Odd Aalen Niels Keiding
Jens Thormann

Interaction Between Life History Events

Preprint
October
1977

8

Institute of Mathematical Statistics
University of Copenhagen

Odd. O. Aalen*, Niels Keiding

and Jens Thormann**

INTERACTION BETWEEN LIFE HISTORY EVENTS

Preprint 1977 No. 8

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

October 1977

Summary

A simple time-continuous Markov chain model is proposed to analyse the interaction between two life history phenomena, where occurrence of one event may change the intensity at which the other occurs. Parametric (assuming time-homogeneity) as well as nonparametric (without this assumption) statistical methods are proposed; these are special examples of recent theory for the statistical analysis of counting processes.

The model is applied to a study concerning possible influence of menopausal hormonal changes on the intensity of outbreak of the chronical skin disease pustulosis palmo-plantaris.

1. Introduction

This note is concerned with the detection of possible interaction between two life history phenomena, where occurrence of one event may change the intensity at which the other occurs. In practice such an interaction is often one-sided, in that one event is hypothesized as a "cause" for changing the occurrence intensity of the order event but has its own intensity left unchanged by the latter phenomenon.

Denote the two phenomena by A and B. A simple time-continuous Markov chain model is the following: Let there be four states denoted 0,A,B,AB and let the transitions between the states correspond to the occurrence of A and B in the manner indicated by Figure 1. If, for instance, A occurs before B then it means that the path 0→A→AB is followed. The Markov chain is given by assuming that the value at time 0 is 0 and by the time-dependent intensities given in Figure 1. Denote by I the set of states {0,A,B,AB}.

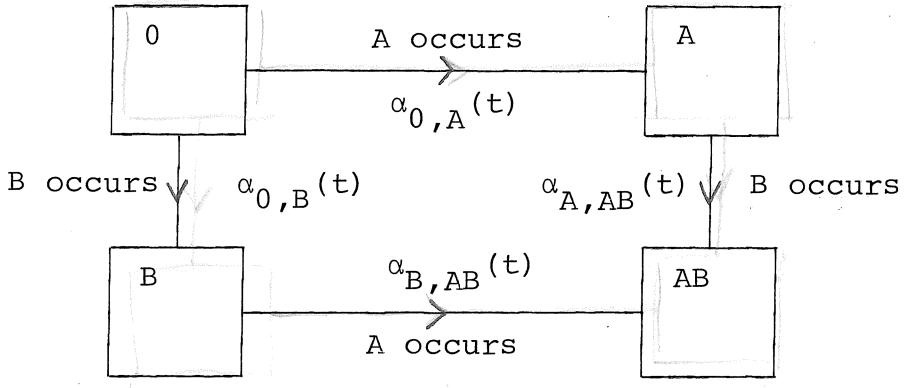


Figure 1. A simple Markov chain model for the occurrence of two life history events.

The model was suggested by Colding-Jørgensen and Simonsen [1940] in a statistical appendix (in Danish) to a study by J.M. Wollesen concerning the possible effect of pregnancy on the formation of gall-stones. Properties of such Markov chains where given by Freund [1961] who also derived maximum likelihood estimates in the case of constant intensities, and by Schweder [1970] who outlined some specific applications, but developed no statistical analysis. A related recent approach to survival data with auxiliary variables was given by Lagakos [1976].

Notice that the model generalizes the simple competing risks situation by also considering the risk of one event <u>after</u> the other has happened.

The question of interaction between the two events A and B corresponds to the statistical problem of estimating the $\alpha_{ij}(t)$ and of testing whether $\alpha_{0,A}(t)=\alpha_{B,AB}(t)$ for all t and $\alpha_{0,B}(t)=\alpha_{A,AB}(t)$ for all t.

In Section 2 we introduce a counting process formulation which is particularly useful when censoring is present. In

Section 3 we give estimators and tests first for a time-homogeneous model and then for a time-inhomogeneous one. We present in Section 4 an extended model and use this to analyse the data that prompted this investigation. These concern the age of appearance of pustulosis palmo-plantaris (a chronical skin disease) and the question is whether the hormonal changes in connection with menopause or similar artificially induced changes in ovarian function might affect the chance of development of this disease in women.

We finally mention that ideas similar to ours are latent in the analyses of Stanford Heart Transplant data by Turnbull, Brown & Hu [1974] and Crowley & Hu [1977].

2. A counting process formulation.

Assume that a number of patients are under observation (we here use the terminology of clinical trials). Each patient may be observed over different periods of time and censoring may take place. Let $Y_i(t)$, $i \in I$, be the number of patients observed to be in state i just before time t (so that $Y_i(t)$ is left continuous) and let $N_{ij}(t)$, $i \neq j$, $i,j \in I$, be the observed number of transitions between the states i and j during the time interval $[0,t]$ (so that $N_{ij}(t)$ is right continuous). Then the vector $\underset{\sim}{N}=\{N_{ij}, i \neq j, i,j \in I\}$ is a multivariate counting process with N_{ij} having intensity process $\alpha_{ij}(t)Y_i(t)$. (We refer to Aalen [1975,1977] for a general framework for statistical analysis of such processes.) The functions α_{ij} are assumed to be left-continuous.

This counting process formulation is particularly useful when censoring occurs or when the number of patients considered

increases during the study. The $Y_{ij}(t)$ indicate the "exposure" set" or "number at risk" and may at any time t be modified on the basis of past experience or as a result of outside random influences, but the $Y_{ij}(t)$ are not allowed to depend on transitions that take place after time t. The inference procedures given below are valid for the general kinds of Y-processes mentioned here.

3. Statistical analysis.

We want to estimate, in some sense, the intensities and then to test the hypotheses $\alpha_{0,A}(t)=\alpha_{B,AB}(t)$ for all t and $\alpha_{0,B}(t)=\alpha_{A,AB}(t)$ for all t. We will consider two models, one with constant intensities and a nonparametric one. Denote the interval of observation by [0,1].

3a. Constant intensities.

Denote the constant intensities by α_{ij}, $i,j\in I$. By Aalen [1975], see also Aalen and Hoem [1978], the maximum likelihood estimator of α_{ij} is given by

$$\hat{\alpha}_{ij} = N_{ij}(1) / \int_0^1 Y_i(s)\,ds.$$

If the number of patients, n, in the study increases in such a way that $a_n^{-1}\int_0^1 Y_i(s)\,ds$, $i\in I$, converges to constants for some increasing sequence $a_n\to\infty$, then $\sqrt{a_n}(\hat{\alpha}_{ij}-\alpha_{ij})$, $i,j\in I$, are asymptotically independent and normally distributed with zero means. Estimators of the asymptotic variances are given by

$$a_n\hat{\alpha}_{ij} / \int_0^1 Y_i(s)\,ds \qquad i,j\in I.$$

The likelihood ratio test statistic for testing $\alpha_{0,A}=\alpha_{B,AB}$ is given by

$$2\left[N_{0,A}(1)\log\left(\frac{N_{0,A}(1)}{\int_0^1 Y_0(s)\,ds}\right)+N_{B,AB}(1)\log\left(\frac{N_{B,AB}(1)}{\int_0^1 Y_B(s)\,ds}\right)\right.$$

$$\left.-(N_{0,A}(1)+N_{B,AB}(1))\log\left(\frac{N_{0,A}(1)+N_{B,AB}(1)}{\int_0^1 Y_0(s)\,ds+\int_0^1 Y_B(s)\,ds}\right)\right]$$

which as $n\to\infty$ is asymptotically χ^2-distributed with 1.d.f. Of course, we get a similar formula for testing $\alpha_{0,B}=\alpha_{A,AB}$.

3b. Nonparametric analysis.

In this section no assumption will be made about the intensities except that they are integrable on [0,1]. As explained by Aalen [1975,1977] one may estimate the integrated intensities

$$A_{ij}(s,t) = \int_s^t \alpha_{ij}(u)\,du \qquad i,j\in I$$

by

$$\hat{A}_{ij}(s,t) = \int_s^t X_i(u)\,dN_{ij}(u) \qquad i,j\in I$$

where

$$X_i(t) = \begin{cases} Y_i^{-1}(t) & \text{when } Y_i(t)>0 \\ 0 & \text{otherwise.} \end{cases}$$

We remark that the Stieltjes integral defining $\hat{A}_{ij}(s,t)$ is just the sum

$$\sum_{k:s<T_k\le t} Y_i^{-1}(T_k)$$

where T_k is the time of the k'th transition from i to j. Similar interpretations apply in the sequel.

The estimators generalize the hazard plots defined for the life testing model by Nelson [1972] and Altshuler [1970]. An estimator of $\mathrm{Var}\hat{A}_{ij}(s,t)$ is given by

$$\int_s^t X_i^2(u)\,dN_{ij}(u) \qquad i,j\in I.$$

Properly normalized the $\hat{A}_{ij}$ will asymptotically be distributed as normal processes with independent increments.

We now turn to the problem of testing the hypotheses mentioned at the beginning of Section 3. To be specific, let us consider the hypothesis $\alpha_{0,A}(t)=\alpha_{B,AB}(t)$ for all t. A generalized Savage statistic was given by Aalen [1975,1977] as

$$S = \int_0^1 (Y_0(u)+Y_B(u))^{-1}[Y_B(u)\,dN_{0,A}(u)-Y_0(u)\,dN_{B,AB}(u)]\,.$$

The variance of S may be estimated by

$$V = \int_0^1 (Y_0(u)+Y_B(u))^{-2}Y_0(u)\,Y_B(u)\,[dN_{0,A}(u)+dN_{B,AB}(u)]\,.$$

The statistic $SV^{-\frac{1}{2}}$ is under the hypothesis asymptotically normal (0,1) when the number of patients under observation increases to infinity. Of course, we get similar results corresponding to the hypothesis $\alpha_{0,B}(t)=\alpha_{A,AB}(t)$.

The simple model described above should of course be viewed as a prototype for studies of interaction of life history events. Each practical situation will need its own modification; in the next section we describe one such extension.

4. Application: Pustulosis palmo-plantaris and menopause.

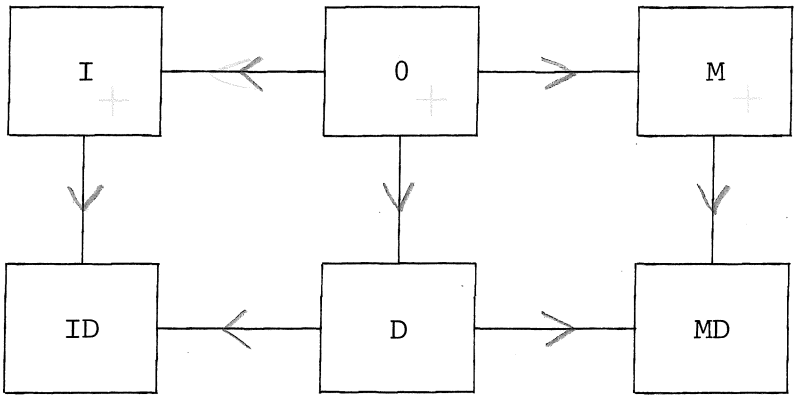
Pustulosis palmo-plantaris is a chronically recurrent skin disease localized to the palms of the hands and the soles of the feet. Owing to differential diagnostic problems vis-a-vis psoriasis the present study includes patients who besides pustulosis palmo-plantaris suffered from psoriasis vulgaris at other localizations. Pustulosis palmo-plantaris is a relatively rare disease with a

reported prevalence of .05 per cent in a Swedish study (Hellgren & Mobacken [1971]). The disease usually lasts for several years with first appearance most often between 35 and 55 years of age. There is a definite surplus of women among the patients (Ashurst [1964], Hellgren & Mobacken [1971]).

The present study concerns the 85 women from a consecutive series of 100 patients at the Department of dermato-venerology, Finsen Institute, Copenhagen. Owing to the above mentioned circumstances and since the etiology of the disease is completely unknown it was considered of interest to search for possible connection between the events of menopause as reported by the women and first appearance of pustulosis palmo-plantaris.

For 19 of the women the "menopause" was induced artificially, surgically or by X-ray treatment. We want to extend the model to allow for separate intensities for natural menopause and what we shall term "induced menopause".

The appropriate extension of the model is given in Figure 2, which should be self-explaining by analogy with Figure 1.



0 - no event has occurred. M - (natural) menopause has occurred.
D - disease has broken out. I - induced menopause has occurred.

Figure 2. A Markov chain model for the occurrence of natural and induced menopause, and the outbreak of pustulosis palmo-plantaris.

We denote the intensities of transition by $\alpha_{ij}(t)$. Clearly, estimation and testing may be performed in precisely the same way as before. The data to be analyzed are given in Table 1.

In order to estimate $A_{ij}(s,t)$ it is obviously a reasonable condition that the state i is not empty on any part of [s,t]. Below we give the age intervals over which the states 0,D,M and I are nonempty:

0: 0-54, D: 17-52

M: 37-72 I: 28-40 and 43-67

For the estimation some problems regarding ties had to be resolved. We scored multiple events of the same type as happening simultaneously, all with "risk set" corresponding to the situation before the event. As usual censoring was assumed to take place after events happen. For the five women who reported (natural or induced) menopause and outbreak of the disease as happening simultaneously, we assumed the outbreak of disease to happen last and we then included the women in the relevant risk set when calculating estimates and test statistics. Unfortunately, as will turn out below, the opposite interpretation will change one of the conclusions. Alternatively, a more careful discretisation of the model along similar lines as the discrete model by Cox [1972] or the grouped observations model by Kalbfleisch and Prentice [1973] could have been attempted.

TABLE 1

Age of first appearance, age of natural or induced menopause and age last seen for 85 female patients with pustulosis palmo-plantaris.

First Appear.	Natural Meno pause	Last seen	First Appear.	Natural Meno pause	Last seen	First Appear.	Natural Meno pause	Last seen	First Appear.	Induced Meno pause	Last seen
17	-	18				53	48	54	34	31	69
19	-	22	43	43	44	53	52	56	35	32	67
23	-	33	43	47	50				37	28	38
24	-	26	44	-	46				39	37	42
27	-	38	44	-	44	53	43	54	40	51	55
27	-	40				54	48	54	41	38	43
29	-	30				54	46	54	45	46	47
29	-	32	45	43	47				45	45	47
30	-	31	45	-	46	56	50	63	50	44	53
30	-	32	45	45	55	57	47	64	50	50	50
31	-	33	47	43	51	57	37	60	50	45	50
31	-	35	48	49	70	58	48	60	51	48	52
33	-	37	48	40	50	59	50	61	53	46	56
			48	47	49	59	47	64	53	50	54
35	-	36	49	46	49	60	50	62	54	43	56
			49	49	55	60	51	62	62	54	67
36	-	37	49	45	74	61	55	66	64	49	64
36	-	40	50	52	54				64	47	65
						64	48	66	67	52	71
37	-	37				64	42	68			
38	39	39				64	40	66			
			51	50	53						
40	50	74									
40	47	48	51	49	51	66	47	68			
40	50	52	51	48	66						
40	46	59	51	46	62	67	49	79			
			52	47	64	70	45	71			
41	-	41	53	48	56	73	45	74			
41	40	54									

Figure 3 shows the estimated integrated intensities $\hat{A}_{0,D}$, $\hat{A}_{M,MD}$ and $\hat{A}_{I,ID}$. Notice that $\hat{A}_{I,ID}$ is "pasted together" across the empty risk set at ages 41 and 42. The graphs give the overall impression that the intensity $\alpha_{I,ID}$ of outbreak of the disease after induced menopause (which is the local slope of $\hat{A}_{I,ID}$) is much higher than $\alpha_{0,D}$ for most of the ages where they may be compared. A similar conclusion seems to hold for $\alpha_{M,MD}$ versus $\alpha_{0,D}$ whereas $\alpha_{I,ID}$ and $\alpha_{M,MD}$ look about similar (the estimates $\hat{A}_{I,ID}$ and $\hat{A}_{M,MD}$ being roughly parallel.)

Correspondingly, the normalized generalized Savage test statistics described in Section 3b are -4.85 for the hypothesis $\alpha_{0,D}=\alpha_{I,ID}$, -3.04 for $\alpha_{0,D}=\alpha_{M,MD}$ (both highly significant) and equals the nonsignificant value of .66 for the hypothesis $\alpha_{I,ID}=\alpha_{M,MD}$. If for the five patients reporting menopause and outbreak of the disease simultaneously menopause was assumed to happen last, the same qualitative conclusions apply except that $\alpha_{0,D}=\alpha_{M,MD}$ now yields -1.24, corresponding to a (one-sided) significance probability of about 10.5%.

Similar comparison may be made between the intensities of getting menopause before or after outbreak of the disease. There seems to be no immediate explanation of a difference if any is found, and in the present case indeed no such difference was indicated.

Our conclusion is that induced menopause increases the chance of appearance of pustulosis palmo-plantaris. There is also an indication, however not a conclusive one, that natural menopause has the same effect.

References.

Aalen, O.O. [1975]. Statistical inference for a family of counting processes. Ph. D. Dissertation, Department of Statistics, University of California, Berkeley.

Aalen, O.O. [1977]. Nonparametric inference for a family of counting processes. Ann. Statist. (to appear).

Aalen, O.O. & Hoem, J.M. [1978]. Random time changes for multivariate counting processes. Scand. Act. J. (to appear).

Altshuler, B. [1970]. Theory for the measurement of competing risks in animal experiments. Math. Biosc. *6*, 1-11.

Ashurst, P.J.C. [1964]. Relapsing pustular eruptions of the hands and feet. British Journal of Dermatology *76*, 169-180.

Colding-Jørgensen, H. and Simonsen, W. [1940]. Statistical appendix (in Danish). In: J.M. Wollesen: Graviditet og galdestensdannelse. Kjøbenhavn: Nyt Nordisk, pp. 81-101.

Cox, D.R. [1972]. Regression models and life-tables. (with discussion) J.R. Statist. Soc. B *34*, 187-220.

Crowley, J. & Hu, M. [1977]. Covariance analysis of heart transplant survival data. J. Amer. Statist. Assoc. *72*, 27-36.

Freund, J.E. [1961]. A bivariate extension of the exponential distribution. J. Amer. Statist. Assoc. *56*, 971-977.

Hellgren, L. & Mobacken, H. [1971]. Pustulosis palmaris et plantaris. Prevalence, Clinical Observations and Prognosis. Acta Dermatovenereologica *51*, 284-288.

Kalbfleisch, J.D. & Prentice, R.L. [1973]. Marginal likelihoods based on Cox's regression and life model. Biometrika *60*, 267-278.

Lagakos, S.W. [1976]. A stochastic model for censored-survival data in the presence of an auxiliary variable. Biometrics *32*, 551-559.

Nelson, W. [1972]. Theory and application of hazard plotting for censored failure data. Technometrics *14*, 945-965.

Schweder, T. [1970]. Composable Markov processes. J. Appl. Prob. *7*, 400-410.

Turnbull, B.W., Brown, B.W., Jr., & Hu, M. [1974]. Survivorship analysis of heart transplant data. J. Amer. Statist. Assoc. *69*, 74-80.

PREPRINTS 1976

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Becker, Niels: Estimation for an Epidemic Model.

No. 2 Kaplan, Norman: A Generalization of a Result of Erdös and Renyi and a Related Problem.

No. 3 Lauritzen, Steffen & Remmer, Ole: An Investigation of the Asymptotic Properties of Geodetic Least Squares Estimates in Certain Types of Network.

No. 4 Asmussen, Søren: Almost Sure Behavior of Linear Functionals of Supercritical Branching Processes.

No. 5 Hald, Anders & Møller, Uffe: On the SPRT of the Mean of a Poisson Process.

No. 6 Lauritzen, Steffen: Projective Statistical Fields.

No. 7 Keiding, Niels: Inference and Tests for Fit in the Birth-Death-Immigration Process.

No. 8 Møller, Uffe: OC and ASN og the SPRT for the Poisson Process.

No. 9 Keiding, Niels: Population Growth and Branching Processes in Random Environments.

No. 10 Andersen, Per K., Andersen, Søren & Lauritzen, Steffen: The Average Noise from a Poisson Stream of Vehicles.

No. 11 Hald, Anders & Møller, Uffe: Multiple Sampling Plans of Given Strength for the Poisson and Binomial Distributions.

No. 12 Johansen, Søren: Two Notes on Conditioning and Ancillarity.

No. 13 Asmussen, Søren: Some Martingale Methods in the Limit Theory of Supercritical Branching Processes.

No. 14 Johansen, Søren: Homomorphisms and general exponential families.

No. 15 Asmussen, Søren & Hering, Heinrich: Some Modified Branching Diffusion Models.

No. 16 Hering, Heinrich: Minimal Moment Conditions in the Limet theory for General Markov Branching Processes.

PREPRINTS 1977

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Asmussen, Søren & Keiding, Niels: Martingale Central Limit Theorems and Asymptotic Estimation Theory for Multitype Branching Processes.

No. 2 Jacobsen, Martin: Stochastic Processes with Stationary Increments in Time and Space.

No. 3 Johansen, Søren: Product Integrals and Markov Processes.

No. 4 Keiding, Niels & Lauritzen, Steffen L. : Maximum likelihood estimation of the offspring mean in a simple branching process.

No. 5 Hering, Heinrich: Multitype Branching Diffusions.

No. 6 Aalen, Odd & Johansen, Søren: An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations.

No. 7 Johansen, Søren: The Product Limit Estimator as Maximum Likelihood Estimator

No. 8 Aalen, Odd & Keiding, Niels & Thormann, Jens: Interaction Between Life History Events.