Dynamic Regression with Recurrent Events

Jae Eui Soh[1,2] Yijian Huang[1]

[1]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University
[2]Division of Viral Hepatitis, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention

Recurrent Events

Multiple events of the same type for a subject in longitudinal studies. Examples inlcude:

- seizures in epileptic patients;
- successive tumors in cancer patients.

By the multiplicity nature, a dependence structure between events is often observed within a subject; not all dependence is captured by observed covariates, i.e., unobserved heterogeneity between individuals.

The Motivation for Dynamic Model: Constant Effect Over Time?

Most models with recurrent events assume constant effects of covariates, e.g., Andersen and Gill (1982) and Lin et al. (2000).

In practice, the effects may vary over time. In a clinical study for AIDS patients, for example,

- a drug may take time to reach its full efficacy,
- the treatment effect may erode over time as drug resistance develops;

e.g., Eshleman et al. (2001) and Wu et al. (2005).

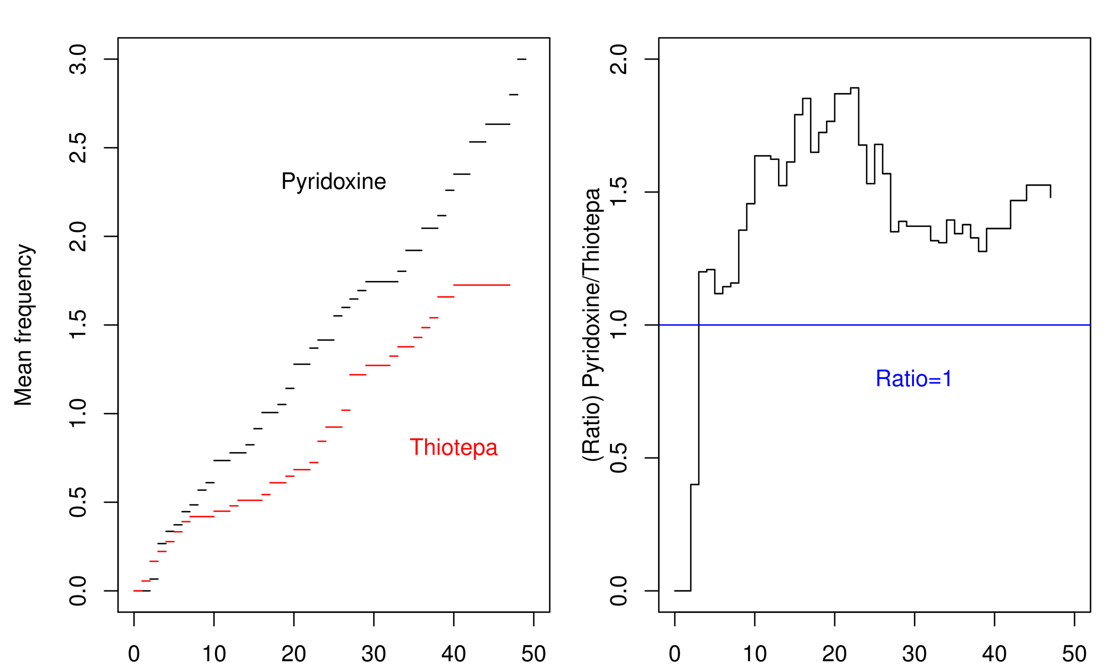

Figure: *Nonparametric Nelson-Aalen type mean frequency functions for two treatment arms in the bladder tumor study (Byar (1980)); and the ratio of these over time*

The Proposed Model

To accommodate time-varying effects, we propose a marginal dynamic regression model to target the mean frequency of recurrent events,

$$E(N^*(t)|Z) = \mu_0(t)\exp\{b_0(t)^\top Z\}$$
$$= \exp\{\beta_0(t)^\top \widetilde{Z}\}, t \geq 0.$$

- $N^*(t)$: the number of events on interval $[0, t]$
- Z: p-dimensional covariate vector
- $\mu_0(t)$: unspecified baseline mean frequency
- $b_0(t)$: p-dimensional time-varying coefficient

We adopt the conditional independence censoring assumption on incomplete follow-ups, that is,

$$N^*(\cdot) \perp C \quad | \, Z.$$

The Proposed Estimating Integral Equation

Under the proposed model and the independent censorship, it follows that

$$E\left(\widetilde{Z}\left[N(t) - \int_0^t Y(s)\, d\exp\{\beta_0(s)^\top \widetilde{Z}\}\right]\right) = 0,$$

where $N(t) = N^*(t \wedge C)$ and $Y(t) = I(C \geq t)$.

Therefore we propose an estimating integral equation, for *all* $t \geq 0$,

$$n^{-1}\sum_{i=1}^{n} \widetilde{Z}_i\left[N_i(t) - \int_0^t Y_i(s)\; d\exp\{\beta(s)^\top \widetilde{Z}_i\}\right] = 0.$$

Based on this, $\beta_0(\cdot)$ is sequentially estimated over ordered observed event times in the sample, cf. Peng and Huang (2007).

Large Sample Properties

Theorem 1: Uniform Consistency

Under regularity conditions C1–C6,
$\sup_{t\in[\kappa,\tau]} \|\widehat{\beta}(t) - \beta_0(t)\| \longrightarrow 0$, *almost surely.*

Theorem 2: Weak Convergence

Under regularity conditions C1–C6,
$n^{1/2}\{\widehat{\beta}(\cdot) - \beta_0(\cdot)\}$ *on* $(\kappa, \tau]$ *weakly converges to a mean-zero Gaussian process.*

The Proposed Bootstrap Inference Procedure

For interval estimation, we propose a multiplier bootstrap, adapting Rubin (1981); for *all* $t \geq 0$,

$$\sum_{i=1}^{n} \xi_i \widetilde{Z}_i \left[dN_i(t) - Y_i(t)\, d\exp\{\beta(t)^\top \widetilde{Z}_i\}\right] = 0.$$

- $\{\xi_i\}_{i=1}^n$: size n random sample from Exp(1)
- $\beta^*(t)$: the stochastic solution at time t

The $100(1-\alpha)\%$ confidence interval for $\beta_0(t)$ can be constructed with the $(\alpha/2)$th and $(1-\alpha/2)$th quantiles of the empirical distribution for $\beta^*(t)$.

Simulation Studies

Monte Carlo simulations with various setups demonstrated the proposed estimator was virtually unbiased and efficient, cp. Fine et al. (2004).

	$h_0(t) = log(t)$								$b_0(t) = \exp\{-t/\exp(1)\}$							
	Proposed Method				Fine et al. (2004)				Proposed Method				Fine et al. (2004)			
t	B	SD	SE	Cov95	B	SD	SE	Cov95	B	SD	SE	Cov95	B	SD	SE	Cov95
Multiplicative unit-mean gamma frailty with variance 1																
0.5	-8	240	234	94.5	-11	257	247	93.9	-6	393	380	95.0	-3	421	401	94.0
1.0	-1	209	203	93.7	2	230	229	94.3	-11	350	335	93.0	-23	388	380	93.8
1.5	0	202	194	93.6	-4	256	243	93.1	-11	339	324	93.3	-22	442	405	93.8
2.0	-5	203	194	93.2	-16	301	278	92.6	-6	342	324	92.8	-8	518	469	92.6
2.5	-9	213	199	93.1	-38	424	370	91.4	0	360	333	92.6	-1	725	625	90.0

NOTE: B: empirical bias (×1000); SD: empirical standard deviation (×1000); SE: average standard error (×1000); Cov95: empirical coverage probability of the Wald 95% confidence interval (×100). Based on 1,000 Monte Carlo replications.

Application to the Bladder Tumor Data

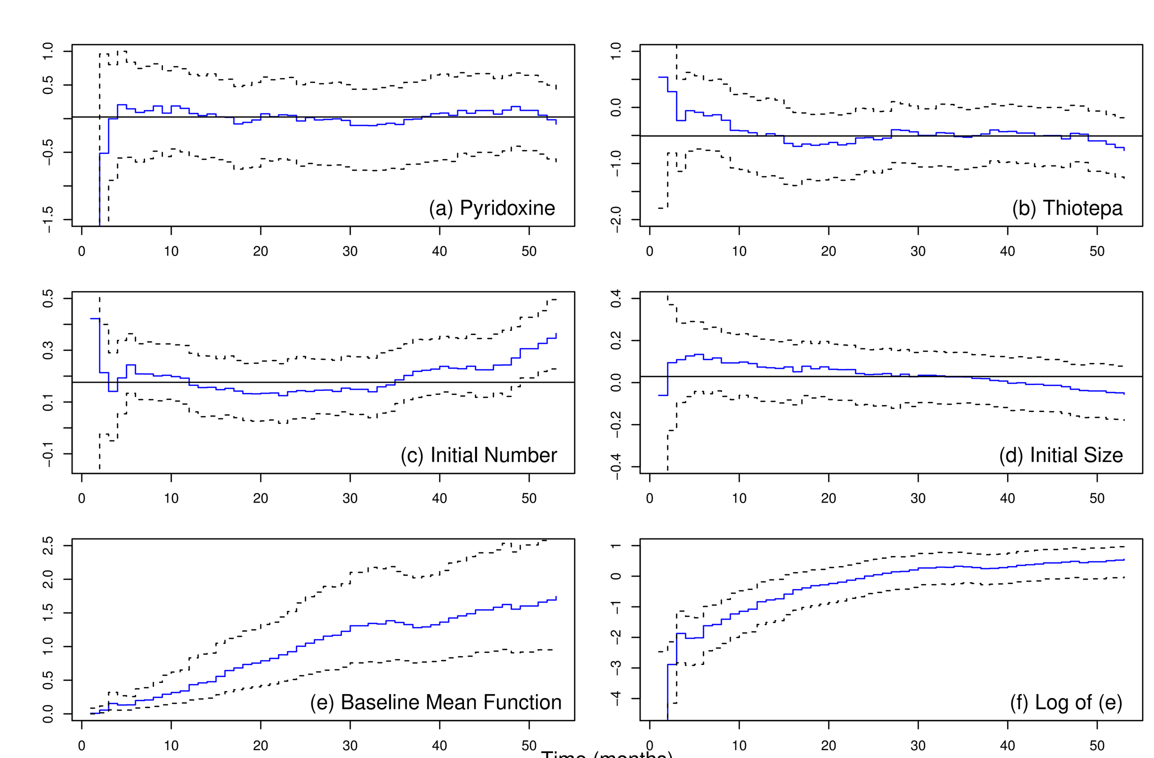

Figure: *Estimates for time-varying effects of covariates and the baseline mean frequency function (blue rugged lines); with the point-wise 95% bootstrap percentile confidence intervals (dashed lines).*

Remarks

- We propose a marginal model on mean frequency of recurrent events, accommodating *dynamic* effects of covariates, cf. the *proportional* means model of Lin et al. (2000);
- It is a *global* model over time for evolving effects of covariates, which facilitates efficient estimation, cf. Fine et al. (2004)'s *local* model.
- Consistency and weak convergence of the proposed estimator are established.
- The proposed nonparametric bootstrap inference procedure provides confidence band construction even for an infinite-dimensional quantity.
- Conducted simulations and two real data analyses illustrated practical utility of the proposed method.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* **10,** 1100–1120.

Byar, D. P. (1980). *The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumours: Comparisons of Placebo, Pyridoxine and Topical Thiotepa*. Springer US, Boston, MA.

Eshleman, S. H., Mracna, M., Guay, L., Deseyve, M., Cunningham, S., Mirochnick, M., et al. (2001). Selection and fading of resistance mutations in women and infants receiving nevirapine to prevent hiv-1 vertical transmission (hivnet 012). *AIDS* **15,** 1951–7.

Fine, J. P., Yan, J., and Kosorok, M. R. (2004). Temporal process regression. *Biometrika* **91,** 683–703.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62,** 711–730.

Peng, L. and Huang, Y. (2007). Survival analysis with temporal covariate effects. *Biometrika* **94,** 719–733.

Rubin, D. B. (1981). The bayesian bootstrap. *The Annals of Statistics* **9,** 130–134.

Wu, H., Huang, Y., P Acosta, E., L Rosenkranz, S., R Kuritzkes, D., J Eron, J., et al. (2005). Modeling long-term hiv dynamics and antiretroviral response: Effects of drug potency, pharmacokinetics, adherence, and drug resistance. *Journal of Acquired Immune Deficiency Syndromes* **39,** 272–83.

Contact Information

Jae Eui Soh
Email: jay.soh@emory.edu