A Likelihood Model That Accounts for Censoring Due to Fetal Loss Can Accurately Test the Effects of Maternal and Fetal Genotype on the Probability of Miscarriage

Colin I. O'Donnell^a Charles J. Glueck^b Tasha E. Fingerlin^a Deborah H. Glueck^a

^aDepartment of Preventive Medicine and Biometrics, University of Colorado Denver, Denver, Colo., and

^bThe Cholesterol Center, Jewish Hospital, Cincinnati, Ohio, USA

Key Words

PAI-1 · Pregnancy loss · Thrombophilia · Hypofibrinolysis · Genetics

Abstract

Objective: Heritable maternal and fetal thrombophilia and/or hypofibrinolysis are important causes of miscarriage. Under the constraint that fetal genotype is observed only after a live birth, estimating risk is complicated. Censoring prevents use of published statistical methodology. We propose techniques to determine whether increases in miscarriage are due to the fetal genotype, maternal genotype, or both. **Methods:** We propose a study to estimate the risk of miscarriage contributed by an allele, expressed in either dominant or recessive fashion. Using a multinomial likelihood, we derive maximum likelihood estimates of risk for different genotype groups. We describe likelihood ratio tests and a planned hypothesis testing strategy. **Results:** Parameter estimation is accurate (bias <0.0011, root mean squared error <0.0780, $n = 500$). We used simulation to estimate power for studies of three gene mutations: the 4G hypofibrinolytic mutation in the plasminogen activator inhibitor gene (PAI-1), the prothrombin G20210A mutation, and the Factor V Leiden mutation. With 500 families, our methods have approximately 90% power to detect an increase in the miscarriage rate of 0.2, above a background rate of 0.2. **Conclusion:** Our statisti-

cal method can determine whether increases in miscarriage are due to fetal genotype, maternal genotype, or both despite censoring.

Copyright © 2008 S. Karger AG, Basel

1. Introduction

Fetal and maternal factors may be equally important in the establishment and maintenance of the placental/maternal arterio-venous anastomoses. When increased risk of thrombosis arising from the physiologic hyperestrogenemia of pregnancy is superimposed on familial thrombophilia and/or hypofibrinolysis, then uterine and placental arteries may become thrombosed, leading to vascular uteroplacental insufficiency, and thence to adverse pregnancy outcomes and fetal loss [1, 2]. Polymorphisms in Factor V Leiden, the 4G mutation in the plasminogen activator inhibitor (PAI-1) gene, the CT mutation in the methylenetetrahydrofolate reductase (MTHFR) gene, and the G20210A prothrombin gene

A portion of this paper was submitted to the University of Colorado Denver in partial fulfillment of the requirements for the Masters of Science in Biostatistics for C.I. O'Donnell.

have all been implicated in sporadic and recurrent miscarriage [3–17]. An increase in placental infarcts has been observed when the fetus carries the same gene polymorphism [18]. Both maternal and fetal clotting problems probably increase miscarriage rates.

Thus, if one genotyped maternal-fetal dyads, one should find miscarriages in the order of probability shown in table 1. Whether the miscarriage rate is higher in the group with only maternal thrombophilia and/or hypofibrinolysis than in the group with only fetal thrombophilia and/or hypofibrinolysis, or when mother and fetus are both affected is open to speculation. The true ordering of the rates of miscarriage for the groups when the mother and the fetus are discordant reflects the relative importance of the fetal and maternal contributions to the establishment of good placental circulation.

A conceptus is technically the diploid bundle of cells that exists after conception but before implantation. A fetus refers to a conceptus that has successfully implanted in the endometrium. Because miscarriage can occur at any time during a pregnancy, we will refer to fetuses throughout the manuscript, even though it is possible that implantation has not yet taken place.

The current biostatistical literature for analyzing any combined maternal-fetal-placental risk assumes that the receptivity of the uterus and the viability of the embryo are independent [19, 20]. If maternal and fetal tendencies to thrombophilia and/or hypofibrinolysis can both affect the pregnancy outcome, this assumption must be invalid. Because the fetus inherits half of its genome from its mother, the uterine receptivity and the fetal viability are dependent.

Because the fetal genotype for miscarriage is censored, and may be observed only after a live birth, estimating genetic risk is complicated. The published statistical methods either do not allow for widespread censoring, or require typing members of the extended family. The current literature does allow assessing the relative contribution of the parents and the offspring in diseases where a live birth occurs, such as sudden infant death syndrome and many birth defects. However, in the case of miscarriage, or genetic defects that are lethal long before birth, the degree of censoring prevents use of published statistical methodology. The transmission/disequilibrium test (TDT) requires genotyping of the parents and the ‘affected’ offspring [21], in this case, the miscarried fetus, which is often impossible. Sebastiani et al. [22] are able to provide envelope boundary estimates, despite missing offspring, but cannot provide estimation of exact miscarriage risk. While Mitchell et al. [23] suggest methods for

Table 1. Hypothesized probabilities of miscarriage by phenotype groups

Probability of miscarriage	Maternal thrombophilia	Fetal thrombophilia
Low	no	no
Medium	no	yes
Medium	yes	no
High	yes	yes

discerning maternal and fetal synergistic effects, their method requires grandparents’ genotype. Finally, Wilcox et al. [24], advocate using a log-linear model for case-parent triads, under the assumption that ‘the parent’s fertility and the survival of the fetus to diagnosis are unrelated to the genotype under study’. This assumption can only be relaxed if ‘the probability of survival among fetuses who have developed the defect is independent’ of the allele counts of mother, father, and child, respectively. Neither of these two assumptions is true in our case.

New statistical methods are needed that explicitly model both maternal and fetal contributions to miscarriage risk, despite censoring due to fetal loss. In this paper, we describe such methods. In Section 2, we propose a study of miscarriage and gene effects. In Section 3 we describe a multinomial model, and show how it can be used to estimate miscarriage risk accurately, and test hypotheses about the causes of miscarriage. In Section 4 we conduct power analyses for three possible studies of the effect of common thrombophilic and/or hypofibrinolytic genes on miscarriage. Section 5 is the discussion and conclusion.

2. Study Design

We make several simplifying assumptions. The matings are independent of the genotype at the locus of interest. That is, people are assumed not to choose their reproductive partner on the basis of their thrombophilic or hypofibrinolytic gene mutations. We assume that the alleles at each locus are inherited according to Mendelian expectations: independent assortment and no segregation distortion. Imprinting is assumed to have no effect. All conceptions, all miscarriages and all live births are observed. That is, even if the genotype of the fetus is not observed, we know the outcome of each pregnancy. Occasionally, germ line mutations occur, leading the fetus to have different alleles than those carried by either the mother or the father. This typically occurs at extremely low frequencies, so we make the simplifying assumption that it will not occur to offspring in the study.

The design we consider largely mirrors that of the North Carolina Early Pregnancy Study [25]. We will recruit mating pairs. Both the man and the woman must be willing to be genotyped for the genetic defect of interest. The women must be of child-bearing age, and must wish to become pregnant. All women entered into the study must have had assessment by reproductive endocrinologists, which may include pelvic ultrasound, hysterosalpingogram, CT and MRI as appropriate, and male partner sperm evaluation to rule out non-inherited causes of miscarriage and infertility. Women who have male partner infertility, anatomic barriers to conception, insufficient hormonal production (peri-menopause and menopause), Mullerian fusion abnormalities, infectious diseases, ovulatory dysfunction or luteal phase defect will be excluded. These are all causes of infertility and/or miscarriage and are not related to heritable thrombophilia-hypofibrinolysis. The women must not be undergoing hormonal treatment with endogenous progesterone, estrogen, or human chorionic gonadotropin.

Women will collect their urine each morning and test it with a First Response, Early Results home pregnancy test. Characteristics of this home pregnancy test have been quantified by the USA hCG Reference Service, Department of Obstetrics and Gynecology, School of Medicine, University of New Mexico, Albuquerque, N.M., USA. The test is accurate to 12.5 mIU/ml hCG [26], and has 95% confidence to detect pregnancy the first day after missed menses. Women whose home pregnancy test yields a positive result will visit their family doctor for verification of the pregnancy and report the clinical findings to us. Follow-up on pregnant women will determine if the pregnancy led to a live birth.

The women and their male partners will be genotyped. For ethical and practical reasons, we cannot genotype fetuses that are miscarried. After any live birth, the neonate will be genotyped using a sample of umbilical cord blood.

To allow one set of likelihood equations for both a dominant and recessive model of inheritance, we adopt the following notation. Under the assumption of a recessive mode of inheritance, let m indicate the thrombophilic-hypofibrinolytic polymorphism, and M indicate the wild type allele. Under the assumption of a recessive mode of inheritance, a person can express the thrombophilic-hypofibrinolytic phenotype only if they are homozygous for the recessive polymorphism and carry mm . Note that we do not assume complete penetrance. Heterozygotes (mM) and wild type homozygotes (MM) are unaffected.

For a dominant model of inheritance, let M indicate the thrombophilic-hypofibrinolytic polymorphism, and m indicate the wild type allele. Under this model form, heterozygotes (mM) and homozygotes (MM) may be affected, and homozygous wild types (mm) are unaffected.

Families can be classified into nine mating types depending on the genotype of the mother and the genotype of the father, both of which can always be observed. For example, the mating type where both parents are heterozygous can be labeled mM , mM , where the first genotype is the mother's and the second is the father's. Let $i \in \{1, 2, \dots, 9\}$ index the mating types. The nine mating types, with their index given in parentheses, are (1) mm , mm ; (2) mm , mM ; (3) mm , MM ; (4) mM , mm ; (5) mM , mM ; (6) mM , MM ; (7) MM , mm ; (8) MM , mM ; and (9) MM , MM . This setup is described in table 2.

Table 2. All possible mating pair and offspring combinations

Maternal genotype	Paternal genotype	Fetal genotype	Number of		
			conceptions	miscarriages	live births
mm	mm	mm	C_1	B_1	N_1
mm	mM	mm	C_2	B_2	N_{21}
mm	mM	mM			N_{22}
mm	MM	mM	C_3	B_3	N_3
mM	mm	mm	C_4	B_4	N_{41}
mM	mm	mM			N_{42}
mM	mM	mm	C_5	B_5	N_{51}
mM	mM	mM			N_{52}
mM	MM	MM	C_6	B_6	N_{62}
mM	MM	MM			N_{63}
MM	mm	mM	C_7	B_7	N_7
MM	mM	mM	C_8	B_8	N_{82}
MM	mM	MM			N_{83}
MM	MM	MM	C_9	B_9	N_9

Notice that the genotype of miscarried fetuses is unobserved, N_i , C_i and B_i are the number of live births, conceptions and miscarriages for the i -th mating type, respectively.

In this study we will observe the number of conceptions, live births, miscarriages, and elective abortions in each of the 9 mating types to account for all possible pregnancy outcomes. Here we introduce some notation for the observed and unobserved data. Let C_i indicate the number of fetuses in each mating type indexed by i . Let B_i indicate the number of miscarriages. C_i and B_i are always observed. Notation is also needed to index live births. Let $j \in \{1, 2, 3\}$ index the genotype of the fetuses, with $j = 1$ indicating a genotype of mm , $j = 2$ indicating a genotype of mM , and $j = 3$ a genotype of MM . Then let N_{ij} indicate the number of live births in the i -th mating type with the j -th genotype, and

$$N_i = \sum_{j=1}^3 N_{ij}$$

be the total number of live births for the i -th mating type. Notice that not all three genotypes of fetuses will be expressed in every mating type, and only the offspring who are born alive can be genotyped. The total number of conceptions is equal to the total number of live births plus the total number of miscarriages in each family category ($C_i = N_i + B_i$).

3. Estimation of Miscarriage Rates

Calculating the miscarriage rates from observed data in table 2 would be simple if all fetuses could be genotyped. Because the miscarried fetuses are not genotyped, the problem becomes more complicated.

Our goal is to estimate the following conditional miscarriage probabilities,

$$\begin{aligned}\theta &= \Pr\{\text{Miscarriage} \mid \text{Mother and fetus are } mm\} \\ \phi &= \Pr\{\text{Miscarriage} \mid \text{Mother is } mm; \text{ fetus is not } mm\} \\ \omega &= \Pr\{\text{Miscarriage} \mid \text{Mother is not } mm; \text{ fetus is } mm\} \\ \psi &= \Pr\{\text{Miscarriage} \mid \text{Neither mother nor fetus are } mm\}.\end{aligned}\quad (1)$$

The null hypothesis is that the conditional miscarriage rates shown in Equation 1 are all the same. This occurs only if maternal and fetal genotype had no effect.

The alternatives of interest are (1) that miscarriage is affected only by the maternal genotype; (2) that miscarriage is affected only by the fetal genotype; (3) that maternal genotype or fetal genotype, or both contribute to the miscarriage rate. Each hypothesis corresponds to constraints on the miscarriage rates. Under the null hypothesis, the miscarriage rate is the same no matter what the maternal or the fetal genotype is, so H_0 is $\theta = \phi = \omega = \psi$. Under alternative 1, only the maternal genotype has an effect, so H_1 is $\{\theta = \phi\} > \{\omega = \psi\} > \{\theta \neq \omega\}$. Under alternative 2, only the fetal genotype has an effect, so H_2 is $\{\theta = \omega\} > \{\phi = \psi\} > \{\theta \neq \phi\}$. Under alternative 3, all the rates could be different, so H_3 is $\theta \neq \phi \neq \omega \neq \psi$.

The derivation of the likelihood under the null and the alternatives depends on the constraints on the parameters and two further assumptions. First, we assume that we can use Mendelian laws to predict the probability of conceiving a fetus of a given genotype, given the parental genotypes. Secondly, we assume that the genotype of the father does not affect the outcome of the pregnancy once the genotype of the fetus has been determined. Mathematically, that means that one can write, for example:

$$\begin{aligned}\Pr\{\text{miscarriage} \mid \text{Mother} = mm; \text{Father} = mM; \text{Fetus} = mm\} \\ = \Pr\{\text{miscarriage} \mid \text{Mother} = mm, \text{Fetus} = mm\}\end{aligned}\quad (2)$$

Then the likelihood under the null, and under each alternative is the product of independent multinomials, and is shown in the Appendix.

The likelihood ratio test allows one to decide which likelihood, and thus which hypothesis, is more probable, given the observed data. To differentiate among the hypotheses, there are several strategies that might be used, depending on the a priori hypothesis. While all strategies are logically equivalent, they can have power against different alternatives. We suggest a planned, backwards stepwise method, using a testing cascade of three likelihood ratio tests. The testing cascade is shown in figure 1. We chose the cascade shown in Figure 1 as it had higher power to detect fetal rather than maternal effects.

Fig. 1. Testing Cascade. The first test, of H_0 vs. H_3 , determines whether there is any genetic effect on miscarriage rate. The second test, of H_1 vs. H_3 , determines whether only the mother's genotype affects the miscarriage risk, or whether the fetal genotype affects the risk. The final test, of H_0 vs. H_1 , determines whether the miscarriage risk is due to the fetal genotype, or if both the maternal and fetal genotype affect the miscarriage risk.

In a study with high power, one can interpret a non-significant result as evidence of the veracity of the null hypothesis. We make the assumption here that we have carefully designed a study with power greater than 90%. The first test compares H_0 versus H_3 , and is a χ^2 test with 3 degrees of freedom. A non-significant result leads us to conclude that neither the maternal nor the fetal genotype affects the miscarriage risk. If the result is significant, we conclude that there is a genetic risk for miscarriage, and continue in our testing, to determine whether the risk is maternal, fetal, or due to both maternal and fetal effects. The second test evaluates the relative likelihood of H_1 and H_3 . It is a χ^2 test with 2 degrees of freedom. A non-significant result means that the fetal genotype has no effect on the risk of miscarriage, and thus that all the genetic risk must be caused by maternal effects.

By contrast, a significant result indicates that the fetal genotype is an important risk factor, and that having only a maternal effect in the model cannot fully explain the pattern of risk. To decide whether a combination of fetal and maternal effects, or only a fetal effect (with no maternal effect) is more likely, we compare the relative likelihood of H_1 and H_0 . This is a χ^2 test with 1 degree of freedom. A non-significant test shows that there is fetal effect, but no maternal effect. A significant result argues that the maternal and fetal genome both act to affect the risk of miscarriage.

Formulae for the likelihoods under the null and under each alternative are given in the Appendix. For the null, and each alternative, parameter estimation is done using maximum likelihood techniques. The maximum likelihood estimators can be found analytically only in special cases involving zero counts or other boundary conditions. Away from the boundaries, we use a 0.005 step-sized, incremental grid search over the support of the parameters $\mathcal{R}^4[0,1]$ to obtain initial estimates of the values. The initial estimates are then refined using Powell's Set Direction method [27], modified to run in Mathematica 5.1, © Wolfram Research, Inc. The algorithm is stopped when it reaches the limits of machine accuracy. Convergence is rapid.

We conducted simulation experiments to test the accuracy of the parameter estimation method. We used a total sample size of 500 conceptions, distributed in the mating types assuming Hardy-Weinberg equilibrium and a fixed frequency of 0.2 for a single genetic polymorphism. We assumed a recessively acting allele with an autosomal mode of inheritance. We then fixed a set of representative constant values in the parameter space for the multinomial probabilities for miscarriage. Live births and miscarriage counts were randomly generated for the number of families in each group using the set multinomial probabilities. The counts of live births and miscarriages in each group were passed to the optimization routine and estimates were returned for the four parameter values. The process was repeated 30,000 times and the bias, standard deviation, and root mean squared error (RMSE) were calculated for each parameter. The standard deviation of the estimate describes the precision of the estimate. Bias is the difference between the expected value of the estimate and the true parameter value. The RMSE is a measure of the overall difference between the estimate and the true value, including both bias and imprecision. Results are shown in table 3, indicating virtually no bias in the estimators and low RMSE in the estimates.

Table 3. True parameter values, mean estimated values, bias, standard deviation (SD), and root mean squared error (RMSE), from 30,000 iterations of a simulated study on 500 families, distributed by Hardy-Weinberg equilibrium into nine family genotypes for a dominant model

Parameter	Mean estimate	[Bias]	SD	RMSE
$\theta = 0.9$	0.901079	0.001079	0.045463	0.04548
$\phi = 0.7$	0.701021	0.001021	0.064648	0.06465
$\omega = 0.5$	0.500534	0.000534	0.077943	0.07794
$\psi = 0.2$	0.199987	0.000013	0.023847	0.02385

Table 4. Validation of the testing cascade under each model

Model	θ	ϕ	ω	ψ	% correctly identified
Combined	0.90	0.70	0.50	0.20	96.8%
Maternal	0.80	0.80	0.20	0.20	94.8%
Fetal	0.65	0.20	0.65	0.20	73.5%
Null	0.20	0.20	0.20	0.20	95.6%

Results are for 5000 simulated randomized trials of each.

The variance of the estimates in table 3 arises from two sources: the variability due to the maximum likelihood estimation, and the variation due to the number of live births and miscarriages observed for each mating type. We have assured that the variation due to the likelihood maximization is small. However, the variation in the number of live births is relatively larger. Multinomial variation is a function of sample size.

We used a simulation to examine how well the testing cascade worked. The true model in nature is either no genetic effect, a maternal only effect, a fetal only effect, or a combination of maternal and fetal effects. For a fixed sample size of 500 families and conceptions, we chose reasonable values for the parameters for each one of these models. We simulated the number of live births and conceptions, and conducted the hypothesis testing cascade. The results are shown in table 4.

The accuracy of the testing cascade depends on several factors. These are (1) the number of families in the study; (2) the population frequency of the allele that raises miscarriage risk; (3) the mode of inheritance of that allele; (4) the difference between the miscarriage risk for unaffected mothers and fetuses, and that for pregnancies

where the mother may be affected, the fetus may be affected, or both may be affected, and (5) the choice of the Type 1 error rate (α level) for each hypothesis test in the testing cascade.

For the choices of parameter values and α levels used in the simulation study, the testing cascade detected the true population model more than 94% of the time for the model with no genetic risk, the maternal only model, and for the model where both the mother and the fetal genome affected risk. It was less accurate for a fetal only model. For a real study, it would seem advisable to first run a pilot to estimate the miscarriage rates, and then design a validation study, with carefully chosen allocation of the Type I error rate to maximize the percentage of times the study would select the correct population model.

Naturally, with a complex hypothesis testing cascade as proposed here, care must be taken to avoid inflated Type 1 Error rates. Splitting the α level between the various tests allows information about the biology to guide where maximum power will be needed. In our simulations we set $\alpha_1 = 0.001$ for the comparison of H_0 and H_3 , set $\alpha_2 = 0.048$ for the comparison of H_1 and H_3 , and set $\alpha_3 = 0.001$ for the comparison of H_0 and H_1 . The overall Type I error for the set of hypothesis tests is at most $\alpha = 0.05 \alpha_1 + \alpha_2 + \alpha_3$.

4. Example Power Analysis

To see if the modeling method described here would be practical for actual studies, power analyses were conducted for the recessively acting PAI-1 4G allele and the dominantly acting prothrombin G20210A allele. These polymorphisms were chosen because they have been implicated in clotting disorders and because the population frequencies are relatively high. To conduct the power analysis, several simplifying assumptions are necessary. PAI-1 homozygous 4G individuals occur in the population at a rate of about 0.20 [3, 28–31]. The alleles are assumed to be in Hardy-Weinberg equilibrium, leading to an allele frequency of $\sqrt{0.2} = 0.447$. We use an allele frequency for prothrombin G20210A of [32, 33].

We wanted to assess the ability of our model to differentiate between the null hypothesis, with no genetic effects on the miscarriage rate, and a special case of H_3 , in which either the mother or the fetus, or both having the at risk genotype(s) increased the risk. For the recessive model, we fixed $\psi = 0.2$, which is roughly the population background miscarriage rate [34]. Recall that is the mis-

carriage rate when neither the mother nor the fetus are *mm*. To estimate power, we varied θ , ϕ , and ω in unison between 0.2 and 0.5, incrementing the value by step sizes of 0.005. For the dominant model, we fixed $\theta = 2$ and varied ψ , ϕ , and ω between 0.2 and 0.5.

For each given set of parameters, and fixed sample size, we estimated the power using the following steps. Each mating pair was assumed to generate a single conception. For example, with mating assumed to be random and independent of genotype, a sample of 500 matings leads to expected counts of 20 homozygote-homozygote matches, 47 homozygote-homozygote wild-type matches, 123 heterozygote-heterozygote matches, 152 heterozygote-homozygote wild-type matches, 98 heterozygote-homozygote matches, and 60 homozygote wild-type homozygote matches. Using multinomial sampling laws, the numbers of miscarriages and live births in each mating type were randomly assigned. From the resulting data set, maximum likelihood estimates of the rate parameters were calculated; a difference in the $-2 \log$ likelihoods for the test of hypotheses was obtained; and finally the acceptance or rejection of the χ^2 statistic for H_3 versus H_0 at the predetermined α level was recorded. These steps were repeated 1000 times for each value of the parameters. Empirical power was calculated by the number of rejections, divided by 1000.

The power was calculated for 500 conceptions for the dominant prothrombin G20210A polymorphism (fig. 2), and 500 conceptions for the recessive PAI-1 polymorphism (fig. 3). For samples of size 500 there is roughly $90 \pm 0.1\%$ power to detect an increase in the miscarriage rate of 0.2, above a background rate of 0.2.

5. Discussion

It is plausible that familial thrombophilia and hypofibrinolysis play an important role in miscarriage. The genetic predisposition to clotting disorders works in synergism with the estrogen-mediated thrombophilia of pregnancy. The miscarriages occur either through thrombosis of the spiral arteries of the uterus, placental infarcts leading to placental insufficiency, or both [1, 2]. However, an estimate of the risk of miscarriage attributable to thrombophilia-hypofibrinolysis is difficult to obtain. Most parents are never genotyped before attempting a pregnancy. In addition, the number of total miscarriages in humans is strongly underestimated. Many pregnancy losses occur early, often before the woman knows she is pregnant. Even pregnancy losses

Fig. 2. Estimated power for a study of PAI-1 4G, a recessively acting allele with an autosomal mode of inheritance. The graph shows the estimated power for the test of no genetic effect against an alternative where there is increased risk if either the mother, the fetus, or both are affected. We assume that the miscarriage rate is 0.2 (the normal background rate) when neither mother or fetus are hypofibrinolytic. When mother or fetus, or both may have thrombophilia or hypofibrinolysis, we vary the miscarriage rate between 0.2 and 0.5. The power to detect an increase in the miscarriage rate of about 0.20 for a sample of 500 families with one conception each is approximately 90%. The power curve is based on 1000 iterations per data point.

Fig. 3. Estimated power for a study of prothrombin G20210A, or V Leiden, a dominantly acting allele with an autosomal mode of inheritance. The graph shows the estimated power for the test of no genetic effect against an alternative where there is increased risk if either the mother, the fetus, or both are affected. We assume that the miscarriage rate is 0.2 (the normal background rate) when neither mother or fetus are affected. When mother or fetus, or both may have hypofibrinolysis or thrombophilia, we vary the miscarriage rate between 0.2 and 0.5. The power to detect an increase in the miscarriage rate of 0.185 for a sample of 500 families with one conception each is approximately 90%. The power curve is based on 1000 iterations per data point.

that are recognized as miscarriages are rarely reported to physicians.

This paper examines the problem of heritable thrombophilia-hypofibrinolysis for only one gene. It is possible that many genes act in concert to promote miscarriage. A simple model was chosen because it seems reasonable to first test the hypothesis that a single gene can affect the mother, the baby, and the miscarriage rate.

This paper describes an accurate and powerful method for testing the hypothesis that fetal genotype does contribute to the risk of miscarriage. If a study of the type proposed in the paper proves that inherited single gene defects do affect the risk of miscarriage, we can build more complex statistical models that will include other genes, epigenetic factors and time dependent covariates such as estrogen levels.

Our results suggest that a prospective clinical study of the PAI-1 4G allele, the prothrombin G20210A polymor-

phism, or the V Leiden mutation, using around 500 families would be successful in determining whether fetal genotype, maternal genotype, or both are risk factors for miscarriage. In addition, since the MTHFR CT mutation occurs at roughly the same allele frequency as the genes considered in our power analysis, studies of those genes would also be possible, with slightly higher sample sizes necessary. Thus, we have the tools to examine the maternal and fetal effects for gene mutations commonly implicated as pathoetiologies for miscarriage.

In a classical linear or log-linear model, one can determine the form of interaction, i.e. whether the increase in risk due to maternal and fetal effects is additive, or multiplicative. In our design, which uses maximum likelihood estimation and likelihood ratio tests, we cannot determine the form of the interaction. However, understanding the form of the interaction would make little clinical difference in treatment decisions.

Information obtained from such a study would lead to a better understanding of how familial hypofibrinolysis or thrombophilia cause miscarriage. Pre-conception genetic counseling would allow the identification of women likely to experience hypofibrinolysis-thrombophilia mediated miscarriage, and also allow the prediction that a fetus could be affected. Knowing the etiology of miscarriage is the key to planning studies of Lovenox thromboprophylaxis [35, 36]. The eventual aim is reduce the incidence of miscarriage.

Appendix

Recall that

$$\begin{aligned}\theta &= \Pr\{\text{Miscarriage} \mid \text{Mother and fetus are } mm\} \\ \phi &= \Pr\{\text{Miscarriage} \mid \text{Mother is } mm; \text{ fetus is not } mm\} \\ \omega &= \Pr\{\text{Miscarriage} \mid \text{Mother is not } mm; \text{ fetus is } mm\} \\ \psi &= \Pr\{\text{Miscarriage} \mid \text{Neither mother nor fetus are } mm\}.\end{aligned}\quad (3)$$

The general log likelihood for all models is the product of nine independent binomial or multinomial likelihoods, one for each mating type $i \in \{1, 2, \dots, 9\}$. We demonstrate the derivation of the log likelihoods for $i = 1$ and 2, as examples, and then give the full log likelihood.

When $i = 1$, the genotypes of the mother, the father and the fetus are all mm . The probability of miscarriage is θ and the probability of a live birth is $1 - \theta$. Thus, the likelihood for this mating type is binomial. Ignoring the constant, the log likelihood is

$$\log[L_1(\theta)] = B_1 \log(\theta) + N_1 \log(1 - \theta). \quad (4)$$

When $i = 2$, the genotype of the mother is mm , the genotype of the father is mM , and the genotype of the fetus is either mm or mM . There are three possible outcomes for the pregnancy. The probability of miscarriage is $(\theta + \phi)/2$. The probability of a live birth for fetal genotype mm is $(1 - \theta)/2$, and the probability of a live birth for fetal genotype mM is $(1 - \phi)/2$. Again ignoring the constant, the resulting trinomial log likelihood is

$$\log[L_2(\theta, \phi)] = B_2 \log\left(\frac{\theta + \phi}{2}\right) + N_{21} \log\left(\frac{1 - \theta}{2}\right) + N_{22} \log\left(\frac{1 - \phi}{2}\right). \quad (5)$$

The log likelihood for the entire study is obtained by adding the nine log likelihoods (one for each mating type), collecting like terms, and dropping most of the additive constants. Thus,

$$\begin{aligned}L(\theta, \phi, \omega, \psi) &= B_1 \log(\theta) + (N_1 + N_{21}) \log(1 - \theta) \\ &\quad + (B_6 + B_7 + B_8 + B_9) \log(\psi) \\ &\quad + (N_{42} + N_{52} + N_{53} + N_6 + N_7 + N_8 + N_9) \log(1 - \psi) \\ &\quad + B_3 \log(\phi) + (N_{22} + N_3) \log(1 - \phi) \\ &\quad + (N_{41} + N_{51}) \log(1 - \omega) \\ &\quad + B_2 \log\left(\frac{\theta + \phi}{2}\right) + B_4 \log\left(\frac{\omega + \psi}{2}\right) + B_5 \log\left(\frac{\omega + 3\psi}{4}\right).\end{aligned}$$

Under H_1 we let $\theta = \phi = \theta'$, and $\omega = \psi = \omega'$. The general log likelihood equation then becomes

$$\begin{aligned}L_{(1)}(\theta', \omega') &= \left(\sum_{i=1}^3 B_i\right) \log(\theta') \\ &\quad + \left(\sum_{i=1}^3 N_i\right) \log(1 - \theta') \\ &\quad + \left(\sum_{i=4}^9 B_i\right) \log(\omega') \\ &\quad + \left(\sum_{i=4}^9 N_i\right) \log(1 - \omega').\end{aligned}$$

Under H_2 we let $\theta = \omega = \theta'$, and $\phi = \psi = \phi'$. The general log likelihood equation under this alternative hypothesis is

$$\begin{aligned}L_{(2)}(\theta', \phi') &= B_1 \log(\theta') + \left(B_3 + \sum_{i=6}^9 B_i\right) \log(\phi') \\ &\quad + (B_2 + B_4) \log\left(\frac{\theta' + \phi'}{2}\right) + B_5 \log\left(\frac{\theta' + 3\phi'}{4}\right) \\ &\quad + (N_1 + N_{21} + N_{41} + N_{51}) \log(1 - \theta') \\ &\quad + (N_{22} + N_3 + N_{42} + N_{52} + N_{53} + N_6 + N_7 + N_8 + N_9) \log(1 - \phi').\end{aligned}$$

Under the null hypothesis H_0 , $\theta = \omega = \phi = \psi$. The general log likelihood with

$$Y = \sum_{i=1}^9 B_i, \quad \text{and} \quad I = \sum_{i=1}^9 N_i,$$

the likelihood under H_0 is

$$L_{(0)} = Y \log(\theta_0) + I \log(1 - \theta_0), \quad (10)$$

with miscarriage rate $\hat{\theta}_0 = 0$ if $Y = 0$, and $\hat{\theta}_0 = 1$ if $I = 0$.

References

- Patnaik MM, Haddad T, Morton CT: Pregnancy and thrombophilia. *Expert Rev Cardiovasc Ther* 2007;5:753-765.
- Gris JC, Lissalde-Lavigne G, Quere I, Dauzat M, Mares P: Prophylaxis and treatment of thrombophilia in pregnancy. *Curr Opin Hematol* 2006;13:376-381.
- Balasa VV, Gruppo RA, Glueck CJ, Stroop D, Becker A, Pillow A: The relationship of mutations in the MTHFR, prothrombin, and PAI-1 genes to plasma levels of homocysteine, prothrombin, and PAI-1 levels in children and adults. *Thromb Haemost* 1999;81:739-744.
- Buchholz T, Lohse P, Rogenhofer N, Kosian E, Pihusch R, Thaler CJ: Polymorphisms in the ACE and PAI-1 genes are associated with recurrent spontaneous miscarriages. *Hum Reprod* 2003;18:2473-2477.

- 5 Coulam CB, Jeyendran RS, Fishel LA, Rousev R: Multiple thrombophilic gene mutations rather than specific gene mutations are risk factors for recurrent miscarriage. *Am J Reprod Immunol* 2006;55:360–368.
- 6 Dossenbach-Glaninger A, van Trotsenburg M, Dossenbach M, Oberkanins C, Moritz A, Krugluger W, Huber J, Hopmeier P: Plasminogen activator inhibitor 1 4G/5G polymorphism and coagulation factor XIII Val-34Leu polymorphism: impaired fibrinolysis and early pregnancy loss. *Clin Chem* 2003;49:1081–1086.
- 7 Glueck CJ, Pranikoff J, Aregawi D, Haque M, Zhu B, Tracy T, Wang P: The factor V Leiden mutation, high factor VIII, and high plasminogen activator inhibitor activity: Etiologies for sporadic miscarriage. *Metabolism* 2005;54:1345–1349.
- 8 Glueck CJ, Wang P, Goldenberg N, Sieve L: Pregnancy loss, polycystic ovary syndrome, thrombophilia, hypofibrinolysis, enoxaparin, metformin. *Clin Appl Thromb Hemost* 2004;10:323–334.
- 9 Glueck CJ, Wang P, Bornovoli S, Goldenberg N, Sieve L: Polycystic ovary syndrome, the G1691A factor V Leiden mutation, and plasminogen activator inhibitor activity: Associations with recurrent pregnancy loss. *Metabolism* 2003;52:1627–1632.
- 10 Glueck CJ, Awadalla SG, Phillips H, Cameron D, Wang P, Fontaine RN: Polycystic ovary syndrome, infertility, familial thrombophilia, familial hypofibrinolysis, recurrent loss of in vitro fertilized embryos, and miscarriage. *Fertil Steril* 2000;74:394–397.
- 11 Glueck CJ, Sieve L, Zhu B, Wang P: Plasminogen activator inhibitor activity, 4G5G polymorphism of the plasminogen activator inhibitor I gene, and first-trimester miscarriage in women with polycystic ovary syndrome. *Metabolism* 2006;55:345–352.
- 12 Guan LX, Du XY, Wang JX, Gao L, Wang RL, Li HB, Wang SX: [Association of genetic polymorphisms in plasminogen activator inhibitor-1 gene and 5,10-methylenetetrahydrofolate reductase gene with recurrent early spontaneous abortion] (in Chinese). *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2005;22:330–3.
- 13 Koleva R, Dimitrova V, Chernev T, Savov A, Karag'ozova Z, Mazneukova V, Kremenski I: [Impact of inherited thrombophilia on the development of some pregnancy complications] (in Bulgarian). *Akush Ginekolog (Sofia)* 2005;44:18–26.
- 14 Grandone E, Colaizzo D, Tiscia G, Margaglione M: More on: factor V Leiden and prothrombin G20210A polymorphisms as risk factors for miscarriage during a first-intended pregnancy: The matched case-control 'NOHA first' study. *J Thromb Haemost* 2006;4:709.
- 15 Gueant-Rodriguez RM, Gueant JL, Debard R, Thirion S, Hong LX, Bronowicki JP, Namour F, Chabi NW, Sanni A, Anello G, Bosco P, Romano C, Amouzou E, Arrita HR, Sanchez BE, Romano A, Herbeth B, Guillard JC, Mutchinick OM: Prevalence of methylenetetrahydrofolate reductase 677T and 1298C alleles and folate status: A comparative study in Mexican, West African, and European populations. *Am J Clin Nutr* 2006;83:701–707.
- 16 Nurk E, Tell GS, Refsum H, Ueland PM, Vollset SE: Factor V Leiden, pregnancy complications and adverse outcomes: The Hordaland Homocysteine Study. *QJM* 2006;99:289–298.
- 17 Ruiz-Quezada S, Vázquez-Del Mercado M, Parra-Rojas I, Rangel-Villalobos H, Best-Aguilera C, Sánchez-Orozco LV, Muñoz-Valle JF: Genotype and allele frequency of PAI-1 promoter polymorphism in healthy subjects from the west of Mexico. Association with biochemical and hematological parameters. *Ann Genet* 2004;47:155–162.
- 18 Dizon-Townson D, Meline L, Nelson L, Varner M, Ward K: Fetal carriers of the factor V Leiden mutation are prone to miscarriage and placental infarction. *Am J Obstet Gynecol* 1997;177:402–405.
- 19 Speirs AL, Lopata A, Gronow MJ, Kellow GN, Johnston WI: Analysis of the benefits and risks of multiple embryo transfer. *Fertil Steril* 1983;39:468–471.
- 20 Zhou H, Weinberg CR: Evaluating effects of exposures on embryo vitality and uterine receptivity in in vitro fertilization. *Stat Med* 1998;17:1601–1612.
- 21 Spielman RS, McInnes RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;52:506–516.
- 22 Sebastiani P, Abad MM, Alpargu G, Ramoni MF: Robust transmission/disequilibrium test for incomplete family genotypes. *Genetics* 2004;168:2329–2337.
- 23 Mitchell LE, Weinberg CR: Evaluation of offspring and maternal effects on disease risk using a family based approach: the 'Pent' design. *Am J Epidemiol* 2005;162:676–685.
- 24 Wilcox AJ, Weinberg CR, Lie RT: Distinguishing the effects of maternal and offspring genes through studies of 'case-parent triads'. *Am J Epidemiol* 1998;148:893–901.
- 25 Wilcox AJ, Weinberg CR, O'Connor JF, Baird DD, Schlatterer JP, Canfield RE, Armstrong EG, Nisula BC: Incidence of early loss of pregnancy. *N Engl J Med* 1988;319:189–194.
- 26 Cole LA, Khanlian SA, Sutton JM, Davies S, Rayburn WF: Accuracy of home pregnancy tests at the time of missed menses. *Am J Obstet Gynecol* 2004;190:1000–1005.
- 27 Press WH, Flannery BP, Teukolsky SA, Vetterling WT: Numerical Recipes in C: The Art of Scientific Computing, ed 2. Cambridge, Cambridge University Press, 1992, pp 412–420.
- 28 Glueck CJ, Kupferminc M, Fontaine RN, Wang P, Weksler BB, Eldor M: Genetic hypofibrinolysis in complicated pregnancies. *Obstet Gynecol* 2001;97:44–48.
- 29 Henry M, Chomiki N, Scarabin PY, Alessi MC, Peiretti F, Arveiler D, Ferrieres J, Evans A, Amouyel P, Poirier O, Cambien F, Juhan-Vaque I: Five frequent polymorphisms of the PAI-1 gene: lack of association between 31 genotypes, PAI activity, and triglycerides levels in a healthy population. *Arterioscler Thromb Vasc Biol* 1997;17:851–858.
- 30 Margaglione M, Cappucci G, D'Addetta M, Colaizzo D, Giuliana N, Vecchione G, Mascolo G, Grandone E, Di Minno G: PAI-1 plasma levels in a general population without clinical evidence of atherosclerosis: Relation to environmental and genetic determinants. *Arterioscler Thromb Vasc Biol* 1998;18:562–567.
- 31 Ridker PM, Hennekens CH, Lindpaintner K, Stampfer MJ, Miletich JP: Arterial and venous thrombosis is not associated with the 4G/5G polymorphism in the promoter of the plasminogen activator inhibitor gene in a large cohort of US men. *Circulation* 1997;95:59–62.
- 32 Bauduer F, Lacombe D: Factor V Leiden, prothrombin 20210A, methylenetetrahydrofolate reductase 677T, and population genetics. *Mol Genet Metab* 2005;86:91–95.
- 33 Zivelin A, Mor-Cohen R, Kovalsky V, Kornbrot N, Conard J, Peyvandi F, Kyrle PA, Bertina R, Peyvandi F, Emmerich J, Seligsohn U: Prothrombin 20210G>A is an ancestral prothrombotic mutation that occurred in whites approximately 24000 years ago. *Blood* 2006;107:4666–4668.
- 34 Torpy JM: Miscarriage. *JAMA* 2002;288:1936.
- 35 Glueck CJ, Wang P, Goldenberg N, Sieve L: Pregnancy loss, polycystic ovary syndrome, thrombophilia, hypofibrinolysis, enoxaparin, metformin. *Clin Appl Thromb Hemost* 2004;10:323–334.
- 36 Gris JC, Mercier E, Quere I, Lavigne-Lissalde G, Cochery-Nouvellon E, Hoffet M, Ripart-Neveu S, Tailland ML, Dauzat M, Mares P: Low-molecular-weight heparin versus low-dose aspirin in women with one fetal loss and a constitutional thrombophilic disorder. *Blood* 2004;103:3695–3699.