**ISTE-600-02-2201** Analytical Thinking

Logic of Problem (LOP)

Title: Who Gets a Chance at the Vaccine?

Team # 4 Tanner Peterson, Nandhini Nagarathinam, and Sai Preetham Reddy Leburu

1. **Problem** **Title**: Socioeconomic Impact on COVID-19 Vaccine Rollout

2. **Statement of the Problem**:

The COVID-19 global pandemic is a problem that has been impacting the world for over a year now. A problem that started of containing, testing, minimizing death, has now become an issue on how to create herd immunity. The problem now is the number of vaccines to help obtain immunity. Do socioeconomic indicators impact the number of people that are receiving the vaccine?

**3.** **Elements of Thought:**

a. The key **Question** at issue is …

With vaccines available to help resolve the COVID-19 global pandemic, the question now is what counties are receiving a vaccine. Do socioeconomic indicators have an impact on the number of vaccines countries are administering to their population?

b. The **Purpose** of reasoning through the issues is …

The goal in reasoning this issue is to see on the global level which countries are administering vaccines to the public. Focus is to educate with publicly available data if socioeconomic factors do impact the vaccine rollout on the global level. Which counties are succeeding in administering the vaccine and do they have an advantage based on socioeconomic indicators?

c. The **Information** to use in answering this question is …

The information needed to solve this problem include datasets that are publicly curated. Data used throughout the project:

1. Vaccine dataset that has been curated by Our World in Data <https://ourworldindata.org/covid-vaccinations> compiled into Kaggle set https://www.kaggle.com/gpreda/covid-world-vaccination-progress
2. World Bank Data
   1. Education:https://data.worldbank.org/topic/4
   2. Economy:https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
   3. Mortality: https://data.worldbank.org/indicator/SH.DYN.NMRT
3. United States county-level vaccination data collected from the Economic Research Service -

<https://covid.cdc.gov/covid-data-tracker/#vaccinations>

<https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/>

There may be a need for additional information that is not in these datasets that could have contributing factors such as is it known that counties are not focusing on vaccines as a source of obtaining immunity. Or are there counties that simply do not have the means to receive a vaccine.

The vaccine dataset from the CDC will have to be preprocessed to transform raw data into an understandable format by removing incomplete and null values. The same procedure will have to be applied to the other datasets on education, poverty, and mortality collected from ERS as well. Merging the data gathered above as a final step, we can measure if the socio-economic factors have an impact on the vaccine distribution rate using a linear regression model and linear regression trees.

d. The **Assumptions** making when reasoning through this issue are …

This is assuming that the data reported from public repositories has been accurately collected into our vaccination dataset that has been published to Our World in Data and the Economic Research service.

Another is that it is assumed that socioeconomic factors do impact the number of vaccines administered because if a county has more means to obtain vaccines, they should have higher numbers of vaccinations.

Aside from poverty, other assumptions such education and population indicators can increase the likelihood counties are putting a vaccine as the focus rather than obtaining herd immunity without a vaccine.

e. The primary **Concepts** guiding the reasoning about the issue are …

The hypothesis to be tested is that counties that have better statistics relating to socioeconomic impacts are more likely to be administering more vaccines to their population.

f. Some important **Implications** of reasoning through this problem well or poorly are…

Implications of using socioeconomic impacts to analyze vaccine administering results may imply that populations that live in more “successful” counties may have an advantage in receiving a vaccine rather than immunity through obtaining the actual virus. The county dataset was used in place of the country, that said we used the same socioeconomic features of education, poverty (income), and mortality rate. These counties that have more favorable socioeconomic indicators would be the ones expected to have a higher rate of vaccination.

g. The **Point of View** from which we are looking at this situation is …

As stated, some counties may not have a vaccine as a focus. This may lead to an alternative solution rather than assuming having means to receive vaccines and administer them. Our point of view is that vaccines are the safest way to gain immunity and it is a focus for all countries. Other solutions may be accepted in other countries so knowing this going forward will help us understand results in a more reasonable way.

h. In reasoning through this issue, the main **Inferences** we have made and **Conclusion(s)** we have come to, in terms of how to deal with the issue are …

We will deal with the fact that data collected for vaccinations may have gaps. However, it is the most complete that has been found and reported. Also, inferring that countries want to vaccine their population rather than gaining immunity in other ways.

Based on statistical analysis, there was some relationship between socioeconomic indicators (education, poverty, mortality), however they are not the best predictors to make a sound model. We reject the null hypothesis because the P values are low, but the fit of the model is not ideal. Better than not having a model at all.

**4.** **Summary:**

The COVID-19 global pandemic has impacted the world for over a year. Now that vaccines have been created and approved, they are being administered to populations. Question is do socioeconomic factors impact the number of vaccines that are administered globally and can these factors group help predict the rate of vaccinations. The hypothesis is that socioeconomic factors do impact the number of vaccines and counties that focus on education, have a lower poverty rate, and lower mortality rate should lead to higher vaccinations.

The data exploration phase has helped in filtering the large data sets with a focus on relevant data by making use of various visualization methods.

I would like to suggest adding “MEASURABILITY” as an element of thought to the Logic of Problem as I believe it is important to have a realistic goal and this element can help us determine if our goals are measurable or not. Would also add what existing information is out there regarding the problem. It can assist us define the goal or move it in a new direction to find something new out relating to the problem.

**Updates made in Checkpoint #3**:

The dataset used in checkpoint #2 being a relatively small one couldn’t produce accurate results. The correlation coefficient obtained on running a linear regression technique was very low and statistically insignificant. To approach this issue from a different dimension, this model was tested on a much larger dataset with similar attributes. The dataset used was a compilation of United States county-level vaccine distribution data, which too showed low correlation.

I would like to suggest adding utility as a standard for evaluation to the logic of the problem. The data gathered should be useful and beneficial to solving the issue at hand. Presence of irrelevant and redundant data can impact accuracy and cause variability.

**Updates made in Final**

The entire LoP was updated based on the larger dataset, United States county vaccination data. The country dataset was too small to perform any data mining algorithms. What was also updated were the features that we selected as socioeconomic indicators. The conclusion was updated with our conclusions that were drawn doing the analysis in Python and in WEKA. Our two recommendations have been added to the LoP process.