

Visualization of Sparse PK Concentration Sampling Data, Step by Step (Improvement by Improvement)

Ilya Krivelevich, Eisai Inc., Woodcliff Lake, NJ, USA
Simon Lin, Eisai Inc., Woodcliff Lake, NJ, USA

ABSTRACT

Pharmacokinetic (PK) data is collected for a variety of purposes in clinical trials. PK data is critical in understanding a drug's safety and determining its dosing frequency. Pharmacokinetic analysis is a major part of clinical trials to obtain information on drug disposition in humans.

Graphics are always a powerful tool in reporting, and Concentration vs. Timeline graphs are standard for reporting PK data. Data visualization tools are widely used in PK analysis, so meaningful graphics play a very important role.

There are many presentations and examples that deal with displaying results of rich sample collections; however, there is a significant shortage of similar presentations that cover the visualization of results of the widely used sparse PK data collection. This article may be interesting from a methodology standpoint; it presents examples and suggestions of step-by-step development of figures for sparse PK concentration using powerful ODS SAS tools (PROC SGPLOT).

INTRODUCTION

Pharmacokinetic (PK) data (plasma drug concentrations) are routinely collected in early drug development clinical studies (with rich, or in another word, intense, sampling collection). From an industry perspective, the PK characteristics in early phase clinical trials aid in the selection of promising molecules and help to make a go/no-go decision.

Sparse concentration sampling is very common in Late Phase (Phase II and III) studies. Sparse PK information implies that few samples are collected during each study or study day from each patient in the trial, while rich PK information implies that complete concentration-time profiles were determined based on the data collected. Under certain circumstances, because of ethical and medical concerns, only a limited number of samples are taken from any given patient, leading to sparse PK sampling. Sparse PK sampling is also very common in pediatric studies to minimize the total volume of blood sampled.

As usual, results of PK sampling might need visual representation. The anticipated graph will have a combination of timepoint boxplots overlaid with scatterplots. Boxplots are commonly used to visually compare distributions and summary statistics of a continuously measured response variable across two or more distinct timepoints. Overlaying a scatterplot on a boxplot can provide additional information about distributional differences or similarities. This type of overlay became compatible with the PROC SGPLOT starting from version 9.4M3. Starting with this version of SAS, bar charts can be done over a linear category axis.

STUDY DESIGN AND INPUT DATA

According to study protocol, each patient received daily administration of oral study drug ABCD. PK blood samples are drawn on:

- Cycle 1 Day 1 (1 - 4 hours postdose)
- Cycle 1 Day 15 (predose; 1 - 4 hours postdose; 6 - 9 hours postdose)
- Cycle 2 Day 1 (predose; 2 - 6 hours postdose)
- Cycle 3 Day 1 (predose)

The collected PK data were mapped into SDTM domain PC (Pharmacokinetics), then later, the corresponding ADaM dataset ADPC was created to support the needed analysis. Please see **Table 1** below for selected variables from the first 10 observations of ADPC. (For the variable ARRLT (Actual Relative Time from Reference TPT), the unit is collected in a different variable, RRLTU (Relative Time from Reference TPT Units) and equals to 'HOURS'). All records were created for the same Parameter (ADPC variable PARAM): 'ABCD Plasma Concentration (ng/mL)'

STEP 1: STARTING BOXPLOT

First, let's draw a simple boxplot as a starting point. Applying below simple SAS code with PROC SGPLOT:

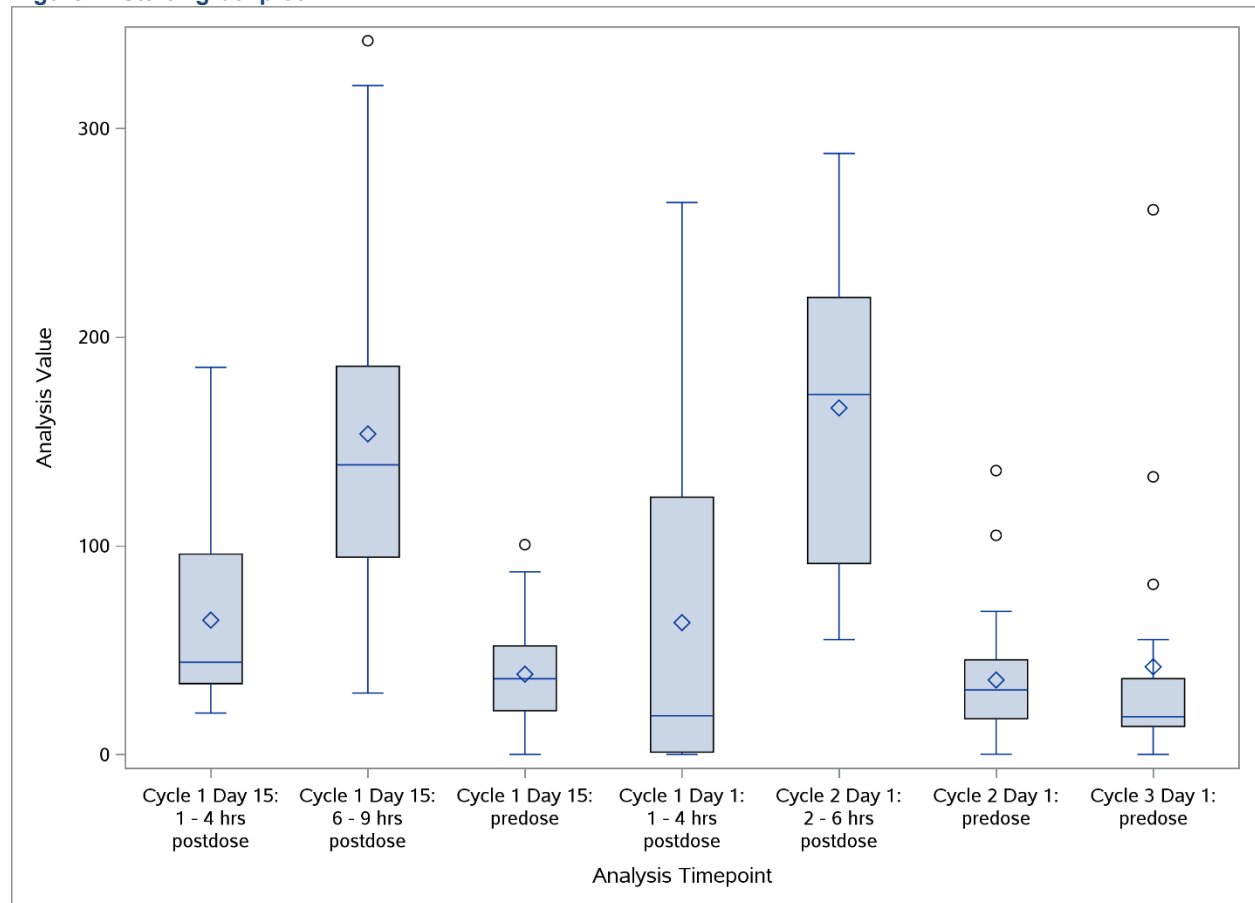
```
proc sgplot data = from_adpc;  
  vbox aval / category = atpt;  
run;
```

the following **Figure 1** was produced.

Table 1: First 10 Observations of Input Analysis Data Set ADPC

Subject Identifier for the Study (SUBJID)	Visit Name (VISIT)	Visit Number (VISITNUM)	Planned Time Point Name (PCTPT)	Planned Time Point Number (PCTPTNUM)	Analysis Timepoint (ATPT)	Analysis Value (AVAL)	Actual Relative Time from Reference TPT (ARRLT)
1001	Cycle 1 Day 1	1.01	1 - 4 hrs postdose	2	Cycle 1 Day 1: 1 - 4 hrs postdose	44.6	0.50000
1001	Cycle 1 Day 15	1.15	predose	0	Cycle 1 Day 15: predose	41.8	
1001	Cycle 1 Day 15	1.15	1 - 4 hrs postdose	2	Cycle 1 Day 15: 1 - 4 hrs postdose	40.9	0.66667
1001	Cycle 1 Day 15	1.15	6 - 9 hrs postdose	4	Cycle 1 Day 15: 6 - 9 hrs postdose	190.5	6.16667
1001	Cycle 2 Day 1	2.01	predose	0	Cycle 2 Day 1: predose	38.2	
1001	Cycle 2 Day 1	2.01	2 - 6 hrs postdose	3	Cycle 2 Day 1: 2 - 6 hrs postdose	66.0	1.50000
1001	Cycle 3 Day 1	3.01	predose	0	Cycle 3 Day 1: predose	261.0	
1002	Cycle 1 Day 1	1.01	1 - 4 hrs postdose	2	Cycle 1 Day 1: 1 - 4 hrs postdose	1.8	0.88333
1002	Cycle 1 Day 15	1.15	predose	0	Cycle 1 Day 15: predose	44.6	
1002	Cycle 1 Day 15	1.15	1 - 4 hrs postdose	2	Cycle 1 Day 15: 1 - 4 hrs postdose	41.8	0.75000

Figure 1: Starting boxplot



STEP 2: ORDERING BOXPLOTS AND ARRANGING THEIR LOCATIONS; REMOVING OUTLIERS AND FILLINGS

As indicated in the Study Design and Input Data section earlier, blood samples were drawn during 4 visits — Cycle 1 Day 1 (one timepoint), Cycle 1 Day 15 (three different timepoints), Cycle 2 Day 1 (two different timepoints), and Cycle 3 Day 1 (one timepoint). The vertical bars for different timepoints (X-axis) in the anticipated graph are expected to be drawn in chronological order. It also makes sense to separate visits in some way. We propose that the updated graph should have the distance between bars from different visits be twice as big as the distance between timepoint bars within the same visit. For this, we can create a new numeric variable, X_ATPT, having the following values, representing the chronological sequence in required order while separating visits. Later, the values of this variable will be used as horizontal coordinate of each boxplot:

- 1 for Cycle 1 Day 1: 1 - 4 hrs postdose
- 3 for Cycle 1 Day 15: predose
- 4 for Cycle 1 Day 15: 1 - 4 hrs postdose
- 5 for Cycle 1 Day 15: 6 - 9 hrs postdose
- 7 for Cycle 2 Day 1: predose
- 8 for Cycle 2 Day 1: 2 - 6 hrs postdose
- 10 for Cycle 3 Day 1: predose

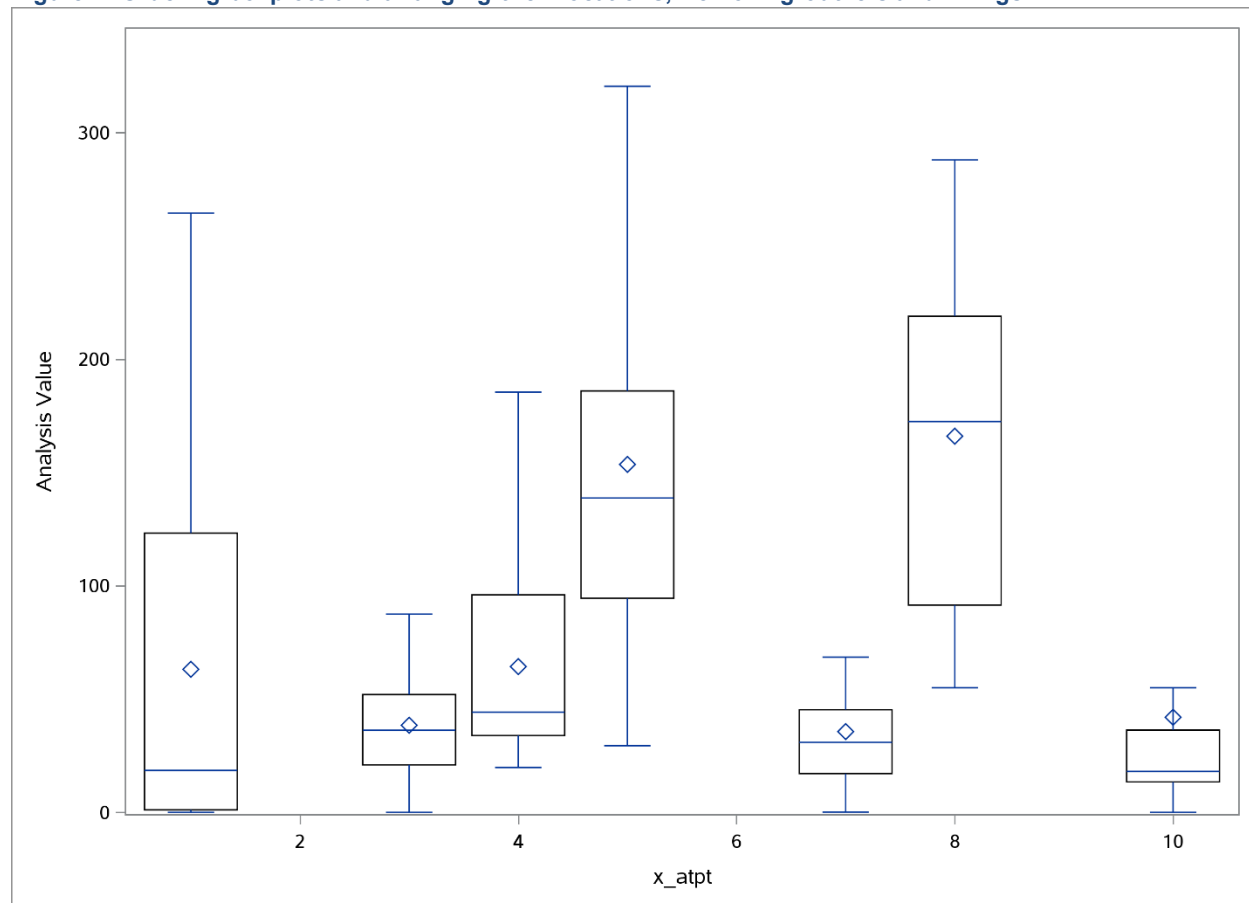
As mentioned above, starting from version 9.4M3 of SAS, bar charts in PROC SGPLOT can be drawn over a linear category axis.

The following steps will display the results of individual PK samplings as dots in scatterplots. In this case we do not need to display outliers; it also makes sense to eliminate the boxes' fillings.

The SAS code below will produce **Figure 2**:

```
proc sgplot data = from_adpc;  
  vbox aval / category = x_atpt nooutliers nofill;  
  xaxis type = linear;  
run;
```

Figure 2: Ordering boxplots and arranging their locations; Removing outliers and fillings



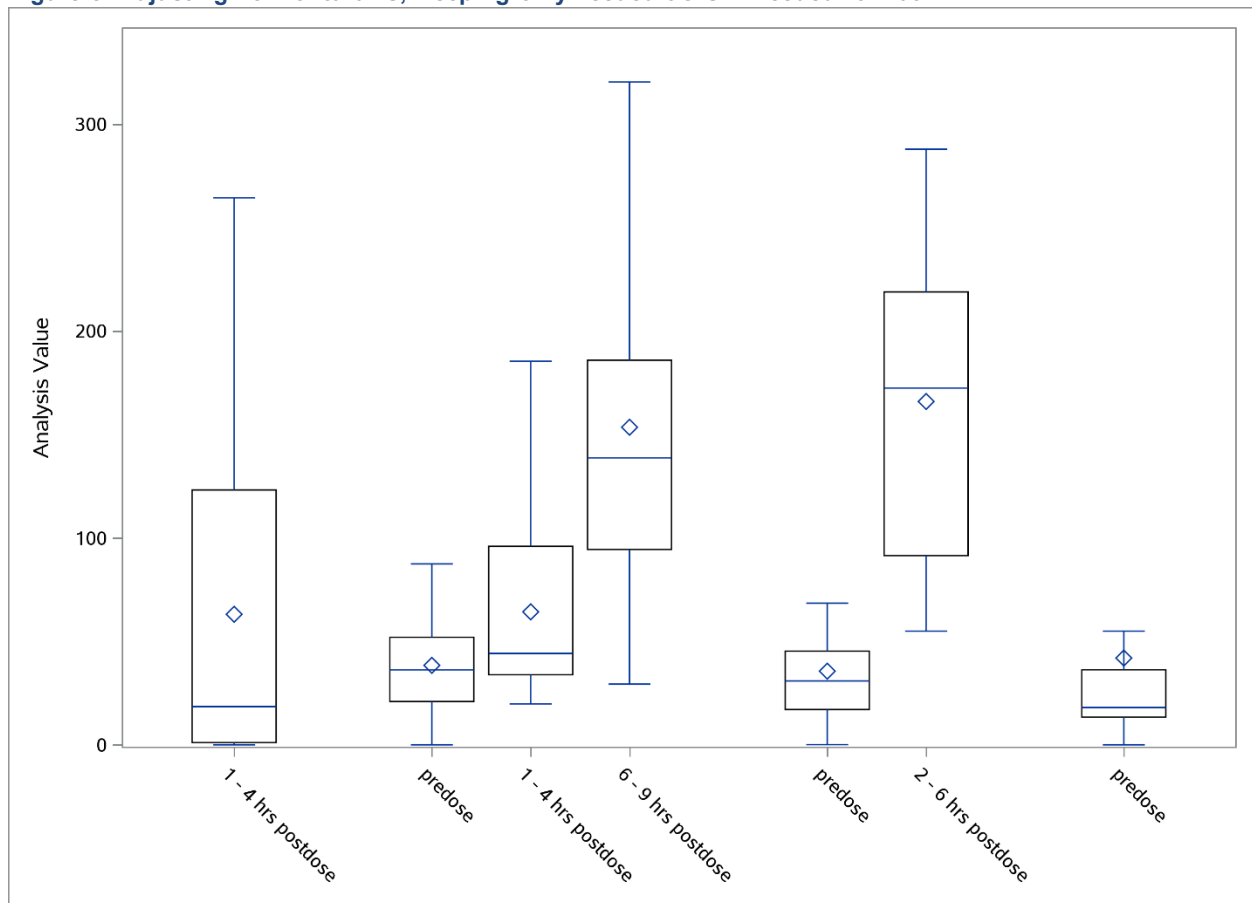
STEP 3: ADJUSTING HORIZONTAL AXIS; KEEPING ONLY NEEDED TICKS IN NEEDED FORMAT

On the horizontal axis, we need to have only the ticks that correspond to boxplots (in our case, values 1, 3, 4, 5, 7, 8, and 10). We would also like to format these values to identify the timepoints with meaningful text.

Let's first collect the values of all needed timepoints to be identified. For this, we have the following sample SAS code:

```
proc sort data = from_adpc nodupkey out = for_hor_axis (keep = x_atpt pctpt visit);
  by x_atpt;
run;
data for_hor_axis;
  set for_hor_axis end = endofset;
  length needed_values $200;
  retain needed_values ' ';
  needed_values = strip (needed_values) || ' ' || strip (put (x_atpt, best.));
  if endofset then call symput ('needed_hor_values', needed_values);
run;
%put Value of macro variable needed_hor_values is equal to &needed_hor_values;
In the log window, we will see:
Value of macro variable needed_hor_values is equal to 1 3 4 5 7 8 10
```

Figure 3: Adjusting horizontal axis; Keeping only needed ticks in needed format



Next time, when we draw a horizontal axis, we need to format these values with the timepoint information collected in the data set ADPC. The new format to be used with the horizontal axis (let's name it `tpt_values`), can be created programmatically:

```
data new_format1;
  set for_hor_axis;
  retain fmtname 'tpt_values';
  length label $60;
  start = x_atpt;
  label = pctpt;
run;
```

```
proc format cntlin = new_format1;
run;
```

We only need to add one statement to the SAS code developed earlier (inside PROC SGPLOT):

```
xaxis type = linear display = (nolabel) values = (&needed_hor_values)
      valuesformat = tpt_values. fitpolicy = rotate offsetmin = 0.1 offsetmax = 0.1;
```

to get **Figure 3**.

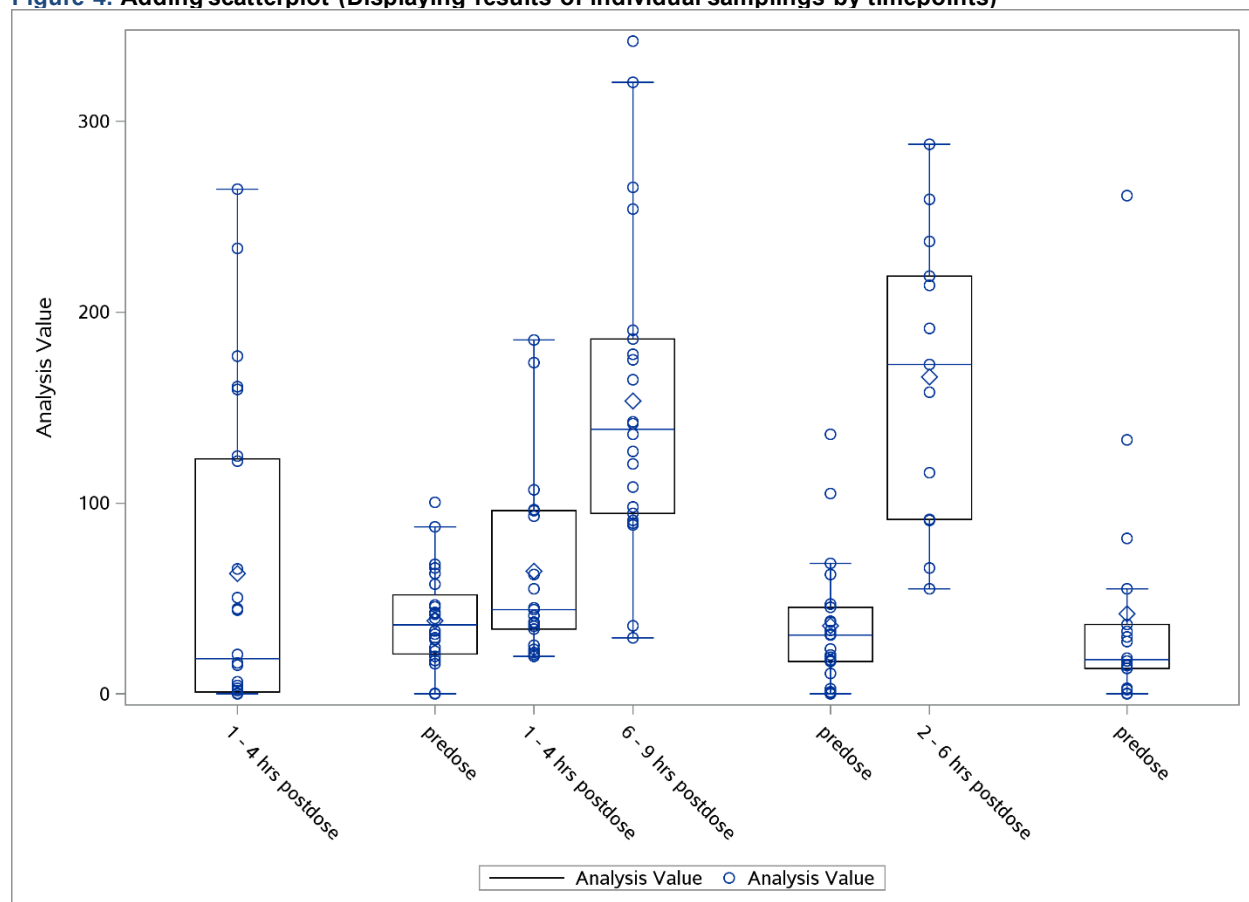
STEP 4: ADDING SCATTERPLOT (DISPLAYING RESULTS OF INDIVIDUAL SAMPLINGS BY TIMEPOINTS)

Now, it is time to add the scatterplot. Each dot will be displayed at the corresponding timepoint. This can be done easily by adding a scatter statement inside of PROC SGPLOT:

```
scatter x = x_atpt y = aval;
```

After this step, we are getting improved figure version of figure: **Figure 4**.

Figure 4: Adding scatterplot (Displaying results of individual samplings by timepoints)



STEP 5: ADJUSTING SCATTERPLOT; SHIFTING VALUES OF INDIVIDUAL SAMPLING RESULTS USING ELAPSED TIME INFORMATION; ELIMINATING AUTOMATICALLY CREATED LEGEND

The ADaM dataset ADPC, which is being used as the source, contains one important variable for pharmacokinetic data analysis - Actual Relative Time from Reference TPT (ARRLT). The unit is collected in a different variable, RRLTU (Relative Time from Reference TPT Units), and equal to "HOURS". In our study, this information was collected only at the postdose timepoints. It is reasonable to shift the dots for results of concentrations of individual sampling according to the information provided by this variable ARRLT (Actual Relative Time from Reference TPT). The location of the boxplots is associated with the center of the corresponding time intervals. In this case, we need one additional variable, that will give us the value of the horizontal coordinates of each dot for the results of individual sampling. Let's name this variable `x_scatter`.

```
data from_adpc;
  set from_adpc;
  if put (x_atpt, tpt_values. ) = 'predose' then x_scatter = x_atpt;
  if put (x_atpt, tpt_values. ) ^= 'predose' and arrlt ^= . then
```

```

do;
  if put (x_atpt, tpt_values. ) = '1 - 4 hrs postdose' then
    x_scatter = x_atpt + (ar1t - 2.5) * 0.375 / 2;
  if put (x_atpt, tpt_values. ) = '2 - 6 hrs postdose' then
    x_scatter = x_atpt + (ar1t - 4) * 0.375 / 2;
  if put (x_atpt, tpt_values. ) = '6 - 9 hrs postdose' then
    x_scatter = x_atpt + (ar1t - 7.5) * 0.375 / 2;
end;
run;

```

In this step, we are also going to eliminate the automatically created legend using the option `noautolegend` in PROC SGPLOT. As another minor improvement, we are going to display the results of individual plasma concentrations as filled circles.

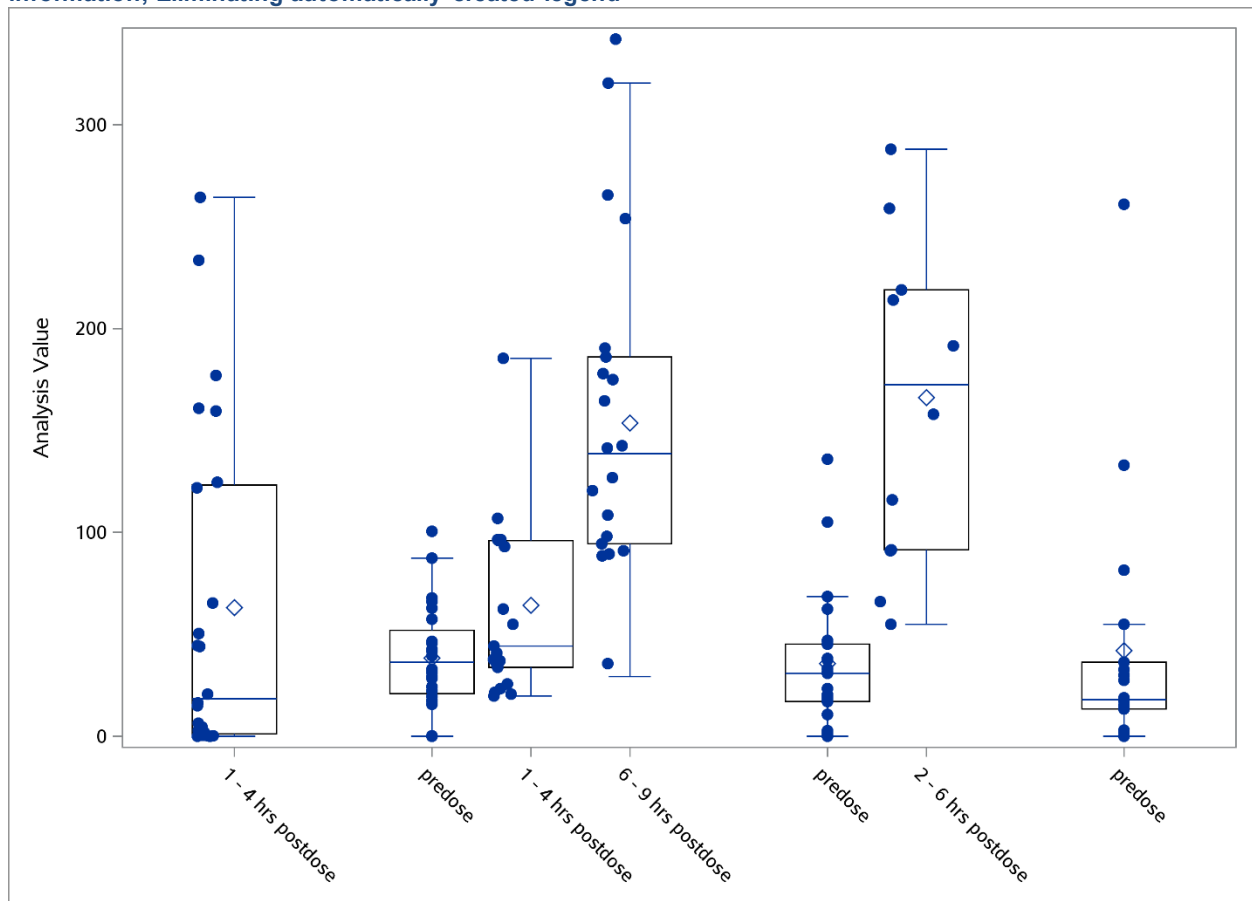
```

proc sgplot data = from_adpc noautolegend;
  vbox aval / category = x_atpt nooutliers nofill;
  scatter x = x_scatter y = aval / jitter markerattrs = (symbol = circlefilled);
  xaxis type = linear display = (nolabel) values = (&needed_hor_values)
    valuesformat = tpt_values. fitpolicy = rotate offsetmin = 0.1 offsetmax = 0.1;
run;

```

The improved PROC SGPLOT below will produce **Figure 5**:

Figure 5: Adjusting scatterplot; Shifting values of individual sampling results using Elapsed Time information; Eliminating automatically created legend



STEP 6: ADJUSTING VERTICAL AXIS

Now it is time to take care of the vertical axis. This can be done easily by using the following statement:

```

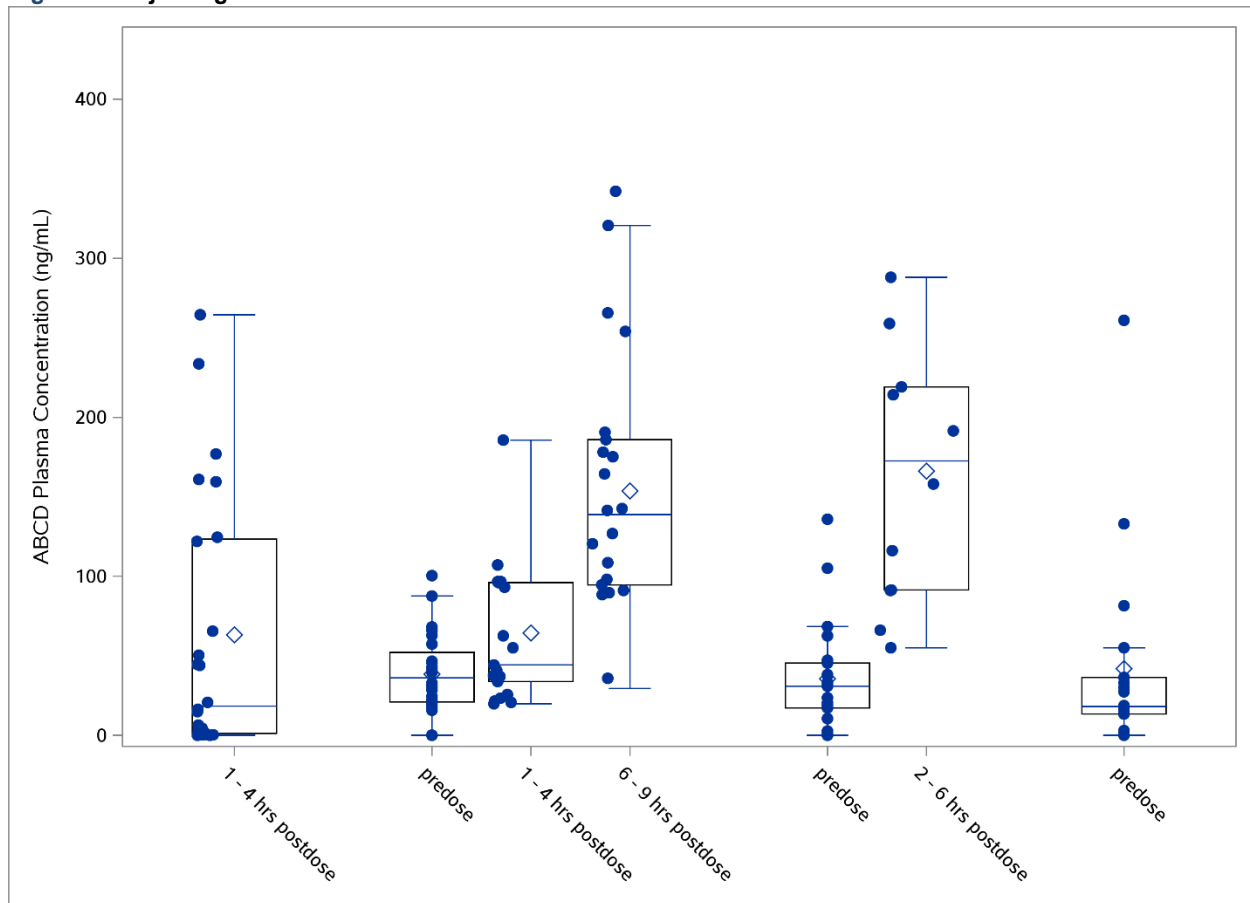
yaxis type = linear label = 'ABCD Plasma Concentration (ng/mL)'
  thresholdmin = 1 thresholdmax = 1 offsetmax = 0.1;

```

Specifying `thresholdmin = 1` and `thresholdmax = 1` ensures that the data range is bounded by tick marks. We need to reserve space in the plot for information about visits. This information will be displayed in the topmost 10% of a future graph; we are currently reserving room for this information, identified by `offsetmax = 0.1`.

After this step, we are getting an updated and improved version of figure: [Figure 6](#).

Figure 6: Adjusting vertical axis



STEP 7: ADDING INFORMATION ABOUT VISITS AS AN ANNOTATION

As we saw above, blood sampling for PK concentration is scheduled during four visits: Cycle 1 Day 1 (one timepoint), Cycle 1 Day 15 (three timepoints), Cycle 2 Day 1 (two timepoints), and Cycle 3 Day 1 (one timepoint). We already took care of displaying the timepoint information (see the tick values on the horizontal axis). Now it's time to display visit information as an annotation; we already reserved the topmost 10% of the figure for this in the previous step. It makes sense to display the visit information in the center, above the timepoints related to each visit.

Now let's create an annotation file. We can use the previously created dataset `for_hor_axis` as the source.

```
data for_visit_annotation;
  set for_hor_axis;
  by visit;
  retain min_visit max_visit . ;
  if first.visit then
    do;
      min_visit = . ;
      max_visit = . ;
    end;
  if first.visit then min_visit = x_atpt;
  if last.visit then max_visit = x_atpt;
  if last.visit then
    do;
      center_visit = (min_visit + max_visit) / 2;
      output;
      keep visit center_visit;
    end;
run;
```

```

data for_visit_annotation;
  set for_visit_annotation;
  x1space = 'datavalue';
  x1 = center_visit;
  y1space = 'wallpercent';
  y1 = 95;
  function = 'text';
  label = visit;
  textweight = 'bold';
  width = 50;
run;

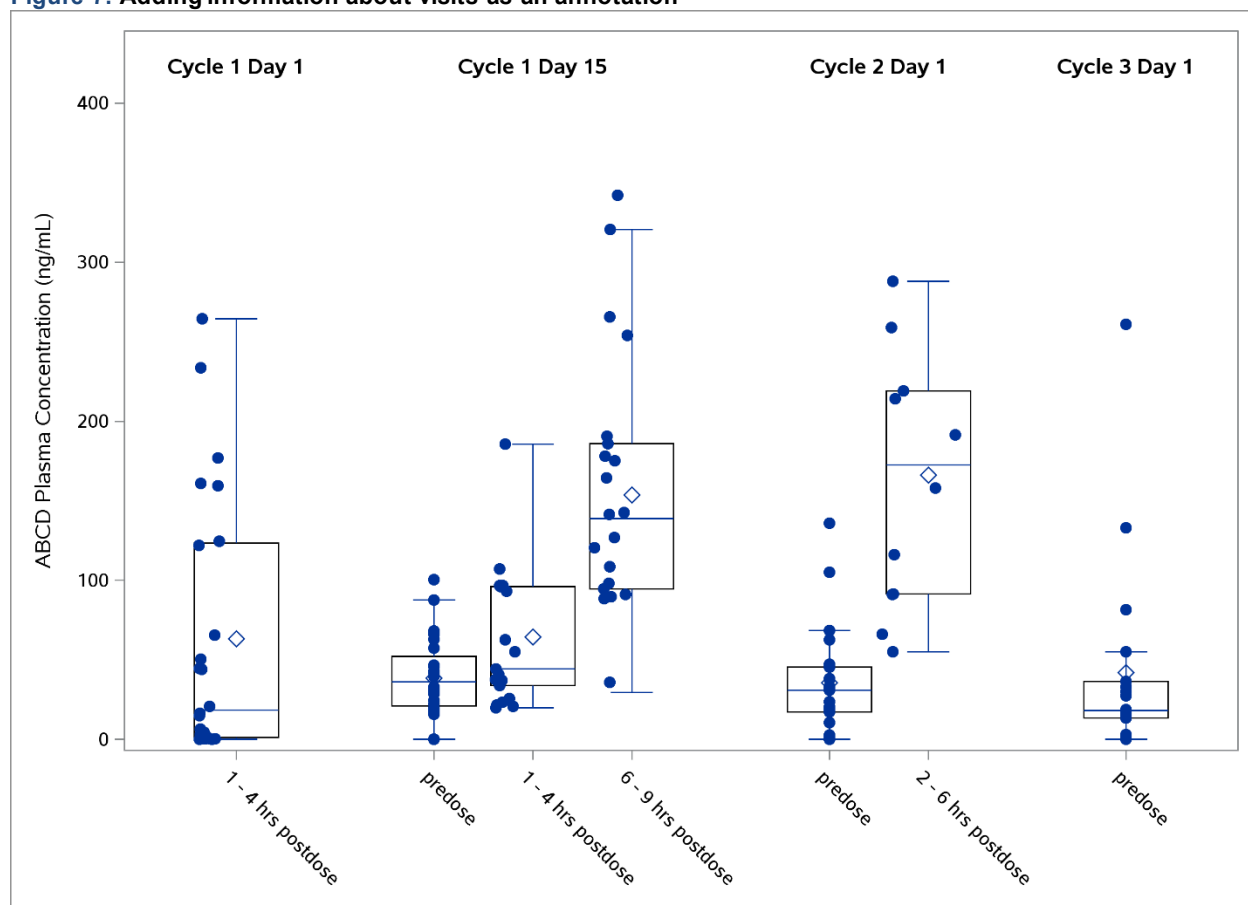
```

This data set `for_visit_annotation` will be applied in PROC SGPLOT statement as:

```
proc sgplot data = from_adpc sganno = for_visit_annotation noautolegend;
```

As a result, our figure will get another improvement - see **Figure 7**.

Figure 7: Adding information about visits as an annotation



STEP 8: ADDING NECESSARY STATISTICS

The next step in improving the figure is adding the necessary statistics for each timepoint. These can be N (Number of subjects), Mean, GeoMean (Geometric Mean), Median, and StD (Standard Deviation). However, each user may add or remove any statistics, according to their Pharmacokinetic Analysis Plan.

The statistics for all the timepoints can easily be obtained from `proc univariate` and saved as the output dataset. All values of the variable AVAL must be positive when the geometric mean is requested. If some values of the AVAL variable are not positive, the study statistician and study pharmacological director must decide how to handle this situation. This mostly occurs when the collected value of concentration is reported as BLQ (Below Level of Quantification) and this value is converted into AVAL in ADPC as a 0. One of the possible solutions in case of BLQ is to convert the provided SDTM concentration into AVAL, which is equal to half of the provided Level of Quantification. We can run a simple SAS code to obtain the necessary statistics. In this code, we will also handle the precision of these statistics.

```

proc univariate data= from_adpc noprint;
  class x_atpt;
  var aval;
  where aval > 0;
  output out = stats n = n mean = mean std = std median = median geomean = geomean;
run;
data stats;
  set stats;
  format mean median geomean 8.1;
  format std 8.2;
run;

```

The dataset final is a concatenation of the previously used datasets `from_adpc` and the newly created dataset `stats`:

```

data final;
  set from_adpc stats;
run;

```

Using two additional statements inside PROC SGPLOT allows us to display the required statistics for each timepoint below the horizontal axis:

```

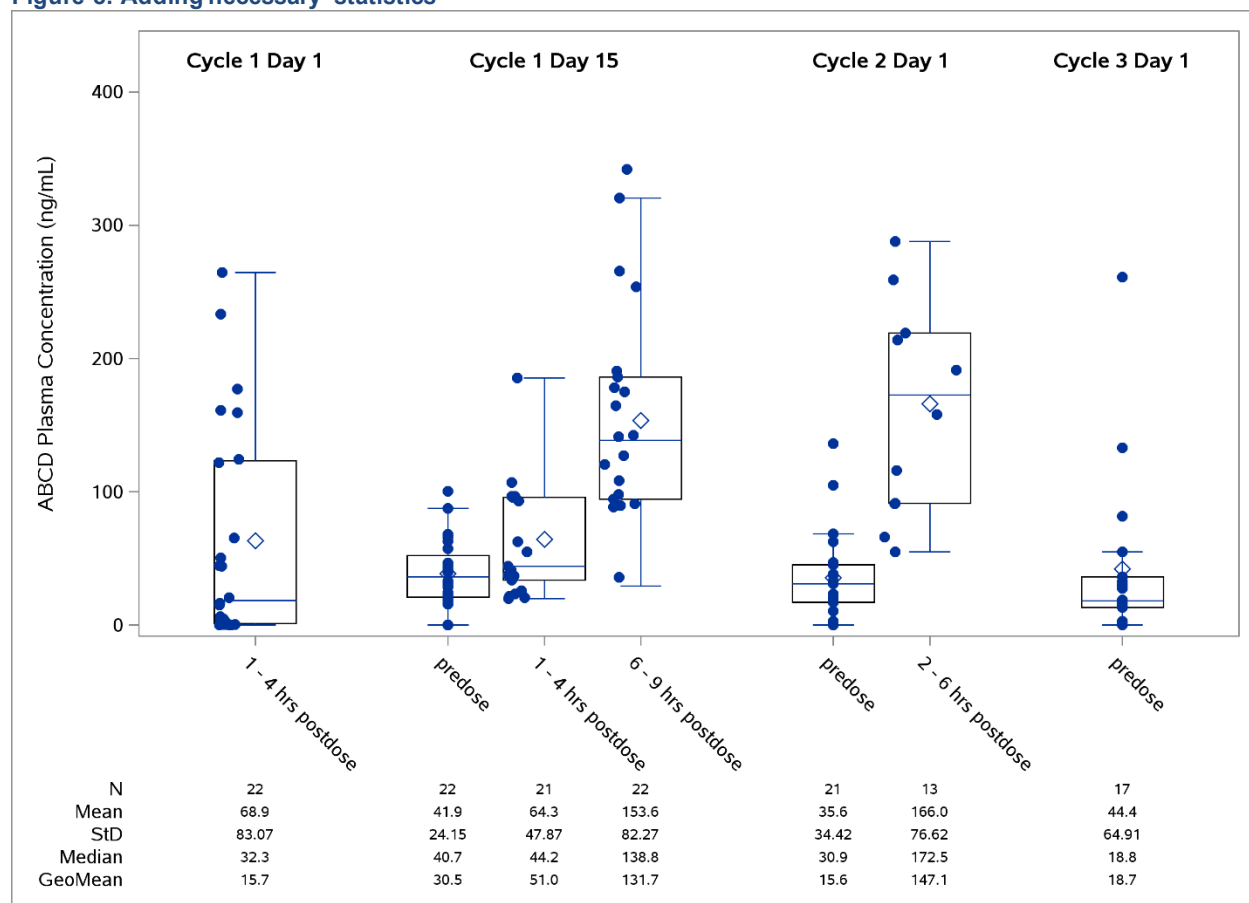
label n = 'N' mean = 'Mean' std = 'StD' median = 'Median' geomean = 'GeoMean';
xaxistable n mean std median geomean / location = outside;

```

Keep in mind that, starting from this step, our figure uses the recently produced dataset `final` as a source (we need to specify `data = final` in the PROC SGPLOT statement).

Now, our figure is almost finalized. See **Figure 8**.

Figure 8: Adding necessary statistics



STEP 9: TITLES / FOOTNOTES

Of course, each plot should contain the required titles and footnotes. They can easily be added by using `title` and `footnote` statements. Afterwards, the entire PROC SGPLOT statement will look like this:

```

proc sgplot data = final sganno = for_visit_annotation noautolegend;

```

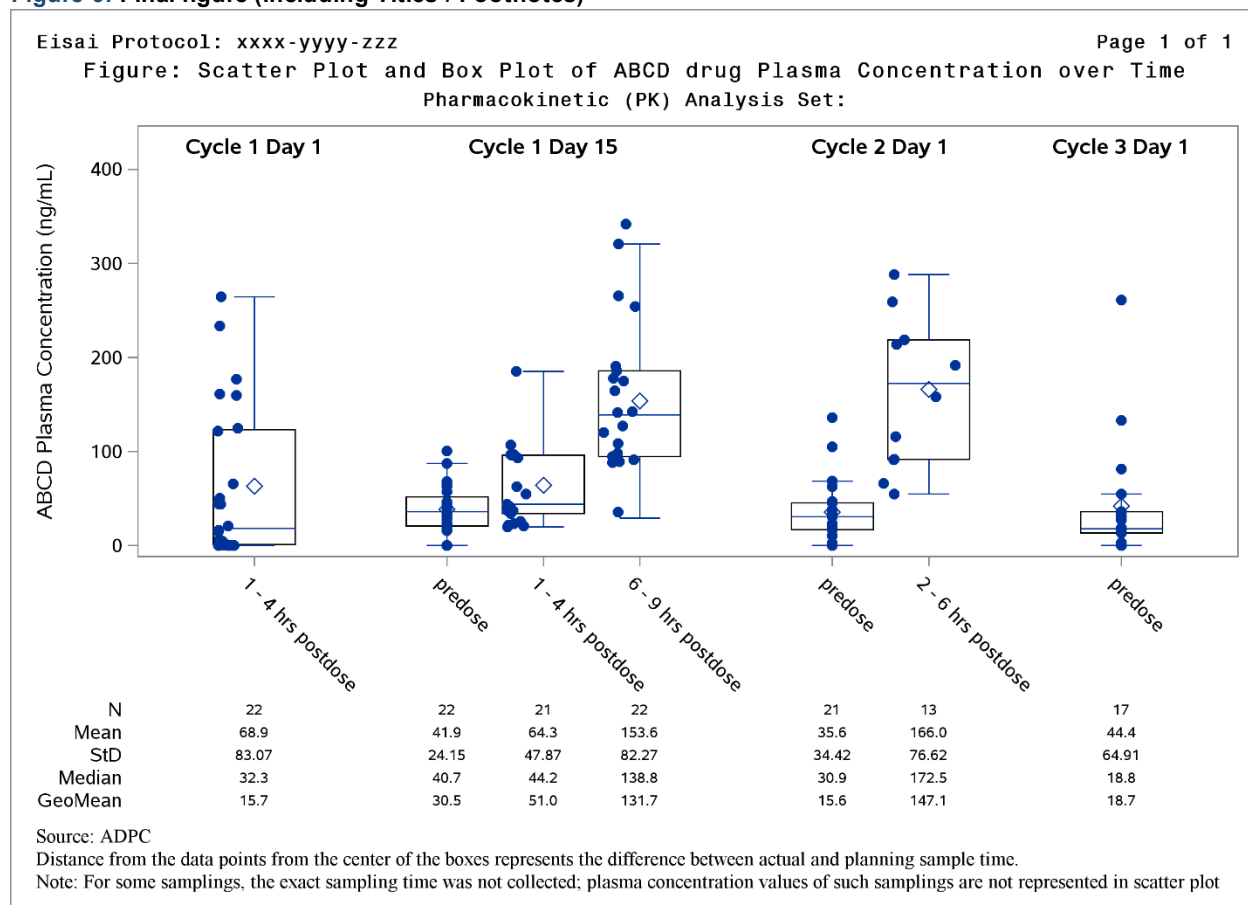
```

vbox aval / category = x_atpt nooutliers nofill;
scatter x = x_scatter y = aval / jitter markerattrs = (symbol = circlefilled);
axis type = linear display = (nolabel) values = (&needed_hor_values)
    valuesformat = tpt_values. fitpolicy = rotate offsetmin = 0.1 offsetmax = 0.1;
yaxis type = linear label = 'ABCD Plasma Concentration (ng/mL)'
    thresholdmin = 1 thresholdmax = 1 offsetmax = 0.1;
label n = 'N' mean = 'Mean' std = 'StD' median = 'Median' geomean = 'GeoMean';
xaxistable n mean std median geomean / location = outside;
title1 font = 'SAS Monospace' color = black height = 9 pt justify = left
    'Eisai Protocol: xxxx-yyyy-zzz' justify = right 'Page 1 of 1';
title2 font = 'SAS Monospace' color = black height = 10 pt justify = center
    'Figure: Scatter Plot and Box Plot of ABCD drug Plasma Concentration over
    Time';
title3 font = 'SAS Monospace' color = black height = 9 pt justify = center
    'Pharmacokinetic (PK) Analysis Set: ';
footnote1 font = 'Times New Roman' color = black height = 9 pt justify = left
    "Source: ADPC";
footnote2 font = 'Times New Roman' color = black height = 9 pt justify = left
    "Distance from the data points from the center of the boxes represents the
    difference between actual and planning sample time.";
footnote3 font = 'Times New Roman' color = black height = 9 pt justify = left
    "Note: For some sampling, the exact sampling time was not collected; plasma
    concentration values of such samplings are not represented in scatter plot";
run;

```

And finally, we are getting finalized version – see **Figure 9**.

Figure 9: Final figure (including Titles / Footnotes)



POSSIBLE FURTHER STEPS

Of course, the reader can continue the improvement process, such as applying different colors, fonts, attribute map, and other features. This is not the subject of this article, but each reader is welcomed to do so. For more details, please see SAS® 9.4 ODS Graphics: Procedures Guide

CONCLUSION

By applying the techniques described in this paper, the reader can successfully visualize results of sparse Pharmacokinetic sampling. We suggest creating this type of graph, a combination of boxplot and scatterplots that also displays the required statistics. Similar graphs (a combination of boxplots and scatterplots with displaying required statistics) can be used in visualizing results of other discrete measurements, such as Vital Signs, ECG, or Laboratory Data. This is one of the first articles to demonstrate that the combination of a boxplot and scatterplot in the same graph can be easily achieved by using PROC SGPLOT without using more complicated Graph Template Language.

This article can be used as methodological guidance, with the main idea: Do not try to create complicated figures all at once, do it step by step (improvement by improvement).

REFERENCES:

SAS® 9.4 ODS Graphics: Procedures Guide, Sixth Edition. Copyright©2016, SAS Institute Inc., Cary, NC, USA

Rajnikanth Madabushi, Jeffry Florian, Fang Li, Christoffer W. Tornøe, Christine Garnett, Hao Zhu, Yaning Wang, Nitin Mehrotra, Anshu Marathe, Jiang Liu, Venkatesh A. Bhattaram, Pravin R. Jadhav, Jogarao V. Gobburu, Joo-Yeon Lee, Kevin Krudys, Justin C. Earp
PK in Late Phase Trials
Applied Clinical Trials - Volume 21, Issue 2
Available from: <https://www.appliedclinicaltrialsonline.com/view/pk-late-phase-trials>

Fred Wood, Octagon Research Solutions, Wayne, PA
Peter Schaefer, Certara, Cary, North Carolina
Richard Lewis, Octagon Research Solutions, Wayne, PA
Considerations in the Submission of Pharmacokinetics (PK) Data in an SDTM Compliant Format
PharmaSUG 2012 - Paper DS10
Available from: <https://www.lexjansen.com/pharmasug/2012/DS/PharmaSUG-2012-DS10.pdf>

Charles G. Minard
Dan L. Duncan Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX
Overlaying Scatter Plots on Box Plots in the Presence of Ties
Available from: <https://www.lexjansen.com/scsug/2012/SCSUG-2012-ScatterBox.pdf>

Kiran Cherukuri, Seattle Genetics Inc., Bothell, WA
Programming Pharmacokinetic (PK) Timing and Dosing Variables in Oncology Studies: Demystified
PharmaSUG 2015 – Paper PO06
Available from: <https://www.lexjansen.com/pharmasug/2015/PO/PharmaSUG-2015-PO06.pdf>

Pooja Trivedi, Cadila Healthcare Limited, Ahmedabad, India
Mean and Individual Subject Graphs of Concentration vs. Time Data Using PROC SGPLOT
PharmaSUG 2017 - Paper DV04
Available from: <https://www.lexjansen.com/pharmasug/2017/DV/PharmaSUG-2017-DV04.pdf>

James R. Johnson, Summit Analytical, LLC., Cary, NC USA
Methods for Handling Concentration Values Below the Limit of Quantification in PK Studies
PhUSE US Connect 2018 - Paper DH05
Available from: <https://www.lexjansen.com/phuse-us/2018/dh/DH05.pdf>

Shallabh Mehta, PPD
Pharmacokinetic Parameters for Sparse and Intensive Sampling - Nonclinical and Clinical Studies
PharmaSUG 2019 - Paper SS-172
Available from: <https://www.lexjansen.com/pharmasug/2019/SS/PharmaSUG-2019-SS-172.pdf>

ACKNOWLEDGMENTS

The authors would like to thank Dr. Robert Shumaker for sharing his ideas and providing scientific guidance

CONTACT INFORMATION

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

Ilya Krivelevich

Eisai Inc.

155 Tice Blvd.

Woodcliff Lake, NJ

Work Phone: 201-949-4204

Email: ilya_krivelevich@eisai.com