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

Helene Charlotte Rytgaard and Thomas Alexander Gerds

October 18, 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: $S_1, \ldots, S_J$. 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. Note that the original SAS macro allowed at most 4 different package types. We combine all packages purchased on the same date that have the same strength. The number of smallest units, $s_j^{\min}$, of type j purchased on date T_k is denoted by $\tilde{n}_{jk}$. Thus, the total amount D_k of the drug purchased on date T_k is given by the formula

$$D_k = \frac{1}{\#J_k} \sum_{j=1}^{J} \tilde{n}_{jk} s_j^{\min} = \sum_{j=1}^{J} n_{jk} s_j^{\min},$$

where $\#J_k = \sum_{j=1}^{J} 1\{n_{jk} > 0\}$ is the number of different drugs purchased on date T_k. Note that $n_{jk} = \tilde{n}_{jk} \,/\, (\#J_k)$.
FIXME: does it actually make sense to divide with number of different drugs (as done in macro)?
FIXME: this way of defining the dosis does not fit with the calculation of dosis sequences… Figure 1 illustrates the data

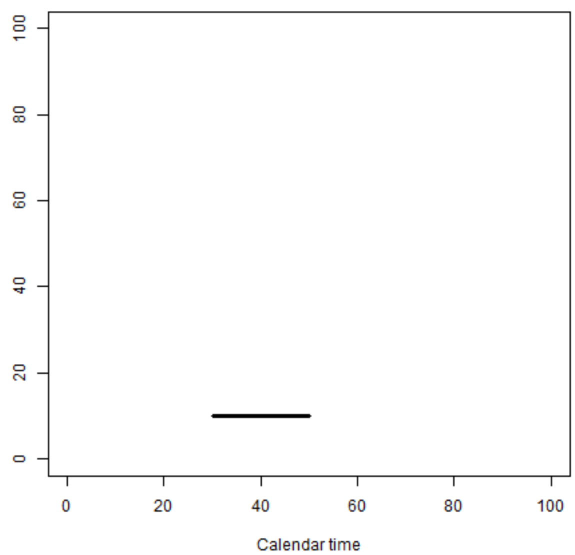

Figure 1: label til figure

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 $L_1, \dots, L_Q$ and the corresponding discharge dates $R_1, \dots, R_Q$.

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

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

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

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

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 the following parameters. An integer N 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.

FIXME: would it not be more logical to use $\tilde{n}_k = n_{jk}\frac{S_j}{s_j^{\min}}$ instead of n_k? Helene: Not relevant any longer.

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

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

We use $\bar{Y}_k$ to denote the history of a variable Y up to period k. For example,

$$\bar{D}_k = (D_k, D_{k-1}, \ldots, D_1)$$

is the history of total drug purchase until and including date T_k.

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

HELENE: we need to adapt the following formula to the new notations.

$$\begin{aligned} X_k = {} & (1 - u_k)\,(1 - u_{k-1})\, d_k^* \\ & + u_{k-1} \Big(1\{D_{k-1} = D_k\}\, z(\bar{S}_k, \bar{D}_k, \bar{C}_k) + 1\{D_{k-1} \neq D_k\}\, w(\bar{S}_k, \bar{D}_k, \bar{C}_k) \Big) \\ & + u_k\,(1 - u_{k-1})\, w(\bar{S}_k, \bar{D}_k, \bar{C}_k). \end{aligned}$$

Here, $z(\bar{S}_k, \bar{D}_k, \bar{C}_k)$ and $w(\bar{S}_k, \bar{D}_k, \bar{C}_k)$ are calculated from the mean dose

$$M_{jk}^{z,w} = \frac{\sum_{i \in I_k^{z,w}} n_{ij} s_j^{\min}}{\sum_{i \in I_k^{z,w}} C_i},$$

where

$$I_k^w = \big\{ \min\{i : u_i = \cdots = u_{k-1} = 1\}, \ldots, k-1 \big\},$$

and,

$$\begin{aligned} I_k^z = \big\{ \max\big(&\min\{i : u_i = \cdots = u_{k-1} = 1\}, \\ &\min\{i : g(s_i) = \cdots = g(s_k)\}\big), \ldots, k-1 \big\}, \end{aligned}$$

where

$$s_i = \frac{1}{\#J_i} \sum_{j=1}^{J} 1\{n_{ij} > 0\}\, S_j, \quad \text{and} \quad g(x) = \sum_{j=1}^{J-1} 1\{S_j \le x < S_{j+1}\}\, S_j,$$

according to

$$\begin{aligned} z(\bar{S}_k, \bar{D}_k, \bar{C}_k) &= \sum_{j=1}^{J} \text{round}\left(M_{jk}^z \,/\, s_j^{\min} \right) \cdot s_j^{\min}, \\ w(\bar{S}_k, \bar{D}_k, \bar{C}_k) &= \frac{1}{\#J_k} \sum_{j=1}^{J} \max\Big(1\{n_{jk} > 0\}\, s_j^{\min},\ 1\{M_{jk}^w > s_j^{\max}\}\, s_j^{\max} + 1\{M_{jk}^w < s_j^{\min}\}\, s_j^{\min} \\ &\quad + 1\{s_j^{\min} \le M_{jk}^w \le s_j^{\max}\}\, s_j^* \Big), \end{aligned}$$

where, again, $\#J_k = \sum_{j=1}^{J} 1\{n_{jk} > 0\}$ is the number of different drugs purchased on date T_k.

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

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

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

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.