

# SAS® Macro to Create ADDL Records To Estimate Dosing Not Represented in the Collected Data, In a NONMEM Population Pharmacokinetic Analysis Dataset

Richard Schneck, Mike Heathman, and Lynn Reynolds,  
Eli Lilly and Company, Indianapolis, IN

## ABSTRACT

Datasets created to be input to the NONMEM software for Population Pharmacokinetic (PK) and Pharmacodynamic (PD) analyses are often built from data which captures only the last few doses prior to a blood draw for PK or PD sampling, when in fact dosing occurs at regular intervals leading up to the visit at which the sample is drawn. This appears in the data as a gap in the dosing, which upon analysis would give the false impression that these last few doses alone contributed to the concentration measured in the sample. This situation typically occurs in studies in which a subject or patient is taking oral medication at home between visits. The remedy is to create supplemental dose records, hereinafter referred to as "ADDL records", to estimate the dosing which has occurred between visits but is not in the collected data.

## INTRODUCTION

The programming required to create these ADDL records can be complex and tedious, and the particular coding requirements can vary between studies depending on study design and data collection methods. We present here the culmination of our efforts to create a macro to automate this task in as many cases as possible where ADDL records are required.

Simply stated, the basic function of macro ADDL is to read in a SAS® dataset, which must be in NONMEM format, and add estimated doses in the form of ADDL records wherever necessary to fill in gaps in dosing. In each case where a gap is found, the macro makes a copy of the dose record ending the gap, i.e., the first dose record after the gap. It then transforms this copy into an ADDL record by adding ADDL-specific information, and changing the date and time of the record to the date and time of the earliest of the estimated doses to be represented by the record. This ADDL record represents all of the doses to be estimated in the gap. The resulting dataset is identical to the input dataset except for the addition of these ADDL records and some ADDL-specific variables.

## THE INPUT NONMEM DATASET

The primary input to the macro is a SAS® dataset which is in NONMEM format. Specifically, this means that it contains the variables and records required for NONMEM analysis. It should be noted that the NONMEM software itself does not read SAS® datasets. At some point downstream, we presume that the calling program or some external process will be used to convert the final SAS® dataset to a form that can be directly utilized by NONMEM. At the point in the processing where macro ADDL is invoked, however, the input and output files are SAS® datasets. When we say that the input dataset must be in NONMEM format, we are strictly referring to the required variables and observations that must be included in this dataset.

For purposes of this paper, we use the term "record" to refer to an "observation" or "row" in a SAS® dataset. This is to avoid confusion with the term "observation", which in the context of NONMEM refers to a PK or PD sample measurement.

A key variable in a NONMEM dataset is the "Event-ID" variable. This variable must be in the input dataset, and it must be called "EVID". Variable *EVID* identifies a record in the dataset as being a "dose record" or an "observation record". A record in the

dataset on which *EVID*=0 is an observation record, which represents a PK or PD sample measurement. A record in the dataset on which *EVID*=1 is a dose record, which represents an administered dose. *EVID* may also take on values other than 0 or 1. While these "other" records are carried to the output dataset unchanged, they are not a factor in determining any aspect of creating ADDL records.

The NONMEM dataset does not have distinct variables for date and time of sample vs. date and time of dose. There is one variable for the calendar date of the record and one variable for the time of day. On observation records (*EVID*=0), the values of the date and time variables are the date and time of a PK or PD sample; and on dose records (*EVID*=1), they are the date and time of a dose.

In addition to the variables containing the date and time of the record, the dataset must have variables containing the date and time of the start of therapy, and a variable containing the total time on therapy in hours. The time on therapy in hours is the key to determining the time of the ADDL records. This variable is a decimal number of hours; for example, thirty minutes is the numeric value 0.5 (not 0:30).

The dataset must also have at least one variable containing a dose amount. All variables named as dose variables in the macro call will have their values imputed to the ADDL record from the appropriate dose record.

## THE OUTPUT NONMEM DATASET - WITH ADDL RECORDS

The output NONMEM dataset is a new version of the input dataset with the addition of the ADDL records and a few ADDL-related variables. ADDL records are created wherever there are gaps in the dosing large enough to require the addition of estimated doses. All of the original records in the input dataset are kept in the output dataset.

Each ADDL record represents all of the estimated doses needed to fill one gap in dosing. Macro ADDL creates a variable called *ADDL*, which tells NONMEM the number of estimated doses that each ADDL record represents. The value of variable *ADDL* is equal to the number of estimated doses to be represented by the ADDL record, *minus one*. When NONMEM reads the dataset, it regards the ADDL record as a regular dose record. It then interprets the value of variable *ADDL* as the number of doses to generate in addition to the ADDL record itself. This is why the value of variable *ADDL* is one less than the number of estimated doses to be represented by the ADDL record.

In addition, the output dataset includes a dose-interval variable called "II", which has a constant value equal to the protocol-specified dose interval in hours. When NONMEM generates the estimated doses, it begins with the ADDL record, and generates *ADDL* more doses going forward in increments of *II* hours. For example, if on a given ADDL record we have *ADDL*=10 and *II*=24, this record represents eleven estimated doses, beginning with the ADDL record itself, followed by ten more doses spaced twenty-four hours apart.

## DETERMINING THE TIME OF THE ADDL RECORDS

To reiterate, the method the macro uses to create an ADDL record is to copy the record ending a gap in dosing and change the copy into the ADDL record. A gap in dosing begins with either

a dose record (EVID=1) or an observation record (EVID=0), and ends with a dose record (EVID=1).

The date and time of each ADDL record, and the time on therapy in hours, is set on the ADDL record to be the earliest of the estimated doses being represented by the ADDL record. This computation is driven by the time on therapy in hours.

The method the macro uses to determine the time of the ADDL record is to take the time of the dose record ending the gap and count backwards in time in increments equal to the dose interval, until it is not possible to go back one more interval without getting too close to the previous dose or observation record. In order to determine the time of the ADDL record, and the number of estimated doses it represents, two key parameters must be used: *DOSEINT* and *MINGAP*. Parameter *DOSEINT* is the protocol-specified dose interval in hours. Parameter *MINGAP* establishes a minimum amount of time, expressed in number of dose intervals, between the record beginning the gap and the ADDL record. (Note: *MINGAP* does not have to be an integer.)

Here is an example of how an ADDL record is created and how the date and time are computed. In this example the user sets *DOSEINT*=24 and *MINGAP*=1.

### Example 1

This is an excerpt from the data prior to calling Macro ADDL:

EVID	Time on Therapy (Hrs)	
0 or 1	6.0	Beginning of Gap
---	---	<GAP IN DOSING>
1	120.4	End of Gap

In this example, the subject is dosed at 24-hour intervals. In the data, there is a dose record at 120.4 hours from start of therapy, which follows a period of 114.4 hours since the previous dose or observation. Since the protocol-specified dose interval is 24 hours, this gap of 114.4 hours is large enough to require the creation of an ADDL record.

The first step in calculating the time of the ADDL record is to determine the earliest possible time at which an estimated dose can be represented without violating the user-specified "minimum gap" between the record beginning the gap in dosing and the ADDL record. In the example, the user has specified *MINGAP*=1. This means that a "cushion" of exactly one full dose interval, in this case 24 hours, must be allowed between the ADDL record and the preceding record. Therefore, without even knowing the time of the dose record ending the gap, we know that the earliest possible time of the ADDL record is calculated as follows:

```

Earliest Time
= Beginning Time + (&MINGAP x &DOSEINT)
= 6.0             + (1 interval x 24 hrs)
= 30 hrs

```

The second step is to determine the number of complete dose intervals that fit between the Earliest Time, and the time of the dose record ending the gap:

```

Number of Intervals
= (Ending Time - Earliest Time) / &DOSEINT
= (120.4 - 30.0) / 24
= 3.76 (truncates to 3)

```

If the number from this step is not an integer, it is truncated to the next lower integer value. In the example, there are three complete 24-hr. intervals that fit in the gap.

The third step is to use this number to compute the time of the ADDL record:

```

Time of ADDL record
= Ending Time -
  (Number of Intervals x &DOSEINT)
= 120.4 - (3 x 24)
= 48.4

```

Finally, the macro assigns values to variables *ADDL* and *II*:

```

ADDL = Number of Intervals - 1
      = 3 - 1 = 2
II    = &DOSEINT = 24

```

Here is the data after calling macro ADDL:

EVID	Time on Therapy (Hrs)	ADDL	
0 or 1	6.0	0	Original Rec.
1	48.4	2	ADDL Record
1	120.4	0	Original Rec.

Although it is just one observation in a SAS® dataset, this ADDL record represents three estimated dose records when it is input to NONMEM: the first one at the calculated time of 48.4 hrs., and two more going forward in increments of 24 hrs. (because *ADDL*=2 and *II*=24), at 72.4 hrs and 96.4 hrs. All ADDL records have an EVID value of 1, and are regarded as dose records.

The variables containing the date and time of the record are correctly calculated to agree with the time on therapy on the ADDL record. In the example, if the observation at 6.0 hours occurred on January 1 at 1400 (2:00 pm), the ADDL record is assigned date and time values of January 3 at 0824 (8:24 AM), which is 48.4 - 6.0 = 42.4 hours later.

### DETERMINING THE DOSE AMOUNT ON ADDL RECORDS

Obviously, a key piece of information to include on ADDL records is the dose amount of the estimated doses to be represented by the ADDL record. As mentioned earlier, all variables named as dose variables in the macro call will have their values imputed to the ADDL record from the appropriate dose record. The question is: From which dose record is it appropriate to impute the dose amount?

By the method used to create ADDL records, as discussed, the ADDL record is an exact copy of the record ending the gap in dosing (except for time and ADDL-specific information). In the above example, this is the record on which the time on therapy is 120.4 hrs. Whatever the dose amount is on this record, that is the dose amount on the ADDL record, by default. This is perfectly acceptable in the typical situation where the only dose records in the data are those representing the last few doses prior to the PK sample, assuming the dose amount is constant throughout the gap in dosing, and assuming this is the only type of dose record found in the dataset.

These assumptions, however, are not always valid. There are cases when dose-adjustment records are present in the dataset. These records represent the very first dose at the given dose amount. Whenever one of these dose records appears at the end of a gap in dosing, the default method of copying the dose amount from this dose record to the ADDL record would produce misleading results. The resulting ADDL record would falsely report that the subject was at the newly adjusted dose amount during the gap, when in reality, the subject was at the prior dose amount.

The solution to this problem is to identify dose-adjustment records in the input NONMEM dataset. Macro ADDL expects a variable called "DSEQ", which should have a numeric value of 1 on dose-adjustment records, and a value of zero on records which represent doses where a subject is continuing at the same dose amount. If *DSEQ*=1, indicating a dose-adjustment record, the macro imputes the previous dose amount for the subject to the ADDL record, rather than copy it from the dose adjustment record ending the gap.

If variable *DSEQ* has a value other than 0 or 1 (or is omitted from the dataset entirely), the macro assumes that the record is not a dose-adjustment record and accordingly sets the dose amount on the ADDL records by the default method. When this occurs, the macro prints the appropriate warning message in the SAS® log file to notify the user that this default action has occurred. In order for this logic to work correctly, all variables containing dose-amount information must be included on parameter *DOSEVRS*. Any variables not specified will derive their values by the default method of copying from the dose record ending the gap in dosing.

#### GAPS REMAINING AFTER STANDARD ADDL PROCESSING

It is possible that unwanted gaps in dosing remain in the data after the standard ADDL processing discussed to this point.

By the process described to this point, any dosing gap to be filled by an ADDL record ends with a dose record, but it begins with either a dose or an observation record. The reason for this is that it would not make sense to create an ADDL record at a point in time prior to an observation record which may be intended to be associated with an existing dose record.

This causes a problem when intervening observation records not preceded by a dose record are present in the data. These observations break up the gap in dosing and prevent ADDL records from being created in cases where the subject is known to be taking doses during the gap. See the following example:

#### Example 2

This is an excerpt from the data, prior to calling Macro ADDL, in which there is a dosing gap with an intervening observation record ending the gap:

<u>EVID</u>	<u>Time on Therapy (Hrs)</u>	
0 or 1	6.0	Beginning of Gap
---	---	<GAP IN DOSING>
0	120.4	End of Gap
1	121.4	Subsequent Dose

In this example, no ADDL record would be created, because there is an intervening observation record breaking up the gap in dosing. The solution is to set parameter *FILLGAPS* to one of two valid values: *PREVIOUS* or *NEXT*. This tells the macro to create an ADDL record to fill the gap in dosing as if the observation record ending the gap were a dose record. In this example, therefore, the resulting time of the ADDL record is the same as in Example 1:

<u>EVID</u>	<u>Time on Therapy (Hrs)</u>	<u>ADDL</u>	
0 or 1	6.0	0	Original Rec.
1	48.4	2	ADDL Record
0	120.4	0	Original Rec.
1	121.4	0	Original Rec.

Then, to determine the dose amount on the ADDL record, the macro imputes the dose amount from the most recent or the next dose amount occurring in the data, depending on whether *PREVIOUS* or *NEXT* is specified.

#### ASSIGNING VISIT INFORMATION

Macro ADDL does not use visit numbers to determine any aspect creating ADDL records. By default, any visit numbers or visit date variables included in the data, like all other variables in the data, will have their values copied from the dose record ending the gap in dosing to the ADDL record. This can result in a minor inaccuracy in the dataset: the time associated with the ADDL record may not fall within the visit indicated by the imputed visit information.

To solve this problem, the user may use optional parameters *VISNUM*, *VISSTART*, and *VISSTOP* in any combination to specify the names of the variables containing the visit number, visit start date, and visit stop date, respectively. The macro uses

available visit information already in the dataset to assign the correct visit information to the ADDL records, based on the date and time of the ADDL records.

#### INTERPRETING VARIOUS DATE AND TIME FORMATS

The input NONMEM dataset must have variables containing the date and time of the dose or observation record, and the date and time of the start of therapy. The names of these variables are communicated to the macro via parameters *SAMPDATE*, *SAMPTIME*, *THSTRDAT*, and *THSTRTIM*, respectively.

In recognition of the fact that dates and times come in many forms, the macro allows a wide range of possible formats. The term "format" is not used here in the literal context of recognized SAS® language formats. Rather, the authors have attempted to compile a list of virtually every numeric or character form that is likely to be encountered in clinical data, and set them up as valid choices for the macro.

These choices are selected by setting a valid value for parameter *DATEFMT*, which gives the formats for variables *&SAMPDATE* and *&THSTRDAT*, and parameter *TIMEFMT*, which gives the formats for variables *&SAMPTIME* and *&THSTRTIM*. For example, one valid value for parameter *DATEFMT* is *CCYY-MM-DD*. If *DATEFMT* is set to this value, then the date variables *&SAMPDATE* and *&THSTRDAT* must be in this character form, including the dashes (another choice without the dashes is available). Likewise, if *DATEFMT* is set to *SASDATE*, then variables *&SAMPDATE* and *&THSTRDAT* must be SAS® numeric date values. Similarly, if parameter *TIMEFMT* is set to *HH:MM24*, then the input time variables *&SAMPTIME* and *&THSTRTIM* must be character variables expressing the time in this 24-hour time form, including the colon.

Complete lists of valid values of parameters *DATEFMT* and *TIMEFMT* are given in Appendices 2(a) and 2(b), respectively. Note that the format given by *DATEFMT* applies to both *&SAMPDATE* and *&THSTRDAT*. Both must be in the same format. Likewise, the format given by *TIMEFMT* applies to both variables *&SAMPTIME* and *&THSTRTIM*.

The resulting date and time variables will be similarly formatted on the ADDL records.

#### COMPUTING OTHER DATE AND TIME VALUES

As discussed, the key variable which drives the time calculation on the ADDL records is the variable representing the total running time on therapy in hours, specified by parameter *TIMEHRS*. The macro also requires a date variable and a time variable to be specified by parameters *SAMPDATE* and *SAMPTIME*, and it automatically determines their value on the ADDL records.

It often happens that there is also a variable in the dataset representing the total running time on therapy in days. Since this variable is not automatically re-calculated on ADDL records, if it exists in the dataset, its value will be erroneously copied from the dose record ending the gap. To ensure that it will be correctly calculated on the ADDL records, the user must name the variable using optional parameter *TIMEDAYS*. The macro will then calculate this variable to be in agreement with the other date and time information on the record.

Similarly, optional parameter *DATETIME* can be used to name a SAS® Datetime numeric variable to ensure that it will be correctly calculated on the ADDL records.

#### IGNORING RECORDS

There are two specific cases in which the macro will ignore records in the input dataset. In each of these cases, the ignored records will be carried to the output dataset unchanged, but will not have an impact on any aspect of the creation of ADDL records.

- The macro has the capability of recognizing "commented" records, i.e., records which are to be ignored by NONMEM for any reason. To flag a record as a commented record, set variable CMMT to a character value of "C" in the input dataset prior to calling the macro. If variable CMMT is not in the dataset, the macro will assume that none of the records are commented.
- Any records in the input NONMEM dataset with an EVID value other than 0 or 1 will be ignored.

#### ADDITIONAL NOTES ABOUT MACRO ADDL OUTPUT

Here are a few additional notes about the output dataset generated by macro ADDL:

- When the value of variable *ADDL* is zero on an ADDL record, indicating that exactly one estimated dose is being represented by the record, then the value of *ll* will also be zero. This is a NONMEM requirement.
- Variable *ADDL* will have a value of zero on all non-ADDL records (records from the original dataset). Variable *ll* will have a missing value on all non-ADDL records. The value of variable EVID will be 1 on all ADDL records, since these records are in effect dose records.
- A flag variable called *\_ADDLFLG* will be added to the output dataset to identify the ADDL records. *\_ADDLFLG* will have a value of 1 to indicate ADDL records created to fill a typical gap in dosing which ends with a dose record. *\_ADDLFLG* will have a value of 2 to indicate an ADDL record which was created to fill a gap in dosing which ends with an observation record, as a result of specifying a value for parameter *FILLGAPS*.

#### CONCLUSION

Macro ADDL should be used with care. This macro has been completely tested and validated with various examples of clinical data, and it will work in the manner prescribed if all of the expected conditions are met. However, we recognize that the business of estimating data is not an exact science, and we concede that this macro has a high level of dependency on pre-existing conditions in the input dataset. The macro cannot verify that the variables on which it depends, such as the time on therapy, are correctly calculated in the input dataset. It cannot verify that the date and time variables are in the specified format, and problems can result if undecipherable date or time values are hiding in the data. (For this reason, SAS® numeric date and time values are the most reliable.) The macro cannot verify that all of the dose variables in the dataset have been named on parameter DOSEVRS, or that the value specified on parameter MINGAP is the most appropriate for the dosing schedule, or that the dose records in the data are spaced consistently with the value specified for parameter DOSEINT. This macro relies heavily on clean input data and proper use. Extra caution is advised when processing datasets with dose adjustments, crossover designs, character date and time formats, or any other complicating factors. The user should review the results carefully to verify that the resulting ADDL records have been created as the user intended.

Nevertheless, macro ADDL has proven most valuable to us in creating NONMEM datasets. It has eliminated many hours of custom coding, and it has proven flexible for a variety of study designs.

#### APPENDIX 1: MACRO ADDL PARAMETERS

INDATA	The name of the Input SAS® dataset in NONMEM format. (Required)
PATKEY	The list of variables defining the unique key identifying the subject or patient. (Required)

DOSEVRS	The list of variables containing dose amount variables to be imputed to the ADDL records. (Required)
SAMPDATE	The variable containing the record date (date of dose, sample, or other). Must be in the format specified by parameter DATEFMT. (Required)
THSTRDAT	The variable containing the date of the start of therapy (subject's first dose). Must be in the format specified by parameter DATEFMT. (Required)
DATEFMT	The format of the date variables specified by parameters SAMPDATE and THSTRDAT. (See APPENDIX 2(a)) (Required)
SAMPTIME	The variable containing the record time (time of dose, sample, or other). Must be in the format specified by parameter TIMEFMT. (Required)
THSTRTIM	The variable containing the time of the start of therapy (subject's first dose). Must be in the format specified by parameter TIMEFMT. (Required)
TIMEFMT	The format of the time variables specified by parameters SAMPTIME and THSTRTIM. (See APPENDIX 2(b)) (Required)
TIMEHRS	The variable containing the total running time since start of therapy in hours. Must be a numeric decimal value. (Required)
TIMEDAYS	The variable containing the total running time since start of therapy in days. Must be a numeric decimal value. (Optional)
DATETIME	The combined date/time variable. Must be a valid SAS® Datetime numeric variable. (Optional)
DOSEINT	The protocol-specified interval between doses in hours. (Required)
MINGAP	The required minimum number of intervals from the previous sample or dose time to the time assigned to the ADDL record. For example, if the previous sample or dose is at 48 hrs from start of therapy, and MINGAP is set to 1 for a 24-hr dose interval, then the time of the ADDL record must be at least 72 hrs from start of therapy. NOTE: does not have to be an integer. (Default=1)
FILLGAPS	The option to create additional ADDL records to fill in gaps between observation records where there is no intervening dose record, but the subject or patient is known to have been taking regular doses during the gap. Valid values: PREVIOUS - Impute the previous dose amount NEXT - Impute the subsequent dose amount (optional)
VISNUM	The variable containing the visit number. (optional)
VISSTART	The variable containing the visit start date. (optional)
VISSTOP	The variable containing the visit stop date. (optional)

#### APPENDIX 2(a):

##### VALUES OF PARAMETER DATEFMT

Following are valid values of parameter DATEFMT, which is the expected input format of the date variables specified by parameters SAMPDATE and THSTRDAT. Following each value is a sample dataset value equivalent to the date Jan 31, 2002.

SASDATE	15371	(SAS® numeric date)
CCYYMMDD	20020131	(numeric or character)
CCYY/MM/DD	2002/01/31	(character)

CCYY-MM-DD	2002-01-31	(character)
YYMMDD	020131	(numeric or character)
YY/MM/DD	02/01/31	(character)
YY-MM-DD	02-01-31	(character)
MMDDCCYY	01312002	(numeric or character)
MM/DD/CCYY	01/31/2002	(character)
MM-DD-CCYY	01-31-2002	(character)
MMDDYY	013102	(numeric or character)
MM/DD/YY	01/31/02	(character)
MM-DD-YY	01-31-02	(character)
DDMONYY	31JAN02	(character)
DDMONCCYY	31JAN2002	(character)
DDMMCCYY	31012002	(numeric or character)
DD/MM/CCYY	31/01/2002	(character)
DD-MM-CCYY	31-01-2002	(character)
DDMMYY	310102	(numeric or character)
DD/MM/YY	31/01/02	(character)
DD-MM-YY	31-01-02	(character)

## APPENDIX 2(b):

### VALUES OF PARAMETER *TIMEFMT*

Following are valid values of parameter *TIMEFMT*, which is the expected input format of the time variables specified by parameters *SAMPTIME* and *THSTRTIM*. Following each value is a sample dataset value equivalent to the time 1:00 PM.

SASTIME	46800	(SAS® numeric time):
HHMM24	1300	(numeric or character)
HH:MM24	13:00	(character)
HHMMSS24	130000	(numeric or character)
HH:MM:SS24	13:00:00	(character)
HHMM12	0100 PM or 100PM	(character)
HH:MM12	01:00PM or 1:00 PM	(character)
HHMMSS12	010000 PM or 10000PM	(character)
HH:MM:SS12	01:00:00PM or 1:00:00 PM	(character)

## REFERENCES

- NONMEM Users Guides, (1989-98). Beal, S.L. and Sheiner L.B. (Eds.) GloboMax LLC, Hanover, MD.
- SAS® Language, Reference Version 6, First Edition, Copyright® 1990 by SAS Institute Inc., Cary, NC, USA

## ACKNOWLEDGMENTS

The authors wish to thank John Labore, Associate Information Consultant, and Fred Standeford, Pharmacokinetics Team Leader, both of Eli Lilly and Company, for technical, procedural, and moral support.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Author Name	Richard Schneck
Company	Eli Lilly and Company
Address	Lilly Corporate Center DC 0734
City state ZIP	Indianapolis, IN 46285
Work Phone:	(317) 433-1721
Fax:	(317) 276-8366
Email:	RLSCHNECK@LILLY.COM

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

