Research Article

Statistics in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Supplementary Materials for *Analysis of* in vitro *fertilization data with multiple outcomes using discrete time-to-event analysis*

Arnab Maity*, Paige Williams†, Louise Ryan†‡, Stacey Missmer§¶‖, Brent Coull†, Russ Hauser††**

* *Correspondence to: Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, North Carolina 27695, U.S.A., amaity@ncsu.edu*
† Department of Biostatistics, Harvard School of Public Health, Boston, MA
‡ Mathematics, Informatics and Statistics, Commonwealth Scientific and Industrial Research Organisation, Australia
§ Department of Epidemiology, Harvard School of Public Health, Boston, MA
¶ Department of Obstetrics, Gynecology, and Reproductive Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA
‖ Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA
** Department of Environmental Health, Harvard School of Public Health, Boston, MA
†† Vincent Memorial Obstetrics and Gynecology Service, Andrology Laboratory and In Vitro Fertilization Unit, Massachusetts General Hospital, Boston, MA

Prepared using simauth.cls *[Version: 2010/03/10 v3.00]*

1. Example of Sample data records

Here we demonstrate the data recording procedure. Consider data corresponding to two women: woman 1 experiences implantation failure in cycle 1, SAB in cycle 2 but has a live birth in cycle 3 while woman 2 has a live birth in cycle 1, as described in Section 2. The tabular form of the data is given in Table S1.

2. Additional results for BWH IVF data analysis

Here we collect some additional results from the IVF data analysis presented in Section 3 in the main manuscript. Table S2 presents results for analysis based only on first IVF cycle where each IVF failure type is considered separately. Results from models allowing for separate effects of BMI category on each IVF outcome are shown in Figure S1.

3. A Simulation Study

We conducted a simulation study to further evaluate the performance of our method. Based on our motivating study, we assumed a sample size $n = 2500$, a maximum number of cycles to be 6 and three types of failures (denoted by 1, 2 and 3). The three failure types could correspond to implantation failure, chemical pregnancy, and SAB or might represent other ordered outcomes during an IVF cycle. In each cycle, if a woman did not fail at any of the three failure points, she is considered to have a live birth. All women experiencing a failure enter the next cycle, for up to 6 cycles. For the purpose of describing our simulation scenarios, it is convenient to move to a different notation, using three different subscripts to denote woman (ℓ), cycle number (c) and failure type (k).

For each woman, we generated three covariates: X_1 from a Normal$(0,1)$ distribution; X_2 from a Bernoulli(p) distribution with $p = 0.4$ or $p = 0.2$, and X_3 from one of the two scenarios:

- Same value in each cycle per person: $X_{3\ell 1} = \ldots = X_{3\ell J_\ell} = \text{Normal}(0,1)$;
- Different values in each cycle: $(X_{3\ell 1}, \ldots, X_{3\ell J_\ell})$ generated from a multivariate Gaussian distribution with mean 0 and covariance $I + \rho(J - I)$, with $\rho = 0.8, 0.6, 0.4$ and 0.2,

where J_ℓ denotes the number of cycles the ℓ-th woman went through. The covariates X_i can be thought of as representing various woman-specific factors in IVF studies which are assumed to remain constant from one cycle to the next. For example, X_1 and X_2 might represent the initial BMI levels at study entry and history of previous live birth, respectively, and X_3 might represent cycle-specific covariates, or exposure to chemicals such as pesticides, which may change over time.

The probability of failure k, in cycle c for person ℓ was modeled using a logistic model:

$$p_{\ell ck} = H(\alpha_c + \beta_k X_{1\ell} + \gamma X_{2\ell} + \eta X_{3\ell c}),$$

where H is the logistic distribution function, α_c's are cycle specific intercepts, β_k are failure type-specific effect of X_1, and γ and η are effects of X_2 and X_3, respectively. We considered the following two sets of true parameter values:

Case I: $(\alpha_1, \ldots, \alpha_6) = (-2, -1.9, \ldots, -1.5)$, $(\beta_1, \beta_2, \beta_3) = (1.5, 2, 2.4)$, $\gamma = -1$, $\eta = 0.3$
Case II: $(\alpha_1, \ldots, \alpha_6) = (-1, -0.9, \ldots, -0.5)$, $(\beta_1, \beta_2, \beta_3) = (1, 1.5, 1.9)$, $\gamma = -1$, $\eta = 0.4$.

These values were taken such that, for both Case I and II, when all other covariates are fixed, (1) the probability of failure within a cycle depends on age and increases as the person advances towards live birth and (2) probability of each failure type increases as a person goes through more cycles. Compared to Case I, Case II assigns higher failure probabilities to all failure types, resulting in more overall cycles, more women entering each cycle and fewer live births. As an example, the number of woman experiencing each failure type for up to six IVF cycles is displayed in Table S3 for a single set of simulated datasets under Case I and Case II. Here, '# Starting Cycle' denotes the number of women entering the corresponding cycle, 'Failure type' denotes the number (and percent) of failures corresponding to the type (1, 2 or 3) among the women undergoing that cycle and 'Live births' indicates the number of live births at the end of that cycle.

We simulated 1,000 datasets each including 2500 women undergoing up to 6 IVF cycles, and then fit the mis-specified model (which assumes a common intercept for all cycles).

$$p_{\ell ck} \quad = \quad H(\alpha + \beta_k X_{1\ell} + \gamma X_{2\ell} + \eta X_{3\ell c} + U_\ell),$$

with common intercept α. Here U_ℓ denotes the subject-specific Gaussian random effect for the ℓ-th subject.

Simulation results are available in Table S4 for X_2 generated from both a Bernoulli($p = 0.4$) and Bernoulli($p = 0.2$) distribution. Despite the mis-specified model, the mean parameter estimates demonstrated little bias, and had low standard errors. For both the cases of $p = 0.4$ and $p = 0.2$, mean parameter estimates were slightly further from 0 than the true parameter values, by 0.14-0.15 for β_1 of Case I and β_2 of Case II, by 0.17-0.18 for β_2 of Case I and β_3 of Case II, by 0.01-0.04 for ν, and 0.08-0.09 lower than the true parameter estimate of -1 for γ. The empirical standard errors of the parameters estimates ranged between 0.02 to 0.09.

Table S1. Sample data records for a hypothetical woman undergoing multiple IVF cycles.

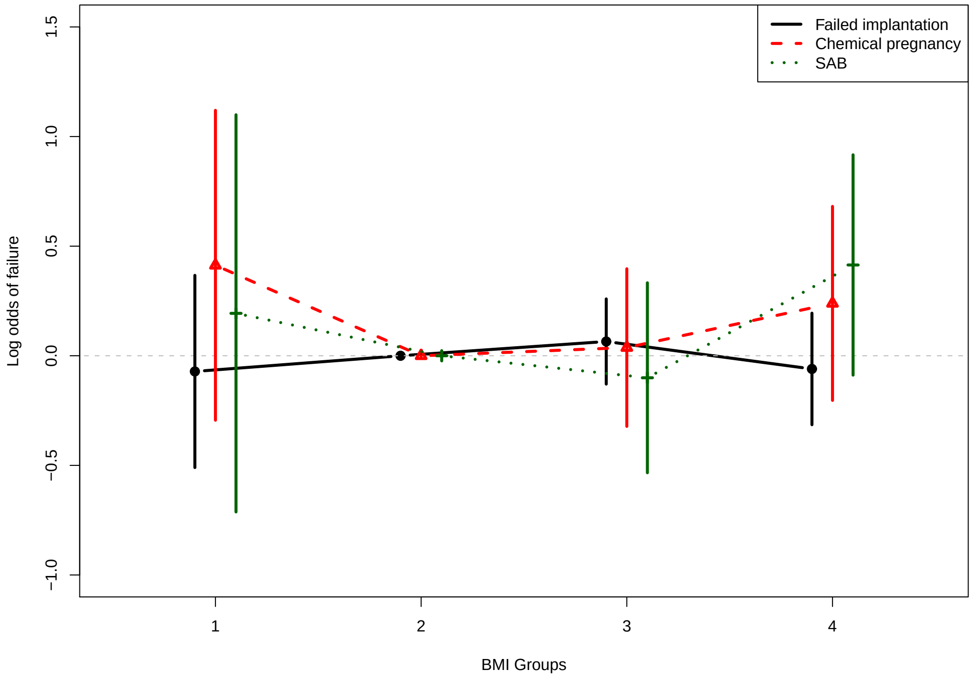
Woman's Id (J)	Cycle number (C)	Failure type (F)[1]	Failure (Y)	Covariates (X)
1	1	1	1	x_1
1	2	1	0	x_2
1	2	2	0	x_3
1	2	3	1	x_4
1	3	1	0	x_5
1	3	2	0	x_6
1	3	3	0	x_7
2	1	1	0	x_8
2	1	2	0	x_9
2	1	3	0	x_{10}

[1] Failure type indicators are: 1=implantation failure, 2=chemical pregnancy and 3=SAB. For this hypothetical example, the woman experiences implantation failure in cycle 1 and SAB in cycle 2 but has a successful live birth in cycle 3.

Table S2. Evaluation of IVF and Participant Characteristics on IVF Failures in Brigham and Women's Hospital IVF Study, for Models Based Only on First IVF Cycle, with Each IVF Failure Type Considered Separately

IVF and Participant Characteristics	Implantation Failure OR (95% CI)	Chemical Pregnancy OR (95% CI)	Spontaneous Abortion OR (95% CI)
Gonadotropin dose	1.31 (1.18, 1.46)	1.17 (0.98, 1.40)	1.08 (0.88, 1.33)
Embryos transferred	0.74 (0.66, 0.83)	0.80 (0.64, 1.00)	0.93 (0.73, 1.87)
Previous livebirth	0.82 (0.65, 1.02)	0.76 (0.50, 1.16)	0.72 (0.43, 1.18)
Down regulation	0.65 (0.50, 0.83)	1.37 (0.83, 2.25)	0.75 (0.41, 1.38)
ICSI	1.16 (0.94, 1.43)	1.20 (0.81, 1.79)	0.74 (0.45, 1.22)
BMI Categories			
< 18.5	0.99 (0.58, 1.69)	1.14 (0.45, 2.92)	1.44 (0.47, 4.47)
$18.5 - 25$	1.00 (referent)	1.00 (referent)	1.00 (referent)
$25 - 30$	1.09 (0.86, 1.39)	0.96 (0.61, 1.52)	1.22 (0.72, 2.06)
> 30	0.97 (0.71, 1.32)	1.33 (0.77, 2.28)	1.01 (0.55, 2.20)
Age Categories			
< 35	1.00 (referent)	1.00 (referent)	1.00 (referent)
$35 - 37$	1.09 (0.86, 1.39)	1.12 (0.72, 1.74)	1.02 (0.59, 1.76)
$38 - 40$	1.45 (1.11, 1.89)	1.39 (0.83, 2.31)	1.69 (0.92, 3.10)
> 40	2.66 (1.85, 3.82)	2.26 (1.11, 4.62)	4.35 (1.93, 9.82)

IVF=in vitro fertilization; OR=odds ratio; CI=confidence interval; SAB=spontaneous abortion; ICSI=intracytoplasmic sperm injection; BMI=body mass index; AIC=Akaike's Information Criterion



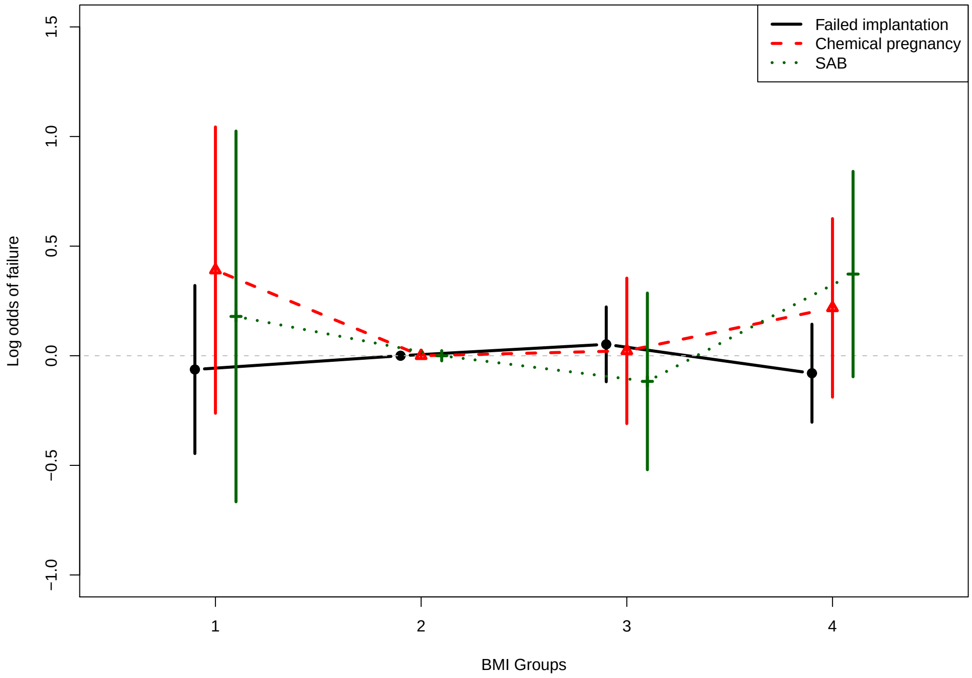


Figure S1. Results of IVF data analysis. Displayed are the log odds ratios for interaction of BMI by failure type with 95% confidence intervals for logistic mixed model (top panel) and transitional model (bottom panel).

Table S3. Two Generated Samples from Simulation Cases I and II Showing Number of Failures by Type Within Each IVF Cycle.

Case I: Lower Failure Probabilities

	IVF Cycle Number					
	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6
# Starting Cycle	2500	882	612	506	449	414
Failure type 1	397 (16%)	299 (34%)	252 (41%)	248 (49%)	255 (57%)	227 (55%)
Failure type 2	285 (11%)	193 (22%)	187 (31%)	135 (27%)	110 (24%)	123 (30%)
Failure type 3	200 (8%)	120 (14%)	67 (11%)	66 (13%)	49 (11%)	36 (9%)
# Live births	1618 (65%)	270 (31%)	106 (17%)	57 (11%)	35 (8%)	28 (7%)

Total number of cycles = 5363
Total number of livebirths = 2114

Case II: Higher Failure Probabilities

	IVF Cycle Number					
	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6
# Starting Cycle	2500	1326	926	747	656	603
Failure type 1	645 (26%)	488 (37%)	405 (44%)	373 (50%)	367 (56%)	348 (58%)
Failure type 2	399 (16%)	297 (22%)	239 (26%)	190 (25%)	161 (25%)	160 (27%)
Failure type 3	282 (11%)	141 (11%)	103 (11%)	93 (12%)	75 (11%)	57 (9%)
# Live births	1174 (47%)	400 (30%)	179 (19%)	91 (12%)	53 (8%)	38 (6%)

Total number of cycles = 6758
Total number of livebirths = 1935

Table S4. Simulation Results for Cases I and II Under Two Scenarios (Covariate X_2 distributed Bernoulli with $p = 0.4$ or 0.2), with Means and Corresponding Standard Errors of Parameter Estimates Based on 1,000 Simulated Datasets.

	Parameters					
	α	β_1	β_2	β_3	η	γ
	Case I, $p = 0.4$, (**true values**)					
X_3 Scenario[1]	-	**(1.5)**	**(2.0)**	**(2.4)**	**(0.3)**	**(-1)**
Constant	-1.98	1.64 (0.04)	2.17 (0.06)	2.60 (0.09)	0.32 (0.03)	-1.08 (0.07)
$\rho = 0.8$	-1.98	1.64 (0.04)	2.17 (0.06)	2.60 (0.09)	0.31 (0.03)	-1.08 (0.07)
$\rho = 0.6$	-1.98	1.64 (0.03)	2.18 (0.06)	2.61 (0.08)	0.32 (0.03)	-1.08 (0.07)
$\rho = 0.4$	-1.98	1.65 (0.04)	2.17 (0.05)	2.61 (0.09)	0.33 (0.03)	-1.09 (0.08)
$\rho = 0.2$	-1.98	1.64 (0.03)	2.17 (0.06)	2.60 (0.08)	0.32 (0.03)	-1.08 (0.07)
	Case I, $p = 0.2$, (**true values**)					
		(1.5)	**(2.0)**	**(2.4)**	**(0.3)**	**(-1)**
Constant	-1.98	1.65 (0.04)	2.18 (0.07)	2.61 (0.10)	0.32 (0.03)	-1.08 (0.08)
$\rho = 0.8$	-1.98	1.64 (0.04)	2.17 (0.07)	2.60 (0.10)	0.32 (0.03)	-1.08 (0.08)
$\rho = 0.6$	-1.98	1.64 (0.04)	2.17 (0.06)	2.61 (0.10)	0.32 (0.03)	-1.08 (0.08)
$\rho = 0.4$	-1.98	1.64 (0.04)	2.17 (0.07)	2.60 (0.09)	0.31 (0.03)	-1.08 (0.08)
$\rho = 0.2$	-1.98	1.64 (0.04)	2.17 (0.07)	2.61 (0.07)	0.31 (0.03)	-1.08 (0.08)
	Case II, $p = 0.2$, (**true values**)					
		(1.0)	**(1.5)**	**(1.9)**	**(0.4)**	**(-1)**
Constant	-0.90	1.11 (0.03)	1.65 (0.05)	2.08 (0.09)	0.44 (0.02)	-1.09 (0.06)
$\rho = 0.8$	-0.90	1.11 (0.03)	1.64 (0.06)	2.07 (0.09)	0.43 (0.02)	-1.09 (0.06)
$\rho = 0.6$	-0.90	1.11 (0.03)	1.64 (0.06)	2.08 (0.08)	0.42 (0.02)	-1.09 (0.06)
$\rho = 0.4$	-0.90	1.10 (0.03)	1.64 (0.05)	2.07 (0.09)	0.42 (0.02)	-1.09 (0.07)
$\rho = 0.2$	-0.90	1.11 (0.03)	1.64 (0.05)	2.08 (0.08)	0.41 (0.02)	-1.09 (0.07)

[1] The row marked by 'Constant' denotes the case where the simulated X_3 covariate remained constant over cycles and the remaining rows correspond to cases where X_3 values vary over cycles, with between cycle correlation of ρ.