

THE IMPACT OF DATA VISUALIZATION IN A STUDY OF CHRONIC DISEASE

South Central SAS Users Group

SAS Educational Forum 2007

Austin, TX

Gabe Cano, Altarum Institute
Brad Smith, Altarum Institute
Paul Cuddihy, GE Global Research

ABSTRACT

A recent of study of chronic disease patients was sponsored by the University of Texas Health Science Center at San Antonio. Participants in the study wore a wrist monitor for 2 weeks. Information gathered by these devices included time and a patient's daily activity per minute.

Before analysis began various hypotheses were established regarding patients in the study. Researchers planned to summarize the activity data and analyze possible connections to health care utilization, hypothesizing that certain patterns of activity could be identified which would predict future hospitalization. In a real world setting, clinicians with access to activity data would then have the ability to foresee/provide early treatment and possibly prevent an expensive hospitalization.

Research into these hypotheses began by taking the large amounts of activity data downloaded from the wrist monitors and converting them into a SAS data warehouse. Various steps were taken to clean and define valid data. Further steps were taken to establish criteria for computing analysis measures.

This poster presentation will outline the steps taken in this pilot research study to organize the data for analysis, data visualization and the eventual impact of visualization.

Keywords

SAS/GRAPH, GPLOT, ANNOTATE

BACKGROUND AND MOTIVATION FOR THE STUDY

Recently, there has been a vested interest in the study of chronic diseases. The rise in the cost of care has led to a push to better manage the nature of a patient's chronic disease. One way this can be done is to better predict patient hospitalizations and thereby avoid preventable hospitalizations. New technologies have been brought into the mainstream of such studies allowing for the continuous measure of a patient's level of activity. Among these strategies are a variety of medical devices have been employed to this end, including wrist-based monitors.

In 2000 The University of Texas Health Science Center at San Antonio sponsored a study of chronic disease to track patients with various types of dispositions. As part of a clinical trial of educational and clinical intervention for chronic disease, researchers asked more than 300 trial participants to wear wristwatch-type activity monitors for as many successive 2 week periods as possible over the 18 months in which they participated in the trial. Information gathered by these devices included time of day and daily activity per minute. With some patients completing more than a dozen such cycles, the resulting dataset totaled more than 2GB for all patients. Data were then downloaded from the watches into ASCII files.

Before analysis began various hypotheses were established regarding the patients in the study. Researchers planned to summarize the activity data and analyze possible connections to health care utilization, hypothesizing that certain patterns of activity could be identified that would predict future hospitalizations. In a real world setting, clinicians with access to activity data would then have the ability to foresee/provide early treatment and possibly prevent an expensive hospitalization.

MOVING FROM ASCII DATA TO MEASURE READY-DATA: DATA CLEANING AND VALIDATION

Much work was done to agree on our definition of a “valid dataset”. This included the following: detecting patient compliance with wearing the monitoring device, valid watch data, choosing a period where data was uniformly available and measurable for all patients, and others.

It took many iterations of computing summary measures to recognize problems with the data and the general methodology. This is a general flow of the computing cycle:

- ◆ Move ASCII data to a SAS dataset
- ◆ Clean/Edit dataset
- ◆ Validate the cleaned/edited data
- ◆ Compute summary measures
- ◆ Render visualization of measures

Most major problems encountered during the study originated with the ASCII data. Such problems included the following: missing data, data format inconsistencies, faulty devices, and others. These problems were not detected until the computation or visualization steps of processing. This then caused processing to begin again with new preprocessing utilities to correct these problems. Clearly, this was a tedious and time consuming process.

DOWNLOADING ACTIVITY DATA

In this study, there were just more than 300 patients who participated in wearing wrist bands. Each patient could have as many as 12 data sets for a two week period. In total there were approximately 3000 data files from wristwatch activity.

Downloading the data began with reading in the raw activity data. Data were organized as shown in Figure 1.

Figure 1. File for Patient 001

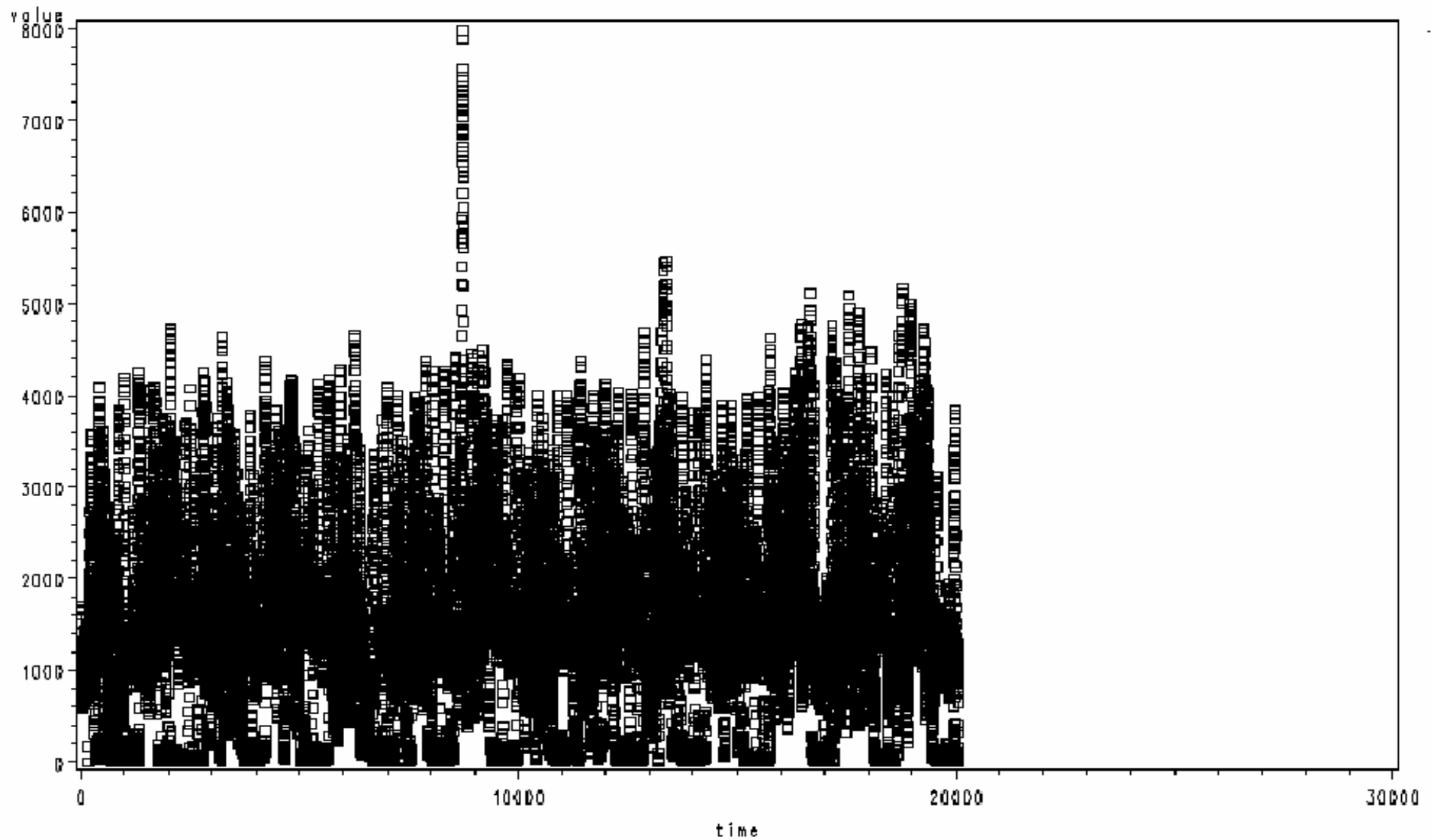
Date	Time	Activity_1	Activity_2	Activity_3
7/15/2000	9:00:00	3166	174	156
7/15/2000	9:01:00	2978	254	240
7/15/2000	9:02:00	3931	258	248
7/15/2000	9:03:00	2596	226	228
7/15/2000	9:04:00	1250	214	178
7/15/2000	9:05:00	2984	200	262
7/15/2000	9:06:00	2711	228	254

Figure 1 represents a sample of wristwatch activity data for a single patient for a single round in the study. In general, this activity data exists for a two week period. This means that each set of raw data would contain more than 20,000 rows of data. As shown, date and time are followed by three activity measures in comma separated files read via an Excel spreadsheet. Downloading the data was not a trivial task.

STEPS TO VISUALIZATION – COMPUTING MOVING AVERAGES

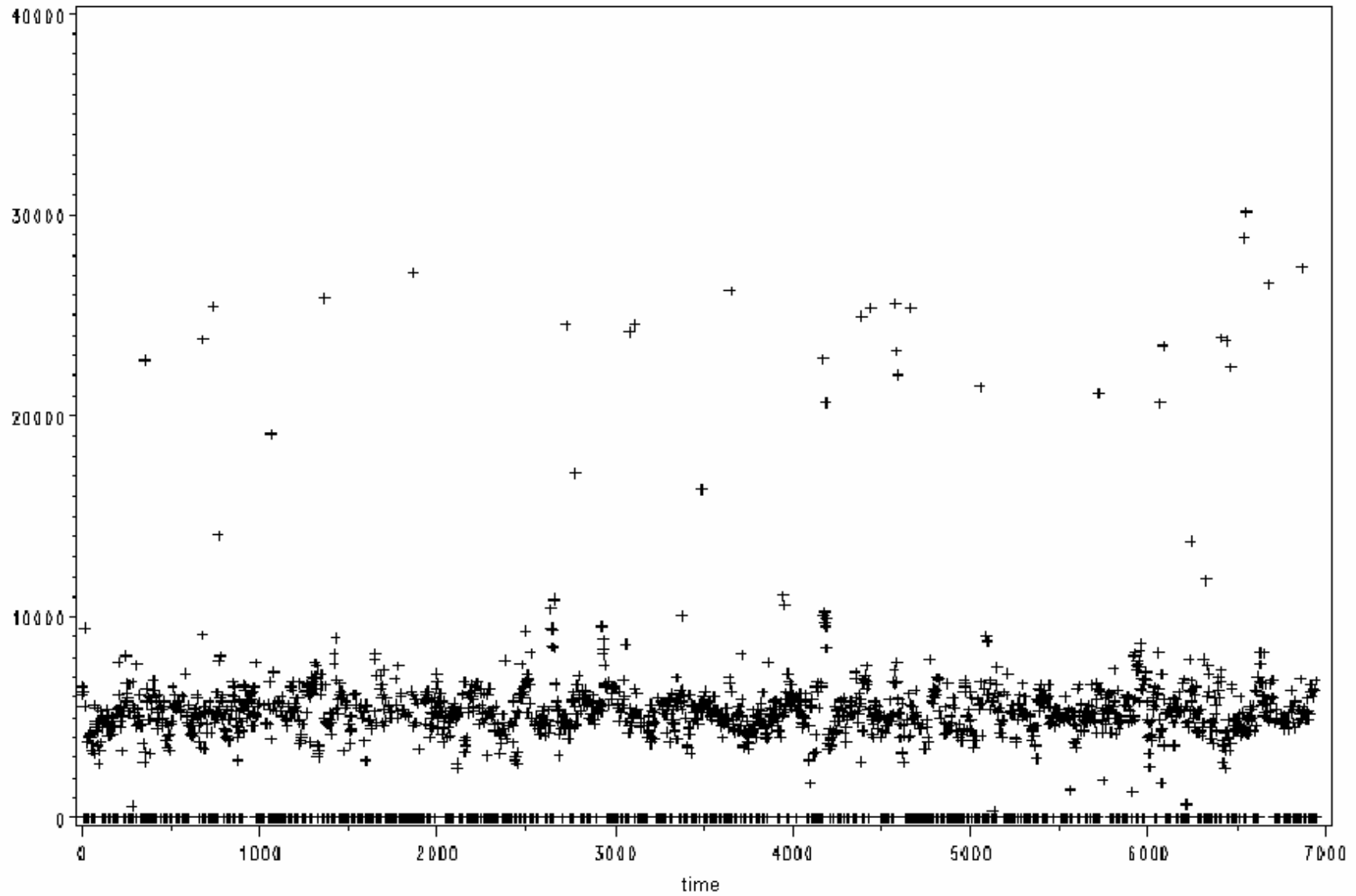
NYHA class=4: PIM - 15 minute moving averages

Diastolic type of heart failure - 1 Patients



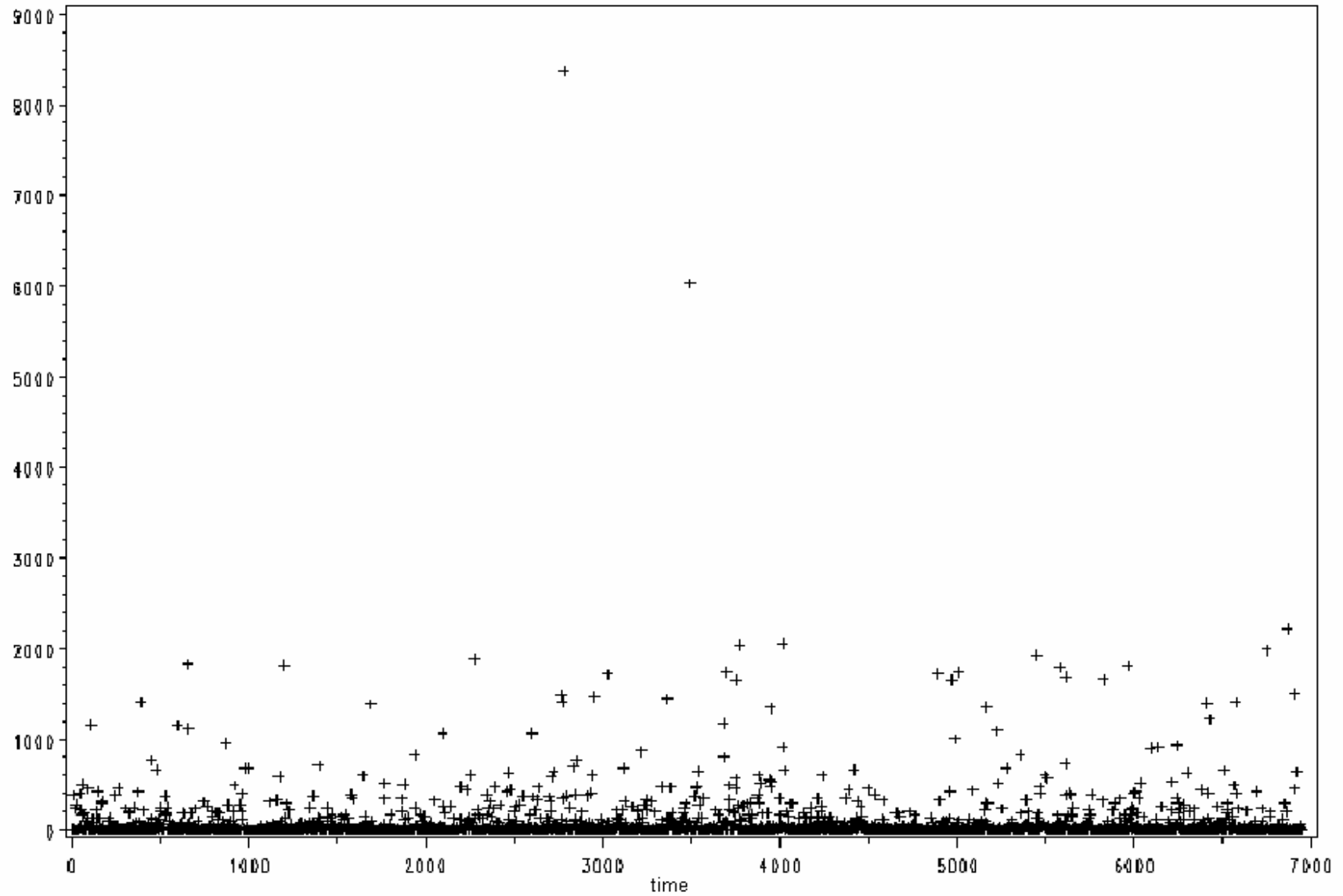
STEPS TO VISUALIZATION – COMPUTING A HIGH ACTIVITY MEASURE

Daytime Activity - PIM



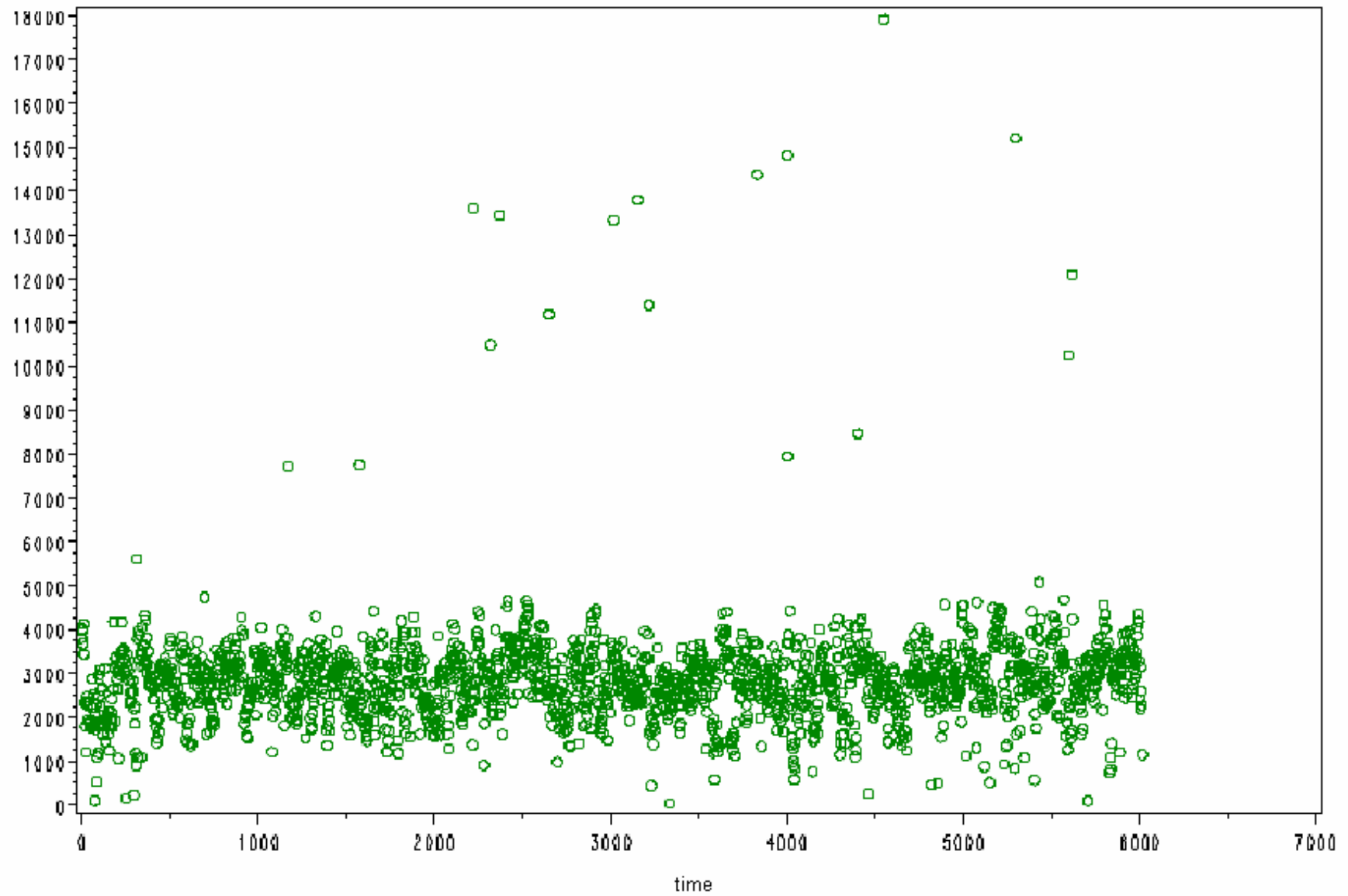
STEPS TO VISUALIZATION – COMPUTING A LOW ACTIVITY MEASURE

Nighttime Activity - PIM



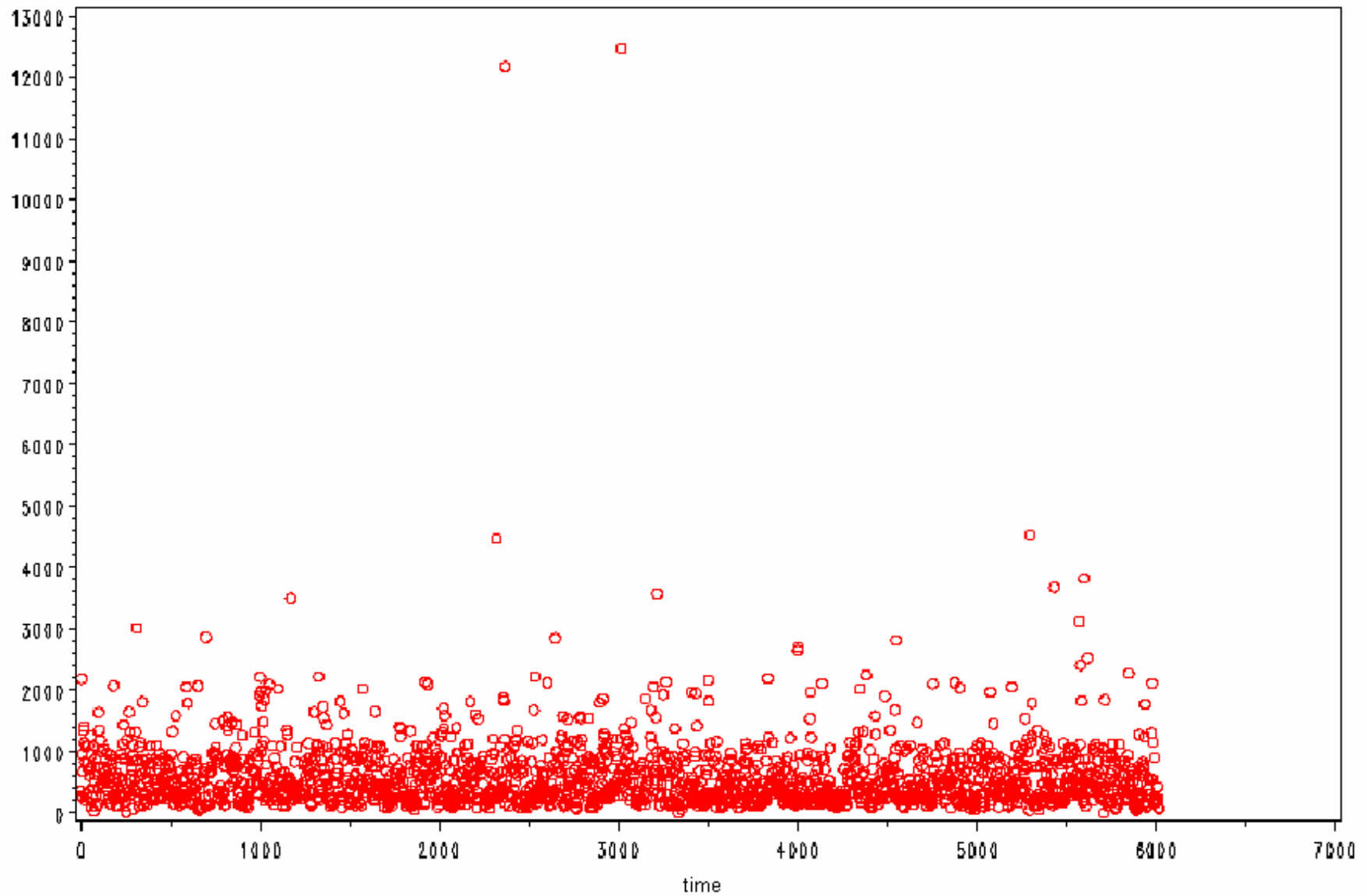
REFINING VISUALIZATION OF A HIGH ACTIVITY MEASURE

PIM: Daytime Activity



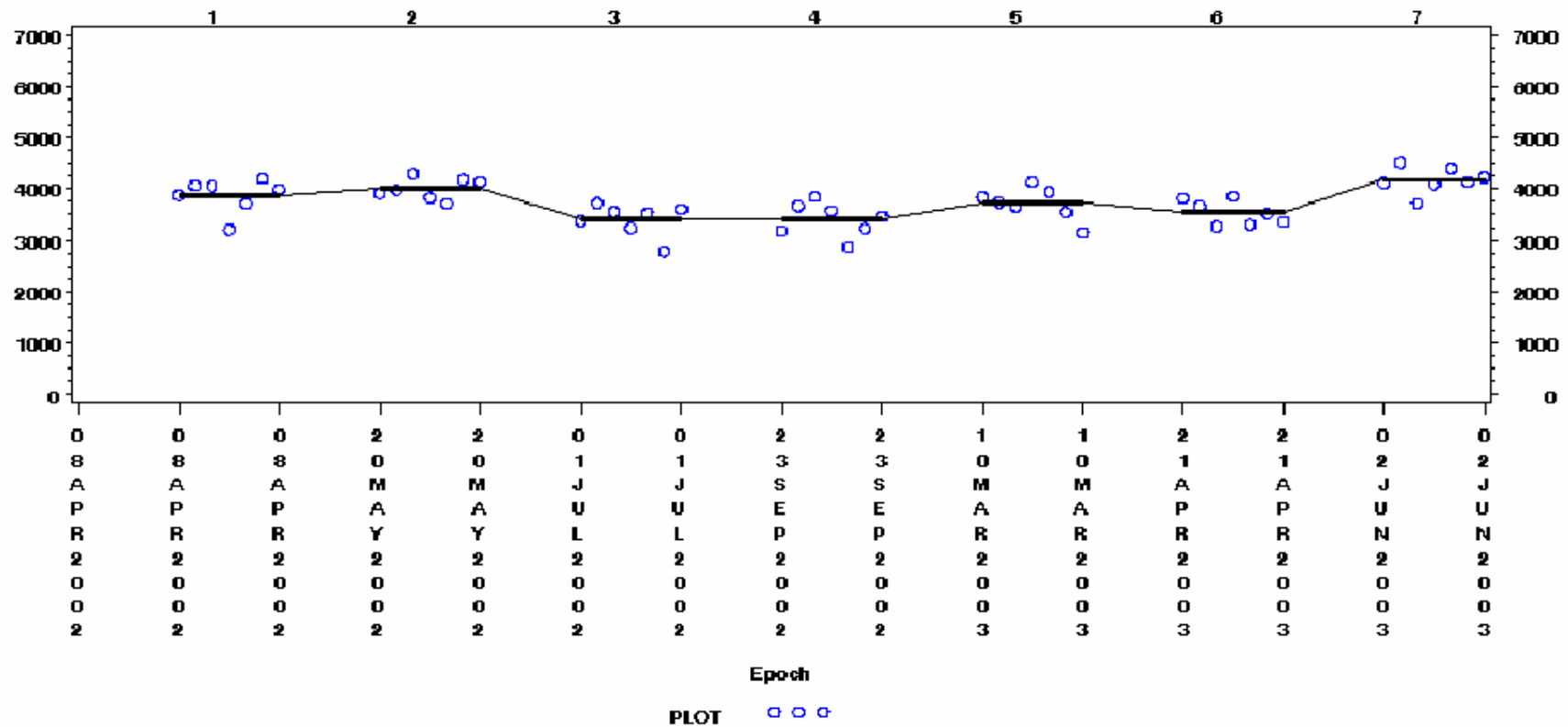
REFINING VISUALIZATION OF A LOW ACTIVITY MEASURE

PIM: Nighttime activity



VISUALIZATION – A COMPLETE HIGH ACTIVITY EPOCH

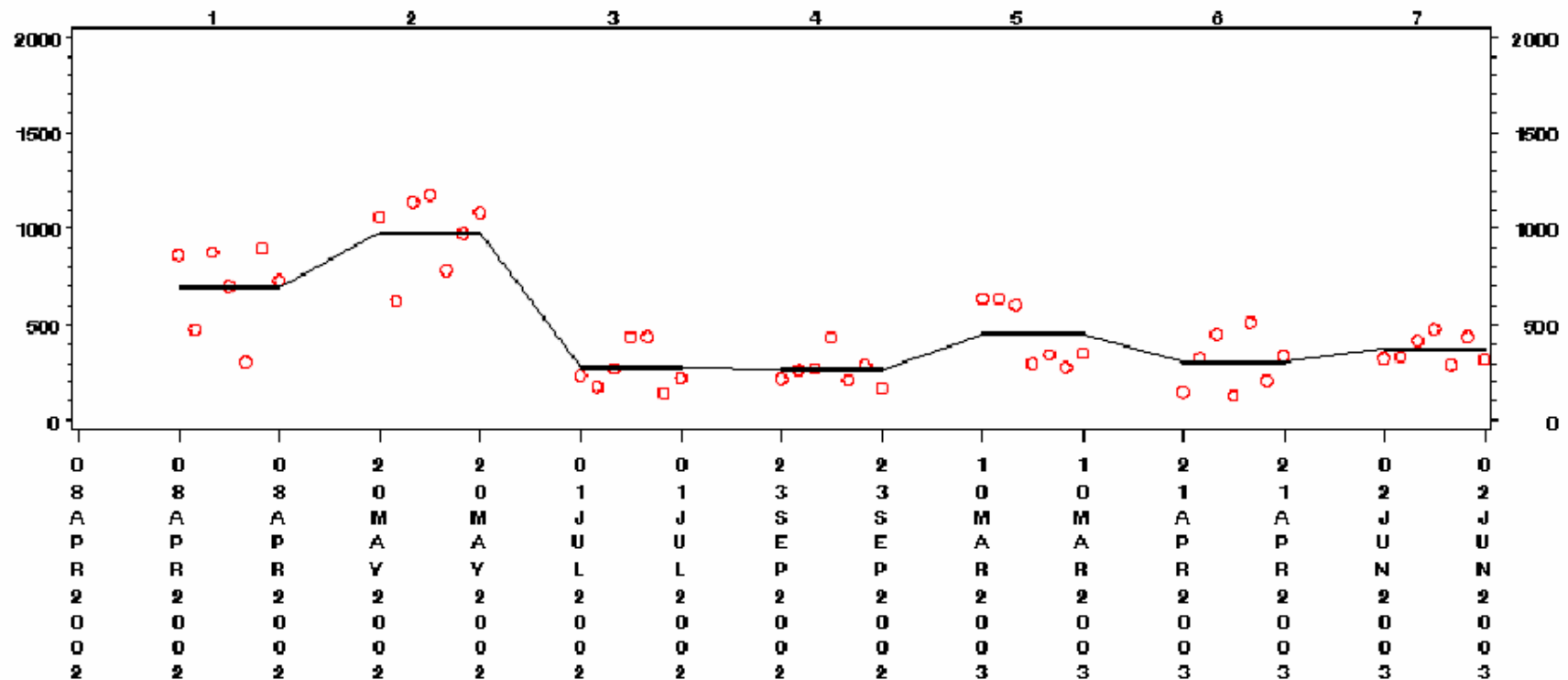
PATID: 4B006EB79A. First date in study: 4/8/2002. Activity component: daytime activity
 NYHA class = 1. Systolic type heart failure



Epoch 1 date 4/8/2002 Activity began on 4/8/2002 : Horizontal line signifies the average of the 7 day := 3520.01
 Epoch 2 date 5/20/2002 Activity began on 5/20/2002 : Horizontal line signifies the average of the 7 day := 3833.42
 Epoch 3 date 7/1/2002 Activity began on 7/1/2002 : Horizontal line signifies the average of the 7 day := 3238.86
 Epoch 4 date 9/23/2002 Activity began on 9/23/2002 : Horizontal line signifies the average of the 7 day := 3089.22
 Epoch 5 date 3/10/2003 Activity began on 3/10/2003 : Horizontal line signifies the average of the 7 day := 3398.11
 Epoch 6 date 4/21/2003 Activity began on 4/21/2003 : Horizontal line signifies the average of the 7 day := 3271.48
 Epoch 7 date 6/2/2003 Activity began on 6/2/2003 : Horizontal line signifies the average of the 7 day := 3807.48

VISUALIZATION – A COMPLETE LOW ACTIVITY EPOCH

PATID: 4B005EB79A. First Date in study = 4/8/2002. Activity component: PIM - nighttime activity
 NYHA class = 1. Systolic type heart failure



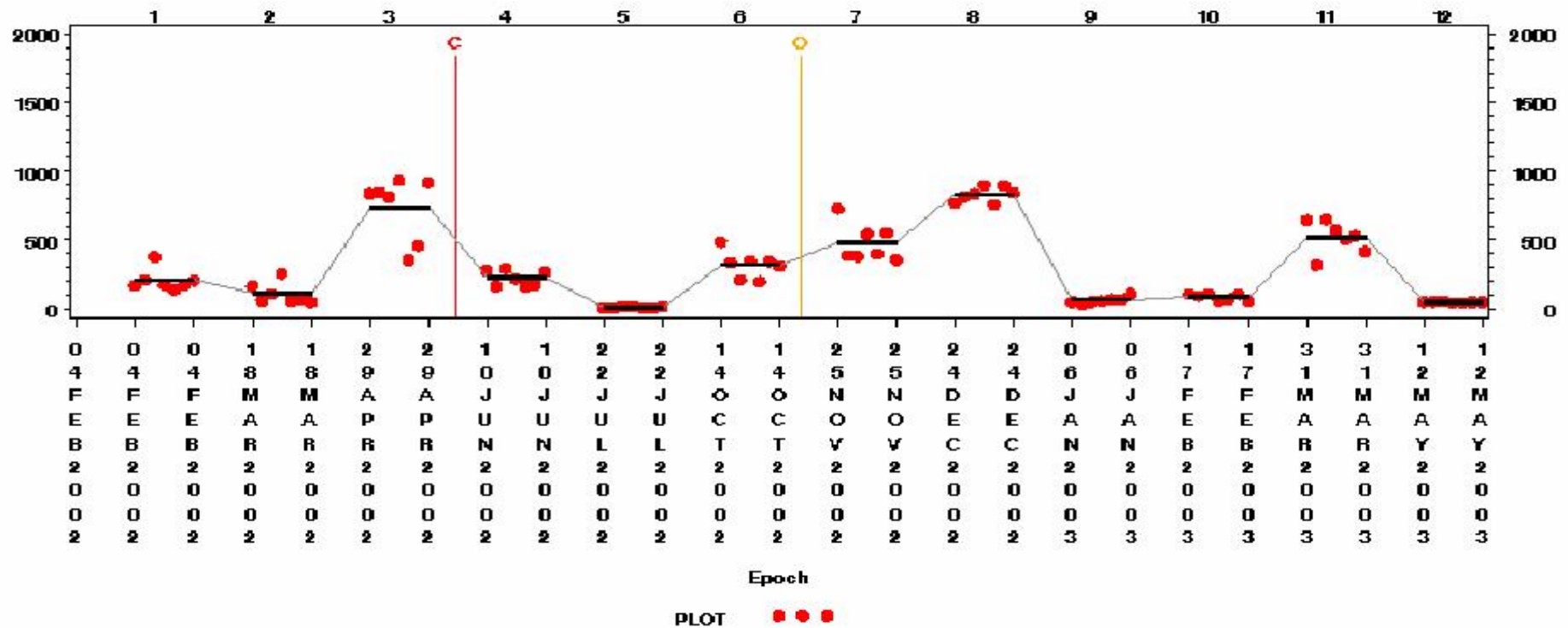
Epoch

PLOT □ □ □

Epoch 1 date 4/8/2002 Activity began on 4/8/2002 : Horizontal line signifies the average of the 7 day := 1053.04
 Epoch 2 date 5/20/2002 Activity began on 5/20/2002 : Horizontal line signifies the average of the 7 day := 1282.67
 Epoch 3 date 7/1/2002 Activity began on 7/1/2002 : Horizontal line signifies the average of the 7 day := 493.22
 Epoch 4 date 9/23/2002 Activity began on 9/23/2002 : Horizontal line signifies the average of the 7 day := 525.29
 Epoch 5 date 3/10/2003 Activity began on 3/10/2003 : Horizontal line signifies the average of the 7 day := 585.58
 Epoch 6 date 4/21/2003 Activity began on 4/21/2003 : Horizontal line signifies the average of the 7 day := 576.63
 Epoch 7 date 6/2/2003 Activity began on 6/2/2003 : Horizontal line signifies the average of the 7 day := 555.35

VISUALIZATION IMPACT – SYSTOLIC TYPE HEART FAILURE WITH EVENTS AND SPIKE

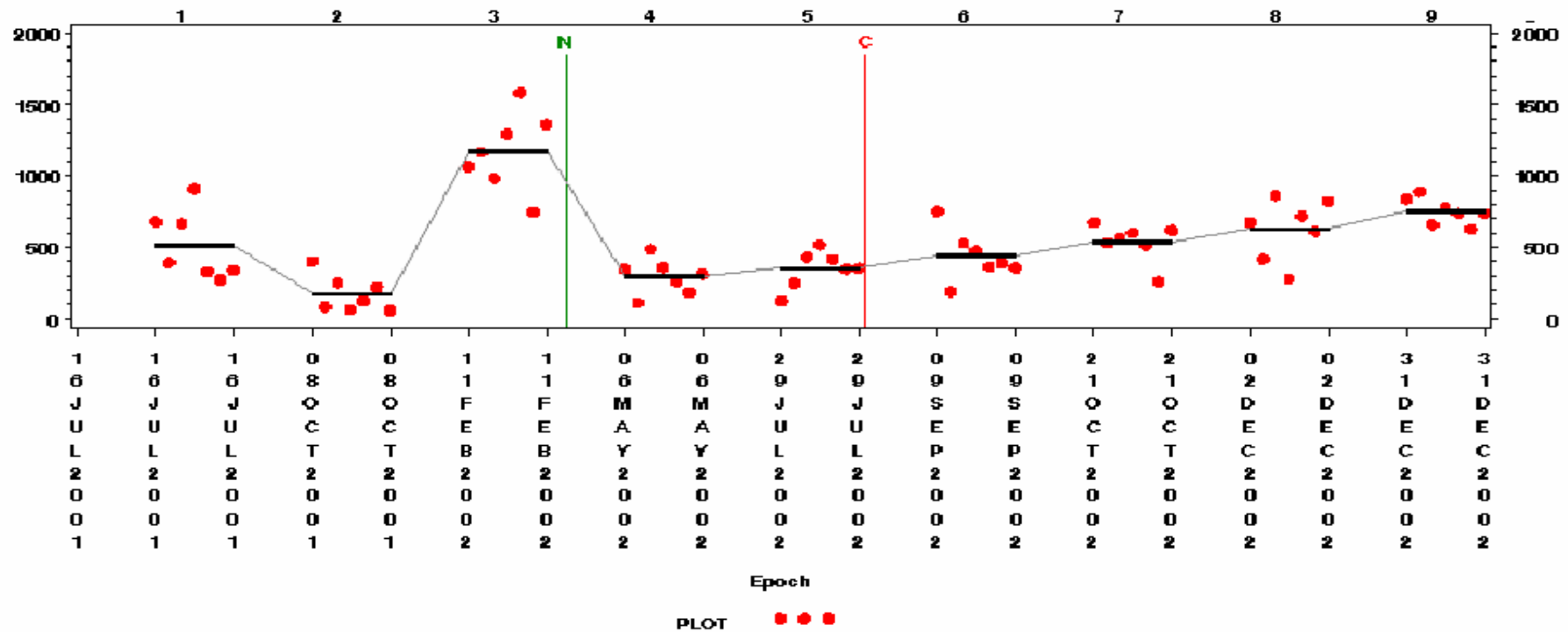
PATID: 12006BF525. First Date in study: 2/4/2002. Activity component: PIM - nighttime activity
 NYHA class= 2. Systolic type heart failure



Epoch 1 date 2/4/2002 — Activity began on 2/4/2002 :: Horizontal line signifies the average of the 7 day := 308.94
 Epoch 2 date 3/18/2002 — Activity began on 3/18/2002 :: Horizontal line signifies the average of the 7 day := 250.59
 Epoch 3 date 4/29/2002 — Activity began on 4/29/2002 :: Horizontal line signifies the average of the 7 day := 1097.9
 Epoch 4 date 6/10/2002 — Activity began on 6/10/2002 :: Horizontal line signifies the average of the 7 day := 304.53
 Epoch 5 date 7/22/2002 — Activity began on 7/22/2002 :: Horizontal line signifies the average of the 7 day := 29.37
 Epoch 6 date 10/14/2002 — Activity began on 10/14/2002 :: Horizontal line signifies the average of the 7 day := 518.93
 Epoch 7 date 11/25/2002 — Activity began on 11/25/2002 :: Horizontal line signifies the average of the 7 day := 626.27
 Epoch 8 date 12/24/2002 — Activity began on 12/24/2002 :: Horizontal line signifies the average of the 7 day := 885.17
 Epoch 9 date 1/6/2003 — Activity began on 1/6/2003 :: Horizontal line signifies the average of the 7 day := 109.41
 Epoch 10 date 2/17/2003 — Activity began on 2/17/2003 :: Horizontal line signifies the average of the 7 day := 280.85

VISUALIZATION IMPACT – SYSTOLIC TYPE HEART FAILURE WITH EVENTS AND NO SPIKE

PATID: 51004A4D40 - First Date in study: 7/16/2002. Activity component: PIM - nighttime activity
NYHA class = 3. Systolic type heart failure



Epoch 1 date 7/16/2001 – Activity began on 7/16/2001 :: Horizontal line signifies the average of the 7 day := 836.49
 Epoch 2 date 10/8/2001 – Activity began on 10/8/2001 :: Horizontal line signifies the average of the 7 day := 618.06
 Epoch 3 date 2/11/2002 – Activity began on 2/11/2002 :: Horizontal line signifies the average of the 7 day := 1642.27
 Epoch 4 date 5/6/2002 – Activity began on 5/6/2002 :: Horizontal line signifies the average of the 7 day := 891.96
 Epoch 5 date 7/29/2002 – Activity began on 7/29/2002 :: Horizontal line signifies the average of the 7 day := 804.69
 Epoch 6 date 9/9/2002 – Activity began on 9/9/2002 :: Horizontal line signifies the average of the 7 day := 928.41
 Epoch 7 date 10/21/2002 – Activity began on 10/21/2002 :: Horizontal line signifies the average of the 7 day := 1014.42
 Epoch 8 date 12/2/2002 – Activity began on 12/2/2002 :: Horizontal line signifies the average of the 7 day := 1097.38
 Epoch 9 date 12/31/2002 – Activity began on 12/31/2002 :: Horizontal line signifies the average of the 7 day := 1126.98

A GENERALIZED TEMPLATE OF THE MACRO USED FOR VISUALIZATION

%MACRO	<pre>%macro display_patient_epochs(IN_VAR , IN_PATID , IN_FILE , PLOT_FILE); /***** /* Create GRAPH_FILE from IN_FILE of patient study information. */ /* Number of times in the study, /* time points, /* disease classifications, /* patient events in the study, if any /* DAYTIME & NIGHTTIME activity measures, MAXs & MINS /* Summary measures for each time in the study *****/</pre>
ODS	<pre>/* Open ODS conduits */ ods listing close; ods html path="&path_data_plot.\&in_patid.\&in_var." gpath="&path_data_plot.\&in_patid.\&in_var." file="&plot_file." ; goptions ftext=swissb rotate=landscape;</pre>
ANNOTATE	<pre>/* create annotate file */ data anno_activity; length color function \$8 text \$25; retain xsys '2' ysys '2' when 'a' ; %do z = 1 %to &total_time_in_study.; function='move'; x=%sysfunc(round(&&first_act_t_&z..,1)); y=%sysfunc(round(&&daily_avg_activity_&z..,1)); output; function='draw'; x=%sysfunc(round(&&last_act_t_&z..,1)); y=%sysfunc(round(&&daily_avg_activity_&z..,1)); color='black'; size=2; output; function='label'; x=%sysfunc(round(%eval(&&last_act_t_&z.. - %eval(4*1440)),1)); y=2000; size=1; position='2'; color='black'; text="&z."; output; %end; run;</pre>
AXIS	<pre>/* X axis definitions */ /* create X-axis definitions and references including patient events as reference lines, if they exist */ axis1 major=none minor=none value=none %if &ref_count > 0 %then %do; rellabel=(h=1 &ref.) %end; ; /* Y axis definitions */ axis4 order=(0 to 2000 by 500) ; symbol1 v=dot color=red;</pre>

A GENERALIZED TEMPLATE OF THE MACRO USED FOR VISUALIZATION - CONT'D

GPLOT	PLOT	<pre> /* plot patient's study data */ proc gplot data=GRAPH_FILE; format Epoch datetime9.; plot ACTIVITY*Epoch / overlay legend vaxis=axis4 anno=anno_activity haxis=axis1 %if &ref_count > 0 %then %do; href=&ref_date. chref=(&href_colors.) %end; ; plot2 ACTIVITY_*Epoch / vaxis=axis4; </pre>
	TITLES	<pre> /* set up titles */ %let month=%sysfunc(month(%sysfunc(datepart(&date_1.)))); %let day =%sysfunc(day(%sysfunc(datepart(&date_1.)))); %let year =%sysfunc(year(%sysfunc(datepart(&date_1.)))); title h=1 "PATID: &in_patid. - First Date in study: &month./&day./&year. - Activity component: %upcase(&in_var.) - Nighttime activity"; %if "&in_thf."="D" %then %do; title2 h=1 "NYHA class=%trim(%left(&in_class.)) - Diastolic type heart failure"; %end; %else %if "&in_thf."="S" %then %do; title2 h=1 "NYHA class=%trim(%left(&in_class.)) - Systolic type heart failure"; %end; %else %do; title2 h=1 "No heart failure information available"; %end; </pre>
	FOOTNOTES	<pre> /* create footnotes */ %do z = 1 %to &total_time_in_study.; %let month1=%sysfunc(month(%sysfunc(datepart(&&date_&z.)))); %let day1 =%sysfunc(day(%sysfunc(datepart(&&date_&z.)))); %let year1 =%sysfunc(year(%sysfunc(datepart(&&date_&z.)))); %let month2=%sysfunc(month(%sysfunc(datepart(&&first_act_t_&z.)))); %let day2 =%sysfunc(day(%sysfunc(datepart(&&first_act_t_&z.)))); %let year2 =%sysfunc(year(%sysfunc(datepart(&&first_act_t_&z.)))); footnote&z. j=1 "Epoch &z. date &month1/&day1/&year1 - Activity began on &month2/&day2/&year2 :: Horizontal line signifies the average of the 7 day := %sysfunc(round(&&activity_&z. ,.01))"; %end; run; quit; </pre>
%MEND	CLEAN UP	<pre> /* clear out footnotes, titles */ %do z = 1 %to &total_time_in_study.; footnote&z. '00'x; %end; title ' '; title2 ' '; /* Close ODS conduits */ ods graphics off; ods html close; %mend display_patient_epochs; </pre>

ACKNOWLEDGEMENTS

Special thanks to The University of Texas Health Science Center at San Antonio for sponsoring this study.

Special thanks to General Electric for funding the research for this contract.

This study was supported by a grant from the US Department of Defense, US Army Medical Research Acquisition Activity.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the U.S. and other countries.

CONTACT INFORMATION

Gabe Cano

Altarum Institute
3737 Broadway, Suite 205
San Antonio, TX 78209-6547
gabe.cano@altarum.org
Phone: (210) 832-3000

Brad Smith

Altarum Institute
3737 Broadway, Suite 205
San Antonio, TX 78209-6547
brad.smith@altarum.org
Phone: (210) 832-3000

Paul Cuddihy

GE Global Research
1 Research Circle, K1-3A24A
Niskayuna, NY 12309
cuddihy@research.ge.com
Phone: (518) 387-734