Creating NONMEM datasets — how to escape the nightmare

Shafi Chowdhury

Shafi Consultancy Limited, London, UK

Pharmacokinetics (PK) is the impact of the body on the drug and pharmacodynamics (PD) is the impact of the drug on the body. The effect of the drug on the target population of patients is predicted using models during population PK/PD analysis. This type of analysis is becoming more and more common in clinical trials, but creating the dataset structure required by the widely used non-linear mixed effects modelling software called NONMEM® is often the nightmare part of the process. It usually takes months to prepare the NONMEM dataset before the pharmacokineticist feels that it is ready for them to use. It is also produced after unblinding, adding to the delay in finalising the dataset before analysis can be performed with it. This delay can lead to holding back decisions about future trials, or those decisions are then made without taking into account the population PK analysis report. This paper will look at the issues which cause problems when creating a NONMEM dataset, and what steps we can take to avoid these problems and minimize the time taken to create the final dataset.

Keywords: NONMEM, Pop PK/PD, Population PK/PD analysis

Introduction

Population pharmacokinetics (PK)/pharmacodynamics (PD) analysis is performed by modelling the drug concentration data to predict the effect of the drug on the target population of patients. This is performed using a software called NONMEM. Data managers and programmers have traditionally created the analysis dataset that is used as the source data for the NONMEM software. However, the variables required are often complex and come from many different source datasets, making them both difficult and time-consuming to create.

The data used to create the NONMEM dataset can be grouped into various categories. By defining how these categories will be treated for the common types of data issues such as missing data, the process for generating NONMEM datasets can be standardized. This will ensure that there is a common understanding between the pharmacokineticist and the programmer, and a consistent expectation from the different pharmacokineticists working in the various trials. A standardized process will also allow the development of standard programs, which will lead to a reduction in time and effort required to generate a NONMEM dataset.

Correspondence to: Shafi Chowdhury, Shafi Consultancy Limited, 7 Blossom Way, Uxbridge, Middlesex UB10 9LL, UK. Email: Shafi@shaficonsultancy.com

Background

Population PK/PD analysis look at PK and PD data in the body over time. These may be affected by many factors, called covariates. What is important is that we know the value of these covariates at the time when the blood samples were taken for measuring the PK and PD results, and here lies the cause of all problems.

Studies are usually set up with either safety or efficacy in mind, not to collect PK or PD data. As a result, these data are cleaned for analysis in the clinical trial report, but not for creating NONMEM datasets. Laboratory data for example may be summarized by visit, and it may have been checked to make sure that the visits are in chronological order, but what about whether they were done before or after dosing at that visit? This cross-checking between the various types of data is often missing from traditional data management checking processes, and leads to problems when trying to create a NONMEM dataset.

If the trial team can see the results of the concentrations, then they will in many cases be able to tell which treatment a subject was given and therefore remove the blinding status of the study. As a result, the concentration details are not released to the trial team until after the database is locked. This adds pressure to the programmers as they are looking at these data for the first time and they have to make sure that it is all as expected, i.e. the values are

Published by Maney
DOI 10.1179/175709310X12847285433037

plausible and the blood samples and results reconcile. Arthur Collins, Mark Peterson, and Greg Silva[1] explain in depth the difficulties related to collecting and transferring PK/PD data. These are on the whole universal, and affect all pharmaceutical companies. However, assuming that the concentrations were measured correctly, which usually they are, the problems with preparing datasets for analysis using NONMEM lies mostly with inconsistencies in the database from data collected in Case Report Forms (CRF) or electronic Case Report Forms (eCRF). Therefore, these data issues can be resolved if they are actively searched for during the conduct of the trial.

In the end, a combination of data collection and transfer issues and a lack of focus on the requirements for PK/PD analysis can lead to data which should be cleaned being left untouched. This means that the programmer spends most of the time identifying and discussing data issues, adding to the delay in finalising the NONMEM dataset.

What Is Required in a NONMEM Dataset

The NONMEM dataset consists of three basic components; the dosing records, observations i.e. the PK/PD results, and covariates. They all have to be correct and consistent in order to create a good quality NONMEM dataset.

Dosing records

Dosing records can consist of every administration date and time of the trial drug, taken both during the clinic visits and at home via diary data, or assumptions can be made about drug administration taken at home and only those at the clinic visits are recorded. This depends on the type of study and the length of study. If the date and time of every dose taken by the patient are recorded, then that is the best option; however, this is not always practical and can lead to a long cleaning process if the patient is recording these dates and times. Most often patients take the first dose at the investigator site and then take the remaining doses at home until the following visit, and only the administration dates and times of doses taken at the visit are recorded.

If only the drug administration at visits are recorded and the patients are expected to take the medication at home at regular intervals, then a steady state flag is often introduced to account for the doses which are not recorded. Steady state is where the amount of drug leaving the body over an interval is the same as the amount introduced into the body with new drug administration. Whether steady state can be assumed or not is usually defined by the pharmacokineticist. The steady state record is usually the planned dose before the first recorded dose at that visit, and is accompanied by an additional variable to indicate the dosing interval. This is the equivalent of stating that the patient was taking this dose at the specified interval since the last recorded dose at the previous visit. If there were missed doses, then the duration of the missed dose period may affect whether steady state can be assumed or not at a particular visit. If two or three additional drug administration dates and times before the visit can be recorded, then a much better dosing profile is available for analysis.

Observation records

Observation records contain the dates and times of all the blood samples taken for PK/PD analysis together with their results from the laboratories. It is possible for combined treatments to have more than one result from each blood sample, and different PD parameters will also have their own result from the same blood sample. As a result, there may be many records with the same date and time but with different results, each containing a flag to indicate which parameter the results belongs to.

Although the blood sample dates and times will be collected in the CRF or eCRF, the results from the laboratory will not be available until after the database is locked. This is to ensure the trial team is not unblinded before the database is locked. For example, if the patients are randomized to 10 and 50 mg, then the drug concentration from a sample taken 5 minutes after the dose may reveal which dose the patients were taking. As a result, the PK results are only released to the trial team after the database is locked.

The observation records are therefore a record of how much dose (when looking at PK concentrations) is in the body at different times around drug administration. Often a sample is taken shortly before drug administration, and then again three of four times afterwards. Owing to this key link with drug administration times, the relationship between the two data types is crucial. For example, if the drug administration time is not available, then it is not possible to analyse the PK results which are planned to be after the dose, in case that they were not afterwards. The observations therefore need to be looked at in two parts:

1. concentrating on the dates and times and their consistency pre database lock;
2. concentrating on the results from the different analytes after database lock.

Covariates

The dosing records and the observation records create the total number of records in the NONMEM dataset. The covariates are variables which may or may not have influence on the result, and therefore, the values of these covariates at the times of blood sample and drug administration are of interest. Covariates do not add to the number of

records in the final dataset, but they are additional variables in the final dataset.

There are different types of covariates which need to be merged to the dosing and observation records. These can be grouped into four general categories:

1. basic demography covariates which are typically recorded once per patient;
2. those recorded once per visit, for example, weight, or those which can be recorded more than once per visit, such as laboratory data, especially if re-tests are done due to abnormal values in the planned lab test;
3. those which are made up of other covariates, for example, body mass index which is a combination of height and weight;
4. those that are not planned but can happen at any point in time, for example, specific adverse events or concomitant therapies.

Example of NONMEM dataset

Example of some of the columns which might be present in a typical NONMEM dataset is shown in Table 1.

What Can Cause Problems?

Dosing data

If drug administrations are missed or there are missing dates or times or there are dose changes, then these can lead to many problems. A clear approach must be specified before hand to take care of these issues so as to cause minimum delay during production. These should be listed and clarified with the data manager to ensure that the data are correct. If there is a dose reduction, then this can influence steady state, so a clear definition is required: e.g. after how many days of a specific dose can steady state be assumed. How the dose reductions are recorded is also very important. If they are not recorded in a consistent manner across studies, then the program will need many changes for each new trial. What should happen to the PK/PD data after missed dose or a dose without date or time should also be specified. If it is decided that PK data are dropped afterwards, then they should be listed.

Observation data — PK/PD blood sample and result

If there are missing blood samples or missing date or time of blood sample, then these can cause many problems when it comes to determining how they are related to the drug administration. PK observations with missing blood sample dates or times are often dropped, and so it is very important that every effort is made to ensure that they exist and that they are consistent with the drug administration dates and times. Consistency checks should be in place to check that relative times of blood sample reflect the planned relative time, and especially that they are on the expected side of drug administration, i.e. if they are planned to be after drug administration, then they actually are after drug administration! If these can be captured early on, then the data manager can ask the site to pay particular attention to these issues.

The results are not released to the data manager until after the database lock. However, it should be possible to arrange for dates and times of the analysed samples to be sent. These can then be used to reconcile the two sets of data and make sure that

Table 1 Example of some of the columns which might be present in a typical NONMEM dataset

ID	Date	Time	Visit	SS	II	Flag	DV (result)	Dose	Sex	Weight	ALT
1	20100507	0900	3	1	24	0	0	30	1	89	12
1	20100508	0908	3	0	0	0	0	30	1	89	12
1	20100509	0915	3	0	0	0	0	30	1	89	12
1	20100510	0855	3	0	0	1	0.89	30	1	89	12
1	20100510	0900	3	0	0	0	0	30	1	89	12
1	20100510	0910	3	0	0	1	14.23	30	1	89	12
1	20100510	0940	3	0	0	1	14.01	30	1	89	12
1	20100510	1100	3	0	0	1	13.52	30	1	89	12
1	20100610	0900	4	1	24	0	0	30	1	92	13
1	20100611	0908	4	0	0	0	0	30	1	92	13
1	20100612	0915	4	0	0	0	0	30	1	92	13
1	20100613	0835	4	0	0	1	0.77	30	1	92	13
1	20100613	0845	4	0	0	0	0	30	1	92	13
1	20100613	0915	4	0	0	1	13.23	30	1	92	13
1	20100613	0940	4	0	0	1	13.03	30	1	92	13
1	20100613	1110	4	0	0	1	12.524	30	1	92	13

Note: Flag=0 indicates dosing records; Flag=1 indicates PK concentrations.
Dose=dose the patient was on at that time.
SS=1 indicates steady state record, and is created artificially.
II=24 indicates dosing interval during the steady state period.
Sex would have been merged by patient number.
Weight would have been merged by patient and visit number.
ALT would have been merged by patient number, visit number, date, and time, and then LOCF would have been applied by going forward and backward to ensure that all the records at that visit are covered.
Date and time:

- when FLAG=0, these are the dosing dates and times;
- when FLAG=1, these are the blood sample dates and times.

for every sample taken there is a result, or a reason as to why it is missing. This is a key step, as without this step it will mean that these data are only looked at after the database lock, when there is pressure to produce the key safety and efficacy tables for the Clinical Trial Report (CTR). It will also mean that they will be difficult to resolve, which may lead to more samples being excluded from the analysis. Using an external team not related to the trial team is an option, and this can allow them to be unblinded and look at the actual data. It also means that after the database is locked, this team is only looking at the NONMEM part, and so can finalize the NONMEM dataset quickly at a time when the trial team will be busy finalizing the tables, listings, and figures for the CTR.

Covariates

Missing covariates are the last part of the process which can cause many problems. If they are missing or out of sequence, for example, not in chronological order when looking at visit numbers, or they have extreme values which are unrealistic, then the analysis may be affected. Covariates such as adverse events or concomitant therapies also suffer from incomplete or missing dates, and so a process should be defined to make it easy to identify times of certainty and times of assumption for these covariates. Once again, the key here is planning, and if the general issues can be pre-defined in the specification, then many of the issues can be simply listed, summarized, and flagged in the NONMEM dataset as problem data.

The method of replacing missing covariates should be defined in the specification. It is most likely that a different approach is used for different types of covariates. Those recorded once per patient may be replaced by the mean, median, or mode. Those recorded at each visit may be replaced by Last Observation Carried Forward, or other standard methods for replacing missing values. The important aspect is that it should be clear as to which covariates were replaced in each record of the final NONMEM dataset. These should also be flagged in the NONMEM dataset and additionally listed and summarized to ensure transparency. They can also be checked against the CTR if required. A standard approach for resolving missing and incomplete covariates will ensure that they can be programmed into macros and used across trials.

Conclusions

Data collection and identifying what is important and must be checked are key parts of the NONMEM dataset creation process. It is therefore very important to involve the PK group within the CRF review process and when database consistency checks are defined. During the conduct of the trial, a detailed specification for the NONMEM dataset should be developed by the PK group. This should identify which dosing and PK/PD data are required, which covariates are required, and what should be done with missing dates and times or missing covariates. The specification should also give details of the tables and listings required to aid the pharmacokineticist in confirming that the data are as expected, and to support their report in the same way tables and listings are usually produced for the CTR. Everything should be designed to ensure that the PK group can receive the NONMEM dataset as soon as possible after the database lock.

Defining a clear process for handling data issues will ensure that they are always treated the same way and allow generic macros to be created for efficiency across studies. Starting programming at least 3 months before the database lock will allow any data issues to be cleaned in time for the database lock, and provide the PK group with an insight into the data before it is locked. Finally, the use of independent programmers can also help to finalize the NONMEM dataset before database lock, so that once the database is locked, the final NONMEM dataset can be produced very quickly.

Reference

1 Collins A, Peterson M, Silva G. Streamlining the PK/PD data transfer process. Pharm Programm 2010; **3**: 24–8.