THOMAS H. SCHEIKE

Analysis of Time to Pregnancy

Introduction

In this paper, I will review some ongoing work on how to study time trends in human fertility. The basic situation is that it is of interest to study and monitor developments in human fertility, but data that can give insight into this is hard to come by because human fertility is influenced by a myriad of important socioeconomic factors generated by society. It is well known that current trends in modern life as well as the economic situation can influence birth rates quite dramatically.

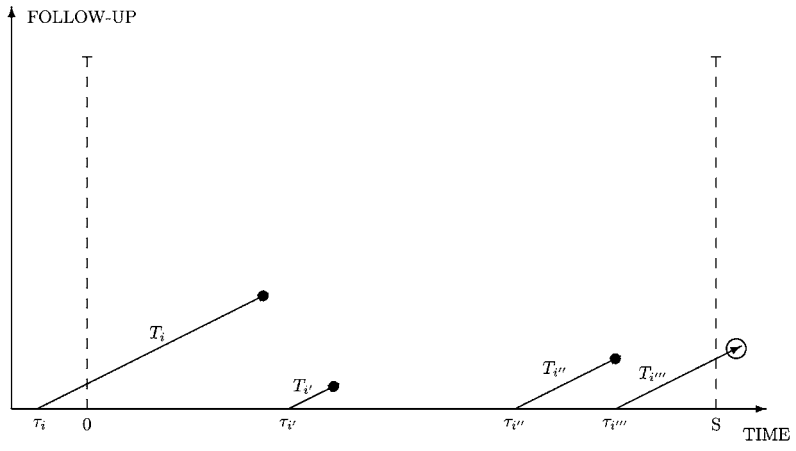
Despite the difficulty in collecting data that contains information solely on biological fertility, it is clear that national and local birth registries might contain relevant information on this. Birth registries typically contain information on all successful pregnancies and how long the couples waited to achieve this pregnancy. Given that the time-to-pregnancy (TTP) contains relevant information on biological fertility, the aim of this review is to describe how one should analyse such data to learn about time trends and covariate effects such as smoking that may also change over time. It turns out that the sampling method in such birth registries must be taken into account to perform such an analysis.

Associate Professor
Thomas Scheike
Department of Biostatistics,
Copenhagen University, Denmark
E-mail: T.Scheike@biostat.ku.dk
CAS Fellow 2005/2006

If data had been prospectively observed, we would simply do a standard logistic regression analysis and then the covariate effects represented by the log-odds-ratio would summarise how smoking or calendar time changed the chance of pregnancy in a particular menstrual cycle.

However, when data are obtained from birth registries and are thus retrospectively sampled, there are various sampling issues to deal with.

Figure 1. Lexis diagram of retrospectively ascertained TTPs. The initiation times are denoted τ, the TTP waiting times T. The full dots indicate TTPs that are sampled and included in the study and open circles indicate TTPs that are not included in the study.



It is easiest to illustrate this point by fixing the initiation time, τ, for a given TTP denoted T. Consider an observation period in calendar time from [0,S] where all conceptions are asked to recall their TTPs and thus also their initiation times. Now, given the initiation, we only observe the TTP when it results in a conception within the period of observation, formally the restriction is that $\tau + T \in [0, S]$. For TTPs initiated at time τ before time 0 they will be observed to be subject to both left truncation (being larger than $-\tau$) and right truncation (being shorter than $S - \tau$). A TTP initiated during [0,S] will be subject to right truncation only ($T < S - \tau$). When looking for time trends, based on retrospectively ascertained TTP, it is obviously crucial to correct for the sampling biases that will lead to long TTPs for those initiated before time 0 and short TTPs for those initiated just before time S.

Models

The basic set-up is as follows. In the ideal setting, we would record the TTPs forward in time from the time of initiation, thus making a prospective study. For a prospective study of TTP, the basic regression model may claim that given a set of covariates then the chance of achieving pregnancy in a given menstrual cycle 't' is modelled by the (discrete time hazard)

$$\lambda_i(t) = P(T_i = t \mid T_i \geq t) = h(x_{it}\beta) \tag{1}$$

where h is a link function and x_{it} are possible time-dependent covariates for subject i that may include information about the initiation time. With the complementary log(-log) (CLL) link we get

$$\lambda_i(t) = 1 - \exp(-\exp(x'_{it}\beta)). \tag{2}$$

When the covariates do not depend on time, the time index is omitted, that is $x_i = x_{it}$. A particular model is given by $\beta = (\gamma_1, \ldots, \gamma_m, \alpha_1, \ldots, \alpha_m)$ where m, $q > 0$, $x'_{it}\beta = \gamma_t + x'_i\alpha$.

The probability function and the survival probability are given by the following expressions

$$\begin{aligned} P(T_i = t) &= \lambda_i(t) \prod_{j=1}^{t-1} (1-\lambda_i(j)) \\ &= \lambda_i(t) \exp(-\sum_{j=1}^{t-1} \exp(x'_{ij}\beta)) \\ &= \exp(-F_i(t-1)) - \exp(-F_i(t)). \end{aligned} \tag{3}$$

where

$$F_i(t) = \sum_{j=1}^{t} \exp(x'_{ij}\beta),$$

with the definition $F_i(0) = 0$, and it follows that $P(T_i \geq t) = \exp(-F_i(t-1))$.

For these formula it is relatively easy to compute maximum likelihood estimates using iteratively reweighted least squares. Standard software can be used to do this.

Retrospectively ascertained TTP

The retrospectively sampled TTPs are not observed from the distribution given by λ_i but are observed from this hazard and, subject to the observation being in some interval [0,S] are thus given by

$$\begin{aligned}\lambda_i^O(t) &= P_{x_i}^O(T_i = t \mid T_i \geq t) \\ &= P_{x_i}(T_i = t \mid T_i \geq t,\ T_i \leq S_i),\end{aligned}$$

that can be expressed in terms of the original prospective model presented in (1) above. Given this representation in terms of the regression effects of the underlying prospective model, a maximum likelihood analysis can be carried out and will give regression effects that are corrected for possible truncation biases.

The above model for the hazard of achieving pregnancy in each cycle leads to the observed probabilities

$$P(T_i = t_1) = \lambda_i^O(t_1) \prod_{j=1}^{t_I - 1} (1 - \lambda_i^O(j)).$$

Using the specific structure for the underlying prospective model with the CLL link we get

$$\begin{aligned}P_{x_i}(T_i = t \mid T_i \geq t,\ T_i \leq S_i) &= \frac{P_{x_i}(T_i = t)}{P_{x_i}(T_i \in [t,\ S_i])} \\ &= \frac{\exp(-F_i(t-1)) - \exp(-F_i(t))}{\exp(-F_i(t-1)) - \exp(-F_i(S_i))}.\end{aligned}$$

We can maximise the likelihood by iteratively reweighted least squares, but there are no standard programs that can do this.

Obviously, when evaluating a time trend, it is crucial to deal with the sampling issues. On the other hand, one pragmatic and simple approach may be to simply limit the data to initiation times within the sampling window and within, say, five years before the end of the study. This should remove any major effects of sampling problems and, if it is possible to disregard a five-year period then it is very simple to do so, and the study can then be thought of as an historically prospective study containing only information on subjects that achieve pregnancy.

If covariate effects such as smoking are evaluated based on birth registries, it is obvious that sampling issues may also lead to biases on the effect of smoking due to the fact that smoking habits are changing over time. If smoking is increasing over time then the one may erroneously conclude that smoking leads to shorter TTPs because the short TTPs are oversampled towards the end of the time period.

Conclusion

The study of time to pregnancy is an important component of reproductive epidemiology. Here, I have presented two main designs with the aim of showing how retrospectively collected data can be analysed. What still remains is to deal with additional sources of bias such as those mentioned in the introduction.

References

Baird D.D., Wilcox A.J., Weinberg C. "Use of time to pregnancy to study environmental exposures". *American Journal of Epidemiology 1986*; 124: 470–479.

Kalbfleisch J.D., Lawless J.F. "Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS". *Journal of the American Statistical Association 1989*; 84: 360–372.

Kalbfleisch J.D., Lawless J.F. "Regression models for right truncated data with applications to AIDS incubation times and reporting lags". *Statistica Sinica 1991*; 1: 19–32.

Lagakos S.W., Barraj L.M, De Gruttola V. "Nonparametric analysis of truncated survival data, with application to AIDS". *Biometrika 1988*; 75: 515–523.

Scheike T.H., Jensen T.K. "A discrete survival model with random effects: An application to time to pregnancy". *Biometrics 1997*; 53: 318–329.

Scheike T.H., Petersen J.H., Martinussen T. "Retrospective ascertainment of recurrent events: An application to time to pregnancy". *Journal of the American Statistical Association 1999*; 94: 713–725.

Weinberg C.S., Baird D.D., Wilcox A.J. "Sources of bias in studies of time to pregnancy". *Statistics in Medicine 1994*; 13: 671–681.

Weinberg C.S., Gladen B.C. "The beta-geometric distribution applied to comparative fecundability studies". *Biometrics 1986*; 42: 547–560.

Weinberg CS, Dunson D.B. "Some issues in assessing human fertility". *Journal of the American Statistical Association 2000*; 95: 300–303.