Paper 71-2022

Using SAS® Gplot Overlay to Effectively Visualize and Compare COVID-19-Sepsis versus Non-COVID Sepsis Post-hospital Discharge Locations Over Time

Xiao Qing Wang, Brenda McGrath, Sarah Seelye, Elizabeth Viglianti, Theodore Iwashyna, Hallie Prescott
Internal Medicine, Division of Pulmonary & Critical Care Medicine, University of Michigan
VA Center for Clinical Management Research

ABSTRACT

In a world of Big Data and results-orientated clinical research, complex results can be difficult to absorb quickly and easily. This is particularity true during the COVID-19 pandemic, when vast amounts of data and statistics were made available to the public. Effective data visualization can offer quicker data insights and processing, help better communicate findings, and facilitate meaningful discussions.

We aim to illustrate the effectiveness of visualization techniques in understanding the unique outcomes of distinct populations. We do this by creating overlay plots using SAS® Gplot to compare post-hospital discharge locations among U.S. Veterans in 2020 with either A) sepsis and COVID-19+ or B) sepsis-only. The plots show clear proportional differences in post-hospital discharge mortality and readmission into acute care or nursing home facilities over time. Our paper describes the data preparation needed for creating Gplots and assessing differences in patient outcomes. The methods presented here have broad applications for diverse fields. Health care professionals may especially benefit by learning effective visualization tools to understand longitudinal clinical outcomes for different patient populations.

INTRODUCTION

The escalating amount of data in healthcare and other sectors create the need for effective visualization that can help communicate complex data quickly. As a picture is worth a thousand words, effective visualization can more effectively increase the processing time to understand complex results and comparisons compared to a mere verbal description. Statistical graphics should not only show the data but also reveal data results.

This paper demonstrates a useful technique to create overlay plots in SAS®. The overlay option from SAS® Gplot's Plot statement can put multiple plots on the same graph with options to scale and specify the number of minor tick marks on the axes. Our team uses these plots to illustrate and compare post-hospital discharge locations for patients hospitalized with various medial conditions, such as COVID-19 and sepsis.

CREATING THE DATASET

A dataset must consist of a cohort with a specific condition discharged alive from hospitals, for example patients hospitalized with COVID-19 or sepsis. To create the plots, first create a dataset with one row per patient and a variable for each of X days (e.g., 90 or 180) following discharge from hospitalization. Each day after discharge should be a field with values indicating location of the patient as illustrated Figure 1. In this instance, we use values 0, 1, 3, and 5 to represent dead, acute inpatient (i.e., hospitalization), non-acute inpatient (e.g. nursing facility) and home settings, respectively.

Figure 1. Example of Dataset with Post-discharge Days

PatientID	AdmitDate	DischargeDate	Dischargeplus1	Dischargeplus2	Dischargeplus3	...	DischargeplusX
111	3/4/2020	4/6/2020	5	0	0		0
222	6/5/2021	7/3/2021	5	5	1		1
333	1/2/2022	1/4/2022	3	3	3		3
444	3/17/2022	3/21/2022	3	3	5		5
555	6/3/2021	6/19/2021	5	5	5		5
666	3/9/2021	3/29/2021	5	5	5		5
Location Value Definitions							
0 = dead							
1 = acute							
3 = non-acute							
5 = home							

DATA MANAGEMENT

STEP 1

To create the plots for X number of days, a summary count of each valued location for each day is needed from the dataset shown in Figure 1. We will need a total of the four locations separately on dischargeplus1, dischargeplus2, and so on. For example, Figure 2 illustrates the total frequency of each of those locations on day X after hospital discharge. The number 7,458 represents (row 1) the total number of patients discharged alive that are in their home one day after hospital discharge. Whereas on day 2 (row 5), it changes to 7,314 patients and day 3 (row 9) with 7,183 patients with home location.

```
/*COVID-Positive only Hospitalizations, post discharge*/
proc summary data=COVID_only;
class DischargePlus1-DischargePlusX;
        ways 1;
        output out=postcounts; run;

/*format dataset and include new variables for location and day*/
data postcounts (compress=yes);
set postcounts;
        array a{*} DischargePlus1-DischargePlusX;
                do _n_=1 to dim(a);
                        if not missing(a{_n_}) then location=a{_n_};
                        if not missing(a{_n_}) then day=_n_;
                        end;
                drop DischargePlus1-DischargePlusX _TYPE; run;

/*combine all health care facilities into one category*/
data postcounts (compress=yes);
set postcounts (rename=(_FREQ_=total));
length loc $8.;
        if location in (1) then loc='acute';
        if location=3 then loc='nonacute';
        if location=0 then loc='dead';
        if location=5 then loc='home'; run;

proc sort data=postcounts; by day loc; run;

proc means data=postcounts sum;
        by day loc;
        var total;
        output out=postcounts2 sum=total2; run;

data postcounts2;
set postcounts2 (keep=day loc total2); run;
```

Figure 2. Postcounts Dataset

	total	location	day	loc
1	7458	5	1	home
2	658	3	1	nonacute
3	122	1	2	acute
4	49	0	2	dead
5	7314	5	2	home
6	631	3	2	nonacute
7	226	1	3	acute
8	97	0	3	dead
9	7183	5	3	home
10	610	3	3	nonacute
11	309	1	4	acute
12	141	0	4	dead
13	7084	5	4	home
14	582	3	4	nonacute
15	372	1	5	acute
16	167	0	5	dead
17	7014	5	5	home
18	563	3	5	nonacute

STEP 2

Then, for each location (i.e., dead, acute, non-acute, home) post-discharge, create a total number of counts for each day. This step re-arranges the total frequencies from Step 1 and create a distinct dataset for each of the four locations. Figure 3A is a dataset with the desired number of days you set for the plots post-discharge, if 180 days then it will 180 numbered rows. Figure 3B illustrates the cumulative total counts in the cohort that patients are no longer alive each day. For example, in the “dead” dataset, on day 4 (row 4) there are 141 patients, on day 5 (row 5) it increases to 167 and by day 19 (row 19) the total patients no longer alive is 423. Similar datasets are created for acute, non-acute, and home settings (i.e., datasets home, nh (non-acute=nursing home), hosp (acute=hospital)).

```
data postcounts2b;
set postcounts2;
        if loc='home' then loc1=5;
        if loc='nonacute' then loc1=3;
        if loc='acute' then loc1=1;
        if loc='dead' then loc1=0; run;

data home;
set postcounts2b;
        where loc1=5; run;

/*nh aka nursing home=non-acute*/
data nh;
set postcounts2b;
        where loc1=3; run;

/*hosp aka acute*/
data hosp;
set postcounts2b;
        where loc1=1; run;

/*recode to full X days*/
proc sql;
        create table hosp_v2 as
        select a.*, b.*
        from positive_XDAYS a
        left join hosp b on a.day=b.day; quit;
```

```
data hosp (compress=yes);
set hosp_v2;
if loc='' then loc='acute';
if total2=. then total2=0;
if loc1=. then loc1=1; run;

data dead;
set postcounts2b;
	where loc1=0; run;

/*recode to full X days*/
proc sql;
	create table dead_v2 as
	select a.*, b.*
	from positive_XDAYS a
	left join dead b on a.day=b.day; quit;

data dead (compress=yes);
set dead_v2;
if loc='' then loc='dead';
if total2=. then total2=0;
if loc1=. then loc1=0;
run;
```

Figure 3A. Positive_XDAYS Dataset

day
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Figure 3B. Example of "Dead" Dataset Total Counts for Each Day

	day	loc	total2	loc1
4	4	dead	141	0
5	5	dead	167	0
6	6	dead	184	0
7	7	dead	210	0
8	8	dead	234	0
9	9	dead	261	0
10	10	dead	274	0
11	11	dead	293	0
12	12	dead	317	0
13	13	dead	336	0
14	14	dead	359	0
15	15	dead	372	0
16	16	dead	391	0
17	17	dead	403	0
18	18	dead	412	0
19	19	dead	423	0

STEP 3

Lastly, before creating the plots, calculate the cumulative frequency for each location to proportions. The plot will stack bar graphs of the 4 locations (dead, acute, non-acute and home) onto the same graph. We want the height of the plot to represent 100% of the data. As shown in Figure 4, those dead (black bar) will be at 100%; then, the graph will be overlayed sequentially, with (1) a crimson-colored bar denoting acute care (~95% of the sample), (2) coral-colored bar for non-acute care (~90% of the sample), and (3) finally, a blue bar for home (~85% of the sample). This is repeated for day 2, day 3, until day X.

```
proc sql; /*hosp aka acute, nh aka nonacute*/
        create table graphpost as
        select a.day, a.total2 as home, b.total2 as hosp, c.total2 as nh, d.total2 as dead
        from home a
        left join hosp b on a.day=b.day
        left join nh c on a.day=c.day
        left join dead d on a.day=d.day; quit;

/*must be cumulative frequencies*/
data graphpost;
set graphpost;
   home1=100*(home/(home+hosp+nh+dead));
   nh1=100*((home+nh)/(home+hosp+nh+dead));
   hospital1=100*((home+hosp+nh)/(home+hosp+nh+dead));
   dead1=100*((home+hosp+nh+dead)/(home+hosp+nh+dead));
run;
```

Figure 4. Graphpost Dataset

	day	home	hosp	nh	dead	home1	nh1	hospital1	dead1
4	4	7084	309	582	141	87.28437654	94.455396747	98.262690981	100
5	5	7014	372	563	167	86.421882701	93.358797437	97.942336126	100
6	6	6963	426	543	184	85.793494332	92.483982257	97.732873337	100
7	7	6930	456	520	210	85.386890094	91.793987186	97.412518482	100
8	8	6919	458	505	234	85.251355347	91.473632331	97.116806309	100
9	9	6912	455	488	261	85.165105964	91.177920158	96.784130113	100
10	10	6907	466	469	274	85.103499261	90.882207984	96.623952686	100
11	11	6914	459	450	293	85.189748645	90.734351897	96.389847215	100
12	12	6933	429	437	317	85.423854115	90.808279941	96.094135042	100
13	13	6944	409	427	336	85.559388862	90.820601281	95.860029571	100
14	14	6947	391	419	359	85.596352883	90.758994579	95.576638738	100
15	15	6972	370	402	372	85.904386397	90.857565303	95.416461311	100
16	16	6978	357	390	391	85.978314441	90.78363726	95.18235584	100
17	17	6983	347	383	403	86.039921143	90.758994579	95.034499754	100
18	18	6989	342	373	412	86.113849187	90.709709216	94.923607689	100
19	19	6998	327	368	423	86.224741252	90.758994579	94.788072942	100
20	20	6985	329	367	435	86.064563825	90.586495811	94.640216856	100
21	21	6991	322	357	446	86.138491868	90.537210448	94.504682109	100
22	22	6998	316	348	454	86.224741252	90.512567767	94.406111385	100

CREATING THE PLOT

Once the "graphpost" dataset in Figure 4 is ready, the following code can be run to create the plot for the cohort. In this example, the X days is set at 180 days post-discharge. The colors can also be changed, and the tick marks may be changed to 365 days.

The two different cohorts in Figure 5A and 5B show clear difference in proportions for all locations over 180 days post hospital discharge. Patients hospitalized with COVID-19 and sepsis have greater proportions of mortality, readmission into acute or non-acute facilities than those with sepsis only. Thus, the proportion of those patients remaining at home is less in the COVID-19+ and Sepsis cohort. Both cohorts have larger proportions of readmission into acute or non-acute facilities within first month post hospital discharge and then gradually declines overtime. Overall, the big picture distinctively comparing the two different cohorts is that post hospital mortality is significantly greater in the COVID-19+ and sepsis cohort.

```
ods listing;

goptions reset=all  DEVICE=PNG300  /*nopclip nopolygonclip nopolygonfill*/
ftext="helvetica"
htext=2
hsize=5in vsize=3in
colors=("royalblue" "coral" "crimson"  black) display; /*can change colors here*/
axis1
color=black
label=("Time (months) Following Acute Hospitalization")
minor=none
order= (0 90 X)
value=(tick=1 "0" tick=2 "3" tick=3 "6" /*tick=4 "9" tick=5 "12"*/) /*I only have up to 180 days (6 months)*/;

axis2
color=black
label= (angle=90 "Percent of Post Discharge Patient Locations")
order= (0 to 100 by 20);

axis3
color=black
label= none
order= (0 to 100 by 20);

symbol1 interpol=join width=1 value=none repeat=5;
ods _ALL_ close;
ods rtf file= "\Overall_post.rtf";

proc gplot data=graphpost;
        plot (home1 nh1 hospital1 dead1)*day/overlay areas=4 haxis=axis1 vaxis=axis2;
        plot2 (home1 nh1 hospital1 dead1)*day/overlay  vaxis=axis3;
run;quit;

ODS RTF CLOSE;
```

Figure 5A. Cohort: COVID-19+ AND Sepsis

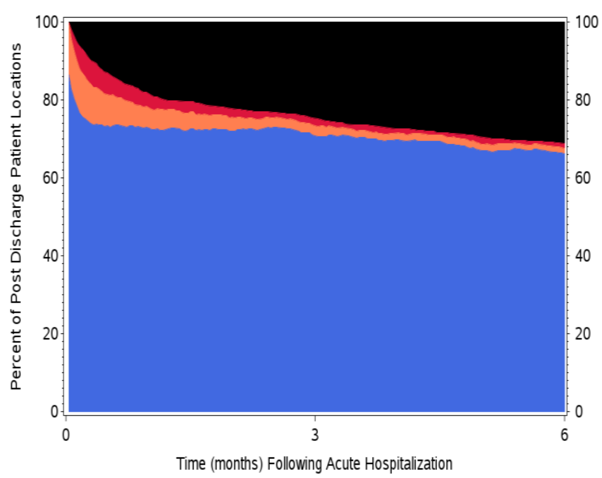
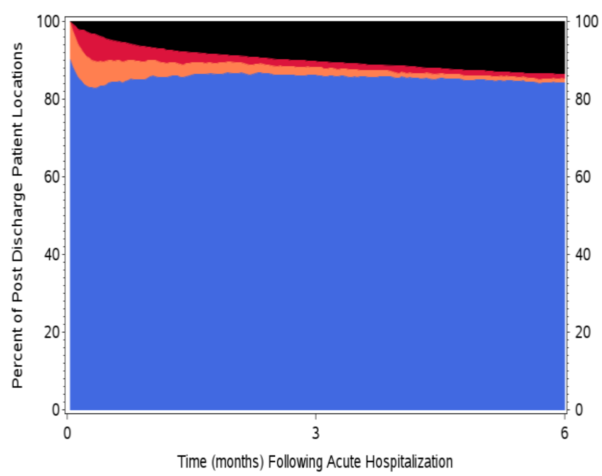


Figure 5B. Cohort: Sepsis ONLY



Color Representations

0 (Dead): Black → On Top

1 (Acute): Crimson → 2nd from top

3 (Non-acute): Coral →3rd from top

5 (Home): Blue → bottom

Figure 5A & B. United States Veterans admitted to a Nationwide Veterans Affairs hospital between March 1 – Aug 31, 2020 and discharged by Oct 31, 2020.

CONCLUSION

The data sources in various institutions are different. While I cannot help you get the data you need to create the dataset, I hope these codes and instructions will help you re-create these plots for your purposes once your dataset is built. These Overlay plots can help your audience effectively visualize post-hospital discharge locations over time. This method can also be applied to pre-hospital healthcare utilization leading up to hospitalization. With these plots, it is easily communicated and compared across groups within a cohort without much verbal description.

REFERENCES

DeMerle KM, Vincent BM, Iwashyna TJ, Prescott HC. Increased healthcare facility use in Veterans surviving sepsis hospitalization. *Journal of Critical Care*. 2017;42:59-64. doi:10.1016/j.jcrc.2017.06.026

Wang XQ, Vincent BM, Wiitala WL, Luginbill KA, Viglianti EM, Prescott HC, Iwashyna TJ. Veterans Affairs patient database (VAPD 2014-2017): building nationwide granular data for clinical discovery. *BMC Med Res Methodol*. 2019 May 8;19(1):94. doi: 10.1186/s12874-019-0740-x. PMID: 31068135; PMCID: PMC6505066.

CONTACT INFORMATION

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

Xiao Qing (Shirley) Wang, MPH
Division of Pulmonary & Critical Care Medicine, University of Michigan
xqwang@med.umich.edu; xiaoqing.wang@va.gov