

# Module 15: Multilevel Modelling of Repeated Measures Data

## MLwiN Practical<sup>1</sup>

*George Leckie, Tim Morris & Fiona Steele*  
Centre for Multilevel Modelling

### Pre-requisites

- MLwiN practicals for Modules 3 and 5

If you find this module helpful and wish to cite it in your research, please use the following citation:

Leckie, G., Morris, T., Steele, F. (2016). Multilevel Modelling of Repeated Measures Data - MLwiN Practical. LEMMA VLE Module 15, 1-76. URL: <http://www.bristol.ac.uk/cmm/learning/online-course/>.

## Contents

### PART I: GROWTH CURVE MODELS

<b>P15.1</b>	<b>Repeated Measures Data .....</b>	<b>1</b>
P15.1.1	Introduction to physical health functioning dataset.....	1
P15.1.2	Restructuring data from wide to long form .....	2
P15.1.3	Summarising longitudinal data.....	6
<b>P15.2</b>	<b>Introduction to Growth Curve Models .....</b>	<b>15</b>
<b>P15.3</b>	<b>Linear Growth Model for Continuous Repeated Measures .....</b>	<b>15</b>
P15.3.1	Preliminary analysis of the physical functioning dataset .....	15
P15.3.2	Random intercept model .....	18
P15.3.3	Random slope model .....	24
<b>P15.4</b>	<b>Nonlinear Growth.....</b>	<b>32</b>
P15.4.1	Quadratic and higher-order polynomials .....	32
P15.4.2	Splines.....	38
P15.4.3	Treating time as categorical: Multivariate response models .....	44

---

<sup>1</sup> This MLwiN practical is adapted from the corresponding Stata practical: Steele, F. (2014). Multilevel Modelling of Repeated Measures Data: Stata Practical. LEMMA VLE Module 15, 1-61. (<http://www.bristol.ac.uk/cmm/learning/course.html>).

<b>P15.5</b>	<b>Adding Explanatory Variables: Fitting Group-specific Growth Curves.....</b>	<b>50</b>
<b>P15.6</b>	<b>Residual Autocorrelation .....</b>	<b>60</b>
<b>P15.7</b>	<b>Introduction to Dynamic Models .....</b>	<b>61</b>
P15.7.1	Introduction to the smoking dataset .....	61
P15.7.2	A simple random effects dynamic model for smoking .....	65
<b>P15.8</b>	<b>The Initial Conditions Problem .....</b>	<b>69</b>
P15.8.1	Incorporating a model for smoking at occasion 1 .....	69
P15.8.2	Fitting joint models in MLwiN .....	70
P15.8.3	Results.....	71
<b>P15.9</b>	<b>Advanced Topics .....</b>	<b>75</b>
<b>References.....</b>		<b>75</b>

## P15.1 Repeated Measures Data

### P15.1.1 Introduction to physical health functioning dataset

In the first part of this practical we will fit growth curve models to data on health functioning from a study of British civil servants called the Whitehall II study (also known as the Stress & Health Study).<sup>2</sup> Health functioning was assessed by the SF-36, a 36 item instrument that comprises eight subscales covering physical, psychological and social functioning. These eight scales can be summarised into physical and mental health components. These are scaled using general US population norms to have mean values of 50 and low scores imply poor functioning. We will study change in physical health functioning which was measured on up to six occasions for each respondent.



The data are in wide form, i.e. with one record per individual and six variables for health functioning at the six measurement occasions. The dataset also includes information on the respondent's age at each occasion, their employment grade at the first occasion, and their gender. The analysis file contains the following variables for 4427 individuals:

Variable	Description and codes
id	Individual identifier (coded 1, 2, . . . , 8815)
female	Gender (1=female, 0=male)
grade	Employment grade at baseline (1=high, 2=intermediate, 3=low)
age1	Age at occasion 1 (years)
phf1	Physical health functioning score at occasion 1
...	...
age6	Age at occasion 6 (years)
phf6	Physical health functioning score at occasion 6

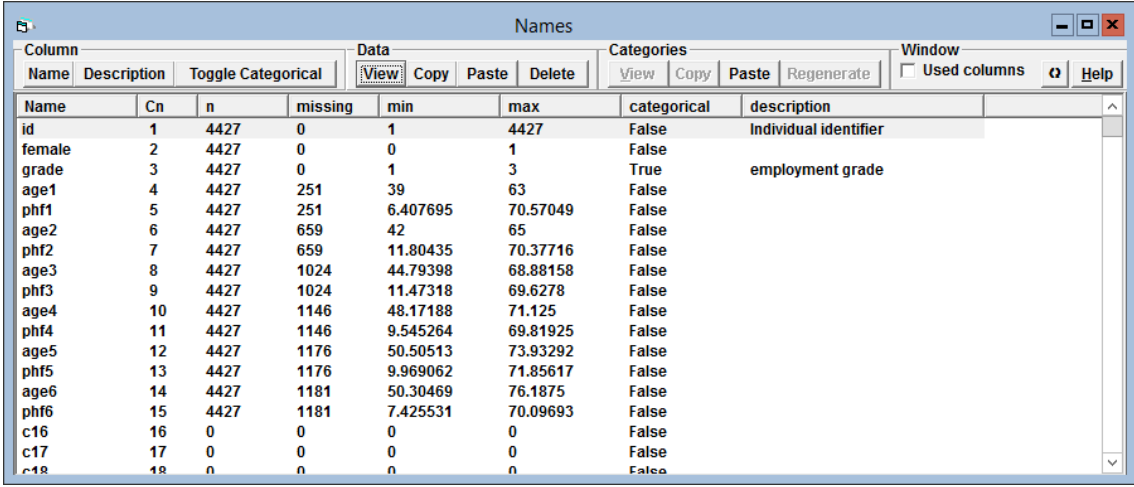
<sup>2</sup> <http://www.ucl.ac.uk/whitehallII>

To open the worksheet:

From within the LEMMA Learning Environment

- Go to **Module 15: Multilevel Modelling of Repeated Measures Data**, and scroll down to  **MLwiN Datafiles**
- If you do not already have MLwiN to open the datafile with, click ([get MLwiN](#)).
- Click “ 15.1.wsz”

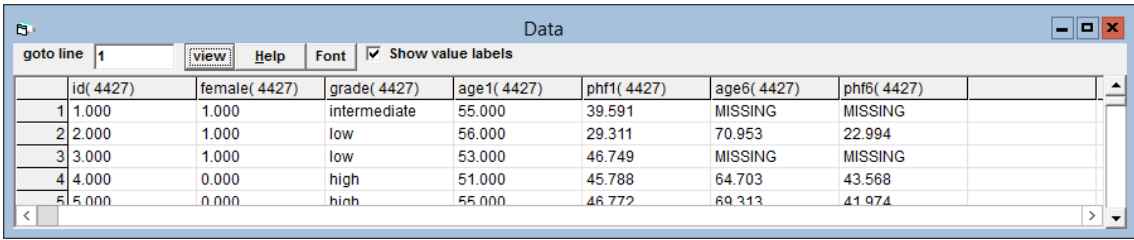
The **Names** window will appear.



The Names window displays a list of variables with their properties. The 'Data' tab is selected, showing columns for Name, Cn, n, missing, min, max, categorical, and description. The 'View' button is highlighted under the 'Data' heading.

Name	Cn	n	missing	min	max	categorical	description
id	1	4427	0	1	4427	False	Individual identifier
female	2	4427	0	0	1	False	
grade	3	4427	0	1	3	True	employment grade
age1	4	4427	251	39	63	False	
phf1	5	4427	251	6.407695	70.57049	False	
age2	6	4427	659	42	65	False	
phf2	7	4427	659	11.80435	70.37716	False	
age3	8	4427	1024	44.79398	68.88158	False	
phf3	9	4427	1024	11.47318	69.6278	False	
age4	10	4427	1146	48.17188	71.125	False	
phf4	11	4427	1146	9.545264	69.81925	False	
age5	12	4427	1176	50.50513	73.93292	False	
phf5	13	4427	1176	9.969062	71.85617	False	
age6	14	4427	1181	50.30469	76.1875	False	
phf6	15	4427	1181	7.425531	70.09693	False	
c16	16	0	0	0	0	False	
c17	17	0	0	0	0	False	
c18	18	0	0	0	0	False	

- To view a selection of the data, in the **Names** window select the variables **id**, **female**, **grade**, **age1**, **phf1**, **age6**, and **phf6** and click the **View** button under the **Data** heading.



The Data window shows a selection of the data for the variables id, female, grade, age1, phf1, age6, and phf6. The 'view' button is highlighted. The data is displayed in a table with columns for each variable and rows for individual observations.

	id( 4427)	female( 4427)	grade( 4427)	age1( 4427)	phf1( 4427)	age6( 4427)	phf6( 4427)
1	1.000	1.000	intermediate	55.000	39.591	MISSING	MISSING
2	2.000	1.000	low	56.000	29.311	70.953	22.994
3	3.000	1.000	low	53.000	46.749	MISSING	MISSING
4	4.000	0.000	high	51.000	45.788	64.703	43.568
5	5.000	0.000	high	55.000	46.772	69.313	41.974

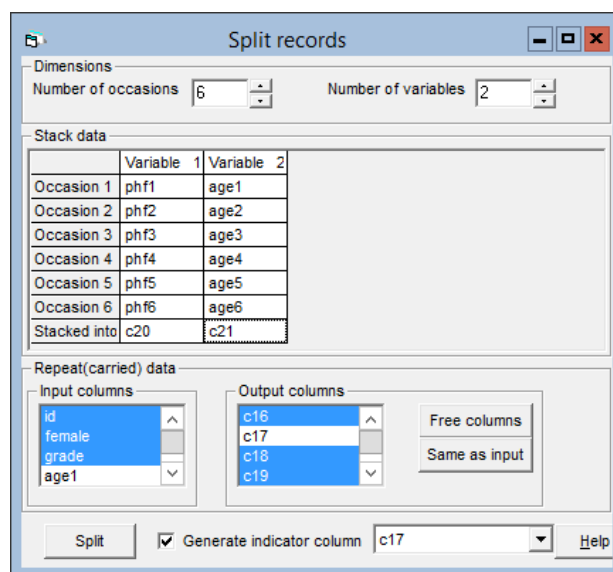
We can see that individuals 1 and 3 have missing data for occasion 6. We will obtain a summary of missing data patterns before fitting any models.

### P15.1.2 Restructuring data from wide to long form

Our first task is to convert the data from wide form (i.e., one record per individual) to long form (i.e., one record per occasion per individual). In other words, we need to create six records for each individual, with each record corresponding to one of the six measurement occasions of the two variables **age** and **phf**.

- From the **Data Manipulation** menu, select **Split Records**
- Change **Number of occasions** to **6**
- Change **Number of variables** to **2**
- Under **Stack data**, use the mouse pointer to fill the first column with the variables **phf1** through to **phf6** and then **c20**, and the second column with the variables **age1** through to **age6** and then **c21**. This tells the window to generate a new variable containing the occasion specific values of **phf** in column **c20** and the corresponding occasion specific values of **age** in column **c21**. These two columns therefore have one record per occasion per individual.
- Under **Repeat(carried) data**, select **id**, **female** and **grade** as the **Input columns**, and **c16**, **c18** and **c19** as the **Output columns**. This tells the window to create long-form versions of the variables **id**, **female**, and **grade** in columns **c16**, **c18**, and **c19**, respectively.
- Finally, check the **Generate indicator column** checkbox and select **c17** in the drop down list to create a measurement occasion indicator variable in **c17**.

Your **Split records** box should look as follows:



- Click **Split**, and when prompted to save the dataset select **No**

Our dataset now stores the original data in both wide and long form. This is confirmed when we inspect the **Names** window. The window shows the original wide-form variables are of length 4427 while the new long-form versions of these variables are of length 26562 (= 6 × 4427).

Names												
Column			Data			Categories				Window		
Name	Description	Toggle Categorical		Copy	Paste	Delete		Copy	Paste	Regenerate	<input type="checkbox"/> Used columns	Help
Name	Cn	n	missing	min	max	categorical	description					
id	1	4427	0	1	4427	False	Individual identifier					
female	2	4427	0	0	1	False						
grade	3	4427	0	1	3	True	employment grade					
age1	4	4427	251	39	63	False						
phf1	5	4427	251	6.407695	70.57049	False						
age2	6	4427	659	42	65	False						
phf2	7	4427	659	11.80435	70.37716	False						
age3	8	4427	1024	44.79398	68.88158	False						
phf3	9	4427	1024	11.47318	69.6278	False						
age4	10	4427	1146	48.17188	71.125	False						
phf4	11	4427	1146	9.545264	69.81925	False						
age5	12	4427	1176	50.50513	73.93292	False						
phf5	13	4427	1176	9.969062	71.85617	False						
age6	14	4427	1181	50.30469	76.1875	False						
phf6	15	4427	1181	7.425531	70.09693	False						
c16	16	26562	0	1	4427	False						
c17	17	26562	0	1	6	True						
c18	18	26562	0	0	1	False						
c19	19	26562	0	1	3	True						
c20	20	26562	5437	6.407695	71.85617	False						
c21	21	26562	5437	39	76.1875	False						
c22	22	0	0	0	0	False						
c23	23	0	0	0	0	False						
c24	24	0	0	0	0	False						

In the interests of keeping the worksheet as simple as possible, we will remove all wide-form variables as we will not analyse these further.

- From the **Data Manipulation** menu, select **Command interface**
- In the text entry area at the bottom of the **Command interface** window, type the following commands one by one, pressing return after each one has been inputted to run the commands:

```
ERAS C1-C15
MOVE
NAME C1 'id'
NAME C2 'occ'
NAME C3 'female'
NAME C4 'grade'
NAME C5 'phf'
NAME C6 'age'
DESC 'id' 'individual identifier'
DESC 'occ' 'measurement occasion'
DESC 'female' 'female civil servant'
DESC 'grade' 'employment grade'
DESC 'phf' 'physical health functioning (from SF-36)'
DESC 'age' 'age (years)'
```

- These commands will remove (ERAS) the wide-form variables in columns c1 to c15, shift (MOVE) all remaining variables (i.e., variables currently in columns c16 to c21) to the now empty columns at the beginning of the worksheet, rename (NAME) the six long-form variables, and lastly add variable labels (DESC) to these six variables

Your **Names** windows should now look as follows:

Names

Column

Name

Description

Toggle Categorical

Data

View

Copy

Paste

Delete

Categories

View

Copy

Paste

Regenerate

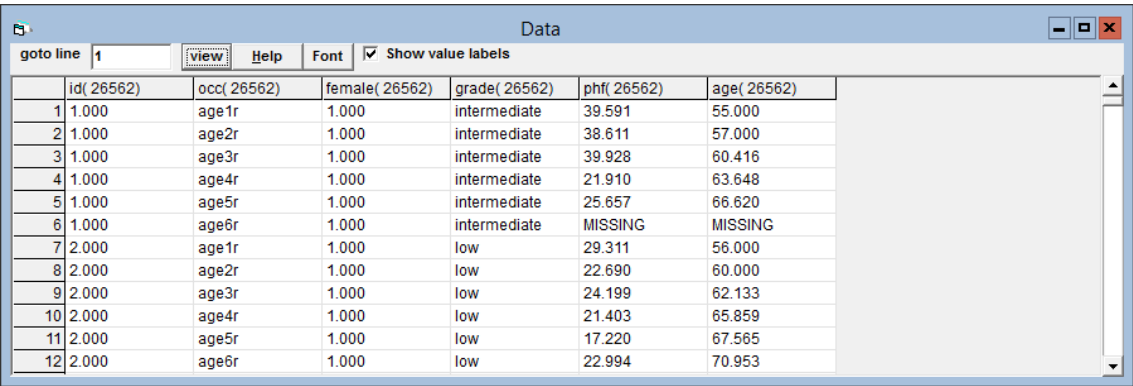
Window

☐ Used columns

Name	Cn	n	missing	min	max	categorical	description
id	1	26562	0	1	4427	False	individual identifier
occ	2	26562	0	1	6	True	measurement occasion
female	3	26562	0	0	1	False	female civil servant
grade	4	26562	0	1	3	True	employment grade
phf	5	26562	5437	6.407695	71.85617	False	physical health functioning (from SF...
age	6	26562	5437	39	76.1875	False	age (years)
c7	7	0	0	0	0	False	
c8	8	0	0	0	0	False	
c9	9	0	0	0	0	False	

Note that the restructured file now only contains the six long form variables **id**, **occ**, **female**, **grade**, **age** and **phf**. Each variable has 26562 records which is six times the number of individuals in the data.

- Select these six variables in the **Names** window and click the **View** button to view the new dataset



Data						
goto line	1	view	Help	Font	<input checked="" type="checkbox"/> Show value labels	
	id( 26562)	occ( 26562)	female( 26562)	grade( 26562)	phf( 26562)	age( 26562)
1	1.000	age1r	1.000	intermediate	39.591	55.000
2	1.000	age2r	1.000	intermediate	38.611	57.000
3	1.000	age3r	1.000	intermediate	39.928	60.416
4	1.000	age4r	1.000	intermediate	21.910	63.648
5	1.000	age5r	1.000	intermediate	25.657	66.620
6	1.000	age6r	1.000	intermediate	MISSING	MISSING
7	2.000	age1r	1.000	low	29.311	56.000
8	2.000	age2r	1.000	low	22.690	60.000
9	2.000	age3r	1.000	low	24.199	62.133
10	2.000	age4r	1.000	low	21.403	65.859
11	2.000	age5r	1.000	low	17.220	67.565
12	2.000	age6r	1.000	low	22.994	70.953

Notice that the occasion variable **occ** has value labels **age1r**, **age2r**, **age3r**, **age4r**, **age5r** and **age6r**. The reason the variable has value labels is because it is defined as a categorical variable (you can see this in the **Names** window in the above screenshot as in the **categorical** column the variable **occ** is set to **True**). We shall change this variable to a numerical variable as this will remove the current value labels revealing the underlying numeric values 1,2,3,4,5,6 which are the corresponding measurement occasions:

- Select the variable **occ** in the **Names** window and click the **Toggle Categorical button** to tell MLwiN that the variable is continuous (i.e. **categorical** = **False**)
- Click the **View** button again and you will see that the underlying numeric values

This document is only the first few pages of the full version.

To see the complete document please go to learning materials and register:

<http://www.cmm.bris.ac.uk/lemma>

**The course is completely free.** We ask for a few details about yourself for our research purposes only. We will not give any details to any other organisation unless it is with your express permission.