From `%x_recepter` to `Rcpp`: a new user interface and an efficient algorithm

Helene Charlotte Rytgaard and Thomas Alexander Gerds

October 28, 2016

1 Introduction

The Danish National Prescription Registry (NCBI) provides individual-level data on all prescription drugs sold in Danish community pharmacies since 1994 (**?**). Effects and side effects of drugs can be assessed in a Danish nation wide registry studies. For this the prescription data are linked to the cause of death register and other Danish registries (**?**).

This document describes a complex algorithm for the computation of drug exposure strength and length relative to a prespecified study period [a,b]. Note that the start of the study period has to be after the start of the registry.

2 Drug prescription data

We describe the data of a single person who has purchased the drug of interest at K different dates in the study period [a,b]. The setup can easily be generalized to multiple individuals. The set of ordered drug purchase dates for one individual is denoted as

$$T_1 < \cdots < T_K.$$

One package of each drug product is defined by the drug strength S of the smallest unit (e.g., one pill or half a pill) and the amount of such units that it contains. For each drug we distinguish $J \geq 1$ different package types according to the different strengths: $S_1 < \cdots < S_J$. Note that the original SAS macro allowed at most 4 different package types. For each drug strength S_j the values $s_j^{\min}$, $s_j^{\max}$, s_j^* define the minimal, maximal and typical dose per day, respectively. Since $s_j^{\min}$ is the smallest dose possible, we must have $S_j = p_j s_j^{\min}$ for some integer $p_j \geq 1$. The number of smallest units, $s_j^{\min}$, of type j purchased on date T_k is denoted by n_{jk}. For each drug strength S_j we have $G_{jk} \geq 0$ purchases on date T_k. For $g = 1, \ldots, G_{jk}$, we let pn_{gk} denote the number of packages and ps_{gk} denote the package size of purchase g. Thus, the total amount D_k of the drug purchased on date T_k is given by the formula

$$D_k = \sum_{j=1}^{J} \left(\sum_{g=1}^{G_{jk}} \mathrm{pn}_{gk}\, \mathrm{ps}_{gk} \right) S_j = \sum_{j=1}^{J} \left(\sum_{g=1}^{G_{jk}} \mathrm{pn}_{gk}\, \mathrm{ps}_{gk} \right) p_j s_j^{\min} = \sum_{j=1}^{J} n_{jk} s_j^{\min},$$

such that $n_{jk} = \left(\sum_{g=1}^{G_{jk}} \mathrm{pn}_{gk}\, \mathrm{ps}_{gk} \right) p_j$.

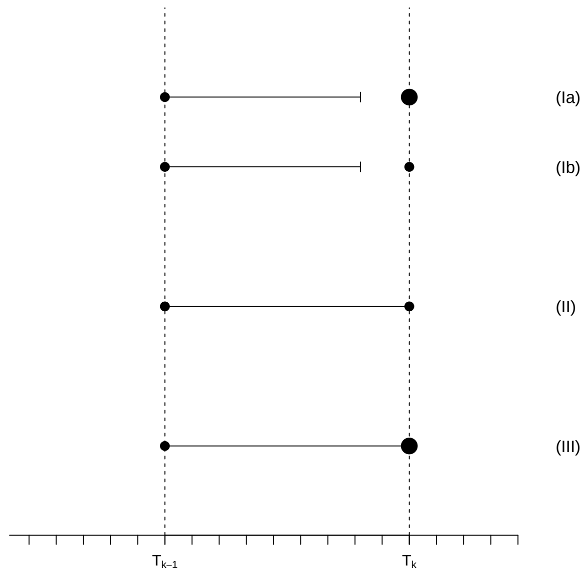

Figure 1: Illustration of the cases to be considered when computing the average daily dosis. The sizes of the circles refers to the dosis strength. That is, if the sizes of the circles are the same, the dosis strengths must be the same.

3 Hospital admission data

Hospitals usually deliver drugs for their patients. It is therefore hensigtsmaessig to take into account periods of hospitalization in the calculation of exposure lengths. For a single individual we define up to Q periods of hospitalization by the admission dates $L_1, \ldots, L_Q$ and the corresponding discharge dates $R_1, \ldots, R_Q$.

We compute the number of days hospitalized in the period $[T_k, T_{k+1})$ as:

$$A_k = \sum_{q=1}^{Q} \max\big(0,\, \min(T_{k+1}, R_q) - \max(T_k, L_q)\big).$$

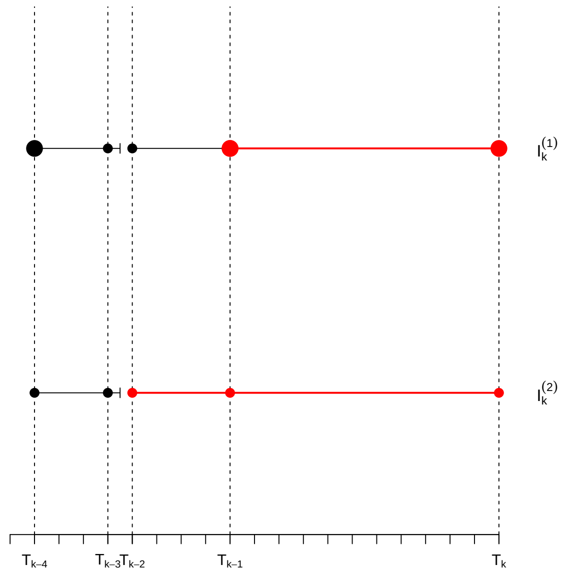

Figure 2: Illustration of the periods to use when computing the average daily dosis. The sizes of the circles refers to the dosis strength. That is, if the sizes of the circles are the same, the dosis strengths must be the same.

Accordingly the number of non-hospitalized days in $[T_k, T_{k+1})$ is:

$$H_k = (T_{k+1} - T_k) - A_k.$$

FIXME:

- need to limit to the study period [a,b]?
- is the day T_{k+1} included in $[T_k, T_{k+1})$or not?}[3cm]

4 Exposure strength and exposure lengths

The aim is to calculate the ends of the exposure periods E_k and for each exposure period the exposure strength per day X_k. The calculations of exposure are based on the drug prescription data and the hospitalization dates and depend further on an integer N that defines the number of prescription dates back in time to use in the calculations of exposure in a given period $[T_k, T_{k+1})$.

To express the exposure in period $[T_k, T_{k+1})$ we first note that based on the total drug purchase on date T_k the individual can be exposed at most $n_k = \sum_{j=1}^{J} n_{jk}$ days.

We use the following notation to indicate if the maximal number of exposure days exceeds the number of non-hospitalized days in period $[T_k, T_{k+1})$:

$$u_k = \begin{cases} 0, & n_k \leq H_k \\ 1, & n_k > H_k \end{cases}.$$

Furthermore, we use $S_{j(k)}$ to denote the nearest of the predefined pill strengths not exceeding the average strength computed for a period T_k, that is, the index $j(k)$ is defined as,

$$j(k) = \max \left\{ \ell \in \{1, \ldots, J\} \, : \, S_\ell \leq \frac{1}{c_k} \sum_{j=1}^{J} 1\{n_{jk} > 0\} \, S_j \right\}.$$

where $c_k = \sum_{j=1}^{J} 1\{n_{jk} > 0\} \, G_{jk}$ is the number of different drug strengths purchased on date T_k and $\frac{1}{c_k} \sum_{j=1}^{J} 1\{n_{jk} > 0\} \, S_j$ is the average strength for period T_k. Recall again, that c_k was the number of different drugs strengths purchased on date T_k, $c_k = \sum_{j=1}^{J} 1\{n_{jk} > 0\}$.

We distinguish between two periods to be used for calculating an average daily dose,

$$\begin{aligned} I_k^{(1)} = \big\{ \max \big(& \min\{\ell \in \{1, \ldots, J\} \, : \, u_\ell = \cdots = u_{k-1} = 1\}, \\ & \min\{\ell \in \{1, \ldots, J\} \, : \, S_{j(\ell)} = \cdots = S_{j(k)}\}\big), \ldots, k-1 \big\}, \end{aligned}$$

and

$$I_k^{(2)} = \big\{ \min\{\ell \in \{1, \ldots, J\} \, : \, u_\ell = \cdots = u_{k-1} = 1\}, \ldots, k-1 \big\},$$

Note that $I_k^{(2)}$ begins on the earliest date T_{i_0} from which the sequence of prescriptions up to time T_k provides a large enough amount of drugs to cover the corresponding number of days between T_{i_0} and T_k, whereas $I_k^{(1)}$ also takes into account that the average doses along the sequence must be the same. The periods $I_k^{(1)}$ and $I_k^{(2)}$ are illustrated in Figure 2. The average daily dose in the period $I_k^{(w)}$ is then defined as,

$$M_k^{(w)} = \frac{\sum_{\ell \in I_k^{(w)}} D_\ell}{\sum_{\ell \in I_k^{(w)}} H_\ell}, \qquad w = 1, 2.$$

At last, we define two binary variables to indicate whether the level of the calculated average dose for $w = 2$ is below/above the relevant minimal/maximal daily doses,

$$v_k^{\max} = 1 \left\{ M_k^{(2)} > s_{j(k)}^{\max} \right\}, \qquad v_k^{\min} = 1 \left\{ M_k^{(2)} < s_{j(k)}^{\min} \right\}.$$

4.0.1 Calculating the doses, $X_1, \ldots, X_K$

The average exposure strength X_k per day for period $[T_k, T_{k+1})$ is computed as follows. An illustration of the three different cases can be found in Figure 1.

$$\begin{aligned}
X_k = {} & (1 - u_{k-1})\, s^*_{j(k)} \\
& + u_{k-1}\, 1\{S_{j(k-1)} = S_{j(k)}\} \left(\min \left\{ \underset{p \in \mathbb{N}}{\operatorname{argmin}} \left| M_k^{(1)} - p \cdot s^{\min}_{j(k)} \right| \cdot s^{\min}_{j(k)} \right\} \right) \\
& + u_{k-1} 1\{S_{j(k-1)} \neq S_{j(k)}\} \left(v_k^{\max}\, s^{\max}_{j(k)} + v_k^{\min}\, s^{\min}_{j(k)} + (1 - v_k^{\max})(1 - v_k^{\min}) s^*_{j(k)} \right).
\end{aligned}$$

Note that $\left(\min \left\{ \underset{p \in \mathbb{N}}{\operatorname{argmin}} \left| M_k^{(1)} - p \cdot s^{\min}_{j(k)} \right| \cdot s^{\min}_{j(k)} \right\} \right)$ is the rounding of the average daily dose $M_k^{(1)}$ to the nearest factor of the relevant minimal dose $s^{\min}_{j(k)}$.

4.0.2 Calculating the end dates, $E_1, \ldots, E_k$

$$\begin{aligned}
E_k = \min \Bigg[& T_{k+1} - 1,\, (1 - u_k)\,(1 - u_{k-1}) \left(T_k - 1 + \operatorname{round}\left(\tfrac{D_k + R_k}{s_k^*} \right) \right) + \\
& (1 - (1 - u_k)\,(1 - u_{k-1})) \left(T_k - 1 + \operatorname{round}\left(\tfrac{D_k + R_k}{X_k} \right) \right) \Bigg]
\end{aligned}$$

4.0.3 Calculating the leftover dose, $R_1, \ldots, R_k$

$$R_k = \Big(D_{k-1} + R_{k-1} - X_{k-1}\,(E_{k-1} - T_{k-1}) \Big)\, u_k.$$

5 User interface

Error: could not find function "dpp" Error in period(obj) <- as.Date("1995-01-01", "2011-01-01") : object 'obj' not found Error: object 'recipe.db' not found Error: object 'lpr.db' not found Error: could not find function "atc" Error: could not find function "package" Error: could not find function "package" Error: could not find function "process"

5.1 Output

The output consists of:

- $B_1, \ldots, B_K$: Starting dates for each prescription period.
- $E_1, \ldots, E_K$: End dates for each prescription period.
- $X_1, \ldots, X_K$: Calculated dose for each prescription period.

References