Appendix to "A General Multilevel Multistate Competing Risks Model for Event History Data, with an Application to a Study of Contraceptive Use Dynamics", *Journal of Statistical Modelling*, 4(2): 145-159.

Fiona Steele, Harvey Goldstein, and William Browne

Data Preparation for a Discrete-time Multilevel Multistate Competing Risks Model

Data preparation is discussed in the context of the contraceptive use dynamics example described in the paper. We consider two states: use (i=1) and non-use (i=2).

Suppose that a woman uses contraception for 18 months, then discontinues and does not use contraception for 14 months; she then resumes contraceptive use for 27 months, before switching to another method and continuing use of this method for 1 month at which point the survey takes place and the episode is censored. The episode-based data file is shown in Table 1.

Table 1. Episode-based file

Individual (k)	Episode (j)	State (i)	Duration (months)	Transition	Censor
1	1	1	18	Discontinue	0
1	2	2	14	Start to use	0
1	3	1	27	Switch	0
1	4	1	1	None	1

Censor indicates whether the episode is right-censored; here, the duration of the first three episodes is completely observed.

The first step in restructuring the data for a multinomial discrete-time model is to create a multinomial response for each time interval, perhaps after grouping durations into broader intervals. We will group the data into six-month intervals, so t=0 for 0-5 months, t=1 for 6-11 months etc. The multinomial response y_{tijk} has R_i categories for state i, where R_i = 2 for i = 1 and R_i = 1 for i = 2. The multinomial response is coded as follows:

y_{tijk} = 0 if no event has occurred

1 if individual discontinues contraceptive use (i = 1), or
if individual starts to use contraception (i = 2)

2 if individual switches to another method (i = 1 only)

In addition, two indicator variables, I_1 and I_2, denoting the state are created. These are interacted with t and covariates. The restructured dataset is shown in Table 2.

Table 2. Data in discrete-time format (1 record per 6-month interval)

k	j	i	t	Exposure	y_{tijk}	I_1	I_2	I_1*t	I_2*t
1	1	1	0	6	0	1	0	0	0
1	1	1	1	6	0	1	0	1	0
1	1	1	2	6	0	1	0	2	0
1	1	1	3	1	1	1	0	3	0
1	2	2	0	6	0	0	1	0	0
1	2	2	1	6	0	0	1	0	1
1	2	2	2	3	1	0	1	0	2
1	3	1	0	6	0	1	0	0	0
1	3	1	1	6	0	1	0	1	0
1	3	1	2	6	0	1	0	2	0
1	3	1	3	6	0	1	0	3	0
1	3	1	4	4	2	1	0	4	0
1	4	1	0	2	0	1	0	0	0

The *exposure* variable is the number of months in a six-month interval for which the woman was at risk of an event. For example, the first episode ends in a discontinuation at month 18. Therefore the woman was at risk of discontinuation (or a switch) for the whole of the first three six-month intervals (0-5, 6-11, and 12-17 months) and just one month in the fourth interval (18-23 months). In the analysis, observations are weighted by the number of months of exposure; the variable *exposure* is defined as the denominator for the multinomial response y_{tijk} .

To fit a multilevel multinomial model in *MLwiN*, the data must be further expanded to obtain a set of binary responses for each multinomial response. This reconstruction is carried out automatically when a multinomial model is specified in *MLwiN*. Therefore data may be read into *MLwiN* in discrete-time format (see Table 2). Note that this further expansion of the data is required only for episodes that originate in state 1, where two types of event are considered; since there is only one type of event that can occur in the non-use state the indicator of event occurrence for episodes originating in state 2 is binary. For *i*=1, the multinomial response y_{t1jk} for each time interval *t* is converted to two binary responses $y_{t1jk}^{(r)}$, where $y_{t1jk}^{(r)}$=1 if y_{t1jk} = *r* and 0 otherwise (*r* = 1, 2). For each time interval, the two binary responses are stacked. Thus, for the first episode in the example above the final data structure is shown in Table 3.

Table 3. Data in *MLwiN* format for a multinomial response model

t	r	$y_{t1jk}^{(r)}$	$\text{I}_1^{(1)}$	$\text{I}_1^{(2)}$	$\text{I}_1^{(1)}*t$	$\text{I}_1^{(2)}*t$
0	1	0	1	0	0	0
0	2	0	0	1	0	0
1	1	0	1	0	1	0
1	2	0	0	1	0	1
2	1	0	1	0	2	0
2	2	0	0	1	0	2
3	1	1	1	0	3	0
3	2	0	0	1	0	3

The indicator for state 1, I_1, is replaced by indicators for *r*, $\text{I}_1^{(1)}$ and $\text{I}_1^{(2)}$. These are multiplied with duration and the covariates to allow duration and covariate effects to vary according to the type of transition from contraceptive use. The destination-specific individual random effects for state 1, $u_{1k}^{(1)}$ and $u_{1k}^{(2)}$, are fitted by allowing the coefficients of $\text{I}_1^{(1)}$ and $\text{I}_1^{(2)}$ to vary randomly across individuals. In addition the random effect for state 2, u_{2k}, is obtained by allowing the coefficient of I_2 to vary across individuals.