Medicin Macro

A matemathical formulation of the algorithm, an efficient implementation and a new R interface

Helene Charlotte Rytgaard

University of Copenhagen, Section of Biostatistics

December 8, 2016

Association studies

- Exposure-outcome analysis
  - lalala
- Example: PPI

Input data: Drug database

Example data:

	pnr	atc	eksd	strnum	packsize	apk
1:	1	A07	1996-06-15	400	30	1
2:	1	A07	1997-04-14	500	60	1
3:	1	A07	1998-03-23	400	30	2
4:	1	A07	1998-11-05	200	300	1
5:	1	A07	1999-08-30	400	100	1
6:	1	A12B	1997-08-21	750	500	1
7:	2	A07	1995-05-03	400	60	2
8:	2	A07	1995-07-10	500	60	1
9:	2	A07	1995-09-02	400	60	2
10:	2	A07	1995-09-21	400	60	3
11:	2	A07	1995-11-19	500	100	1
12:	2	A07	1996-04-26	400	100	1
13:	2	A07	1996-09-21	200	60	1
14:	2	A07	1997-07-18	400	300	1
15:	2	A07	1999-05-19	400	100	2

Input data: Admission database

Example data:

	pnr	inddto	uddto
1:	1	1999-02-27	1999-03-13
2:	1	2008-01-14	2008-01-28
3:	1	2010-12-29	2011-01-30
4:	2	1996-10-14	1996-10-14
5:	2	2003-08-25	2003-09-06
6:	2	2004-02-21	2004-04-03
7:	2	2007-11-22	2007-11-28
8:	2	2010-05-20	2010-06-14
9:	3	1999-06-23	1999-07-20
10:	3	2005-06-03	2005-06-22
11:	4	2008-04-07	2008-04-11
12:	4	2014-07-20	2014-08-17
13:	5	1999-03-01	1999-03-09
14:	5	2007-06-19	2007-07-06
15:	5	2011-05-07	2011-06-18

Current interface (medicin macro)


```
%x_recepter(recept_data, /* forventes at indeholde variable - skulle gerne passe med DST-standarder:
                  pnr - cpr/patientidentifikation
                  atc - ATC kode
                  eksd - udleveringsdato som sas-dato
                  strnum - numerisk styrke
                  apk - antal udleverede pakker
                  packsize - antal piller i hver pakke*/

         datoer, /* Et produkt af medicin-hjlpe-macro eller andet program som ordner ALLE indlggelser pr PNR p
             EEN record med fortlbende inddto uddto */


         out, /* tabel over behandlingsperioder - navn p SAS datast valgt af brugeren*/
         1a, /* atc kode - den behandling som der skal beregnes p*/
         5, /* antal recepter der indgr i beregninger - testet med 5, alts op til 2 fr og 2 efter interesserecept */
         50,  75, 100, 125, /* Doser svarende til de flgende variable - det er pillestrrelser
            - her og de flgende variable skal ALLE have en vrdi. Hvis der findes frre skal der blot gentages*/
         10, 50, 25, 50,    /* Mindst accepterede dosis af lgemidler p hver pillestyrke*/
         75, 200, 150, 150,    /* Max accepterede dosis*/
         50, 100,  75, 100,    /* Typiske doser - en slags "default" dosis - og startdosis altid ved left_only */
         10, /* Maximum sktrrelse af "restdosis" som kan overfres til flgende receptperioder. Denne giver mulighed for
          at forhindre excessiv ophobning hvis sm antagelser om maxdosis medfrer til tiltagende stort depot
          Max_depot er piller*styrke - Hvis der hjst m gemmes 100 piller a 10 mg, s er max_depot 1000
          */
         '01sep12'd, /* frste og sidste dato som har interesse kan angives som en "SAS-dato" eller med konventionen
            'ddmmyy'd     */
         '02may20'd,
         1, /* Hvis vrdien er 1 s kommer der tracking udskrift i loggen - hvis nul, s ikke. Tilsvarende slettes en rkke
           temporre datast hvis vrdien er 0 */
         1, /* Hvis vrdien er 1 s kommer der grafer */
         test, /* prfix p generede variable som kan benyttes til at skelne fra lignende variable genereret i andre trin
   */
         0 /* danner tabeller "l_" hvor doser og sluttider KUN regnes bagud*/
         );
```

Output data

Continuing example: (FIXME: Except NOT the same)

	pnr	dosis	startdag	slutdag
1	1000	50	17SEP15	06OCT15
2	1000	100	23JAN20	28JAN20
3	1000	100	08APR20	22MAY20
4	2000	20	15MAY13	05AUG13
5	2000	75	04NOV15	16NOV15
6	2000	100	15MAR17	21MAY17
7	3000	100	16MAR13	21MAR13
8	3000	100	26APR13	02MAY13
9	3000	50	10MAR16	08MAY16
10	3000	75	04JAN19	16JAN19
11	3000	100	14JUL19	04AUG19
12	3000	150	05AUG19	16AUG19
13	3000	75	17AUG19	22AUG19
14	3000	50	01NOV19	19NOV19
15	3000	75	20NOV19	16DEC19

Immediate limitations

- Slow
- Lack of transparency
- Other issues:
  - Dependence on the future
  - Only possible to specify four different doses
  - Possible to run for parts of data? Only specific individuals?

Input data: Drug database (unchanged)

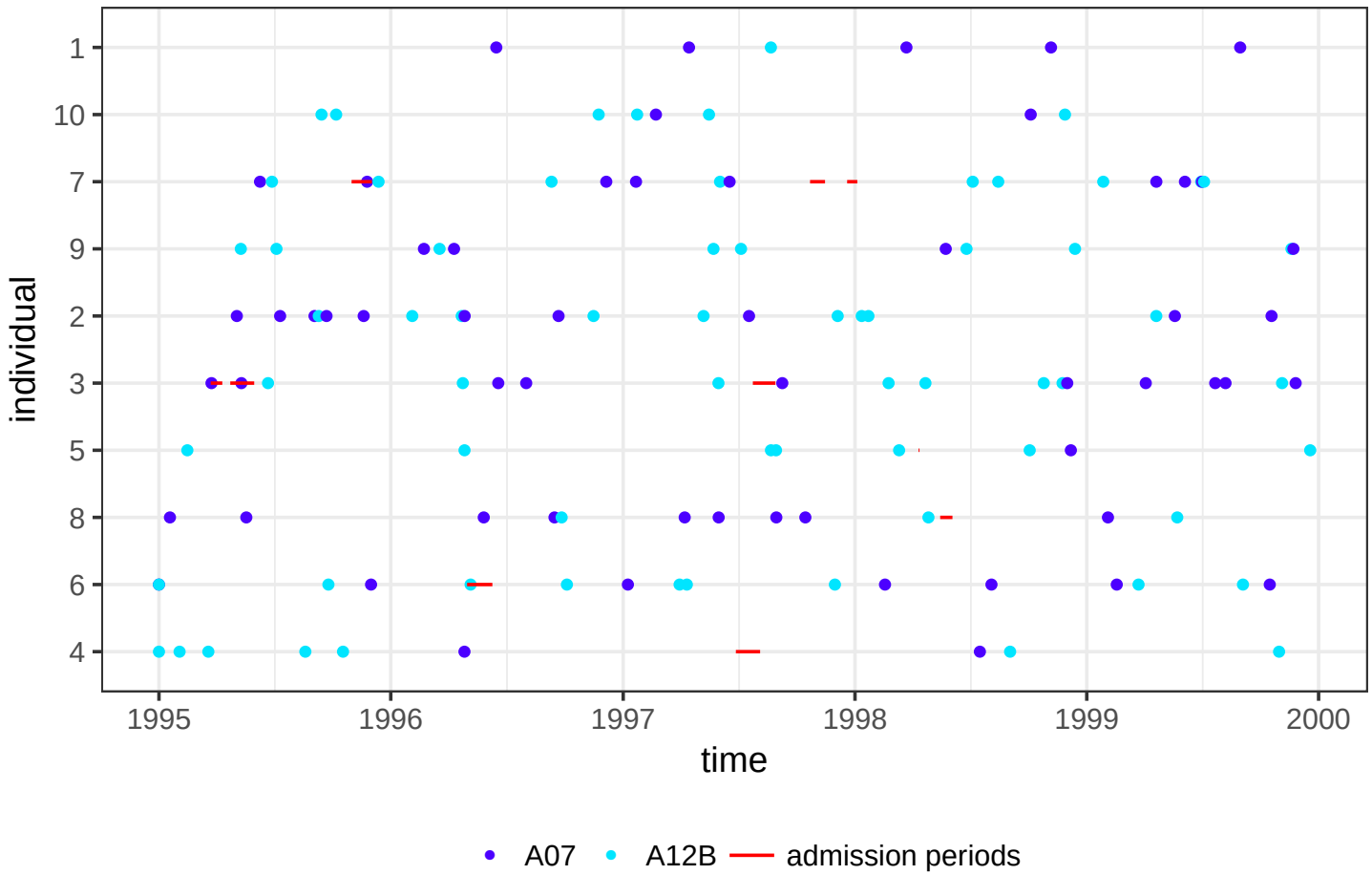
	pnr	atc	eksd	strnum	packsize	apk
1:	1	A07	1996-06-15	400	30	1
2:	1	A07	1997-04-14	500	60	1
3:	1	A07	1998-03-23	400	30	2
4:	1	A07	1998-11-05	200	300	1
5:	1	A07	1999-08-30	400	100	1
6:	1	A12B	1997-08-21	750	500	1
7:	2	A07	1995-05-03	400	60	2
8:	2	A07	1995-07-10	500	60	1
9:	2	A07	1995-09-02	400	60	2
10:	2	A07	1995-09-21	400	60	3
11:	2	A07	1995-11-19	500	100	1
12:	2	A07	1996-04-26	400	100	1
13:	2	A07	1996-09-21	200	60	1
14:	2	A07	1997-07-18	400	300	1
15:	2	A07	1999-05-19	400	100	2

Input data: Admission database (unchanged)

	pnr	inddto	uddto
1:	1	1999-02-27	1999-03-13
2:	1	2008-01-14	2008-01-28
3:	1	2010-12-29	2011-01-30
4:	2	1996-10-14	1996-10-14
5:	2	2003-08-25	2003-09-06
6:	2	2004-02-21	2004-04-03
7:	2	2007-11-22	2007-11-28
8:	2	2010-05-20	2010-06-14
9:	3	1999-06-23	1999-07-20
10:	3	2005-06-03	2005-06-22
11:	4	2008-04-07	2008-04-11
12:	4	2014-07-20	2014-08-17
13:	5	1999-03-01	1999-03-09
14:	5	2007-06-19	2007-07-06
15:	5	2011-05-07	2011-06-18

Tools for visualizing the data (input)

A `plot()`-function to show purchases and admission periods for the patients (i.e., visualizing input data):



How to use the interface

Load package:

```
library(heaven)
```

Create empty object:

```
d <- dpp()
```

Attach relevant data:

```
drugdb(d) <- drugdata
admdb(d) <- admissiondata
```

Add treatments:

```
drug(d, "treatment1") <- atc("A12B")
drug(d, "treatment1") <- pack(c(750, 75),
                  min = c(250, 25),
                  max = c(1000, 100),
                  def = c(750, 100))
```

Specify window of prescription dates to use in calculations:

```
pwindow(d) <- 3 ## use 3 prescriptions back in time
```

How to use the interface

When everything is specified, we perform the calculations by running:

```
process(d)
```

```
$treatment1
   id   X          B          E
1   1 100 1997-08-21 2007-11-26
2   2 100 1995-09-09 2030-02-05
3   3 100 1995-06-21 1997-08-12
4   3   0 1997-08-13 1998-02-21
5   3 100 1998-02-22 2010-02-08
6   4 100 1995-01-01 2030-08-17
7   5 100 1995-02-14 1996-02-23
8   5   0 1996-02-24 1996-04-25
9   5  75 1996-04-26 1997-08-20
10  5 100 1997-08-21 2000-03-01
11  6 100 1995-01-01 1995-03-16
12  6   0 1995-03-17 1995-09-23
13  6  25 1995-09-24 1996-05-04
14  6 100 1996-05-05 2015-01-26
15  7 100 1995-06-27 1999-09-16
16  8 100 1996-09-26 2009-08-27
17  9 100 1995-05-09 1999-06-18
18  9   0 1999-06-19 1999-11-18
19  9 100 1999-11-19 2001-06-03
20 10 100 1995-09-13 2014-04-21
```

How to use the interface

We may add treaments:

```
drug(d, "treatment2") <- atc("A07")
drug(d, "treatment2") <- pack(c(200, 400, 500),
             min = c(100, 100, 250),
             max = c(400, 500, 1000),
             def = c(300, 200, 500))
```

And then perform calculations again:

```
process(d)
```

```
$treatment1
  id   X          B          E
1  1 100 1997-08-21 2007-11-26
2  2 100 1995-09-09 2030-02-05
3  3 100 1995-06-21 1997-08-12
4  3   0 1997-08-13 1998-02-21
5  3 100 1998-02-22 2010-02-08
6  4 100 1995-01-01 2030-08-17

$treatment2
  id   X          B          E
1  1 200 1996-06-15 1996-08-13
2  1   0 1996-08-14 1997-04-13
3  1 500 1997-04-14 1997-06-12
4  1   0 1997-06-13 1998-03-22
5  1 200 1998-03-23 1998-07-20
6  1   0 1998-07-21 1998-11-04
```

How to use the interface

The function can be used treatment and/or id speficic:

```
process(d, treatment = "treatment2")
```

```
$treatment2
  id   X          B          E
1  1 200 1996-06-15 1996-08-13
2  1   0 1996-08-14 1997-04-13
3  1 500 1997-04-14 1997-06-12
4  1   0 1997-06-13 1998-03-22
5  1 200 1998-03-23 1998-07-20
6  1   0 1998-07-21 1998-11-04
```

```
process(d, id = 9)
```

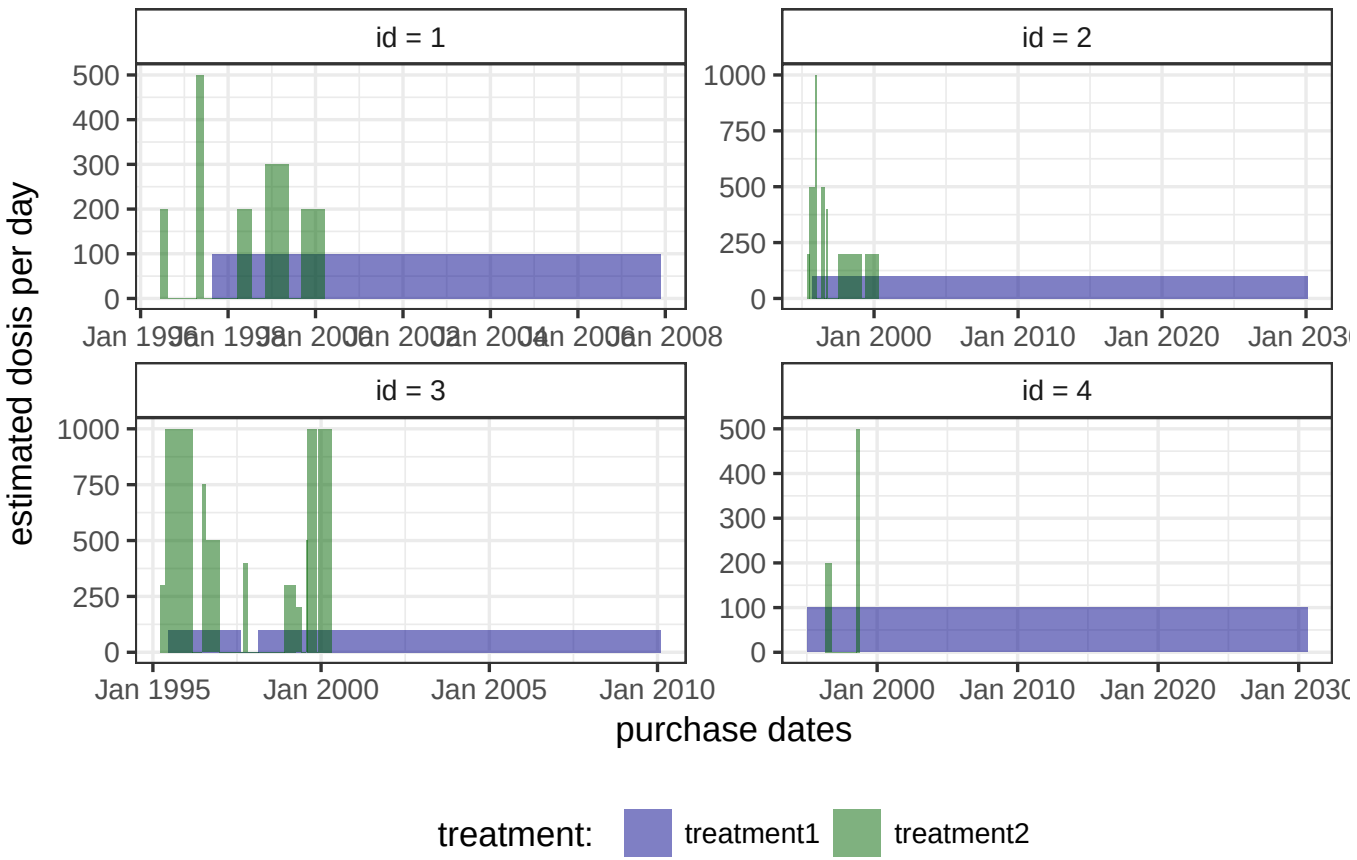
```
$treatment1
  id   X          B          E
1  9 100 1995-05-09 1999-06-18
2  9   0 1999-06-19 1999-11-18
3  9 100 1999-11-19 2001-06-03

$treatment2
  id   X          B          E
1  9 200 1996-02-22 1996-04-08
2  9 500 1996-04-09 1996-05-26
3  9   0 1996-05-27 1998-05-22
4  9 300 1998-05-23 1998-06-11
5  9   0 1998-06-12 1999-11-21
6  9 500 1999-11-22 2000-09-16
```

Built-in tools for varieties of output visulizations

A `plot()`-function to visualize the output is defined in the package:

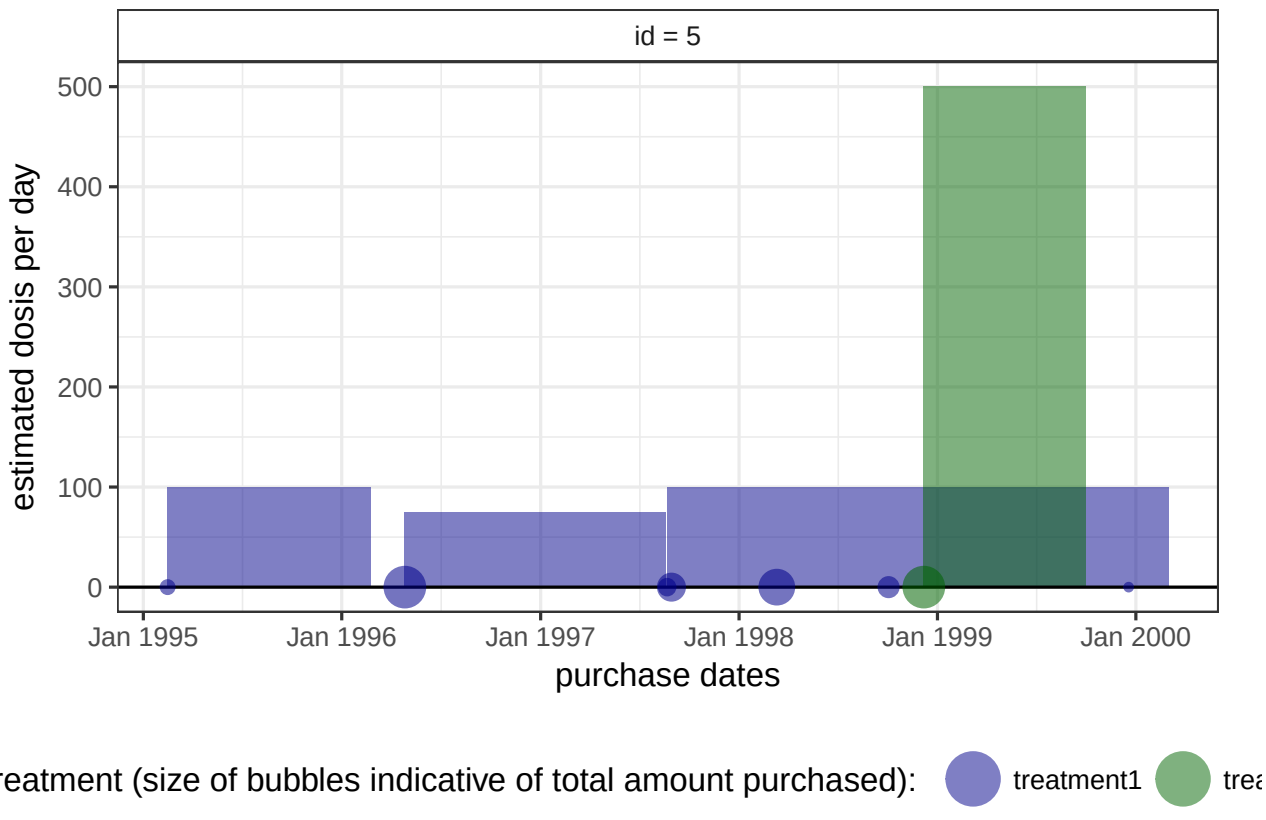
```
out <- process(d)
plot(out)
```



Built-in tools for varieties of output visulizations

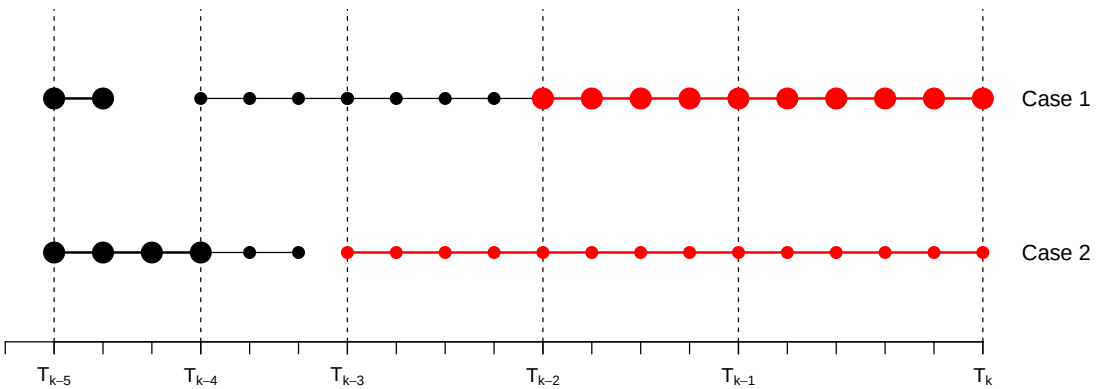
We may also take a closer view on the underlying purchases behind the final exposures estimated:

```
out1 <- process(d, keep_data = TRUE)
plot(out1, id = 5, trace = TRUE)
```



Technical details

- The R-interface and the following formulas are all based on the implementation of `medicin macro` (`left_only`).
- The computations performed consists basically of an averaging over a set of prescriptions back in time (set by the user)
- A number of things will for each prescription date help us determine how many dates back in time we should use for the calculations:
  - The number of days of supply of a certain drug is calculated based on the minimal possible doses for a drug
  - The actual number of dates between the prescription periods (where the number of days hospitalized is subtracted)
  - Whether or not the total amount of drug purchased at time k is approximately the same as purchased at earlier times



Final formula

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

First slide

aa

```
cat(1+1,"\n")
```

code output


```
library(heaven)
set.seed(8)
drugdata <- simPrescriptionData(10, startDate = "
    2006-01-01")
org(drugdata)
```

```
      pnr   atc        eksd strnum packsize apk
  1:    1   A07 2007-06-16    400       30   1
  2:    1   A07 2008-04-14    500       60   1
  3:    1   A07 2009-03-23    400       30   2
  4:    1   A07 2009-11-05    200      300   1
  5:    1   A07 2010-08-30    400      100   1
 ---
121:   10  A12B 2006-10-07    750      500   3
122:   10  A12B 2007-11-24    750      100   3
123:   10  A12B 2008-01-23    750      100   1
124:   10  A12B 2008-05-16    750      500   2
125:   10  A12B 2009-11-27    750      250   3
```

slide 2


```
2+2
```

```
[1] 4
```

Formula

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

Titled column

some explanations

Discussion

- cumulative exposure
  - andre macroer?
- tradition in other registry data research groups