PhUSE 2010

Paper SP07

Getting your data in good shape for Non Linear Mixed Effects Modeling

Vincent Buchheit, Novartis, Basel, Switzerland

ABSTRACT

NONMEM® is the industry standard for Non Linear Mixed Effects Modeling and still represents the main methodology used for modeling and simulation (M&S) in the Pharmaceutical industry. M&S has been identified as a key component to speed up drug development and reduce failures in late phases. The NONMEM® software requires data in a very specific pre-defined format. In this abstract, the author will provide an overview of some of the data challenges that a programmer may face when delivering ready-to-use NONMEM® modeling input datasets.

INTRODUCTION

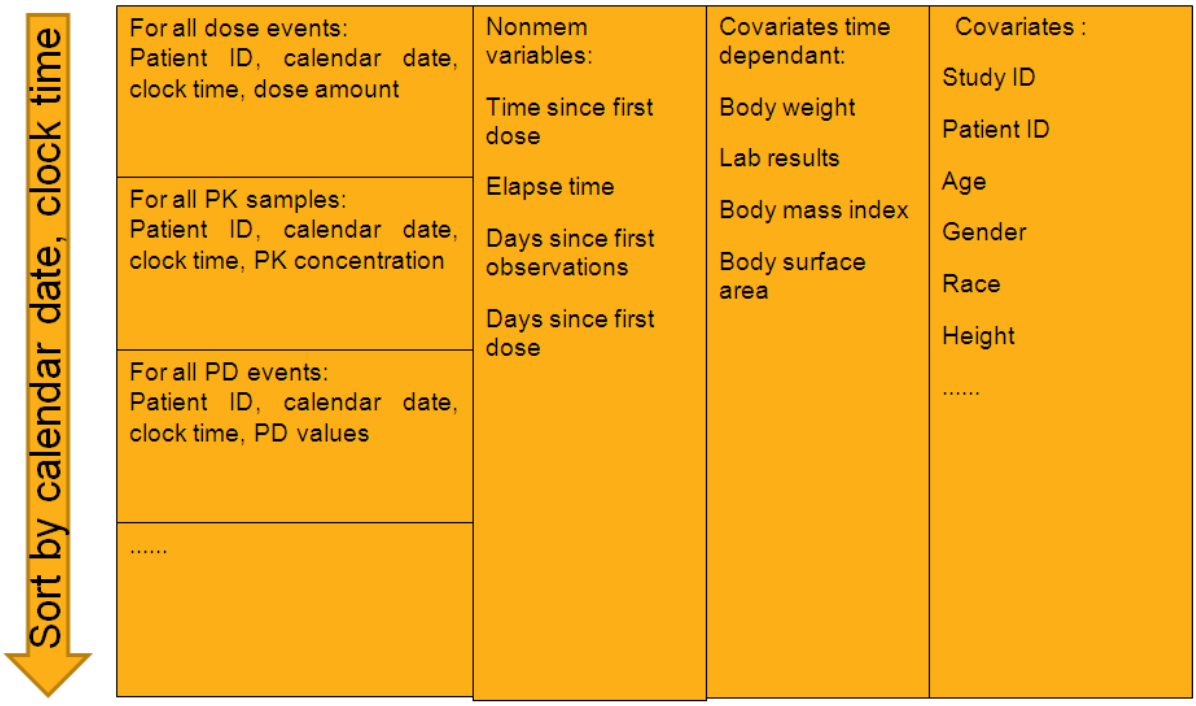
The NONMEM® input dataset is an ASCII file that contains time dependent information where each row is one single event e.g. dose event, pharmacokinetic (PK) event (drug concentration in the body), pharmacodynamic (PD) event (markers for safety and efficacy).

In the example below, 1 color represent one type of event (1 color for the pharmacodynamic (PD) assessment, 1 different color for the pharmacokinetic (PK) assessment, 1 different color for the dose observations....).

STUD	ID	ID2	EVDT	EVID	CMT	DV	LNDV	MDV	AMT	DOSE	DUR	RATE	TIMH	TSLD	DAY	DAY2	AGE0	SEX	RACE	HT0	WT0	BMI0	BSA0
12101	1001	10131101	24MAR06:10:09:00	0	4	63.52346	5.945421	0	0	33.5	0	0	0	0	0	0	43	1	1	181	105	32.1	2.3
12101	1001	10131101	27MAR06:09:00:00	0	3	30.65293	0	0	0	33.5	0	0	0	0	2.952083	0	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:09:00:00	0	3	75.03406	0	0	0	33.5	0	0	0	0	4.952083	0	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:09:01:00	0	4	57.07707	6.021023	0	0	33.5	0	0	0	0	4.952778	0	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:09:10:00	0	2	85.5143	-0.46362	1	0	33.5	0	0	0	0	4.959028	0	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:09:12:00	0	4	87.48237	5.966147	0	0	33.5	0	0	0	0	4.960417	0	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:09:15:00	0	4	86.61708	6.003887	0	0	33.5	0	0	0	0	4.9625	0	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:09:17:00	0	4	52.6286	5.97381	0	0	33.5	0	0	0	0	4.963889	0	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:09:23:00	0	4	36.75099	6.003887	0	0	33.5	0	0	0	0	4.968056	0	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:09:45:00	1	2	0	0	1	33.5	33.5	0.5	67	0	0	4.983333	0	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:10:00:00	0	2	82.56956	6.57647	0	0	33.5	0	0	0.25	0.25	4.99375	0.010417	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:10:12:00	0	4	39.16724	6.003887	0	0	33.5	0	0	0.45	0.45	5.002083	0.01875	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:10:13:00	0	2	58.40729	6.173786	0	0	33.5	0	0	0.466667	0.466667	5.002778	0.019444	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:10:20:00	0	2	27.80137	5.411646	0	0	33.5	0	0	0.583333	0.583333	5.007639	0.024306	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:10:30:00	0	2	93.88114	4.770685	0	0	33.5	0	0	0.75	0.75	5.014583	0.03125	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:10:45:00	0	2	71.15088	4.325456	0	0	33.5	0	0	1	1	5.025	0.041667	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:11:42:00	0	4	94.34376	5.966147	0	0	33.5	0	0	1.95	1.95	5.064583	0.08125	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:11:45:00	0	2	8.031077	3.817712	0	0	33.5	0	0	2	2	5.066667	0.083333	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:13:50:00	0	2	85.76386	3.602777	0	0	33.5	0	0	4.083333	4.083333	5.153472	0.170139	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:13:57:00	0	4	14.78104	5.958425	0	0	33.5	0	0	4.2	4.2	5.158333	0.175	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:15:41:00	0	4	52.58056	5.97381	0	0	33.5	0	0	5.933333	5.933333	5.230556	0.247222	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:15:45:00	0	2	95.04444	3.367296	0	0	33.5	0	0	6	6	5.233333	0.25	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:17:30:00	0	2	64.45107	3.152736	0	0	33.5	0	0	7.75	7.75	5.30625	0.322917	43	1	1	181	105	32.1	2.3
12101	1001	10131101	29MAR06:17:33:00	0	4	93.07029	6.003887	0	0	33.5	0	0	7.8	7.8	5.308333	0.325	43	1	1	181	105	32.1	2.3
12101	1001	10131101	30MAR06:09:30:00	0	2	55.2461	1.876407	0	0	33.5	0	0	23.75	23.75	5.972917	0.989583	43	1	1	181	105	32.1	2.3
12101	1001	10131101	30MAR06:09:58:00	0	4	95.55382	5.958425	0	0	33.5	0	0	24.21667	24.21667	5.992361	1.009028	43	1	1	181	105	32.1	2.3
12101	1001	10131101	30MAR06:12:12:00	0	4	8.448673	5.948035	0	0	33.5	0	0	26.45	26.45	6.085417	1.102083	43	1	1	181	105	32.1	2.3
12101	1001	10131101	30MAR06:14:10:00	0	4	28.52941	5.942799	0	0	33.5	0	0	28.41667	28.41667	6.167361	1.184028	43	1	1	181	105	32.1	2.3
12101	1001	10131101	30MAR06:16:15:00	0	4	32.38035	6.059123	0	0	33.5	0	0	30.5	30.5	6.254167	1.270833	43	1	1	181	105	32.1	2.3
12101	1001	10131101	30MAR06:17:18:00	0	4	33.71462	6.042633	0	0	33.5	0	0	31.55	31.55	6.297917	1.314583	43	1	1	181	105	32.1	2.3
12101	1001	10131101	31MAR06:10:00:00	0	2	68.45663	1.235471	0	0	33.5	0	0	48.25	48.25	6.99375	2.010417	43	1	1	181	105	32.1	2.3
12101	1001	10131101	31MAR06:10:23:00	0	4	29.88976	5.988961	0	0	33.5	0	0	48.63333	48.63333	7.009722	2.026389	43	1	1	181	105	32.1	2.3
12101	1001	10131101	31MAR06:12:13:00	0	4	70.38069	5.926926	0	0	33.5	0	0	50.46667	50.46667	7.086111	2.102778	43	1	1	181	105	32.1	2.3
12101	1001	10131101	31MAR06:14:12:00	0	4	67.6967	5.953243	0	0	33.5	0	0	52.45	52.45	7.16875	2.185417	43	1	1	181	105	32.1	2.3
12101	1001	10131101	31MAR06:16:13:00	0	4	92.5106	5.910797	0	0	33.5	0	0	54.46667	54.46667	7.252778	2.269444	43	1	1	181	105	32.1	2.3
12101	1001	10131101	31MAR06:17:16:00	0	4	44.92123	5.899897	0	0	33.5	0	0	55.51667	55.51667	7.296528	2.313194	43	1	1	181	105	32.1	2.3

A NONMEM® file can be considered as a combination of:

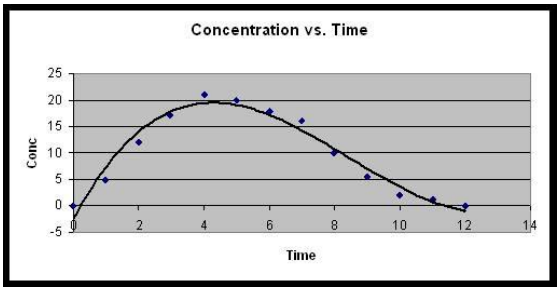
- covariates (age, gender, height)
- time dependant covariates (body weight, lab results, body mass index, body surface area...)
- time event observations (dose events, PK samples, PD events....)
- NONMEM® specific variables



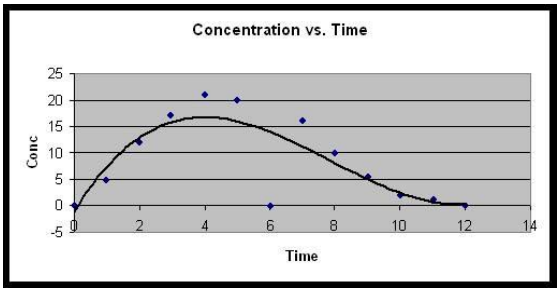
The quality of the NONMEM® input dataset drives the quality of the modeling outputs and hence the quantitative input to drug development decision making. Consequently careful attention needs to be paid to processes for harmonization and cleaning the data. This presentation will highlight some representative examples of typical issues.

OUTLIERS

Despite the best efforts at data collection and management, no database will be 100% correct and accurate. Mistakes happen and will always exist e.g. data programmers always encounter outliers in their studies. What are the impacts from a modeling point of view? If you plot the pharmacokinetic profile for a patient you can expect a curve like this one.

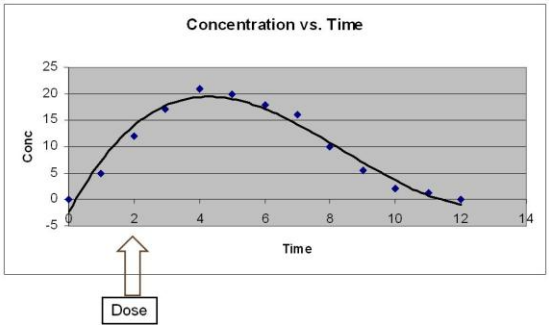


Now if for some reasons one of the values is not correct, the model that fits the data will be different and the predictions based on the model are likely to be incorrect. What do you do? What is wrong? The time? The PK concentration value itself? Does this PK concentration belong to the (14 hours) post-dose profile? Or is the time correct but the PK value itself is wrong?

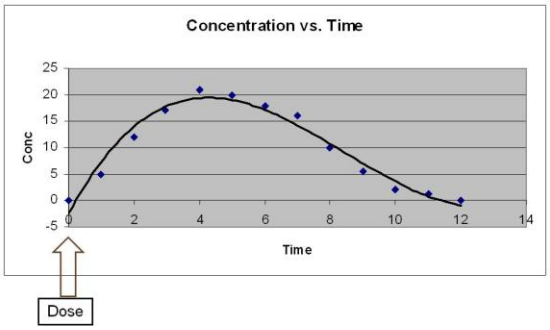


DATA CONSISTENCY

Because we often combine dosing history with other data like PK or PD data, and because we look at the entire profile, the programmer uniquely placed to identify these types of discrepancies. In the example below if one compares the calendar date and clock time from the dose administration records and the calendar date and clock time from the pharmacokinetic data profile, one finds positive PK concentration before the first dose. In theory this is not possible.



What is wrong with the data? Is it the calendar date and clock time from the dose administration records incorrect, and should be 2 hours earlier, or should the PK profile be shifted by 2 hours? Both scenarios are possible.



MISSING DATA

Because not all data are known, you may need to impute some of them. All dose calendar dates and clock times are not consistently collected during a study. Let's imagine a study where patients are treated for 6 months. Because of the duration of the study, we only collect the dose clock time at certain visits. The patients come to the clinic, we give him/her the treatment and then we collect PK data for the next 24 hours. In theory the next dose should be given once the last PK sample has been taken. You cannot just replace all missing dose clock time by a default value for all patients, like 09:00. If the last PK sample has been taken at 09:15, then your dose event will come before your last PK sample in your NONMEM® dataset. This is not correct. So programmers need to be creative and derive an algorithm that works for all patients. In this specific case, we could do something like: replace missing dose event by the last PK sample +1 minute.

Now what do you do if the dose clock time prior to your PK profile is missing? 09:00 for all of them? It may not work for all patients. You can estimate it by the PK pre-dose clock time +1 minute

OPEN ENTRY FIELDS IN CASE REPORT FORM

In some clinical studies, you may find useful information in some of the open entry fields. In one of the current project I'm working on, some dose regimen were entered in the comments field. Why? Simply because the dose regiment for some of the patients did not match what was specified in the protocol, the database being created from the protocol. In such a case you can exclude the patients from the analysis or integrate the comments in program and include those patients in your analysis.

```
else if dosreg1c in (42) then do;
  if (index(dosreg1a,"(MO,WE,FR) 3 TIMES PER WEEK EV") gt 0 or index(dosreg1a,"3 TIMES MONDAY WEDNESDAY AND F") gt 0 or index(dosreg1a,"3 TIMES, MONDAY WEDNESDAY FRID") gt 0 or
      index(dosreg1a,"30 MG MO-WE-FR.  EVERY 2 WEEKS") gt 0 or index(dosreg1a,"EVERY OTHER WEEK M, W, F") gt 0 or index(dosreg1a,"EVERY OTHER WEEK M-W-F") gt 0 or
      index(dosreg1a,"M, W, F EVERY OTHER WEEK") gt 0 or index(dosreg1a,"M, W, F EVERY OTHER WEEK") gt 0 or index(dosreg1a,"M, W, FR. EVERY OTHER WEEK") gt 0 or
      index(dosreg1a,"M,W,F EVERY OTHER WEEK") gt 0 or index(dosreg1a,"M,W,FR EVERY OTHER WEEK") gt 0 or index(dosreg1a,"M-W-F EVERY OTHER WEEK") gt 0 or
      index(dosreg1a,"MO - WE - FR EVERY 2 WEEKS (AM") gt 0 or index(dosreg1a,"MO, WE, FR (EVERY OTHER WEEK)") gt 0 or index(dosreg1a,"MO, WE, FR (THREE TIMES A WEEK") gt 0 or
      index(dosreg1a,"MO, WE, FR (THREE TIMES EVERY") gt 0 or index(dosreg1a,"MO, WE, FR - EVER OTHER WEEK") gt 0 or index(dosreg1a,"MO, WE, FR ALTERNATE WEEKS") gt 0 or
      index(dosreg1a,"MO, WE, FR EVERY 2 WEEKS") gt 0 or index(dosreg1a,"MO, WE, FR EVERY OTHER WEEK") gt 0 or index(dosreg1a,"MO, WE, FR QOW") gt 0 or
      index(dosreg1a,"MO, WE, FR, EVERY OTHER WEEK") gt 0 or index(dosreg1a,"MO, WE, FRI ALTERNATE WEEKS") gt 0 or index(dosreg1a,"MO,WE,FR (THREE TIMES A WEEK E") gt 0 or
      index(dosreg1a,"MO-WE-FR EACH OTHER WEEK") gt 0 or index(dosreg1a,"MO-WE-FR EVERY OTHER WEEK") gt 0 or index(dosreg1a,"MO-WE-FRI EVERY 2 WEEKS (AMEND") gt 0 or
      index(dosreg1a,"MO/WE/FRI EVERY OTHER WEEK") gt 0 or index(dosreg1a,"MON, WE, FR ALTERNATE WEEKS") gt 0 or index(dosreg1a,"MO, WE, FR ALTERNATIVE WEEKS") gt 0 or
      index(dosreg1a,"O OTHER WEEK M, W, F") gt 0 or index(dosreg1a,"MON, WED FR ALTERNATE WEEKS") gt 0 or index(dosreg1a,"MO, WE, FR - EVERY OTHER WEEK") gt 0 or
      index(dosreg1a,"MON, WED, FRI EVERY OTHER WEEK") gt 0 or index(dosreg1a,"EVERY OTHER WEEK MO, WE, FR") gt 0 or index(dosreg1a,"MO, WE; FR (EVERY OTHER WEEK) ") gt 0 or
      index(dosreg1a,"MONDAY, WEDNESDAY, FRIDAY EVER") gt 0)
      and weekday(datepart(et))in (2,4,6) then output;

  else if index(dosreg1a,"SA/MO/WE EVERY OTHER WEEK") gt 0 and weekday(datepart(et))in (7,2,4) then output;
  else if index(dosreg1a,"TU, TH, SA EVERY OTHER WEEK") gt 0 and weekday(datepart(et))in (3,5,7) then output;
  else if (index(dosreg1a,"FRI SUN TUE EVERY OTHER WEEK") gt 0 or index(dosreg1a,"EVERY OTHER WEEK FRI SUN TUE") gt 0) and weekday(datepart(et))in (6,1,3) then output;
  else if index(dosreg1a,"ALTERNATE THUR/SAT/MON.  WEEK.") gt 0 and weekday(datepart(et))in (5,7,2) then output;
  else if index(dosreg1a,"ALTERNATE WEEK - WED/THUR/SAT") gt 0 and weekday(datepart(et))in (4,5,7) then output;
  else if (index(dosreg1a,"ALTERNATE WEEKS WED/FRI/SUN") gt 0 or index(dosreg1a,"ALTERNATE WEEKS - WED, FRI, SU") gt 0) and weekday(datepart(et))in (4,6,1) then output;
  else if index(dosreg1a,"SU, TU, THU EVERY OTHER WEEK") gt 0 and weekday(datepart(et))in (1,3,5) then output;
  else if (compress(dosreg1a)=compress("EVERY OTHER WEEK") or index(dosreg1a,"PER PROTOCOL AMENDMENT N: 2") gt 0 or index(dosreg1a,"MO, MI, FR EVERY OTHER WEEK") gt 0) and weekday(datepart(et)) in (2,4,6) then output;
end;
```

CONCLUSION

Data quality drives the quality of modeling results. If you do not pay attention to all potential data issues, you may lose some important information. The role of the programmer is fundamental. The role of the modeling and simulation programmer is not just only about following a set of rules containing data set specification! It goes beyond that.

ACKNOWLEDGMENTS

I'd like to thank all my colleagues who supported me to put this talk together.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Vincent Buchheit
Novartis
WSJ.27.6.067
Forum 1
Novartis Campus
CH-4002 Basel
Switzerland
Work Phone: +41 61 32 46072
Email: vincent.buchheit@novartis.com