

Calculating time-to-event parameters using a single DATA step and a RETAIN statement

Andrew L Hulme, PRA, Lenexa, KS

ABSTRACT

During oncology studies, a typical measure of efficacy is the amount of time elapsed until a particular response is achieved for the first time (time to response) and how long that response is maintained (duration of response). For subjects that do not meet the response criteria before the end of the study, a censoring flag and time must also be calculated. Thus, typical methods of calculating these parameters use several programming steps for each particular variable of interest. This paper presents a method to calculate all necessary parameters in a single DATA step using a RETAIN statement. A subject's records are ordered by date and the response criteria are evaluated during each iteration of the DATA step. If the criteria are met the time-to-event values are passed to the retained variables and kept until a subject's last record; if not, the censored time is calculated on the last record and outputted.

INTRODUCTION

Time-to-event parameters are an important component in determining the safety and efficacy of a drug during a clinical trial. Since the derivation of these parameters typically utilizes multiple data points across several domains and time points, calculations can be complex and difficult to program. An additional complication is the inclusion of a censoring variable that is used for subjects that do not meet the criteria of the event. This paper presents a simple method to calculate a variety of time-to-event parameters using only a single DATA step with a RETAIN statement.

One event of interest in a large percentage of studies is the response of the subject to the study medication. Response is usually collected on a case report form (CRF) at the end of each treatment cycle and is separated into several different levels (e.g. partial remission, complete remission). It may also depend on other parameters like tumor size and laboratory assessments. Each study will have slightly different criteria for when a positive response¹ is achieved. Nonetheless, the most important parameter to calculate is time to response (TTR) which is defined as the time elapsed between some starting point at the beginning of the study, usually date of first dosing, and the day on which the positive response was first achieved. Another important parameter is the duration of response (DOR) which is the time between the date of the first positive response and relapse. Additional parameters of interest may be the best level of response achieved and the time to each individual level of response.

Normally, not all subjects in a study achieve a positive response, but most analyses (e.g. Kaplan–Meier) require that the days spent in the study by the non-responding subjects are counted. This is accomplished by censoring subjects on the last day they were in the study (or some other end point) and giving them a time-to-event duration equal to the number of days from first dosing to the censoring date. A binary censoring variable is then determined with one value for those that met the response criteria and the opposite value for those that are censored. Note that the values for this variable will typically be 0 and 1 but which value is given to which set of subjects varies (in this paper, a value of 1 will be given to those with the event and a value of 0 to those who were censored). Thus, the end result of deriving a time-to-event parameter should be a data set containing the time-to-event duration variable and the censor variable for each subject.

DESCRIPTION OF METHOD

Table 1 displays data of a typical oncology study with response collected at the end of each cycle or about every four weeks. Investigators indicate on the CRF whether the subject has complete response (CR), partial response (PR), clinical improvement (CI), stable disease (SD), or progressive disease (PD). Relapse can also be indicated if a positive response has already been achieved. For this study, a positive response is defined as any instance of CR, PR, or CI and time-to-response is defined as the time from the date of first dose to the first instance of CR, PR or CI.

To calculate this parameter by hand, one would look at a subject's first cycle and determine if the criteria for a positive response have been achieved (Figure 1). If yes, the time-to-event duration would be calculated and, for the given subject, the derivation would stop. If not, the process would start again with the next cycle continuing in chronological order until either a positive response is found or a subject's reaches the end of the study. At the end of

¹ The term response can be used to describe two distinct things: 1.) the data point collected on the CRF or derived from other criteria that is a representation of the effect of the drug (e.g. complete remission, progressive disease), 2.) the period in which a subject meets the study's pre-defined criteria for determining a positive impact of the drug. This paper will use positive response when referring to the second definition.

USUBJID	PESEQ	PETESTCD	PEORRES	PESTRESC	VISIT	PEDTC	PEDY
ABC-XYZ-002	1	RESPONSE	SD	Stable Disease	End of Cycle 1	2012-01-31	31
ABC-XYZ-002	2	RESPONSE	CI	Clinical Improvement	End of Cycle 2	2012-02-26	57
ABC-XYZ-002	3	RESPONSE	PR	Partial Remission	End of Cycle 3	2012-03-25	85
ABC-XYZ-002	4	RESPONSE	PR	Partial Remission	End of Cycle 4	2012-04-22	113
ABC-XYZ-002	5	RESPONSE	CR	Complete Remission	End of Cycle 5	2012-05-21	142
ABC-XYZ-002	6	RESPONSE	RELAPSE	Relapse	End of Cycle 6	2012-06-19	171
ABC-XYZ-054	1	RESPONSE	SD	Stable Disease	End of Cycle 1	2012-01-29	29
ABC-XYZ-054	2	RESPONSE	SD	Stable Disease	End of Cycle 2	2012-02-26	57
ABC-XYZ-054	3	RESPONSE	SD	Stable Disease	End of Cycle 3	2012-03-25	85
ABC-XYZ-054	4	RESPONSE	PD	Progressive Disease	End of Cycle 4	2012-04-21	112
ABC-XYZ-074	1	RESPONSE	CI	Clinical Improvement	End of Cycle 1	2012-01-27	27
ABC-XYZ-074	2	RESPONSE	CI	Clinical Improvement	End of Cycle 2	2012-02-26	57
ABC-XYZ-074	3	RESPONSE	CI	Clinical Improvement	End of Cycle 3	2012-03-28	88
ABC-XYZ-074	4	RESPONSE	CI	Clinical Improvement	End of Cycle 4	2012-04-22	113
ABC-XYZ-074	5	RESPONSE	CI	Clinical Improvement	End of Cycle 5	2012-05-17	138

Table 1. Typical response data for an oncology study. Note that only variables that are pertinent are displayed. First dose date and RFSTDTC for all subjects is 2012-01-01.

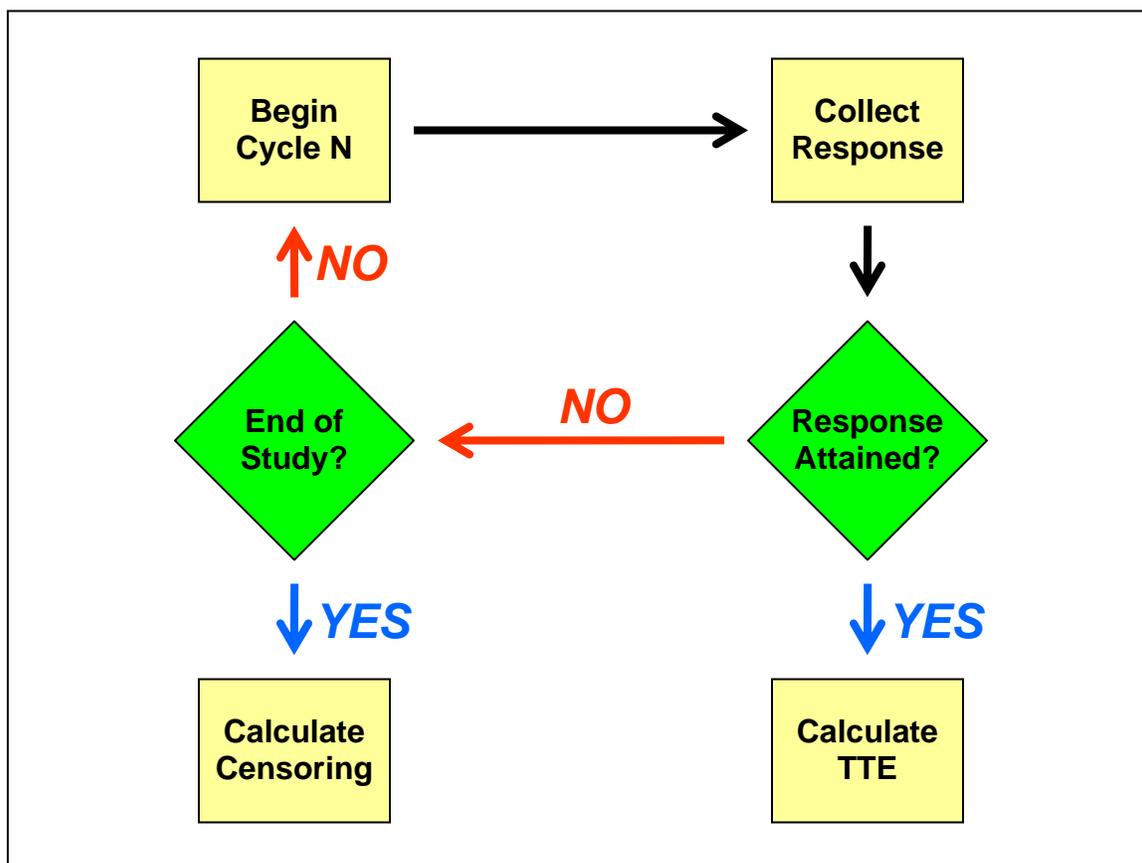


Figure 1. Schematic for calculating response.

the study, if no positive response had been found, the censoring rules would be used to calculate a censored time-to-event for that subject. This process would be repeated for all subjects in the population being analyzed.

The method described in this paper mimics the above calculation since searching record-by-record in chronological order for a response resembles the single record at a time input property of a DATA step. Instead of stopping the calculation when a positive response is found, the values are passed to variables initialized by a RETAIN statement to keep them until a subject's last record. The end of study record is found by using a LAST variable and if no response is found by this time, the appropriate censoring time is calculated. Lastly, upon reaching this final record an OUTPUT statement creates one record for each subject with the time-to-response and censor variable values.

EXAMPLES

1. BASIC TIME-TO-RESPONSE EXAMPLE

This example uses the data in Table 1 ordered by subject number and date of collection. The first step is to initialize the time-to-response variable `TIMERESP` and set it to missing when a subject's first record is input. Next, for each record the variable containing the response information (`PEORRES`) is checked to see if the criteria for a positive response are met. If this condition is true, time to response is calculated as the difference in days between the date of collection (`PEDTC`) and the reference start date (`RFSTDTC`). Note that the check for missing `TIMERESP` ensures that the calculation is done only the first time a response is found and that the step will be skipped once `TIMERESP` is calculated. The program will proceed to check for a positive response until it reaches a subject's last record. At this point, if a positive response has not been found and hence `TIMERESP` will still be missing, then the censored time-to-response will be calculated. In this case, the last collection date of response data is used as the censoring date but the rules for the censoring calculation will vary by study. Additionally, the censoring variable `CENSRESP` is given a value of 0 for the subjects that do not have a response. If `TIMERESP` was calculated earlier in the program due to a positive response, the censoring variable gets a value of 1. The last step is to output one record per subject with the time-to-response and censor variables.

```
data ttel(keep=usubjid timeresp censresp);
  set pe;
  by usubjid pedtc;

  /* 1. INITIALIZE RETAIN VARIABLES */
  retain timeresp;
  if first.usubjid then timeresp=.;

  /* 2. EVALUATE RESPONSE CRITERIA */
  if ( peorres in("CR","PR","CI") ) and timeresp=. then
    timeresp=input(pedtc,yymmdd10.)-input(rfstdtc,yymmdd10.);

  /* 3. CALCULATE CENSORED TIME ON LAST RECORD */
  if last.usubjid then do;
    if timeresp=. then do;
      timeresp=input(pedtc,yymmdd10.)-input(rfstdtc,yymmdd10.);
      censresp=0; end;
    else censresp=1;
  end;

  /* 4. OUTPUT LAST RECORD WITH PARAMETERS */
  if last.usubjid then output;
run;
```

The end result of this DATA step is displayed in Table 2. Note that as expected subjects 002 and 074 have a censoring variable of 1 since they achieved a positive response, whereas subject 054 has a censoring variable of 0 since no positive response was attained. A data set formatted in this way can be passed easily into the LIFETEST procedure to be used for Kaplan–Meier analysis using `TIMERESP*CENSRESP(0)` for the TIME option.

<u>USUBJID</u>	<u>TIMERESP</u>	<u>CENSRESP</u>
ABC-XYZ-002	56	1
ABC-XYZ-054	111	0
ABC-XYZ-074	26	1

Table 2. Results of Example 1.

2. CALCULATING SEVERAL VARIABLES

While time-to-response may be the most important parameter to calculate, additional measures regarding the efficacy of a study drug may be the time to the first instance of each of the individual response levels. Using arrays, the method described above can be adapted to calculate several parameters at once. The single response evaluation step in Example 1 is replaced by a DO loop that calculates a different time-to-response value during an iteration of the loop. Everything else is the same with the duration and censoring variables being replaced by the appropriate array values from the array declaration statement. Note that since a subject meeting the criteria for a more complete response would also meet the criteria for the less complete response, the response criteria for each level of response includes the responses more complete than it (i.e. a subject that has a first positive response of CR will also be counted as having a PR and CI response on that same date).

```

data tte2(keep=studyid usubjid time: cens:);
  set pe;
  by usubjid pedtc;

  retain timecr timepr timeci;
  array time{3} timecr timepr timeci;
  array cens{3} censcr censpr censci;

  do i=1 to 3;
    if first.usubjid then time{i}=.;

    if i=1 and ( peorres in("CR") ) and time{i}=. then
      time{i}=input(pedtc, yymmdd10.)-input(rfstdtc, yymmdd10.);
    if i=2 and ( peorres in("CR","PR") ) and time{i}=. then
      time{i}=input(pedtc, yymmdd10.)-input(rfstdtc, yymmdd10.);
    if i=3 and ( peorres in("CR","PR","CI") ) and time{i}=. then
      time{i}=input(pedtc, yymmdd10.)-input(rfstdtc, yymmdd10.);

    if last.usubjid then do;
      if time{i}=. then do;
        time{i}=input(pedtc, yymmdd10.)-input(rfstdtc, yymmdd10.);
        cens{i}=0; end;
      else cens{i}=1;
    end;
  end;

  if last.usubjid then output;
run;

```

The end result is displayed in Table 3. Note that TIMECI has the same response criteria as the time-to-response calculated in Example 1; hence TIMERESP is not repeated here.

USUBJID	TIMECR	CENSCR	TIMEPR	CENSPR	TIMECI	CENSCI
ABC-XYZ-002	141	1	84	1	56	1
ABC-XYZ-054	111	0	111	0	111	0
ABC-XYZ-074	137	0	137	0	26	1

Table 3. Results of Example 2.

3. DURATION OF RESPONSE

Another important parameter in determining the efficacy of a study drug is how long a positive response is maintained once it is achieved. This parameter, called duration of response, is defined as the number of days between the date when response was achieved and the date when the subject either relapses or ends the study. For the method presented here, this parameter introduces an additional complication because the starting point for the duration is not the starting date of the study but the day the positive response is achieved. Thus in the below example, not only is the duration of response variable TIMEDOR added to the RETAIN statement, but the date of first response is also retained (RESPDTC). This response date is then used to calculate duration of response when the criteria for a relapse are reached. If no relapse was recorded, the duration is calculated similarly with the last collection date being used as a censoring date and the censoring variable being given a value of 0. Note that if a subject did not have a

positive response during the study, their duration of response will be missing, hence the check for CENSRESP equal to 1 before calculating TIMEDOR.

```

data tte3(keep=studyid usubjid timeresp censresp timedor censdor);
  set pe;
  by usubjid pedtc;

  retain timeresp timedor respdtc; length respdtc $ 19;
  if first.usubjid then do; timeresp=.; timedor=.; respdtc=""; end;

  if ( peorres in("CR","PR","CI") ) and timeresp=. then do;
    timeresp=input(pedtc,yymmdd10.)-input(rfstdtc,yymmdd10.);
    respdtc=pedtc; end;

  if ( timeresp ne . and peorres="RELAPSE" ) and timedor=. then
    timedor=input(pedtc,yymmdd10.)-input(respdtc,yymmdd10.);

  if last.usubjid then do;
    if timeresp=. then do;
      timeresp=input(pedtc,yymmdd10.)-input(rfstdtc,yymmdd10.);
      censresp=0; end;
    else censresp=1;

    if censresp=1 and timedor=. then do;
      timedor=input(pedtc,yymmdd10.)-input(respdtc,yymmdd10.);
      censdor=0; end;
    else if timedor ne . then censdor=1;
  end;

  if last.usubjid then output;
run;

```

The end result is displayed in Table 4. Note that as expected subject 054 has missing values for DOR since the subject never had a positive response.

<u>USUBJID</u>	<u>TIMERESP</u>	<u>CENSRESP</u>	<u>TIMEDOR</u>	<u>CENSJOR</u>
ABC-XYZ-002	56	1	114	1
ABC-XYZ-054	111	0	.	.
ABC-XYZ-074	26	1	111	0

Table 4. Results of Example 3.

4. OUTPUTTING ADTTE

The above examples produce a data set that contains one record per subject. While output like this is easy to input into the LIFETEST procedure, it does not meet the new CDISC standards for the Analysis Data Model (ADaM), namely the Analysis Time-to-Event (ADTTE) standard. This method can be adapted to create a compliant data set by simply moving the output statement into the last DO loop and performing a transpose. The time-to-event durations are set to AVAL and the censoring variables become CNSR. The start and end dates of each duration are also included as STARTDT and ADT respectively. Lastly, the values of the censoring variables are used to populate the event description variable EVNTDESC.

```

data adtte(keep=studyid usubjid param paramcd aval startdt adt cnshr evntdesc);
  set pe;
  by usubjid pedtc;
  length param $ 20 paramcd $ 8 aval startdt adt cnshr 8 evntdesc $ 20;
  format startdt adt date9.;

  retain timeresp timedor respdtc relpsdte; length respdte relpsdte $ 19;
  if first.usubjid then do; timeresp=.; timedor=.; respdte=""; relpsdte=""; end;

  if ( peorres in("CR","PR","CI") ) and timeresp=. then do;
    timeresp=input(pedtc,yymmdd10.)-input(rfstdtc,yymmdd10.);
    respdte=pedtc; end;

```

```

if ( timeresp ne . and peorres="RELAPSE" ) and timedor=. then do;
    timedor=input(pedtc,yymmdd10.)-input(respdte,yymmdd10.);
    relpsdte=pedtc; end;

if last.usubjid then do;
    if timeresp=. then do;
        timeresp=input(pedtc,yymmdd10.)-input(rfstcdte,yymmdd10.);
        censresp=0; end;
    else censresp=1;

    if censresp=1 and timedor=. then do;
        timedor=input(pedtc,yymmdd10.)-input(respdte,yymmdd10.);
        censdor=0; end;
    else if timedor ne . then censdor=1;

    if timeresp ne . then do;
        param="Time to Response"; paramcd="TTR"; aval=timeresp;
        cnshr=censresp; startdt=input(rfstcdte,yymmdd10.);
        adt=input(coalescec(respdte,pedtc),yymmdd10.);
        if censresp=1 then evntdesc="Response Achieved";
        else if censresp=0 then evntdesc="End of Study";
        output; end;
    if timedor ne . then do;
        param="Duration of Response"; paramcd="DOR"; aval=timedor;
        cnshr=censdor; startdt=input(respdte,yymmdd10.);
        adt=input(coalescec(relpsdte,pedtc),yymmdd10.);
        if censdor=1 then evntdesc="Relapse";
        else if censdor=0 then evntdesc="End of Study";
        output; end;
end;
run;

```

The end result is displayed in Table 5. Note that two records are created for both subjects that had a positive response, but the duration of response record was not created for subject 054 because TIMEDOR was missing in the previous example.

USUBJID	PARAMCD	PARAM	AVAL	CNSR	ADT	STARTDT	EVNTDESC
ABC-XYZ-002	TTR	Time to Response	56	1	26FEB2012	01JAN2012	Response Achieved
ABC-XYZ-002	DOR	Duration of Response	114	1	19JUN2012	26FEB2012	Relapse
ABC-XYZ-054	TTR	Time to Response	111	0	21APR2012	01JAN2012	End of Study
ABC-XYZ-074	TTR	Time to Response	26	1	27JAN2012	01JAN2012	Response Achieved
ABC-XYZ-074	DOR	Duration of Response	111	0	17MAY2012	27JAN2012	End of Study

Table 5. Results of Example 4.

5. RECIST CRITERIA

This method can be adapted to calculate several different complex time-to-event parameters. One of these is calculating RECIST criteria which are unique because a subject must have two instances of a CR response eight weeks apart to qualify as a CR (note that this example is highly simplified). This example is nearly identical to Example 1 except that a condition is added requiring another CR at least 56 days before the confirmatory CR. To accomplish this the first CR date FCRDTC is added to the RETAIN statement and the variable is given a value when the first CR response is found. The first CR date is then used to check that there was a prior CR and is compared with the next collection date to ensure that 56 days have passed.

```

data recist(keep=studyid usubjid timecr cnshr);
    set recist0;
    by usubjid pedtc;

    retain timecr fcrdte; length fcrdte $ 19;
    if first.usubjid then do; timecr=.; fcrdte=""; end;

```

```
if ( peorres="CR" and fcrdtc ne "" and
      input (pedtc, yymmdd10.) - input (fcrdtc, yymmdd10.) >= 56 ) and timecr=.
  then timecr=input (pedtc, yymmdd10.) - input (rfstdtc, yymmdd10.);

if peorres="CR" then fcrdtc=pedtc;
  else fcrdtc="";

if last.usubjid then do;
  if timecr=. then do;
    timecr=input (pedtc, yymmdd10.) - input (rfstdtc, yymmdd10.);
    censcr=0; end;
  else censcr=1;
end;

if last.usubjid then output;
run;
```

CONCLUSION

Further examples for using the method described in this paper include survival analysis where the survival information is collected in a scheduled manner on the CRF and time to adverse event parameters where the first instance of a set of adverse events is the event in question. More complex event criteria can be used by adding additional conditions to the evaluation step as shown in Example 5. Additionally, the preferred method of creating ADTTE for analyses that require data from multiple domains is to put all of the information into one data set ordered by date. This input data set could be run through this method to easily create ADTTE.

Typical methods for calculating time-to-event parameters require several complicated steps to ensure a result. This paper presents a simple method that can calculate several time-to-event parameters in one step and is easily-adaptable to several different kinds of input data. By leveraging the iteration characteristic of a DATA step, this method mimics how time-to-event parameters would be calculated by hand allowing one to ensure that the parameters have been calculated correctly.

REFERENCES

Brucken, Nancy, Sandra Minjoe, and Mario Widel, Using the ADaM Basic Data Structure for survival analysis. *PharmaSUG 2011, Nashville, TN, May 2011, TU09*. Available at <http://www.pharmasug.org/proceedings/2011/TU/PharmaSUG-2011-TU09.pdf>.

RECOMMENDED READING

The ADaM Basic Data Structure for Time-to-Event Analyses. Available at cdisc.org.

CONTACT INFORMATION

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

Name: Andrew L Hulme
Enterprise: PRA
Address: 9755 Ridge Drive
City, State ZIP: Lenexa, KS 66219
Work Phone: 913-410-2083
E-mail: hulmeandrew@praintl.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.