1999

A new model for predicting probabilities of viable pregnancy and multiple gestations for in Vitro fertilization

Ray Vernon Haning Jr.

A New Model for Predicting Probabilities of Viable Pregnancy and Multiple Gestations for *in Vitro* Fertilization

Thesis submitted to
The Graduate School of
Marshall University

In partial fulfillment of the
Requirements for the Degree of
Master of Arts
in Mathematics with a Concentration in Statistics

by

Ray Vernon Haning, Jr.

Marshall University

Huntington, West Virginia

June 7, 1999

This thesis was accepted on June 7 1999
Month Day Year

as meeting the research requirements for the master's degree.

Advisor Laura Adkins

Department of Mathematics

Leonard Deutsch
Dean of the Graduate College

ACKNOWLEDGMENTS

I thank the other physicians of the Section of Reproductive Endocrinology of Women and Infants' Hospital, Jacquelyn S. Loughlin, MD, Carol A. Wheeler, MD, Gary N. Frishman, MD, David B. Seifer, MD, and Andrew Blazar, MD, and the staff of the *in vitro* fertilization laboratory for their assistance in maintaining adherence to the demanding protocol required for providing care to these patients.

TABLES

CONTENTS

1. INTRODUCTION

In vitro fertilization (IVF) can be viewed as high stakes gambling with more than one winning outcome and more than one outcome leading to failure. Knowing the odds of successful outcomes, the odds of outcomes leading to failure, and the various costs is a tremendous asset in any form of gambling. Four important variables have been reported to affect pregnancy rates in IVF programs: age of patient, number of unsuccessful prior attempts, embryo morphology, and number of embryos transferred.[1-10]

While it is desirable to optimize pregnancy rates, such optimization may often result in a high incidence of multiple gestation. Data from the Society for Assisted Reproductive Technology show that the pregnancy rate per retrieval for standard IVF increased from 21.9% in 1991 to 29.1% in 1994 and that the U.S. national incidence of IVF high order (triplet or more) multiple gestation also increased from 5% to 6.5%.[6,11-13] Any trend toward a higher incidence of high order multiple gestations is alarming since the obstetric and neonatal risks for high order multiple gestation are significantly increased.[14-17] Multiple gestation is also associated with an increased risk of severe ovarian hyperstimulation syndrome,[18] a life-threatening medical condition.

When I began the present study in 1993, I hoped to discover a way to help physicians assist patients in deciding the optimum number of embryos to transfer to maximize the probability of pregnancy while limiting the probability of multiple gestation. I suspected that this

would require building a model based on effective predictors such as the number and quality of embryos transferred as well as other factors that affect the probability of an individual embryo implanting in a given patient's uterus. The function resulting from such model building could be evaluated after substitution of the actual predictor values for the specific patient at hand to yield prospective probability predictions for pregnancy and multiple gestation for individual patients.

It would seem reasonable, *a priori*, to think that after transferring a number of embryos, say n, to the uterus with each having the probability of implanting, p, the events of implantation of $y \in [0,n]$ embryos could be described by the binomial probability distribution,

$$p(y)=[n!/((y!)(n-y)!)]p^{y}(1-p)^{(n-y)}, \qquad (1.1)$$

as long as the implantations were independent events. Unfortunately, a literature search showed that the binomial distribution seriously underestimated the probability that multiple gestations would occur.[19-21] This observation suggests that in some way the implantation of embryos into the uterine endometrium does not conform to the mathematical concept of independent trials. It was clear that to be of use the new model must not only be able to successfully predict the probability of pregnancy but also the probability of multiple gestations. The introduction of conditional probabilities for the first time to describe the probabilities of various multiple gestations combined with logistic regression analysis to estimate parameters permitted development of a new model which met these goals.

2. MATERIALS AND METHODS

Human Subjects and Source of the Data Set

2.1 Human Subjects

Women and Infants' Hospital, Providence, Rhode Island, is a teaching hospital affiliated with Brown University Medical School. I served as director of the Division of Reproductive Endocrinology at Women and Infants' Hospital and Director of the IVF laboratory during the study interval, and the IVF protocol and laboratory practices did not vary during this time. Our IVF protocol[18] and IVF population demographics[16] have been described elsewhere in detail. All classic IVF cycles performed at Women and Infants' Hospital from the time of its inception in 1988 to December of 1994 were reviewed retrospectively. This study was approved by the Research and Human Rights Committee of Women and Infants' Hospital.

2.2 Source of the Data Set

Only classic IVF cycles with transfer of only fresh embryos and with all eggs originating from the embryo recipient herself were utilized for statistical analysis. Any subsequent cycles contributed by patients who had once become pregnant via IVF and then

attempted to achieve additional pregnancies were also excluded. This left data from 1113 cycles available for evaluation in the **full data set**.

Since a significant number of our patients with high order multiple gestation underwent multifetal reduction,[16] data for deliveries would not be a true reflection of the number of implantations. Therefore, a clinical pregnancy was defined as a positive fetal heart in an intrauterine gestational sac on ultrasound obtained 6 weeks after embryo transfer.

Mathematical and Statistical Considerations

2.3 The Data and Considerations About Modeling

The observed data for this study consisted of the number of implantations which had occurred after transfer of from 1 to 10 embryos to the uterus of individual patients. Failure to conceive or success (implantation of 1, 2, 3, 4, or 5 embryos) was observed to result. These degrees of success are commonly referred to as singletons, twins, triplets, quadruplets, and quintuplets and are hereafter referred to collectively as pregnancy order. Such data could theoretically be modeled by the binomial distribution (Equation 1.1) if these implantations were considered the outcomes of from 1 to 10 independent trials of equal probability. However, I knew that previous investigations had demonstrated that the binomial model seriously underestimated the probabilities of the pregnancy orders higher than one.[19-21] The most obvious explanation for this is that trials for an individual patient are not independent.

Therefore, it was imperative that the new model must permit the probability of "subsequent" implantations to be influenced by the occurrence of "prior" implantations. Note that I have used "subsequent" more as a linguistic crutch than as a way of indicating cause and effect since the alteration in probabilities may well occur through a winnowing of the population rather than through directly causal relationships. This concept is important enough to explore some of the potential biological explanations in simple terms.

2.4 Potential Biological Explanations of Apparent Lack of Independence of Implantations

One or more fertilized embryos were *transferred* to the uterus of these patients transcervically using a small plastic catheter. However, for a pregnancy to occur, the embryo must undergo further growth in the patient's uterus and *implant* by physically invading the endometrial lining of the uterus. The occurrence of implantation is first detected by observation of a rising concentration of human chorionic gonadotropin (hCG) in the patient's blood. Later, at about eight weeks after the theoretical last menstrual period, the viable early implantation can be observed by detection of a fluid filled sac containing an embryo with a beating heart using a vaginal ultrasound probe. Thus, "implanted embryo" refers to an embryo which has been physically been transferred to the uterus and which has managed to invade the endometrial lining and undergo further growth. Furthermore, the words "implant", "implants", and "implantation" refer to the physical process by which the growing embryo invades the uterine endometrium.

It is well established that hCG secreted by the implanted embryo is able to increase

ovarian secretion of the steroid hormones, estradiol and progesterone, and the protein hormone, relaxin, prior to the time that menstruation would be expected to occur.[18,22-23] It is even possible that an implanted embryo could directly affect the implantation of other embryos through secretion of known or unknown hormones or other factors acting locally on the uterus or endometrium or more distantly on the ovary or other structures. Based on the apparent lack of statistical independence of implantation events, it has been speculated by a number of authors that there may be a "helper" effect which assists implantation of additional embryos if one embryo implants,[19-21,24] but no proof of the existence of a direct causal mechanism has been presented. It is possible that a winnowing effect explains the statistical observation. Older women are less likely to become pregnant than younger ones. So embryos of apparent equal quality are less likely to implant in the uterus of an older women than that of a younger one. The quality of one embryo in a group of embryos tends to be similar to that of others in the same group. So when a group of embryos is transferred to the uterus it is reasonable to expect that if one embryo is healthy enough to implant, then others from the same batch are also likely to be able to implant as well. Therefore, the implantation of at least one embryo can be viewed as a bioassay result suggesting both that the endometrium was capable of allowing implantation and that the embryos were healthy enough to implant. A positive result of this bioassay can be viewed as suggesting that the odds of additional implantations are increased.

2.5 Definitions of the Events and Conditional Probabilities

This approach requires the use of conditional probabilities associated with the following events:

A1) Implantation of at least ***one*** embryo

A2) Implantation of at least ***two*** embryos

A3) Implantation of at least ***three*** embryos

The (unconditional) probabilities for these events will be denoted by P(A1), P(A2), and P(A3), respectively. Event A3 includes triplets, quadruplets, and other higher order gestations. It was not useful to define events for implantation of 4 and 5 embryos since the number of such events was small, and the risks posed to both the mother and infants by triplet pregnancies are severe enough to warrant attempting to avoid triplets as well as multiple gestations of even higher order.

In order to provide a mathematical means of describing the effect of each event on the probabilities for the other events, it was necessary to also define two conditional probabilities. P(A2|A1) was defined as the probability of 2 or more embryos implanting given that at least one embryo had implanted. P(A3|A2) was defined as the probability of 3 or more embryos implanting given that at least two embryos had implanted. Mathematically, these conditional probabilities are defined by

$$P(A2|A1) = P(A1 \cap A2)/P(A1) \qquad (2.1)$$

and

$$P(A3|A2) = P(A2 \cap A3)/P(A2), \qquad (2.2)$$

where $\cap$ indicates the intersection of two events, that is, that both events occur.

2.6 Calculation of the Probabilities of Singleton, Twin, and Triplet and Higher Order Gestations

My use of conditional probabilities in this way was a novel addition to the IVF probability literature.[25] These conditional probabilities have the useful properties that they can be estimated from appropriate subsets of the data as will be described concretely in ***Results*** (Section 3.2). Furthermore, because event A2 can occur only when event A1 occurs, $P(A1 \cap A2) = P(A2)$. And since event A3 can occur only when event A2 occurs, $P(A2 \cap A3) = P(A3)$. As a result, the following equations hold.

$$P(A2|A1) = P(A1 \cap A2)/P(A1) = P(A2)/P(A1) \quad (2.3)$$

$$P(A3|A2) = P(A2 \cap A3)/P(A2) = P(A3)/P(A2) \quad (2.4)$$

Solving Equation (2.3) for P(A2) and Equation (2.4) for P(A3) yields

$$P(A2) = P(A1) \times P(A2|A1) \quad (2.5)$$

and

$$P(A3) = P(A2) \times P(A3|A2). \quad (2.6)$$

It will now be necessary to define three more events:

B1) Implantation of ***exactly one*** embryo

B2) Implantation of ***exactly two*** embryos

B3) Implantation of at ***least three*** embryos

The (unconditional) probabilities for these events will be denoted by P(B1), P(B2), and P(B3), respectively. Event B3 includes triplets, quadruplets, and other higher order

gestations.

From Equation (2.5) it can be determined that

$$P(B1) = P(A1) - P(A2) = P(A1) \times (1 - P(A2|A1)) \quad (2.7)$$

Equations (2.5) and (2.6) imply that

$$P(B2) = P(A2) - P(A3) = P(A1) \times P(A2|A1) \times (1 - P(A3|A2)) \quad (2.8)$$

and

$$P(B3) = P(A3) - P(A2) = P(A1) \times P(A2|A1) \times P(A3|A2) \quad (2.9)$$

2.7 Calculation of the Proportion of All Pregnancies which are Expected to be Singleton, Twin, or Triplet or Higher Order Gestations

In order to describe the proportions of pregnancies which fell into the categories singleton, twin, and triplet or higher order gestation, it was necessary to describe the following three conditional probabilities. $P(B1|A1)$ was defined as the probability of a singleton pregnancy given that pregnancy had occurred. $P(B2|A1)$ was defined as the probability of a twin pregnancy given that pregnancy had occurred. $P(B3|A1)$ was defined as the probability of a triplet or higher order pregnancy given that pregnancy had occurred. Mathematically, these conditional probabilities are defined by

$$P(B1|A1) = P(B1 \cap A1)/P(A1), \quad (2.10)$$

$$P(B2|A1) = P(B2 \cap A1)/P(A1), \quad (2.11)$$

and

$$P(B3|A1) = P(B3 \cap A1)/P(A1) \quad (2.12)$$

Since events B1, B2, and B3 can only occur when event A1 occurs, $P(B1 \cap A1) = P(B1)$, $P(B2 \cap A1) = P(B2)$, and $P(B3 \cap A1) = P(B3)$. From Equations (2.7) and (2.10) it can be determined that

$$P(B1|A1) = P(B1 \cap A1)/P(A1) = P(B1)/P(A1) =$$

$$[P(A1) \times (1 - P(A2|A1))]/P(A1) = 1 - P(A2|A1). \quad (2.13)$$

Equations (2.8) and (2.11) imply that

$$P(B2|A1) = P(B2 \cap A1)/P(A1) = P(B2)/P(A1) =$$

$$[P(A1) \times P(A2|A1) \times (1 - P(A3|A2))]/P(A1) =$$

$$P(A2|A1) \times (1 - P(A3|A2)). \quad (2.14)$$

From Equations 2.9 and 2.12 it can be shown that

$$P(B3|A1) = P(B3 \cap A1)/P(A1) = P(B3)/P(A1) =$$

$$[P(A1) \times P(A2|A1) \times P(A3|A2)]/P(A1) = P(A2|A1) \times P(A3|A2). \quad (2.15)$$

2.8 Probabilities that Must be Estimated to Properly Inform Patients About Probabilities of Singleton, Twin, and Triplet or Higher Order Gestation

As can be seen from Equations (2.7) - (2.9) and (2.13) - (2.15) the probabilities required to properly inform patients about the probabilities of conception and of conceiving singleton, twin, or at least triplets can all be calculated from 3 probabilities: P(A1), P(A2|A1), and P(A3|A2). The outcomes for each event were encoded as 1 or 0 for success or failure, respectively.

As mentioned above, previous studies had identified 4 variables which had been

associated with success.[1-10] These four variables could potentially be used to adjust the predicted probability of success for each patient, although equations using more than one of these predictors had not been published prior to our preliminary report.[25] Later, these four potential predictors will be defined, and their suitability for use as predictors will be examined.

2.9 The Logistic Regression Model

The multiple linear regression model uses $\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k$ to predict the value of the response variable. Since the result could be any real number, and our response variable, a predicted probability, ϵ (0,1), some transformation will be necessary. The predicted probability is indicated by $\hat{p}$. The logit function

$$g(\hat{p}) = \text{logit}(\hat{p}) = \log(\hat{p}/(1-\hat{p})) \qquad (2.16)$$

has (0,1) as its domain and $(-\infty, \infty)$ as its range, so predicting logit($\hat{p}$) rather than $\hat{p}$ will solve the problem. The predicted value of $\hat{p}$ can then be determined by using logit^{-1}

$$\hat{p} = e^{g(\hat{p})}/(1 + e^{g(\hat{p})}). \qquad (2.17)$$

Because $\hat{p}$ must be calculated from Equation (2.17), $0 < \hat{p} < 1$, a range consistent with the definition of probabilities. The linear logistic regression model has the form

$$g(\hat{p}) = \text{logit}(\hat{p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k, \qquad (2.18)$$

where β_0 is the intercept parameter, the β_1, β_2, β_2,... β_k are the slope parameters, and the X_1, X_2,...,X_k are the predictor values. All statistical calculations were performed using Minitab release 5.1.1 (Minitab Inc., State College PA) except for logistic regression for which SAS release 6.04 was utilized (SAS Institute Inc., Cary NC).

2.10 The Potential Predictor Variables

Prior studies had identified 4 variables as potential predictors of IVF success:[1-10] number of unsuccessful prior attempts, age of patient, embryo morphology, and number of embryos transferred. I will now examine each of the potential predictor variables.

The *cycle number* is the way in which the number of unsuccessful IVF attempts is encoded. The cycle number used in the preliminary report[25] differed from that in the present study since it was the number assigned by the laboratory for purposes of keeping track of patient activity. Under the laboratory system an IVF attempt was classified as a cycle even though it failed to result in embryo transfer. Although this numbering system could have little effect on cycle 1 data (the subject of the preliminary report), it resulted in classification of some cycles as 2 or 3 even though there had never been an embryo transfer. Inclusion of such IVF attempts in the analysis of cycle 2 or cycle 3 probabilities would presumably interfere with accurate determination of the change in probabilities induced by failed prior IVF cycles, an effect which appears to reflect a winnowing process.

Cycle number for the present study was redefined as the current cycle plus the number of consecutive cycles *with transfer of embryos* that had been performed without viable pregnancy. All cycles from patients with prior IVF pregnancies were excluded from analysis in order to eliminate concern about autocorrelation. Because any change in probability of conception from one attempt to the next is apparently influenced only by removal of a more fertile subset of patients from the population attempting subsequent cycles, attempts which did not result in transfer of embryos for any reason were not counted as completed cycles (since no pregnancy could result). The cycle numbers for the present

study were assigned after detailed review of each individual IVF attempt. This careful review resulted in discovery of 38 additional cycle 1 attempts in the preliminary data set which had been excluded from the initial report[25] because a prior attempt had been cancelled due to excessively high serum estradiol levels, poor follicular recruitment, or poor fertilization results. Under either numbering system, a given patient could contribute only 1 time to the estimation of implantation probability for any given cycle number. Due to the way cycle number has been defined for the present analysis, all cycle 2 patients represent a subset of cycle 1 patients, and all cycle 3 patients represent a subset of cycle 2 patients. Similar considerations also apply to subsequent cycles. Combining data from different cycles into a single analysis would have resulted in all cycle 3 patients being included 3 times and all cycle 2 patients being included at least two times, introducing the possibility of autocorrelation. This problem was side-stepped by analyzing data from each cycle number separately. In that way patients appeared only once in the analysis of an individual cycle.

The *patient's age* in years at the time of the egg retrieval was included as a potential predictor. *Embryo morphology* was encoded according to the criteria of Puissant[2] which assigns a score of 0 - 6 to each embryo (with higher scores reflecting better morphology and correlating with higher chances of success). The Puissant embryo score was not used prospectively in deciding which embryos to transfer. The laboratory selected the "best" embryos for transfer, preferring those with higher numbers of blastomeres and the least fragmentation. The *total* and *mean* Puissant *embryo scores* for each transfer group were calculated.[8]

The *Number of embryos transferred* ranged from 1 to 10. The final decision about

how many embryos to transfer was made by the couple after receiving a clinical estimate of the potential for pregnancy and multiple gestation. It is obvious mathematically that the total embryo score is the product of the number of embryos transferred and the mean embryo score, so the total embryo score would be expected to be highly correlated with both mean embryo score and the number of embryos transferred.

Model Building

2.11 Potential Predictors and Criteria for Assessing Model Fit

I investigated age, total embryo score, number of embryos transferred, and mean embryo score as well as their cross products and squares as potential predictors. I used the SAS Logistic procedure to fit the various models to the data. According to the SAS Technical Report P-200,[26] the Logistic Procedure fits logistic regression models for binary or ordinal response data by the method of maximum likelihood. It uses iteratively reweighted least squares to find the maximum likelihood estimates of the parameters in the model. See the SAS Technical Report P-200[26] for a detailed description of the technique. Several criteria for assessing model fit are provided by the program. Three that were useful in the present project were -2 log likelihood (-2 log L), Akaike Information Criterion (AIC), and Schwartz Criterion (SC). The estimates of each of the $\hat{p}_j$ are found using the maximum likelihood estimates of the regression coefficients.

$$-2 \text{ Log } L = -2 \sum_j [y_j \log(\hat{p}_j) + (1-y_j)\log(1-\hat{p}_j)], \qquad (2.19)$$

where y_j is the (0,1) outcome of the jth observation.

$$AIC = -2 \log L + 2(k + s), \qquad (2.20)$$

where k is the number of ordered values for the response (1 in this case), and s is the number of predictor variables.

$$SC = -2 \log L + (k + s)\log(N), \qquad (2.21)$$

where N is the number of observations.

The Logistic Procedure provides four model-selection methods.[26] "None" fits the model specified by the investigator. "Forward" selects the predictor with the highest adjusted Chi-Square for variables not in the model if it is significant at the level specified; once a variable is entered it is never removed. "Backward" enters all variables specified in the model and removes at each step the least significant variable which fails to meet the criteria specified; once a variable is removed it is not re-entered. "Stepwise" enters variables using a method similar to "Forward" and removes any which are no longer significant using criteria similar to "Backward". It is possible to force variables into the model using the "None", "Start", or "Include" options. I used 0.05 as the criterion for entry ("Slentry") or removal ("Slstay") of a predictor. I always specified the "Details" option to print statistics useful in choosing among the various models.

2.12 Four Model Selection Methods

The "analysis of variables not in the model" generated by the "Stepwise" selection method provided an excellent means for assessing the predictive power of potential indepen-

dent variables which was not provided by the "None" option. The "Stepwise" method had the advantage of rapidly constructing most models of interest and then printing out diagnostics on the models. The "Forward" selection method offered no advantage here, since the "Details" option made it clear when any variables that had been entered were subsequently removed. I used "Backward" to explore whether this approach generated other good models which I had overlooked. After inspection of the various outputs, I used a knowledge gained by prior experience in reproductive endocrinology research and clinical experience to construct any further models of interest using the "None" selection method. The various models for the same data set were compared using the AIC and the SC statistics. To make interpretation of the regression slopes clear I planned to select first order simple models unless selection of the higher order model was justified by lower AIC and SC statistics in both the preliminary and holdout data sets (which will be explained in detail below in Section (2.11)).

Model Validation

2.13 Splitting the Data into Model-Building and Model-Validation Sets

A major potential problem in fitting a model to a set of data is the question of whether the model has been fit to some peculiarity unique to the data set or whether the model is generalizable to other data from a similar population. According to Neter et al.[27] "The best means of model-validation is through collection of new data". Another way

according to Neter et al.[27] is through use of a holdout sample to check validity. This is accomplished by data splitting, which amounts to an attempt to simulate replication of the study.

> "When data are collected sequentially in time, it is often useful to pick a point in time to divide the data. Generally, the earlier data are selected for the model-building set and the later data for the validation set."[27]

The split in the present data set was made in 1993, when I decided to analyze all data available up to that time. However, due to the difficulty in finding an adequate way to model the data, the project did not come to fruition until after I received further training at Marshall University. Additional IVF treatment cycles accumulated under the same protocols and procedures through the end of 1994. The data from the initiation of the program up to the split in 1993 were used for model building and the data from the split to the end of 1994 (hereafter referred to as the ***holdout*** data) were used for the model validation set as described by Neter et al.[27]

2.14 Sizes of the Various Data Sets and How They were Used

In this study, there were 667 first cycles, 448 of which were used for initial model development and the preliminary report (for which I provided the statistical analysis and served as the corresponding and senior author until after acceptance of the revised draft).[25] Another 219 first IVF cycles from the holdout data were not used for the original model-building. Data from 285 second and 107 third IVF cycles after prior failure constitute two additional sets of data not used in the preliminary report which were available for model

validation. As suggested by Neter et al.,[27] I used the holdout data to re-estimate all of the "good" models that had been considered originally in development of the new model to see if the new model was still the preferred model according to the holdout data. Similarly, I used the other two data sets (the data from cycle 2 and the data from cycle 3) in the same way. The fit of the new model was assessed using the Chi-Square goodness of fit test.

3. RESULTS

Overview

3.1 Overview of the Data

There were 1113 cycles suitable for extended analysis. To provide an overview, data on number of embryos transferred and implanted and numbers of observations are presented in Table 1. Of the 306 patients who became pregnant, only 3 (0.98%) became pregnant after transfer of a single embryo; the other 303 had transfer of at least two embryos. Of the pregnant patients with transfer of two or more embryos, 118 (38.9%) had at least 2 implantations. Of the 118 patients with at least 2 implantations, all but one had transfer of 3 or more embryos. Summary data on the predictor variables for Cycles 1-3 are presented in Table 2.

3.2 Overview of Model-Building and Model-Validation

Model building of the logistic regression model began with selection of the model-building data set, 448 patients who attempted IVF for the first time, as described above. The first

Table 1

Number of Embryos Implanted, Tabulated by Number of Embryos Transferred in 1113 Classic IVF Cycles

Embryos transferred	----------No. embryos implanted-------						Total cycles	With any pregnancy		Average implantation
No.	0	1	2	3	4	5	No.	No.	%	%
1	56	3	0	0	0	0	59	3	5	5
2	84	7	1	0	0	0	92	8	9	5
3	144	36	8	5	0	0	193	49	25	12
4	255	76	36	13	3	0	383	128	33	13
5	206	54	25	15	4	1	305	99	32	11
6	28	7	3	2	0	0	40	12	30	8
7	20	4	1	0	0	0	25	5	20	3
8	8	1	0	1	0	0	10	2	20	5
9	4	0	0	0	0	0	4	0	0	0
10	2	0	0	0	0	0	2	0	0	0
ALL	807	188	74	36	7	1	1113	306	27	11

Table 2

Summary Data for Predictor Variables for Cycles 1 - 3

	Mean	Median	S.D.	Minimum	Maximum
Cycle 1 (N=667)					
Age	33.72	34	3.96	23	46
Mean Embryo score	4.25	4	1.16	1	6
Embryos Transferred	3.72	4	1.17	1	8
Total Embryo score	15.84	16	6.47	1	40
Cycle 2 (N=285)					
Age	34.17	34	4.14	24	46
Mean Embryo score	4.19	4	1.17	1	6
Embryos Transferred	4.16	4	1.39	1	8
Total Embryo score	17.51	18	7.49	2	46
Cycle 3 (N=107)					
Age	34.664	35	4.005	25	44
Mean Embryo score	3.958	4	1.057	2	6
Embryos Transferred	4.495	5	1.488	1	9
Total Embryo score	17.822	18	7.664	4	42

fit undertaken was for P(A1) since fewer trials were available for fitting P(A2|A1) or P(A3|A2). Next, the potential predictors were selected: age, total embryo score, number of embryos transferred, and average embryo score. (Inclusion of all the latter 3 predictors, when only one or two would suffice, was based on a determination not to make any assumptions about the form of the model). All second order terms were also evaluated (cross products and squares of the variables). After extensive evaluation of the predictors for P(A1) using the "Stepwise", "Backward", and "None" options as discussed above, the "best" models were chosen. Only models where all predictors were statistically significant at $P < 0.05$ were considered. The AIC and SC statistics were used to evaluate models based on the same data. The best models were then fit to the holdout data and the data from cycle 2 and cycle 3. There were insufficient data to fit any models to cycles beyond cycle 3. First order models were preferred over second order models because of their easier interpretation, and there were no second order models which were superior to the first order models. The best model selected on the model-building data set[25] was also an excellent model for the holdout data, the cycle 2 data, and the cycle 3 data. Accordingly, it was refit on the full cycle 1 data set. The model selected utilized the total embryo score and age as predictors without any interaction or other second order terms.

Building the logistic regression model for P(A2|A1) was done in the same way. But several mathematical corrections were necessary for the predictors. The number of embryos available for implantation was corrected by subtracting one since one embryo had already implanted. Similarly, the corrected total embryo score is the product of the average embryo score and the corrected number of embryos available for a second implantation. These

calculations are based on the following analysis of the realistic problems I faced in building the model. When a group of embryos has been transferred to the uterus and some of them implant, there is no way of knowing which ones have implanted unless they all do. That happens in only a tiny fraction of all IVF cycles. In the vast majority of cycles we are left with only knowing the number of embryos transferred, the total embryo score, and the average embryo score. If we are using the total number of embryos as a predictor of the chance of implantation, it is appropriate to use the total number only in predicting the chance of the first implantation. Once at least one has implanted, the chance of a second one implanting would logically depend on the number remaining, the total-1. This is nowhere more obvious than where only one embryo has been transferred. Clearly, there is no chance that a second implantation can occur. Similar considerations would dictate using the total-2 as the logical predictor for the chance of a third implantation given that at least two have implanted. The average embryo score (calculated as

(total embryo score)/(total number of embryos transferred))

is the best we can do at estimating the embryo morphology of the implanting embryos absent the ability to know which one actually implants. Therefore, if the total embryo score is to be used as the predictor for P(A1), the corrected total embryo score (calculated as described above) should logically be used as the predictor for P(A2|A1). The proper subset for estimating regression coefficients for predictors of P(A2|A1) consists of pregnant women with at least two embryos transferred. Again the best models developed on the model-building data set were also applied to the holdout data, the cycle 2 data, and the cycle 3 data. The best logistic regression model for P(A2|A1) consisted of one first order predictor, the

corrected total embryo score.

A similar approach was taken to develop the model for predicting P(A3|A2). Again, corrected predictors were calculated including correcting the number of embryos by subtracting 2 and calculation of the corrected total embryo score as the product of the corrected number of embryos transferred and the mean embryo score. The subset used to search for predictors was women pregnant with at least two implantations who had transfer of at least 3 embryos. However, there were no statistically significant predictors in any of the data sets. Accordingly, the uniform probability defined by Equation (3.1) was assigned.

$$P(A3|A2) = \text{(number with at least 3 implantations)/(number with at least 2 implantations)} \quad (3.1)$$

The final estimates for the regression coefficients for the logistic regression models for all three cycles are given in Table 3. The results of fitting the new model to the data from cycles 1 - 3 are provided below.

Results of the Model Fitting Process for the Cycle 1 Data

3.3 Logistic Regression Model for P(A1)

A total of 667 trials resulted in 199 successes (29.8%).

$$\text{Logit}(\hat{p}(A1)) = 0.3839 - 0.0728(\text{Age}) + 0.0727(\text{Total Embryo Score}) \quad (3.2)$$

Table 3

Parameter Estimates from Logistic Regression Analysis of Cycles 1 - 3

Cycle 1

Fit	Trials N	Success N(%)	Intercept	S.E.	Total Embryo Score	S.E.	Age	S.E.	P^2
1	667	199(29.8)	0.3839	0.8120	0.0727	0.0143	-0.0728	0.0225	0.0001
2	196	89(45.4)	-1.8942	0.4923	0.1239	0.0338	....	....	0.0001
3	88	30(34.1)	....	....	....	....	....	....	....

Cycle 2

Fit	Trials N	Success N(%)	Intercept	S.E.	Total Embryo Score	S.E.	Age	S.E.	P^2
1	285	74(26.0)	1.5617	1.2641	0.0505	0.0193	-0.1048	0.0354	0.0001
2	74	21(28.4)	-2.9609	0.9471	0.1271	0.0544	....	....	0.0141
3	21	11(52.4)	....	....	....	....	....	....	....

Cycle 3

Fit	Trials N	Success N(%)	Intercept	S.E.	Total Embryo Score	S.E.	Age	S.E.	P^2
1	107	22(20.6)	1.8340	2.2865	0.0595	0.0327	-0.1267	0.0657	0.0207
2	22	6(27.3)	-5.0951	2.1489	0.2308	0.1130	....	....	0.0094
3	6	2(33.3)	....	....	....	....	....	....	....

Where fit = 1 indicates fit for logit(P(A1), fit = 2 indicates fit for logit(P(A2|A1)), and fit = 3 indicates fit for logit(P(A3|A2)).

†Significance of the combined effects of the explanatory variables based on -2 log likelihood test.

3.4 Logistic Regression Model for P(A2|A1)

A total of 196 trials resulted in 89 successes (45.4%).

$$\text{Logit}(\hat{p}(A2|A1)) = -1.8942 + 0.1293(\text{Total Embryo Score}) \qquad (3.3)$$

3.5 Uniform Probability Model for P(A3|A2)

A total of 88 trials resulted in 30 successes (34.1%).

$$\hat{p}(A3|A2) = 0.3409 \qquad (3.4)$$

3.6 Effectiveness of the Predictors in Cycle 1

The regression coefficients for Equation (3.2) were highly significant ($P=0.0012$ for age and $P=0.0001$ for total embryo score). The regression coefficient for total embryo score in Equation (3.3) was also highly significant ($P=0.0002$). The combined effects of the regression coefficients were highly significant ($P<0.0001$) for both Equations (3.2) and(3.3). There were no effective predictors for $\hat{p}(A3|A2)$ ($P \geq 0.32$).

Results of the Model Fitting Process for the Cycle 2 Data

3.7 Logistic Regression Model for P(A1)

A total of 285 trials resulted in 74 successes (26.0%).

$$\text{Logit}(\hat{p}(A1)) = 1.5617 - 0.1048(\text{Age}) + 0.0505(\text{Total Embryo Score}) \qquad (3.5)$$

3.8 Logistic Regression Model for P(A2|A1)

A total of 74 trials resulted in 21 successes (28.4%).

$$\text{Logit}(\hat{p}(A2|A1)) = -2.9609 + 0.1271(\text{Total Embryo Score}) \quad (3.6)$$

3.9 The Uniform Probability Model for P(A3|A2)

A total of 21 trials resulted in 11 successes (52.4%).

$$\hat{p}(A3|A2) = 0.524 \quad (3.7)$$

3.10 Effectiveness of the Predictors in Cycle 2

The regression coefficients for Equation (3.5) were highly significant ($P=0.0031$ for age and $P=0.0088$ for total embryo score). The regression coefficient for total embryo score in Equation (3.6) was also highly significant ($P=0.0196$). The combined effects of the regression coefficients were statistically significant ($P=0.0001$ for Equation (3.5) and $P=0.014$ for Equation (3.6). There were no effective predictors for $\hat{p}(A3|A2)$ ($P \geq 0.26$).

Results of the Model Fitting Process for the Cycle 3 Data

3.11 Logistic Regression Model for P(A1)

A total of 107 trials resulted in 22 successes (20.6%).

$$\text{Logit}(\hat{p}(A1)) = 1.8340 - 0.1267(\text{Age}) + 0.0595(\text{Total Embryo Score}) \quad (3.8)$$

3.12 Logistic Regression Model for P(A2|A1)

A total of 22 trials resulted in 6 successes (27.3%).

$$\text{Logit}(\hat{p}(A2|A1)) = -5.0951 + 0.2308(\text{Total Embryo Score}) \quad (3.9)$$

3.13 The Uniform Distribution Model for P(A3|A2)

A total of 6 trials resulted in 2 successes (33.3%).

$$\hat{p}(A3|A2) = 0.333 \quad (3.10)$$

3.14 Effectiveness of the Predictors in Cycle 3

The number of trials for cycle 3 was quite small. Nevertheless, the regression coefficients for Equation (3.8) were close to achieving statistical significance ($P=0.054$ for age and $P=0.069$ for total embryo score). The regression coefficient for total embryo score in Equation (3.9) was statistically significant ($P=0.04$). The combined effects of the regression coefficients were significant ($P<0.02$) for Equation (3.8) and $P< 0.01$ for Equation (3.9). There were no effective predictors for $\hat{p}(A3|A2)$ ($P\geq 0.39$).

Calculating Expected Probabilities

3.15 Calculating Expected

Probability of Conception (Tables 4-6)

After calculating the $g(\hat{p}(A1)) = \text{logit}(\hat{p}(A1))$ from Equation (3.1), $\hat{p}(A1)$ was obtained by substituting $g(\hat{p}(A1))$ into Equation (3.11):

$$\hat{p} = (e^{g(\hat{p})})/(1+e^{g(\hat{p})}). \quad (3.11)$$

The expected probability of pregnancy (of all orders), P(A1), in a cycle (as calculated using the appropriate regression coefficients from Table 3) and multiplied by 100 to yield percent appears in Tables 4-6.

3.16 Using the Expected Probabilities to Check for Lack of Fit

First, the appropriate cycle-specific constants from Table 3 were inserted into Equations (3.2), (3.3), and (3.4) for cycle 1, Equations (3.5), (3.6), and (3.7) for cycle 2, or (3.8), (3.9), and (3.10) for cycle 3. The appropriate cycle-specific Equations were then used to estimate the logits of the predicted probabilities for each patient using the patient specific values for each predictor variable. After calculating the $g(\hat{p}(A1)) = \text{logit}(\hat{p}(A1))$ and $g(\hat{p}(A2|A1)) = \text{logit}(\hat{p}(A2|A1)$ from the appropriate pair of logistic regression equations, the logit^{-1} function was used to find the estimated value for $\hat{p}(A1)$ and $\hat{p}(A2|A1)$. These results were then used to calculate the expected probability for singletons, twins, and triplets for each patient using Equations (2.7), (2.8), and (2.9). The final results of this series of calculations were then summed over all of the patients in a given cycle to yield the expected values appearing in Table 7. The Chi-square lack of fit test showed that the new model produced predictions which conformed closely to the observed values for cycle 1, cycle 2, and cycle 3 (Table 7).

Table 4

Probability of Conception in Cycle 1
Tabulated by Age, Mean Puissant
Embryo Score, and Number of
Embryos Transferred (N)

		-------Mean Puissant Embryo Score-------					
Age	N	1	2	3	4	5	6
28	1	17	18	19	20	22	23
	2	18	20	23	25	28	31
	3	19	23	27	31	36	41
	4	20	25	31	38	45	52
	5	22	28	36	45	54	63
32	1	13	14	15	16	17	18
	2	14	16	18	20	23	25
	3	15	18	22	25	30	35
	4	16	20	25	31	38	45
	5	17	23	30	38	47	56
36	1	10	11	12	12	13	14
	2	11	12	14	16	18	20
	3	12	14	17	20	24	28
	4	12	16	20	25	31	38
	5	13	18	24	31	40	49
40	1	8	8	9	10	10	11
	2	8	10	11	12	14	16
	3	9	11	13	16	19	23
	4	10	12	16	20	25	31
	5	10	14	19	25	33	41

All probabilities are multiplied by 100 to yield percent.
Age = the patient's age in years at the time of the in vitro fertilization attempt; N = the number of embryos transferred to the uterus; mean Puissant embryo score = the mean Puissant score of the embryos transferred to the uterus.

Table 5

Probability of Conception in Cycle 2
Tabulated by Age, Mean Puissant
Embryo Score, and Number of
Embryos Transferred (N)

		-------Mean Puissant Embryo Score-------					
Age	N	1	2	3	4	5	6
28	1	21	22	23	24	25	26
	2	22	24	26	28	30	32
	3	23	26	29	32	35	39
	4	24	28	32	36	41	46
	5	25	30	35	41	47	54
32	1	15	16	16	17	18	18
	2	16	17	18	20	22	23
	3	16	18	21	23	26	29
	4	17	20	23	27	31	36
	5	18	22	26	31	37	43
36	1	10	11	11	12	12	13
	2	11	12	13	14	15	17
	3	11	13	15	17	19	21
	4	12	14	17	20	23	27
	5	12	15	19	23	28	33
40	1	7	7	8	8	8	9
	2	7	8	9	10	11	12
	3	8	9	10	12	13	15
	4	8	10	12	14	17	19
	5	8	11	13	17	20	25

All probabilities are multiplied by 100 to yield percent.
Age = the patient's age in years at the time of the in vitro fertilization attempt; N = the number of embryos transferred to the uterus; mean Puissant embryo score = the mean Puissant score of the embryos transferred to the uterus.

Table 6

Probability of Conception in Cycle 3
Tabulated by Age, Mean Puissant
Embryo Score, and Number of
Embryos Transferred (N)

		------Mean Puissant Embryo Score-------				
Age	N	2	3	4	5	6
28	1	17	18	19	20	20
	2	19	20	22	25	27
	3	20	24	27	31	34
	4	22	27	32	37	43
	5	25	31	37	44	52
32	1	11	11	12	13	13
	2	12	13	15	16	18
	3	13	16	18	21	24
	4	15	18	22	26	31
	5	16	21	26	32	39
36	1	7	7	8	8	9
	2	8	9	10	11	12
	3	9	10	12	14	16
	4	10	12	14	18	21
	5	11	14	18	22	28
40	1	4	4	5	5	5
	2	5	5	6	7	7
	3	5	6	7	9	10
	4	6	7	9	11	14
	5	7	9	11	15	19

All probabilities are multiplied by 100 to yield percent.
Age = the patient's age in years at the time of the in vitro fertilization attempt; N = the number of embryos transferred to the uterus; mean Puissant embryo score = the mean Puissant score of the embryos transferred to the uterus.

Table 7

Lack of Fit Analysis for the New Model in Cycles 1-3

	Predicted	Observed
Cycle 1		
Non-pregnant†	468.1	468
Total Pregnancies	198.90	199
Singleton†	110.02	110
Twins†	59.227	59
Triplets or more†	29.653	30
Lack of Fit Chi-square	0.0050	
Degrees of freedom	1	
	N.S††	
Cycle 2		
Non-pregnant†	211.074	211
Total Pregnancies	73.926	74
Singleton†	52.266	53
Twins†	10.476	10
Triplets or more†	11.185	11
Lack of Fit Chi-square	0.0350	
Degrees of freedom	1	
	N.S	
Cycle 3		
Non-pregnant†	85.031	85
Total Pregnancies	21.969	22
Singleton†	14.975	16
Twins†	4.6901	4
Triplets or more†	2.3039	2
Lack of Fit Chi-square	0.2118	
Degrees of freedom	1	
	N.S	

†The lack of fit was calculated using the 4 expected values on the lines marked with "†".
††N.S. indicates not statistically significant ($P > 0.05$).

Expected Distribution of Pregnancies Among the Various Orders

3.17 Calculation of Expected Probabilities of Singleton, Twin and Triplet Pregnancies Given that Pregnancy had Occurred (Tables 8 - 10)

The expected distribution of pregnancies among singleton, twin, and triplet and higher order gestations (Tables 8 - 10) required a series of calculations. First, the appropriate cycle-specific constants from Table 3 were inserted into Equations (3.3) and (3.4) for cycle 1, Equations (3.6) and (3.7) for cycle 2 or (3.9) and (3.10) for cycle 3. The appropriate cycle-specific equations were then used to estimate the logits of the predicted probabilities for $P(A2|A1)$ for each patient using the patient specific values for the total embryo score variable. Then the logit^{-1} function was used to find the estimated value for $\hat{p}(A2|A1)$. As can be seen from Equations (2.13) - (2.15) and Equations (3.12) - (3.14), these three expected values can all be calculated from the expected values for $P(A2|A1)$ and $P(A3|A2)$,

$$P(B1|A1) = 1 - P(A2|A1) \tag{3.12}$$

$$P(B2|A1) = P(A2|A1) \times (1 - P(A3|A2)) \tag{3.13}$$

and

$$P(B3|A1) = P(A2|A1) \times P(A3|A2) \tag{3.14}$$

and they do not depend on the value of $P(A1)$. Since $P(A2|A1)$ depends only on the total embryo score and $P(A3|A2)$ is a uniform probability specific to each cycle, $P(B1|A1)$, $P(B2|A1)$, and $P(B3|A1)$ do not depend on age even though $P(A1)$ is a function of age.

Table 8

Expected Pregnancy Order in Cycle 1

N	X	Mean Puissant Embryo Score 1	2	3	4	5	6
1	1	100	100	100	100	100	100
	2	0	0	0	0	0	0
	3	0	0	0	0	0	0
2	1	85	84	82	80	78	76
	2	15	16	18	20	22	24
	3	0	0	0	0	0	0
3	1	84	80	76	71	66	60
	2	11	13	16	19	23	26
	3	6	7	8	10	12	14
4	1	82	76	69	60	51	42
	2	12	16	21	26	32	38
	3	6	8	11	14	17	20
5	1	80	71	60	48	36	25
	2	13	19	26	34	42	49
	3	7	10	14	18	22	25

All probabilities are multiplied by 100 to yield percent of pregnancies of the specified order tabulated by number of embryos transferred and mean Puissant embryos score, where N = the number of embryos transferred to the uterus, X = the order of the resulting pregnancy (1=singleton, 2 = twins, 3 = triplets or higher order), and mean Puissant embryo score = the average score of the embryos transferred.

Table 9

Expected Pregnancy Order in Cycle 2

N	X	------Mean Puissant Embryo Score------ 1	2	3	4	5	6
1	1	100	100	100	100	100	100
	2	0	0	0	0	0	0
	3	0	0	0	0	0	0
2	1	94	94	93	92	91	90
	2	6	6	7	8	9	10
	3	0	0	0	0	0	0
3	1	94	92	90	87	84	81
	2	4	6	7	9	11	14
	3	2	2	3	4	4	5
4	1	93	90	86	81	74	66
	2	5	7	10	14	19	24
	3	2	3	4	5	7	10
5	1	92	87	81	72	60	48
	2	6	9	14	20	28	37
	3	2	4	5	8	11	15

All probabilities are multiplied by 100 to yield percent of pregnancies of the specified order tabulated by number of embryos transferred and mean Puissant embryos score, where N = the number of embryos transferred to the uterus, X = the order of the resulting pregnancy (1=singleton, 2 = twins, 3 = triplets or higher order), and mean Puissant embryo score = the average score of the embryos transferred.

Table 10

Expected Pregnancy Order in Cycle 3

N	X	------Mean Puissant Embryo Score------ 1	2	3	4	5	6
1	1	100	100	100	100	100	100
	2	0	0	0	0	0	0
	3	0	0	0	0	0	0
2	1	99	99	99	98	98	98
	2	1	1	1	2	2	2
	3	0	0	0	0	0	0
3	1	99	98	98	96	94	91
	2	1	1	2	2	4	6
	3	0	1	1	1	2	3
4	1	99	98	95	91	84	72
	2	1	2	3	6	11	19
	3	0	1	2	3	5	9
5	1	98	96	91	80	62	39
	2	1	2	6	13	25	41
	3	1	1	3	7	13	20

All probabilities are multiplied by 100 to yield percent of pregnancies of the specified order tabulated by number of embryos transferred and mean Puissant embryos score, where N = the number of embryos transferred to the uterus, X = the order of the resulting pregnancy (1=singleton, 2 = twins, 3 = triplets or higher order), and mean Puissant embryo score = the average score of the embryos transferred.

P(B2|A1), and P(B3|A1) do not depend on age even though P(A1) is a function of age.

Interpretation of the Regression Coefficients from the New Model

3.18 An Example using Equation (3.1)

Regression coefficients of logistic models which are linear in the predictor variables can be interpreted in terms of odds ratios.[28] Thus, the regression coefficients of Equation (3.15) for cycle 1,

$$\text{Logit}(\hat{p}(A1)) = 0.3839 + 0.8120(\text{total embryo score}) - 0.0728(\text{age}) \qquad (3.15)$$

can be interpreted using the relationship between the odds ratio and the regression coefficient implicit in the logit function.

$$[\text{odds in favor/odds against}] = e^{(\text{regression coefficient})} \qquad (3.16)$$

Thus, every unit increase in total embryo score increases the odds ratio by 7.5% (since $e^{0.0727} = 1.0754$), while transfer of an additional perfect embryo (a perfect score = 6) increases the odds ratio by 55% (since $e^{(6 \times 0.0727)} = 1.5468$). Aging one year decreases the odds ratio to 93% of that for a person 1 year younger (since $e^{-0.0728} = 0.9298$) while a person 40 has an odds ratio only 48% as high as that of a person age 30 (since $e^{(10 \times -0.0728} = 0.4829$).

Other Logistic Regression Models of Interest

3.19 Two Other Logistic Regression Models

There are two other models, Models II and III, (Table 11) closely related to the logistic

Interpretation of the Regression Coefficients from the New Model

3.18 An Example using Equation (3.1)

Regression coefficients of logistic models which are linear in the predictor variables can be interpreted in terms of odds ratios.[28] Thus, the regression coefficients of Equation (3.15) for cycle 1,

$$\text{Logit}(\hat{p}(A1)) = 0.3839 + 0.8120(\text{total embryo score}) - 0.0728(\text{age}) \qquad (3.15)$$

can be interpreted using the relationship between the odds ratio and the regression coefficient implicit in the logit function.

$$[\text{odds in favor/odds against}] = e^{(\text{regression coefficient})} \qquad (3.16)$$

Thus, every unit increase in total embryo score increases the odds ratio by 7.5% (since $e^{0.0727} = 1.0754$), while transfer of an additional perfect embryo (a perfect score = 6) increases the odds ratio by 55% (since $e^{(6 \times 0.0727)} = 1.5468$). Aging one year decreases the odds ratio to 93% of that for a person 1 year younger (since $e^{-0.0728} = 0.9298$) while a person 40 has an odds ratio only 48% as high as that of a person age 30 (since $e^{(10 \times -0.0728} = 0.4829$).

Other Logistic Regression Models of Interest

3.19 Two Other Logistic Regression Models

There are two other models, Models II and III, (Table 11) closely related to the logistic regression models used above (Model I) which are also potentially of use for estimating P(A1) and P(A2|A1). Model II, which uses age, number of embryos transferred, and

Table 11

Three Logistic Regression Models for Logit($\hat{p}$(A1))

Fitted to the Cycle 1 Data

Criterion	Model I[1]		Model II		Model III	
AIC	775.574		777.076		815.020	
SC	789.082		795.087		819.523	
-2 Log L	769.574		769.076		813.020	
Association of Predicted Probabilities and Observed responses						
Concordant	64.4%		65.4%		62.5%	
Discordant	32.4%		34.1%		35.6%	
Tied	3.2%		0.6%		1.9%	
Parameter Estimates		S.E.		S.E.		S.E.
Intercept	0.3839[2]	0.8120	-0.7755[2]	0.9325	0.7264[2]	0.8105
Age	-0.0728[3]	0.0225	-0.0739[3]	0.0225	-0.0825[3]	0.0223
No. Transfered	....	....	0.3214[3]	0.0810	0.3099[3]	0.0790
Average Embryo Score	....	....	0.2693[3]	0.0795	....	....
Total Embryo Score	0.0727[3]	0.0143	....	....	....	....

[1]Logistic Regression Models:

Model I: Logit($\hat{p}$(A1)) = β_0 + β_1(Age) + β_2(Total Embryo Score)

Model II: Logit($\hat{p}$(A1)) =

β_0 + β_1(Age) + β_2(Number of Embryos Transferred) + β_3(Average Embryo Score)

Model III: Logit($\hat{p}$(A1)) = β_0 + β_1(Age) + β_2(Number of Embryos Transferred)

[2]Not significant

[3]$P \leq 0.001$

(all indicating a poorer model). This simply reflects the fact that information about embryo quality is not utilized by Model III. However, age and number of embryos transferred were still highly statistically significant predictors, and similar results were obtained for the cycle 2 and cycle 3 data as well (data not shown). Potential advantages of one model over another will be discussed below in Section (4.3).

4. DISCUSSION

4.1 Overview of Achievements of the Study

Fitting a model to an existing data set with the intention of applying the result to future trials suffers from a number of potential pit-falls. Fitting a model to a specific set of data will obtain the best fit for that particular set of results. In general, its predictive value for future trials should be expected to be lower than it is for the data to which it has been fitted.[27] By using the hold-out data and the data from cycles 2 and 3 (which had not been used to construct the new model),[27] I was able to demonstrate that the form of the new model could be generalized to other sets of trials from the same IVF program. Furthermore, poor predictive results may occur if there is a change in the conditions under which the trials occur or a difference in the patient population, changes which the program had made every effort to prevent. Despite these potential problems, the present study has provided useful information.

The present study has confirmed the reports of others that 1) the probability of success increases as the total embryo score increases,[2,8] that 2) the chance of success is higher if larger numbers of embryos are transferred,[20,24] and that 3) the chance of success decreases with increasing age.[1,4,7,9,13] Furthermore, the preliminary report[25] from the present study was the first to use conditional probabilities to estimate probabilities of singleton, twin,

and triplet or higher order gestation and to simultaneously adjust for all 4 factors (age, embryo quality, number of embryos transferred, and number of prior unsuccessful IVF attempts with embryo transfer) to calculate patient-specific probabilities for singletons, twins and triplets or higher order gestation. The new model is also the first to successfully overcome the problem of systematically underestimating the risk of multiple gestation which had plagued the models based on the binomial probability distribution.[19-21]

4.2 Comparison of the New Model to Prior Models

Previous models have been based on a variety of simpler assumptions which resulted in their failure to adjust for the multiple predictors of success.[1,3,20,24] None corrected simultaneously for age, number of embryos transferred, and embryo quality. The model used most frequently was the two-parameter model of Speirs et al.[24] This is a simple binomial model with an additional "uterine" factor to account for a hypothetical proportion of the population which was unable to permit implantation. But Speirs et al. provides no way to determine prospectively which women have been disadvantaged in this way. The Speirs model becomes a pure binomial model when the uterine factor is assumed to be 1, at which point all women and all embryos are assumed to be equal under the binomial model of probability. Evaluation of the binomial model[24] demonstrated that it systematically underestimated the percentage of pregnancies that were in the triplets or higher order category.[19-21] The present studies' demonstration that there are multiple important predictors of success shows that the prior models failed because they were overly simplistic.

4.3 Potential Innovative Uses of an Alternative Logistic Regression Model

The alternative Logistic Regression model, Model III (using only age and number of embryos transferred) would still allow computation of tables which would be of assistance for physicians and patients if no embryo scoring data were available. There would be some loss of accuracy since no information about embryo quality would be used. A major advantage of Model III is that it would permit age and embryo number adjusted comparisons between IVF protocols with embryo transfers at different stages of maturity and/or between different IVF programs even if appropriate embryo scoring data were not available.

Making comparisons among several, say C, IVF programs or protocols would only require adding C-1 variables to the Model III, each taking on values of 0 or 1. Such variables are known as indicator, binary variables, or dummy variables.[29] An accurate method for comparing the chance of success achieved by different IVF protocols adjusted for age, number of embryos transferred, and number of prior unsuccessful IVF attempts is urgently needed. Such a method would allow for accurate performance-based assessment of different IVF protocols within the same program or even between different programs.

4.4 The Expected Probabilities of Multiple Gestations in Pregnant Patients are Independent of Age

Although age was a statistically significant negative predictor for P(A1), age was not a statistically significant predictor for either P(A2|A1) or P(A3|A2). So, the probability of pregnancy (of any order), P(A1), decreases with age, but the expected distribution of those

4.3 Potential Innovative Uses of an Alternative Logistic Regression Model

The alternative Logistic Regression model, Model III (using only age and number of embryos transferred) would still allow computation of tables which would be of assistance for physicians and patients if no embryo scoring data were available. There would be some loss of accuracy since no information about embryo quality would be used. A major advantage of Model III is that it would permit age and embryo number adjusted comparisons between IVF protocols with embryo transfers at different stages of maturity and/or between different IVF programs even if appropriate embryo scoring data were not available.

Making comparisons among several, say C, IVF programs or protocols would only require adding C-1 variables to the Model III, each taking on values of 0 or 1. Such variables are known as indicator, binary variables, or dummy variables.[29] An accurate method for comparing the chance of success achieved by different IVF protocols adjusted for age, number of embryos transferred, and number of prior unsuccessful IVF attempts is urgently needed. Such a method would allow for accurate performance-based assessment of different IVF protocols within the same program or even between different programs.

4.4 The Expected Probabilities of Multiple Gestations in Pregnant Patients are Independent of Age

Although age was a statistically significant negative predictor for P(A1), age was not a statistically significant predictor for either P(A2|A1) or P(A3|A2). So, the probability of pregnancy (of any order), P(A1), decreases with age, but the expected distribution of those

Others may decide to accept a higher risk of a multiple gestation in order to increase their chance of achieving pregnancy.[10] Patients come with strong beliefs about many aspects of IVF: beliefs about risks and benefits of transferring certain numbers of embryos and beliefs about what to do if triplets or even more are conceived. It seems reasonable to hope that providing estimates of the probabilities involved will help physicians assist them in making the wisest possible decisions.

5. REFERENCES

1. Hughes EG, King C, Wood EC: A prospective study of prognostic factors in in vitro fertilization and embryo transfer. Fertil Steril 1989;51:838-844.1

2. Puissant F, Van Rysselberge M, Barlow P, Deweze J, Leroy F. Embryo scoring as a prognostic tool in IVF treatment. Hum Reprod 1987;2:705-8.

3. Hershlag A, Kaplan EH, Loy RA, DeCherney AH, Lavy G. Heterogeneity in patient populations explains differences in vitro fertilization programs. Fertil Steril 1991;56:913-7.

4. Hull MGF, Fleming CF, Hughes AO, McDermott A. The age-related decline in female fecundity: a quantitative controlled study of implanting capacity and survival of individual embryos after in vitro fertilization. Fertil Steril 1996;65:783-90.

5. Padilla SL, Garcia JE. Effect of maternal age and number of in vitro fertilization procedures on pregnancy outcome. Fertil Steril 1989;52:270-3.

6. Society for Assisted Reproductive Technology, The American Fertility Society. 1991 results from the Society for Assisted Reproductive Technology generated from the American Fertility Society Registry. Fertil Steril 1993;59:956-62.

7. Tan SL, Royston P, Campbell S, Jacobs HS, Betts J, Mason B, Edwards RG. Cumulative conception and livebirth rates after in-vitro fertilsation. Lancet 1992; 339:1390-4.

8. Leroy F, Puissant F, Barlow P, de Maertelaer G. Guidelines for the prevention of multiple pregnancy in treatment by in vitro fertilization. Acta Genet Med Gemellol 1990;39:371-7.

9. Svendsen TO, Jones D, Butler L, Muasher SJ. The incidence of multiple gestations after in vitro fertilization is dependent on the number of embryos transferred and maternal age. Fert Steril 1996;65:561-5.

10. Azem F, Barak Y, Yaron y, Peyser MR, Amit A, David MP, Yovel I, Lessing JB. Transfer of six or more embryos improves success rates in patients with repeated in vitro fertilization failures. Fertil Steril 1995:63:1043-6.

11. Society for Assisted Reproductive Technology, The American Fertility Society: Assisted reproductive technology in the United States and Canada: 1992 results generated from the American Fertility Society/Society for Assisted Reproductive Technology Registry. Fertil Steril 1994;62:1121-8.

12. Society for Assisted Reproductive Technology, The American Fertility Society: Assisted reproductive technology in the United States and Canada: 1993 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology Registry. Fertil Steril 1995;64:13-21.

13. Society for Assisted Reproductive Technology and the American Society for reproductive Medicine, Assisted reproductive technology in the United States and Canada: 1994 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology Registry. Fert Steril 1996;66:697-705.

14. Seoud, MAF, Toner JP, Kruithoff C, Muasher SJ. Outcome of twin, triplet, and quadruplet in vitro fertilization pregnancies: the Norfolk experience. Fertil Steril 1992;57:825-34.

15. Tallo CP, Vohr B, Oh W, Rubin L, Seifer DB, Haning RV Jr. Maternal and neonatal morbidity associated with in vitro fertilization (IVF). J Pediatr 1995;127:794-800.

16. Haning, RV, Seifer DB, Wheeler CA, Frishman GN, Silver H, Pierce DJ. Effects of fetal number and multifetal reduction on length of in vitro fertilization pregnancies. Obstet Gynecol 1996;87:964-68.

17. Doyle P: The outcome of multiple pregnancy. Hum Reprod 1996;11 Sup 4:110-120

18. Dahl-Lyons CA, Wheeler CA, Frishman GN, Hackett RJ, Seifer DB, Haning RV Jr. Early and late presentation of the ovarian hyperstimulation syndrome (OHSS): Two distinct entities with different risk factors. Hum Reprod 1994;9:792-9.

19. Walters DE: An assessment of two mathematical models of embryo implantation. In Edwards RG, Purdy JM, Steptoe PC: Implantation of the Human Embryo, 1985, Academic Press, pp 219-231.

20. Walters DE, Edwards RG, Meistrich ML. A statistical evaluation of implantation after replacing one or more human embryos. J Reprod Fert 1985;74:557-63.

21. Acosta AA, Muasher SJ, Moon SY, Rosenwaks Z, Oehninger S, Matta JF: Implantation potential of each pre-embryo in multiple pregnancies obtained by in vitro fertilization seems to be different. Fertil Steril 1988;50:906-911.

22. Haning, RV Jr, Goldsmith LT, Seifer DB, Wheeler CA, Frishman G, Sarmento J, Weiss, G: Relaxin secretion in IVF pregnancies. Am J Obstet Gynecol 174:233-40, 1996

23. Haning, RV Jr, Canick JA, Goldsmith LT, Shahinian KA, Erinakes NJ, Weiss G: The effect of ovulation induction on the concentration of maternal serum relaxin in twin pregnancies. Am J Obstet Gynecol 174: 227-232, 1996

24. Speirs AL, Lopata A, Gronow MJ, Kellow GN, Johnston WIH: Analysis of the benefits and risks of multiple embryo transfer. Fertil Steril 1983;39:468-71.

25. Wheeler CA, Cole BF, Frishman GN, Seifer DB, Lovegreen SB, Hackett RJ: Predicting probabilities of Pregnancy and Multiple gestation from in vitro fertilization - a new model. Obst Gynecol 1998;91:696-700.

26. SAS Institute Inc., Sas Technical Report P-200, SAS/STAT Software: Calis and Logistic Procedures, Release 6.04., 1990, SAS Institute Inc., Cary, NC, pp 175-230.

27. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W: Applied Linear Statistical Models, fourth edition. Irwin, Chicago, 1996. pp 434-439.

28. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W: Applied Linear Statistical Models, fourth edition. Irwin, Chicago, 1996. pp 567-615.

29. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W: Applied Linear Statistical Models, fourth edition. Irwin, Chicago, 1996. pp 456-457.

30. Walters DE: The statistical implication of 'number of replacements' in embryo transfer. Hum Reprod 1996;11:10-2.