

Longitudinal Data Exploration With Stacked Cumulative Percent Plots for Categorical Data

Theresa Gilligan, Cheryl Hill

RTI Health Solutions, Research Triangle Park, NC, United States

INTRODUCTION

Longitudinal clinical trial data present analysts with the unique challenge of summarizing large trends over time without losing the detail of changes from one time point to the next. Analysts must find a balance between the preservation of details at each time point and parsimonious output that can be easily summarized and interpreted.

For instruments that measure outcomes using nominal or ordinal response categories, analysts may want to preserve categorical information by reporting response category frequencies over time. In this case, the frequency of change across the course of treatment or disease progression is more informative, and perhaps more appropriate, than mean change. Collapsing responses into means draws attention to overall trends, while losing the ability to detect movement from one category response to another. Stacked cumulative percent plots offer two distinct solutions for researchers:

1. Data visualization: Allows analysts to synthesize large amounts of ordinal longitudinal data when it is necessary to retain the individual categories in addition to collapsing into means.
2. Data exploration: Allows analysts to sift through enormous amounts of data in which meaning must be found.

HYPOTHETICAL STUDY

A hypothetical research study will be used to describe the use and interpretation of stacked cumulative percent plots.

Study Population

Pediatric patients with seasonally exacerbated asthma

Study Description

- Interested in easing the worst asthma symptoms during seasonal exacerbations
- Compare two treatments and a placebo to decide which drug, if any, meets expectations to minimize the worst asthma symptoms
- Wish to minimize resource utilization (levels 3 and 4 of response scale below) so that fewer patients visit the ER or doctor throughout the year

METHODS

Three hundred study participants, randomized evenly to placebo, Treatment 1, or Treatment 2, rated their asthma symptoms daily for 365 days.

Daily Response Scale

- 1 = had little or no asthma symptoms today.
- 2 = My asthma symptoms were noticeable today, but they did not affect my usual daily activities.
- 3 = My asthma symptoms were pronounced today, and they affected my ability to perform my usual daily activities. I had to use my rescue medication or visit the doctor.
- 4 = My asthma symptoms were severe today, and I couldn't perform my usual daily activities. I had to go to the ER or be admitted to the hospital.

The Plots

The four types of plots described were developed based on the hypothetical study outcomes.

1. Individual trajectory plot: displays individual response trajectories
 2. Means plot: displays daily mean response over time
 3. Stacked cumulative percent plot (also known as categorical plot): displays daily ordinal response percentages
 4. Change plot: displays ordinal response change frequencies
- In each plot, the x-axis represents time. The months are labeled one through twelve, January through December. The y-axis represents response levels or cumulative response percentages, depending on the type of plot. For cumulative response percentage plots, response categories are indicated by colors.

The hypothetical study outcomes are presented in multiple plots per treatment (individual trajectories, means plots, categorical plots, or change plots).

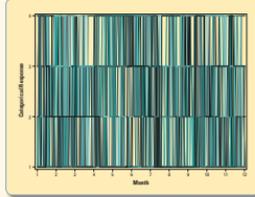
Placebo

Individual Trajectory Plot

An individual trajectory plot shows the following:

- Raw individual patient response values over time
 - Trends over time or onset of action
- In Figure 1, for simplicity, the individual trajectory plot below displays only 50 of the 100 placebo patients

Figure 1. Placebo: Individual Trajectories

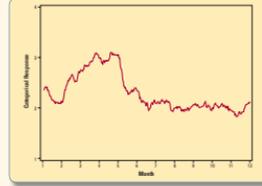


Even with only half the placebo patients, the trajectory plot is muddled and provides little useful information. Individual trajectory plots are best suited for examining small samples or highlighting a small selection from a larger population.

Means Plot

Use of means plots is a more typical and better option to display longitudinal data.

Figure 2. Placebo: Mean Daily Response



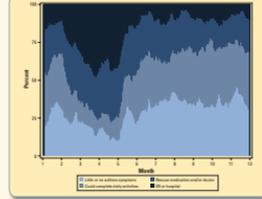
The spring months yield more serious asthma symptoms and therefore, more resource utilization in the spring months compared to other seasons.

While this plot is helpful in understanding general response trends over time, it does not show distribution of symptom levels over time. In other words, researchers cannot compare the distribution of responses, and it is impossible to tell if the spring increase is a result of less serious doctor visits and rescue medication usage (level 3) or more serious ER visits or hospitalization (level 4).

Categorical Plot

A categorical plot (or stacked cumulative percent plot) is used to compare the distribution of ordinal responses over time. In this plot, the y-axis displays the stacked percent of each symptom level (1 through 4).

Figure 3. Placebo: Ordinal Response



The seasonal influences on exacerbations are still evident in this plot. It becomes more obvious that the increase in mean values during the spring was a result of an increase in both symptoms

that resulted in doctor visits or rescue medication usage (level 3), and even more so, symptoms that resulted in ER visits or hospitalization (level 4). Also a lower number of patients reported no noticeable asthma symptoms in the spring (level 1).

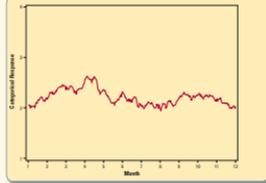
Comparison of Means Plot With Categorical Plot

The maximum value in the means plot is around 3, suggesting that the spring months are dominated by a number of level 3 responses. However, the categorical plot clearly illustrates a large portion of level 4 responses during this time. Use of both types of plots together provides a more thorough understanding of the distribution of data.

Treatment 1

Means Plot

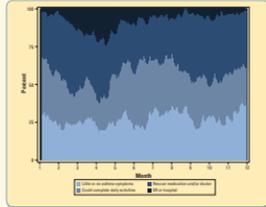
Figure 4. Treatment 1: Mean Daily Response



The seasonal trend is evident in the means plot for Treatment 1. A general decline in mean values for Treatment 1 is apparent when comparing the Treatment 1 means plot with the placebo means plot. The shift downward is promising for the treatment; however, it is not clear how the actual symptoms differ between placebo and Treatment 1 just by comparing the means plots.

Categorical Plot

Figure 5. Treatment 1: Ordinal Response

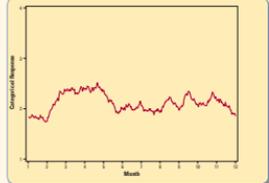


The categorical plot shows that the proportion of level 4 symptoms during the spring indicates that seasonal exacerbations do not result in as many ER visits or hospitalizations as in the placebo. Asthma symptoms that affect daily activities still occur during this time in Treatment 1, but they do not result in such extreme resource utilization. This categorical plot clarifies the decreased mean value seen in the means plot.

Treatment 2

Means Plot

Figure 6. Treatment 2: Mean Daily Response

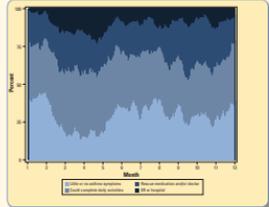


The Treatment 2 means plot also illustrates a downward shift in the mean during the spring months, indicating an increase in lower severity symptoms.

The means plots for Treatment 1 and Treatment 2 are similar, but both showing an increase in means during the spring months, although these mean values do not reach the magnitude seen in the placebo group. While the two treatments look similar, it is not clear how they differ in terms of resource utilization.

Categorical Plot

Figure 7. Treatment 2: Ordinal Response



Similar to Treatment 1 and placebo categorical plots, the seasonal components are still visible. As with Treatment 1, the seasonal symptoms are not as severe as with placebo. There are fewer level 4 symptoms compared with the placebo, although the level 4 frequencies are similar in the two treatment groups. More interestingly, level 3 responses decrease dramatically, while level 2 responses are similar to the two groups (as well as in placebo). This scenario suggests that the frequency of asthma symptoms in the spring is consistent across placebo and treatment groups, but the effects of the symptoms on daily outcomes, including daily activities and resource utilization, differ among the three groups. The placebo group shows a large amount of resource utilization (levels 3 and 4), while Treatment 1 shows a large decrease in ER visits and hospitalizations (level 4). Treatment 2 shows a large decrease in doctor visits and rescue medication usage (level 3). In exchange for the decrease in level 3 values, Treatment 2 yields more level 2 values, indicating a presence of symptoms that do not impact daily activities.

Given the ability to distinguish the proportion of each ordinal response over time with stacked cumulative percent plots, it is clear that Treatment 2 is preferable for decreasing resource utilization.

Change Plot

An additional tool for data exploration, the change plot, is a simple variation on the stacked cumulative percent plot. This plot is used to visualize the proportion of change in symptoms from time point to time across the course of a study, instead of viewing percentages over time. The table below demonstrates how the difference from day to day is calculated. (Analysts can calculate differences on any time interval that is most applicable to the project at hand.)

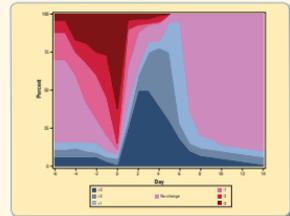
Table 1. Example Dataset

Patient ID	Day	Asthma Symptom Level	Difference in Asthma Symptom Level
1	-6	1	0
1	-4	1	0
1	-3	1	0
1	-2	1	0
1	-1	2	1
1	0	4	2
1	1	2	-2
1	2	1	-1
1	3	1	0
1	4	1	0
1	5	1	0
1	6	1	0
1	14	1	0

If no change has occurred, the difference calculated (in gray) is equal to 0. If symptoms got worse from one day to the next by one level, the difference calculation would be equal to 1 for that day. On the other hand, if symptoms declined from one day to the next from a level 4 to a level 2, the difference calculation would be -2 for that day.

Patients can make seven types of changes from time point to time point in this example $\{-3, -2, -1, 0, +1, +2, +3\}$. Therefore, there are seven colors representing each one of these change levels in Figure 8.

Figure 8. Placebo: Change in Ordinal Response



This plot is centered on asthma exacerbations taken from the placebo group. It covers 20 days surrounding a random sample of exacerbations. The plot begins 6 days prior to the exacerbation and ends 14 days postexacerbation. The exacerbations are centered at day 0 (i.e., the x-axis does not correspond to the calendar day).

Six days before the exacerbation, most patients do not experience any change in their symptoms. This symptom stability is illustrated by the large portion of lavender on day -6. As time moves forward towards the exacerbation, the proportion of red tones increases, indicating increases in symptom severity. On day 0, the exacerbation, the red tones indicate that most patients described their symptoms worsening from the days prior, with many patients indicating a change from level 1 to level 4 (the only way to score +3 on daily change). From day 1 to day 8, the majority of the plot is blue. The increase in blue indicates that patients reported a negative change in symptom severity. In other words, their symptoms were getting better. After day 8, the lavender portion of the plot increases, representing no changes from day to day for most patients and a stabilized condition.

CONCLUSION

Stacked cumulative percent plots are a helpful addition to longitudinal analysis. In this example, researchers were able to differentiate among seemingly similar treatments and identify the treatment that met their needs.

CONTACT INFORMATION

Theresa Gilligan, MS
Senior Health Outcomes Analyst
E-mail: gilligan@rti.org

Phone: 919 541 5946
Fax: 919 541 7222

RTI Health Solutions
200 Park Office Drive
Research Triangle Park,
NC 27709
Orlando, Florida, United States

ACKNOWLEDGEMENT

The authors would like to thank Lori McLeod and Carla Demuro-Marcos for their guidance with this work and Peter Haskak for his assistance with data generation.