Article

Scand J Work Environ Health 1999;25(1):10-11

Analysis of time-to-pregnancy data

by Keiding N

The following article refers to this text: 2002;28(6):418-428

Key terms: cross-sectional sampling; discrete-time survival analysis; frailty; prospective sampling; retrospective sampling; survival analysis; truncation

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/10235398

Scand J Work Environ Health 1999;25 suppl 1:10—11

Analysis of time-to-pregnancy data

by Niels Keiding, MSc[1]

Keiding N. Analysis of time-to-pregnancy data. *Scand J Work Environ Health* 1999;25 suppl 1:10—11.

Statistical models for time-to-pregnancy data are usually formulated in discrete time. Two approaches were surveyed, including methods of accounting for both known heterogeneity (covariates, possibly time-dependent) and unknown heterogeneity (frailty). Delicate censoring and truncation patterns arose for prospective and, particularly, retrospective sampling designs. The inclusion of several pregnancies per couple presents new challenges and possibilities.

Key terms cross-sectional sampling, discrete-time survival analysis, frailty, prospective sampling, retrospective sampling, truncation in survival analysis.

Models for time-to-pregnancy data

Statistical models for time-to-pregnancy data belong to the general area of survival analysis. Often the focus is on menstrual cycle as the time unit, which makes time intrinsically discrete and has motivated the consistent use of discrete-time survival models. However, the actual measurement is often not made in cycles but rather in months, either as a surrogate or as itself of interest. (And then there would seem to be no harm in using continuous time-survival models.)

The statistical models need to be able to accommodate known heterogeneity between couples in the form of covariates, and one will often also want to incorporate residual random heterogeneity.

Let t=1,2, . . . be the number of menstrual cycles since "initiation" (ie, since attempts at getting pregnant started) and let x_t be a vector of covariates at time t. The task is to model the discrete hazard rate $\lambda(t|x_t)=P(T=t|T\geq t,x_t)$, which is the probability of becoming pregnant at cycle t, given that it did not happen before t and given the covariates.

An early influential model by Weinberg & Gladen (1) assumed the following model:

$$\log[\lambda(t|x_t)]=x'_t\beta,$$

where log is the natural logarithm and $x'_t\beta = x_{1t}\beta_1 +\ldots+ x_{kt}\beta_k$.

The model has the disadvantage that, when β varies across $(-\infty, \infty)$, $\lambda(t|x_t)$ is not restricted to the range [0,1] of a probability. This problem was avoided by Scheike & Jensen (2) using the following model:

$$\log\{-\log[1-\lambda(t|x_t)]\} = x'_t\beta,$$

in line with common practice in current discrete-time survival models. [See, for example, Fahrmeir & Tutz (3).] This model can be interpreted as a grouped-time version of the Cox regression model.

Weinberg & Gladen incorporated *unobserved heterogeneity* into a model with no covariates by assuming that the hazard for each given couple was constant over time ($\lambda(t)=r$) and that r follows a beta distribution across the population. The resulting marginal hazard in the population is as follows:

$$\lambda(t)=1/[\alpha+\mu(t-1)],$$

with parameter α and μ given by the beta distribution. Weinberg & Gladen extended this model to accommodate covariates in the following model:

$$\lambda(t)=1/[\alpha+\mu(t-1)]+x'_t\beta;$$

although the interpretation of a mixture across the population is then lost, indeed the parameter β has no interpretation at the individual level, only marginally for the population.

In contrast, Scheike & Jensen (2) extended their model to incorporate a random effect R_i for couple i according to the following assumption:

$$log\{-log[\lambda(t|R_i, x_{it})]\}=R_i+x'_{it}\beta.$$

1 Department of Biostatistics, Institute of Public Health, University of Copenhagen, Copenhagen, Denmark.

Reprint requests to: Professor Niels Keiding, Department of Biostatistics, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark. [E-mail: N.Keiding@biostat.ku.dk]

In this model the individual interpretation of β is conserved. Scheike et al (4) formulated generalizations of this model to allow several times to pregnancy per couple, and they discussed these generalizations in the context of current literature on frailty models and multivariate survival analysis.

Prospective sampling

As is so often the case, inference is easiest in prospective sampling in that it leads to standard right-censored survival data, in which the couples who have not conceived at the end of follow-up are counted as *right-censored.* Note that couples recruited into a prospective study at known time t after initiation will have to be counted with *delayed entry* (left truncation) at t. In practice, prospective studies are not very common, usually rather small, and often marred by considerable self-selection difficulties. A particular difficulty with the assessments of the effect of calendar time is whether to score it at initiation, at conception (which creates difficulties at least for the censored couples), or as current calendar time along the way.

Retrospective sampling

Large time-to-pregnancy surveys are often retrospective, data being gathered from pregnant women. These surveys have obvious weaknesses, primarily the biased sampling based on fecundity, particularly the nonpresence of sterile or nonfecund couples.

However even beyond these unavoidable difficulties, the correct analysis of retrospective time-to-pregnancy data is more intricate than often realized, particularly when due account is taken of the observed dramatic secular trends in birth rates, which must be generated by similar trends in initiation intensity. As an example, in a common design the data are gathered from interviews in a fixed time window. It is then clear that, if calendar time is related to initiation, long times to pregnancy will be overrepresented in the early phase and short times to pregnancy in the late phase, with intricate patterns of left and right truncations (2). These phenomena were defined away by a tacit (hardly tenable) assumption of stationarity in the classical work of Weinberg & Gladen (1), as made explicit on page 679 of Weinberg et al (5). Incorporating several time-to-pregnancies per couple in a retrospective design is possible through careful likelihood constructions. [See Scheike et al (4) for details.] Basso et al (6) discussed recall bias in retrospective fertility studies.

Cross-sectional sampling

A simple procedure would be to ask a cross-sectional sample of a population (or subpopulation) of women whether they are currently attempting to get pregnant and, if so, for how long they have attempted to do so. With this design it would seem reasonably realistic to minimize selection bias. There is no a priori exclusion of sterile couples as in retrospective sampling, and self-selection is minimal, in contrast to most prospective studies. Although the design was (briefly) mentioned at least by Weinberg & Gladen (1), there seems to be little or no discussion of the relevant statistical tools.

In the simplest case the observed distribution of "current waiting times" could be modeled as the backward recurrence time in a stationary renewal process, from which Denby & Vardi (7) described a nonparametric estimator of the waiting-time distribution, and generalizations from there would follow lines similar to those for prevalent cohort and current status data (8—10).

References

1. Weinberg CS, Gladen BC. The beta-geometric distribution applied to comparative fecundability studies. Biometrics 1986;42:547—60.
2. Scheike TH, Jensen TK. A discrete survival model with random effects: an application to time to pregnancy. Biometrics 1997;53:349—60.
3. Fahrmeir L, Tutz G. Multivariate statistical modelling based on generalised linear models. New York (NY): Springer-Verlag, 1994.
4. Scheike TH, Martinussen T, Petersen JH. Retrospective ascertainment of recurrent events: an application to time to pregnancy. Copenhagen: Department of Biostatistics, University of Copenhagen, 1997. Research report 97/14.
5. Weinberg CS, Baird DD, Wilcox AJ. Sources of bias in studies of time to pregnancy. Stat Med 1994;13:671—81.
6. Basso O, Olsen J, Bisanti L, Karmaus W, and the European Study Group on Infertility and Subfecundity. The performance of several indicators in detecting recall bias. Epidemiology 1997;8:269—74.
7. Denby L, Vardi Y. The survival curve with decreasing density. Technometrics 1986;28:359—67.
8. Keiding N. Age-specific incidence and prevalence: a statistical perspective (with discussion). JR Stat Soc A 1991;154:371—412.
9. Diamond ID, McDonald JW. Analysis of current-status data. In: Trussell J, Hankinson R, Tilton J, editors. Demographic applications of event history analysis. Oxford: Clarendon Press, 1992:231—52.
10. Keiding N, Begtrup K, Scheike TH, Hasibeder G. Estimation from current-status data in continuous time. Lifetime Data Anal 1996;2:119—29.