Next Step Manuscript Figure Questions

Chart, scatter chart

Description automatically generated

Above: Scatterplots of data from combined cycle 2019-20 and 2020-21. **Panel A** - Step 1 vs Step 2. **Panel B** – Step 1 vs number of peer reviewed paper. **Panel C** – Step 1 vs number of oral presentations. **Panel D** – Step 1 vs number of poster presentations.

A few things I would like to discuss:

1. **Regression line:** Having a line in Panel A makes sense, but should we include one in the other panels?
2. **Outliers:** There are some data points in each panel that are obvious outliers. If there isn’t a legitimate reason to remove them, we shouldn’t remove them. I want to hear your thoughts on whether to include the outliers in the figures. Also, if we use the conventional definition of outliers, we would find quite a number of outliers in each of the variable dataset (see the boxplots below). If we decide to remove some outliers for whatever reason, we might need to define our own boundaries for outliers so that we don’t exclude as many of them.
3. **R and p values:** Would you like me to include the R and p values for the correlation in each individual panel?
4. **x- and y-axis labels:** In panel B, the y-axis label is too long that it overlaps with the panel label, “B.” Do you have any idea how we can rephrase and shorten the y-axis label?
5. **Axis limit:** I included axis limits below. Please review them and see if you approve.
6. **Other comments?** Any adjustment needed on other parameters, e.g. color of the correlation line, size/shape of the dots, capitalization of the x- and y-axis labels, etc.?

**Axis limits**

|  |  |  |
| --- | --- | --- |
| Variable | Data (min, max) | Axis limit |
| Step 1 | (198, 270) | (191, 284) in panel A  (193, 275) in all others |
| Step 2 | (196, 279) | (191, 284) – panel A |
| Peer reviewed abstract | (0, 35) | (0, 36) – panel B |
| Oral | (0, 25) | (0, 26) – panel C |
| Poster | (0, 46) | (0, 47) – panel D |

**Boxplots illustrating outliers**

Chart, scatter chart

Description automatically generated Graphical user interface, application

Description automatically generated with medium confidenceChart, scatter chart

Description automatically generated Chart, box and whisker chart

Description automatically generated

Chart, scatter chart

Description automatically generated Chart

Description automatically generated with medium confidence