

Unbalanced versus missing data

In the context of, for example, a 2-level longitudinal study, the difference between missing Level 1 observations and the missing data referred to by HLM's **Missing data?** options available when making an MDM file may be confusing. First of all, we need to deal with the concept of unbalanced data versus balanced (complete) data. It is easier to explain this using an example, so suppose we have a repeated measures design, in which measurements were made at 6 different time points for each respondent on three different variables. Now let us suppose that the following data were obtained for a respondent:

obs	v1	v2	v3
1	9	8	.
2	7	.	6
3	5	4	3
4	.	2	5
5	3	4	.

First of all, we notice that for this respondent we only have 5 of the possible 6 measurements. In terms of classical statistical analysis, we would not have been able to use this respondent at all, as you need each respondent to have the same number of measurements. This would mean everybody should have 6 measurements each, or they should all have 5, etc. In multilevel models, however, all respondents need not have the same number of measurements. So we can use this respondent even though we have less than 6 measurements. A next respondent may have a full set of 6, or even just 2 measurements. This is what is referred to as being one of the advantages of ML models in that we can use the data from this respondent even though the set of measurements is incomplete. This is the "unbalanced" data concept.

The '.' on variable V3 on the first occasion, on V2 on the second occasion, etc. indicate missing data. Data values are absent for specific variables. This is what we refer to as "missing data". Within HLM, listwise deletion is used to deal with missing data problems. The user has the option to use listwise deletion either when making the MDM file, or when running the analysis. Suppose that deletion is requested when the MDM file is created. Listwise deletion in this context implies that after inclusion of this respondent in our analysis, we need to check each separate record (equivalent here to occasion of measurement) for completeness. In this case, there is only one line of data (record) with complete data- observation no 3. Thus, while the respondent will be retained when making the MDM file, only the data for observation 3 will be used. On the other hand, if deletion is requested at the time of analysis, only variables included in the model specified will be considered when the listwise deletion is implemented. For example, if only V1 and V2 were included in the model specified, observations 1, 3, and 5 will be retained for this respondent.

In previous versions of HLM, the option to use pairwise deletion was also available. While pairwise deletion would retain more of the data than listwise deletion, use of the method of deletion is likely to cause problems during the estimation procedure. In addition, the calculation of degrees of freedom and $-2 \ln L$ are debatable, and thus it becomes risky to base any conclusions on these. Due to these concerns, HLM 6 no longer offers the option to implement pairwise deletion.