



University of
Zurich^{UZH}

Epidemiology, Biostatistics and Prevention Institute

"Match of the day":
Finding least proximal
measurements to a given date
with ***fmatch***

Viktor von Wyl

Head of Swiss MS Registry

Co-Head of Division of Chronic Disease Epidemiology @ EBPI



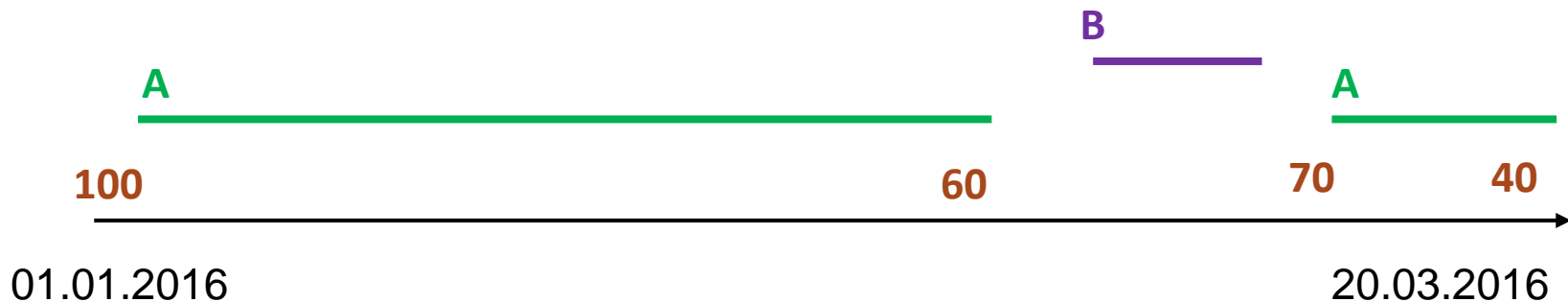
Imagine the following scenario...

Laboratory Measurements

ID	Lab Date	Lab Value
1	01.01.2016	100
1	18.02.2016	60
1	14.03.2016	70
1	20.03.2016	40
...

Treatment Episodes

ID	Start Date	End Date	Treatment
1	05.01.2016	20.02.2016	A
1	01.03.2016	10.03.2016	B
1	15.03.2016	(ongoing)	A
...



How can these data be merged efficiently?

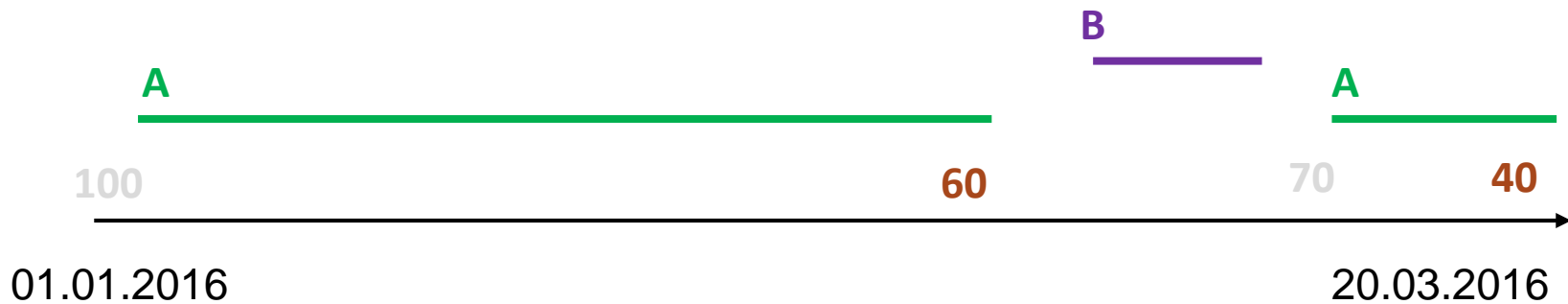


This is why we came up with *fmatch* ...

- Provides an easy and efficient way to combine data from different tables based on dates and date ranges
- The “engine” is the ***mmerge*** command (written by J. Weesie); please `net instal mmerge`, from(<http://fmwww.bc.edu/RePEc/bocode/m>) before first use
- Includes options for further specification of eligibility criteria for matching records (if, sorting, max. number of records to be retrieved)



Use Case 1: Filter out all laboratory values that belong to treatments with A



```
use "Treatment Episodes", clear
```

```
keep if Treatment == "A"
```

```
fmatch id using "Laboratory M.",  
umatchby(id Lab_Date) urange2(Start_Date, End_Date)
```



Let's take a look at the syntax of *fmatch* in this example

- **fmatch** *varlist*
specifies the matching key in the master data.
- **using** *filename*
specifies the “using data”, i.e. file used for merging
- **umatchby**(*varlist*)
specifies corresponding matching key in using data. Required even if match keys have same variable names in master and using data.
- **urange2**(*date, date*)
defines range by start date/end date from master data;
must be in date format (%d)
If a treatment has no End Date (still ongoing):
replace End Date = `mdy(12, 31, 2099)` if `mi(End Date)`



Use Case 2: Find the last laboratory value prior to start of the very first treatment



```
use "Treatment Episodes", clear
```

```
bysort id (Start_Date): gen First_Trtr_Start = Start_Date[1]
```

```
gen Dummy_Date = mdy(01,01,1900)
```

```
format Dummy_Date %d
```

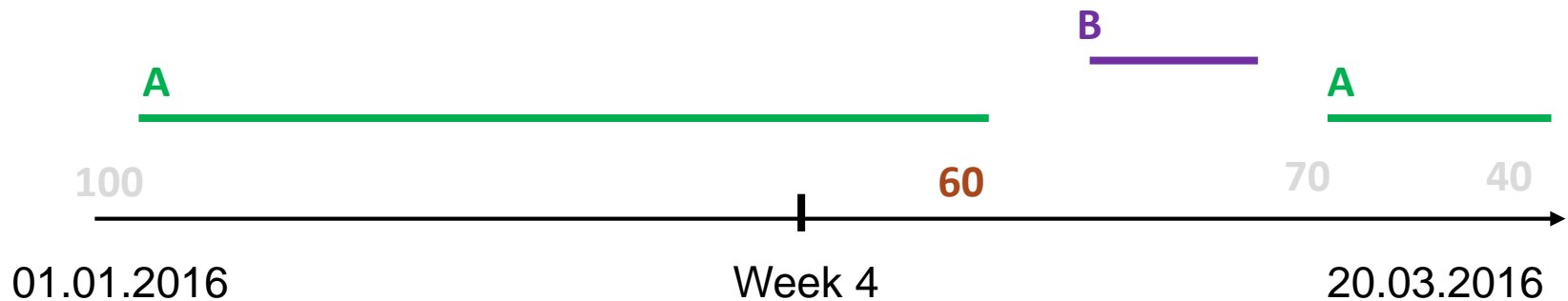
```
fmatch id using "Laboratory M.", unmatched(id Lab_Date)  
urange2(Dummy_Date, First_Trtr_Start)
```

Variable with date of very first treatment start

Dummy variable with date far in the past must be in %d format



Use Case 3: Find the closest laboratory value around a given time point



```
use "Treatment Episodes", clear
```

```
gen week_4 = Start_Date + 28
```

New variable with week 4 date;
must be in %d format

```
format week_4 %d
```

```
fmatch id week_4 using "Laboratory M.", unmatched(id Lab_Date) urange(-20, 20)
```

```
urange(-#, #)
```

Defines time window for eligible measurements;
here: 20 days. Order of values is important!



Additional options of *fmatch*

- **ukeep**(*varlist*)

variables to be kept from using data

- **uif**(*expression*)

restriction criteria for using records.

Example: `uif(lab_value < 60)` will only extract measurements smaller than 60

- **strict**(*varlist*)

results will only include records from using data where values in *varlist* are not missing

Example: `strict(lab_value)` will only consider non-missing measurements for merging



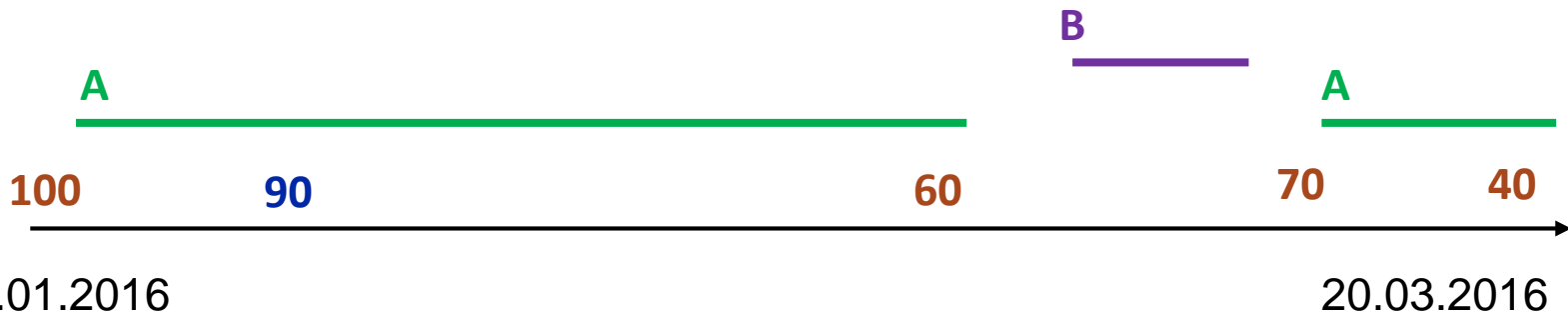
What if there are more than 1 lab measurements per ID / treatment episode combination

Laboratory Measurements

ID	Lab Date	Lab Value
1	01.01.2016	100
1	10.01.2016	90
1	18.02.2016	60
1	14.03.2016	70
1	20.03.2016	40
...

Treatment Episodes

ID	Start Date	End Date	Treatment
1	05.01.2016	20.02.2016	A
1	01.03.2016	10.03.2016	B
1	15.03.2016	(ongoing)	A
...





Options for dealing with multiple merge records

- `ufct (+/-varname)`

defines sorting (if more than one record) or using data record selection (min/max of specified variable)

Example: `ufct (+Lab_Value)` selects the largest measurement

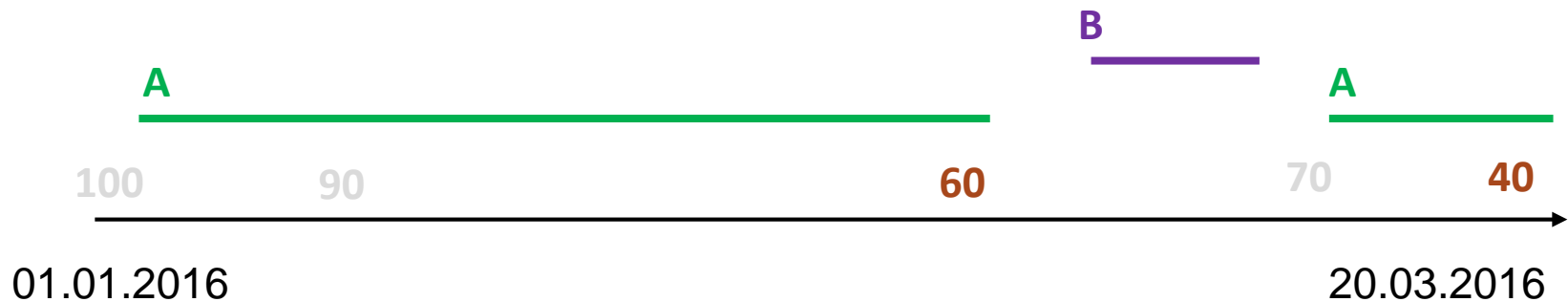
- `urecs (#)`

maximum number of records to be kept from using data

Example: `urecs(2)` selects up to 2 eligible measurement



Use Case 5: Find the smallest laboratory value per treatment



```
use "Treatment Episodes", clear
```

```
fmatch id using "Laboratory M.",  
umatchby(id Lab_Date) urange2(Start_Date, End_Date)  
ufct(-Lab_Value)recs(1)
```



What does the data look like...

```
fmatch id using "Laboratory M.", unmatched(id Lab_Date)
urange2(Start_Date, End_Date) recs(2)
```

ID	Start Date	End Date	Treat ment	Lab Date 1	Lab Val 1	Lab Date 2	Lab Val 2
1	05.01.2016	20.02.2016	A	10.01.2016	90	18.02.2016	60
1	01.03.2016	10.03.2016	B				
1	15.03.2016	(ongoing)	A	20.03.2016	40		
...				

```
fmatch id using "Laboratory M.", unmatched(id Lab_Date)
urange2(Start_Date, End_Date) recs(2) vert
```

ID	Start Date	End Date	Treat ment	Lab Date	Lab Val
1	05.01.2016	20.02.2016	A	10.01.2016	90
1	05.01.2016	20.02.2016	A	18.02.2016	60
1	01.03.2016	10.03.2016	B		
1	15.03.2016	(ongoing)	A	20.03.2016	40
...		



One final advice... monitor the fmatch report

```
*****fmatch-report*****
Time elapsed (secs): 1
Matching C:\Users\Viktor\AppData\Local\Temp\ST_0a000003.tmp with C:\temp\DATA\lab
Matched by id => id labdate
Keeping rna from C:\temp\DATA\lab
Date boundaries: moddate (lower); dummydate (upper)
Non-missing values: rna

-----
Total number of master records:          229
Master records with match for specified timeframe:  221
-----
*****
```

For additional information see `help fmatch` or email me directly: viktor.vonwyl@uzh.ch