ELSEVIER
Preventive Veterinary Medicine 26 (1996) 263-274
PREVENTIVE VETERINARY MEDICINE

The application of event-time regression techniques to the study of dairy cow interval-to-conception

J.L. Harman [a,*], G. Casella [b], Y.T. Gröhn [a]

[a] *Section of Epidemiology, Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA*
[b] *Biometrics Unit, Cornell University, Ithaca, NY 14853 , USA*

Accepted 14 August 1995

Abstract

This paper explains the use of event-time techniques, also known as survival analysis, to assess a relationship involving dairy cows: the effect of early lactation milk yield on the interval from calving to subsequent conception. Possible confounding factors, such as disease history, parity, calving season, and herd production level, were controlled by their inclusion in the model equations.

Both Cox proportional hazards regression models and Weibull regression models were employed, with similar results, indicating that a Weibull distribution may be a good approximation of the distribution of conception times in dairy cattle.

Keywords: Milk production; Event-time analysis; Survival analysis; Dairy cows; Reproduction

1. Introduction

Event-time analysis allows the inclusion of data which lack an observed endpoint for an interval-type outcome variable. This outcome variable is often the measure of time elapsed from some starting point until an awaited event occurs. The length of this interval may not be known because, prior to the event-time, competing events may intervene and preclude further observation, as occurs, for example, when animals are culled or sold, or a study is terminated. However, partial information is often available, i.e. one might know the last day when it was certain that the event had not yet occurred; event-time analysis allows the use of this partial information, so that some of the

* Corresponding author. Tel.: (301) 436-7050; fax: (301) 436-4233.

Elsevier Science B.V.
SSDI 0167-5877(95)00553-6

individuals which are usually lost to studies of this sort may influence the answer to the question being posed.

Many of the analytical techniques of event-time analysis originated from research in industrial-life testing; this field advanced rapidly during the post-World War II era due to growing interest in ways to evaluate the reliability and estimate the lifetime of manufactured product components. A frequently encountered synonym, failure-time analysis, remains as a legacy of methodological origins in components testing. Soon, however, the technique engaged medical statisticians as a natural way to study chronic disease survival rates and the outcome of clinical trials (Gehan and Schneiderman, 1990). Employed in this way, it is called a survival analysis, the event in question being the death of the patient. These methods may, as easily, be applied to situations where the outcome event is favorable, as, in the current example, postparturient conception by dairy cows.

This paper discusses how event-time analysis was used to examine factors influencing conception time in 44450 Finnish Ayrshire cows that calved during a single 12-month period in 1986–1987. The event of interest was conception, the studied interval was days past parturition until subsequent conception, examined as a function of early-lactation milk production. The subjects were cows that had received at least one insemination after parturition, and for which early-lactation milk and milkfat records were available. Each record included monthly milk tests; bimonthly milk fat tests; calving, culling, insemination, and disease treatment dates; reason for culling; and diagnosis codes for 43 different diseases. The origins and quality of these data are described in detail in Gröhn et al. (1986).

1.1. Describing the event-time distribution

When T, time to event, is characterized as a continuous random variable, its instantaneous probability distribution, $f(t)$, the probability that the event occurs at time t, is

$$f(t) = \lim_{\Delta t \to 0^+} \frac{P(t \le T < t + \Delta t)}{\Delta t}$$

The distribution of T may also be described by its cumulative distribution function, $F(t) = P(T \le t)$, the probability that the event occurs at or before time t. Equivalently, an event time distribution is often referred to by its survival function, $S(t) = (1 - F(t))$, the probability that event will not have occurred by time t.

Very important to note, however, is a third way that is commonly used to represent the distribution of an event-time random variable: that is, by the hazard function, $h(t)$. The hazard function is nothing more than the ratio of the value of $f(t)$ to $S(t)$ at any time t and, as such, a hazard function is not something unique to event-time data. But the hazard function has an intuitive meaning in event-time analysis: it represents the probability of failure at time t, conditional upon survival (i.e. no event) until time t and, as such, represents the way hazard is generally conceptualized in everyday situations. As shall be shown, the hazard function is also a convenient construct from which to analyze event-time data.

Event-time analysis usually has as its goal the resolution of one of these two questions. (1) How do the event-time functions of two or more groups of subjects compare with one another? (2) How does time-until-event depend upon multiple covariates that describe population heterogeneity or treatment differences? The first question is often addressed by forming Kaplan–Meier estimates of the survival function (Kaplan and Meier, 1958; see Thysen, 1988 for other veterinary applications). In the current study, techniques of the first type were exploited for univariable screening of possible covariates to include in subsequent multivariable regression models. However, our analysis focused on the second issue, as we attempted to describe how interval-to-conception depends upon milk production, while controlling for other variables.

1.2. Censoring

A major motivation behind the use of event-time analysis is the need to incorporate information from observations that are right-censored. Using the convention that the time axis proceeds from left to right, left-censored subjects are those for whom the event in question occurred before the beginning of the observation period. Subjects have been right-censored when observation ceases before the event occurs; the time of last observation is then recorded as the subject's censoring time. Ignoring information from these subjects usually means losing important clues to reasons for long survivability (or, in the present case, long interval-to-conception). For example, end-of-study censoring occurs for those cows that have not yet conceived by the last day of the study.

Event-time analysis in veterinary field studies almost always involves variable censoring times, as animals are culled, sold, die, or do not reach the desired event (e.g. conception) before the end of the study. In the current study, uncensored cows were those for which a calving was recorded immediately subsequent to the study lactation; for these cows, insemination dates and dates of next calving were used to determine the date of conception. For cows with no subsequent calving recorded, the record was considered to have been censored at the last date for which any herd record existed for that animal, e.g. a milk recording date or a culling date.

It is, of course, inevitable that some pregnant cows will have been culled. This misclassification of a few pregnant culled cows as non-pregnant culled cows would have affected the results of the analysis in the following way: characteristics that were more likely to have been expressed by culled cows, compared with non-culled cows, would have shown slightly more association with longer interval-to-conception than was the true situation; likewise, characteristics of non-culled cows would have appeared to be more strongly associated with short intervals-to-conception than was the true situation. We felt that this bias would be of much smaller magnitude than the opposite bias that would have resulted from using last insemination date as the censoring date for culled cows (when, in reality, most of them remained non-pregnant beyond that date). Also, this choice of censoring assignments gave our analysis a conservative aspect when judging whether or not high milk production (which is generally associated with low probability of culling) would be detrimentally associated with longer intervals-to-conception.

2. Techniques

2.1. Nonparametric screening tests

The log-ranks test and the Wilcoxon test for event-time data are two nonparametric tests that compare the survival of groups of individuals, as defined by different values of a single variable (Thysen, 1988; SAS Institute Inc., 1988). These relatively simple procedures can be used as univariable analysis methods to test whether or not time-until-outcome is related to a single independent variable. It is prudent to use both tests, as the log ranks test places more emphasis on larger intervals-to-outcome, and the Wilcoxon test gives more weight to smaller intervals (SAS Institute Inc., 1988).

Based on these preliminary analyses, we offered all disease occurrence variables to the subsequent regression models, as well as variables representing three categories of calving season, three categories of parity for the multiparous cow model, and five categories of 60-day milk and milkfat production.

2.2. Semi-parametric modeling: the Cox proportional hazards model

In a biomedical setting, the distribution of event times, T (the random variable finding its expression in the n sample failure times, t_1, $t_2, \dots t_n$) can be only vaguely surmised and rarely well-tested. The requirement of having to specify the nature of the underlying event-time distribution in a fully parametric model, therefore, makes the use of parametric models problematic in these situations. A more generally applicable model was developed by Sir David Cox (1975); Cox's contribution was a major step leading to widespread use of event time analysis in disease studies. The Cox model makes the assumption that covariates act multiplicatively on the hazard function in such a way that the relative effect on the hazard of event occurrence can be compared between subjects without the need to specify an underlying distribution function for T. In other words, the hazard of one subject always remains at some constant proportion of the hazard of another subject.

Here, we imagine that there is some baseline hazard function, $h_0(t)$, describing the hazard for a hypothetical animal for which all covariate values equal zero. Then

$$h(t;z) = \psi(z)h_0(t)$$

where $\psi(t)$ is some arbitrary function of t (although, as shall be shown, some functions are better than others), and z is the vector of covariate values for that animal. When comparing the hazard of two animals at the same time, t, the baseline hazard, $h_0(t)$, cancels out of the expression, and so is superfluous. However, this same superfluity is both the strength and the weakness of the Cox model: this model can perform quite well without a specified distribution function, but it can only be used to compare the hazard between subjects, and cannot be used to predict the absolute value of the hazard for any particular subject.

Although it cannot estimate absolute measures of risk, the Cox model has several virtues. Censoring is not difficult to accommodate within a model of this type. Also, the dispensability of prior knowledge of the underlying distribution simplifies inference

about regression coefficients (Cox and Oakes, 1984). This type of model accommodates the many instances when easily explained biological or physical phenomena support a proportionality of hazard over time. For example, it is reasonable to surmise that the age of a dairy cow influences the time from parturition to subsequent conception, such that an older cow may have a conception hazard that is some percentage of the conception hazard of a younger cow, and that this ratio of hazard remains constant over the range of time being considered.

The function, $\psi(z)$, may be parameterized in different ways, but the log-linear form, $\psi(z) = e^{\beta z}$, has become the form one usually encounters; in fact, most texts refer solely to this parameterization when discussing Cox proportional hazard models. One of the attributes of this log-linear form is that it ensures that the hazard always remains positive. Also, inferences concerning the vector, β, of regression coefficients are readily made when using this form, owing to the relative simplicity of the first and second derivatives of the log-likelihood functions; solving the parameter estimates requires maximization of the log-likelihood function by examining these two derivatives.

2.3. *Testing the proportional hazards assumption*

Before invoking a proportional hazards model, one must test whether the proportional hazards assumption applies to the variables under consideration. A valid way to test the proportional hazards assumption for a covariate is by plots of the estimated survival functions for different strata of that variable, explained in detail in Cox and Oakes (1984). Briefly, in the usual log-linear parameterization of the Cox model

$$S(t;z) = S_0(t)^{\exp(z\beta)}$$

in which $S(t)$ represents the probability that the event in question has not occurred by time t. The baseline survival function, $S_0(t)$, is obtained in a manner somewhat analogous to that used in the Kaplan–Meier estimate (Kaplan and Meier, 1958; Thysen, 1988). Strata are constructed for the variable in question so that within a stratum, the effect of change in the value of the covariate is of no practical importance. Plots of $\log(-\log S(t))$ vs. t are then constructed for each stratum of subjects. It can be shown algebraically that, if the proportional hazards assumption holds for that particular covariate, these stratum-by-stratum plots should all lie approximately parallel to one another when plotted onto the same grid. Therefore, the plotted lines representing truly different strata of the same variable should not cross one another nor diverge too strongly within the time period being tested. Sequential estimations of S(t) can be calculated so as to account for other variables that have already been shown to comply with the proportional hazards assumption.

Fig. 1 shows an example of this type of plot, as generated from the data for the dairy cow interval-to-conception problem. For testing the proportional hazards assumption for production variables, the combined 60-day milk and milkfat production was described by a five-category variable assigned to each multiparous cow (Harman, 1994). Fig. 1 shows that the plots are approximately parallel for the five categories of this variable, Cow 5 classification representing the highest 60-day production level, and Cow 1 representing the lowest production level. The crossing-over of levels Cow 3 and Cow 5

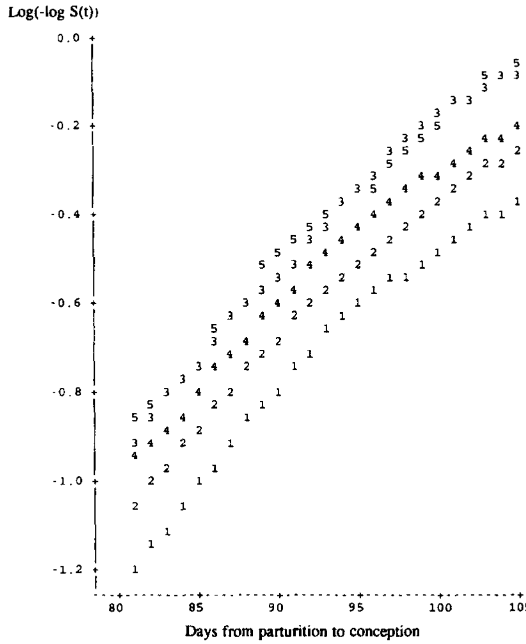

Fig. 1. Log(− log $\hat{S}$(time)) vs. time by production category for nondiseased Finnish Ayrshire multiparous cows to test the assumption of proportionality of conception hazard across production strata; data from 1986–1987. Time is number of days after parturition and $\hat{S}$(time) is the estimated survival function plotted at each observed conception time ($n = 11000$).

categories foreshadows the finding in the multiple regression model of no significant difference in chance of conception between cows of these two production categories. Similar plots were constructed for parity and diseases.

2.4. Generating the Cox proportional hazards models

Considering only those individuals that had not already conceived at the start of the study interval, the conception hazard was then modeled according to the Cox proportional hazards regression for the interval 56–120 days postpartum. A disease was coded as present only if it occurred within the period beginning 30 days before parturition and ending 120 days after parturition.

Two different proportional hazards models were generated: one for heifers and one for cows. For each of these two groups, the all-subsets proportional hazards regression procedure of SAS Proc PHREG (SAS Institute Inc., 1992) was used to select the models with the highest score χ^2 statistic for a given number of variables. Although the all-subsets procedure was allowed to include or exclude disease variables, all variables describing milk and milkfat production, calving season, and parity were forced into each regression, in order to require each model to account for any variance contributed by these factors.

From the results of the all-subsets regressions, the best model was considered to be the one for which an additional variable would cause a non-significant ($\alpha = 0.05$) improvement in the (-2 log likelihood) statistic. Two best models were selected by this method, one each for heifers and cows. These two best models were then re-analyzed, stratified by community. The martingale residuals were plotted and the plots were visually examined for outlying data points of potentially high influence. Martingales residuals are, like all residuals, a measure of observed vs. expected data values, and can be used as a tool evaluate event-time regression models. The computation of a martingale residual is discussed in more detail in Therneau et al. (1990). Ten to twenty observations with residuals that indicated high influence were removed from each of the models; examination of these records showed them each to be unusual animals regarding either their disease or reproductive record, or both, relative to the other tens of thousands of animals in the models. Omitting these few observations from such a large study population did not have a noticeable effect on parameter estimates; owing to the unusual circumstances of these few cows, the records of these animals were omitted from both the final Cox and Weibull models.

Here is a brief example of the interpretation of these Cox regression model results. Each cow was assigned by its 60-day milk and milkfat production into one of five categories, increasing in production level from the lowest, Cow 1 category, to the highest, Cow 5 category. The main group of mid-producing cows, those in the Cow 3 production group, were used as the comparison group in the regression model, presented in Table 1. We will compare the results for two hypothetical cows that had not conceived as of day 56: a cow in the Cow 4 production group, relative to a cow in the Cow 3 production group. The corresponding coefficient, -0.08, reported in Table 1 indicated that, at any particular time between 56–120 days postpartum, a cow in the Cow 4 production group was 0.92 times as likely to conceive as a cow in the Cow 3 category ($e^{-0.08} = 0.92$), assuming that all other covariates were equal. For a Cow 4 subject, the cumulative probability of remaining nonpregnant at any particular day in this range would be the 0.92th power of the cumulative probability of a Cow 3 animal remaining nonpregnant. Because the cumulative probability is less than one, raising this probability figure to the 0.92th power will increase its value; thus the Cow 4 animal is more likely to remain nonpregnant than the Cow 3 animal at any time, t, within 56–120 days postpartum. In interpreting these results, it must be remembered that the proportional hazards model can only provide comparisons between subjects, and cannot furnish absolute probabilities of the occurrence of an event.

2.5. *Sensitivity analysis*

Construction of a likelihood function for the model requires that some assumptions are made about the censoring mechanism operating on the population. Random censoring allows the generation of the type of likelihood function usually described. The concept of independent censoring, while being less restrictive in its definition than completely random censoring, still allows for the validity of the likelihood function used with random censoring. Independent censoring requires that, for any given time t and covariate vector z_i, a subject may not be censored because of appearing at particularly

Table 1
Parameter estimates of proportional hazards model vs. a Weibull model for conception hazard in multiparous cows, 56–120 days after parturition, controlling for parity, calving season, and significant diseases that were diagnosed −30 through 120 days after parturition ($n = 30026$)

	Proportional hazards model	Weibull model
Mar. 16–May 1	−0.03	−0.03
May 16–Sept. 15	−0.03	−0.03
Parity 2	−0.03	−0.03
Parity 3–4	−0.03	−0.03
Herd 1 [a]	−0.034	−0.034
Herd 2	−0.049	−0.050
Herd 3	−0.0034	−0.0040
Cow 1 [b]	−0.22	−0.23
Cow 2	−0.13	−0.13
Cow 4	−0.080	−0.082
Cow 5	−0.060	−0.059
Anestrus	−0.53	−0.54
Ovulatory dysfunction	−0.50	−0.50
Other infertility	−0.68	−0.69
Late metritis	−0.96	−0.98
Ketosis	−0.12	−0.13
Acute mastitis	−0.12	−0.12
Teat injury	−0.18	−0.18
Uterine prolapse	−0.75	−0.77
Other disease	−0.24	−0.24
Laminitis	−0.38	−0.39

[a] Herd 1–Herd 4 designate four categories of herd production level, based on 305-day mature equivalent milk production, from lowest to highest quartile of herds.
[b] Cow 1–Cow 5 designate five categories of individual cow 60-day milk and milk fat production, from lowest to highest ranking of cows.
Weibull parameter estimates were adjusted by factor of (−1/scale) (scale parameter estimate 0.76; intercept estimate 4.14).

high or low risk of event occurrence. Independent censoring allows that the failure mechanism may depend on the history of previous events or previous censoring times, on the values of covariates included in the model, or on random processes external to the event mechanism—as long as it does not depend on the future outcome of the event in question (Kalbfleisch and Prentice, 1980; Cox and Oakes, 1984). Thus, we need to make the assumption that for a culled cow with a particular set of covariate values, had she continued to be subjected to insemination, its chance of future conception would not be lower or higher than any other non-culled cow with the same covariates. In order to invoke this assumption with some confidence, we must include in our model all covariates that may influence censoring (usually due to culling, in our particular example).

It is not possible to test, nonparametrically, for independence between censoring and event mechanisms (Cox, 1962; Tsai, 1990). This judgment must be made as carefully as possible, given knowledge of the situations being modeled. It is possible, however, to find the extreme bounds for model parameters in the case that censoring and event

mechanisms are totally dependent on one another. The extreme cases modeled by such a sensitivity analysis would be: first, the scenario that censoring (culling) implies imminent occurrence of the event (conception), and second, the opposite extreme, that censoring (culling) indicates a future of long survival (non-occurrence of conception). These two extreme scenarios were simulated by manipulating the data employed in the standard model regressions. The first simulation re-coded data to reflect a presumption that all cows censored between 56–120 days conceived immediately upon censoring; the second simulation re-coded these data as though conception actually occurred later than day 120, i.e. as though these cows were censored by end-of-study censoring at day 120. Cows or heifers that were censored because conception occurred sometime after day 120 were not re-coded for these analyses.

Theoretically, the more independent conception and censoring really were in this population, the more closely the standard model estimates—those obtained when all censoring was coded as it actually occurred—approximate the true population parameter values. Those covariates that are, themselves, positively or negatively associated with culling will be expected to have their regression coefficients vary between simulations, the true value lying somewhere between, subject to the influences of other covariates. We found that all of our parameter estimates remained within two standard deviations of their value as estimated by the standard model. Most importantly, for our purposes, the nature of the relationship of early-lactation milk production to conception was not altered in these extreme simulation models, compared with the standard model.

2.6. A parametric failure time model—the Weibull distribution

Because the Cox semiparametric model does not require specification of a distribution for the event time, it is a useful tool. However, precisely because of the non-specification of this underlying distribution, the Cox model has a major limitation: this model can only provide comparisons between subjects, and cannot furnish absolute probabilities of the occurrence of an event. A reason for generating a fully parametric model, a model that requires the specification of the type of event time distribution, is that a parametric model can then be used to produce an expected value for the outcome variable when all covariate values are known.

The probability distribution function of event times may take many forms. One distribution that is mathematically tractable and is often appropriate for event time data is the Weibull distribution

$$f(t) = \lambda p(\lambda t)^{p-1}\exp\left[-(\lambda t)^{p}\right]$$

where $\lambda, p > 0$. The hazard function for the Weibull failure model is

$$h(t) = \lambda p(\lambda t)^{p-1}$$

The parameter λ acts as a scale parameter for this function; the shape is changed by changing the value of p (Fig. 2). The hazard is constant over time when $p = 1$ (letting $p = 1$ effectively reduces the Weibull distribution model to the exponential distribution model, a very simple event-time model with a constant hazard rate). A value of p

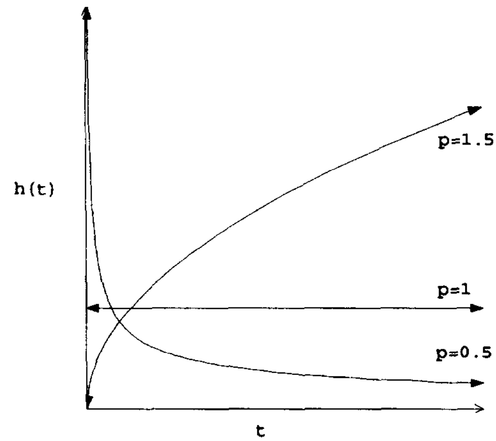

Fig. 2. Weibull hazard function, $h(t) = \lambda p(\lambda t)^{p-1}$.

greater than unity indicates that the hazard increases monotonically with time; the hazard decreases with time if p is less than one.

Because the fertility of cows usually increases steadily following parturition, we suspected that a Weibull model with p parameter greater than one may be a good approximation of the distribution of intervals-to-conception.

The Weibull event-time model will accommodate a vector, z, of covariate values such that when these covariate values are specified, the model will predict a value for the outcome variable. To incorporate this vector, assume that the hazard at any time t is the product of the baseline hazard (that is, the hazard when all covariate values are zero) multiplied by some linear function of the covariate vector. This function of the covariate vector, z, is usually chosen to be $e^{z\beta}$, because of its simplicity and consistent non-negativity. Now the hazard function is expressed as

$$h(t;z) = \lambda p(\lambda t)^{p-1} e^{(z\beta)}$$

and the probability density function conditional on the covariate vector, $f(t;z)$, becomes

$$f(t;z) = \lambda p(\lambda t)^{p-1} e^{z\beta} \exp\left[-(\lambda t)^{p} e^{z\beta}\right]$$

2.7. Comparison of the Weibull and Cox proportional hazards regression models

The regression analysis for cows during days 56–120 was re-run as a Weibull regression model using the same covariates selected by the best proportional hazards model. In addition, 79 covariates representing community were included in the model, and the population's baseline hazard function was specified as that of a Weibull distribution. SAS Proc LIFEREG (SAS Institute Inc., 1988) generated the Weibull model.

The parameter estimates obtained from a Weibull regression model included scale and intercept estimates that described the shape of the underlying Weibull distribution of

conception times, as well as a regression parameter estimate for each independent variable in the model. To compare covariate parameter estimates between Weibull and Cox proportional hazards models, one must first multiply each Weibull estimate by (-1/scale estimate). Table 1 shows the Weibull scale and intercept estimates and comparisons of regression parameter point estimates between the two models. The remarkable similarity in regressions parameter estimates between these models indicated that the Weibull function was indeed a good approximation of the baseline distribution of intervals-to-conception for this population.

Because the parameters of the underlying distribution are estimated by this type of regression, the Weibull model can be used to produce an expected value for the outcome variable when all covariate values are specified. This analysis showed that the Weibull event-time regression model is a good tool for describing the relationship between production, health, and reproductive delay for these dairy cows; the less predictive Cox semiparametric models may be unnecessary for this type of population. Subsequent economic studies should be able to exploit the Weibull parametric model directly in order to better assess the costs of disease and of extremely high production.

Acknowledgements

This work was partly supported by the American Veterinary Medical Association Foundation Research Grant Program, and the US Department of Agriculture Animal Health and Disease Program.

Computations were performed on the Cornell Production Supercomputer Facility, which is supported in part by the National Science Foundation, New York State, and the IBM Corporation.

References

Cox, D.R., 1962. Renewal Theory. Methuen, London, 142 pp.

Cox, D.R., 1975. Partial likelihood. Biometrika, 62: 269–276.

Cox, D.R. and Oakes, D., 1984. Analysis of Survival Data. Chapman and Hall, London, 201 pp.

Gehan, E.A. and Schneiderman, M.A., 1990. Historical and methodological developments in clinical trials at the National Cancer Institute, Bethesda, Maryland USA. Stat. Med., 9 (8): 871–880.

Gröhn, Y.T., Saloniemi, H. and Syväjärvi, J., 1986. An epidemiological and genetic study on registered diseases in Finnish Ayrshire cattle. 1. The data, disease occurrence and culling. Acta Vet. Scand., 27: 182–195.

Harman, J.L., 1994. The effect of early-lactation milk yield on dairy cow interval-to-conception: a survival analysis. Ph.D. Thesis, Cornell University, 164 pp.

Kalbfleisch, J.D. and Prentice, R.L., 1980. The Statistical Analysis of Failure Time Data. John Wiley, New York, 321 pp.

Kaplan, E.L. and Meier, P., 1958. Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc., 53: 457–481.

SAS Institute Inc., 1988. SAS/STAT User's Guide, Release 6.03 edn. Statistical Analysis Systems Institute Inc., Cary, NC.

SAS Institute Inc., 1992. SAS Technical Report P-229, SAS/STAT Software: Changes and Enhancements, Release 6.07. Statistical Analysis Systems Institute Inc., Cary, NC.

Therneau, T.M., Grambsch, P.M. and Fleming, T.R., 1990. Martingale-based residuals and survival models. Biometrika, 77: 147–160.

Thysen, I., 1988. Application of event time analysis to replacement, health and reproduction data in dairy cattle research. Prev. Vet. Med., 5: 239–250.

Tsai, W.-Y., 1990. Testing the assumption of independence of truncation and failure time. Biometrika, 77 (1): 169–177.