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

January 23, 2017

1 Introduction

The Danish National Prescription Registry (NCBI) provides patient-level data on all prescription drugs sold in Danish community pharmacies since 1994 (Kildemoes et al., 2011). 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 (Thygesen et al., 2011).

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 unique dates 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. 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

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

and the maximal number of days of supply based on D_k is

$$n_k = \sum_{j=1}^{J} n_{jk}.$$

2.1 Example

Consider a patient for who we have recorded seven purchase dates, $T_1, \ldots, T_7$, on which the patient purchased a drug in two different strengths, $S_1 = 50$ and $S_2 = 80$. On the first date, the patient buys two packages with drug strength $S_1 = 50$, the first one containing 15 pills and the other one containing 10 pills, and then also one package of another drug strength $S_2 = 80$ with 10 pills. 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 $s_1^{\min} = 25$ and $s_2^{\min} = 20$. This gives

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

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

$$D_1 = 50 \cdot 25 + 40 \cdot 20 = 2450,$$

and the maximal number of days of supply is $n_1 = 50 + 40 = 90$.

The following is a print of the data for this considered patient.

```
org(d$drugdb)
```

	id	atc	pdate	strength	npack	ppp
1	1	A07	2012-05-08	50	10	1
2	1	A07	2012-05-08	50	15	1
3	1	A07	2012-05-08	80	10	1
4	1	A07	2012-08-11	50	15	1
5	1	A07	2013-03-15	50	10	1
6	1	A07	2013-03-15	50	10	2
7	1	A07	2013-05-05	80	10	1
8	1	A07	2013-05-25	80	15	3
9	1	A07	2013-10-01	50	10	2
10	1	A07	2013-10-01	80	10	1
11	1	A07	2013-12-28	80	10	1

In the table below, we have the unique list of dates, `B`, the total doses, `D`, and the days of supply, `nk`, calculated according to Section 2.

```
org(ex$exdrug[, names(ex$exdrug) %in% c("B", "D", "nk")])
```

	B	D	nk
1	2012-05-08	2050	90
2	2012-08-11	750	30
3	2013-03-15	1500	60
4	2013-05-05	800	40
5	2013-05-25	3600	180
6	2013-10-01	1800	80
7	2013-12-28	800	40

Note that the i^{th} row of each of the columns `D` and `nk` refers to the period between the date in row i and row $i + 1$.

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:

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

3.1 Example (continued)

We consider again the patient of the example in section 2.1 and now also assume that the patient was hospitalized two times, i.e., $Q = 2$. See Figure 1 for an illustration of the prescription and admission dates.

```
org(d$admdb)
```

	id	inddto	uddto
1	1	2012-06-11	2012-06-28
2	1	2013-01-03	2013-01-18

We compute the number of days hospitalized, `DH`, and the number of days non-hospitalized, `H`.

```
org(ex$exdrug[, names(ex$exdrug) %in% c("B", "DH", "H")])
```

	B	H	DH
1	2012-05-08	78	17
2	2012-08-11	201	15
3	2013-03-15	51	0
4	2013-05-05	20	0
5	2013-05-25	129	0
6	2013-10-01	88	0
7	2013-12-28	1	0

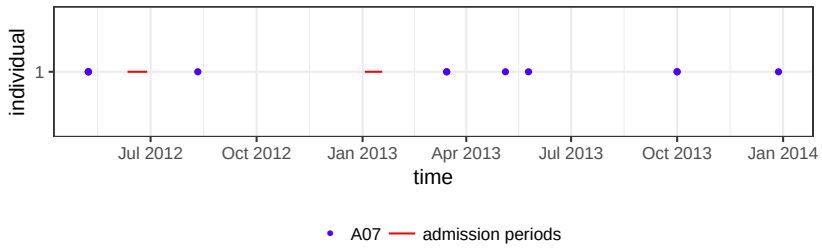

Figure 1: Illustration of the four prescription dates and the two periods of hospitalization of our example patient.

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 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 $[T_k, T_{k+1})$:

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

4.2.1 Example (continued)

Figure 2 shows again the data of section 2.1, together with an illustration of the number of days of supply. The black lines show the actual days of medicin supply, and the red lines are the days hospitalized. Note that if the concatenation of the black and red line in each period reaches the next date, then $u_k = 1$.

4.2.2 Preliminary average dose

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

$$A_k = \frac{1}{c_k} \sum_{j=1}^{J} G_{jk}\, S_j,$$

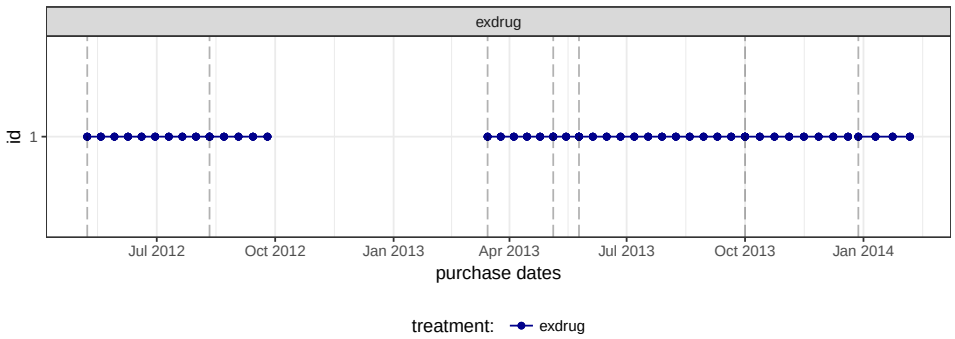

Figure 2: For our example patient the figure shows the maximal number of days of supply calculated at each time point based on the formula in Section 2. Also, the size of the bubbles indicates the value of $S_{b(k)}$, see Section 4.2.2.

where $c_k = \sum_{j=1}^{J} G_{jk}$ is the total number of purchases 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

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

identifies the nearest drug strength $S_{b(k)}$ which does not exceed the first preliminary average strength. Note that in this notation, $S_{b(k-1)}$ refers to the nearest drug strength of the previous prescription date.

4.2.3 Example (continued)

For the patient of our example we have

$$A_1 = \frac{1}{2+1}\,(2 \cdot 50 + 80) = 60.$$

We see that $b(1) = 1$, as $S_1 = 50$ is the nearest drug strength not exceeding the average of $A_1 = 60$ computed above.

4.2.4 Calculation of an average daily dose

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 3. Which case to be used will be made clear later.

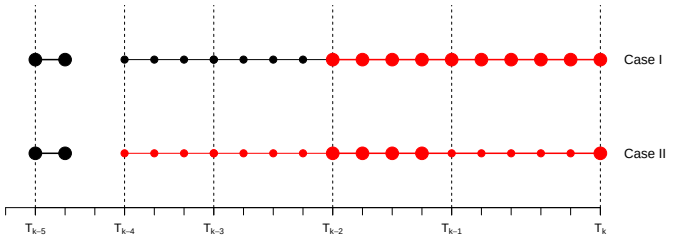

Figure 3: Illustration of the periods back in time to include into the final estimate of the average daily dosis at T_k. Shown are two independent examples illustrating Case 1 and Case 2, respectively. The size of the dots indicates the preliminary average strength. The red periods are included in the final estimate of the average daily dosis in period $[T_k, T_{k+1})$. See also Figure 4.

Case I: $T_{I_k^{(1)}}$ *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*

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

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 II: $T_{I_k^{(2)}}$ *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 $s_{b(k)}^{\min}$ (the index $b(k)$ defined in equation (1)) smaller than the average dose, that is

$$W_k = \left\lfloor \frac{M_k^{(1)}}{s_{b(k)}^{\min}} \right\rfloor s_{b(k)}^{\min}.$$

Note that $\lfloor x \rfloor$ denotes the largest integer value not exceeding x, that is, the downward rounded value.

4.2.5 Example (continued)

The following output shows the computed average daily doses for our example patient.

4.3 Final 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 cases for computations are illustrated in Figure 4.

$$
\begin{aligned}
X_k = {} & (1 - u_{k-1})\, s^*_{b(k)} && \text{(No overlap)} \\
& + u_{k-1} \Bigg[&& \text{(Overlap)} \\
& \quad 1\Big\{S_{b(k-1)} = S_{b(k)}\Big\} \Bigg(1\left\{W_k > s^{\max}_{b(k)}\right\} s^{\max}_{b(k)} \\
& \quad + 1\left\{W_k < s^{\min}_{b(k)}\right\} s^{\min}_{b(k)} && \text{(I)} \\
& \quad + 1\left\{W_k \le s^{\max}_{b(k)}\right\} 1\left\{W_k \ge s^{\min}_{b(k)}\right\} W_k \Bigg) \Bigg]. \\
& \quad + 1\Big\{S_{b(k-1)} \neq S_{b(k)}\Big\} \Bigg(1\left\{M^{(2)}_k > s^{\max}_{b(k)}\right\} s^{\max}_{b(k)} \\
& \quad + 1\left\{M^{(2)}_k > s^{\min}_{b(k)}\right\} s^{\min}_{b(k)} && \text{(II)} \\
& \quad + 1\left\{M^{(2)}_k \le s^{\max}_{b(k)}\right\} 1\left\{M^{(2)}_k \ge s^{\min}_{b(k)}\right\} s^*_{b(k)} \Bigg) \Bigg].
\end{aligned}
$$

4.3.1 Example (continued)

The following table shows the final estimated daily doses for the example individual.

```
org(ex$exdrug[, names(ex$exdrug) %in% c("B", "A", "Sjk", "M", "X")])
```

	X	B	M	A	Sjk
1	50	2012-05-08	50.00000	60	50
2	25	2012-08-11	26.28205	50	50
3	50	2013-03-15	50.00000	50	50
4	60	2013-05-05	29.41176	80	80
5	40	2013-05-25	40.00000	80	80
6	50	2013-10-01	29.50000	65	50
7	60	2013-12-28	60.00000	80	80

We here describe in detail how $X_2 = 25$ was computed.

On date T_2, the individual bought one package with 15 pills of strength $S_1 = 50$. From Figure 2 we see that the dose from date T_1 reaches T_2 and that $S_{b(1)} = S_{b(2)}$. This means that we are in case (I). We thus compute X_2 as,

$$M_2^{(1)} = \frac{D_1}{H_1} = \frac{2050}{78} \approx 26.28.$$

This value is then rounded to the nearest multiple of the minimal corresponding dosis $s^{\min}_{b(2)} = 25$ and hence $X_2 = W_2 = 1 \cdot s^{\min}_{b(2)} = 25$.

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.

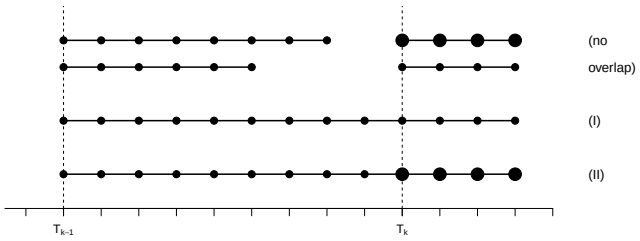

Figure 4: Illustration of the formula for the final estimate of the daily dosis (section 4.3). The size of the dots indicates the preliminary average strength $S_{b(k)}$. The upper most two lines illustrate the cases without overlap and the other two lines the cases with overlap.

4.4 Calculating the leftover doses, $R_1, \ldots, R_k$

The leftover doses are computed as,

$$R_k = u_{k-1} \cdot \min \Big[\text{maxdepot}, \max \Big\{ 0,\, D_{k-1} + R_{k-1} - X_{k-1} \Big(E_{k-1} - T_{k-1} - \sum_{q=1}^{Q} \max\left(0,\, \min\left(T_{k+1}, R_q\right) - \max\left(T_k, L_q\right)\right) \Big) \Big\} \Big],$$

where maxdepot is some user-specified maximum amount of dosis to be "stored" from one prescription date to the next, and

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

is again the number of hospitalized days in the period.

4.5 Calculating the dates of end of exposure, $E_1, \ldots, E_k$

Note that since $D_k + R_k$ is the total prescription dose at T_k and X_k is the estimated dose per day the result of $\frac{D_k+R_k}{X_k}$ has days as unit. However, the result is not necessarily an integer, which is why we round off to the largest integer value not exceeding the decimal value. The estimated end of the exposure period in which the daily dose is estimated as X_k is,

$$E_k = \min\left(T_{k+1} - 1,\, T_k - 1 + \left\lfloor \frac{D_k + R_k}{X_k} \right\rfloor\right).$$

4.5.1 Example (continued)

At last, we show here the full set of estimates for the example considered throughout this report.

```
org(ex$exdrug[, names(ex$exdrug) %in% c("id", "B", "E", "X", "R")])
```

	id	X	B	E	R
1	1	50	2012-05-08	2012-06-17	0
2	1	25	2012-08-11	2012-09-09	0
3	1	50	2013-03-15	2013-04-13	10
4	1	60	2013-05-05	2013-05-18	10
5	1	40	2013-05-25	2013-08-22	10
6	1	50	2013-10-01	2013-11-05	0
7	1	60	2013-12-28	2014-01-09	0

4.6 Final concatenation of treatment periods

The last step of the estimation of treatment periods concatenates treatment periods with same final dose. Furthermore, if there is a gap between two periods, a period filling out this gap is defined and the final dose here is set to 0. That way, a period will always begin after the last day of the previous period.

4.6.1 Example (continued)

The following table shows the estimated estimated periods and doses for the example individual.

```
org(ex1$exdrug)
```

	id	X	B	E
1	1	25	2012-08-11	2012-09-09
2	1	0	2012-09-10	2013-03-14
3	1	50	2013-03-15	2013-04-13
4	1	0	2013-04-14	2013-05-04
5	1	60	2013-05-05	2013-05-18
6	1	0	2013-05-19	2013-05-24
7	1	40	2013-05-25	2013-08-22
8	1	0	2013-08-23	2013-09-30
9	1	50	2013-10-01	2013-11-05
10	1	0	2013-11-06	2013-12-27
11	1	60	2013-12-28	2014-01-09

5 User interface

In the following, we show how to use the implemented R-interface. Firstly, the package is loaded and an empty object `d` is created.

```
library(heaven)
d <- dpp()
```

For this example, we consider simulated data. The following code generates data using functions supplied in the `heaven` package.

```
set.seed(8)
drugdata <- simPrescriptionData(10, startDate = "2006-01-01")
drugdb(d) <- drugdata
org(head(d$drugdb))
```

id	atc	pdate	strength	npack	ppp
1	A07	2007-06-16	400	1	30
1	A07	2008-04-14	500	1	60
1	A07	2009-03-23	400	2	30
1	A07	2009-11-05	200	1	300
1	A07	2010-08-30	400	1	100
1	A12B	2008-08-21	750	1	500

```
admdata <- rbind(simAdmissionData(10, startDate = "2006-01-01"))
admdb(d) <- admdata
org(head(d$admdb))
```

id	inddto	uddto
1	2014-12-20	2014-12-24
1	2017-07-20	2017-09-01
1	2018-04-30	2018-05-15
1	2021-11-27	2022-01-02
1	2024-05-11	2024-05-16
2	2015-01-20	2015-03-03

Then we add treatments to the object.

```
drug(d, ex1) <- atc("A12B")
drug(d, ex2) <- atc("A07")
drug(d, ex1) <- atc(c("A12B"))
drug(d, ex2) <- atc(c("A07"))
drug(d, ex1) <- pack(c(750, 75),
             min = c(250, 25),
             max = c(1000, 100),
             def = c(750, 100))
drug(d, ex2) <- pack(c(200, 400, 500),
             min = c(100, 100, 250),
             max = c(400, 500, 1000),
             def = c(300, 200, 500))
```

Note how the default, minimum and maximum doses are specified for each treatment. In the above, we have two treatments, one named "ex1" and one named "ex2".

We may plot the data for any of the drugs and any of the individuals in the data.

We decide to use 3 periods back in time to calculate the mean doses.

```
N(d) <- 3
```

Then we perform the calculations by calling the function `process()` on the object.

```
ex <- process(d, out=FALSE)
```

Error in order(dat$pdate) : argument 1 is not a vector
This produces an output data set for each of the specified treatments, `ex1` and `ex2`.

```
ex1out <- ex$ex1
ex2out <- ex$ex2
org(head(ex1out[, names(ex1out) %in% c("id", "X", "B", "E", "R")]))
```

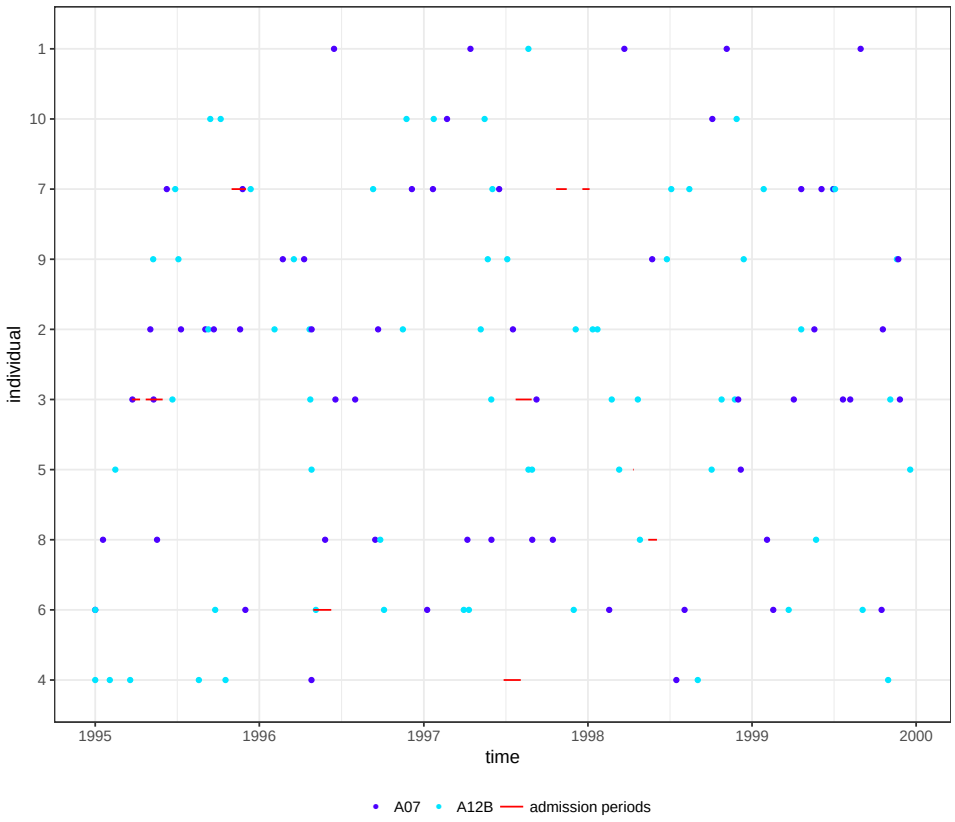

Figure 5: Illustration of the prescription dates and the periods of hospitalization of one our simulated example patients.

	id	X	B	E	R
1	1	100	2008-08-21	2018-11-26	0
2	2	100	2006-09-09	2007-02-02	0
3	2	7650	2007-02-03	2007-04-21	10
4	2	8325	2007-04-22	2007-05-14	10
5	2	4775	2007-11-16	2008-05-06	10
6	2	4500	2008-05-07	2008-07-28	10

For any individual and any of the treatments we may also plot the periods to use for calculations, as was seen in Figure 2.

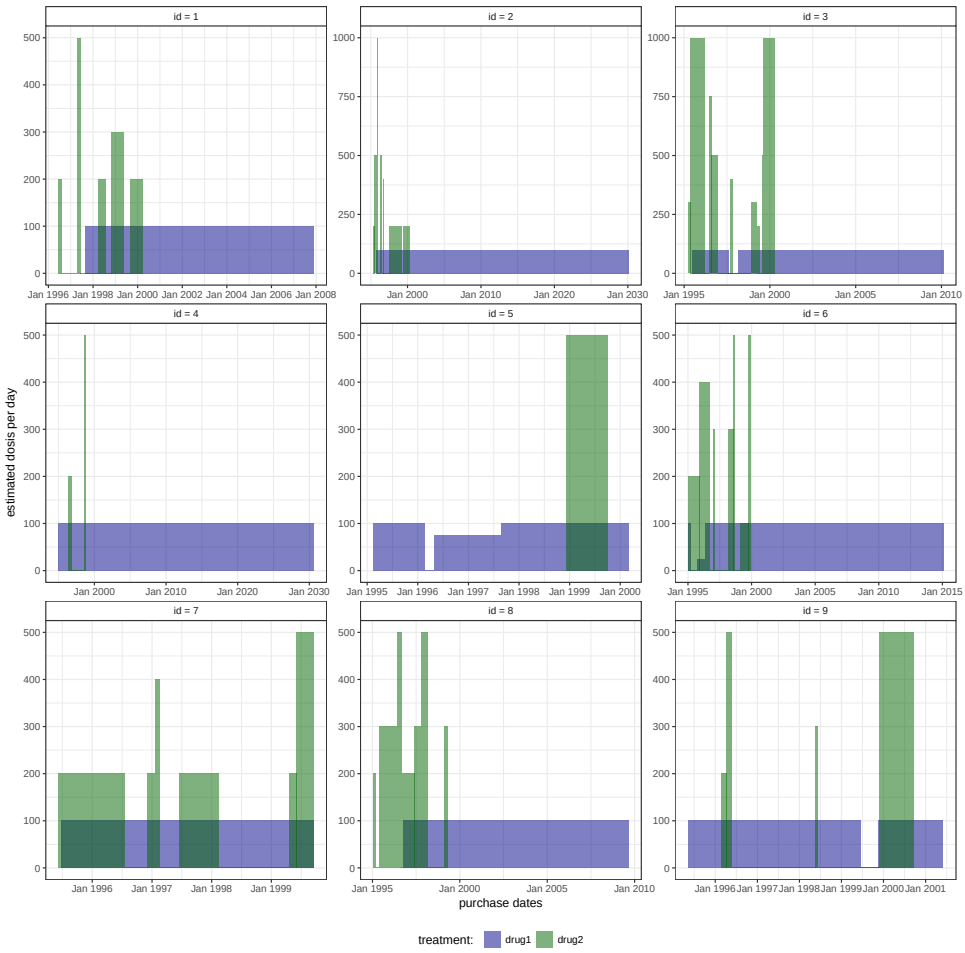

Figure 6: Days of supply and periods to be used to calculate mean doses for some of the individuals in the simulated data.

References

Kildemoes, H. W., H. T. Sørensen, and J. Hallas (2011). The Danish national prescription registry. *Scandinavian journal of public health 39*(7 suppl), 38–41.

Thygesen, L. C., C. Daasnes, I. Thaulow, and H. Brønnum-Hansen (2011). Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. *Scandinavian Journal of Public Health 39*(7 suppl), 12–16.