Supplementary Appendix for the paper

"Medication for ADHD and Criminality"

Paul Lichtenstein, Linda Halldner, Johan Zetterqvist, Arvid Sjölander, Eva Serlachius, Seena Fazel, Niklas Långström, Henrik Larsson

Table of contents

The main statistical analyses methods used in the paper "Medication for ADHD and Criminality"

In the paper, we have used three main statistical analyses methods; Kaplan-Meier curves, (non-stratified) Cox regression, and stratified Cox regression. These are standard analysis methods for time-to-event (aka survival) data. However, there are two features in our data that complicate the analyses, as well as the interpretation of the results. 1) The covariate of main interest (ADHD medication) is time-varying, and 2) the event of interest (conviction) is recurrent (in opposite to death, for instance, which can only occur once). In this Appendix we explain our main analyses in detail. We assume that the reader is already familiar with Kaplan-Meier curves and Cox regression for time-stationary covariates and non-recurrent events.

We illustrate the theory with a simple toy example. Consider Figure S1, which displays the follow-up for two fictitious individuals, A and B. At enrollment (month 0), neither A nor B is medicated ($X = 0$). Individual A starts taking the medication at 12 months after enrollment ($X = 1$), and continues until he is censored, at 20 months after enrollment. Individual B starts taking the medication at 10 months after enrollment, and continues until he is censored, at 32 months after enrollment. A is convicted at 2 months after enrollment, and at 8 months after enrollment. B is convicted at 9 months after enrollment, and at 17 months after enrollment. We use unfilled circles for censoring, and filled circles for conviction.

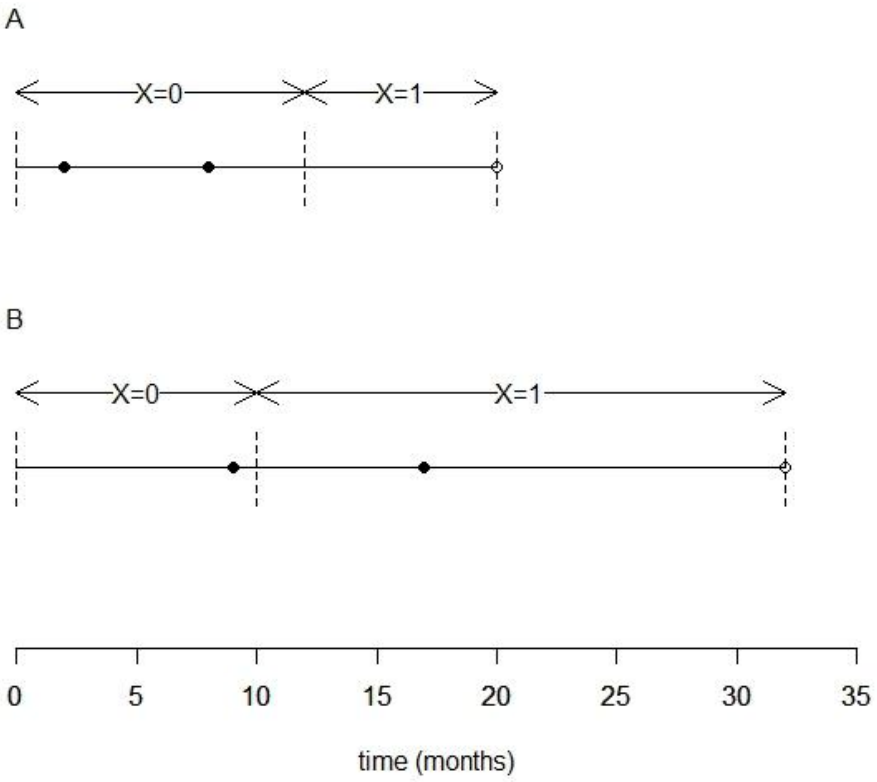

Figure S1

In what follows, we assume that there are D observed events (i.e. convictions) at distinct times $t_1 < t_2 < \cdots < t_D$. In our toy example, we have that $D = 4$, $t_1 = 2$, $t_2 = 8$, $t_3 = 9$, and $t_4 = 18$. To make notation simple we assume that there are no "ties", i.e. that there is at most one event per time-unit. Extensions to data with ties are straight-forward. For a time-stationary covariate X, the Kaplan-Meier curve for covariate level x is calculated as

$$\hat{S}_x(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_{jx}}{n_{jx}}\right) \qquad (1)$$

where d_{jx} and n_{jx} is the is the number of events at time t_j and the number of individuals in the risk set at time t_j, respectively, among those individuals with $X = x$ (that is, d_{jx} is either 0 or 1). Breslow and Crowley (1974) formally showed that the Kaplan-Meier curve consistently estimates the conditional survival function

$$S_x(t) = \Pr(T > t | X = x) = \exp\left\{-\int_0^t \lambda(u|X = x)du\right\} \qquad (2)$$

where

$$\lambda(t|X=x) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T < t+\Delta t | T \ge t, X=x)}{\Delta t} \qquad (3)$$

is the conditional hazard function (rate) at time t, given $X = x$. Heuristically, the convergence of $\hat{S}_x(t)$ to $S_x(t)$ can be seen by noting that $\hat{S}_x(t)$ can be written as

$$\hat{S}_x(t) = \exp\left[\log\left\{\prod_{j:t_j \le t}\left(1-\frac{d_{jx}}{n_{jx}}\right)\right\}\right] = \exp\left\{\sum_{j:t_j \le t}\log\left(1-\frac{d_{jx}}{n_{jx}}\right)\right\} \approx \exp\left(-\sum_{j:t_j \le t}\frac{d_{jx}}{n_{jx}}\right), \qquad (4)$$

and noting that $\frac{d_{jx}}{n_{jx}}$ is the "sample analog" of $\lambda(t_j|X=x)dt$.

There is no universally accepted method to deal with time-varying covariates in Kaplan-Meier curves. Snappin et al (2005) gave an overview of various ad hoc methods that have been used in the medical literature, and showed that these methods have serious limitations. They proposed an alternative method, which they called the "extended Kaplan-Meier curve". They showed that the extended Kaplan-Meier curve has a more principled underpinning than the ad hoc methods, and that it can be viewed as a nonparametric analog to Cox regression with time-varying covariates, which is standard in the literature (Klein and Moeschberger, 1997). Let $X(t_j)$ denote the covariate level at event time t_j, for a given individual. Snappin et al (2005) defined the extended Kaplan-Meier curve for covariate level x as

$$\hat{S}'_x(t) = \prod_{j:t_j \le t}\left(1-\frac{d'_{jx}}{n'_{jx}}\right) \qquad (5)$$

where d'_{jx} and n'_{jx} is the is the number of events at time t_j and the number of individuals in the risk set at time t_j, respectively, among those individuals with $X(t_j) = x$. The crucial distinction between $\hat{S}_x(t)$ and $\hat{S}'_x(t)$ is that in the latter, the risk set is updated at each event time t_j, including only those who happen to have covariate level x at the particular event time t_j. Thus, in the computation of $\hat{S}'_x(t)$ individuals can enter and exit the risk set several times during follow-up. Using arguments similar to those in Breslow and Crowley (1974), it can be shown that $\hat{S}'_x(t)$ consistently estimates the function

$$S'_x(t) = \exp\left\{-\int_0^t \lambda\{u|X(u)=x\}du\right\}, \qquad (6)$$

where

$$\lambda\{t|X(t)=x\} = \lim_{\Delta t \to 0} \frac{\Pr\{t \le T < t+\Delta t | T \ge t, X(t)=x\}}{\Delta t} \qquad (7)$$

is the conditional hazard function (rate) at time t, given $X(t) = x$. Heuristically, the convergence of $\hat{S}'_x(t)$ to $S'_x(t)$ can be seen by noting that $\hat{S}'_x(t)$ can be written as

$$\hat{S}'_x(t) = \exp\left[\log\left\{\prod_{j:t_j \le t}\left(1 - \frac{d'_{jx}}{n'_{jx}}\right)\right\}\right] = \exp\left\{\sum_{j:t_j \le t}\log\left(1 - \frac{d'_{jx}}{n'_{jx}}\right)\right\} \approx \exp\left(-\sum_{j:t_j \le t}\frac{d'_{jx}}{n'_{jx}}\right), \qquad (8)$$

and noting that $\frac{d'_{jx}}{n'_{jx}}$ is the "sample analog" of $\lambda\{t_j | X(t_j) = x\}dt$.

The function $S'_x(t)$ is not a true survival function, in the sense that it cannot generally be written as a probability of surviving time t, for any fixed subpopulation. The reason is that it integrates hazard functions which apply to different subsets of the population, defined by the current covariate level at each time u. Rather, it can be viewed as a non-parametric summary of covariate-specific hazards. There is an important special case though, in which interpretation can be simplified. Suppose that the conditional hazard at time t, given current covariate level $X(t) = x$, is independent of the covariate history. We can then write

$$\lambda\{t | X(t) = x\} = \lambda\{t | X(u) = x \; for \; u \le t\} \qquad (9)$$

I.e. the hazard at time t for those who had covariate level x at time t is the same as the hazard at time t for those that had covariate level x all the way from baseline and up to time t. Under assumption (9) we have that

$$S'_x(t) = \exp\left[-\int_0^t \lambda\{u | X(v) = x \; for \; v \le u\}du\right] \qquad (10)$$

which by definition is the survival function for the fixed subpopulation that have covariate level x all through follow-up.[1] In terms of our application; if the conviction hazard is independent of medication history, given current medication level, then the extended Kaplan-Meier curves can be interpreted as the (estimated) survival functions for those who are on and off medication, respectively, all through follow-up. We note that assumption (9) can be empirically tested, for instance by running a Cox regression analysis with both current and (some summary of) previous medication as covariates in the model. We did this for our data (see the section labeled "Long-term associations" in the paper), detecting no statistically significant association with previous medication, given current medication. Thus, we tentatively conclude that assumption (9) may be reasonable for our data.

Neither the ordinary Kaplan-Meier curve, nor the extended Kaplan-Meier curve, is developed for recurrent events. To deal with recurrence of convictions, we reset time to 0 for each individual, at every conviction. We emphasize that this procedure does not imply that individuals are assumed to have the same hazard functions regardless of the number of previous convictions; it simply implies that we are "pooling" (e.g. averaging) the hazard functions over previous history of convictions. To use a simple

[1] Snappin et al (2005) stated, without motivation, that extended Kaplan-Meier curves "represent hypothetical cohorts whose covariate values remain constant during follow-up". Our argument shows that this interpretation is only valid under assumption (9).

analogy, suppose we pool men and women together, and compute the risk in the pooled population. This does not imply that we assume the risk to be the same for men and women, only that we average the sex-specific risks. For consistency, we have treated recurrent events in the same fashion in the Cox regression analyses described below. We note that in the stratified Cox regression analysis (see below), we have adjusted for number of previous convictions by including number of previous convictions as a categorical covariate in the regression model.

To compute the extended Kaplan-Meier curves for our toy example by hand, it is useful to first "pre-process" the data so that the follow-up is divided into sub-periods for each individual. Specifically, every conviction and every switch of medication level initiates a new sub-period. A sub-period that follows a conviction is considered as starting at "baseline" (time 0 in the analyses), whereas a sub-period that follows a switch in medication level is considered left-truncated at the time of the switch. A sub-period that ends with a medication switch is considered as censored. Figure S2 displays the sub-periods produced by the fictitious individuals A (left panel) and B (right panel). For instance, the topmost line to the left in Figure S2 corresponds to the first sub-period for A, that starts when A is enrolled, and ends with the conviction at 2 months after enrollment. The bottommost line to the left in Figure S2 corresponds to the last sub-period for A, that starts immediately after the switch in medication level at 12 months after enrollment, and ends with the censoring at 24 months after enrollment. Figure S3 displays the same sub-periods, but divided into medicated (left panel) and non-medicated (right panel) sub-periods.

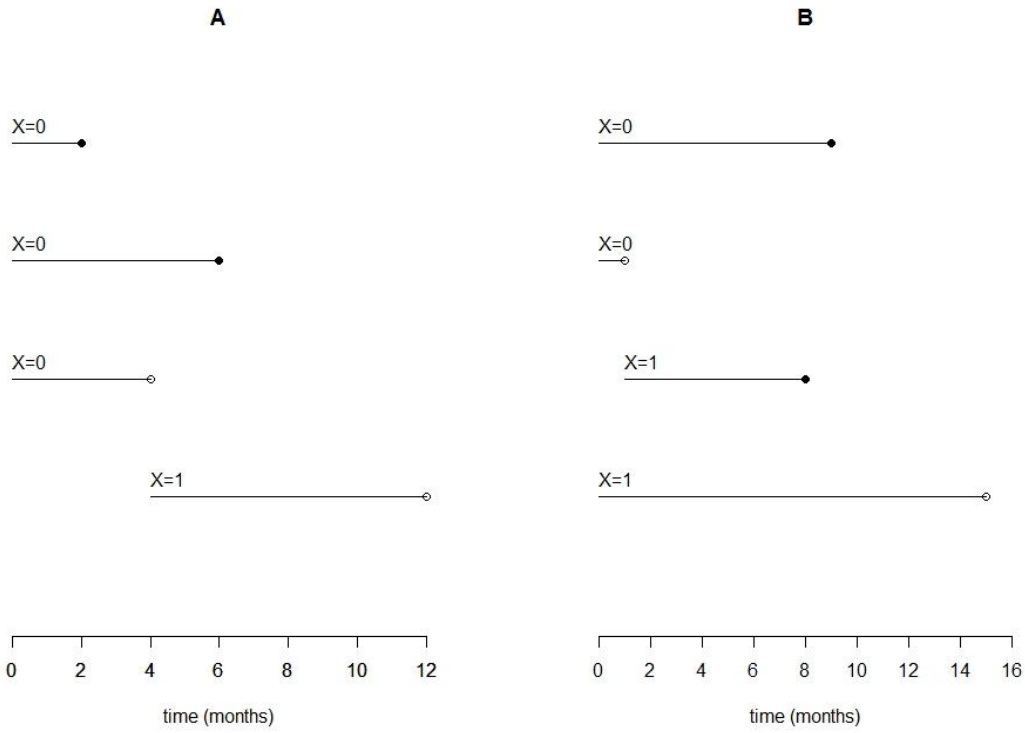
A
X=0
X=0
X=0
X=1
0
2
4
6
8
10
12
time (months)
B
X=0
X=0
X=1
X=1
0
2
4
6
8
10
12
14
16
time (months)

Figure S2

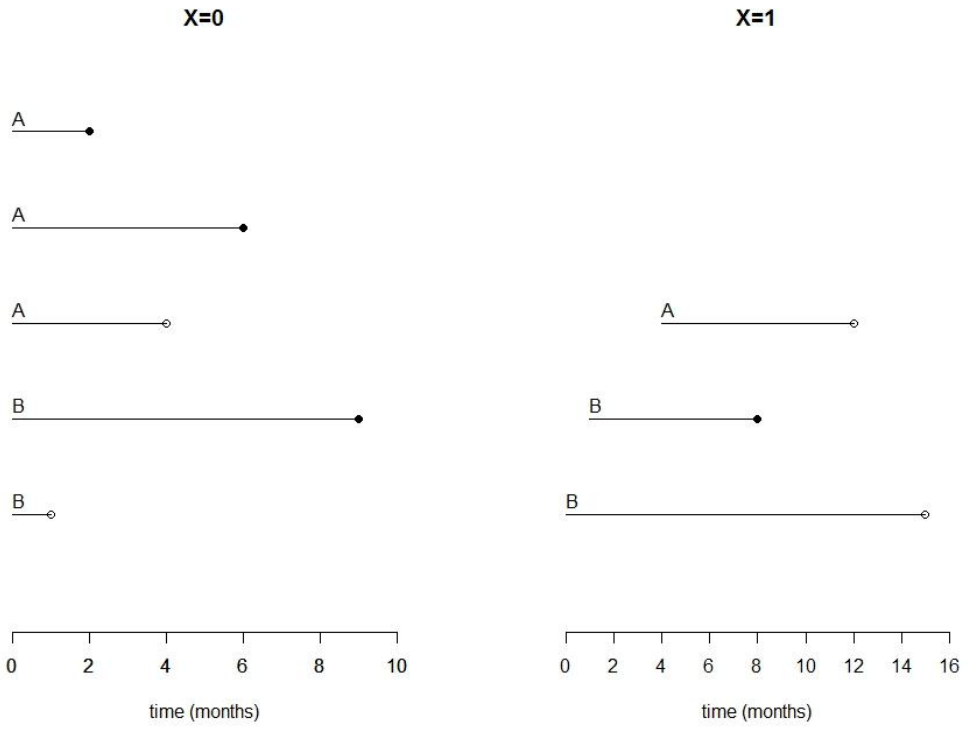

Figure S3

After this pre-processing, the extended Kaplan-Meier curves can be conveniently computed by applying the formula for the ordinary Kaplan-Meier curves (equation (1)) to the sub-periods, as if they would come from unrelated individuals with time-stationary covariates. Among the set of medicated sub-periods, there is only one conviction, at month 8. At this time, there are three medicated sub-periods at risk. Thus, the extended Kaplan-Meier curve for the medicated sub-periods starts at 1 at month 0, then takes a step down to (1-1/3) = 2/3 at month 8, and remains at 2/3 through the rest of follow-up. Among the set of non-medicated sub-periods, there are three convictions, at months 2, 6 and 9. At month 2 there are 4 sub-periods at risk, at month 6 there are 2 sub-periods at risk, and at month 9 there is only one sub-period at risk. Thus, the extended Kaplan-Meier curve for the medicated sub-periods starts at 1 at month 0, then takes a step down to (1-1/4) = 3/4 at month 2, followed by a step down to 3/4 x (1-1/2) = 3/8 at month 6, and finally takes a step down to 3/8 x (1-1/1) = 0 at month 9. Figure S4 displays the extended Kaplan-Meier curves for medicated (solid line) and non-medicated (dashed line) sub-periods.

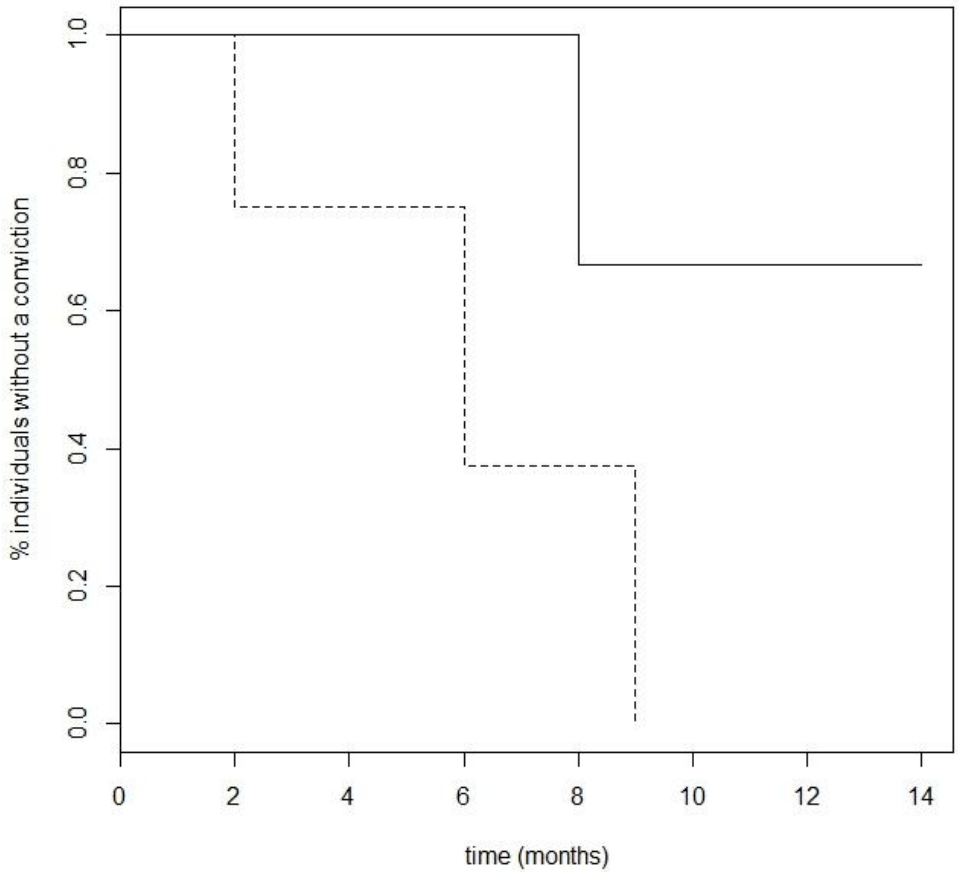

Figure S4

We next turn to the Cox regression analysis. We used the standard model

$$\log[\lambda\{t|X(t) = x\}] = \log\{\lambda_0(t)\} + \beta x \qquad (11)$$

This model assumes that the ratio of the conditional hazard functions is constant over time, and equal to $\exp(\beta)$. Let $X_{(j)}(t_j)$ be the medication level at time t_j for the individual whose conviction time is t_j. Let $R(t_j)$ be the risk set at time t_j, and let $X_k(t_j)$ be the medication level at time t_j for individual k in the risk set $R(t_j)$. We estimated β by maximizing the standard partial likelihood

$$\prod_{j=1}^{D} \frac{\exp\{\beta X_{(j)}(t_j)\}}{\sum_{k\in R(t_j)} \exp\{\beta X_k(t_j)\}} \qquad (12)$$

(Klein and Moeschberger, 1997). To account for the fact that sub-periods from the same individual may be correlated, we computed a robust ("sandwich") standard error for $\hat{\beta}$, treating each individual as a "cluster" (Lee et al, 1992).

With medication coded as 0 (no medication) and 1 (medication), the partial likelihood for our toy example equals

$$\frac{1}{4+2\exp(\beta)} \cdot \frac{1}{2+3\exp(\beta)} \cdot \frac{\beta}{1+3\exp(\beta)} \cdot \frac{1}{1+2\exp(\beta)} \qquad (13)$$

For instance, the first term in the likelihood corresponds to the conviction that occurred for individual A at time 2. At this time, A was not medicated so the numerator in the term becomes $\exp(\beta \cdot 0) = 1$. At time 2, the risk set consists of all sub-periods in Figure S2 that begin before time 2 and end after time 2, that is, all sub-periods except the bottommost left and the second topmost right. Four of these sub-periods are non-medicated, whereas two are medicated, so that the denominator in this term becomes equal to $4\exp(\beta \cdot 0) + 2\exp(\beta \cdot 1) = 4 + 2\exp(\beta)$. The partial likelihood is maximized at $\exp(\hat{\beta}) = 0.21$. Thus, for the data in our toy example, the estimated hazard function for those that are currently on medication is only 21% of the estimated hazard function for those that are currently off medication.

Both the extended Kaplan-Meier curves and the ordinary Cox regression are "unadjusted" analysis. Thus, the observed inverse association in these analyses between medication and conviction may be explained by confounding. A powerful way to reduce confounding is to let each individual "serve as his own control". That is, by comparing the conviction rate for the same individual under medicated and non-medicated periods. Such a "within-individual" analysis is free from confounding due to all factors that are constant within the individual across follow-up, e.g. genetic make-up and early childhood environment (Allison, 2009). For time-to-event data, within-individual analyses can be conveniently carried out with stratified Cox regression. Let $\lambda_i\{t|X(t) = x\}$ denote the conditional hazard at time t, given covariate level $X(t) = x$, at time t, for individual i. The stratified Cox regression analog to the model in equation (11) is given by

$$\log[\lambda_i\{t|X(t) = x\}] = \log\{\lambda_{i0}(t)\} + \beta x \qquad (14)$$

In equation (14), $\lambda_{i0}(t)$ is an individual-specific baseline hazard, and $\exp(\beta)$ is the within-individual hazard ratio. The standard way to estimate β is to maximize the partial likelihood, which is obtained by computing the likelihood in equation (12) for each individual separately, and multiplying the individual contributions (Klein and Moeschberger, 1997).

The partial likelihood for our toy example is given by

$$\frac{1}{3} \cdot \frac{1}{1+\exp(\beta)} \cdot \frac{\beta}{1+2\exp(\beta)} \cdot \frac{1}{1+\exp(\beta)} \qquad (15)$$

For instance, the first term in the likelihood corresponds to the conviction that occurred for individual A at time 2. At this time, A was not medicated so the numerator in the term becomes $\exp(\beta \cdot 0) = 1$. At

time 2, the risk set consists of all sub-periods in the **left** panel of Figure S2 that begins before time 2 and end after time 2, that is, all sub-periods except the bottommost. None of these three sub-periods are medicated, so that the denominator in this term becomes equal to 3. The partial likelihood is maximized at $\exp(\hat{\beta}) = 0.39$. Thus, for the data in our toy example, the estimated hazard function for those that are currently on medication is 39% of the estimated hazard function for those that are currently off medication.

In the analysis of our real data, we adjusted for age, previous number of convictions, and previous number of switches in medication levels, by adding these factors as categorical variables in the stratified Cox regression model.

All analyses were carried out with SAS Version 9.2 (Cary, NC. SAS Institute Inc, 2008), except for the calculations of the extended Kaplan-Meier curves for where package `survival` for R was used (Therneau, 2012).

References

Allison PD. (2009). *Fixed effects regression models.* Sage Publications, Thousand Oaks.

Breslow N, Crowley J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics* **3**, 437-453.

Klein JP, Moeschberger ML. (1997). *Survival Analysis: techniques for censored and truncated data.* Springer-Verlag, New York.

Lee EW, Wei LJ, Amato DA. (1992). Cox-type regression for large numbers of small groups of correlated failure time observations. *Survival Analysis: State of the Art*, JP Klein and PK Goel (ed.) Kluwer Academic: Dordrecht, pp. 237-247.

SAS Version 9.2, Cary, NC. SAS Institute Inc; 2008

Snappin MS, Jiang Q, Iglewicz B. (2005). Illustrating the impact of a time-varying covariate with an extended Kaplan-Meier estimator. The American Statistician **59**(4), 301-307.

Therneau T (2012). A package for survival analysis in S. R package version 2.36-12