THE medicine macro

Anders Munch

October 8, 2019

1 Introduction

Something something...

2 Formulas

2.1 The data and user input

The data consists of N rows of observations of individual purchases for individual patients. For each patient and each drug purchase, the ATC and strength of the drug is recorded along with the number of units in the purchased packs and the number of packs purchased. We also need information about the hospitalization periods for each individual, summarized as a sequence of admission and discharge dates. All the information in the data is summarized in table **??**; the abbreviations are the same as the ones used in DST. The information about the hospitalization are displayed below the dashed line because they are given separately to the program and are indexed differently from the rest of the data.

Besides the data the algorithm also needs some user-specified values. First of all we have to specify which period and which ATC values we want to calculate the individual exposure times for. We also have to tell the program what the minimal, typical (default), and maximal doses are for each of the possible drug strengths. The possible drug strengths are in turn determined by the chosen ATC's. Finally, we have to specify two slightly more technical values: A cap on the amount of drug a patient can store, and how many prescription dates back in time we want to use in our estimation. Thus, the needed user-specified information is summarized in **??**. Of course, this information is common for all the individual patients. Note that both the number A of specified ATC's and the number J of drug strengths (and corresponding doses) are given by the user, but that all possible doses arising from the specified ATC's should be present among the J drug strengths; if not, the program will fail.

The exposures are calculated individually so we will now consider data for one isolated individual. Using the user-specified values for the ATC's to use and the period to consider, we get N_i purchases for each individual i. Ordering the purchase dates for this individual, we can summarize the data for individual i

Table 1: The data needed for the algorithm.

All the data		
ID	:	$\text{pnr}_1, \text{pnr}_2, \ldots, \text{pnr}_N$
ATC	:	$\text{atc}_1, \text{atc}_2, \ldots, \text{atc}_N$
Purchase dates	:	$\text{eksd}_1, \text{eksd}_2, \ldots, \text{eksd}_N$
Drug strength	:	$\text{strnum}_1, \text{strnum}_2, \ldots, \text{strnum}_N$
Number of units in pack	:	$\text{packsize}_1, \text{packsize}_2, \ldots, \text{packsize}_N$
Number of packs	:	$\text{apk}_1, \text{apk}_2, \ldots, \text{apk}_N$
ID	:	$\text{pnr}_1, \text{pnr}_2, \ldots, \text{pnr}_{\tilde{N}}$
Admission dates	:	$\text{inddto}_1, \text{inddto}_2, \ldots, \text{inddto}_{\tilde{N}}$
Discharge dates	:	$\text{uddto}_1, \text{uddto}_2, \ldots, \text{uddto}_{\tilde{N}}$

Table 2: The information needed from the user.

User input		
ATC's to use	:	$a_1, a_2, \ldots, a_A$
Highest amount of stored drugs	:	$md \in \mathbb{R}$
Prescription window	:	$pw \in \mathbb{N}$
Period	:	$[d_{\text{start}}, d_{\text{end}}]$
Drug doses	:	(see table below)

Strength	Minumum	Default	Maximum
s_1	$\min(s_1)$	$\text{def}(s_1)$	$\max(s_1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
s_J	$\min(s_J)$	$\text{def}(s_J)$	$\max(s_J)$

as in table **??**. Note that one individual can still make several purchases on the same date (for example, this would be the case if an individual purchases a drug of different strengths on the same date). Again, we display the hospitalization information below a dashed line because this data is on a different format (for example, a patient might not have been hospitalized at all, and thus the last row would be empty).

From these data and user-specified input we want to calculate individual exposure period. The output we get out will be on the form shown in table **??**. For each individual i we get a collection of G_i exposure periods, which are reported by a start date (SD) and end date (ED) together with the (estimated) daily dose and the length of the period. The G_i periods cover the whole time span specified by the user in table **??**; therefore, some of the daily doses might be 0, reflecting a period with no exposure.

Table 3: The data for a single individual.

Data for subject i		
ID	:	$\text{pnr}_1 = \text{pnr}_2 = \cdots = \text{pnr}_{N_i}$
ATC	:	$\text{atc}_1, \text{atc}_2, \ldots, \text{atc}_{N_i} \in \{a_1, a_2, \ldots, a_A\}$
Purchase dates	:	$d_{\text{start}} \leq \text{eksd}_1 \leq \text{eksd}_2 \leq \cdots \leq \text{eksd}_{N_i} \leq d_{\text{end}}$
Drug strength	:	$\text{strnum}_1, \text{strnum}_2, \ldots, \text{strnum}_{N_i}$
Number of units in pack	:	$\text{packsize}_1, \text{packsize}_2, \ldots, \text{packsize}_{N_i}$
Number of packs	:	$\text{apk}_1, \text{apk}_2, \ldots, \text{apk}_{N_i}$
Hospitalization	:	$d_{\text{start}} \leq \text{inddto}_1 \leq \text{uddto}_1 \leq \text{inddto}_2 \leq \cdots \leq \text{uddto}_{\tilde{N}_i} \leq d_{\text{end}}$

Table 4: The output for a single individual.

Output for individual i		
Start dates	:	$SD_1 < SD_2 < \cdots < SD_{G_i}$
End dates	:	$ED_1 < ED_2 < \cdots < ED_{G_i},\ ED_l = SD_{l+1}$
Daily dose	:	$X_1, X_2, \cdots, X_{G_i},\ X_l \neq X_{l+1}$
Number of exposed day	:	$\text{Days}_1, \text{Days}_2, \cdots, \text{Days}_{G_i}$

2.2 Estimating the daily dose

To find the daily dose and the exposure periods for an individual i, we first find the unique K_i purchase dates for this individual, which we denote by T_k. For each of these dates we then collect the total amount of drugs purchased on that date in the variables D_k. We also calculate the maximal number of days of supply by normalizing the drug strength with their corresponding minimal daily dose, and denote this quantity by M_k. In formulas we have:

$$\begin{aligned}
T_1 < T_2 < \cdots < T_{K_i}, \quad & T_k \in \{\text{eksd}_1, \ldots, \text{eksd}_{N_i}\}, \\
D_1, D_2, \ldots, D_{K_i}, \quad & D_k := \sum_{l\colon \text{eksd}_l = T_k} \text{apk}_l \cdot \text{packsize}_l \cdot \text{strnum}_l, \\
M_1, M_2, \ldots, M_{K_i}, \quad & M_k := \sum_{l\colon \text{eksd}_l = T_k} \text{apk}_l \cdot \text{packsize}_l \cdot \frac{\text{strnum}_l}{\min(s_l)}.
\end{aligned}$$

We also calculate two intermediate values for the final estimation: We calculate A_k as the average strength purchase on date T_k, and use this quantity to estimate the nearest *possible* drug strength, $\hat{S}_k$, for the period $[T_k, T_{k+1})$. That is,

$$\begin{aligned}
A_1, A_2, \ldots, A_{K_i}, \quad & A_k := \frac{1}{\#\{l \mid \text{eksd}_l = T_k\}} \sum_{l\colon \text{eksd}_l = T_k} \text{strnum}_l, \\
\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_{K_i}, \quad & \hat{S}_k := \max\{s_1, \ldots, s_J \mid s_j \leq A_k\}.
\end{aligned}$$

From the admission data, the number of non-hospitalized days in the period $[T_k, T_{k+1})$ are easily calculated. We denote these as

$$H_1, H_2, \dots, H_{K_i}.$$

In the following we use $\lfloor x \rfloor$ to denote the rounded down value of x, i.e., the largest integer smaller than x. Similarly we shall use the notation

$$[x]_a := \begin{cases} a, & x < a \\ x, & x \geq a \end{cases}, \quad [x]^b := \begin{cases} b, & x > b \\ x, & x \leq b \end{cases}, \quad [x]_a^b := \begin{cases} b, & x > b \\ a, & x < a \\ x, & a \leq x \leq b. \end{cases}$$

This simply means that, e.g., $[x]^b$ returns x whenever x is smaller than b, and caps the value to b if x is larger than b.

With the above quantities defined we can calculate the estimate of the end of the exposure periods (ED_k) and the daily dose (X_k). These estimates are calculated recursively over the K_i unique purchase dates, and depend on two auxiliary variables: A variable, Reach_k, indicating whether or not the supply of drugs at date T_k is enough to cover the time until the next purchase date T_{k+1}; and a variable, Stash_k, giving the amount of leftover drug from the previous purchases. In formulas, we can write these as:

$$\begin{aligned} 0 = \text{Stash}_1, \text{Stash}_2, \dots, \text{Stash}_{K_i}, &\quad \text{Stash}_k := [\text{stash}_{k-1} + D_{k-1} - \text{cons}_{k-1}]^{md} \\ \text{Reach}_1, \text{Reach}_2, \dots, \text{Reach}_{K_i - 1}, &\quad \text{Reach}_k := 1_{\{\text{stash}_k + M_k \geq H_k\}}, \end{aligned}$$

where cons_k is simply defined as the amount of drugs consumed in period $[T_k, T_{k+1})$, which is calculated as

$$\text{cons}_k := X_k(ED_k - SD_k - H_k).$$

Remember that the value md, by which we cap the value in the calculation of the stash, denotes the user-specified, assumed maximum amount of stored drugs (see table **??**).

Finally, we are ready to calculate the estimated daily dose. The calculation is divided into 3 different cases, depending on how much information (we think) we can use from the previous purchases. The idea is that if the drug supply from one purchase date is enough to reach the next, then we assume that the purchases are related to the same treatment, and thus we can use both periods to estimate the daily dose; exactly how we calculate the dose then again depends on whether or not the drug strength seems to change from one period to the next. Concretely, the 3 cases are:

1. The drug supply from the last purchase date does *not* reach this period, so we have no useful information from previous purchases.

2. The drug supply from the last purchase date *does* reach this period, *but* the drug strength seems to change; thus we have some information to use, but this can only be used partially as there seems to be some change in the level of treatment.

3. The drug supply from the last purchase date *does* reach this period, *and* the drug strength seems to stay the same; thus we guess that the patient is continuing with exactly the same kind and level of treatment, and we are therefore more confident in using the previous purchases in the estimation.

The exact calculations for the three cases go as follows and depend on the preliminary values we have calculated above in this section and the user-specified values from table **??**.

1. No information: In formulas, this means that $\text{Reach}_{k-1} = 0$. In this case, we simply put
$$X_k := \text{def}(\hat{S}_k),$$
because we do not have any better guess.

2. Some information: In formulas, this means that $\text{Reach}_{k-1} = 1$ but $\hat{S}_k \neq \hat{S}_{k-1}$. In this case, we use the values as long back in time as we have a continuous supply and until we reach the allowed, user-specified value of maxium number of previous prescriptions to use, pw. For this, we first calculate
$$X'_k := \frac{\sum_{l=I_k}^{k-1} D_l}{\sum_{l=I_k}^{k-1} H_l}, \quad \text{where } I_k := [\min\{l \leq k-1 \mid \text{Reach}_l = 1\}]_{pw},$$
and then put
$$X_k := \begin{cases} \min(\hat{S}_k), & X'_k < \min(\hat{S}_k) \\ \max(\hat{S}_k), & X'_k > \max(\hat{S}_k) \\ \text{def}(\hat{S}_k), & \min(\hat{S}_k) \leq X'_k \leq \max(\hat{S}_k) \end{cases}.$$

3. Most information: In formulas, this means that $\text{Reach}_{k-1} = 1$ and $\hat{S}_k = \hat{S}_{k-1}$. In this case, we use a similar approach but now also demand that the values used back in time have the same estimated drug strengths. We thus first calculate
$$X'_k := \frac{\sum_{l=\tilde{I}_k}^{k-1} D_l}{\sum_{l=\tilde{I}_k}^{k-1} H_l}, \quad \tilde{I}_k := \left[\min\{l \leq k-1 \mid \text{Reach}_l = 1, \hat{S}_l = \hat{S}_k\}\right]_{pw},$$
and then normalize this quantity to the minimal daily dose scale
$$X''_k := \left\lfloor \frac{X'_k}{\min(\hat{S}_k)} \right\rfloor \min(\hat{S}_k).$$
We put the final estimate equal to
$$X_k := [X''_k]_{\min(\hat{S}_k)}^{\max(\hat{S}_k)} = \begin{cases} \min(\hat{S}_k), & X''_k < \min(\hat{S}_k) \\ \max(\hat{S}_k), & X''_k > \max(\hat{S}_k) \\ X''_k, & \min(\hat{S}_k) \leq X''_k \leq \max(\hat{S}_k) \end{cases}.$$

Note the difference between case 2 and 3: In both cases our estimate is a simply sum over the amount of drugs purchased in a suitable time span, normalized by the number of days the patient has needed to supply himself in this time span. Besides the difference in how these times spans are defined, we also have that in case 2, the final estimate can only take on one of the three possible values of either maximal, minimal or typical dose for the preliminary estimated drug strength $\hat{S}_k$, while in case 3 we are confident enough to let the final estimate take on values that lie within the minimal and maximal dose, but might be different from the typical dose.

With the estimated daily dose in hand, the end of the k'th exposure period is simply calculated as

$$ED_k := \left[T_k + \left\lfloor \frac{D_k + \text{Stash}_k}{X_k} \right\rfloor\right]^{T_{k+1}},$$

that is, we just use the estimated dose, X_k, for the period to normalize the stash and the amount of drug purchased.

As a final step we concatenate periods with the same estimated daily dose. That is, if $ED_l = SD_{l+1}$ and $X_l = X_{l+1}$, we join the periods $[SD_l, ED_l)$ and $[SD_{l+1}, ED_{l+1})$ to one period $[SD_l, ED_{l+1})$ with daily exposure X_l. Also, if we have gaps between the periods, meaning $ED_l < SD_{l+1}$, we define a new period $[ED_l, SD_{l+1})$ with estimated daily exposure 0. This shapes the output into the form of **??**.

3 Examples

3.1 One purchase

The simplest setting is one individual making one purchse. The calculations are then straightforward. We can specify the data as follows.

```
lmdb0 <- data.table(pnr=8,
            eksd=as.Date("2005-01-25"),
            apk=1,
            atc="C10AA01",
            strnum=10,
            packsize=3)
```

This gives data of the form:

pnr	eksd	apk	atc	strnum	packsize
8	2005-01-25	1	C10AA01	10	3

We then give the user specified information corresponding to table **??** as follows.

```
user <- list(atc="C10AA01",
         maxdepot=8000,
```

```
         period=as.Date(c("2004-01-01","2015-12-31")),
         prescriptionwindow=2,
         doses=list(value=c(10,20,30),
            min=c(5,10,15),
            max=c(20,40,60),
            def=c(10,20,30)))
```

We can then plug this into the program. For now, we do not have any admission data so we set this equal to NULL.

```
x <- medicinMacro(drugs=list(C10AA01=user),drugdb=lmdb0,admdb=NULL)
```

The results are

pnr	dose	firstday	lastday	exposure.days
8	10	2005-01-25	2005-01-28	3 days

3.2 Two purchases

A more complicated example is one individual making two purchases. We consider examples with and without overlap.

3.2.1 Two purchases without overlap

The data and user specified information is given in the code below. The form of the data and the estimation output is in the table below.

```
lmdb1 <- data.table(pnr=c(8,8),
            eksd=as.Date(c("2005-01-16","2005-01-23")),
            apk=c(1,1),
            atc=c("C10AA01", "C10AA01"),
            strnum=c(10,20),
            packsize=c(5,5))
simva1 <- list(atc="C10AA01",maxdepot=8000,
           period=as.Date(c("2004-01-01","2015-12-31")),
           maxdepot = 10000,
           prescriptionwindow=2,
           doses=list(value=c(10,20,30),min=c(10,10,15),max=c
    (10,40,60),def=c(10,20,30)))
org("Purchase data:")
org(lmdb1)
org("Results from the progam:")
org(medicinMacro(drugs=list(simva1=simva1),drugdb=lmdb1,admdb=NULL)$
    simva1)
```

Purchase data:

pnr	eksd	apk	atc	strnum	packsize
8	2005-01-16	1	C10AA01	10	5
8	2005-01-23	1	C10AA01	20	5

Results from the progam:

pnr	dose	firstday	lastday	exposure.days
8	10	2005-01-16	2005-01-21	5 days
8	0	2005-01-21	2005-01-23	2 days
8	20	2005-01-23	2005-01-28	5 days

3.2.2 Two purchases with overlap

Here, we change the second date to introduce an overlap. This changes the total exposure time from 10 to 9 days.

```
lmdb1$eksd[2] <- as.Date("2005-01-19")
org("Purchase data:")
org(lmdb1)
org("Results from the progam:")
org(medicinMacro(drugs=list(simva1=simva1),drugdb=lmdb1,admdb=NULL)$
    simva1)
```

Purchase data:

pnr	eksd	apk	atc	strnum	packsize
8	2005-01-16	1	C10AA01	10	5
8	2005-01-19	1	C10AA01	20	5

Results from the progam:

pnr	dose	firstday	lastday	exposure.days
8	10	2005-01-16	2005-01-19	3 days
8	20	2005-01-19	2005-01-25	6 days

3.3 Comparing SAS-macro with R-macro

We use a slightly more involved example to compare the output of the R-program with the old SAS-macro. We have the following data and user input.

```
lmdb <- data.table(pnr=rep(8,5),
           eksd=as.Date(c("2005-01-25","2005-03-03","2006-01-11","
    2006-04-15","2006-07-31")),
           apk=rep(1,5),
           atc=rep("C10AA01",5),
           strnum=c(10,10,10,10,20),
           packsize=rep(100,5))
simva <- list(atc="C10AA01",maxdepot=8000,
          period=as.Date(c("2004-01-01","2015-12-31")),
          prescriptionwindow=1,
          doses=list(value=c(10,20,30,40),
             min=c(5,10,15,20),
             max=c(20,40,60,80),
             def=c(10,20,30,40)))
org("Purchase data:")
org(lmdb)
```

Purchase data:

pnr	eksd	apk	atc	strnum	packsize
8	2005-01-25	1	C10AA01	10	100
8	2005-03-03	1	C10AA01	10	100
8	2006-01-11	1	C10AA01	10	100
8	2006-04-15	1	C10AA01	10	100
8	2006-07-31	1	C10AA01	20	100

The results differ a bit. What is most correct should be checked.
Results using SAS:

```
adm <- simAdmissionData(4)
y <- xrecepter(drugdb=lmdb,
           adm=adm,
           atc="C10AA01",
           period=c("'2004jan01'd","'2015dec31'd"),
           maxdepot=8000,
           value=c(10,20,30,40),
           min=c(5,10,15,20),
           max=c(20,40,60,80),
           def=c(10,20,30,40))
org("Estimated doses data using SAS:")
org(y)
```

pnr	dose	firstday	lastday	exposure.days
8	10	2005-01-25	2005-03-02	36 days
8	20	2005-03-03	2005-05-23	81 days
8	5	2006-01-11	2006-04-14	93 days
8	10	2006-04-15	2007-04-06	356 days

Results using R:

```
x <- medicinMacro(drugs=list(simva=simva),drugdb=lmdb,admdb=NULL)
org(x$simva[])
```

pnr	dose	firstday	lastday	exposure.days
8	10	2005-01-25	2005-03-03	37 days
8	20	2005-03-03	2005-05-23	81 days
8	0	2005-05-23	2006-01-11	233 days
8	5	2006-01-11	2006-04-15	94 days
8	10	2006-04-15	2007-04-03	353 days