# From `%x_recepter` to `Rcpp`: a mathematical formulation of the algorithm, a new R-user interface and an efficient implementation

Helene Charlotte Rytgaard and Thomas Alexander Gerds

November 1, 2016

## 1 Introduction

The Danish National Prescription Registry (NCBI) provides patient-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 study. For this the prescription data are linked to the cause of death register and other Danish registries (**?**).

This document describes the *medicin macro* by Christian Torp-Pedersen, a complex algorithm for the computation of drug exposure strength and length relative to a prespecified study period $[a, b]$ in calendar time. Note that the start of the study period has to be after the start of the registry in 1994.

## 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 patients. The set of ordered drug purchase dates for one patient is denoted as

$$T_1 < \cdots < T_K.$$

Here $K$ is the number of times the patient has purchased one or more packages of the same drug product in the period $[a, b]$. One package of each drug product is defined by the drug strength of the smallest unit (e.g., one pill or half a pill) and the amount of such units that it contains. For each drug product 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 but that the new implementation does not have this limitation. 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.

We now describe the total amount of drug purchased on date $T_k$ and the corresponding maximal number of days of supply. First, note that since $s_j^{\min}$ is the minimal dosis, one unit of strength $S_j$ can supply a maximal number of $(S_j/s_j^{\min})$ days. Next, consider that on date $T_k$ the patient purchases $G_{jk}$ many packages of strength $S_j$. If the patients purchases no package of strength j on date $T_k$ then $G_{jk} = 0$. Since each package may include a different number of units, the total amount of units of strength $S_j$ purchased on date $T_k$ is given by

$$m_{jk} = \sum_{g=1}^{G_{jk}} (\text{number of units in package } g)$$

We convert this number to the scale defined by smallest units and define the total amount of smallest units of strength $S_j$ purchased on date $T_k$ as

$$n_{jk} = m_{jk} \frac{S_j}{s_j^{\min}}.$$

The total amount $D_k$ of the drug purchased on date $T_k$ is given by the formula

$$D_k = \sum_{j=1}^{J} m_{jk} S_j = \sum_{j=1}^{J} n_{jk} s_j^{\min}.$$

The maximal number of days of supply based on $D_k$ is $n_k = \sum_{j=1}^{J} n_{jk}$.

Let us consider a small example of an individual $y$ with four purchase dates, $T_1, T_2, T_3, T_4$, and a drug "z" with two different strengths, $S_1 = 50$ and $S_2 = 80$. On the first date, individual $y$ buys two packages with drug strength $S_1$, the first one with 15 units and the other with 10 units, and one package with drug strength $S_2$ and 10 units. This means that

$$m_{1,1} = 15 + 10 = 25, \qquad m_{2,1} = 10.$$

We let the minimal doses corresponding to $S_1, S_2$ be $s_1^{\min} = 10, s_2^{\min} = 20$, such that

$$n_{1,1} = 25 \cdot \tfrac{50}{10} = 125, \qquad n_{2,1} = 10 \cdot \tfrac{80}{20} = 40.$$

Thus, the total amount of drug purchased on date $T_1$ is,

$$D_1 = 125 \cdot 10 + 40 \cdot 20 = 2050,$$

and the maximal number of days of supply is $n_1 = 125 + 40 = 165$.

# 3 Hospital admission data

Hospitals usually deliver drugs for their patients. It therefore seems reasonable to take into account periods of hospitalization in the calculation of exposure lengths. For a single patient 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 a patient is not hospitalized in the period $[T_k, T_{k+1})$ as:

$$H_k = (T_{k+1} - T_k) - \sum_{q=1}^{Q} \max\left(0,\, \min\left(T_{k+1}, R_q\right) - \max\left(T_k, L_q\right)\right)$$

We consider again individual $y$ from the previous section. Let us say that $y$ was hospitalized twice.

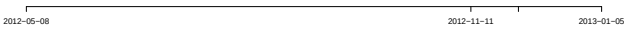


Figure 1: Illustration of the periods back in time to include into the estimate of the average daily dosis. The size of the dots indicates the second preliminary average strength $\mathrm{B_k}$. The red periods are included in two cases of the estimate of the average daily dosis in period $[T_k, T_{k+1})$. Which case to be used is determined by the cases in Figure 3, such that case (II) in Figure 3 uses case 1 and case (III) in Figure 3 uses case 2.

FIXME:

- what if $\mathrm{L_q}$ <a eller $\mathrm{R_q}$>b? limit to the study period [a,b]?
- should the day $T_{k+1}$ be included $[\mathrm{T_k},\, \mathrm{T_{k+1}}]$ or not $[T_k, T_{k+1})$?

# 4 Exposure strength and exposure lengths

The aim is to estimate the ends of the exposure periods $E_k$ and for each exposure period to estimate the exposure strength per day $X_k$. It is important to note that the estimates are only based on the data of the current patient and based on specific assumptions which may or may not be valid for a given patient and a given drug. The estimates are based on the drug prescription data (Section 2) and the hospitalization dates (Section 3) 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 $[T_k, T_{k+1})$.

## 4.1 Remark

The original SAS macro also uses prescription dates in the future to estimate the current exposure strength. However, since usually the aim is to use the exposure in Poisson and Cox regression where this would violate the mathematical framework the authors of this report hesitate to implement this feature. To motivate the feature we would very much like to see an example which demonstrates that the results of the Cox or Poisson regression can be improved when estimates of the current exposure depend on future purchases of the drug.

## 4.2 Definition of periods included in the estimates

To express the exposure in period $[T_k, T_{k+1})$ recall from section 2 that based on the total drug purchase on date $T_k$ the patient can be exposed at most $n_k = \sum_{j=1}^{J} n_{jk}$ days. We use the following notation to define potential overlap, i.e., to indicate if the maximal number of exposure days exceeds the number of non-hospitalized days in period $[T_k, T_{k+1})$:

$$u_k = \begin{cases} 0, & n_k \leq H_k, \text{ in words: } \textit{the supply at } T_k \textit{ is empty before } T_{k+1} \\ 1, & n_k > H_k, \text{ in words: } \textit{the supply at } T_k \textit{ can be sufficient to reach } T_{k+1}. \end{cases}$$

A first preliminary version of the average dosis per day in period $[T_k, T_{k+1})$ is calculated as

$$A_k = \frac{1}{c_k} \sum_{j=1}^{J} 1\{n_{jk} > 0\} S_j$$

where $c_k = \sum_{j=1}^{J} G_{jk}$ is the number of different drug strengths purchased on date $T_k$. Since the preliminary average $A_k$ may lie between two of the available drug strengths we define a second still preliminary version of the average dosis per day as the nearest drug strengths which does not exceed the average strength. That is, the index

$$j(k) = \max\{\ell \in \{1, \ldots, J\} \,:\, S_\ell \leq A_k\} \tag{1}$$

identifies the nearest drug strength which does not exceed the first preliminary average strength, and $S_{j(k)}$ is nearest drug strength. Note that in this notation, $S_{j(k-1)}$ refers to the nearest drug strength of the previous prescription date. On the following still quite long remaining part of the pilgrim trail towards the final estimate of the average daily dosis in period $[T_k, T_{k+1})$ the next thing to do is to decide how many purchase dates back in time should be used. We distinguish between two cases which are also illustrated in Figure 2.

**Case 1** $T_{I_k^{(1)}}$ *is the closest purchase date back in time, such that there is both continuous potential overlap and average dosis match. The index is defined as*

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

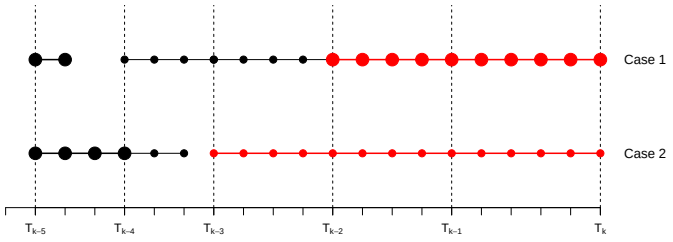


Figure 2: Illustration of the periods back in time to include into the estimate of the average daily dosis. The size of the dots indicates the second preliminary average strength $\mathrm{B_k}$. The red periods are included in two cases of the estimate of the average daily dosis in period $[T_k, T_{k+1})$. Which case to be used is determined by the cases in Figure 3, such that case (II) in Figure 3 uses case 1 and case (III) in Figure 3 uses case 2.

*The average daily dose in the period* $[T_{I_k^{(1)}}, T_{k+1})$ *is defined as*

$$M_k^{(1)} = \frac{\sum_{\ell=I_k^{(1)}}^{k-1} D_\ell}{\sum_{\ell=I_k^{(1)}}^{k-1} H_\ell}.$$

**Case 2:** $T_{I_k^{(2)}}$ *is the closest purchase date back in time, such that there is continuous potential overlap. The index is defined as*

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

*The average daily dose in the period* $[T_{I_k^{(2)}}, T_{k+1})$ *is defined as*

$$M_k^{(2)} = \frac{\sum_{\ell=I_k^{(2)}}^{k-1} D_\ell}{\sum_{\ell=I_k^{(2)}}^{k-1} H_\ell}.$$

At last, we define the rounding of the average daily dose $M_k^{(1)}$ to the nearest multiple of the minimal dose $s_{j(k)}^{\min}$ which corresponds to index $j(k)$ defined in equation (1) as

$$W_k = \max\left\{ \operatorname*{argmin}_{p\in\mathbb{N}} \left| M_k^{(1)} - p s_{j(k)}^{\min} \right| s_{j(k)}^{\min} \right\}.$$

### 4.2.1 Estimate of the daily dosis

The final estimate of the average daily dosis $X_k$ per day in period $[T_k, T_{k+1})$ is computed as follows, the computations are illustrated in Figure 3.

$$
\begin{aligned}
X_k = {} & (1 - u_{k-1})\, s^*_{j(k)} && \text{(No overlap)} \\
& + u_{k-1} 1\{S_{j(k-1)} = S_{j(k)}\} \Bigg[ && \text{(Overlap)} \\
& \quad W_k && \text{(II)} \\
& \quad + 1\left\{M_k^{(2)} > s^{\max}_{j(k)}\right\} s^{\max}_{j(k)} + 1\left\{M_k^{(2)} > s^{\min}_{j(k)}\right\} s^{\min}_{j(k)} \\
& \quad + 1\left\{M_k^{(2)} \le s^{\max}_{j(k)}\right\} 1\left\{M_k^{(2)} \le s^{\min}_{j(k)}\right\} s^*_{j(k)} \Bigg]. && \text{(III)}
\end{aligned}
$$

**Remark**: Note that the original SAS macro (even under the left-only option) also conditioned on the dosis at time $T_{k+1}$ but that we do not want to condition on the future until we are convinced by means of real examples that the potential damage (the mathematics of the Cox and Poisson regression are violated) can be counterbalanced by potential benefit.

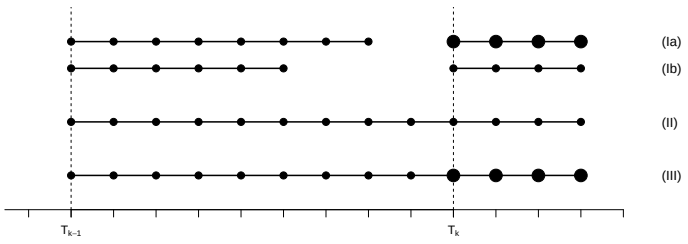


Figure 3: Illustration of the 4 cases with overlap that enter the estimate of the average daily dosis. The size of the dots indicates the second preliminary average strength $\mathrm{B_k}$.

### 4.2.2 Calculating the end dates, $E_1, \ldots, E_k$

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

### 4.2.3 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

work in progress

```
d <- dpp()
drug(d, firstdrugname) <- atc("a")
drug(d, firstdrugname, add=TRUE) <- atc("b")
drug(d, firstdrugname) <- pack(c(50, 75, 100, 125),
                   min = c(10, 50, 25, 50),
                   max = c(75, 200, 150, 150),
                   def = c(50, 100, 100, 125))
drugdb(d, add=FALSE, id=pnr) <- recept_data
admdb(d) <- datoer
period(d) <- c("2012-01-01", "2015-05-01")
N(d) <- 2
process(d,id=3000)
```

## 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

# 6