

Predicting reconviction : do some predictors fade with length of follow-up?

Brian Francis, Keith Soothill and Les Humphreys
Lancaster University, UK

LANCASTER
UNIVERSITY



American Society of Criminology Conference, Chicago, 14 November, 2012

Long term follow-up studies

Two types of longitudinal designs in recidivism studies in criminology

Type 1 Repeated contact longitudinal studies

- a) Individuals in the study are selected at time $t=0$.
- b) Their previous criminal history and other covariates (eg age of offence, demographic variables etc) are measured at this point
- c) The offenders are followed up repeatedly
- d) Covariates such as marital status, employment) are also updated at the subsequent follow-up points.

Prospective design.

Example Farrington and West Cambridge delinquency study.

Long term follow-up studies (2)

Type 2. Administrative follow-up studies

- a) Individuals in the study are selected at time $t=0$.
- b) Their previous criminal history and other covariates (age of offence, demographic variables etc.) are measured once at this point

In this type of study, the offenders are followed up using **administrative data** (police records, court conviction records etc.). While conviction and arrest information is generally available, information on socio-demographic covariates beyond $t=0$ is generally not available. Design can be retrospective as well as prospective.

Methods available for recidivism studies

Logistic regression fixed follow-up time $t=T_f$ taken (e.g. two years) and effect of covariates at $t=0$ on whether an individual has been reconvicted or not can be assessed.

Ordinal regression an extension of logistic regression, where interest is in building models for recidivism at more than one follow up time. (e.g. one year, two years etc). Estimates of *covariate effects* are the same for each follow-up time but intercept changes.

Survival analysis models models the changing hazard of reconviction over time and the effect of covariates on the hazard. No need for a fixed follow-up time. Can also incorporate **time-dependent effects** where covariates are changing over time - typically collected in Type 1 studies. **A more flexible approach**

The focus of this talk

The focus is **not on time-dependent covariates** in survival analysis (where the covariates are changing over time) **but rather on time-varying coefficients**- where the *effects* of the covariates on the hazard are changing over time. (Martinussen et al, 2002)

In other words, for some covariates we might expect **strong effects** on the hazard of reconviction in **the short term** (say up to two years), but these effects might **weaken** over **the longer term**. For some other covariates, the parameters might stay unchanged.

Such methods are particularly relevant for Type 2 long term follow-up studies. (although they can also be used for type 1 studies)

The statistical models

We define the hazard of recidivism for an offender i at time t to be $\lambda_i(t)$ with baseline hazard $\lambda_0(t)$.

- A. Covariates which are time constant - no time varying parameters.

$$\lambda_i(t) = \lambda_0(t)\exp(\beta^T X_i)$$

- B. Covariates which are time dependent- no time varying parameters

$$\lambda_i(t) = \lambda_0(t)\exp(\beta^T X_i(t))$$

- C. Covariates which are time constant- **time varying parameters**

$$\lambda_i(t) = \lambda_0(t)\exp(\beta(t)^T X_i)$$

Our interest is in model C. We use a Cox formulation and treat the baseline hazard as non-parametric.

Does Model C provide extra insight over the standard model A?

Illustration - A study of UK middle class offenders.

A consecutive series of 388 offenders who were seeking white-collar employment between 1 January 1970 and 31 March 1973 with Apex Trust acting as a specialist employment agency for ex-offenders.

Focus on middle-class rather than just on white-collar offenders. All offenders had CVs and were seeking white collar employment.

Some very different types of middle-class offender with very different likelihoods of being convicted of further crime. Some are murderers and sex offenders, others property offenders.

Nearly all have been in prison.

The five clusters

Using **latent class analysis**, we have identified five clusters based on the offender's prior offending frequency, offending pattern and age.

1. 'Low-rate *white-collar* offenders.' (31%)
2. 'Low-rate *general* offenders.' (25%)
3. 'Medium-rate acquisitive specialists.' (22%)
4. 'Medium-rate generalists.' (14%)
5. 'High-rate generalists.' (9%)

Soothill, Humphreys and Francis (2012), *Brit J Crim.*

8 year recidivism of middle class male offenders

We might expect that some covariates will act on short term risk, and others will act on longer term risk.

We model the hazard of recidivism using a time varying effects Cox model. As covariates, we use

- Modal cluster membership.(based on age and previous criminal history)
- Target offence type
- Education at APEX interview (three levels)
- Experience of previous custody
- Marital status at APEX interview.(ever married/never married).
- Problems identified at APEX interview (eg Alcohol, drugs use) - two or more
- Whether placed by APEX into a job

Procedure

- We can identify whether a specific term is important.
- If it is important, we can test whether the parameter varies over time, or is time constant.
- If the parameter is time varying, we examine cumulative beta plots to gain some insight into how beta is changing over time.
- Use the `timecox` function in the `timereg` library in \mathbb{R}
- Interpretation is based on **cumulative betas** over time (Martinussen and Scheike, 2006).

Significance of individual terms(p-values)

Offending_Cluster- 2	0.620
Offending_Cluster- 3	0.001
Offending_Cluster- 4	0.001
Offending_Cluster- 5	<0.001
Previous_Custody- Yes	0.659
Placed Yes	0.442
Ever_Married- Yes	0.087
Target_Offence- Sexual	<0.001
Target_Offence- Burglary	<0.001
Target_Offence- Theft	<0.001
Target_Offence= Other	<0.001
Target_Offence- White-collar	<0.001
Number_of_Problems-2 or more	0.069
Education olevel+other	0.085
Education degree. a-level	0.182

Examine overall p-values for each parameter.

We remove the variables “Placed ”, “Education”, and “Previous_Custody” as uninformative for subsequent reconviction. At this stage, P-values for “Ever_Married” and “Number_of_problems” are 0.069 and 0.037 and are retained.

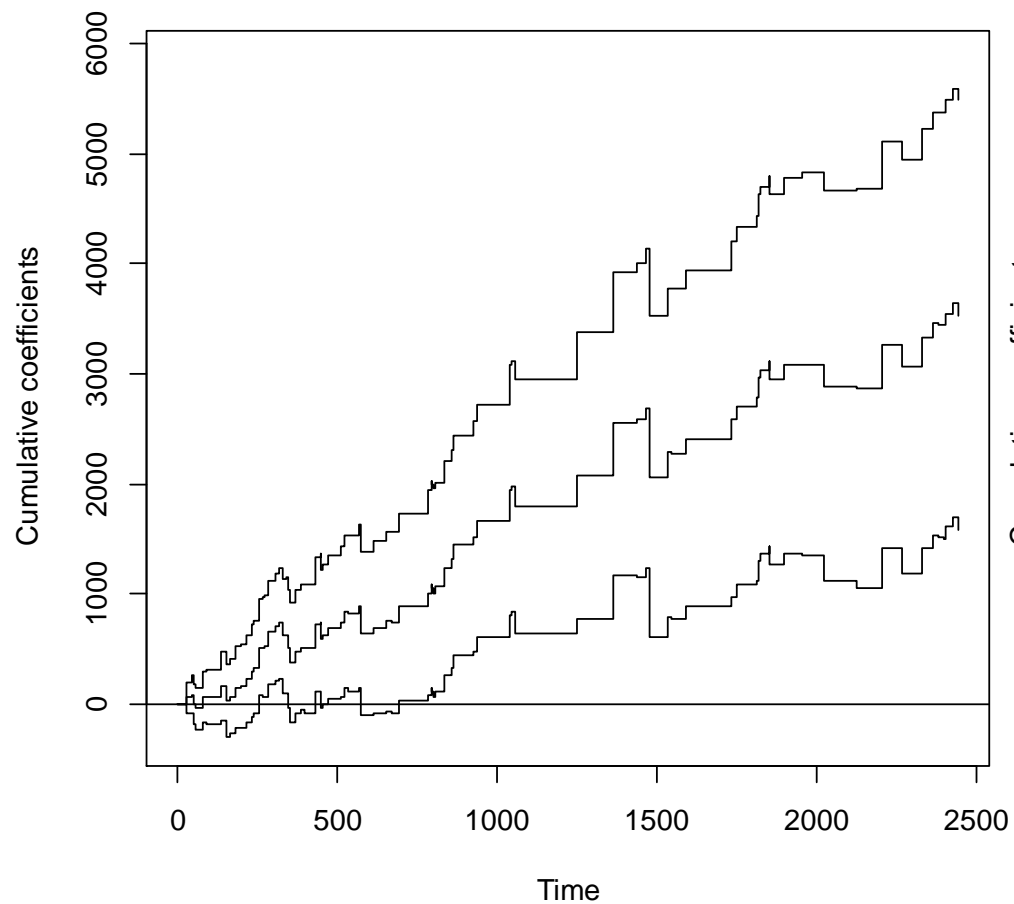
Time varying or time constant coefficients?

R gives two tests. We report the Cramer Von Mises test.

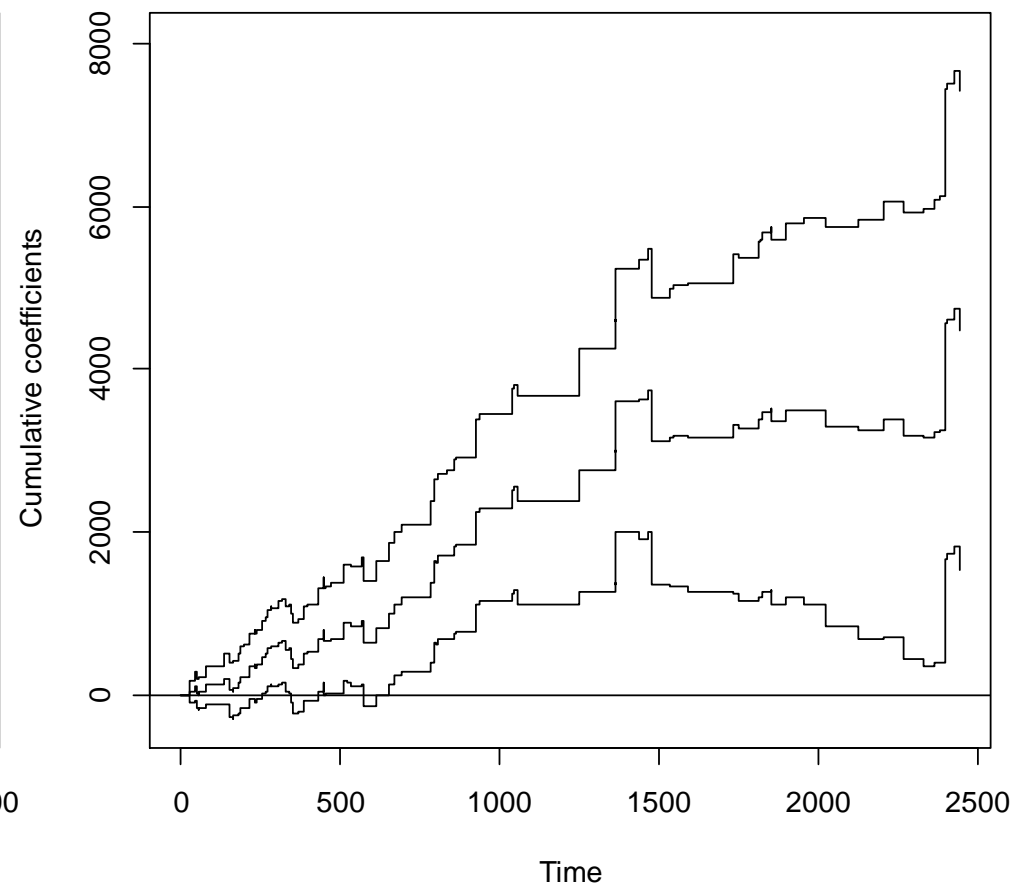
	Cramer von Mises test	p-value	H ₀ :constant effect
(Intercept)	1.39e+10		0.000
Offending_Cluster- 2	1.21e+08		0.925
Offending_Cluster- 3	1.51e+08		0.793
Offending_Cluster- 4	1.22e+08		0.885
Offending_Cluster- 5	5.71e+08		0.661
Ever_Married- Yes	1.38e+08		0.575
Target_Offence- Sexual	1.49e+10		0.000
Target_Offence- Burglary	1.68e+10		0.000
Target_Offence- Theft	1.37e+10		0.000
Target_Offence= Other	5.57e+10		0.000
Target_Offence- White-collar	1.41e+10		0.000
Number_of_Problems-2 or more	1.25e+08		0.583

Results show that effect of previous criminal history (via the offending clusters) marital status, and number of problems are all constant over time. However the effect of the *type* of the target conviction is time varying. We look at the cumulative beta plots.

Offending_Cluster-4

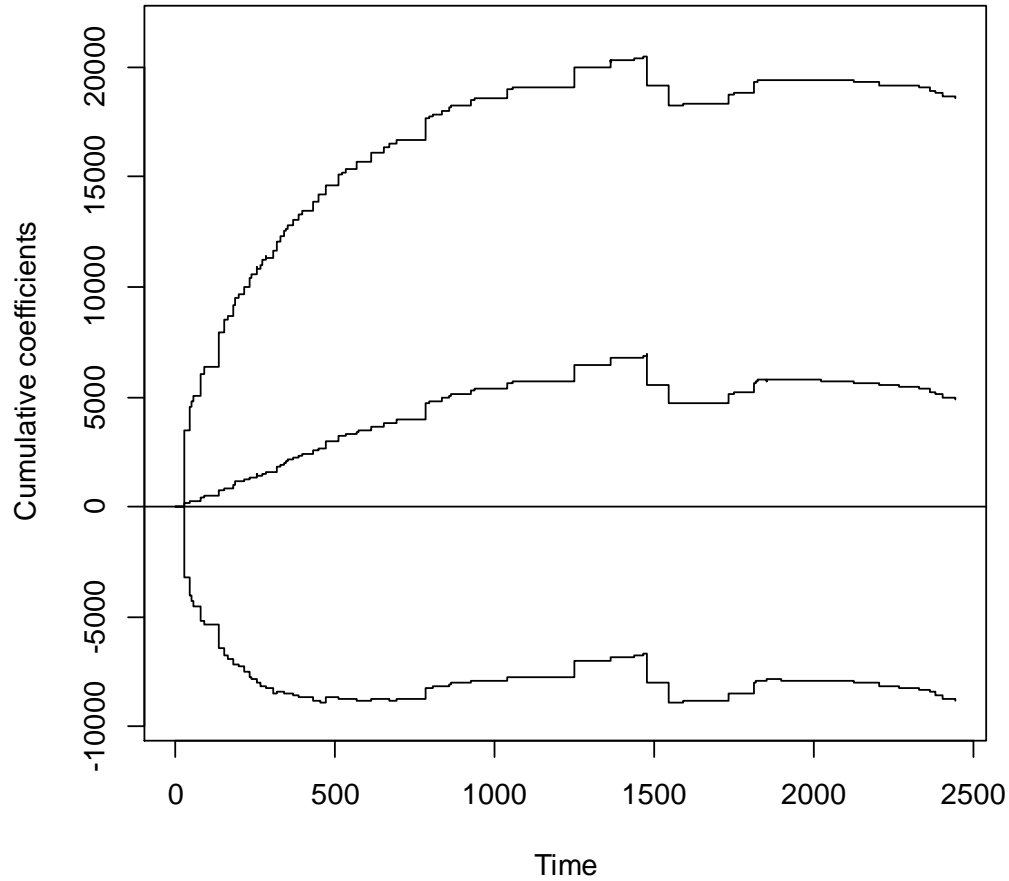


Offending_Cluster-5

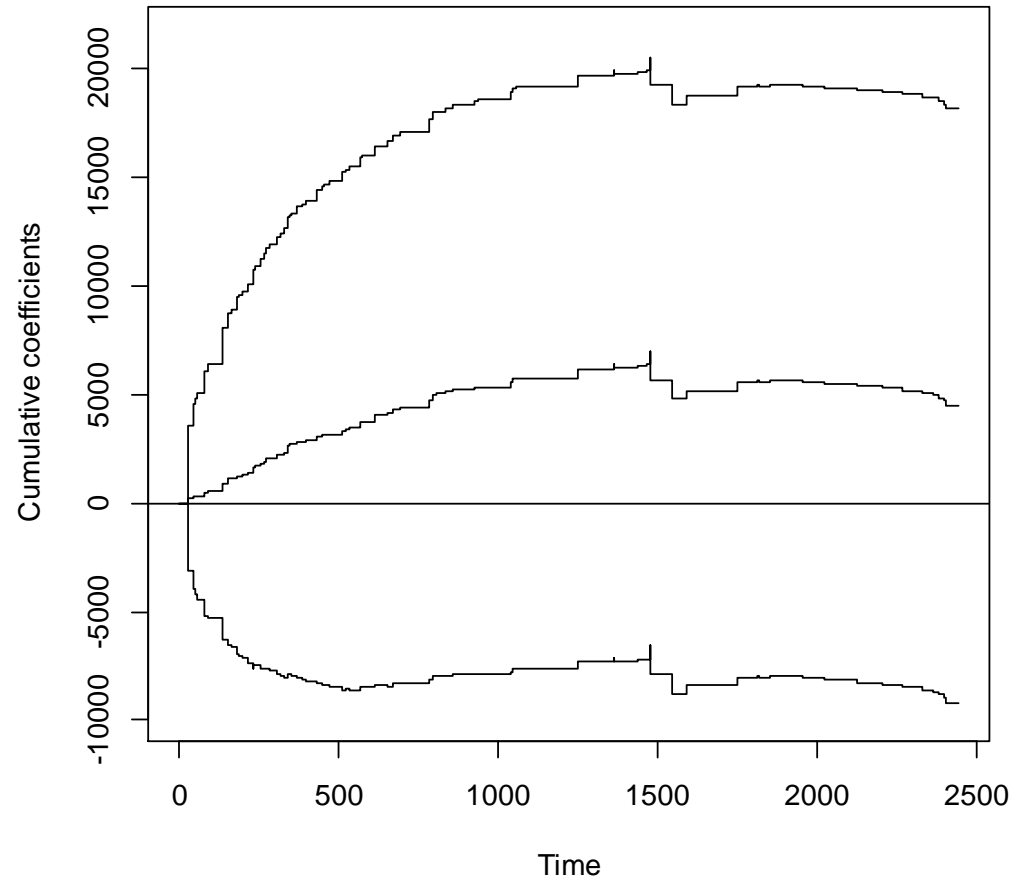


Cumulative beta plots with 95% CIs for offending clusters show (nearly) straight lines, showing the effect of previous history persists.

Target_Offence- Sexual

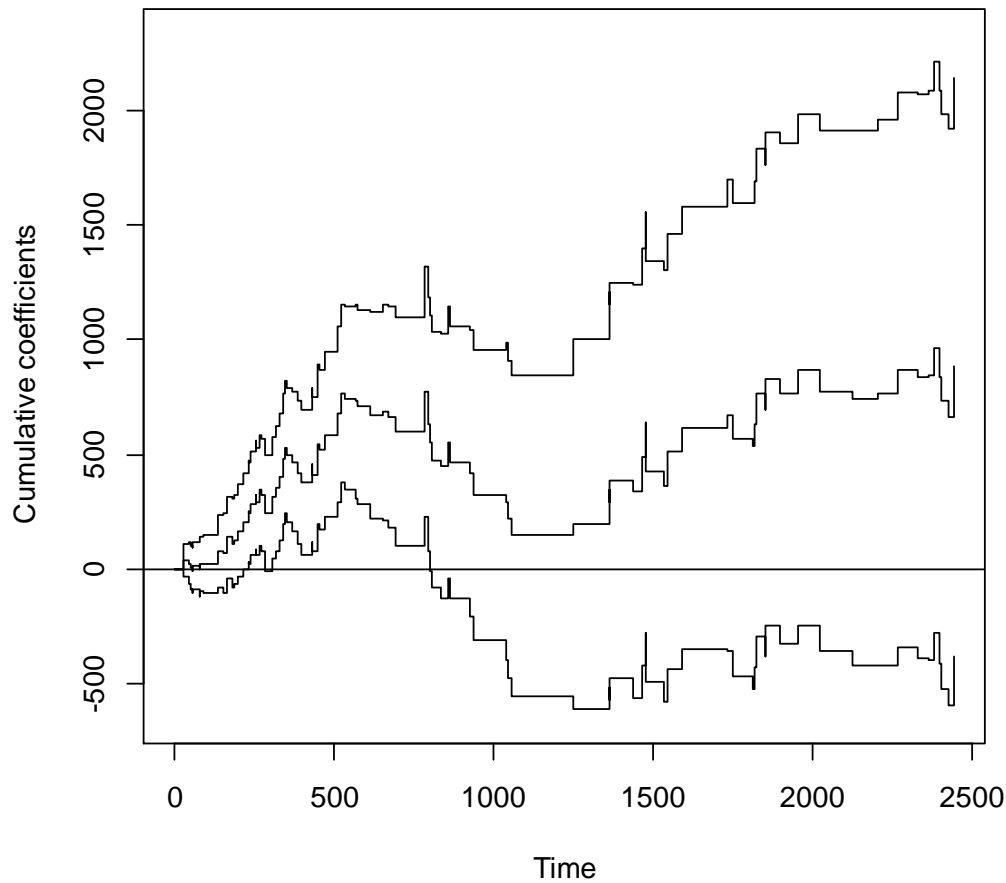


Target_Offence- Burglary

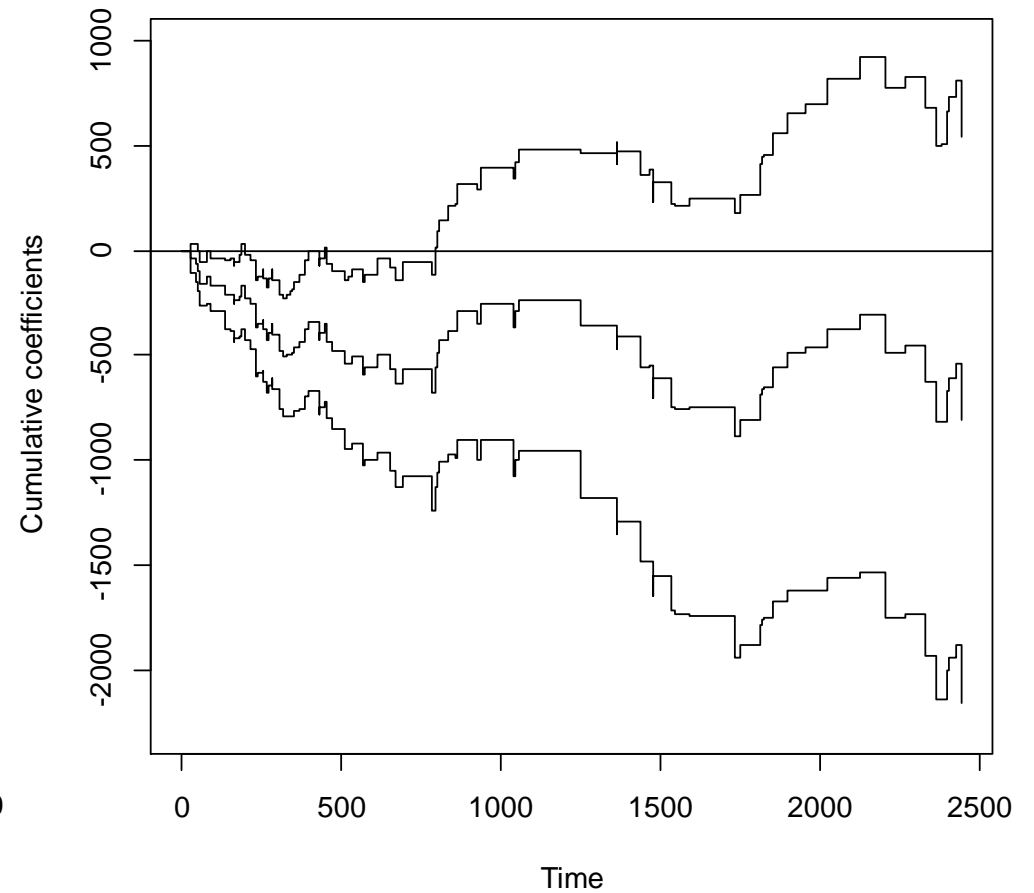


Effect of “target offence” gradually fades. Curves become flat (showing no positive beta effect) after around 1000 days or three years.

Number_of_Problems-2 or more



Ever_Married- Yes



For both marital status and number of problems, the graphical output show a short term effect up to about 500 days (~18 months) then a flattening. However, the Von Mises test did not identify these as time varying.

Technical discussion

Can simplify model by replacing any time-varying effect by a time constant one. This then becomes a semi-parametric hazard model with two sets of covariates - one set with time constant effects and the other set with time varying effects.

The implementation in R is useful, but smoothed **beta plots** rather than **cumulative beta** plots are probably better to present to a criminological audience.

No likelihood or deviance or AIC statistics are given in as part of the `timecox` function.

Work in progress - are there better ways of fitting this model?

Criminologically---

Analysis has identified that the effect of certain covariates do fade over time. The effect of **type of target offence fades** after **three years** if there has been no conviction up to that point.

There is also graphical evidence that the effects of **marital status** and **life problems fade** after about **eighteen months**.

One way of proceeding might be to use the method as **exploratory**, and to carry out separate hazard analyses for different time windows. 0 -18 months, 18 months to three years, more than three years.

References

Aalen, O.O., Borgan, O., and Gjessing, H.K. (2008) *Survival and Event History Analysis: A Process Point of View* Springer

Martinussen, T. and Sheike, T.H. (2006) *Dynamic Regression Models for Survival Data* Springer.

Martinussen, T., Scheike, T. H., and Skovgaard, I. M. (2002), “Efficient Estimation of Fixed and Time-Varying Covariate Effects in Multiplicative Intensity Models,” *Scandinavian Journal of Statistics*, 29, 57-74.

Soothill, K, Humphreys, L. and Francis, B. (2012) Middle Class Offenders - a 35 year follow-up. *Brit. J. Criminol.*, 52, 765-785