Using the Pandemic Vulnerability Index Model to Examine the Risk Factors Associated with COVID-9

Lesson 1

Subject: High School Math 1

Lesson Title: Examining a Population’s Susceptibility to COVID-19 by Using the Pandemic Vulnerability Index Model

Alignment to the NGSS:

- Develop and use models.
- Use mathematics and computational thinking.
- Analyze and interpret data.
- Construct explanations (for science) and design solutions (for engineering).

Statement of the NC Essential Standards

- NC Math 1 Standards, NC.M1.F-LE.5—Interpret expressions for functions in terms of the situation they model.

  https://files.nc.gov/dpi/documents/curriculum/mathematics/scos/current/math-1.pdf

Student Learning Objectives

- Describe what a mathematical model is and what it is for.
- Interpret each risk factor’s (the variables in the model) impact on the spread of COVID-19 verbally and numerically (or qualitatively and quantitatively).
- Analyze a county’s PVI score and identify the areas that make this county less or more vulnerable to the spread of COVID-19.

Lesson Summary

Models describe our understanding about how the world functions; in mathematical modeling, we translate those beliefs into the precise and concise language of mathematics. The majority of interacting systems in the real world are far too complicated to model in their entirety. Hence, it is important to identify the most important parts of the system and use them to build the model. Mathematical modeling can be used to achieve the following objectives:

- Develop scientific understanding of a system.
- Test the effect of change in a system.
- Aid decision making.

The COVID-19 Pandemic Vulnerability Index (PVI) model was jointly developed by researchers at NIEHS, North Carolina State University, and Texas A&M University. This lesson will provide a brief description and analysis of the PVI mode, which examines and visualizes the vulnerability to COVID-19 for each county in the U.S. from multiple perspectives. These factors include infection rate, intervention measures such as testing, environmental determinants such as residential density, and social determinants such as health disparities associated with race and socioeconomic status. Students will use this model to compare susceptibility between counties to COVID-19 and discuss the factors that contribute to these differences. In addition, students will propose prevention and intervention strategies to reduce or avoid exposure to risk factors and their adverse health impacts.

Factors used in the creation of the PVI score and the weights applied to them were informed by epidemiological modeling as well as known contributors to general health morbidities.

The following factors were used in the creation of an overall PVI score.

- X_1: Transmissible cases, weight 0.20 (20%)
- X_2: Disease spread, weight 0.04 (4%)
- X_3: Population mobility, weight 0.08 (8%)
- X_4: Residential density, weight 0.08 (8%)
- X_5: Social distancing, weight 0.08 (8%)
- X_6: Testing, weight 0.08 (8%)
- X_7: Population demographics, weight 0.08 (8%)
- X_8: Air pollution, weight 0.08 (8%)
- X_9: Age distribution, weight 0.08 (8%)
- X_{10}: Co-morbidities, weight 0.08 (8%)
- X_{11}: Health disparities, weight 0.08 (8%)
- X_{12}: Hospital beds, weight 0.04 (4%)

There are 3,142 counties in the United States. The PVI is based on a given county's rank compared to each of the other counties. We first rank the individual factors by taking the raw value and scaling it zero to 1. While doing this we also reverse scale some factors such as the

testing %, as less testing contributes to higher risk. For each of the 3,142 counties and each factor (X_1 through X_{12}), the rank of the i^{th} county for the j^{th} factor is expressed as:

$$Rank\ Calculation\text{:}\ \ X_{ij} = \frac{X_{ij} - \min\,(X_j)}{\max\,(X_j) - \min\,(X_j)}$$

In this equation, j represents one of the twelve factors listed above and X_{ij} is the value of a specific factor j (for example: X_{i1} represents the value for *Transmissible Cases* and X_{i12} is the value for *Hospital Beds*) of the i^{th} county. Min(X_j) is the lowest value for factor j among the 3,142 counties and max(X_j) is the highest value of factor j among the 3,142 counties. This ranking allows each factor to be expressed in the same units (zero to one) and removes the difficulties associated with different units applied to different factors.

For ***each county*** (i = 1 to 3,142) we have:

- X_i – a vector of 12 calculated ranks associated with the 12 factors ($X_i = X_{i1}, X_{i2}, X_{i3}, \ldots, X_{i12}$ and every county will have a unique X_i vector)
    - X_{i1} would be the calculated rank for the i^{th} county's *Transmissible Cases* factor.
    - ...
    - X_{i12} would be the calculated rank for the i^{th} county's *Hospital Beds* factor.
    - Again, we'll use j to represent the 12 factors listed above with their ranks.

For ***all counties*** the weights (W) for the 12 factors are the same:

- W – a vector of 12 *Factor Weights* that sum to 1 ($W_{j=1} + W_{j=2} + W_{j=3} + \ldots + W_{j=12} = 1$)
    - $W_{j=1} = 0.20$ (*Transmissible Cases* Factor Weighting)
    - ...
    - $W_{j=12} = 0.04$ (*Hospital Beds* Factor Weighting)
    - Note: The 12 values in W are the same for ***every*** county. They are invariant with respect to county.

We get the PVI for the i^{th} county (i = 1 to 3,142) by summing the products of these two sets of 12 values and the mathematical formula is represented as:

$$PVI_i = \sum_{j=1}^{12} W_j X_{ij}$$

For example, the PVI for Wake county would be calculated as follows:

- Get the scaled factor ranks for wake county (X_{Wake} – this is a vector of Wake county's calculated ranks for each of the 12 factors ($X_{wake\_transmissible\ cases}$...$X_{wake\_hospital\ beds}$). All rank values are between zero and one.
- Multiply those 12 values to their respective weights (W) and sum the 12 products
- $PVI_{(Wake)}$ = (0.20 * Transmissible Case_$Rank_{Wake}$) + … + (0.04 * HospitalBeds_$Rank_{Wake}$)

NOTE: The weighted sum of factors ranks that are all scaled from zero to 1 also results in a zero to 1 quantity as well. This means that the PVI score is now calculated such that every county has a value between zero and 1, with values closer to zero representing the least risk, and values closer to 1 representing the highest risk.

Materials and Equipment

Computers and internet connection

Procedure

Engagement (15-20 minutes)

Teacher provides a brief introduction of the COVID-19 timeline and Pandemic Vulnerability Index (PVI) model (https://covid19pvi.niehs.nih.gov/). Students receive a copy of the PowerPoint slide "COVID-19 Timeline."

Students are divided into groups and discuss the following:

- What is a math model?
- What is the purpose of building this model? (Understand how a system works, make predictions, plan future actions, etc.)
- Suggest a simple, real-life scenario which can use a math model to aid the decision-making.

  Optional:

  Depending on the level of students, teacher will briefly discuss the stages of building a math model: variable and predictor selection, score construction, score validation, and assessment of accuracy.

Exploration (15-20 minutes)

Students will check the PVI model and examine the risk factors (independent variables) for this model and their weight (expressed in %). Students will then be divided into four groups and randomly assigned to discuss one of the following four groups of risk factors:

1. Infection Rate
2. Population Concentration
3. Intervention Measures
4. Health and Environment

Group 1—Infection Rate (24%):

- **Transmissible cases (20%)**: Population size divided by cases from the last 14 days which are the most likely to be transmissible. This metric is the number of “contagious” individuals relative to the population, so a greater number indicates more likely continued spread of disease.

Ask students to answer the following questions:

a. Explain why “a greater number indicates more likely continued spread of disease.”
COVID-19 spreads from person to person. A greater number means a bigger population that facilitates and sustains the infection process.

b. How frequently should Group 1 data be updated—daily, weekly, or monthly?
Group 1 data should be updated on daily basis.

- **Disease spread (4%)**: Fraction of total COVID-19 cases that are from the last 14 days (one incubation period). The incubation period of an infectious disease is the time between infection and symptom onset.

Ask students to answer these questions:

a. There is evidence some people infected by SARS-CoV-2 (the virus that causes COVID-19) never developed any symptoms (asymptomatic). Therefore, it is possible that these patients did not know they were infected and were never diagnosed. Will this change the “disease spread score?”

b. If your answer to question a is yes, do you think the “real infection rate” is higher or lower than the number you acquired from the data source? Explain.

c. Suggest what can be done to more accurately determine the number of positive cases during the last 14 days.
Widespread testing for COVID-19 and other plausible suggestions.

Group 2—Population Concentration (16%):

- **Population mobility (8%):**
  - Daytime density: Greater daytime population density is expected to increase the spread of infection because more people are in closer proximity to each other.
  - Baseline traffic: Average traffic volume per meter of major roadways in the county. Greater traffic volume is expected to increase the spread of infection due to more people moving and interacting. Baseline traffic is a metric to examine how densely populated the county is in a normal workday environment. More people going to and from work and to and from stores increase the likelihood of interaction and spread.
- **Residential density (8%)**: Data of families in multi-unit structures, mobile homes, over-crowding (more people than rooms), being without a vehicle, and persons in institutionalized group quarters. All of these variables are associated with greater residential density, which is expected to increase the spread of infection because more people are in closer proximity to each other.

Ask students to answer these questions:

a. Does high population density inevitably increase the spread of infectious disease?
High population density facilitates the spread of infectious diseases. This is particularly true for densely populated areas in developing countries that lack infrastructure for sanitation and hygiene. However, people's behavior (such as practicing social distancing and other precautionary measures) can mitigate this effect.

b. Assume you are developing this model. In addition to density, are there other characteristics of this population (e.g., behavior, age, educational attainment) that you may want to include in the model? Explain.
Yes. Individual behavior, age, accessibility to medical care, and other characteristics all have impact on the spread of the disease. Group 4 data include these factors.

c. Should Group 2 data be treated as static data or updated frequently (daily or weekly)?
Group 2 data are static and updated quarterly.

Group 3—Intervention Measures (16%):

- **Social distancing (8%):** Social distancing grade is assigned by looking at the change in overall distance traveled and the change in nonessential visits relative to baseline (previous year), based on cell phone mobility data. The grade is converted to a numerical score, with higher values indicating less social distancing (worse score). This is one of the factors that we mentioned were "reverse scaled," which means that a larger value represents decreased adherence to social distancing and therefore increased risk.
- **Testing (8%):** Population divided by tests performed (currently only statewide statistics are available, rather than county statistics). This is the inverse of the tests per population, so greater numbers indicate less testing. Lower testing rates mean it is more likely that infections are undetected, so we would expect to see an increase the spread of infection due to less information and awareness concerning the spread of the disease through the population.

Ask students to answer the following questions (or performing tasks):

a. How frequently should Group 3 data be updated—daily, weekly, or monthly?
   Daily

b. Do you expect the social distancing score to increase or decrease as an individual county moves from "Stay-at-Home" to "Reopening" phase?
   Increase

c. What are other intervention strategies that might be implemented to reduce the vulnerability to spread of COVID-19?
   Contact tracing, wearing facial masks, and frequent hand-washing combined with social distancing.

d. If you have other intervention strategies, should they be considered as additional variables for this model?
   An important factor to consider is whether there are readily available data that demonstrate the implementation level of the intervention. For example: How many people wear facial masks in the public?




Group 4—Health and Environment (44%):

- **Population demographics (8%)**
  - % Black: Percentage of population who self-identify as Black or African American.
  - % Native: Percentage of population who self-identify as American Indian or Alaska Native.
- **Air pollution (8%)**: Average daily density of fine particulate matter in micrograms per cubic meter (PM2.5). Air pollution has been associated with more severe outcomes from COVID-19 infection.
- **Age distribution (8%)**: Age 65 or older. Older ages have been associated with more severe outcomes from COVID-19 infection.
- **Co-morbidities (8%)**
  - Premature death: Years of potential life lost before age 75 per 100,000 population (age adjusted). This is a broad measure of health, and a proxy for cardiovascular and pulmonary diseases that have been associated with more severe outcomes from COVID-19 infection.
  - Smoking: Percentage of adults who are current smokers. Smoking has been associated with more severe outcomes from COVID-19 infection and also causes cardiovascular and pulmonary disease.
  - Diabetes: Percentage of adults age 20 and older with diagnosed diabetes. Diabetes has been associated with more severe outcomes from COVID-19 infection.
  - Obesity: Percentage of the adult population age 20 and older that reports a body mass index (BMI) greater than or equal to 30 kg/m2. Obesity has been associated with more severe outcomes from COVID-19 infection.
- **Health disparities (8%):**
  - Uninsured: Percentage uninsured in the total civilian noninstitutionalized population estimate. Individuals without insurance are more likely to be undercounted in infection statistics, and may have more severe outcomes due to lack of treatment.
  - Social Vulnerability Index (SVI) socioeconomic status: This score integrates data on percent of a given population below poverty, percent unemployed (historical), income, and percent without a high school diploma. Populations with lower socioeconomic status are more likely to be undercounted in infection statistics and may have more severe outcomes due to lack of treatment.
- **Hospital beds (4%):** Summation of hospital beds for hospitals with “open” status and “general medical and surgical” description.

Suggestions:

Teachers will provide definitions for all variables in this group and ask students what relationship they expect between each variable and the vulnerability to COVID-19 (for example, higher percentage of smoking among the population is associated with higher vulnerability to COVID-19). Ask students to review the following Centers for Disease Control and Prevention (CDC) webpages:

https://www.cdc.gov/diabetes/disparities.html
https://www.cdc.gov/tobacco/disparities/low-ses/index.htm
https://www.cdc.gov/nchs/data/nhis/earlyrelease/insur201905.pdf, Figure 6

Ask students to summarize the health disparities illustrated by these data.

Explanation and Elaboration (40-45 minutes)

- Each group explains what they learn from "Exploration" to the entire class.
- Pick one county from the following list and check its PVI score:
  a. New York, New York
  b. Scotland, North Carolina
  c. Boise, Idaho
  d. Student's home county

  The list might change as the pandemic continues.
- Analyze the PVI score of the chosen county and present the results to the entire class:
  - In terms of the vulnerability to the spread of COVID-19, what are the strengths and weaknesses of the county? For example, the county's strength is low level of co-morbidities and the weakness is low social distancing.
  - Examine the county's PVI score over the past __ months. What was the trend?

Do you have any suggestions to reduce the vulnerability of your county to the spread of COVID-19?

Evaluation

- **Option 1:** Task students with writing a concluding paragraph for this lesson to demonstrate their understanding and practice their science communication skills.
- **Option 2:** Task students with picking a county from the website, analyzing its PVI score, examining the trend, making suggestions to reduce the county's vulnerability, and justifying the suggestions.
- **Option 3:** Task students with assessing the model and evaluating whether adding additional variables will improve the predictive power of the model.