Dynamic Analysis of Covariate Influence in Aalen's Linear Hazard Regression Model

Jochen Mau
University Hospital Duesseldorf, Department of Statistics in Medicine,
University Street 1, Building 23.02, Level 03,
40225 Duesseldorf, Germany
ismmau@uni-duesseldorf.de

In medical follow-up studies, the survival of n patients is typically influenced by the presence of or the exposure to some covariates $Z_1, \ldots, Z_k$. This may be modelled (Aalen, 1980) through the intensity process

$$\lambda_i\,(t;\, Z_1^{(i)}, \ldots, Z_k^{(i)}) = (\gamma_0(t) + \gamma_1(t)Z_1^{(i)}(t) + \ldots + \gamma_k(t)Z_k^{(i)}(t))\, Y_i(t)$$

associated with a multivariate counting process $\{N_1(t), \ldots, N_n(t);\, t \geq 0\}$; here $\{Y_i(t);\, t \geq 0\}$ are binary processes that indicate whether an individual i is still at risk in the study at time t-. This linear regression model is distinct from the more commonly known proportional hazards model (Cox, 1972) in that it does not assume constant effects of the covariates on the hazard function. Estimation of the time-integrated (cumulative) regression functions $\{\Gamma_j(t);\, t \geq 0\}$, $j = 0, \ldots, k$, proceeds from linearity properties in multivariate martingale theory (cf. Aalen, 1980; Anderson et al., 1983, for details). The estimates $\{G_j(t);\, t \geq 0\}$, $j = 0, 1, \ldots, k$, are typically represented componentwise for $j = 1, \ldots, k$ as Aalen plots (Mau, 1986), one for each covariate. A global assessment of the effect of a particular covariate will then imply a test statistic that involves the sequence of events (deaths) and the numbers at risk along the time interval of observation.

By the dynamic nature of the estimation procedure, successively incremented estimates with accumulating numbers of events, one would like an accordingly dynamic statistical assessment of a covariate's influence on the hazard function. The observed process of the evolving (global) test statistic (Aalen, 1989) will, however, miss the point since it presents only the "cumulative evidence". Several applications suggest that an influence of a covariate may come up rather among late events than among the early occurrences (Mau, 1988). Hence, an analysis by time segments may generally be more meaningful.

An immediately viable approach uses the theoretical properties of partitioned counting processes (Mau, 1985) and defines the segmentation points as either a fixed preselected time point or the (first) occurrence of a fixed number of events. Then Aalen's (1989) test statistic will be computed separately within each time segment; on any segment, a significant effect will be claimed at level α whenever a critical value according to an appropriately adjusted nominal level of significance, α', will be claimed, e.g. $\alpha' = \alpha/2$ when only two segments had been created from a single segmentation point. Though power properties in accord with the globally computed test statistic can be expected, Siegler (2001) recently demonstrated in a simulation study that the segmented approach has a higher power when the covariate effect is actually restricted to an early time interval: The lack of effect later „dilutes" the evidence from the early phase when only the globally computed test statistic is used.

Though Aalen's (1989) globally computed test statistic may be less powerful in the presence of an early terminating covariate effect, its handicap, this dilutive effect, may be used to suggest the location of such a break point from a successively evolving profile. In fact, the maximum of the test statistic's profile is an estimator of such a break point. Siegler (2001) found by simulations that, in his scenarios, the inherent positive bias (the break point is always seen too late, of course) is actually rather small.

The approaches discussed so far focus on effects that are restricted to one interval on the time axis; one would of course also like to have more flexibility. One aspect of such flexibility is to detect effects that come up early, disappear, and come up later again. Such pattern of covariate influence will not be frequently seen but may well arise in large data sets. Another aspect of flexibility is to overcome the preselection of segmentation points. In general, one would like to move a fixed-width window from left to right along the time axis and to assess, by the test statstic, the strength of evidence for a covariate effect from all events and the sample at risk within this window. "Fixed-width" can relate to a fixed length of the time-interval window, but it will be more appropriate to define it in terms of a fixed number of events: Simulation results by Siegler (2001) suggest that such windows should comprise at least ten events for the early parts of time axis and up to 50 events for the late parts, when sample sizes are around 300 and there is between 30 and 50% censoring. The resulting process of repeated evaluation of a test statistic will require a deeper investigation of its properties to derive critical values that comply with a prespecified overall level of significance.

REFERENCES

Aalen, O.O. (1980) A model for nonparametric regression analysis of counting process. Springer Lect. Notes Statist. 2, 1-25.

Aalen, O.O. (1989) A linear regression model for the analysis of life times. Statist. Med. 8, 907-925.

Andersen, P.K., Borgan, Ø, Gill, R.D., Keiding, N. (1993) Statistical Methods Based on Counting Processes, Springer, New York.

Cox, D.R. (1972) Regression models and life-tables (with discussion). J. Roy. Statist. Soc. B4, 187-220.

Mau, J. (1985) Statistical modelling via partitioned counting processes. J. Statist. Planning Inf. 12, 171-176.

Mau, J. (1986) On a graphical method for the detection of time-dependent effects of covariates in survival data. Appl. Statist. 35, 245-255.

Mau, J. (1988) A comparison of counting process models for complicated life histories. Applied Stochastic Models and Data Analysis 4, 283-298.

Siegler, K.E. (2001) Zur statistischen Analyse der zeitlichen Variabilität des Einflusses von Kovariablen im linearen Regressionmodell von Aalen, Dissertation. Fachbereich Statistik der Universität Dortmund, Dortmund.

ACKNOWLEDGMENT

Supported by Deutsche Forschungsgemeinschaft DFG grant Ma 1013/2-1.