The Beta-Geometric Distribution Applied to Comparative Fecundability Studies

Clarice Ring Weinberg, Beth C. Gladen

Biometrics, Volume 42, Issue 3 (Sep., 1986), 547-560.

Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28198609%2942%3A3%3C547%3ATBDATC%3E2.0.CO%3B2-3

BIOMETRICS 42, 547–560
September 1986

The Beta–Geometric Distribution Applied to Comparative Fecundability Studies

Clarice Ring Weinberg and Beth C. Gladen

National Institute of Environmental Health Sciences, Mail Drop B3-02, Statistics and Biomathematics Branch, P.O. Box 12233, Research Triangle Park, North Carolina 27709, U.S.A.

SUMMARY

A convenient measure of fecundability is time (number of menstrual cycles) required to achieve pregnancy. Couples attempting pregnancy are heterogeneous in their per-cycle probability of success. If success probabilities vary among couples according to a beta distribution, then cycles to pregnancy will have a beta–geometric distribution. Under this model, the inverse of the cycle-specific conception rate is a linear function of time. Data on cycles to pregnancy can be used to estimate the beta parameters by maximum likelihood in a straightforward manner with a package such as GLIM. The likelihood ratio test can thus be employed in studies of exposures that may impair fecundability. Covariates are incorporated in a natural way. The model is illustrated by applying it to data on cycles to pregnancy in smokers and nonsmokers, with adjustment for covariates. For a cross-sectional study, when length-biased sampling is taken into account, the pre-interview attempt time is shown to follow a beta–geometric distribution, so that the same methods of analysis can be applied even though all of the available data are right-censored.

For a cohort followed prospectively, there will be some couples enrolled whose fecundability is effectively 0, and for such applications, the beta could be considered to be contaminated by a distribution degenerate at 0. The mixing parameter (proportion sterile) can be estimated by application of the expectation–maximization (EM) algorithm. This, too, can be carried out using GLIM.

1. Introduction

Environmental exposures may affect human reproduction by many diverse mechanisms. Sperm production may be suppressed in the male, or subtle abnormalities in the sperm may impair their ability to fertilize the ovum. The exposed female may experience anovulation, or may produce ova that are nonviable. Her cervical secretions may become hostile to the passage of sperm, or the uterine lining may become abnormal so that the developing blastocyst has trouble implanting. The blastocyst itself may be blighted and be lost in a very early, unrecognized miscarriage.

All such mechanisms lead to a common observable effect: longer time is required on average for affected couples to achieve pregnancy. Information on the number of menstrual cycles required to achieve pregnancy can be gathered with little inconvenience or embarrassment to couples under study. Consequently, this provides a useful epidemiological screening method for detecting effects of human exposures to reproductive toxins.

In this paper, we presume two or more groups are to be compared because of an exposure or condition that may reduce fecundability. Couples attempting pregnancy are to be followed for up to K menstrual cycles, or until pregnancy occurs. We assume K is fairly small, so that aging during the follow-up interval will have negligible effects on the fecundability of any given couple. In practice, K is usually some number less than or equal

Key words: Beta–geometric; EM algorithm; Fecundability.

to 12. Couples will be assumed to vary in their fecundability, so that a given couple has a per-cycle conception probability that stays constant throughout the follow-up interval, but these probabilities vary across couples according to a beta distribution. This approach resembles that of Maruani and Schwartz (1983), who extended to comparative applications a model already familiar to demographers (Sheps and Menken, 1973; Suchindran and Lachenbruch, 1975; Bongaarts, 1975). We will reparametrize and extend the method to allow incorporation of continuous or categorical time-independent covariates. Cross-sectional and retrospective designs will also be considered.

The second section discusses a general model of heterogeneity and derives some of the basic relationships. The third section specializes to the beta distribution. It also describes use of a standard statistical package such as GLIM to do maximum likelihood estimation of the beta parameters. Section 4 describes the use of the model in retrospective studies. The techniques are illustrated in Section 5, where we describe results obtained by applying the model to fecundability data from a recent study comparing smokers with nonsmokers. The example also illustrates the incorporation of covariates. Section 6 introduces the cross-sectional design and shows that the time in cycles from beginning of a pregnancy attempt to interview follows a beta–geometric distribution whose parameters relate in a simple way to the parameters governing time to pregnancy in a prospective study. An extension of the model to allow contamination of the beta by a proportion of completely sterile couples in the sample is discussed in Section 7, where an EM-type algorithm is given that allows the proportion sterile to be estimated iteratively by maximum likelihood. This algorithm is applied to the data on smoking and fecundability, after modifying the data to include a hypothetical subsample of sterile couples.

2. Effects of Heterogeneity

If all noncontracepting, sexually active couples had the same per-cycle conception probability, p, then the number of cycles required to achieve pregnancy would be distributed as geometric with parameter p. In fact, there is ample evidence that couples vary in their fecundability (Tietze, 1968; Leridon and Spira, 1984). About 30% of sexually active couples achieve pregnancy in their first noncontracepting cycle, a smaller proportion of the remaining couples achieve pregnancy in the second, and with each additional unsuccessful cycle, the conception rate continues to decline, as the risk sets become further depleted of the relatively fecund couples. The pronounced decrease in conception probability over time is not properly viewed as a time effect, but as a sorting effect in a heterogeneous population.

Suppose the per-cycle conception probability, p, which we will also refer to as fecundability, is fixed for a given couple, but across couples varies according to some unspecified underlying distribution. Let X denote the number of cycles required for conception, and let $q = 1 - p$. For a given couple with fecundability p, X will follow a geometric distribution:

$$\Pr(X = x \mid q) = q^{x-1}(1 - q).$$

Removing the conditioning on q by integrating over its distribution yields the probability that conception occurs at x for a randomly selected couple:

$$\Pr(X = x) = \mathrm{E}[q^{x-1}(1 - q)] = \mathrm{E}(q^{x-1}) - \mathrm{E}(q^x).$$

Similarly, one can show that

$$\Pr(X > x - 1) = \mathrm{E}(q^{x-1}).$$

Thus, the distribution of X can be written entirely in terms of the moments of q (or p). If K is the maximum number of cycles of follow-up for the study, then the likelihood of the data is completely specified by the first K moments of q. The maximum likelihood estimates

for these moments are simple functions of the number of couples conceiving at each cycle, $E(q)$ being estimated by the proportion not conceiving at cycle 1, $E(q^2)$ by the proportion not conceiving at cycle 1 or at cycle 2, and so on. This was noted previously by Sheps (1964).

An alternate way to write this likelihood is to decompose each term into a series of cycle-specific conditional rates. Define μ_j as the conditional mean of p after $j-1$ failures:

$$\mu_j = E(p \mid X > j-1) = \Pr(X = j \mid X > j-1) = 1 - E(q^j)/E(q^{j-1}).$$

This cycle-specific mean is the analogue of a hazard rate. We may then write

$$\begin{aligned}\Pr(X = x) &= \Pr(X = x \mid X > x-1) \prod_{j=1}^{x-1} \Pr(X > j \mid X > j-1) \\ &= E(p \mid X > x-1) \prod_{j=1}^{x-1} E(q \mid X > j-1) \\ &= \mu_x \prod_{j=1}^{x-1} (1 - \mu_j).\end{aligned}$$

In similar fashion, we have

$$\Pr(X > x) = \prod_{j=1}^{x} (1 - \mu_j).$$

Thus, each couple's contribution to the likelihood can be written as if the data were the result of a sequence of Bernoulli trials with parameters $\mu_1, \mu_2, \ldots$. If factors are grouped by cycle rather than by couple, this is seen to be identical (ignoring the combinatoric factor) to a likelihood arising from a sequence of binomial trials, one at each cycle, where the binomial parameters are the number of couples at risk and μ_j. We will see later that this expression for the likelihood has some computational advantages.

Using these expressions, we can estimate the first K moments for a cohort followed for K cycles. If there are two cohorts, they can be compared by a likelihood ratio test. Unfortunately, this would not be expected to yield a powerful testing procedure, since there are many moments to estimate and we are left with a chi-squared statistic with K degrees of freedom. Thus, it seems reasonable to consider possible parametric descriptions of the distribution of fecundabilities.

3. The Beta–Geometric Model

We will assume now that fecundability follows a beta distribution. We will use the parametrization suggested by Griffiths (1973), involving a mean parameter, μ, and a "shape" parameter, θ, which is zero when there is no heterogeneity. The variance of p is $\mu(1-\mu)\theta/(1+\theta)$. The density of p is

$$p^{(\mu-\theta)/\theta}(1-p)^{(1-\mu-\theta)/\theta} B\left[\frac{\mu}{\theta}, \frac{(1-\mu)}{\theta}\right]^{-1}.$$

The probability that a couple with fecundability p conceives at cycle x is

$$\Pr(X = x \mid p) = (1-p)^{x-1}p.$$

Removing the conditioning on p by integrating over the beta yields the probability that conception occurs at x for a randomly selected couple:

$$\Pr(X = x) = \frac{\int_0^1 (1-p)^{x-1} p(1-p)^{(1-\mu-\theta)/\theta} p^{(\mu-\theta)/\theta}\, dp}{B[\mu/\theta, (1-\mu)/\theta]} = \frac{\mu \prod_{i=1}^{x-1} [1 - \mu + (i-1)\theta]}{\prod_{i=1}^{x} [1 + (i-1)\theta]}.$$

This distribution has mean $(1 - \theta)/(\mu - \theta)$ and variance $\mu(1 - \mu)(1 - \theta)/[(\mu - \theta)^2 \cdot (\mu - 2\theta)]$, whenever these terms are finite. We will call this distribution the beta–geometric. It is a special case of the inverse Polya–Eggenberger distribution described, for example, by Johnson and Kotz (1969, Chap. 9, §4.2); in their notation, set $k = 1$. It is also related to a discrete decreasing failure rate model given in unpublished work by Padgett and Spurrier (University of South Carolina Statistics Technical Report No. 97, 1984) and has a prominent place in the demography literature (e.g., Suchindran and Lachenbruch, 1975).

As was pointed out by Sheps and Menken (1973), we do not need to assume that p is constant across cycles within each couple. No matter what the distribution of p within couples across time, if there is independence from cycle to cycle within each couple and if the couple-specific means follow a beta distribution across couples, then time to pregnancy is beta–geometric.

What will the distribution of p be for couples who have experienced j unsuccessful cycles? If we condition on j failures having occurred, the conditional density of p is:

$$\begin{aligned} f(p \mid X > j) &\propto \Pr(X > j \mid p) f(p) \\ &\propto (1 - p)^j p^{(\mu-\theta)/\theta}(1 - p)^{(1-\mu-\theta)/\theta} \\ &\propto p^{(\mu-\theta)/\theta}(1 - p)^{[1-\mu+(j-1)\theta]/\theta}. \end{aligned}$$

This is again a beta distribution, but now with parameters $[\mu/(1 + j\theta), \theta/(1 + j\theta)]$, instead of (μ, θ). Note that the mean, μ_{j+1}, i.e., the pregnancy rate at cycle $j + 1$, moves toward 0 as the number of unsuccessful cycles increases, and this time-dependence has a simple functional relationship to the beta parameters. Another direct consequence of this result is that the expected remaining waiting time for couples who have experienced j failures increases linearly with j:

$$\mathrm{E}(X - j \mid X > j) = (1 - \theta)/(\mu - \theta) + j\theta/(\mu - \theta).$$

Recall from Section 2 that the likelihood can be expressed as if it were the result of a series of binomial trials whose parameters are the number of couples at risk and μ_j. The product form of $\Pr(X = x)$ given in Section 2 specializes to the beta–geometric distribution given above, following substitution of $\mu_j = \mu/[1 + \theta(j - 1)]$.

A reparametrization is useful at this point. Define new parameters c and d by

$$c = 1/\mu, \quad d = \theta/\mu.$$

The mean rate after $j - 1$ unsuccessful cycles is then

$$\mu_j = 1/[c + d(j - 1)].$$

In terms of these new parameters, the mean of the underlying beta distribution is $1/c$, and its variance is $d(c - 1)/[c^2(c + d)]$. If we assume the mean conception probability is greater than 0 and less than 1, then this variance can be 0 if and only if d is 0, so that the null hypothesis $d = 0$ is equivalent to the hypothesis that the cycles to pregnancy are truly geometrically distributed, i.e., there is no heterogeneity in fecundability among couples in the group under study.

Maximum likelihood estimates for the beta parameters can be computed using packages such as GLIM, which allow the user to specify a function of the (binomial) mean parameter (which in this case is the cycle-specific conception probability) that is linearly related to the predictor variables. In the present case, for a single population, we have the "link" function:

$$h(\mu_j) = 1/\mu_j = c + d(j - 1),$$

i.e., inverse linearity in time.

If two populations are to be compared, so that we estimate c_1 and d_1 for one group and also c_2 and d_2 for another, it is possible that the risk ratio, $(c_2 + d_2 j)/(c_1 + d_1 j)$, crosses 1 in time. This implies that one of the groups may be less fecund initially but more fecund after enough time has passed. One can easily show that the risks remain proportional over time if and only if $c_1/d_1 = c_2/d_2$. Similarly, the odds ratio for pregnancy remains constant over time (as would be required to apply the discrete Cox model) if and only if $(c - 1)/d$ is the same for the two beta distributions. Thus, the beta–geometric model assumes neither proportional risks nor constant odds ratios.

To use this technique for two populations, the above model is extended as follows:

$$1/\mu_j = c_1 + d_1(j - 1) + (c_2 - c_1)I + (d_2 - d_1)(j - 1)I,$$

where I is an indicator variable for the second population. The two beta distributions can now be compared by means of a likelihood ratio test, yielding a chi-squared statistic with 2 degrees of freedom.

Covariates, continuous or categorical, can be incorporated and tested by including them as linear terms in the above model in a fashion similar to the incorporation of group indicators. This will be illustrated in Section 5.

4. Retrospectively Collected Data

The above discussion implicitly assumed the time-to-pregnancy data arose in the context of a prospective study, where two groups of couples attempting pregnancy were followed forward through time. Alternatively, a study can be conducted by interviewing women who are already pregnant to ascertain, for those with planned pregnancies, how many noncontracepting cycles preceded their successful conception. The difference between this retrospective design and the prospective one discussed previously is that the sampling unit for a retrospective study is the pregnancy and not the attempt at pregnancy.

If there is a one-to-one correspondence between attempts and pregnancies, i.e., if every attempt is eventually successful, then the two study designs should provide equivalent information, in the absence of sampling biases. In fact, it is not true in most populations that every decision to begin trying for pregnancy is successful. In a prospective study, there will be a proportion of couples with fecundability so low that this group can be considered to be sterile. The problem of extending the model to allow for this sterile subpopulation will be treated in Section 7. With retrospectively collected time-to-pregnancy data, the proportion sterile is known to be 0, since we begin with women currently pregnant; this simplifies the analysis.

5. Example: Smoking and Fecundability

The beta–geometric model was applied to data from a retrospective study (Baird and Wilcox, 1985) where women who were pregnant with planned pregnancies were asked how many cycles it took them to get pregnant. Only women who had gotten pregnant within 24 cycles of trying were admitted to the study, since it was felt that recall beyond that would be unreliable. (Alternatively, such women could have been included as providing right-censored data.) Since medical interventions tend to begin after 12 unsuccessful cycles (clinical "infertility"), we elected to truncate at cycle 12, treating times reported as longer than that as right-censored.

A total of 678 women with planned pregnancies were interviewed, of whom 654 became pregnant within 12 cycles after discontinuing contraception. Women were classified as current smokers if they reported smoking at least an average of 1 cigarette a day during at least the first cycle they were trying to get pregnant. This yielded 135 smokers.

Because the model we are applying assumes that variation in pregnancy rates over time can be attributed to heterogeneity in fecundability among couples, and not to true time effects, it would be improper to apply it to any group in which true time effects are suspected. Since there is evidence, both in these data and in the literature (Linn et al., 1982; Harlap and Baras, 1984; Spira, Spira, and Schwartz, 1985) that women going off oral contraceptives experience reduced fecundability for several cycles, 92 women whose most recent method of contraception was given as the pill were excluded from these analyses. This left 586 women, contributing a total of 1844 cycles, for analysis. The cycles to pregnancy for smokers and nonsmokers are shown in Table 1.

The first question of interest is whether there is evidence for heterogeneity in this sample of couples. If there is no variability to explain, we need not bother looking at explanatory variables. The maximum likelihood estimates (with standard errors) for c and d are 2.59 (.12) and .362 (.067), respectively. The evidence for heterogeneity among these fecund couples is thus strong. The estimated beta fecundability distribution has mean .386 and standard deviation .17. These estimates were obtained by GLIM, and are displayed in Table 2, which also gives the deviances.

Next we included an indicator term for current smoking status. This allowed the mean of the beta distribution to be different for smokers and also changed the variance. The parameter estimates and their standard errors are shown in the second model in Table 2.

Table 1
Observed cycles to pregnancy

Cycle	Smokers	Non-smokers	Cycle	Smokers	Non-smokers
1	29	198	8	5	9
2	16	107	9	1	5
3	17	55	10	1	3
4	4	38	11	1	6
5	3	18	12	3	6
6	9	22	> 12	7	12
7	4	7			

Table 2
Models for smoking data

Parameter	Estimate	Standard error	Deviance
Constant	2.59	.12	
Time	.36	.07	2231.5
Constant	2.45	.12	
Time	.33	.07	
Smoking	1.18	.40	2218.4
Constant	2.45	.12	
Time	.34	.07	
Smoking	1.18	.49	
Smoking × Time	−.00	.19	2218.3
Constant	2.55	.13	
Time	.32	.06	
Smoking	1.19	.39	
Coital frequency	−.55	.23	2213.9

The Wald-type chi-squared statistic for including the smoking term was 8.7, and the likelihood ratio test gave a chi-squared value of 13.1. To allow for two completely different beta distributions for the two exposure groups, an interaction is required between smoking status and time. The results of this more general model are shown in the third model in Table 2. The interaction term did nothing to improve the fit over the previous model (both the Wald and the likelihood ratio chi-squared values being less than .1), suggesting that in this case simply including a "main effect" would have been adequate. The likelihood ratio test for two different beta distributions versus a single beta distribution yielded a chi-squared value of 13.2 on 2 degrees of freedom.

The estimated beta densities for smokers and nonsmokers are shown in Figure 1. Note that we have by no means explained all of the variability in fecundability by means of this single exposure variable, so that conception rates will still decrease with time in each of the exposure groups. There is, however, a striking difference in fecundability by smoking status. Figure 2 shows the fitted and observed cycle-specific conception probabilities for the first 12 cycles, for smokers and nonsmokers. The fitted values have been joined as an aid to the eye; time is truly discrete in this context.

We were concerned that there may be bias related to the possibility that some of the women reporting pregnancy in the first cycle were pregnant by accident. It was straightforward to repeat the analysis, ignoring the cycle 1 data. The resulting maximum likelihood estimates were very similar to those given in Table 2.

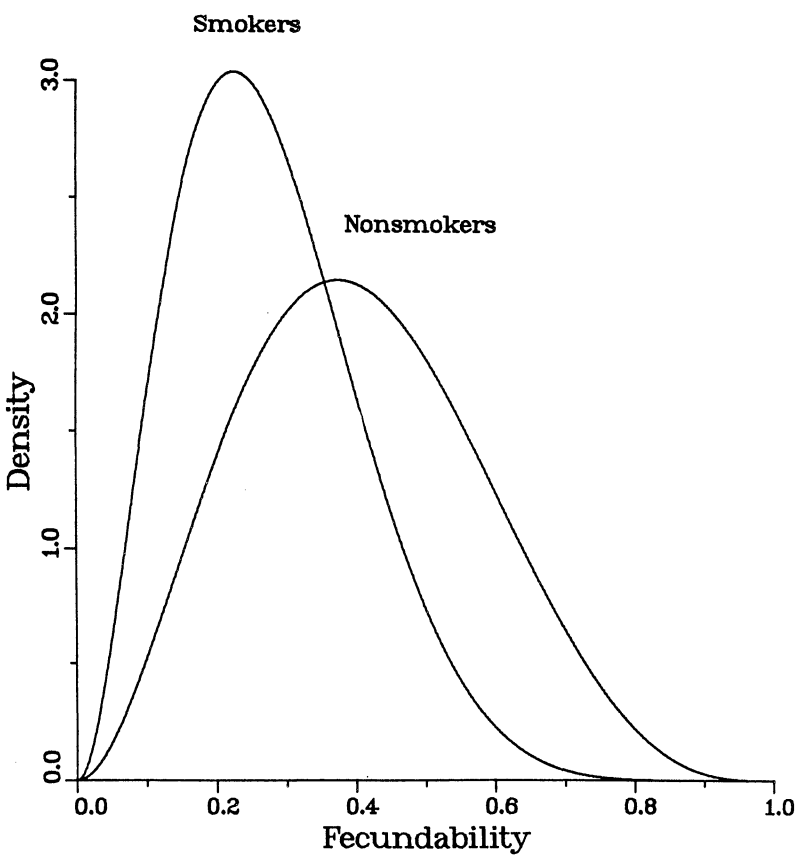

Figure 1. Estimated beta densities as fit by maximum likelihood for the fecundability of smokers and nonsmokers.

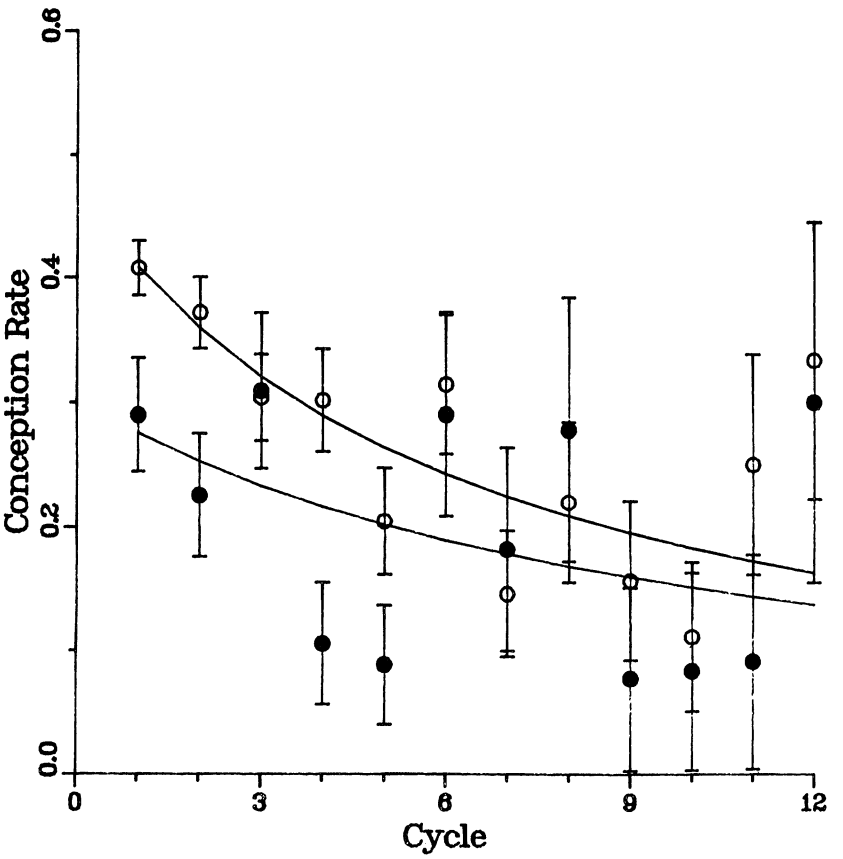

Figure 2. Observed and fitted cycle-specific conception rates for smokers (solid circles) and non-smokers (open circles). Vertical bars indicate plus and minus one standard error.

Covariates available included age (dichotomized at 30), gravidity (number of previous pregnancies), and reported coital frequency. Age had no significant effect and neither did gravidity. However, coital frequency, as one might expect, seemed to be important. When coital frequency was reported to be greater than 4 times a week, fecundability was increased (likelihood ratio chi-squared value = 4.4). The estimated effect of smoking remained stable under dichotomous adjustment for coital frequency, as is shown in the last model in Table 2.

To see how well the beta-geometric fits these data, we compared the moments as estimated from the fitted beta parameters with the moments fitted by nonparametric maximum likelihood for the entire data set. Table 3 shows the results. The two sets of

Table 3
Moments for the distribution of $q = 1 - p$

Moment	Beta-geometric fit	Non-parametric fit	Moment	Beta-geometric fit	Non-parametric fit
1	.615	.613	7	.095	.101
2	.407	.403	8	.077	.077
3	.284	.280	9	.063	.067
4	.207	.208	10	.052	.060
5	.156	.172	11	.044	.048
6	.120	.120	12	.037	.032

moments are remarkably close, suggesting the beta–geometric adequately describes the data in this case.

6. Cross-Sectional Studies

One other type of design that could be employed in comparative fecundability studies would involve a population-based random sample of women who are asked whether they are currently at risk of becoming pregnant. Those who are sexually active and not contracepting are then asked how many cycles have passed since they discontinued contraception, and may be questioned about exposures of interest. The elapsed number of cycles since discontinuing contraception, to be referred to as "attempt time," will be recorded as the exact ordinal reported number of cycles, if less than or equal to some maximal K, and as right-censored if greater than K. Denote the attempt time by T and the total number of cycles to pregnancy by X.

Of what use is the resulting attempt time data? Suppose a nonsterile population is sampled randomly, so that, for example, a couple destined to take 8 cycles to achieve pregnancy is 8 times as likely to be sampled as a couple destined to take 1 cycle. Suppose also that there are no systematic trends in the population, so that beginnings of attempts are uniform in time. Then a couple destined to require 8 cycles is equally likely to be at cycles $1, \ldots, 8$ at the time of interview; i.e., the attempt time, conditional on $X = 8$, is uniformly distributed on the integers $1, \ldots, 8$.

Under the above assumptions, the sampling distribution of T, the attempt time, can be shown in general to be

$$\Pr(T = j) = \Pr(X > j - 1)/\mathrm{E}(X).$$

Similar relations were noted by Freeman and Hutchison (1980). Note that one consequence of this relation is that a nonparametric estimate of the mean of X is the inverse of the proportion of attempts that happen to be sampled at cycle 1.

If the distribution of X is beta–geometric, then a corollary of the above equation is that T is also beta–geometric, but with parameters $[(\mu - \theta)/(1 - \theta), \theta/(1 - \theta)]$ instead of (μ, θ). This means that the methods described above for prospective and retrospective studies of cycles to pregnancy apply just as well to cross-sectional studies of attempt times, even though the attempt time data is, in a sense, entirely right-censored. Similar phenomena have been described by Allison (1985).

The caveat here is that a cross-sectional study of most populations will include sterile couples, heavily overrepresented by length-biased sampling because their recruitable duration may span years. However, methods to be described in the following section can be applied to extend the beta–geometric to allow for contamination by couples with 0 fecundability.

7. Allowing for a Sterile Subpopulation

In a prospective study of couples attempting pregnancy, there will exist couples whose fecundability is low enough to be approximated as 0. Thus, it may be more realistic to consider the distribution of p to be a beta contaminated by a second distribution degenerate at 0, as in Maruani and Schwartz (1983). Let π_0 denote the proportion of sterile couples in the population under study, i.e., the mixing parameter.

After j cycles without conception, the distribution of p in the population remaining at risk is still a mixture of a beta distribution with the degenerate distribution. However, the beta distribution and the mixing parameter both depend on j. A larger and larger fraction of the remaining couples will be from the sterile subpopulation. Let the cycle-specific mixing parameter after j failures be denoted π_j. If the original underlying beta distribution

has parameters c and d, then π_j is governed by the following recursion:

$$\pi_j = \frac{\pi_{j-1}}{\pi_{j-1} + (1 - \pi_{j-1})\{1 - 1/[c + d(j-1)]\}},$$

where the denominator represents the fraction of the risk set at cycle j who are still at risk at cycle $j + 1$.

The problem can be viewed in the missing data framework described by Dempster, Laird, and Rubin (1977); the data can be represented as multinomial with two cells collapsed, which belongs to a curved exponential family. That is, for each woman we know the cycle at which she conceived, if she did. For couples who have not achieved pregnancy in K cycles of trying, we do not know which did not because they are sterile. For concreteness, suppose $K = 12$. The sufficient statistics for the complete-data likelihood are the numbers pregnant at each cycle, $X_1, \ldots, X_{12}$, together with the number who are fecund but unsuccessful, X_{13}, and the number who are sterile, X_{14}. What we are able to observe is $X_1, \ldots, X_{12}, X_{13} + X_{14}$. The full-data likelihood, aside from the combinatoric factor, can be written as

$$\pi_0^{X_{14}}\left[(1 - \pi_0)\left(1 - \sum_1^{12} \alpha_j\right)\right]^{X_{13}} (1 - \pi_0)^{\sum_1^{12} X_j} \prod_1^{12} \alpha_j^{X_j},$$

where α_j, which is a function of the beta parameters, denotes the probability that a couple with nonzero fecundability achieves pregnancy at cycle j. The EM algorithm can now be applied. The expectation step of the algorithm simply solves for $\tilde{X}_{14}$ in the equation

$$\frac{\tilde{X}_{14}}{X_{13} + X_{14}} = \frac{\hat{\pi}_0}{\hat{\pi}_0 + (1 - \hat{\pi}_0)(1 - \sum_1^{12} \hat{\alpha}_j)}.$$

The maximization step then sets $\hat{\pi}_0 = \tilde{X}_{14}/N$, where N is the total number of women, and separately maximizes the beta part of the likelihood.

This algorithm can be implemented in GLIM, at least approximately, by noting that we can remove (by giving them zero weights) a fraction, $\hat{\pi}_{12}$, of the nonpregnant remnant $X_{13} + X_{14}$, and then maximize the part of the likelihood involving the beta distribution, as described above. Thus, weights are recomputed in each iteration following the expectation step.

The beta family is rich enough to include densities that go to infinity at zero, and it was not obvious to us that contamination by a subpopulation with 0 fecundability would be readily detectable. It would not be surprising if the beta mean were biased downward, the variance biased upward, and the proportion sterile seriously underestimated. To address this concern, we decided to append 90 hypothetical nonconceiving women to the Baird and Wilcox (1985) data and find maximum likelihood estimates based on the three-parameter model. This corresponds to a sterility rate of .133. Using the algorithm described above, we estimated the sterility rate to be .102; c and d were estimated as 2.633 and .573, respectively. This corresponds to a beta mean of .380 and standard deviation of .205, compared to the uncontaminated mean of .386 and standard deviation of .17. Thus, while the biases were all in the expected direction, the estimates were not seriously distorted.

Based on the likelihood ratio criterion, a confidence interval for π_0 can be constructed as follows. For a specified fixed value of π_0, one can maximize the likelihood over the beta parameters by applying the EM-type algorithm, as above, except that π_0 is not reset. The upper and lower limits of a 95% confidence interval for π_0 are then derived by finding specific values that yield maximized log likelihoods which differ from the maximized

likelihood value by 3.84/2. For the data with 90 sterile subjects added, the maximum likelihood estimate for π_0 was .10, with an approximate 95% confidence interval of (.02, .15). Similar strategies would yield confidence intervals for the other parameters.

With maximum likelihood methods, two cohorts can be compared as above, now allowing for possibly different sterility rates, by means of a chi-squared likelihood ratio test with 3 degrees of freedom. Again, covariates can be incorporated.

If there is censoring at times less than K, i.e., if couples drop out during the study, the reweighting technique described above can be applied, but with the proportional exclusions carried out on a cycle-by-cycle basis. Alternatively, the likelihood equations can be solved by other means. The likelihood equations, allowing for interim dropouts, for the general three-parameter model, together with the information matrix, are given in the Appendix. This has been given before, for a different parametrization, but the reader is cautioned that there are typographical errors in the published expressions (Maruani and Schwartz, 1983).

For a cross-sectional study of attempt times, the parameter π_0 will be overestimated, due to length-biased sampling. If all reported attempts are included, correction for this bias is not possible. One strategy is to restrict recruitment (i.e., analysis) to couples whose attempt time is at most some L. Then sterile couples will be at risk of being sampled during a known and finite total of L cycles, as compared with an average of

$$\mathrm{E}(X \mid X < L + 1)\Pr(X < L + 1) + L\Pr(X > L)$$

cycles for fecund couples. The resulting distribution of T will be a truncated beta–geometric (truncated at L) mixed with a uniform distribution on $1, \ldots, L$, arising from the sterile subpopulation. The EM algorithm can be applied here again to compute maximum likelihood estimates for μ, θ, and π_0.

8. Discussion

The general notion underlying all of this is that apparent time effects may really just be sorting effects within a population of heterogeneous risk. In effect, the composition of the population changes over time, even when the individuals do not. Stated more generally, both in continuous-time and discrete-time failure-time settings, heterogeneity of risk across individual experimental units is intrinsically aliased with changes in risk over time. Similar difficulties have recently been discussed by Vaupel and Yashin (1985).

Although this paper has focused on fecundability studies, similar issues arise in many other settings where risk may be heterogeneous. For example, the risk of spontaneous abortion varies from woman to woman. This causes a sorting effect analogous to the one discussed here, when looking at the spontaneous abortion risk after a series of spontaneous abortions have already occurred. One might mistakenly conclude from the observed patterns of risk that having had a spontaneous abortion causally increases a woman's risk of having another one. This issue is discussed by Wilcox and Gladen (1982), who give further references. The beta–geometric model could perhaps be applied to reproductive history data, where the outcome is taken as the number of pregnancies before the first live birth. The general model we propose may also be appropriate for certain reliability applications.

The approach described in this paper can be viewed as a regression model with mixed effects. Alternative strategies have been proposed. Stiratelli, Laird, and Ware (1984) have described linear logistic regression with random effects allowed, under the assumption that the random effects are normally distributed. In contexts where a proportion of the population has a response probability of 0, this becomes unworkable since the logit of 0 is

negative infinity and the random effect thus has a very badly behaved distribution. The procedure we have outlined allows for heterogeneity among subjects and allows for a subpopulation with 0 response probability.

If two groups are to be compared with no covariates, one might consider the use of standard survival analysis techniques such as the log-rank test. This can be done, for example, when comparing the smokers and the nonsmokers in the Baird and Wilcox data; the chi-squared value produced is 10.8. However, the log-rank test (equivalent in the discrete case to the Mantel–Haenszel test) is designed for a situation where the odds ratio is constant. Under the beta–geometric model, the odds ratio is nonconstant and may even cross 1, so that the log-rank test may perform poorly.

The procedure we have proposed is easy to apply, provided one already has access to a program such as GLIM that does iteratively reweighted least squares for user-specified "link" functions. Unfortunately, while easy, the EM algorithm is rather slow to converge, and this may be a problem in some computing environments. There are, of course, many standard maximization algorithms that could be adapted to these models; they might provide faster convergence. Unfortunately, however, the incomplete-data likelihood is algebraically quite cumbersome to work with, especially when there are covariates to be included.

The flatness of the likelihood in the mixing parameter suggests that the beta family is rich enough that contamination by a subpopulation with 0 response is difficult to distinguish from a slightly more variable beta. Thus, the mixing parameter is inherently difficult to estimate. However, the estimates of the beta parameters remained quite stable in the presence of contamination: the estimated mean conception probability (for the fecund subpopulation) changed from .386 to .380, and the standard deviation changed from .17 to .20, when the estimation was carried out in the presence of a rather large group of sterile couples.

The discussion in this paper has presumed that two groups are to be compared, with adjustment for time-independent covariates. Extensions to comparisons of more than two groups can be done in the obvious way with the same inverse-μ linear model. Time-dependent covariates could also formally be included, without any modification of the fitting macros, but the interpretation of the results becomes problematic. For example, women who begin to smoke after two cycles of trying to get pregnant may not have the same distribution of conception probabilities at cycle 3 as other women who are also at cycle 3 but have been smoking all along.

ACKNOWLEDGEMENT

We are grateful to Dr Donna Baird for many helpful discussions and for the use of her data.

RÉSUMÉ

Une mesure commode de la fécondité est le temps (nombre de cycles menstruels) nécessaire pour arriver à la grossesse. Les couples désireux d'une grossesse sont hétérogènes du point de vue de la probabilité de succès par cycle. Si les probabilités de succès varient parmi les couples selon une distribution beta, alors, les cycles jusqu'à la grossesse auront une distribution beta–géométrique. Avec ce modèle, l'inverse du taux de conception spécifique d'un cycle est une fonction linéaire du temps. Utilisant ces données sur les cycles jusqu'à la grossesse, les paramètres beta peuvent être estimés simplement par maximum de vraisemblance en utilisant un logiciel tel que GLIM. La rapport de vraisemblance peut donc être employé dans les études de situations qui peuvent diminuer la fécondité. Les covariables sont incluses naturellement. On illustre le modèle en l'appliquant à des données sur le nombre de cycles jusqu'à la grossesse chez des fumeurs et des non-fumeurs, avec ajustement de

covariables. Pour une étude transversale, quand le biais lié à l'échantillonnage est pris en compte, le temps d'attente pour une première entrevue suit une distribution beta–géométrique; par suite, les mêmes méthodes d'analyse peuvent être appliquées quand bien même les données disponibles sont censurées à droite. Pour une cohorte suivie de façon prospective, il y aura des couples dont la fécondité est nulle, et pour de telles applications, on pourra considérer que la distribution beta est contaminée par une distribution dégénérée en 0. Le paramètre du mélange (proportion de stériles) peut être estimé par application de l'algorithme EM de maximisation de l'espérance. On peut aussi utiliser le logiciel GLIM.

References

Allison, P. D. (1985). Survival analysis of backward recurrence times. *Journal of the American Statistical Association* **80,** 315–322.

Baird, D. D. and Wilcox, A. J. (1985). Cigarette smoking associated with delayed conception. *Journal of the American Medical Association* **253,** 2979–2983.

Bongaarts, J. (1975). A method for the estimation of fecundability. *Demography* **12,** 645–660.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39,** 1–38.

Freeman, J. and Hutchison, G. B. (1980). Prevalence, incidence and duration. *American Journal of Epidemiology* **112,** 707–723.

Griffiths, D. A. (1973). Maximum likelihood estimation for the beta–binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics* **29,** 637–648.

Harlap, S. and Baras, M. (1984). Conception-waits in fertile women after stopping oral contraceptives. *International Journal of Fertility* **29,** 73–80.

Johnson, N. L. and Kotz, S. (1969). *Distributions in Statistics: Discrete Distributions.* Boston: Houghton Mifflin.

Leridon, H. and Spira, A. (1984). Problems in measuring the effectiveness of infertility therapy. *Fertility and Sterility* **41,** 580–586.

Linn, S., Schoenbaum, S. C., Monson, R. R., Rosner, B., and Ryan, K. J. (1982). Delay in conception for former "pill" users. *Journal of the American Medical Association* **247,** 629–632.

Maruani, P. and Schwartz, D. (1983). Sterility and fecundability estimation. *Journal of Theoretical Biology* **105,** 211–219.

Sheps, M. C. (1964). On the time required for conception. *Population Studies* **18,** 85–89.

Sheps, M. C. and Menken, J. A. (1973). *Mathematical Models of Conception and Birth.* Chicago: University of Chicago Press.

Spira, N., Spira, A., and Schwartz, D. (1985). Fertility of couples following cessation of contraception. *Journal of Biosocial Sciences* **17,** 281–290.

Stiratelli, R., Laird, N., and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics* **40,** 961–971.

Suchindran, C. M. and Lachenbruch, P. A. (1975). Estimates of fecundability from a truncated distribution of conception times. *Demography* **12,** 291–301.

Tietze, C. (1968). Fertility after discontinuation of intrauterine and oral contraception. *International Journal of Fertility* **13,** 385–389.

Vaupel, J. W. and Yashin, A. I. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *The American Statistician* **39,** 176–185.

Wilcox, A. J. and Gladen, B. C. (1982). Spontaneous abortion: The role of heterogeneous risk and selective fertility. *Early Human Development* **7,** 165–178.

Received December 1985; *revised April and June* 1986.

Appendix

Suppose that fecundabilities have a beta distribution with parameters (μ, θ) contaminated with a proportion π of sterile couples. Assume that dropouts may occur. Then the data can be summarized by $n(y)$, the number who become pregnant at cycle y, and $c(y)$, the number who drop out after y

unsuccessful cycles. Define the following terms for convenience:

$$Q(y) = \prod_{0}^{y-1} [(1 - \mu + j\theta)/(1 + j\theta)],$$

$$A_1(y) = \sum [1/(1 - \mu + j\theta)],$$

$$A_2(y) = \sum [j/(1 - \mu + j\theta)],$$

$$A_3(y) = \sum [j/(1 + j\theta)],$$

$$A_4(y) = \sum [1/(1 - \mu + j\theta)^2],$$

$$A_5(y) = \sum [j/(1 - \mu + j\theta)^2],$$

$$A_6(y) = \sum [j^2/(1 - \mu + j\theta)^2],$$

$$A_7(y) = \sum [j^2/(1 + j\theta)^2].$$

The above sums all run from 0 to $y - 1$ and the usual conventions that empty products equal 1 and empty sums equal 0 are followed. Denote by T the cycle in which pregnancy occurs. Then the model may be expressed by

$$g(y) = \Pr(T = y) = (1 - \pi)Q(y)\mu/[1 - \mu + (y - 1)\theta],$$

$$G(y) = \Pr(T > y) = \pi + (1 - \pi)Q(y),$$

$$L = \text{log-likelihood} = \sum_{1}^{\infty} \{n(y)\log[g(y)] + c(y)\log[G(y)]\}.$$

The likelihood equations are derived by setting the first derivatives of the log-likelihood to zero. These derivatives are:

$$\partial L/\partial\mu = \sum \{n(y)[1/\mu - A_1(y - 1)] - c(y)(1 - \pi)Q(y)A_1(y)/G(y)\},$$

$$\partial L/\partial\theta = \sum \{n(y)[A_2(y - 1) - A_3(y)] + c(y)(1 - \pi)Q(y)[A_2(y) - A_3(y)]/G(y)\},$$

$$\partial L/\partial\pi = \sum \{-n(y)/(1 - \pi) + c(y)[1 - Q(y)]/G(y)\}.$$

The observed information matrix is the negative of the matrix of second derivatives of the log-likelihood. Straightforward calculation yields:

$$I_{\mu\mu} = \sum \{n(y)[1/\mu^2 + A_4(y - 1)] + c(y)(1 - \pi)Q(y)[G(y)A_4(y) - \pi A_1(y)^2]/G(y)^2\},$$

$$I_{\mu\theta} = \sum (-n(y)A_5(y - 1) + c(y)(1 - \pi)Q(y)\{\pi A_1(y)[A_2(y) - A_3(y)] - G(y)A_5(y)\}/G(y)^2),$$

$$I_{\mu\pi} = -\sum [c(y)Q(y)A_1(y)/G(y)^2],$$

$$I_{\theta\theta} = -\sum (n(y)[A_7(y) - A_6(y - 1)] + c(y)(1 - \pi)Q(y)\{G(y)[A_7(y) - A_6(y)] + \pi[A_2(y) - A_3(y)]^2\}/G(y)^2),$$

$$I_{\theta\pi} = \sum \{c(y)Q(y)[A_2(y) - A_3(y)]/G(y)^2\},$$

$$I_{\pi\pi} = \sum \{n(y)/(1 - \pi)^2 + c(y)[1 - Q(y)]^2/G(y)^2\}.$$