PharmaSUG 2019 - Paper DV-323

# Fine-tuning your swimmer plot: another example from oncology

Steve Almond, Bayer Inc.


## ABSTRACT

Swimmer plots are an effective graphical presentation of longitudinal data such as periods of treatment/observation, occurrences of events, and the durations of events or subject status. These types of plots are particularly popular for oncology trials, where the treatment and follow-up periods are displayed along with tumor assessment results and the duration of various responder criteria. The ease of creating swimmer plots has increased in successive SAS versions within the ODS Graphics procedures. This paper reviews the basics for constructing the elements of such plots, and provides tips and tricks for implementing aesthetic touches simply with the main SGPLOT procedure (SAS 9.4) statements and without the use of annotations or the Graphics Template Language.


## INTRODUCTION

The general idea of a swimmer plot is to show subject-level data and events along a unidimensional timeline. The parallel presentation of subjects resembles the racing lanes of a swimming pool, hence the name of the plot.

The concept can accommodate many types of data. Subject disposition events can be plotted to show the administrative progression through different phases / epochs of a study. A natural use for safety evaluation could display the on-treatment time of a subject and the occurrences of adverse events of interest, as well as indicating severity and outcomes.

A popular application of swimmer plots is for oncology studies. There can be numerous endpoints to evaluate the efficacy of a treatment and this type of plot allows for a quick, visual representation of the data that may be easier to assess than plain listings of values. Creating such plots has been the topic of several papers and presentations (Phillips, 2014; Mantage, 2014; Huang, 2016; Wang, 2016).

The purpose of this paper is to not only demonstrate the ease with which basic swimmer plots can now be created in recent versions of SAS (i.e., 9.4), but also to provide tips for fine-tuning the aesthetics without using relatively more complicated concepts such as annotations or the Graphics Template Language (GTL). Rather, enhancements can be achieved simply through a combination of data set design and available SGPLOT functionality. Many of the tips will address complications that arise when working with interim data analyses.

Efficient and straightforward programs are important within the pharmaceutical industry. Decision making based on clinical trial data is occurring earlier and more frequently over the course of a study. Data needs to be presented effectively and with shorter turnaround times. The concepts discussed in this paper require only familiarity with data set processing and simple SGPLOT statements and options. While many powerful features are available by using annotations or GTL, they can be more time-consuming to implement and are far less familiar to programmers which is problematic for timely validation.

In the motivating example for this paper, a swimmer plot will be created in steps and will ultimately include the following data:

- Duration of treatment; indication if ongoing
- Duration of active follow-up; indication if ongoing
- Start / end / duration of response; indication if ongoing
- Best overall response
- Indicator of durable disease control response

## CONTRUCTING THE LANES OF A SWIMMER PLOT

### TREATMENT DURATION

The initial backbone of the swimmer plot is to display the treatment duration for each subject. Table 1 shows part of a data set used for this purpose, with variables for subject ID, treatment start and end days, and one for ordering of the subjects. Generally, subjects are ordered by increasing length of their treatment duration (i.e., treatment end day here, since all treatment starts on day 1).

| SUBJID | TRTSTDY | TRTENDY | ORDER |
|---|---|---|---|
| 1002 | 1 | 28 | 1 |
| 1010 | 1 | 28 | 2 |
| 1008 | 1 | 35 | 3 |
| 1003 | 1 | 56 | 4 |
| | ⋮ | | |
| 1011 | 1 | 210 | 13 |
| 1014 | 1 | 270 | 14 |
| 1005 | 1 | 290 | 15 |

**Table 1 Data Set for Plotting Treatment Duration**

The basic starting code uses the HIGHLOW statement of the SGPLOT procedure to plot bars representing the treatment duration:

```
PROC SGPLOT DATA=swimmer;
  /* bars for treatment duration */
  HIGHLOW Y=order LOW=trtstdy HIGH=trtendy / LOWLABEL=subjid TYPE=BAR;

  /* axes */
  XAXIS LABEL='Study Day' VALUES=(0 TO 300 BY 30);
  YAXIS LABEL='Subject ID' DISPLAY=(NOTICKS NOVALUES);
RUN;
```

By specifying the Y argument with the ordering variable, this will give horizontal bars. The treatment start and end days are assigned to the LOW and HIGH arguments, respectively. The LOWLABEL option uses a data set variable whose values are displayed at the lower ends of the bars; here we use it to label each bar with the subject ID. Since the default action of HIGHLOW is to display lines, we use the TYPE option to display bars instead. The resulting plot is shown in Figure 1.

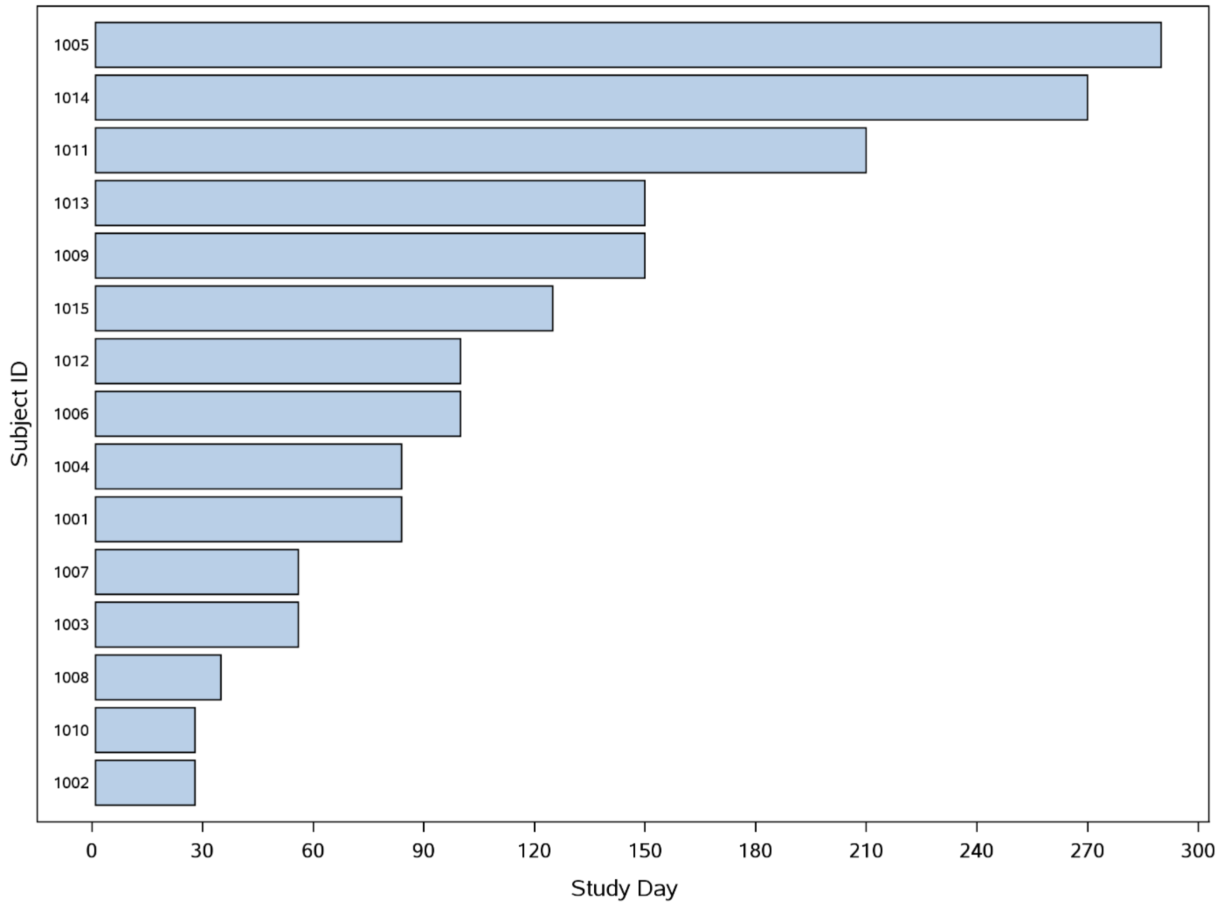


**Figure 1   Treatment Duration**

| **Tips:** |
|---|
| • When individual subject-level data is plotted it is often desirable to be able to identify a particular subject, hence the inclusion of subject IDs in the plot. The placement at the lower ends of the bars (as opposed to using the HIGHLABEL option) provides a consistent and aligned way of displaying this information. |
| • Setting the NOTICKS and NOVALUES options for the y-axis suppresses display of the tick marks and values of the ORDER variable. The subject IDs are sufficient labelling of the bars. |
| • The increment of the x-axis is set to 30 days, which makes for a natural set of values approximately corresponding to monthly intervals. |

Data from oncology studies are often analyzed at interim time points and so there will be a mixture of subjects who have discontinued and those who are still ongoing with treatment. An expected request for a plot is to clearly indicate the latter. The HIGHLOW statement has an option that is particularly suited for this: a keyword – or variable whose value contains a keyword – can be assigned to HIGHCAP which specifies how the higher end of the plot will be capped. For bars, this essentially provides the option of capping with an arrowhead which serves as a clear indicator of ongoing treatment. Since the treatment status will vary from subject to subject, we use the HIGHCAP option with a data set variable which is populated only for subjects with ongoing treatment. In Table 2, this variable is TRTCAP and is used in the code that immediately follows with results shown in Figure 2.

| SUBJID | TRTSTDY | TRTENDY | ORDER | TRTCAP |
|---|---|---|---|---|
| 1002 | 1 | 28 | 1 | |
| 1010 | 1 | 28 | 2 | |
| 1008 | 1 | 35 | 3 | |
| 1003 | 1 | 56 | 4 | FilledArrow |
| ⋮ | | | | |
| 1011 | 1 | 210 | 13 | |
| 1014 | 1 | 270 | 14 | |
| 1005 | 1 | 290 | 15 | FilledArrow |

**Table 2 Data Set Including Cap Indicating Ongoing Treatment**

```
PROC SGPLOT DATA=swimmer;
  /* bars for treatment duration */
  HIGHLOW Y=order LOW=trtstdy HIGH=trtendy / LOWLABEL=subjid TYPE=BAR
    HIGHCAP=trtcap;

  /* axes */
  XAXIS LABEL='Study Day' VALUES=(0 TO 300 BY 30);
  YAXIS LABEL='Subject ID' DISPLAY=(NOTICKS NOVALUES);
RUN;
```

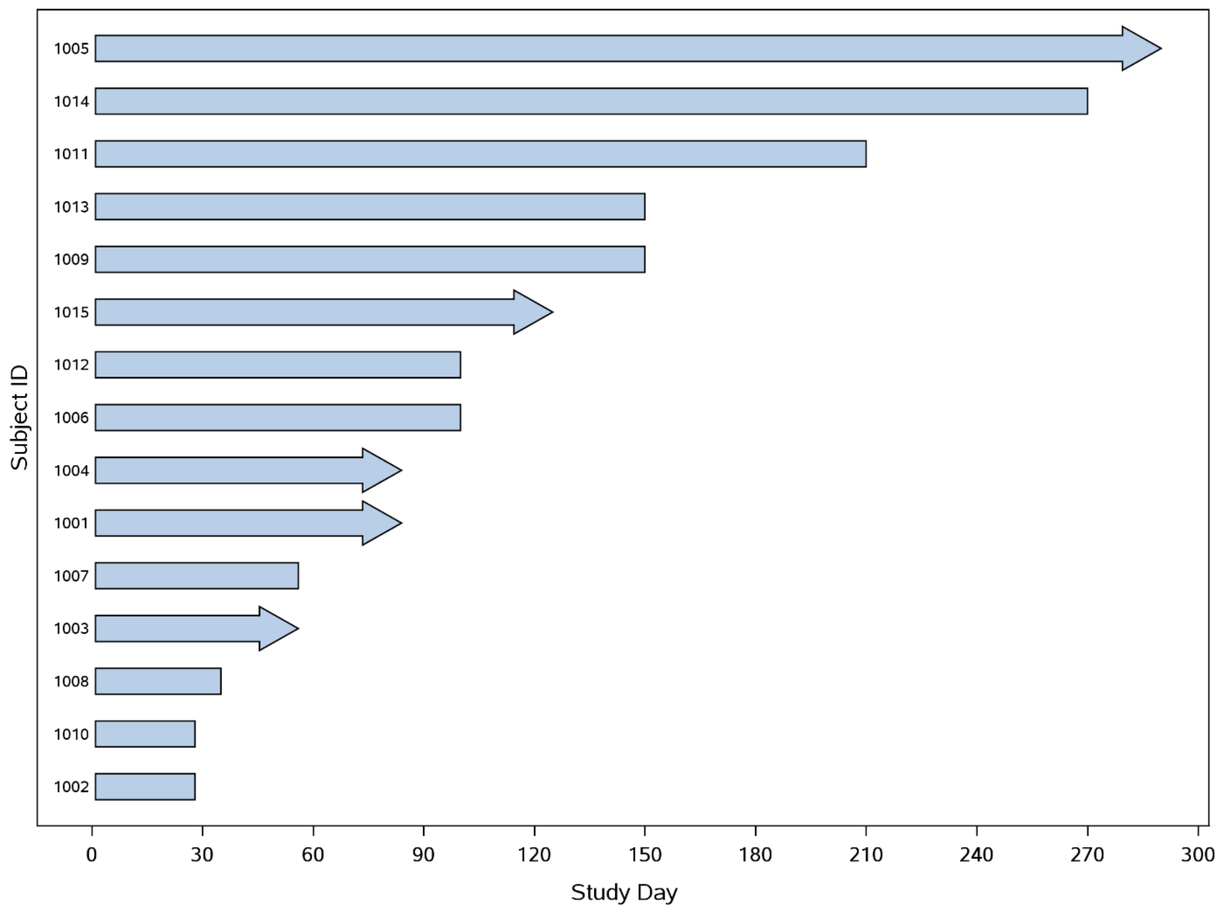


**Figure 2 Treatment Duration and Ongoing Indicator**

The arrowhead caps fit within the width of the original bars and so the widths of the updated bars are reduced accordingly. As seen in Figure 1, there was already some space between the bars; this is because the default setting of the HIGHLOW statement is for bars to occupy 85% of the available width of the “lane.”

**Tips:**

- The bar width can be increased to a value that reduces extraneous space between bars but still allows for some separation should adjacent arrowheads coincide (e.g., BARWIDTH=0.95). The difference, although minor, is shown in Figure 3 with the default (0.85) and in Figure **4** with increased width (0.95).

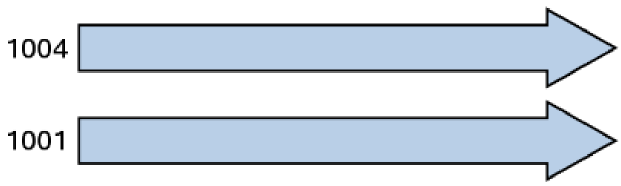


**Figure 3 Default Bar Width**

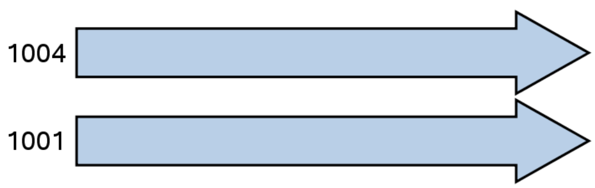


**Figure 4 Increased Bar Width**

- Although not explicitly shown in the data set of Table 2, it is assumed there is a variable available to indicate whether treatment is still ongoing at the time of TRTENDY. If so, then during data set programming the variable TRTCAP is set accordingly.

## ACTIVE FOLLOW-UP

In some study designs, subjects continue to have tumor assessments even after study drug has been discontinued. This would be the case if the reason for discontinuation is something other than disease progression, e.g., due to an adverse event or some administrative reason. Since the subject’s disease has not worsened by this time, there may in fact be residual benefit after the end of treatment. This period of continued assessment is referred to as “active follow-up” (AFU).

In the swimmer plot, we would like to include this period of AFU but clearly distinguish it from the on-treatment period. Table 3 shows an expanded data set with variables for start and end day of AFU; missing values indicate that the subject did not have an AFU period. Similar to the treatment period, there is a variable AFUCAP that will be used to indicate that AFU is ongoing.

| SUBJID | TRTSTDY | TRTENDY | ORDER | TRTCAP | AFUSTDY | AFUENDY | AFUCAP |
|---|---|---|---|---|---|---|---|
| 1002 | 1 | 28 | 1 | | | | |
| 1010 | 1 | 28 | 2 | | | | |
| 1008 | 1 | 35 | 3 | | 35 | 181 | FilledArrow |
| 1003 | 1 | 56 | 4 | FilledArrow | | | |
| ⋮ | | | | | | | |
| 1011 | 1 | 210 | 13 | | 210 | 214 | FilledArrow |
| 1014 | 1 | 270 | 14 | | | | |
| 1005 | 1 | 290 | 15 | FilledArrow | | | |

**Table 3 Data Set Including Active Follow-Up**

We can now add the AFU period to the swimmer plot using a second HIGHLOW statement, very similar to the first, but using the AFU-related variables instead. Another difference is to distinguish this period by using dashed lines for the border of the bars rather than the solid lines of the treatment period. This is done by setting the PATTERN option within the LINEATTRS argument as shown:

```
PROC SGPLOT DATA=swimmer;
    /* bars for treatment duration */
    HIGHLOW Y=order LOW=trtstdy HIGH=trtendy / LOWLABEL=subjid TYPE=BAR
      BARWIDTH=0.95 HIGHCAP=trtcap;

    /* bars for active follow-up duration */
    HIGHLOW Y=order LOW=afustdy HIGH=afuendy / TYPE=BAR BARWIDTH=0.95
      LINEATTRS=(PATTERN=2) HIGHCAP=afucap;

    /* axes */
    XAXIS LABEL='Study Day' VALUES=(0 TO 300 BY 30);
    YAXIS LABEL='Subject ID' DISPLAY=(NOTICKS NOVALUES);
RUN;
```

The resulting plot is shown in Figure 5.

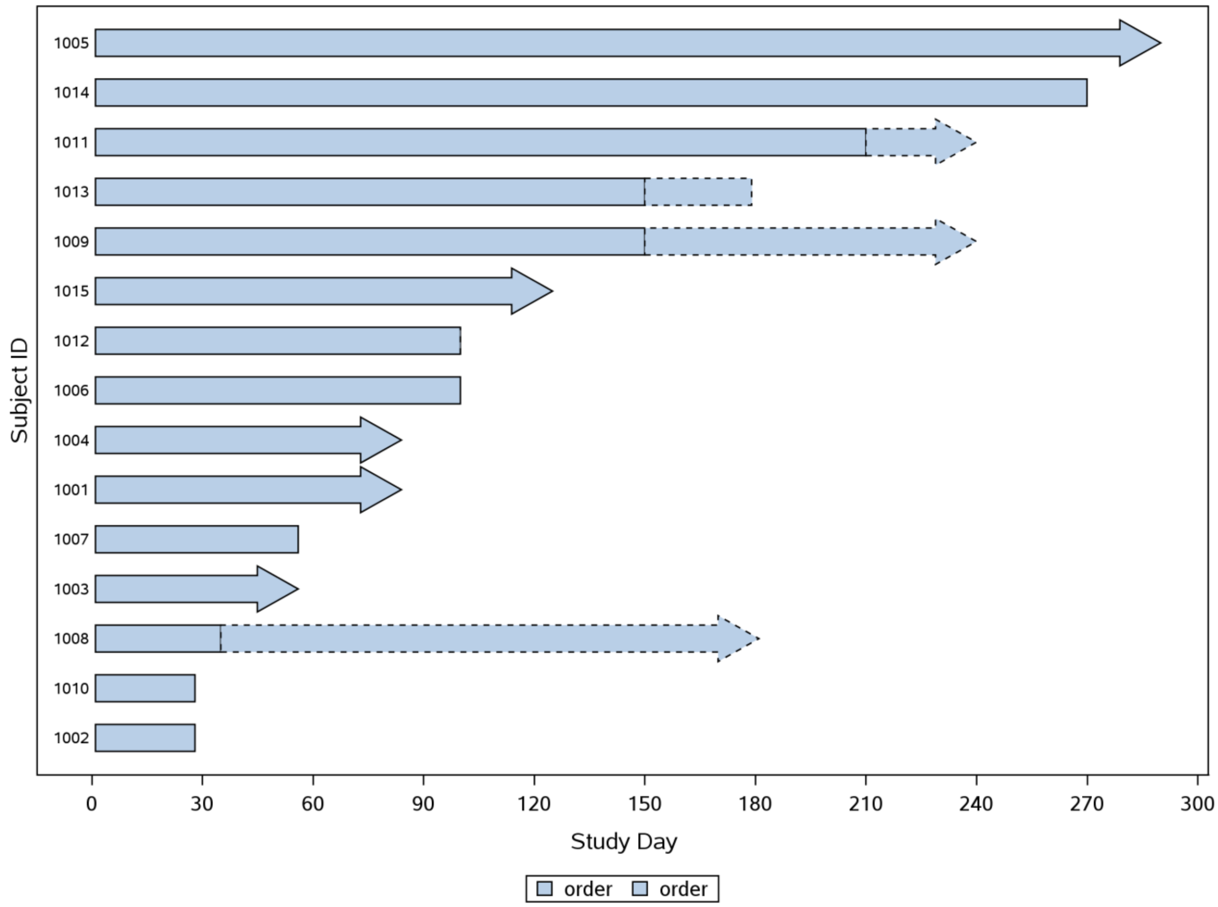


**Figure 5 First Iteration of Treatment Duration and Active Follow-up**

With this iteration of the figure, there are two immediate issues to be addressed as well as some subtle ones which may not be readily apparent.

First, the ordering variable should be re-derived based on the total duration of treatment plus active follow-up. Then the bars will be ordered based on this combined duration.

Second, by adding another plot to the figure, a legend has been automatically included by the procedure in an attempt to differentiate them. It is not helpful in this case since both plots are using the ORDER variable. Irrespective of that, we can suppress the legend for now by using the NOAUTOLEGEND option. Legends for the swimmer plot will be discussed later.

It is tempting to sequentially specify the HIGHLOW plots for treatment and then for AFU. Due to the order of rendering, a consequence is that the dashed lines of the AFU bar overlay the solid line at the end of the treatment period. It is necessary to have close-up inspection to notice this issue, which is shown in Figure 6. This can be resolved by switching the order of the HIGHLOW plots so that the solid lines are rendered last and overlay the dashed ones.

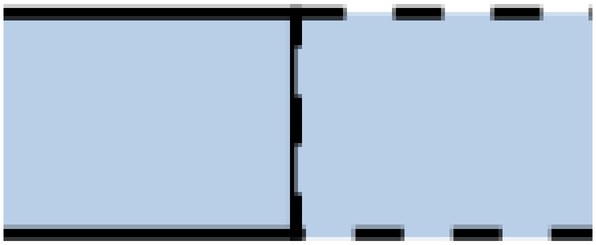

**Figure 6 Dashed Lines Overlaying Solid Lines**

Even more subtle is noticing this phenomenon at the end of the bar for subject 1012, even though there is no apparent period of AFU. The close-up is shown in Figure 7.

**Figure 7 Dashed Lines and Zero-length Bar**

Upon closer inspection of the data, this subject is ongoing for AFU but the start and end days are the same value. That is, the subject's status is only known as of the start of AFU and there have been no further visits at the time of the interim data availability.

In terms of the plot and what is seen in Figure 7, it is explained by a restriction noted in the documentation for the HIGHCAP (and similarly, LOWCAP) argument:

> "Caps are not displayed for very short bars. Bar height must be at least twice the size of the cap in order for the cap to appear" (SAS, 2018).

There is no warning or note in the logs to alert you to this.

To accommodate for this, a programming check can be done on the difference between the start and end days of AFU. If it is within a certain margin, then some "buffer" duration can be added to the end so that the difference is sufficiently large for the cap to be displayed.

This raises the question about whether the plot then accurately reflects the data, since the arrowhead would extend beyond the known timeframe of AFU. While it is a matter of discretion, it can be argued that the visual indication of ongoing AFU is more informative and relevant than the precise positioning of a stylistic element like an arrowhead.

**Tips:**

- Where end cap arrows are to be used, add “buffer” space if necessary to ensure the bar is sufficiently long to display the cap.
- Order subjects based on combined duration of treatment and AFU (including any additional buffer added by the previous tip).
- Sequence the HIGHLOW plots so that AFU is specified first and then followed by treatment period, so that solid lines overlay dashed lines.
- Suppress automatic legend with the NOAUTOLEGEND option of the procedure.

The resulting plot is shown in Figure 8. In particular, we see that the vertical transition point between treatment and AFU has a nice solid line with no interfering dashes, and a case such as subject 1012 has a sufficient arrow to indicate that AFU has just started and is ongoing.

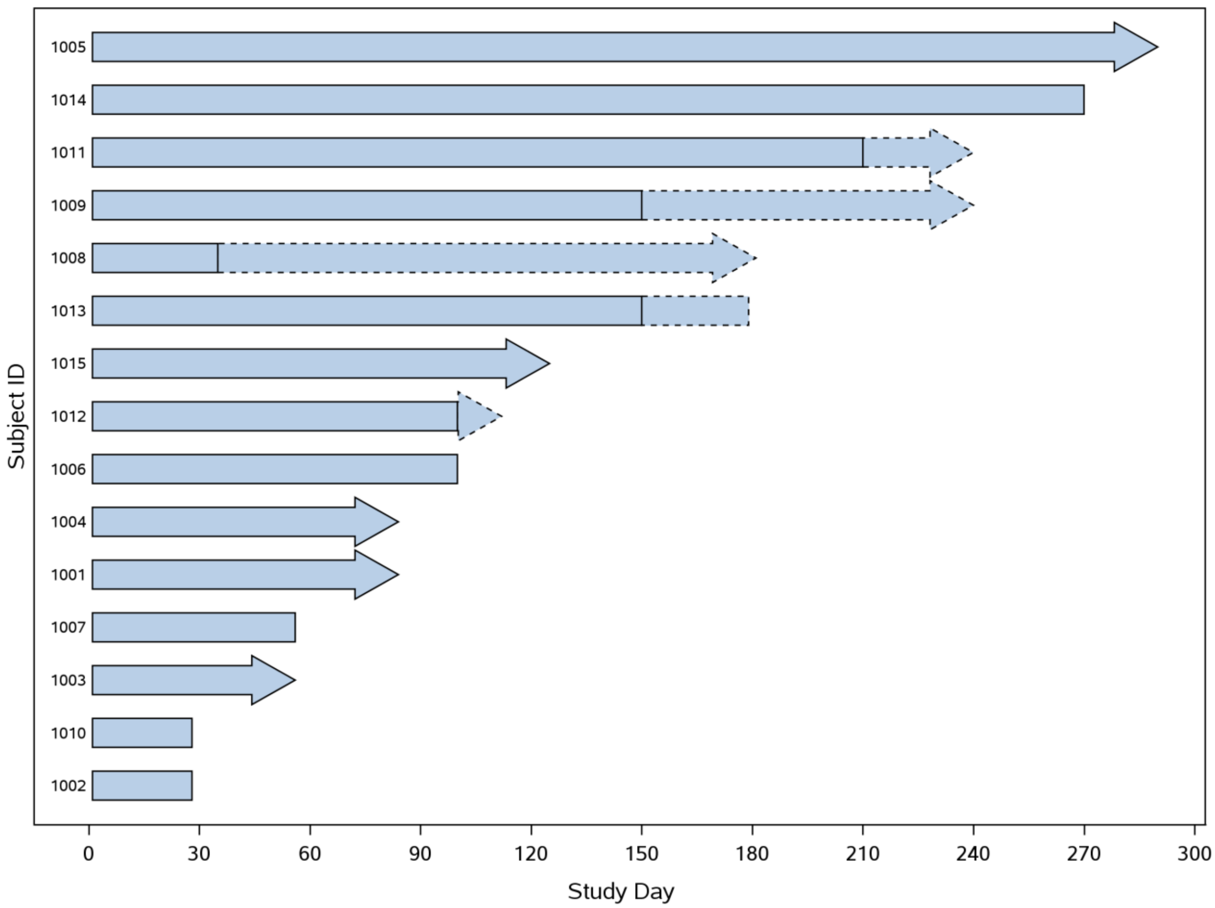


**Figure 8 Updated Treatment Duration and Active Follow-up**

**Further Notes:**

- LOWLABEL is not needed for the HIGHLOW statement of the AFU period since we only want the subject ID labels to be displayed at the far left of the plot, i.e., as labels at the start of the bars for the treatment period.
- There is no conflict between TRTCAP and AFUCAP. If a subject has entered AFU, then by necessity their treatment cannot still be ongoing.
- Since AFU starts immediately after treatment ends, the values in TRTENDY and AFUSTDY are redundant for subjects who continue into that period. But for clarity, AFUSTDY is used.
- As like the derivation of TRTCAP, it is assumed there is a variable to indicate that AFU is ongoing at the time of AFUENDY and this is used to set AFUCAP accordingly.

## PLOTTING THE PERIODS OF RESPONSE

We next consider the data set with variables for the start and end days of some response criteria (e.g., “objective response” according to tumor evaluation criteria called RECIST) and a variable to indicate if the response is ongoing or has ended. These are shown in Table 4 (with other variables omitted, for space).

| SUBJID | ORDER | RESPSTDY | RESPENDY | RESPONGO |
|---|---|---|---|---|
| 1002 | 1 | | | |
| 1010 | 2 | | | |
| 1003 | 3 | 56 | 56 | Response ongoing |
| | | ⋮ | | |
| 1011 | 13 | | | |
| 1014 | 14 | 125 | 235 | Response end |
| 1005 | 15 | 60 | 290 | Response ongoing |

**Table 4 Data Set Variables for Periods of Response**

Symbols for the start and end of response can now be added to the figure by scatter plots using SCATTER statements.

The first statement is for the start of response which will be plotted for each subject (at y-axis position given by ORDER) at the value of RESPSTDY; the marker symbol and style (i.e., “attributes”) are common for all subjects and controlled explicitly via the MARKERATTRS argument and its options. For this plot, a filled diamond marker will be the symbol indicating start of response. The size is chosen for the symbol to fit comfortably within the size of the bar.

For the end of response, different symbols will be used depending on whether the response is ongoing or not. We can differentiate these within a single SCATTER statement by specifying RESPONGO as the GROUP variable. The change of plotting symbols across group values would be controlled by the default style element attributes, but we can override these with our own specific choices. One way to do this directly, without defining or editing styles, is with the STYLEATTRS statement. For purposes of this data, a right-pointing triangle is used for an ongoing response and left-pointing triangle will be used to indicate the end of response. A drawback of this approach is that the order of the values within the input data set must be known in advance and the values for DATASYMBOLS listed in the corresponding matching order.

To minimize other style differences across the plots, the NOCYCLEATTRS option of the procedure is specified. Then, only explicitly-specified attributes will differ as controlled by other options such as within STYLEATTRS.

The code including the SCATTER statements and previously mentioned tips is:

```
PROC SGPLOT DATA=swimmer NOAUTOLEGEND NOCYCLEATTRS;
    /* bars for active follow-up duration */
    HIGHLOW Y=order LOW=afustdy HIGH=afuendy / TYPE=BAR BARWIDTH=0.95
      LINEATTRS=(PATTERN=2) HIGHCAP=afucap;

    /* bars for treatment duration */
    HIGHLOW Y=order LOW=trtstdy HIGH=trtendy / LOWLABEL=subjid TYPE=BAR
      BARWIDTH=0.95 HIGHCAP=trtcap;

    /* markers for start of response */
    SCATTER Y=order X=respstdy / MARKERATTRS=(SYMBOL=DIAMONDFILLED SIZE=10);

    /* markers for end of response */
    STYLEATTRS DATASYMBOLS=(TRIANGLERIGHTFILLED TRIANGLELEFTFILLED);
    SCATTER Y=order X=respendy / MARKERATTRS=(COLOR=BLACK SIZE=10)
      GROUP=respongo;

    /* axes */
    XAXIS LABEL='Study Day' VALUES=(0 TO 300 BY 30);
    YAXIS LABEL='Subject ID' DISPLAY=(NOTICKS NOVALUES);
RUN;
```

The initial attempt at plotting the response start and end points gives Figure 9.

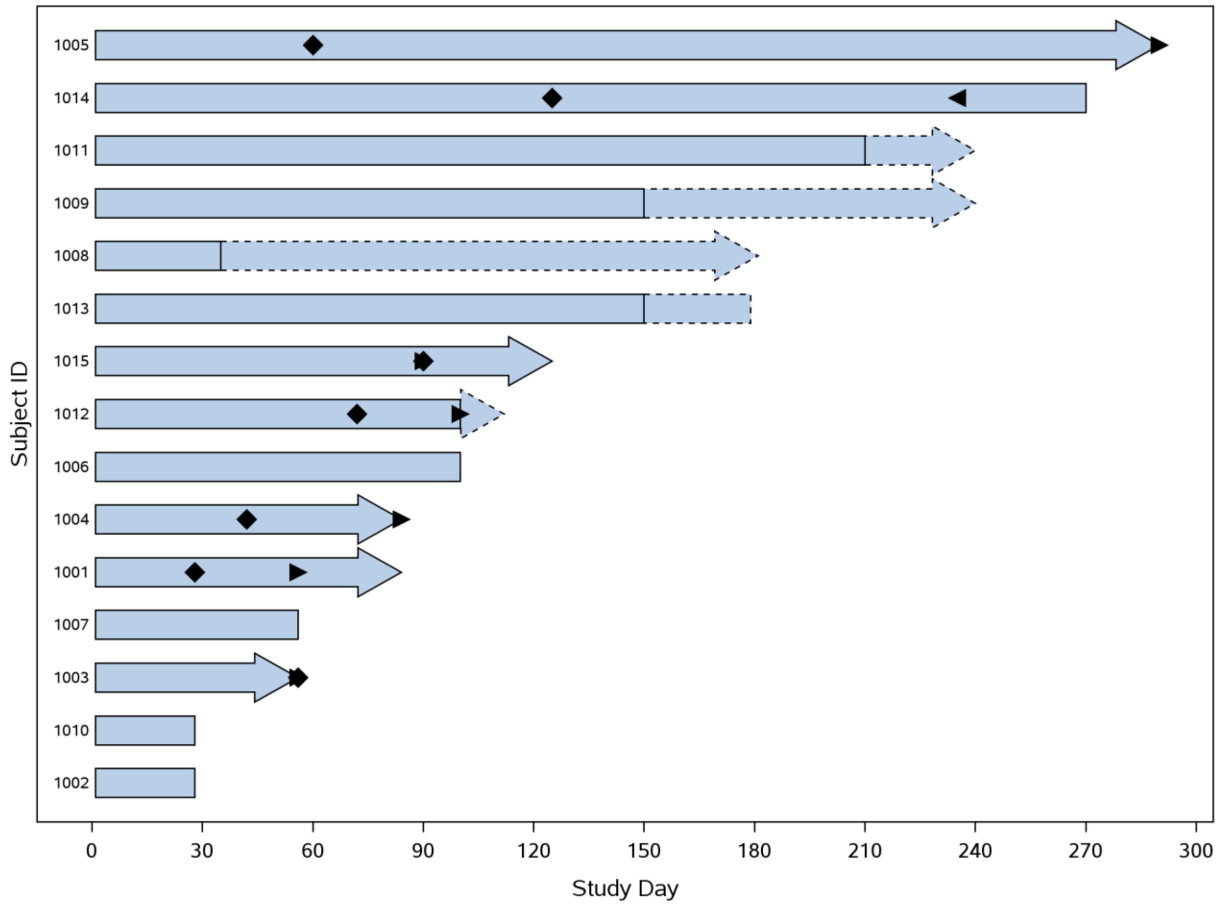


**Figure 9 First Iteration of Including Response Periods**

There are two issues to address with this result: when response start and end days are the same and so their plotting symbols overlap (Figure 10); and for the visually unappealing placement of ongoing response symbols relative to ongoing treatment or AFU arrowheads (Figure 11)

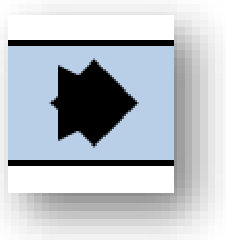

**Figure 10 Overlapping Markers**

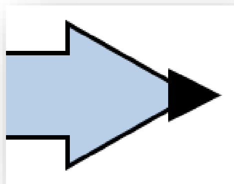

**Figure 11 Overlapping Markers and Bar Arrowheads**

The overlapping symbols can be resolved similarly to the “short bar” issue discussed earlier. If the difference between the start and end of response is within a certain threshold, then a “buffer” can be added to the end of response value to avoid overlap. It is a conscious decision to alter only this value (i.e., as opposed to moving the start value only or splitting the difference between the two). The start of response is a definitive event and should act as the anchor; overlap should only logically occur when no subsequent tumor assessment has been performed yet. Hence, the symbol to indicate ongoing response in this case is largely artificial anyway and is the more appropriate one to move to avoid overlap.

The other issue is a matter of aesthetics. The tip of the bar cap arrowhead corresponds to the underlying data value, whereas the marker symbol is centered about this point. In cases where the end of treatment / AFU and the end of response are the same (or within a certain amount), the resulting visual placements may appear awkward. A more appealing look would be to have the arrowhead centered above the marker symbol – again accomplished by programmatically checking values and then adding some “buffer” to the end of treatment / AFU value to achieve the desired result.

These issues may also intersect, as illustrated by subject 1003 in Figure 9. So care should be taken in the order of applying the suggested adjustments. After implementing such adjustments, the same swimmer plot code can be used with a result as shown in Figure 12.

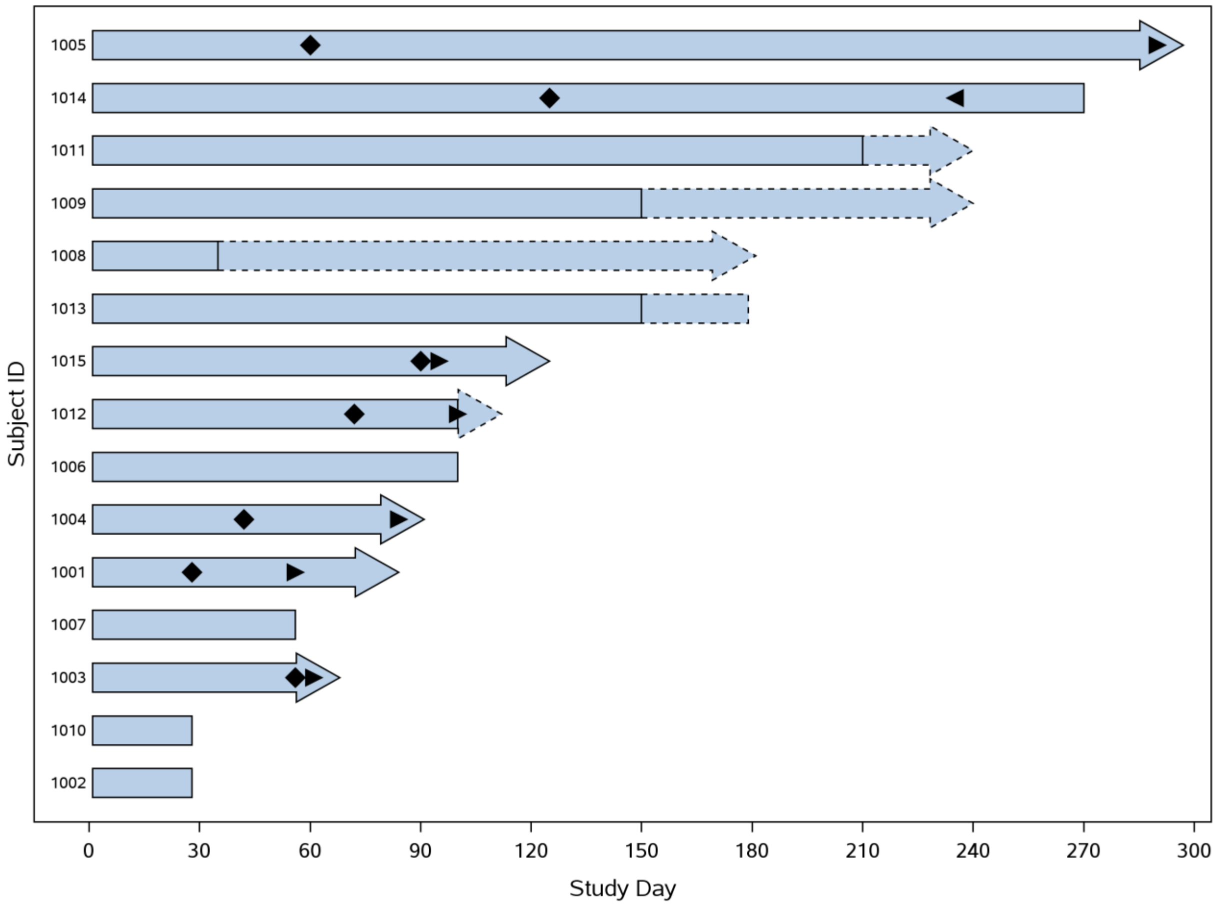


**Figure 12 Resolving Overlapping Markers and Arrowheads**

The final element to emphasize the duration of response is to join the start and end points. This too is accomplished with a HIGHLOW plot, now using the default type which displays line segments. One modification is to increase the thickness of the lines for emphasis. The code with the additional HIGHLOW statement is below, followed by the result in Figure 13:

```
PROC SGPLOT DATA=swimmer NOAUTOLEGEND NOCYCLEATTRS;
    /* bars for active follow-up duration */
    HIGHLOW Y=order LOW=afustdy HIGH=afuendy / TYPE=BAR BARWIDTH=0.95
      LINEATTRS=(PATTERN=2) HIGHCAP=afucap;

    /* bars for treatment duration */
    HIGHLOW Y=order LOW=trtstdy HIGH=trtendy / LOWLABEL=subjid TYPE=BAR
      BARWIDTH=0.95 HIGHCAP=trtcap;

    /* markers for start of response */
    SCATTER Y=order X=respstdy / MARKERATTRS=(SYMBOL=DIAMONDFILLED SIZE=10);

    /* markers for end of response */
    STYLEATTRS DATASYMBOLS=(TRIANGLERIGHTFILLED TRIANGLELEFTFILLED);
    SCATTER Y=order X=respendy / MARKERATTRS=(COLOR=BLACK SIZE=10)
      GROUP=respongo;

    /* lines for duration of response */
    HIGHLOW Y=order LOW=respstdy HIGH=respendy / LINEATTRS=(THICKNESS=2);

    /* axes */
    XAXIS LABEL='Study Day' VALUES=(0 TO 300 BY 30);
    YAXIS LABEL='Subject ID' DISPLAY=(NOTICKS NOVALUES);
RUN;
```

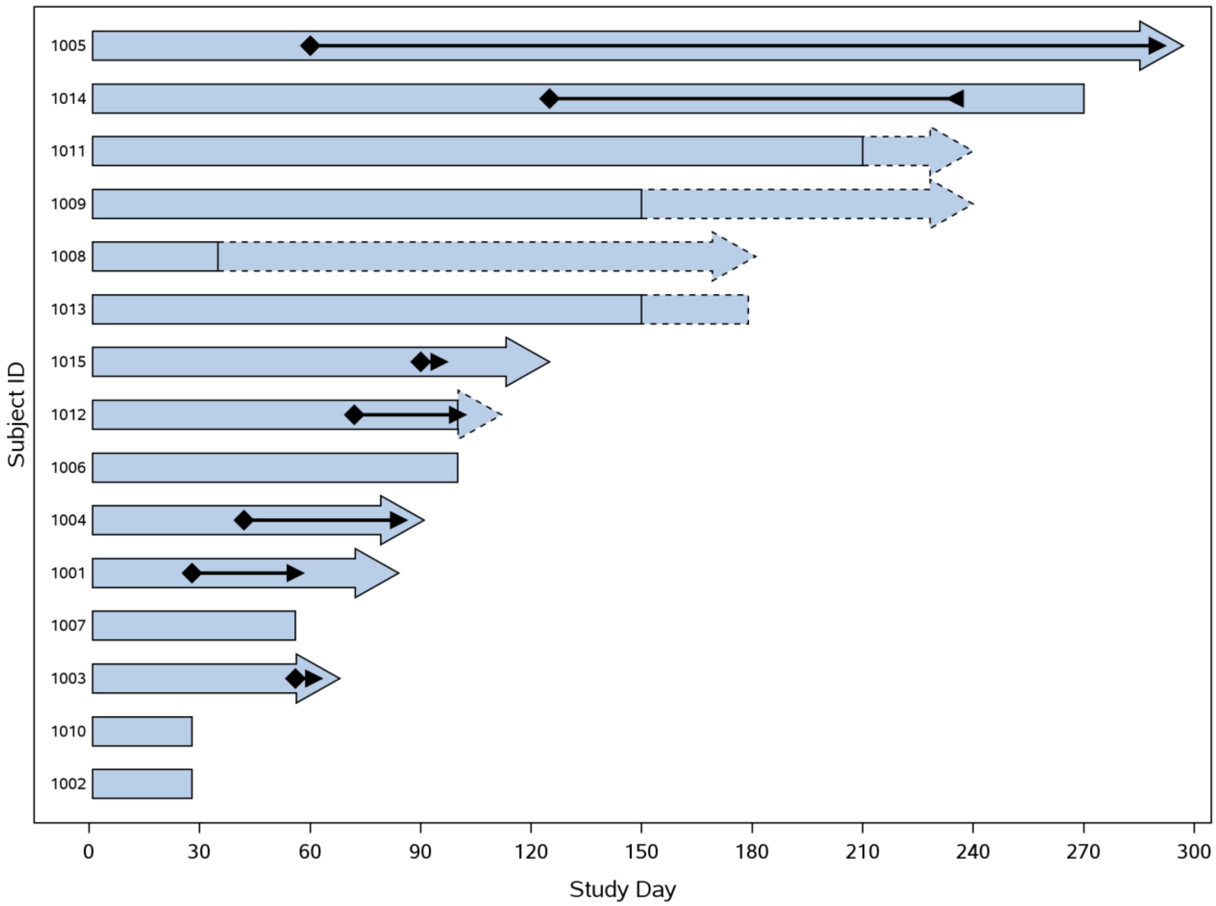


**Figure 13 Joining Start and End Points of Response**

Although the HIGHCAP argument could have been used with these line segments in order to indicate ongoing or end of response (instead of the scatter plot markers), the options are limited. There is an arrowhead as the natural cap for ongoing response, but neither NONE or SERIF serve well as emphatic display for the end of response.

**Tips:**

- Add offset to response end value if needed to avoid overlap with response start.
- If ongoing response symbol and ongoing treatment / AFU arrowhead are too close, offset the value of the latter so that the arrowhead is centered above the symbol.
- Reordering of subjects may be necessary after these adjustments.
- Use STYLEATTRS statement and/or MARKERATTRS options to explicitly control style elements; the use of STYLEATTRS is dependent on the order in which the values of the grouping variable appear in the input data.

## INDICATING TYPES OF RESPONSE

In the context of the plots shown so far, a “responder” corresponds to a best overall response of ‘Complete Response’ or ‘Partial Response’ according to the RECIST criteria for evaluating tumors. It would be desirable to distinguish the responder subjects in the plot by whether it is a complete or partial response. Even better would be to identify the categories of the non-responders as well. Table 5 shows an expanded dataset with variable BOR to indicate each subject’s best overall response.

| SUBJID | ORDER | RESPSTDY | RESPENDY | RESPONGO | BORCD | BOR |
|---|---|---|---|---|---|---|
| 1002 | 1 | . | . | | 4 | Progressive Disease |
| 1010 | 2 | . | . | | 4 | Progressive Disease |
| 1003 | 3 | 56 | 56 | Response ongoing | 1 | Complete Response |
| ⋮ | | | | | | |
| 1011 | 13 | . | . | | 3 | Stable Disease |
| 1014 | 14 | 125 | 235 | Response end | 2 | Partial Response |
| 1005 | 15 | 60 | 290 | Response ongoing | 2 | Partial Response |

**Table 5 Data Set with Best Overall Response**

The swimmer plot is then modified to specify BOR as a group variable for the treatment / AFU bars so that they become color-coded based on best overall response. As with the marker symbols for ongoing or end of response, a STYLEATTRS statement can be used to fix the colors corresponding to each of the response categories (with the same caveat that the order of the category variables within the data must be known).

Both of the HIGHLOW statements for treatment and AFU will use BOR as the grouping variable to receive the same coloring. The first legend for the plot is also generated. We assign a NAME value to the HIGHLOW statement of the treatment duration and then add a KEYLEGEND statement with that plot’s name. Note that for grouped data, legend entry values are sorted according to their appearance in the data. Since there is a meaningful relative ordering of best overall response values (e.g., from best to worse outcomes), a sorting variable such as BORCD can be used to sort the data prior to the procedure call. This also affects the order to be used in the STYLEATTRS mentioned above.

Note that by specifying a grouping variable, certain visual properties will be differentiated for each group value. In order to keep these constant, other than the color of the bars, explicit values are specified. In particular, the COLOR of LINEATTRS is used to keep the outline of the bars as black and LABELATTRS is used for the treatment duration plot so that the text labels of the group values shown in the legend are a constant black color.

Incorporating the best overall responses into the program is as follows, with results shown in Figure 14:

```
PROC SGPLOT DATA=swimmer NOAUTOLEGEND NOCYCLEATTRS;
    /* colors of the grouped bars knowing the best overall response data is
      sorted as CR, PR, SD, PD */
    STYLEATTRS DATACOLORS=(DAGR MEGR LIGR WHITE);

    /* bars for active follow-up duration */
    HIGHLOW Y=order LOW=afustdy HIGH=afuendy / TYPE=BAR BARWIDTH=0.95
      LINEATTRS=(COLOR=BLACK PATTERN=2) HIGHCAP=afucap GROUP=bor;

    /* bars for treatment duration */
    HIGHLOW Y=order LOW=trtstdy HIGH=trtendy / LOWLABEL=subjid TYPE=BAR
      LINEATTRS=(COLOR=BLACK) LABELATTRS=(COLOR=BLACK) BARWIDTH=0.95
      HIGHCAP=trtcap GROUP=bor NAME='bor';

    /* markers for start of response */
    SCATTER Y=order X=respstdy / MARKERATTRS=(SYMBOL=DIAMONDFILLED SIZE=10);

    /* markers for end of response */
    STYLEATTRS DATASYMBOLS=(TRIANGLERIGHTFILLED TRIANGLELEFTFILLED);
    SCATTER Y=order X=respendy / MARKERATTRS=(COLOR=BLACK SIZE=10)
      GROUP=respongo;

    /* lines for duration of response */
    HIGHLOW Y=order LOW=respstdy HIGH=respendy / LINEATTRS=(THICKNESS=2);

    /* axes */
    XAXIS LABEL='Study Day' VALUES=(0 TO 300 BY 30);
    YAXIS LABEL='Subject ID' DISPLAY=(NOTICKS NOVALUES);

    KEYLEGEND 'bor' / TITLE='Best Overall Response';
RUN;
```

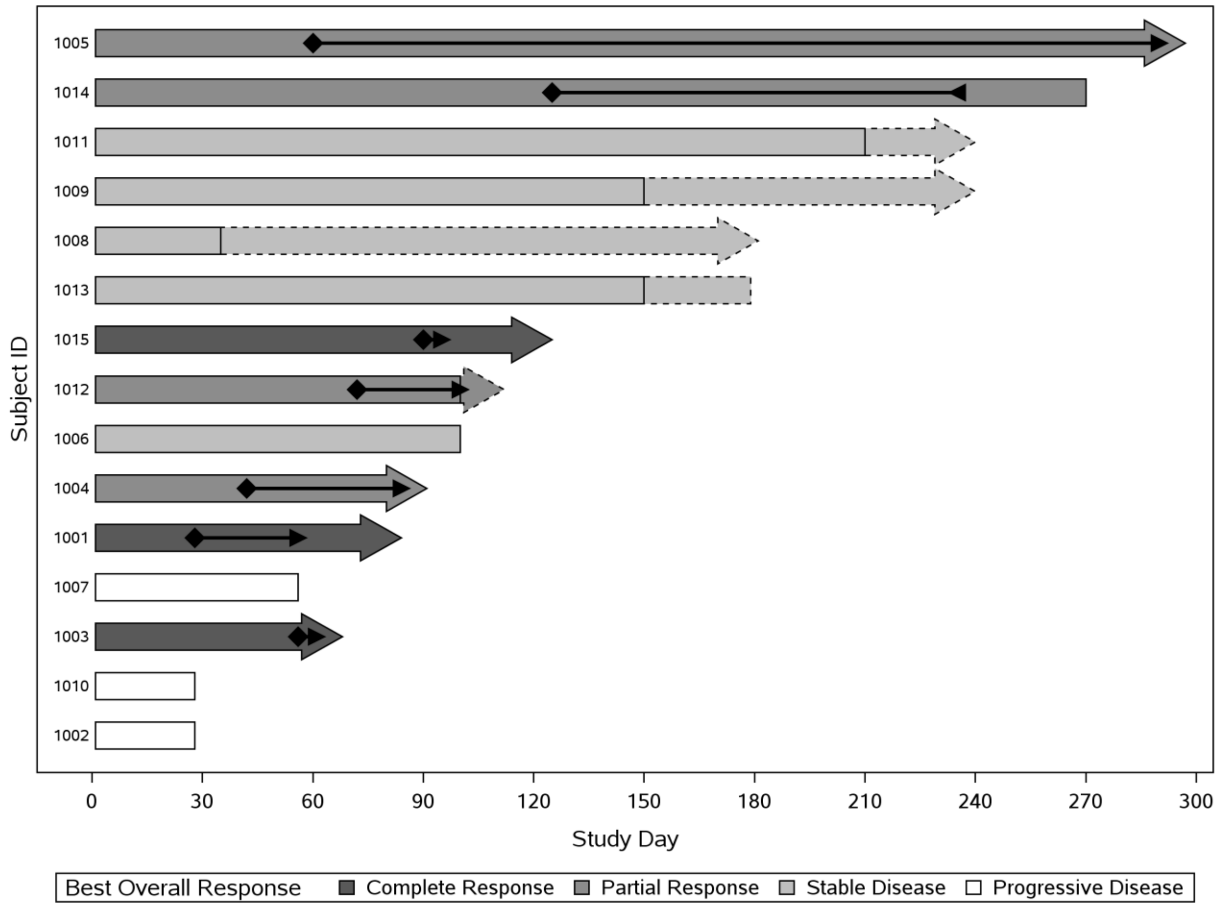


**Figure 14 Indicating Best Overall Response**

Each subject can be further marked for other types of endpoint responses or any characteristics of the user's choice. For example, a study may define a "durable disease control responder" (DDCR) as a subject with a response or with stable disease of duration at least 180 days. Subjects meeting such conditions could be marked at the end of their treatment / AFU bar with a particular marker symbol. If the input dataset contains a variable which indicates DDCR, then a new variable with a value corresponding to the end of the bar plus small offset could be plotted with another SCATTER statement. An example of identifying DDCR subjects with a square marker symbol is shown in Figure 15. While most DDCR subjects can be easily identified anyway, this is a useful technique to be able to accurately count the number of DDCR subjects; indeed, subject 1013 falls just short of meeting the response criteria and this is not possible to gauge definitively based on a visual check alone.

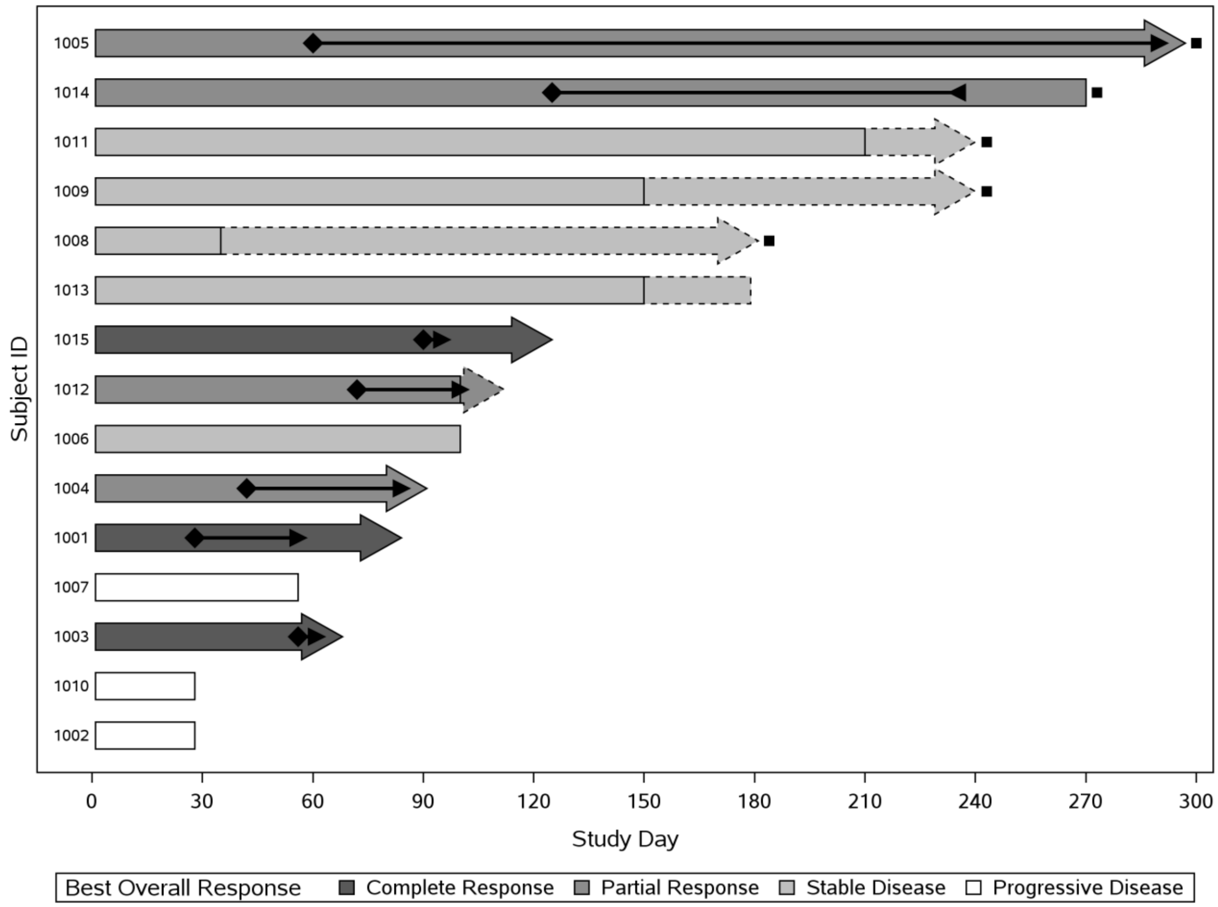


**Figure 15 Indicating Durable Disease Control Responders**

## DATASKINS AND COLORS FOR ENHANCING VISUAL APPEAL

If desired (and permitted), figures can be enhanced by other visual effects and/or color. These are most often purely subjective preferences, although some considerations can be pointed out to keep in mind.

As of SAS 9.4M1, the DATASKIN option is available and can be used with HIGHLOW plots to create a special visual effect for the filled bars. The CRISP option is versatile for both B&W or color, shown in Figure 16 and Figure 17 respectively. The skin is applied only to the treatment duration bar, which further differentiates it from the AFU period.

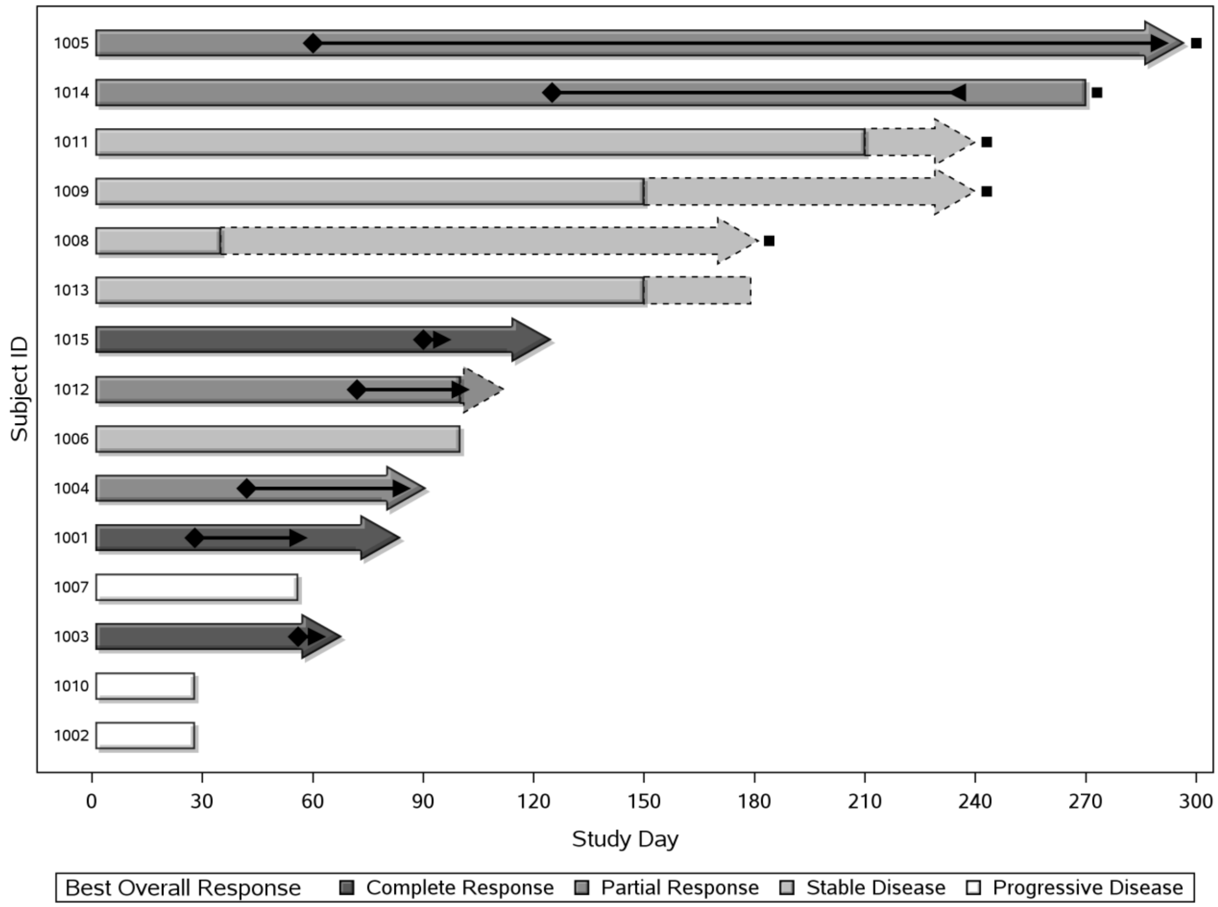


**Figure 16  Applying Crisp Data Skin**

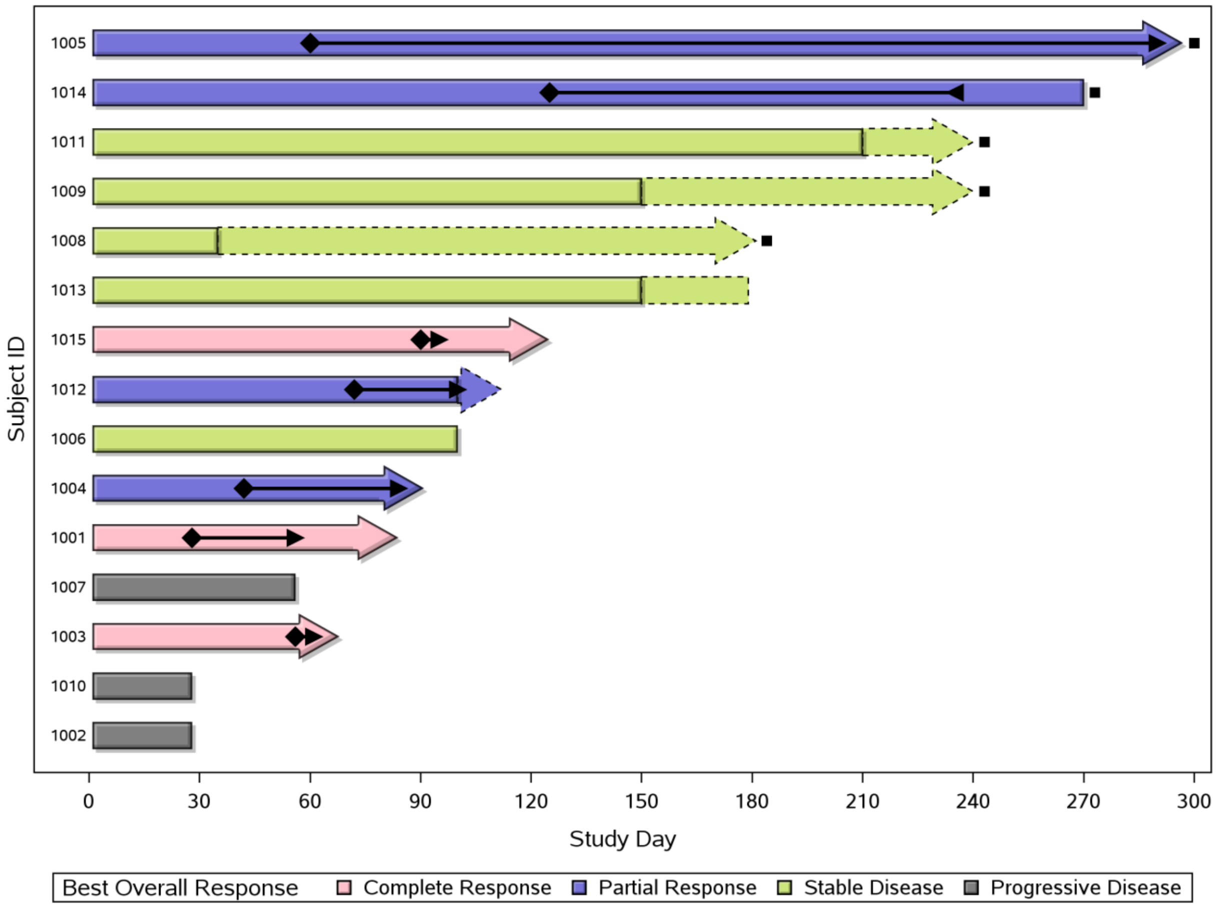


**Figure 17   Applying Crisp Data Skin and Color**

Skin styles that include some gradient effect may be popular due to their ornate appearance, but care should be taken with how these are impacted by the use of arrowhead caps to the bars. It seems that gradients are applied based on the displayed width of the bars and not the boundaries of the “swim lane” they occupy; so a bar with an arrowhead is “wider” than one without and the relative spread of gradients will differ. Figure 18 shows an example with the SHEEN skin applied. Although both subjects 1005 and 1014 have the same Partial Response, at first glance it appears that the coloring is somehow indicating some sort of difference between the two, further illustrated in Figure 19.

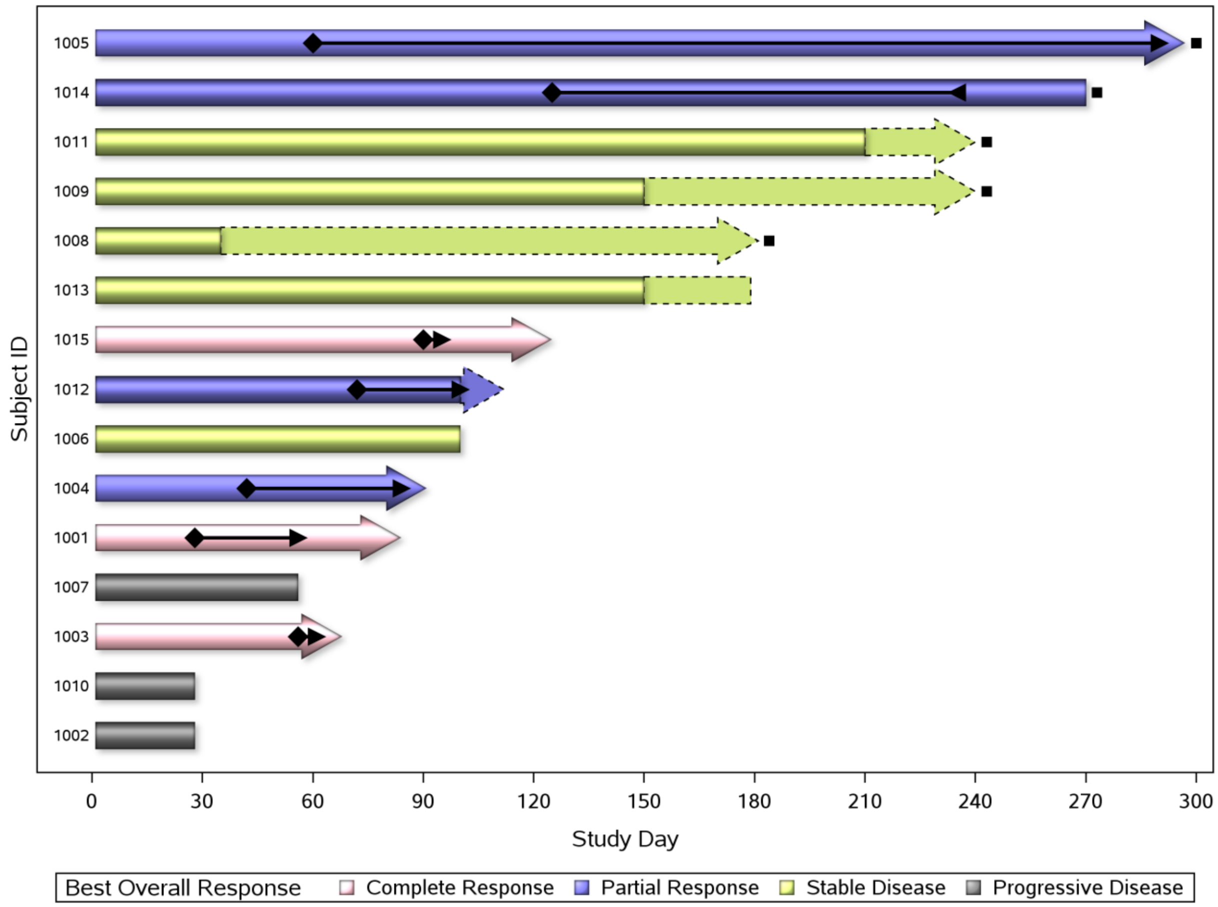


**Figure 18 Applying Sheen Data Skin**

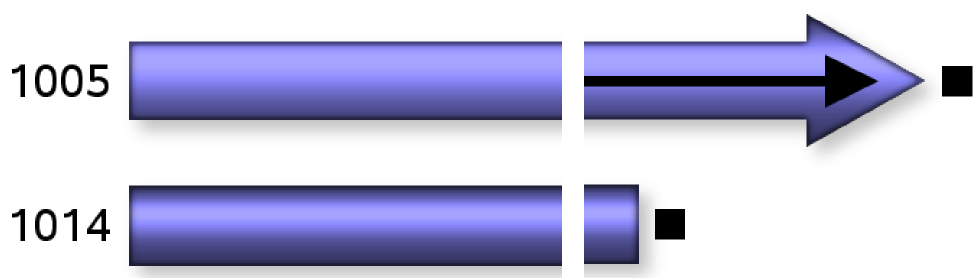


**Figure 19 Differences in Applying Gradient**

## LEGENDS

Thus far, only the Best Overall Response values have been represented in a legend, at the default position outside the plotting area. Other aspects of the plot should be similarly explained for the benefit of the viewer. Due to the natural ordering used for the swimmer plot, the bottom right quadrant will often have sufficient whitespace to be an appropriate location to use for a second KEYLEGEND statement.

The SCATTER statements are each assigned a NAME value for reference in the KEYLEGEND. The response start and DDCR markers are fixed and so constant LEGENDLABEL values for each can be specified. For the response end markers it was previously unremarked that the values of RESPONGO were descriptive text strings, rather than some shorthand or coded value; this was actually a purposeful choice as these values become the default text labels used for the legend. Hence, we do not need to specify a LEGENDLABEL for the corresponding SCATTER statement; but it *is* necessary to also use the NOMISSINGGROUP option so that only non-missing values are considered valid for inclusion in the legend.

The modifications for including the legend, as well as data skins and the DDCR markers, are in the following code, with the result in Figure 20:

```
PROC SGPLOT DATA=swimmer NOAUTOLEGEND NOCYCLEATTRS;
    /* colors of the grouped bars knowing the best overall response data is
      sorted as CR, PR, SD, PD */
    STYLEATTRS DATACOLORS=(PINK VLIB LIY GRAY);

    /* bars for active follow-up duration */
    HIGHLOW Y=order LOW=afustdy HIGH=afuendy / TYPE=BAR BARWIDTH=0.95
      LINEATTRS=(COLOR=BLACK PATTERN=2) HIGHCAP=afucap GROUP=bor;

    /* bars for treatment duration */
    HIGHLOW Y=order LOW=trtstdy HIGH=trtendy / LOWLABEL=subjid TYPE=BAR
      LINEATTRS=(COLOR=BLACK) LABELATTRS=(COLOR=BLACK) BARWIDTH=0.95
      HIGHCAP=trtcap GROUP=bor NAME='bor' DATASKIN=SHEEN;

    /* markers for start of response */
    SCATTER Y=order X=respstdy / MARKERATTRS=(SYMBOL= DIAMONDFILLED SIZE=10)
      NAME='resp_start' LEGENDLABEL='Response start';

    /* markers for end of response */
    STYLEATTRS DATASYMBOLS=(TRIANGLERIGHTFILLED TRIANGLELEFTFILLED);
    SCATTER Y=order X=respendy / MARKERATTRS=(COLOR=BLACK SIZE=10)
      GROUP=respongo NOMISSINGGROUP NAME='resp_end';

    /* lines for duration of response */
    HIGHLOW Y=order LOW=respstdy HIGH=respendy / LINEATTRS=(THICKNESS=2);

    /* markers for indicating durable disease control responders */
    SCATTER Y=order X=ddcrdy / MARKERATTRS=(SYMBOL=SQUAREFILLED SIZE=6)
      NAME='ddcr' LEGENDLABEL='Durable disease control responder';

    /* axes */
    XAXIS LABEL='Study Day' VALUES=(0 TO 300 BY 30);
    YAXIS LABEL='Subject ID' DISPLAY=(NOTICKS NOVALUES);

    KEYLEGEND 'bor' / TITLE='Best Overall Response';
    KEYLEGEND 'resp_start' 'resp_end' 'ddcr' / NOBORDER LOCATION=INSIDE
      POSITION=BOTTOMRIGHT ACROSS=1;
RUN;
```

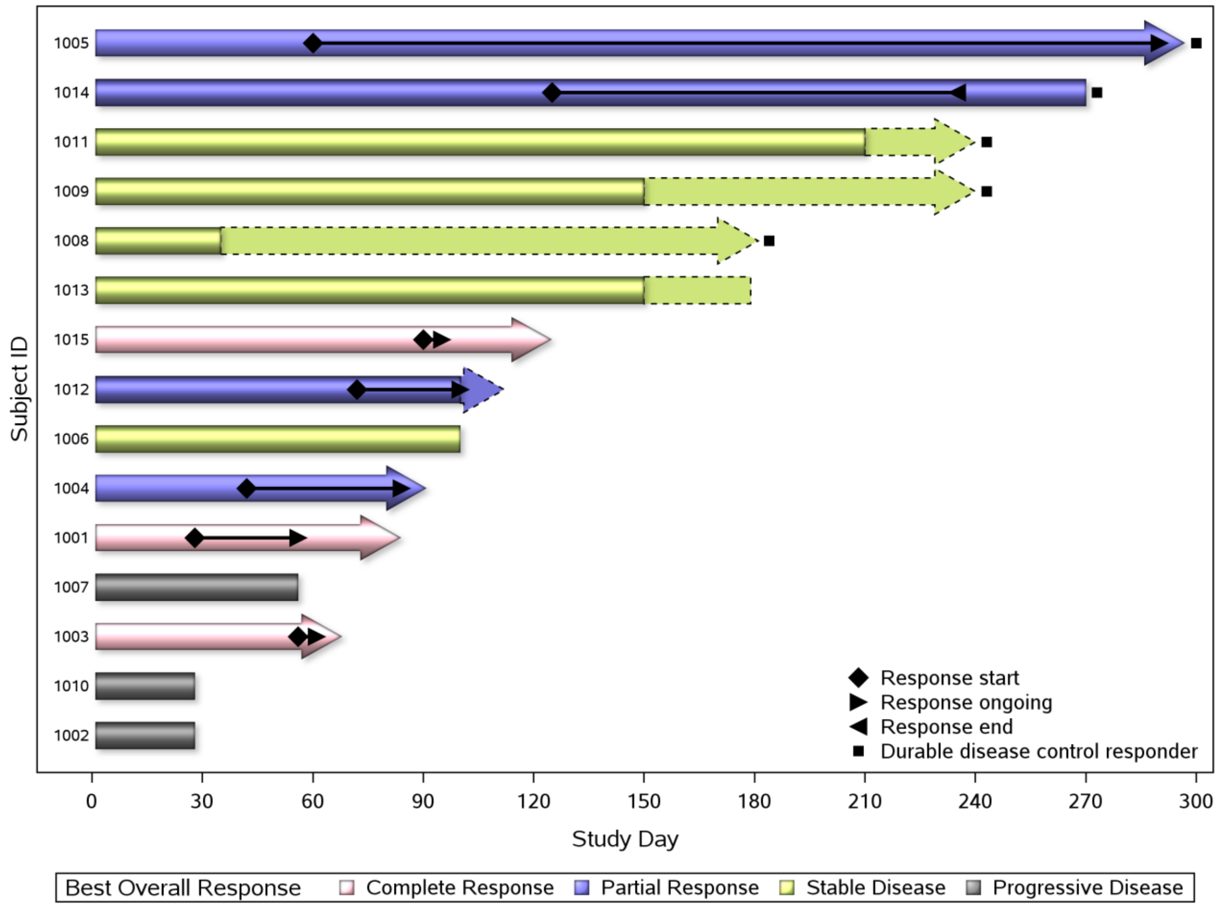


**Figure 20 Final Swimmer Plot with Legends**

## OTHER CONSIDERATIONS

There are numerous other kinds of events or outcomes which could be included on swimmer plots, such as:

- Surgeries
- Surgical complete response
- Disease progression
- Differentiating periods of treatment occurring after disease progression
- Start of new anti-cancer therapies
- Other types of responders defined for the study

The possibilities are many, with the main restriction being consideration of readability and not overwhelming the plot with too much information.

There is also the question of how many subjects (i.e., "lanes") is too many to fit into a figure.

These points are also subjective and should take into account what are the primary data points of interest. In general, the more subjects are included then the more difficult it is to discern the finer details and so unnecessary clutter should be avoided.

## CONCLUSION

By using an appropriately constructed data set and the available functionality of PROC SGPLOT statements, information-dense swimmer plots can be easily constructed. Only a handful of plot statements are necessary, which facilitates quick programming and as importantly allows for simple validation when necessary.

This paper has presented tips for a number of elements which can be implemented and/or customized as the finishing touches for an aesthetically-pleasing final graph.

## REFERENCES

Huang, Jui-Fu. 2016. “A Different Approach to Create Swimmer Plot Using Proc Template and SGRENDER.” *Proceedings of PharmaSUG Conference 2016*, Denver, CO. Available at https://www.pharmasug.org/proceedings/2016/DG/PharmaSUG-2016-DG14.pdf.

Mantage, Sanjay. 2014. “Swimmer plot.” Accessed April 28, 2019. Available at https://blogs.sas.com/content/graphicallyspeaking/2014/06/22/swimmer-plot/.

Phillips, Stacey. 2014. “Swimmer Plot: Tell a Graphical Story of Your Time to Response Data Using PROC SGPLOT.” *Proceedings of PharmaSUG Conference 2014*, San Diego, CA. Available at http://www.pharmasug.org/proceedings/2014/DG/PharmaSUG-2014-DG07.pdf.

Wang, Baiming. 2016. “Swimmer Plot by Graphic Template Language (GTL).” *Proceedings of PharmaSUG Conference 2016*, Denver, CO. Available at https://www.pharmasug.org/proceedings/2016/DG/PharmaSUG-2016-DG08.pdf.

SAS Institute Inc. 2018. “PROC SGPLOT HIGHLOW Statement”. *SAS® 9.4 ODS Graphics: Procedures Guide, Sixth Edition*. Available at https://documentation.sas.com/?docsetId=grstatproc&docsetTarget=n0mjz9ktgnse58n14deqdvnnxarp.htm&docsetVersion=9.4&locale=en.

## CONTACT INFORMATION

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

Steve Almond
Bayer Inc.
2920 Matheson Blvd E
Mississauga, ON, Canada, L4W 5R6
steve.almond@bayer.com