Paper SP01

Checking Symmetry in Predefined Time Assessments

Antonio Nieto (PharmaMar, Madrid, Spain) and José Luis Jiménez-Moro (PharmaMar, Madrid, Spain)

ABSTRACT

Overall survival (OS) is usually the gold standard endpoint in oncology phase III clinical trials comparing a new treatment with the current standard of care. In several situations, OS can be substituted by progression-free survival (PFS), which has to be measured in pre-specified time points. These assessments must be done consistently at the same intervals in both treatment arms to avoid imbalance that would result in a biased global comparison. Consequently, a crucial check (if not the first) would be to ensure that symmetry cannot be rejected. Different ways to conduct this check using SAS© are described herein, from Box-Plots graphs to Wilcoxon test to compare time to disease assessments. In addition, Kaplan-Meier curves of the time from randomization to the first and second disease assessments are also shown.

KEYWORDS

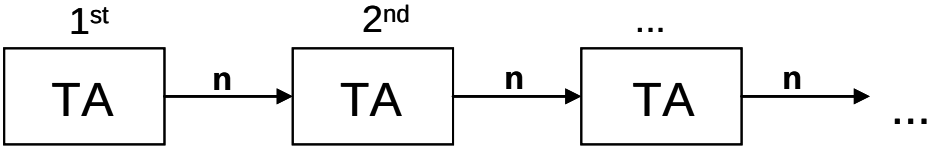
Symmetry; Progression Free Survival; Overall Survival;

INTRODUCTION

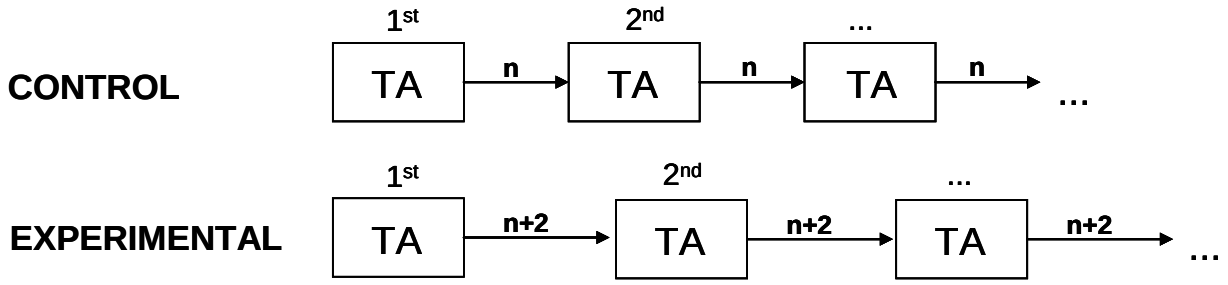
In Oncology, overall survival (OS) is the most reliable endpoint to assess clinical benefit when an experimental arm is compared with a control arm. Main disadvantages of OS endpoint are that it may involve larger sample size requirements and longer follow-up, it may be affected by crossover and sequential therapy and it includes non-cancer deaths[1].

Consequently, regulatory agencies several times accept other time to events endpoints (i.e. progression free survival [PFS]) assuming its surrogacy to OS. PFS is usually defined as the time between the randomization date and the progression date or the death date, irrespective of the cause of death. The progression date is calculated according to published international standards like the RECIST for solid tumors where the tumor extent during the study is assessed, e.g. an increase of 30% in the sum of the diameters in the target lesions respect to baseline value is considered as progression, other common criteria are for example WHO criteria for leukemia, Cheson criteria for lymphoma, Blade criteria for myeloma...

To avoid bias in the comparison, the protocol has to predefine tumor assessments which should be followed by the clinical trial conduct investigators. Therefore tumor assessments will be performed at times TA(n), TA(n+1), TA(n+2) where n is the established period…



Let's imagine that for whatever reason the patients treated on the experimental arm are assessed consistently two weeks later than the patients treated on the control arm. The comparison between treatment arms would be biased in favor of the treatment arm.



Real life shows us that it is not always possible to follow the time schedule rules so one of the first checks will be to ensure the symmetry between treatments. Herein we show different ways to do it with SAS.

HYPOTHETICAL SCENARIO

Suppose a new clinical trial where a new treatment is going to be compared with the standard of care. The primary endpoint selected is PFS and 'a priori' assumptions are α=0.025 (one-sided), ß=0.1 (power=90%), median PFS for the control arm is 24 weeks and 30% risk reduction with the experimental arm is expected, so HR=0.7 and median PFS expected for the treatment arm is 34.3 weeks. According to these assumptions, 331 events out of 400 patients are necessary. So let's start simulating two samples, each of 200 patients, following the expected exponential distributions[2].

```
Data auxiliar1;
 do pat=1 to 200;
 group=0;
 pfs = 24*ranexp(4054/*chosen seed to repeat simulation*/);
 event=1; /*all events, other option might be to create a random variable*/
 output;
 end;
 do pat=201 to 400;
 group=1;
 pfs = 34.3*ranexp(4054/* chosen seed to repeat simulation */);
 event=1; /*all events, other option might be to create a random variable*/
 output;
 end;
run;
```

Then imagine that according to the protocol, assessments have to be performed every six weeks (±1 week). Let's simulate this assessment scenario[3].

```
Data auxiliar2;
set auxiliar1;
assessments=1+int(pfs/6);
do assessment=1 to assessments;
time_assessment=assessment*(2*ranuni(3657)+5);
output;
end;
run;
```

To maintain simplicity throughout the document only the first two assessments are going to be used.

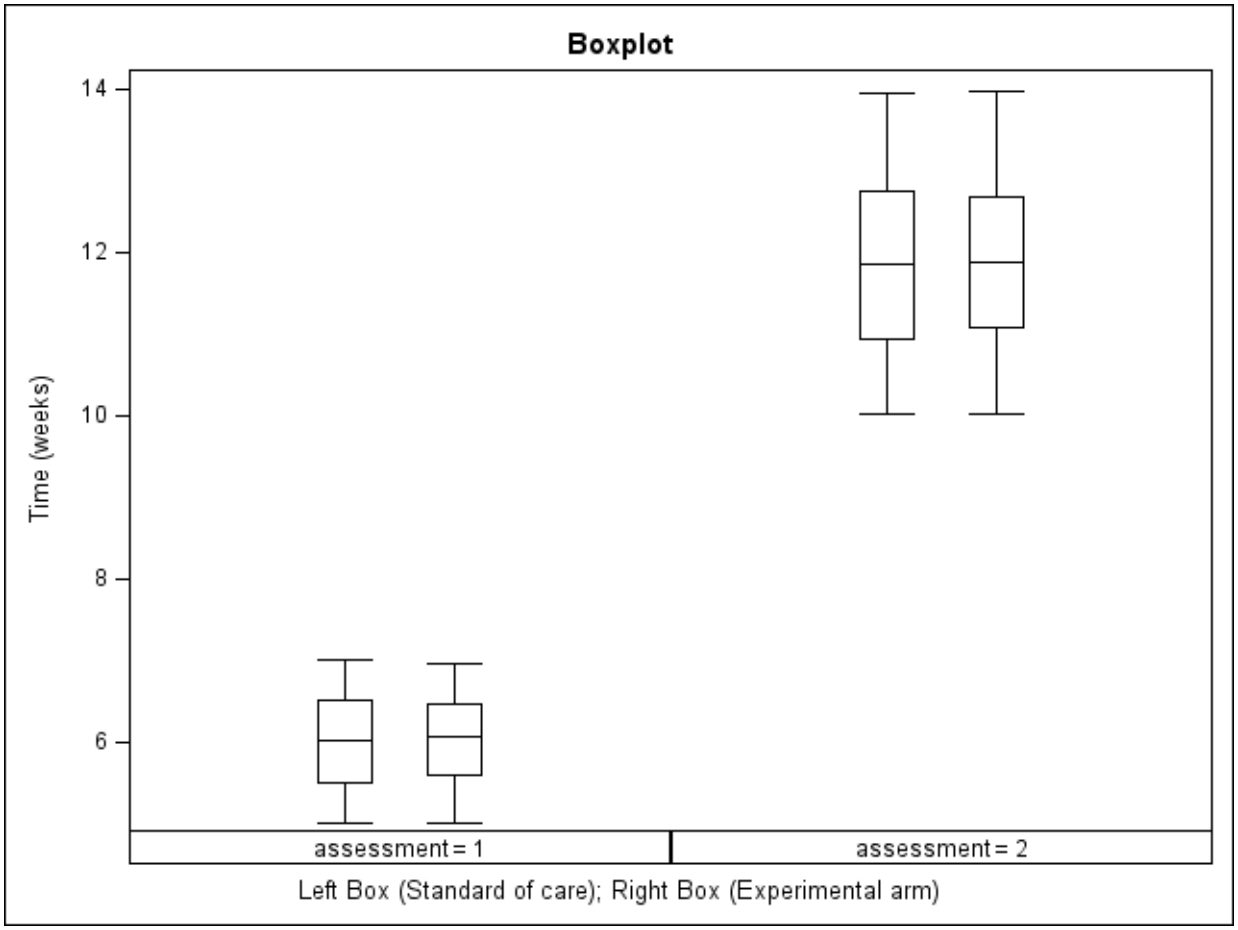
```
/*Restricted to the first 2 assessments*/
Data auxiliar2;
set auxiliar2;
If assessment<=2;
run;
```

BOXPLOT

The first graphical test to check if the symmetry is fulfilled can be performed by means of a Boxplot[4].

```
proc sort data = auxiliar2;
by assessment Group;
run;
proc means data = auxiliar2 noprint;
by assessment Group;
var time_assessment;
output out = BoxPlotStats mean = mean median = median q1 = q1 q3 = q3
qrange = qrange min = min max = max;
run;
data auxiliar3;
merge auxiliar2 BoxPlotStats;
by assessment Group;
run;
data auxiliar4;
set auxiliar3;
XBoxLeft = group - 0.25;
XBoxRight = group + 0.25;
XWhiskerLeft = group - 0.25;
XWhiskerRight = group + 0.25;
YWhiskerTop = (1.5*qrange) + q3;
YWhiskerBottom = q1 - (1.5*qrange);
if max le YWhiskerTop then
YWhiskerTop = max;
if min ge YWhiskerBottom then
YWhiskerBottom = min;
label median='Time (weeks)';
run;

%macro DrawBoxPlots;
* draw median line;
vector x = XBoxRight y = median/
noarrowheads xorigin = XBoxLeft yorigin = median;
* draw quantile box;
vector x = XBoxRight y = q1/
noarrowheads xorigin = XBoxLeft yorigin = q1;
vector x = XBoxRight y = q3/
noarrowheads xorigin = XBoxLeft yorigin = q3;
vector x = XBoxLeft y = q3/
noarrowheads xorigin = XBoxLeft yorigin = q1;
vector x = XBoxRight y = q3/
noarrowheads xorigin = XBoxRight yorigin = q1;
* draw whiskers;
vector x = group y = YWhiskerTop/
noarrowheads xorigin = group yorigin = q3;
vector x = group y = YWhiskerBottom/
noarrowheads xorigin = group yorigin = q1;
vector x = XWhiskerRight y = YWhiskerTop/
noarrowheads xorigin = XWhiskerLeft yorigin = YWhiskerTop;
vector x = XWhiskerRight y = YWhiskerBottom/
noarrowheads xorigin = XWhiskerLeft yorigin = YWhiskerBottom;
%mend;

title 'Boxplot';
footnote 'Left Box (Standard of care); Right Box (Experimental arm)';
proc sgpanel data = auxiliar4;
* define the structure of the panel;
panelby assessment/
layout = columnlattice colheaderpos = bottom noborder;
* leave some room between the columns of the panel;
colaxis display = none offsetmax = 0.35 offsetmin = 0.35;
```

```
keylegend/position = top noborder;
%DrawBoxPlots;
run;
```



As expected by the means, the assessments have been created and it seems they have been done correctly so there is no room to think that if the null hypothesis is rejected in the global comparison of PFS, the observed improvement in efficacy can be due to asymmetrical assessments between arms.

WILCOXON TEST

A non-parametric test as the Wilcoxon (also know as U-Mann-Whitney test in other statistical packages) can be implemented to check symmetry between assessments adding an associated p-value to the visual review.

```
Proc format;
value group
0='Standard of care'
1='Experimental';
run;

Proc means data=auxiliar2 noprint;
output out=auxiliar5 n=n p50=median;
var time_assessment;
by assessment Group;
run;

proc npar1way data=auxiliar2 wilcoxon;
var time_assessment;
```

```
class Group;
by assessment;
ods output WilcoxonTest=auxiliar6;
run;

Data auxiliar7;
set auxiliar6;
If Name1='P2_WIL';
keep assessment cValue1;
run;

Data auxiliar8;
merge auxiliar5 (in=recent) auxiliar7;
by assessment;
If recent~=0;
run;

title "Time to disease assessments";footnote;
proc report data=auxiliar8 nowindows headline headskip missing nocenter
split='@';
columns assessment group n median cValue1;
define assessment/group width=7 center order=data 'Disease assessment';
define group/display width=7 left order=data format=group. 'Treatment arm';
define n/display center order=data;
define median/display center order=data 'Median (weeks)' format=6.1;
define cValue1/group center order=data 'p-value';
run;
```

Time to disease assessments

Disease assessment	Treatment arm	n	Median (weeks)	p-value
1	Standard of care	200	6.0	0.7624
	Experimental	200	6.1	
2	Standard of care	149	11.8	0.9616
	Experimental	164	11.9	

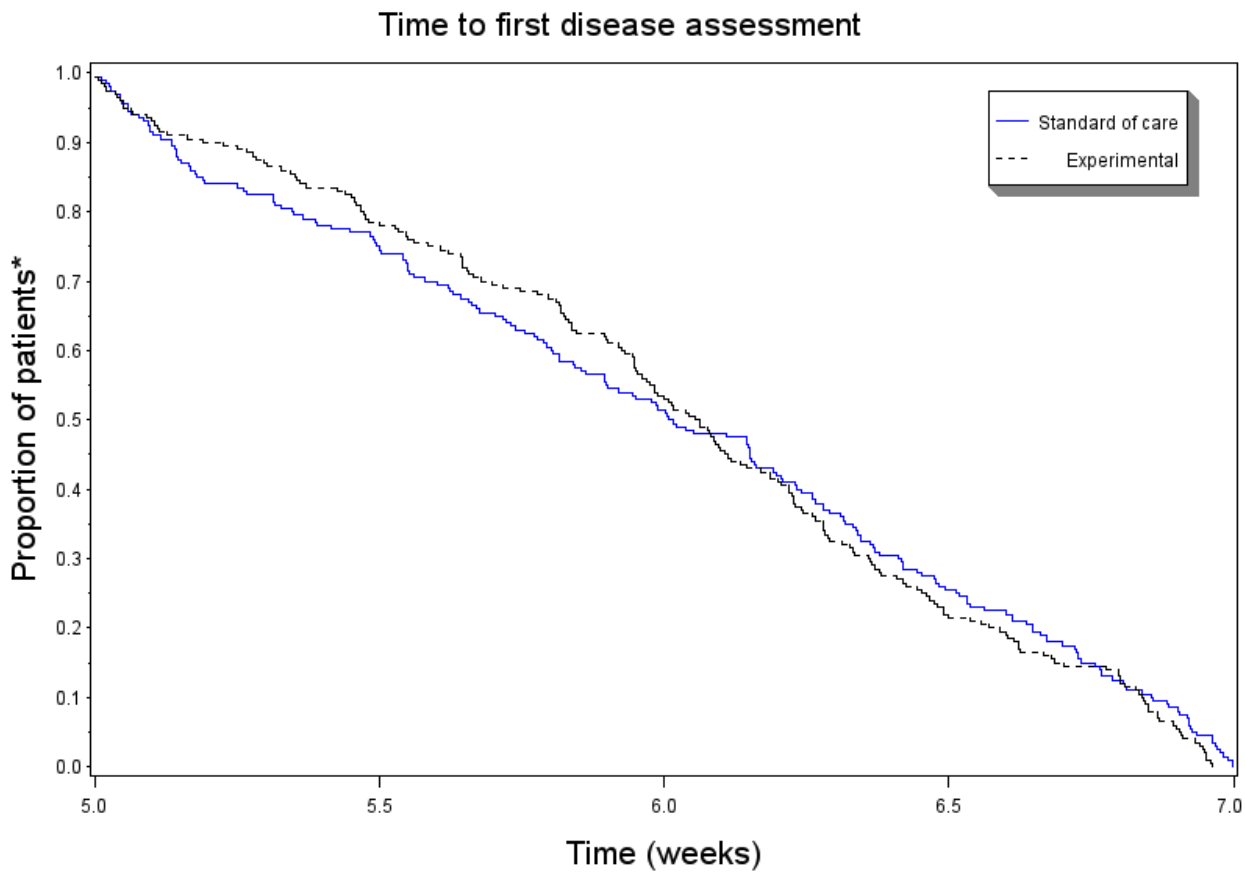
To reinforce the graphic results we can see that the null hypothesis cannot be rejected in any of the disease assessments.

"KAPLAN-MEIER" PLOT

Another visual approach to check time to disease assessments symmetry between treatment arms can be performed by means of a trick (using all times as events) Kaplan-Meier curve, where all the times between randomization and patients'assessments by each disease assessment can be plotted (e.g. time between randomization and first scheduled assessment, second assessment, third assessment,…)

```
title;footnote;
proc lifetest data=auxiliar2 noprint OUTSURV = salida;
  time time_assessment*assessment(0);
  strata group;
where assessment=1;
run;

goptions RESET=ALL gsfmode=replace device=png ftext="Arial" htext=1
htitle=2.5;
TITLE JUSTIFY=CENTER FONT='Arial /i' HEIGHT=16pt COLOR='BLACK' "Time to first
disease assessment";

AXIS1 COLOR=BLACK
 LABEL=(FONT='Arial' HEIGHT=16pt JUSTIFY=CENTER
        'Time (weeks)' )
 MAJOR=(COLOR=BLACK NUMBER=21)
```

```
 MINOR=(NUMBER=1)
 ORDER= 5 TO 7 BY 0.5;

AXIS2  COLOR=BLACK
 LABEL=(FONT='Arial' HEIGHT=16pt JUSTIFY=CENTER
 ANGLE=90 'Proportion that intervals were greater')
  ORDER= 0 TO 1 BY 0.1
 minor=(NUMBER=1);
    symbol1 interpol=join i=stepj ci=blue cv=blue l=1 width=1;
    symbol2 interpol=join i=stepj ci=black cv=black l=2 width=1;

legend1 label=none
        shape=symbol(4,2)
            across=1
        position=(TOP RIGHT INSIDE)
        mode=share cblock=grey offset=(-3 pct,-3 pct) frame;
proc gplot data=salida;
 PLOT
  SURVIVAL * time_assessment = group /
LEGEND=LEGEND1    FRAME HAXIS=AXIS1 VAXIS=AXIS2 HMINOR=0;
format group group.;
RUN;
QUIT;
TITLE;
RUN;
goptions RESET=ALL;
```



* Proportion of patients for whom the time from randomization to the disease assessment is greater that a given time t (in weeks)

In our example as expected there is not any difference between treatment arms, although the KM seems to have a small pattern this is due to chance and differences if any would be negligible. In Appendix I simulations where assessments have not been performed symmetrically can be found, if this would be the case one possible sensitivity analysis is to recalculated the PFS but using the projected tumor assessments timepoints instead of the real assessments performed.

CONCLUSION

This paper describes some SAS® codes to check if disease assessments between treatment arms have been performed symmetrically. Consequently, before reaching any conclusion in the global treatment comparison, the assumption that time to event variable affected by set assessments are asymmetrically measured should be discarded.

REFERENCES

1. Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. U.S. Department of Health and Human Services Food and Drug administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER) May 2007 Clinical/Medical.

2. https://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000202893.htm (accessed on 27DEC2013).

3. http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000202926.htm (accessed on 27DEC2013).

4. Grouped Jittered Box Plots in SAS 9.2 and SAS 9.3. Volker Harm and Catalina Mejía Herrera. Phuse 2012.

Contact Information

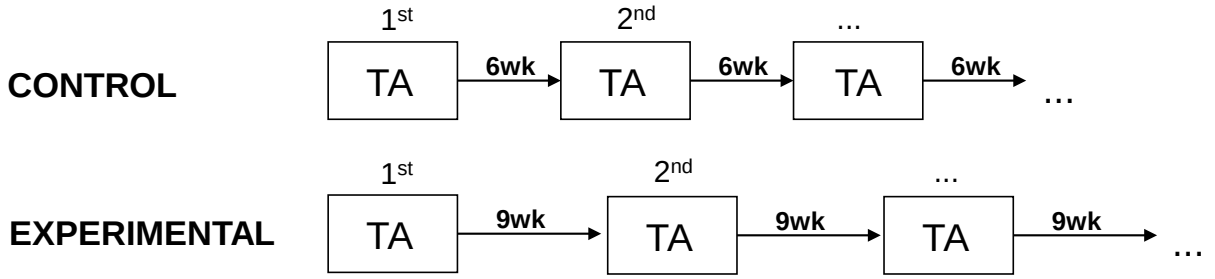
Your comments and questions are valued and encouraged.

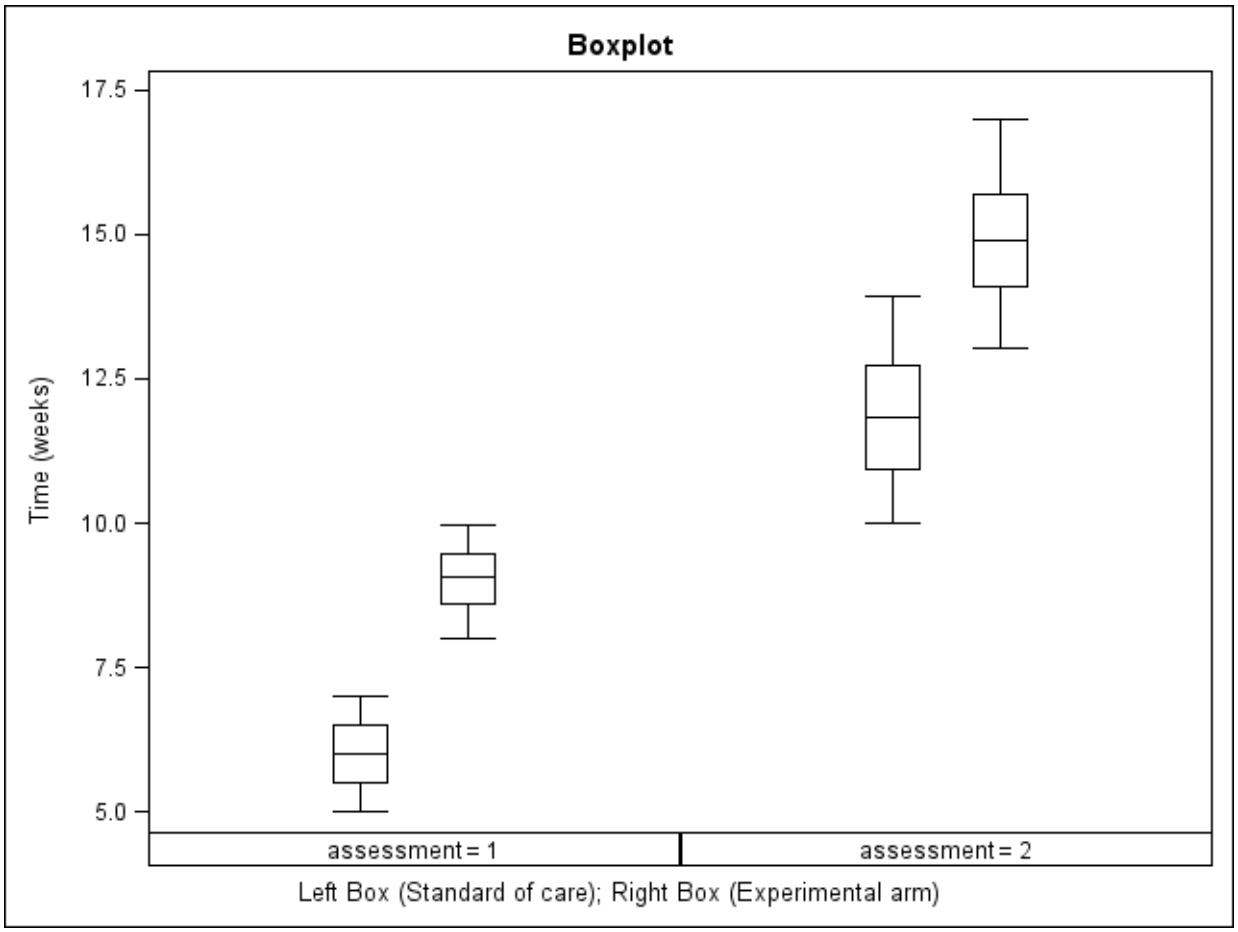
Antonio Nieto Archilla
Clinical Development. PharmaMar S.A.
Avda. de los Reyes, 1
Polígono Industrial La Mina
28770 Colmenar Viejo. Madrid (SPAIN)
anieto@pharmamar.com

José Luis Jiménez-Moro
Clinical Development. PharmaMar S.A.
Avda. de los Reyes, 1
Polígono Industrial de la Mina
28770 Colmenar Viejo. Madrid (SPAIN)
jljimenez@pharmamar.com

APPENDIX I

Fictitious example where assessments are done every 6 weeks in control arm and every 9 weeks in the experimental arm.

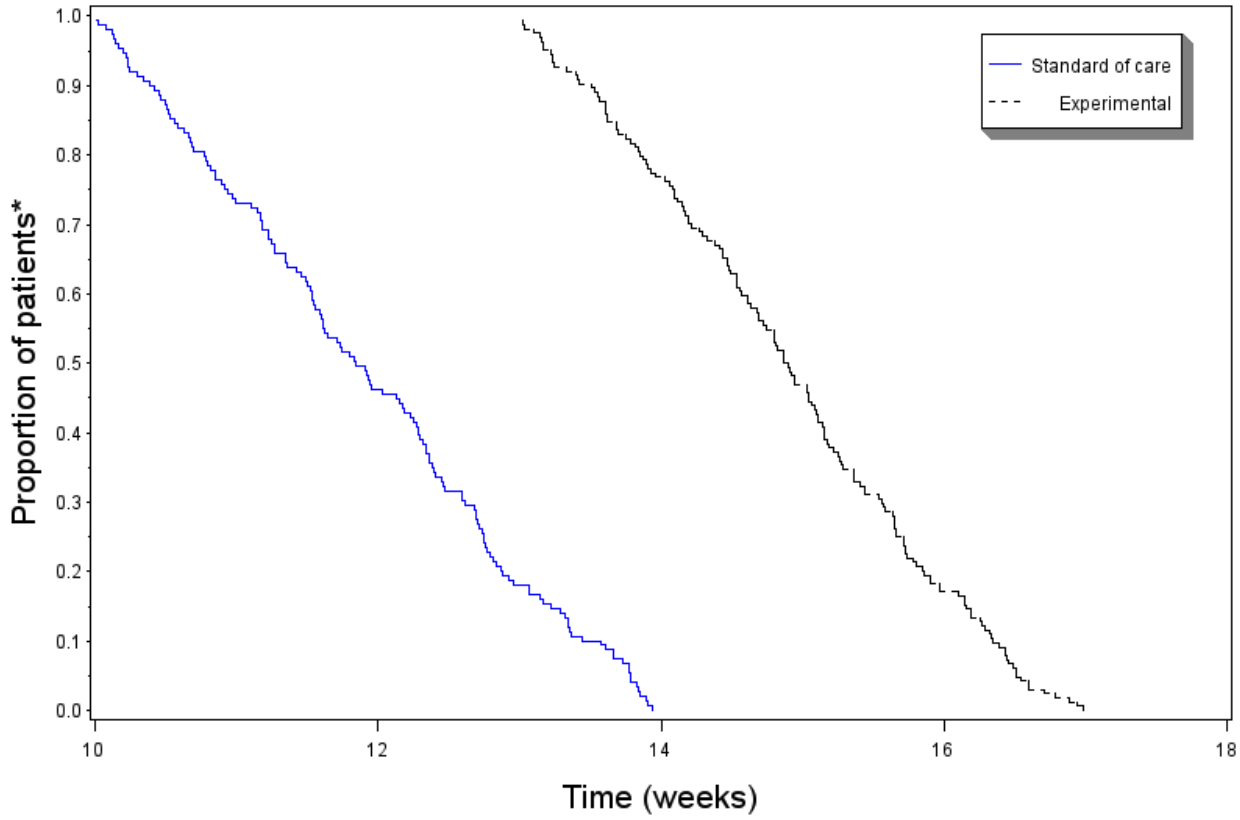




Time to disease assessments

Disease assessment	Treatment arm	n	Median (weeks)	p-value
1	Standard of care	200	6.0	<0.0001
	Experimental	200	9.1	
2	Standard of care	149	11.8	<0.0001
	Experimental	164	14.9	

Time to second disease assessment



* Proportion of patients for whom the time from randomization to the disease assessment is greater that a given time t (in weeks)