

## Data Preparation

***FLoSC***

Forecasting Length of Stay and Cost

*Health and Social Care Modelling Group  
University of Westminster, London*

# Outline

---

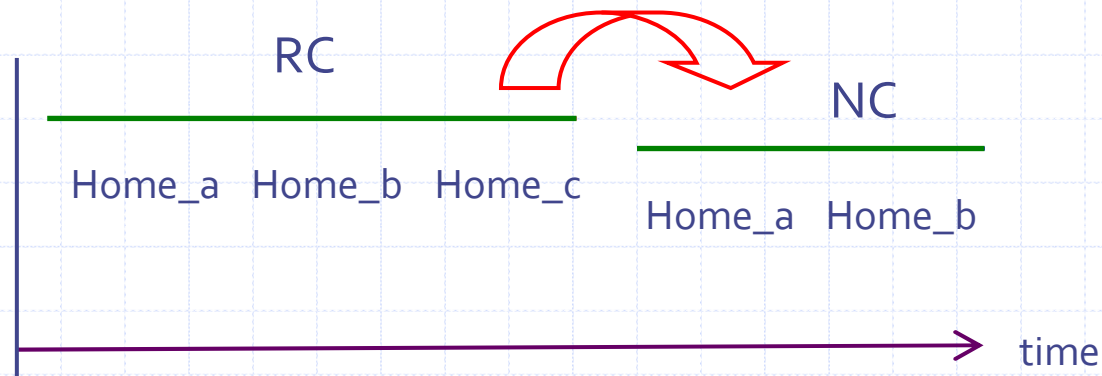
- ◆ Required data for FLoSC
  - ◆ Things that you need to bear in mind
  - ◆ What kind of problems you could be encountering
  - ◆ Solving these problems with examples
  - ◆ Tutorial
    - Hands on session
-

## Required data for FLoSC

- ◆ Data required for FLoSC are relatively simple and are generally derived from administrative data
- ◆ The essential information needed are as follows:
  - ID
  - *Gender (not strictly essential)*
  - Date of admission
  - Type of care admitted to
  - Date of discharge (if discharged)
  - Destination of discharge, including death (if applicable)

## Important issues

- ◆ FLoSC only considers two types of care – RC and NC in the LTC system
- ◆ Resident movements in the system are often recorded at the change of care home level (i.e. within type of care)



## Important issues (continued)

So,

- ◆ FLoSC models the system by type of care rather than homes
- ◆ Data preparation process **MUST** take these movements of residents within and between types of care into account

## Problems (1)

- ◆ How do you determine los and plos
  - 1) los is simple, how about plos?

x2

- ◆ Movement between homes

- 2) Suppose a resident moved between homes in 3 occasions.
- 3) What if a resident is discharged from a RC home and admitted to another RC home, say 30 days later

x3

- ◆ Movement between types of care

- 4) Suppose a resident moved from RC to NC
- 5) What if a resident is discharged from RC, stays at home (or hospital) for 3 months, and then is admitted to NC

# Problems with examples (2)

10/02/01-15/04/08  
Data availability

Suppose your data is in the following format – starting date is specified as 10/02/2001. Today is 15/04/2008.

x4

ID	Gender	date of admission	type of care	name of care home	date of discharge	end reason
1	F	10/04/2000	Nursing Care	home_a	27/03/2001	death
2	F	03/10/2001	Residential Care	home_a	21/10/2001	change care home
2	F	21/10/2001	Residential Care	home_b	08/11/2001	change care home
2	F	08/11/2001	Residential Care	home_b	31/03/2002	death
3	M	17/05/2001	Residential Care	home_a	04/06/2001	change care home
3	M	13/08/2001	Residential Care	home_b	20/08/2001	death
4	F	24/01/2002	Residential Care	home_a	02/04/2002	change care home
4	F	02/04/2002	Residential Care	home_b	07/08/2002	change type of care
4	F	07/08/2002	Nursing care	home_b	NA	NA
5	M	18/07/2001	Residential care	home_a	14/09/2001	change care home
5	M	14/09/2001	Residential care	home_b	02/05/2002	change type of care
5	M	12/11/2002	Nursing care	home_b	NA	NA

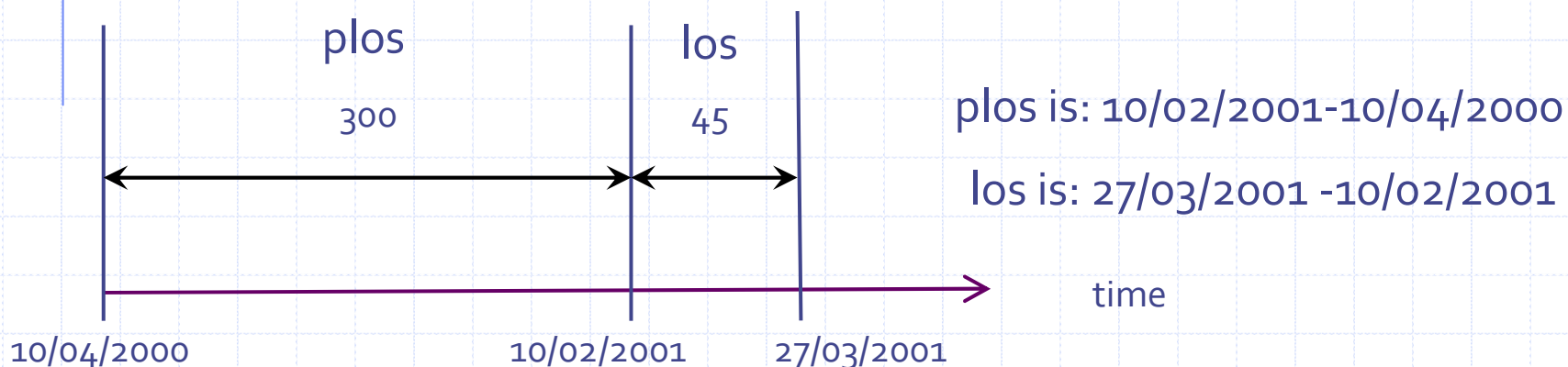
# Data preparation (1) – Example 1

10/02/01-15/04/08  
Data availability

x5

◆ The simple case: specified starting date is 10/02/2001

ID	Gender	date of admission	type of care	name of care home	date of discharge	end reason
1	F	10/04/2000	Nursing Care	home_a	27/03/2001	death



ID	Gender	type of care	plos	los	end reason
1	F	Nursing Care	300	45	death

# Data preparation (2)- Example 2

10/02/01-15/04/08  
Data availability

- ◆ Movement between homes with ix6 same type of care
  - Suppose a resident moved between homes in 3 occasions

ID	Gender	date of admission	type of care	name of care home	date of discharge	end reason
2	F	03/10/2001	Residential Care	home_a	21/10/2001	change care home
2	F	21/10/2001	Residential Care	home_b	08/11/2001	change care home
2	F	08/11/2001	Residential Care	home_b	31/03/2002	death

◆ plos = 0

◆ los is 31/03/2002 – 03/10/2001 = 179 days

ID	Gender	type of care	plos	los	end reason
2	F	Residential Care	0	179	death

# Data preparation (3)- Example 3

10/02/01-15/04/08  
Data availability

- ◆ Change of care between homes
  - What if a resident is discharged from a RC home and admitted to another RC home, say 30 days later

ID	Gender	date of admission	type of care	name of care home	date of discharge	end reason
3	M	13/08/2001	Residential Care	home_b	20/08/2001	death

- ◆ 30 days is an arbitrarily chosen threshold

x8

- ◆ Therefore, los is  $20/08/2001 - 13/08/2001 = 7$  days

ID	Gender	type of care	plos	los	end reason
3	M	Residential Care	0	7	death

# Data preparation (4)- Example 4

10/02/01-15/04/08  
Data availability

- ◆ Movement between types of care
  - from RC to NC

x9

ID	Gender	date of admission	type of care	name of care home	date of discharge	end reason
4	F	24/01/2002	Residential Care	home_a	02/04/2002	change care home
4	F	02/04/2002	Residential Care	home_b	07/08/2002	change type of care
4	F	07/08/2002	Nursing care	home_b	NA	NA

- ◆ Step 1 : Determine los between homes in RC
  - los is 07/08/2002 – 24/01/2002 = 195 days
- ◆ Step 2 : Capture movement between types of care
  - This resident is still alive (NA) today. So, what is the los?
  - Difference between the specified TODAY'S date (i.e. 15/04/2008) and date of admission to Nursing care

# Data preparation (4)- Example 4 (continued)

**10/02/01-15/04/08  
Data availability**

ID	Gender	date of admission	type of care	name of care home	date of discharge	end reason
4	F	24/01/2002	Residential Care	home_a	02/04/2002	change care home
4	F	02/04/2002	Residential Care	home_b	07/08/2002	change type of care
4	F	07/08/2002	Nursing care	home_b	NA	NA

- ◆ Step 1 : Determine los between RC homes
  - los is 195 days
  
- ◆ Step 2 : Capture movement between types of care
  - los is 15/04/2008 – 07/08/2002 = 2078 days
  
- ◆ Finally,

ID	Gender	type of care	plos	los	end reason
4	F	Residential Care	0	195	to nursing care
4	F	Nursing Care	0	2078	NA

# Data preparation (5)- Example 5

10/02/01-15/04/08  
Data availability

- ◆ Change of type of care
  - What if a resident discharged from RC stays at home (or hospital) for 3 months and then is admitted to NC

ID	Gender	date of admission	<u>type of care</u>	name of care home	date of discharge	end reason
5	M	12/11/2002	Nursing care	home_b	NA	NA

- ◆ The difference in between changing types of care is greater than 90 days.
- ◆ We probably cannot ignore this.
  - What do we do?
    - ◆ We delete these records.

# Data preparation (5)- Example 5 (continued)

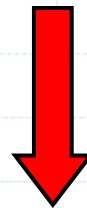
**10/02/01-15/04/08  
Data availability**

◆ Having removed these records we now have

ID	Gender	date of admission	type of care	name of care home	date of discharge	end reason
5	M	12/11/2002	Nursing care	home_b	NA	NA

◆ This patient is still alive. So, what is the los?

- How long has he been staying until TODAY'S date.
- los is  $15/04/2008 - 12/11/2002 = 1981$  days



ID	Gender	type of care	plos	los	end reason
5	M	Nursing care	0	1981	NA

# Data preparation (6) – **NOT FINISHED YET**

## Raw Data

ID	Gender	date of admission	type of care	name of care home	date of discharge	end reason
1	F	10/04/2000	Nursing Care	home_a	27/03/2001	death
2	F	03/10/2001	Residential Care	home_a	21/10/2001	change care home
2	F	21/10/2001	Residential Care	home_b	08/11/2001	change care home
2	F	08/11/2001	Residential Care	home_b	31/03/2002	death
3	M	17/05/2001	Residential Care	home_a	04/06/2001	change care home
3	M	13/08/2001	Residential Care	home_b	20/08/2001	death
4	F	24/01/2002	Residential Care	home_a	02/04/2002	change care home
4	F	02/04/2002	Residential Care	home_b	07/08/2002	change type of care
4	F	07/08/2002	Nursing care	home_b	NA	NA
5	M	18/07/2001	Residential care	home_a	14/09/2001	change care home
5	M	14/09/2001	Residential care	home_b	02/05/2002	change type of care
5	M	12/11/2002	Nursing care	home_b	NA	NA

## Prepared Data

ID	Gender	type of care	plos	los	end reason
1	F	Nursing care	300	45	death
2	F	Residential care	0	179	death
3	F	Residential care	0	7	death
4	F	Residential care	0	195	to nursing care
4	F	Nursing care	0	2078	NA
5	M	Nursing care	0	1981	NA

## Data preparation (7) - coding

### ◆ Coding for input data

gender	
code	meaning
0	female
1	male
-999	missing value

type of care	
code	meaning
1	residential care
2	nursing care
-999	missing value

end reason	
code	meaning
1	to residential care
2	to nursing care
3	died
-1	still in care
-2	left the system alive
-999	missing value

- ◆ FLoSC expects columns in this exact order: id, gender, type of care, plos, los, end reason

## Data preparation (8) – and finally ...

ID	Gender	type of care	plos	los	end reason
1	0	2	300	45	3
2	0	1	0	179	3
3	0	1	0	7	3
4	0	1	0	195	2
4	0	2	0	2078	-1
5	1	2	0	1981	-1

- ◆ We must admit this process is demanding and might require help from your IT team
- ◆ We have illustrated some of the problems that you may encounter, however, each local authority may have its own set of issues related to data preparation.

## Tutorial – Data preparation

- ◆ Download the raw data that we have presented from our website on [www.healthcareinformatics.org.uk](http://www.healthcareinformatics.org.uk)
- ◆ Follow the steps for preparing the data (use the presentation slides)
- ◆ In Microsoft Excel the difference between two dates in days can be calculated by:
  - = DATEDIF(A1,A2,"D")
  - = 1981 days

	A	B
1	12/11/02	1981
2	15/4/08	