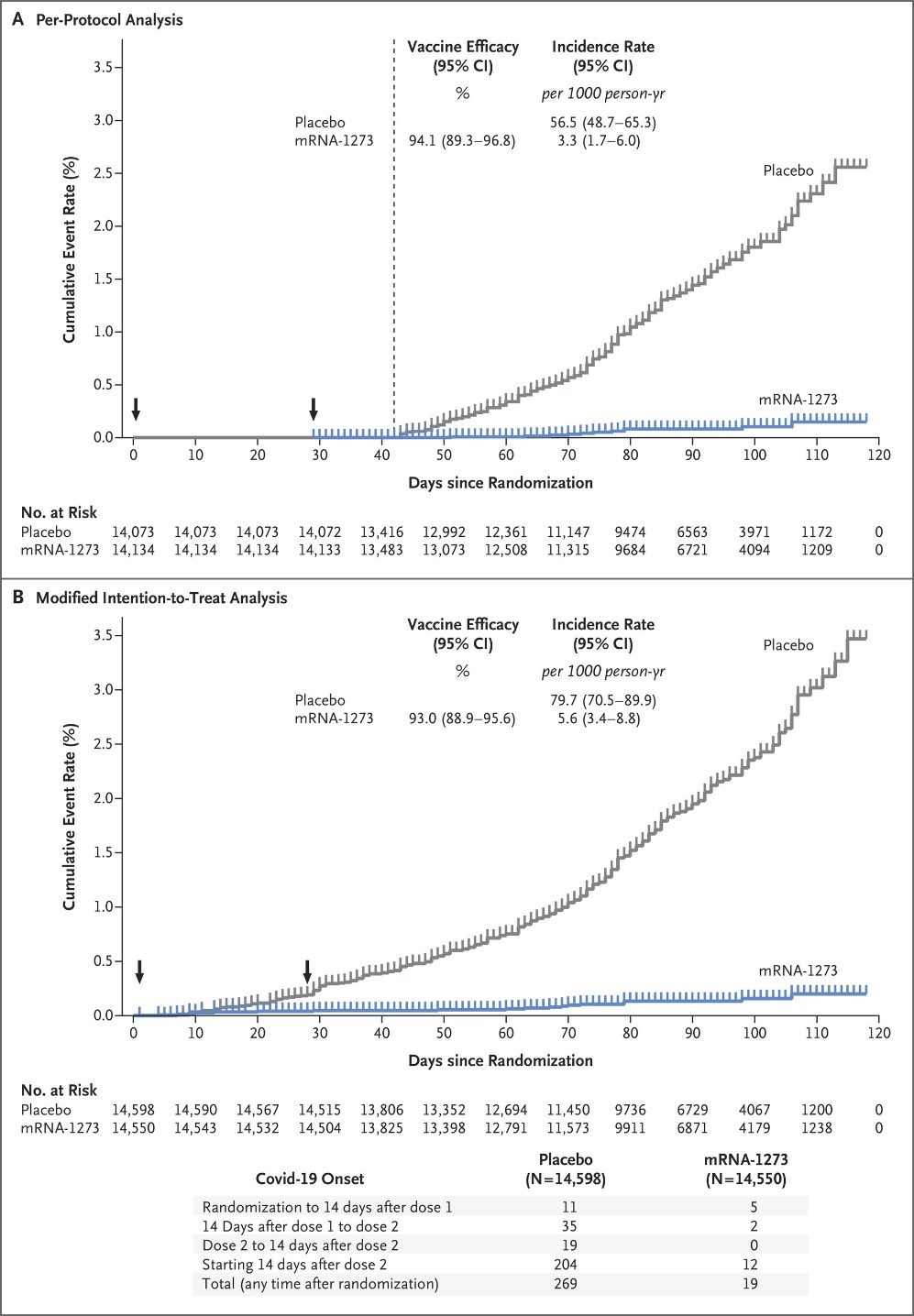
# Lawrence Zhou – Critique Data Visualisations Assignement

## Visualisation 1 – mRNA-2173 Covid-19 Vaccine

**Introduction:**

The mRNA-1273 vaccine is a COVID-19 vaccine being developed by The United States National Institute of Allergy and Infectious Diseases (NIAID), the Biomedical Advanced Research and Development Authority (BARDA), and Moderna. The purpose of the visualisation is to show quantitative data about efficacy of the vaccine and is directed towards the general public reading the New England Journal of Medicine.

**Body:**

The graph is split into two panels, showing the Cumulative event rate that someone from the Placebo and the mRNA-1273 Vaccine group contract the virus after the second dose. The first point of contention with the data visualisation is the usage of hue. They have separated between the Placebo and the mRNA-1273 Vaccine group, using a grey for the Placebo and the blue for the mRNA-1273 group. The grey blends in more with the white background than the blue, creating a distinction of what is being tested for. It is also black arrows to indicate the days in which the vaccine/placebo was taken. The colour choices are also effective in that they do not hinder anyone that is Red-Green colour-blind; the blue colour will stay a blue, whilst the grey colour will remain like its original colour as well. One way to improve the visualisation is to change the colour indicating either the Placebo or the arrows, as they are like each other; a different colour would be able to create more distinction between the two.

The graph itself is a line graph, showing the number of cases increasing each day. In the first panel, there is a dashed line indicating the 42 days after the first vaccination, when the follow-up began. The ticks above each graph also marks some censored data, as described by the New England Journal of Medicine. Meanwhile the Symptomatic Covid-18 case accrual for placebo and vaccine, showing the modifications if all asymptomatic patient from Day 1 of the Vaccination. One improvement that could be made to the graph is to remove the ticks shown. As they contain “censored data”, there is no indication of what the information may comprise of, and thus provides no use to the general viewer of the data visualisation. This would also save ink and create a higher data ink ratio, with the removal of each tick from each visualisation.

There is a small chart located above each graph, showcasing the quantitative data of the Vaccine Efficacy (as a percentage) and Incidence Rate (per 1000 people), alongside the 95% confidence interval for each of the two groups. The headers have been bolded to indicate the type of statistics they are showing, with the rate being shown underneath in italics. The dashed line from earlier can also be used to separate between the two groups and their respective statistics in the first panel. However, there is none in the second panel, as there is no indication of the 42nd Day after the first vaccination. One recommendation is to make a small, dotted line there to keep the separation between the Placebo/mRNA-2173 groups from the statistics. Another recommendation is that there is some separation between the statistics of each group, using a light grey background to do so.

The x and y-axes are in black, potentially providing too much contrast between itself and the background. Whilst the words “Cumulative Event Rate (%)” and “Days since Randomisation” do provide significance in that it provides information about what the axes mean, the axes itself does not provide as much information. The y-axis (and the percentages associated with them) and the x-axis should instead use a grey colour to be able to blend into the background, allowing the user to be able to focus on the contents of the line graph itself. Each tick on the y-axis is separated by 0.5 percent, with the maximum percentage is 3.5 percent for both graphs. This creates separation between the two graphs, ensuring that there is no confusion between the “Modified Intention-To-Treat Analysis” and “Per-Protocol Analysis” graphs. This would ensure that the statistics are not deceptively like each other. Along the x-axis, the distance between the ticks is 10 days. This relates to a common pattern of the English-speaking world, allowing for easier understanding between the days. One recommendation is to have long, but still dashed lines separating the two graphs instead of the solid line, as well at the table underneath the two-line graphs. This still creates some distinction between the “Modified Intention-To-Treat Analysis” and “Per-Protocol Analysis” graphs. The lines along the sides should also be removed entirely, if not greyed out. All of these are used to save ink, improving the data ink-ratio of the visualisation overall. Another recommendation would be to decrease the distance between each tick of the x-axis to 7 days instead, allowing the viewer to associate the distance as a reference to a singular week.

**Conclusion:**

The first visualisation does well in portraying the difference in the two graphs between the Vaccine and the Placebo groups. However, there is still improvements that could be made in terms of the table inside the graphs distinguishing between the various headers and statistics. Improvements could also be made to the data visualisation overall to improve the data-ink ratio.

## Visualisation 2 – BNT162b2 mRNA Covid-19 Vaccine

**Introduction:**

The BNT162b2 mRNA Covid-19 Vaccine is a COVID-19 vaccine being developed by Pfizer and BioNTech SE. The purpose of the visualisation is to show quantitative data about the efficacy of the vaccine and is directed towards the general public reading the New England Journal of Medicine.

## 

**Body:**

This data visualisation is a single panel, showing the Cumulative Incidence rate that someone from the Placebo and the BNT162b2 mRNA Covid-19 Vaccine group contract the virus after the first dose. This does not split between the first and the second doses of the vaccine. Instead of using grey and blue, like the first visualisation, this visualisation opted to use blue for the Placebo group and red for the Vaccine group. There is more of a distinction created between the two groups in comparison to the first data visualisation, more effectively differentiating between them. Like the first visualisation, the colour choice is also good for Red-Green colour-blind people. In this case, the blue will stay blue, whilst the red will turn into an approximate orange colour. The blue of the Placebo group still blends into the background a bit more than the red of the Vaccine group. Unlike the first visualisation, however, it does not mark the dates in which the vaccination takes place. The inclusion of this (in a black or grey colour) may be one way of improving the visualisation overall.

The graph is a line graph, showing the cumulative number of cases per day. Each symbol represents Covid-19 cases starting on a given day, with filled symbols representing severe cases. Some symbols represent more than one case, which means more overlapping cases. It is more detailed, providing more information than the first visualisation especially about the number of severe COVID-19 cases. However, it may be too ink-inefficient, in that there are lots of overlaps between the different shapes, with an over clustering of information. One recommendation for an improvement to the data visualisation is for the usage of a regular line instead of the shapes, which would increase the data-ink ratio dramatically. There may still be a shaded in circle to represent all the severe cases of COVID-19 to keep the information there. The first 21 days is split off into another section of the graph. This allows the user of the information to better understand the data within the first 21 days after the first dose. The same recommendation can be applied here, using a regular line graph without any shapes, except for a shaded in circle to mark the severe cases of COVID-19 between the Placebo and Vaccine groups.

The “Cumulative Incidence (%)” and “Days after Dose 1” is emphasised with the bold in comparison to the x and y-axis. Normally, there may be too much distinction between the black x and y-axes, contrasting the white background, like the first visualisation. The axes on the two graphs need to remain distinct against the lines that separate the full and dissected graphs in the data visualisation. A recommendation for this visualisation is to grey out the lines separating the two graphs instead of the axes, allowing more emphasis on the scales of the two x and y-axes. For the full line graph, the distance between each tick on the y-axis is 0.4%, whilst for the x-axis, it is 7 days. The 7 days scale represent a per week basis, instead of the arbitrary 10 days scale that is used in the first visualisation. On the dissected graph, the distance between each tick on the y0axis is 0.1%, and 2 days on the x-axis. This allows for a more in-depth visualisation on that section, as the cumulative percentage of cases during the first 21 days is very similar, no matter which group.

The x and y-axes are in black, potentially providing too much contrast between itself and the background. Whilst the words “Cumulative Event Rate (%)” and “Days since Randomisation” do provide significance in that it provides information about what the axes mean, the axes itself does not provide as much information. The y-axis (and the percentages associated with them) and the x-axis should instead use a grey colour to be able to blend into the background, allowing the user to be able to focus on the contents of the line graph itself. Each tick on the y-axis is separated by 0.5 percent, with the maximum percentage is 3.5 percent for both graphs. This creates separation between the two graphs, ensuring that there is no confusion between the “Modified Intention-To-Treat Analysis” and “Per-Protocol Analysis” graphs. This would ensure that the statistics are not deceptively like each other. One recommendation is to have long, but still dashed lines separating the two graphs instead of the solid line, as well at the table underneath the two-line graphs. This still creates some distinction between the “Modified Intention-To-Treat Analysis” and “Per-Protocol Analysis” graphs. The lines along the sides should also be removed entirely, if not greyed out. All of these are used to save ink, improving the data ink-ratio of the visualisation overall.

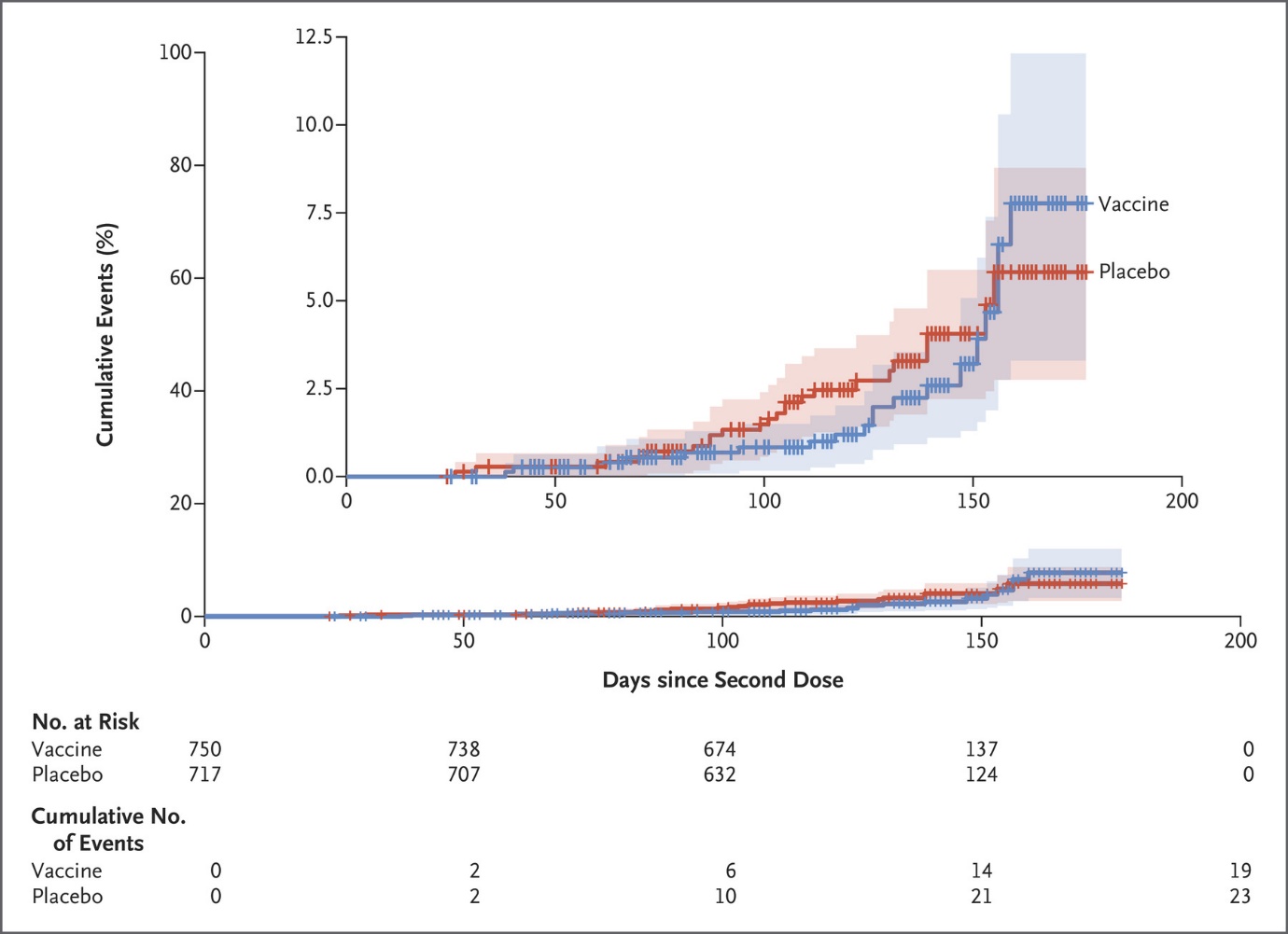
**Conclusion:**

Overall, the colour of hue does more to make itself contrast with the background in comparison to the first visualisation. However, improvements could come in the form of the separation between the axes and the lines separating the full and dissected graphs, creating more of a distinction between the two. One more improvement can come from the lines itself, having too many overlapping shapes instead of potentially using a line to show the increase in COVID-19 cases.

## Visualisation 3 – ChAdOx1 nCoV-19 Covid-19 Vaccine

**Introduction:**

The ChAdOx1 nCoV-19 Covid-19 Vaccine is a COVID-19 vaccine being developed by the University of Oxford and AstraZeneca. The purpose of the visualisation is to show quantitative data about the efficacy of the vaccine and is directed towards the general public reading the New England Journal of Medicine.



**Body:**

The third data visualisation is of a single panel, showing the Cumulative Events rate that someone from the Placebo and Vaccine group contracts the virus after the second dose. This is only showcasing data for after the second dosage of the vaccine, not analysing anything until the 2nd dosage is taken. Like the second visualisation, it is using a blue/red colour to represent the Vaccine and the Placebo groups, respectively. Like the second visualisation, there is more of a distinction created between the two groups and the white background. Same as the first two visualisations also, the colour choice is good for people who are Red-Green colour-blind. In contrary to the first visualisation, but like the second, there are no dates marked in which the vaccination takes place. A black of grey colour arrow indicating these dates may improve the visualisation overall, as it shows when the data is first starting to be tracked.

There are two-line graphs shown, one with the y-axis scaled to 100%, with the other only being scaled to 12.5%. This is done so the viewer of the information would be able to delineate both a more general idea and a more detailed idea about the Cumulative Events percentage. The tick marks indicate cate censored data from either a Symptomatic Covid-19 illness, withdrawal from the trial, or from death. However, it does not do anything to delineate between those 3, only showing red and blue lines based on whether the person was a part of the vaccine group. This may be more useful when there is a low separation between the Vaccine and Placebo group’s Cumulative Events percentage, even in the graph with the much smaller scale. One recommendation for this is to remove it completely, because it is hard to differentiate between the three reasons for the ticks and would make it harder to accurately understand the data visualisation. It may also be too ink inefficient, whilst providing only some, but not a great amount of detail about the Cumulative Events of COVID-19 cases.

Unlike the first visualisation, where the 95% confidence interval and the rate per 1000 people were shown in a table form, this visualisation shows the 95% confidence interval through the bar located on each side of each line. This allows for a better understanding of where the 95% confidence interval overlaps. There is also the usage of hue with this, showing a light red and a light blue where the Placebo and Vaccine group does not overlap. Meanwhile, there is a mix of where the two groups do overlap. For people who are red-green colour-blind, this May also work, with the lighter shade of red and shade of blue remaining relatively the same. This is whilst the darker mix of the two colours will have a darker shade to contrast the two brighter ones. One recommendation to improve the visualisation is that the per 1000 people rate is not shown on the visualisation at all. This may provide a little bit more information, whilst also improving the data density as well.

The “Cumulative Events (%)” and “Days after Second Dose” is showcased in bold, highlighting the meaning of the values on the x and y-axis. Unlike the first graph, the black colour of the axes is used to separate the two graphs. If this were not the case and were both a grey colour instead of a black, then the viewer may be inclined to ignore the difference in scale, only to assume that they are two different graphs. The graph with the expanded scale is used to allow the user to more closely differentiate the between the two graphs, as well as get more precise readings of accuracy. This is dissimilar to the second visualisation, which is only zooming into one section of the graph. One improvement that could be made is through the removal of the larger line graph entirely, as it provides little to no information about the details of the COVID-19 cases, also improving on the data-ink ratio.

**Conclusions:**

Like the second visualisation, the colour of hue contrasts with the background in comparison to the first visualisation. There could still be an improvement of the data in that the rates of COVID-19 cases could be shown (per 1000 people). Improvements could also be made to the data-ink ratio, with the removal of the larger graph and the ticks on each line of the graphs.

## References

Le, T.T., Andrewadakis, Z., Kumar, A., Roman, R.G., Tollefsen, S., Saville and Mayhew, S. (2020). *The COVID-19 vaccine development landscape*. [online] nature.com. Available at: https://www.researchgate.net/profile/Tung-Le-10/publication/340535627\_The\_COVID-19\_vaccine\_development\_landscape/links/5ead65c5a6fdcc7050a1c089/The-COVID-19-vaccine-development-landscape.pdf [Accessed 1 Apr. 2021].

Baden, L.R., El Sahly, H.M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S.A., Rouphael, N., Creech, C.B., McGettigan, J., Khetan, S., Segall, N., Solis, J., Brosz, A., Fierro, C., Schwartz, H., Neuzil, K., Corey, L. and Gilbert, P. (2020). Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine* [Accessed 9 Apr. 2021].

‌Colour Blind Awareness. (n.d.). *Types of Colour Blindness*. [online] Available at: https://www.colourblindawareness.org/colour-blindness/types-of-colour-blindness/#:~:text=People%20with%20deuteranomaly%20and%20protanomaly [Accessed 9 Apr. 2021].

Madhi, S.A., Baillie, V., Cutland, C.L., Voysey, M., Koen, A.L., Fairlie, L., Padayachee, S.D., Dheda, K., Barnabas, S.L., Bhorat, Q.E., Briner, C., Kwatra, G., Ahmed, K., Aley, P., Bhikha, S., Bhiman, J.N., Bhorat, A.E., du Plessis, J., Esmail, A. and Groenewald, M. (2021). Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. *New England Journal of Medicine* [accessed 10 Apr. 2021].

‌Study.com. (2021). *What is Data Density? - Definition & Small Multiples | Study.com*. [online] Available at: https://study.com/academy/lesson/what-is-data-density-definition-small-multiples.html#:~:text=Data%20density%20can%20be%20thought [Accessed 10 Apr. 2021].