Meeker**[5], **Daniel W. Cramer**[1,2], and **Russ Hauser**[6,7]

[1]Department of Obstetrics, Gynecology, and Reproductive Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts

[2]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts

[3]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts

[4]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts

[5]Department of Environmental Health Sciences, University of Michigan, Ann Arbor, Michigan

[6]Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts

[7]Vincent Memorial Obstetrics and Gynecology Service, Andrology Laboratory and In Vitro Fertilization Unit, Massachusetts General Hospital, Boston, Massachusetts

Abstract

The number of in vitro fertilization (IVF) cycles in the U.S. increased from fewer than 46,000 in 1995 to more than 120,000 in 2005. IVF and other assisted reproductive technology (ART) data are routinely collected and used to identify outcome predictors. However, researchers do not always make full use of the data due to its complexity. Design approaches have included restriction to first-cycle attempts only, which reduces power and identifies effects only of those factors associated with initial success. Many statistical techniques have been utilized or proposed for analysis of IVF data, ranging from simple t-tests to sophisticated models designed specifically for IVF. We apply several of these methods to data from a prospective cohort of 2687 couples undergoing ART from 1994 through 2003. Results across methods are compared and the appropriateness of the various methods is discussed with the intent to illustrate methodologic validity. We observed a remarkable similarity of coefficient estimates across models. However, each method for dealing with multiple cycle data relies on assumptions that may or may not be expected to hold in a given IVF study. The robustness and reported magnitude of effect for individual predictors of IVF success may be inflated or attenuated due to violation of statistical assumptions, and should always be critically interpreted. Given that risk factors associated with IVF success may also advance our understanding of the physiologic processes underlying conception, implantation, and gestation, the application of valid methods to these complex data is critical.

Address for Correspondence: Stacey Missmer; Department of Obstetrics, Gynecology, and Reproductive Biology, Brigham and Women's Hospital; 75 Francis Street, Boston, MA 02115, Telephone: (617) 525-2021; Fax: (617) 525-2008; stacey.missmer@channing.harvard.edu.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Accounting for 1% of all births in the U.S., [1] in vitro fertilization (IVF) is an increasingly common method of assisted reproductive technology (ART). The growth in popularity of IVF has been accompanied by an increased number of research studies aimed at identifying and assessing factors that affect IVF success. However, IVF-related studies often present only two-sided p-values calculated from chi-square, Student's t-test, or life table analyses.[2] These methods are limited in the information conveyed; they do not account for confounding; and—given that censorship by treatment cessation may be predictive of the likelihood of successful outcome (livebirth)—the validity of these methods is questionable.[3,4]

IVF data present several statistical challenges.[2,4,5] First, the IVF process involves a sequence of steps, and the outcome at one step may influence the next. Oocyte retrieval must be successful in order to proceed to fertilization, fertilization must be successful in order to proceed to embryo transfer, and so on; recent work of Penman[6,7] describes a method for modeling these stages simultaneously. Second, successful pregnancy depends upon both the health of the mother and the viability of the transferred embryos; the "EU" (embryo/uterus) model[8] addresses this dual nature of successful gestation. Third, a majority of women pursuing IVF undergo more than one treatment cycle. Because women tend to have similar outcomes in successive pregnancies (e.g. low birth weight) in the natural setting,[9] it is likely that the result across IVF cycles will be correlated. This lack of independence of cycles must be taken into account in the data analysis.[10,11]

We focus on methods for addressing the issue of multiple cycles, applying several statistical methods to a dataset from three IVF clinics to illustrate variation in validity and interpretation. This paper does not address the statistical issues arising from the multi-step nature of IVF; for simplicity we focus on the last stage only—beginning at the transfer of at least one embryo, with success defined as the delivery of at least one live birth.

METHODS

Study population

Data collection and demographics—We defined eligibility as all married couples newly enrolled for IVF treatment from 1994 through 2003 at three clinics in the Boston area, excluding those using donor gametes or gestational carriers. Self-administered baseline questionnaires were given to the women and men to assess the following: demographic history; menstrual, contraceptive, fertility, and medical history; physical activity; and lifestyle factors. This study was approved by the Brigham and Women's Hospital and Harvard School of Public Health Institutional Review Boards, and all participants gave informed consent.

ART cycle-specific data were abstracted from medical records, including the planned procedure (IVF or gamete intra-fallopian transfer [GIFT]), use of gonadotropin-releasing hormone agonists, gonadotropins used, intracytoplasmic sperm injection (ICSI), and monitoring details (follicle counts and estradiol levels, number of oocytes retrieved, semen characteristics, number and quality of embryos created, details of the embryo transfer, and post-transfer outcomes). The cycle reference date was defined as the day gonadotropin treatment was begun. Ten percent of records were re-abstracted with agreement of >99%.

The original study population included 2687 couples undergoing a variety of ART procedures. For the purposes of this methodologic paper, only IVF cycles with the transfer of at least 1 embryo were included. Any cycles subsequent to a non-IVF cycle were omitted, as were 69 cycles for which data on covariates were missing and 3 implausible records.

Typical IVF cycle methods—IVF cycles began with gonadotropin-releasing hormone agonists in a long (down-regulation) or short (flare) regimen. Dosing of lupron for flare generally commenced on day one of the menstrual cycle and continued during gonadotropin administration, while down-regulation commenced on day 21 of the prior menstrual cycle and continued for at least two weeks before gonadotropin administration. Various formulations of gonadotropins were used at doses usually ranging from 150 to 600 ampules per day by intramuscular or subcutaneous injection.

Ovarian response was monitored by measuring serum estradiol levels and using ultrasound to monitor the number and size of follicles, typically on cycle day six and then every 1–3 days depending on the patient's response. Generally, when there were 2 or more follicles with a diameter of at least 18 millimeters, human chorionic gonadotropin (hCG) was administered to replicate the pre-ovulatory surge of luteinizing hormone. Transvaginal oocyte recovery generally occurred 36 hours after hCG administration, and oocytes were inseminated by mixing with ~50,000–300,000 sperm or by ICSI (intracytoplasmic sperm injection).

If insemination was successful, the embryos were generally cultured for 2–5 days. Some or all of the embryos were transferred to the uterine cavity via transfer catheter. About 18 days after embryo transfer, a serum pregnancy test (β-hCG) was performed.

If the beta-hCG pregnancy test was positive, a pelvic ultrasound was performed to determine whether there was clinical evidence of a pregnancy (at least one fetus with heartbeat). If a clinical pregnancy was detected on ultrasound, follow-up of couples documented whether the pregnancy ended in a livebirth or in miscarriage, ectopic pregnancy, molar pregnancy, or stillbirth. The IVF cycle was considered to be successful only if the pregnancy ended in the delivery of at least 1 live newborn.

Statistical methodology

We surveyed several methods: logistic regression using a single IVF cycle from each subject, generalized estimating equations (GEE), non-linear mixed effects models, discrete-time survival analysis, and "EU" (embryo-uterus) models. All analyses were performed using Statistical Analysis Software (SAS®) version 9.1 (SAS Institute, Inc., Cary, NC).

Repeated measures—It is unreasonable to expect IVF outcomes of different cycles from the same woman to be independent. To ignore correlations among cycle outcomes can lead to invalid inferences or loss of statistical power.

A common statistical technique used in the IVF literature is to discard all but one cycle— typically the first or the last. The first-cycle data are viewed as the cleanest, while the last-cycle data include every ultimate success; neither considers all IVF treatments. In our dataset, selecting only one cycle would discard approximately half the data, decreasing statistical power to detect associations.

To account for correlations of outcomes, generalized estimating equations (GEEs) can be used to make population-averaged inferences and do not require distributional assumptions about the observations. GEEs can be fit using Proc Genmod in SAS, estimating both model-based and empirical standard errors, and the user may select from a variety of correlation structures. We present the GEE model with an exchangeable correlation structure using empirical standard errors to construct confidence intervals.

In contrast to GEE, non-linear mixed effects models are used to make subject-specific inferences. These models are well suited for analysis of unbalanced data (in which subjects

have different numbers of observations). Mixed effects models can also accommodate multiple layers of data clustering, and Ecochard and Clayton[10] suggest use of those models for ART data with sperm donation. A number of estimation methods can be used to fit these models[12]; here, Gaussian quadrature was used with the NLMIXED procedure.

Despite the capability of GEE and mixed effects models to cope with correlated data, GEE methods can perform poorly in the situation of nonignorable cluster size (e.g., when the number of attempts is related to the outcome of interest).[13] GEE estimates may be biased when missing data are not missing completely at random (a necessary model assumption),[14] although adjustments such as weighted GEE are possible remedies.[15] Similarly, the "conditional independence" assumption in mixed models (i.e., the independence of a given woman's cycle outcomes from each other), may not be tenable with an IVF dataset, as discussed above. Furthermore, since treatment choices such as gonadotropin dose are likely to be guided by clinical responses on previous cycles, the outcome of a particular cycle may affect covariate values at subsequent cycles, adding further to cycle dependence.

Survival analysis—Another alternative is the survival analysis approach, considering event time to be the number of cycles until success .[16,17] This framework obviates the need to account for within-subject correlation because the woman, rather than the cycle, is the unit of observation. However, instead of the traditional Cox proportional hazards model, which assumes the underlying time scale is continuous, the following discrete survival analysis is tailored to the situation of a discrete time unit (here, IVF attempt number).

Letting p_k be the probability that a woman succeeds on the kth IVF attempt given that she has not succeeded on previous attempts, the model stipulates

$$logit(p_k) = \log[p_k/(1-p_k)] = \beta_k + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k = \beta'X. \tag{1}$$

The model is equivalent to performing unconditional logistic regression including cycle number as a categorical (i.e., nominal) covariate. Comparatively, the Cox proportional hazards model is related to conditional logistic regression.[17] The primary disadvantage to any survival analysis is the independent censoring assumption—e.g., women who have 2 failed IVF procedures and then discontinue treatment must be representative of all women in the study who had 2 failed IVF procedures.[18] Whether this assumption is tenable for a particular IVF dataset must be considered critically.

We used these three methods to fit models for successful IVF that included treatment center (coded 1, 2, 3), study enrollment period (1994–1998, 1999–2003), history of previous livebirth (yes, no), woman's age (<35, 35–37, 38–40, >40 years), primary infertility diagnosis (female, male, idiopathic), ampules of gonadotropins divided by 10 (continuous), gonadotropin releasing hormone agonists regimen (down regulation, flare), ICSI (yes, no), and number of embryos transferred (count).

EU models—Generally, the probability of success p is modeled as a function of a linear combination of predictors:

$$g(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k = \beta'X. \tag{2}$$

Most commonly, logistic regression, choosing the logit

$$g(p) = \log[p/(1-p)]$$

as the link function g. Proposed by Spiers[8] and developed further by others[19,20,21], the „EU" model posits that successful conception depends on two independent factors: the viability of the embryo and the receptivity of the uterus. Let E indicate if an embryo is viable (E=1 if viable, E=0 if not) and let e be the probability of viability (e=P[E=1]). Similarly, let U indicate if the uterus is receptive (U=1 if receptive, U=0 if not), and u be the probability of receptivity (u=P[U=1]). Each of e and u may depend on a combination of covariates:

$$g_u(u) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \ldots + \gamma_k x_{k} = \gamma' X, \tag{3}$$

$$g_e(e) = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \ldots + \delta_k x_{k} = \delta' X. \tag{4}$$

A successful IVF cycle (the EU method was initially proposed for the outcome of "pregnancy attainment" but can be adapted to model livebirth) requires that the uterus be receptive (U=1) and that at least one of the embryos be viable (E_j=1 for at least one j). The probability of success is expressed as

$$p = u[1 - \Pi(1 - e_j)], \tag{5}$$

with the product taken over all embryos transferred in that cycle. If embryo-level predictors such as daily cell counts or embryo grade are not available (as in our dataset), then the model simplifies to

$$p = u[1 - (1 - e)^n], \tag{6}$$

where n is the number of embryos transferred.

With the probability of successful IVF decomposed into the E and U factors, this model facilitates investigation of predictors associated with the two primary components of success. Most importantly, the EU model is well suited to IVF data that include cycles with donor eggs or gestational carriers, whereas the covariate values for the woman providing oocytes differ from those for the woman to whom the resulting embryos are transferred. On the other hand, interpretation of overall association with success of IVF is difficult in the case when one predictor is believed to be related to both viability and receptivity. Also, embryo implantation is observed only in the aggregate, and furthermore, the model assumes that embryo viabilities are independent.

We explored how the EU model compares to the simpler model for the probability of success using the first cycle. Either model is difficult if the study includes more than one IVF attempt per subject, and we additionally present results utilizing data from all cycles undertaken.

In contrast with the illustration of repeated-measures statistical methods, we observed that with the EU models pared-down models were more stable, and therefore fewer covariates are included, based upon lack of observed association and retention for comparative purposes. In the EU model, logit links were used on both the E and U factors, and each linear predictor (equations 3 and 4) included history of previous livebirth and woman's age. Embryo viability, but not uterine receptivity, was allowed to depend additionally on ampules of gonadotropins and gonadotropin-releasing hormone agonists regimen (i.e., these variables were included in equation 4 but not 3). Because the likelihood for the EU model (equation 6) automatically incorporates the number of embryos transferred, it is unnecessary to include this as a covariate; if all women had the same number of embryos transferred, the model

could not be fit. The non-linear mixed models procedure using the likelihood function from equation 6 was used to fit the EU model. In total, the EU model was applied to 1) data from the first cycle only, 2) data from all cycles ignoring between cycle correlations, 3) data from all cycles with cycle number included as indicator variables, 4) with a Gaussian random effect on the E (embryo viability) factor only, 5) with a Gaussian random effect on the U (uterine receptivity) factor only, and 6) with a Gaussian random effect on both the E and U factors.

RESULTS

After exclusions, data were available on 2318 couples contributing 3913 cycles (Tables 1 and 2). Women's mean age was 35.2 years (SD = 4.3; range = 20–49 years); men's mean age was 36.9 years (SD = 5.6; range = 20–69). Male, female, and idiopathic infertility diagnoses were equally represented in this population. The median number of cycles per couple was two and the maximum, six.

The percent of failed implantations increased (and livebirth percentage decreased) with cycle number (Table 2). Other outcomes such as spontaneous abortion did not change noticeably across number of cycles (results not shown). The proportion of cycles failing prior to transfer decreased with cycle number, while the proportion of couples discontinuing IVF treatment among those who did not have a livebirth increased with cycle number.

Standard errors from any of the models using all cycles tended to be smaller than those from the model using only the first cycle, resulting in tighter confidence intervals (Table 3). For example, in the first cycle model the estimated odds ratio (OR) of success for women over 40 compared to the referent age group (<35) was 0.21 (95% confidence interval [CI] = 0.13–0.33), while in the discrete survival model the estimate was 0.21 (0.15–0.29). The GEE model with exchangeable correlation structure yielded parameter estimates and standard errors that were extremely close to those for discrete survival analysis. On average the estimates in the two models differed by 7%, and none varied by more than 15%. A mixed effects model was fit with subject-level random intercepts assumed to come from a Gaussian distribution. Comparing these results with those from other models is problematic as the coefficients in a non-linear mixed effects model have subject-specific rather than population-averaged interpretations.[22] Although the interpretation of the odds ratios are not the same, qualitatively the results are similar. The differences in effect magnitude and direction can be striking and non-intuitive; in particular, the subject-specific effect of cycle number seen in the analysis by Hogan and Blazar[11] was found to be similar in our dataset (results not shown). Typically, however, within our data the directionality of effects remained the same. For example, the estimated odds ratio of success for down regulation compared with flare was 1.22 (1.00–1.48) in the discrete survival model 1.32 (1.02–1.71) in the mixed effects model.

We attempted a total of six EU models. Overall, the first-cycle-only model yielded inflated estimates and wider confidence intervals compared with the multicycle models (Table 4). The models that included a Gaussian random effect on either the E factor or the U factor, or that included a Gaussian random effect on both the E and U factors, failed to parameterize. For the model with a random effect on E, optimization could not be completed; with a random effect on U, there were negative eigenvalues; and with a random effect on both E and U, the model did not converge – highlighting the complexity of the multicycle EU-model statistical methodology.

We conducted additional analyses not tabulated here. To illustrate that logistic regression analyses using only one IVF cycle can vary by choice of cycle, we fit models using the first

or last cycles. Using first-cycle data, the estimated odds ratio of successful IVF for women with a previous livebirth compared with those without was 1.39 (1.10–1.75), while using last-cycle data produced an estimate of 1.17 (0.94–1.46). To gauge the extent to which within-woman observations were correlated, we computed intraclass correlations (ICC). In the mixed effects model, the estimated ICC was 0.18 (-0.50 to 0.85). In addition, the discrete survival model was refit including number of embryos transferred on the log scale. The likelihood ratio test statistics for improved fit after including number of embryos transferred were 43.14 (standard scale) and 65.43 (log scale), indicating an advantage if the number of embryos transferred variable is log-transformed. This transformation makes intuitive sense as it incorporates a decreasing marginal benefit to transferring additional embryos. As an aside, a second motivation for the transformation is that if the log link is applied to the embryo viability factor in equation 6, and a Taylor series expansion is applied, the number-of-embryos-transferred variable appears as an offset term on the log scale.

Additionally, for the mixed effects model, we computed empirical Bayes estimates of the random effects. These were found to be somewhat bi-modal with a longer tail on the negative side(results not tabulated), suggesting that the normality assumption for random effects may be suspect for IVF data. Tests for departure from normality such as Cramér von-Mises also indicated that the Gaussian assumption may be suspect. To address the issue of dependent censoring in the discrete survival model, data for women who dropped out after cycles 1–3 were regressed on the type of IVF failure (e.g., implantation failure) controlling for predictors used (e.g., woman's age). Rare failpoints (such as stillbirth) were excluded in these analyses. None of the failpoints was associated at every cycle with a statistically significant increase or decrease in dropout. However, on cycle 1 only, women who experienced spontaneous abortion were more likely to discontinue treatment, and were also more likely to have a later successful cycle. It therefore may be apropos to include attainment of clinical pregnancy on a previous cycle as a covariate; however, Buck-Louis et al.[5] have advised caution when including previous reproductive outcomes as model predictors as doing so may mask effects.

DISCUSSION

Researchers do not always make full use of IVF data due to its complexity. ART data present challenges in statistical analysis, including the fact that many couples require more than one treatment cycle to achieve a successful livebirth. While there are several potential approaches to this challenge, some techniques may introduce bias or reduce statistical power. To date, the literature is composed largely of analyses that present non-parametric Chi-square, Fisher's exact, Student's T-test, and lifetable analyses. Many of these analyses quantify statistical significance without an evaluation of magnitude of effect or variability of the association, do not make use of the full dataset, or do not control for confounders. Use of multivariate regression methods can address many of these concerns, but must account for correlation in the data and consider the tenability of regression model assumptions. To address these issues, we applied several statistical methods to identify predictors of livebirth within a prospective cohort of 2687 couples treated with IVF.

By decomposing the probability of IVF success into embryo viability and uterine receptivity factors, the EU (embryo-uterus) model allows for mechanism-focused inferences, can improve model fit, and has become substantially easier to program with the advent of packages such as the NLMIXED procedure in SAS®. However, the richness of the model can lead to complexity in interpretation of results. To fit EU models to all-cycle data, the hierarchical Bayesian approach of Dukic and Hogan[21] can be used, but this requires expertise to implement. We observed instability in the EU models with large changes in parameter estimates comparing the single cycle to multiple cycle analyses. Each covariate is

entered into the model twice – once for embryo viability and once for uterine receptivity, doubling the parameterization relative to that of the simple logistic regression or discrete survival models. Stability was improved when model covariates were pared down, but this may limit interpretability and data utilization. As an alternative, we fit EU models with random effects on the E factor only, U factor only, and both E and U factors. Computational time was lengthy and convergence of parameter estimates for the single random-effect models was sensitive to initial values and ultimately did not converge. In general, we observed greater woman-to-woman and temporal variation in uterine receptivity, while embryo viability appeared to be more stable.

Power to detect effects can be diminished if data are discarded; standard errors were noticeably larger in the first-cycle-only logistic regression compared with models that used all cycles. Furthermore, it should be noted that our sample size was large; discarding cycles is likely to have more severe consequences in a typical single-site IVF study.

We fit three types of models using all IVF cycles, with various mechanisms to account for within-couple dependence of outcomes: the logistic-normal mixed effects model, GEE with exchangeable correlation structure, and the discrete survival model. Because the survival model is equivalent to ordinary logistic regression with cycle number included as a nominal covariate, this approach is straightforward to interpret and the easiest to implement with standard software. Furthermore, the model is well suited to features of IVF data, such as censoring and cessation of treatment upon success. Use of either discrete survival analysis or GEE requires assumptions about dropout: independent censoring in the case of survival analysis,[17] or non-informative cluster size in the case of GEE.[13] Here, the parameter estimates and standard errors in the two models were very similar. The mixed effects model is well suited to women undergoing different numbers of IVF cycles, and is perhaps the least likely to rely on untenable assumptions. However, its coefficients have subject-specific interpretations, and drawing population-level inferences can be complex. The choice of a marginal model compared with a subject-specific model (such as non-linear mixed effects) should be based primarily on the target of study.[23]

We confirmed previous observations for the following predictors of IVF success given that an embryo transfer occurred: history of previous livebirth, woman's age, gonadotropin releasing hormone agonists regimen, number of gonadotropin ampules used, and number of embryos transferred. Other covariates (such as body mass index, man's age, primary infertility diagnoses, day-3 estradiol level, number of oocytes retrieved and ICSI) were considered, as well, but they were not found to be important predictors of IVF success. (For purposes of method comparison, diagnosis and ICSI were retained in all models) (eTable, http://links.lww.com).

No model is without disadvantages.[23] Techniques that utilize only one IVF cycle per woman can over-simplify the outcome of the overall IVF experience, either by counting only first-cycle successes, or by disregarding the number of attempts required to achieve success. However, if all cycles are used, the likely interdependence of outcomes from the same couple must be taken into account. Each method examined in this paper for dealing with multiple-cycle data relies on various assumptions that may or may not be expected to hold in a given IVF study. Probability models such as the EU and similar methods allow for greater flexibility, but this comes at a cost of increased model complexity. The target of inference and tenability of model assumptions must guide the choice of analytic method.

An encouraging conclusion of this methodologic investigation is that the magnitude of the effects observed and the strengths of the associations did not vary dramatically across models. In fact, the similarity of coefficient estimates across models was surprisingly

striking in this exercise with multicenter IVF data. Though it is beyond the scope of this paper, a simulation study could shed more light on these issues. Those conducting ART research and practicing evidence-based medicine can be assured that previously published research that has critically considered and transparently detailed study design, data collection, and confounding control is likely to have produced results within the range of valid effect estimation regardless of the model that was applied. However, the robustness and reported magnitude of effect for individual predictors of IVF success may be inflated or attenuated due to violation of statistical assumptions, and should be critically interpreted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

1. Wright VC, Chang J, Jeng G, Macaluso M. Assisted reproductive technology surveillance--United States, 2003. MMWR Surveill Summ. 2006; 55(4):1–22. [PubMed: 16723970]

2. McDonough PG. Life-table analysis falls short of the mark. Fertil Steril. 2003; 79(6):1468. [PubMed: 12798909]

3. Land JA, Courtar DA, Evers JL. Patient dropout in an assisted reproductive technology program: implications for pregnancy rates. Fertil Steril. 1997; 68(2):278–281. [PubMed: 9240256]

4. Olive DL. Analysis of clinical fertility trials: a methodologic review. Fertil Steril. 1986; 45(2):157–171. [PubMed: 3512311]

5. Buck Louis GM, Schisterman EF, Dukic VM, Schieve LA. Research hurdles complicating the analysis of infertility treatment and child health. Hum Reprod. 2005; 20(1):12–18. [PubMed: 15489239]

6. Penman, RHG.; Tyler, J. Modelling IVF Data using an Extended Continuation Ratio Random Effects Model; Proceedings of the 22nd International Workshop on Statistical Modelling; 2–6 July 2007; Barcelona. ISBN 978-84-690-5943-2

7. Penman, RHG.; Tyler, J. Modelling assisted reproductive technology data using an extended continuation ratio model; Statistical Solutions to Modern Problems. Processings of the 20th International Workshop on Statistical Modelling; July 10–15 2005; Sydney, Australia.

8. Speirs AL, Lopata A, Gronow MJ, Kellow GN, Johnston WI. Analysis of the benefits and risks of multiple embryo transfer. Fertil Steril. 1983; 39(4):468–471. [PubMed: 6832402]

9. Louis GB, Dukic V, Heagerty PJ, Louis TA, Lynch CD, Ryan LM, Schisterman EF, Trumble A. Analysis of repeated pregnancy outcomes. Stat Methods Med Res. 2006; 15(2):103–126. [PubMed: 16615652]

10. Ecochard R, Clayton DG. Multi-level modelling of conception in artificial insemination by donor. Stat Med. 1998; 17(10):1137–1156. [PubMed: 9618774]

11. Hogan JW, Blazar AS. Hierarchical logistic regression models for clustered binary outcomes in studies of IVF-ET. Fertil Steril. 2000; 73(3):575–581. [PubMed: 10689015]

12. Pinheiro JC, Bates DM. Approximations to the Log-likelihood Function in the Nonlinear Mixed-effects Model. Journal of Computational and Graphical Statistics. 1995; 4(1):12–35.

13. Hoffman EB, Sen PK, Weinberg CR. Within-cluster resampling. Biometrika. 2001; 88:1121–1134.

14. Diggle, PHP.; Liang, K-Y.; Zeger, S. Analysis of Longitudinal Data. 2nd ed.. Oxford, UK: Oxford University Press; 2002.

15. Preisser JSLK, Rathouz PJ. Performances of weighted estimating equations for longitudinal binary data with drop-outs missing at random. Stat Med. 2002; 21:3035–3054. [PubMed: 12369080]

16. Cox DG, Hankinson SE, Kraft P, Hunter DJ. No association between GPX1 Pro198Leu and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2004; 13(11 Pt 1):1821–1822. [PubMed: 15533915]

17. Cox, DROD. Analysis of Survival Data. London, UK: Chapman & Hall; 1984.

18. Kalbfleisch, JDPR. The statistical analysis of failure time data. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2002.

19. Baeten S, Bouckaert A, Loumaye E, Thomas K. A regression model for the rate of success of in vitro fertilization. Stat Med. 1993; 12(17):1543–1553. [PubMed: 8235177]

20. Zhou H, Weinberg CR. Evaluating effects of exposures on embryo viability and uterine receptivity in in vitro fertilization. Stat Med. 1998; 17(14):1601–1612. [PubMed: 9699232]

21. Dukic V, Hogan JW. A hierarchical Bayesian approach to modeling embryo implantation following in vitro fertilization. Biostatistics. 2002; 3(3):361–77. [PubMed: 12933603]

22. Heagerty PJZS. Marginalized multilevel models and likelihood inference. Stat Sci. 2000; 15:1–26.

23. Fitzmaurice, GMLN.; Ware, JH. Applied Longitudinal Analysis. Hoboken, NJ: John Wiley & Sons; 2004.

Table 1

Characteristics of study population in analytic data set (n = 2318 couples).

Categorical subject covariates	%
Primary Diagnosis	
Idiopathic	32
Male	34
Female	34
Woman's age (years)	
<35	45
35–37	23
38–40	20
>40	12
Previous livebirth	
No	77
Yes	23

Table 2

Outcomes and characteristics of IVF cycles that included the transfer of at least 1 embryo, stratified by IVF attempt number (n = 3913 cycles).

IVF Cycle Number	No. Couples	% Failed implantation	% Livebirth	Average no. ampules gonadotropins	% Down-regulation	% Intracytoplasmic sperm injection	Median no. of embryos transferred
1	2092	52	33	32	77	27	3
2	1040	58	27	40	66	36	3
3	478	62	22	44	59	44	3
4	199	63	22	47	50	49	3
5	76	70	18	47	49	55	3
6	28	75	4	52	32	61	4

Table 3

Association between patient and treatment characteristics and the likelihood of livebirth from four regression models (n = 2318 couples, 3913 IVF cycles).

Covariate	First cycle only[a] Adjusted OR (95% CI)[b]	Discrete survival[c] Adjusted OR (95% CI)[b]	GEE[d] Adjusted OR (95% CI)[b]	Mixed effects[e] Adjusted OR (95% CI)[b]
Previous livebirth				
No[f]	1.00	1.00	1.00	1.00
Yes	1.39 (1.10–1.75)	1.27 (1.06–1.52)	1.29 (1.07–1.56)	1.34 (1.04–1.74)
Woman's age (years)				
<35[f]	1.00	1.00	1.00	1.00
35–37	0.91 (0.71–1.16)	0.93 (0.77–1.12)	0.94 (0.77–1.14)	0.94 (0.76–1.16)
38–40	0.53 (0.40–0.71)	0.59 (0.47–0.72)	0.58 (0.47–0.72)	0.60 (0.40–0.82)
>40	0.21 (0.13–0.33)	0.21 (0.15–0.29)	0.21 (0.15–0.29)	0.19 (0.08–0.44)
Primary Diagnosis				
Idiopathic[f]	1.00	1.00	1.00	1.00
Male	1.18 (0.88–1.57)	1.06 (0.86–1.31)	1.05 (0.85–1.31)	1.10 (0.86–1.40)
Female	0.91 (0.71–1.62)	0.93 (0.77–1.13)	0.92 (0.76–1.12)	0.93 (0.75–1.15)
Gonadotropin dose per 10 ampules	0.84 (0.77–0.91)	0.88 (0.84–0.93)	0.88 (0.84–0.92)	0.85 (0.77–0.93)
Down-regulation				
No[f]	1.00	1.00	1.00	1.00
Yes	1.29 (0.97–1.72)	1.22 (1.00–1.48)	1.26 (1.04–1.52)	1.32 (1.02–1.71)
Intracytoplasmic sperm injection	0.79 (0.60–1.04)	0.94 (0.78–1.14)	0.94 (0.78–1.13)	0.86 (0.69–1.08)
No. Embryos transferred	1.30 (1.17–1.43)	1.24 (1.16–1.32)	1.24 (1.16–1.31)	1.22 (1.08–1.38)
Cycle # (Cox only)				
Cycle 1[f]		1.00		
Cycle 2		0.86 (0.72–1.03)		
Cycle 3		0.67 (0.52–0.87)		
Cycle 4		0.69 (0.47–1.00)		
Cycle 5		0.53 (0.29–0.98)		
Cycle 6		0.08 (0.01–0.59)		

[a] Multivariate logistic regression using first-cycle data only (n = 2092 couples, 2092 cycles).

[b] All models were also adjusted for clinic site and study enrollment period.

[c] Discrete survival (Cox proportional odds model) fit using Proc Logistic, including cycle number as a class variable.

[d] Generalized estimating equations, using empirical estimates for standard errors and exchangeable working correlation, fit using Proc Genmod.

[e] Mixed effects model, using the logit link and Gaussian random intercept, fit using Proc NL Mixed.

[f] Reference category.

Table 4

Association between patient and treatment characteristics and the odds of embryo viability (the E component) as well as the odds of uterine receptivity (the U component), (n= 2092 couples).[a]

Covariate[b]	First Cycle Only		Multicycle Ignoring Correlations		Multicycle with Cycle Number as Indicator Variables	
	EU model: Embryo Viability factor Adjusted OR (95% CI)[ez]	EU model: Uterine Receptivity factor Adjusted OR (95% CI)[ez]	EU model: Embryo Viability factor Adjusted OR (95% CI)[ez]	EU model: Uterine Receptivity factor Adjusted OR (95% CI)[ez]	EU model: Embryo Viability factor Adjusted OR (95% CI)[ez]	EU model: Uterine Receptivity factor Adjusted OR (95% CI)[ez]
Women's age (years)						
<38 [d]	1.00	1.00	1.00	1.00	1.00	1.00
38–40	1.31 (0.48–3.56)	0.54 (0.38–0.79)	1.00 (0.47–2.13)	0.64 (0.48–0.85)	0.95 (0.43–2.12)	0.65 (0.48–0.89)
>40	0.84 (0.09–8.12)	0.29 (0.14–0.61)	0.64 (0.20–1.98)	0.30 (0.19–0.47)	0.60 (0.18–2.01)	0.30 (0.18–0.50)
Previous livebirth						
No [d]	1.00	1.00	1.00	1.00	1.00	1.00
Yes	2.58 (0.96–6.94)	1.17 (0.82–1.67)	1.71 (0.85–3.43)	1.25 (0.96–1.62)	1.71 (0.82–3.56)	1.20 (0.90–1.60)
Gonadotropin						
Dose						
Per 10 ampules	0.78 (0.67–0.90)		0.88 (0.81–0.95)		0.90 (0.83–0.98)	
Down-Regulation						
No [d]	1.00		1.00		1.00	
Yes	5.51 (2.97–10.2)		4.35 (2.56–7.38)		3.75 (2.28–6.19)	
Embryos transferred						

[a] Of the 2318 couples, some failed prior to embryo transfer, and thus did not contribute data to the analysis.

[b] In the EU model, number of embryos transferred is built into the likelihood.

[c] Gonadotropin dose and down-regulation were included as predictors only for the embryo viability factor. The EU model was fit using Proc NLMIXED and all initial parameter values for coefficients were set to zero.

[d] Reference category.