COVID-19 INCIDENCE IN THE INDIANA'S SECONDARY SCHOOL SYSTEM THROUGH A CONDITIONAL GAUSSIAN MODEL AND AN AGE-STRUCTURED COMPARTMENTAL MODEL

Kale Menchhofer [1], Nathan Mills [1], Kathleen Lois Foster [2], Alessandro Maria Selvitella [1]

[1]Purdue University Fort Wayne [2]Ball State University

Motivation:

- COVID-19 emerged in Wuhan (China) at the end of 2019 and was declared a pandemic by the World Health Organization in March 2020.
- With more than 2.5 million deaths worldwide as of late February 2021, COVID-19 has been a defining health crisis and has impacted people's everyday lives in countless ways.
- One of the most noteworthy circumstances of the COVID-19 outbreak in the United States was the closure of virtually all schools throughout the country.
- Since their closure, one of the most pressing issues pertaining to COVID-19 is how to properly reopen schools without sparking a surge in cases throughout the community.
- Currently, the situation is highly heterogeneous with even nearby schools adopting alternative strategies.
- The prolonged school closure has been shown to negatively affect student learning experience and to be the cause of serious mental illnesses, such as anxiety and depression.

Dataset and Software

- Data is taken from the Indiana Data Hub, updated to Dec. 28th, 2020. This dataset includes COVID-19 student cases broken down by school.
- The analysis and the simulations utilized the software *R* and the package *deSolve*.

Our Analysis

We will concentrate on a couple of distinct models with the intent of capturing important factors in the diffusion of the coronavirus in Indiana's secondary school system. For the sake of interpretability, we confined our analysis to the simplest models capturing the phenomenon under study.

Conditional Gaussian Model.
In the first model, we analyze the number of cases in each school, subdividing them by county. The distribution of the number of cases in schools within a given county is modeled with a Conditional Gaussian Distribution; namely, we model the number of cases in each county as a linear function of the sum of the student cases in that county plus a Gaussian error.

Age Structured Compartmental Model.
The second model is a compartmental model with age structure (4 compartments of young interacting with 4 compartments of adults). Compartmental models are models in which the population is divided into mutually exclusive and exhaustive classes, and the spread is modeled through a system of coupled ODEs describing the evolution of the disease across compartments.

Conditional Gaussian Model

We considered the number of student cases y_i in Indiana's county i and the sum number of cases per secondary school x_i in county i for $i = 1, \dots, 92$, with 92 the number of counties in Indiana. Our model is a simple linear regression model of the form $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ with the x_i considered non-stochastic, $E[\epsilon_i] = 0$, $\epsilon_i \sim N(0, \sigma^2)$ and ϵ_i independent and identically distributed for $i = 1, \dots, 92$. Although our analysis was comprehensive of 1) mean and sum for students/teachers/employees/all of them [8 models], 2) Conditional Gaussian/Poisson/Negative Binomial for each model with Outlier detection at 1-2-3 st. dev., 3) Cooks distance for all models, and 4) Non- parametric outlier detection tests for all models, for space reasons, we report in this poster only the result on the relationship between the sum of student cases per county.

Age-structured SEIR model

We considered the following SEIR model with two age-groups: children vs adults. We have the following system of coupled differential equations:

$$(SEIR2) \quad \begin{cases} \frac{dS_1}{dt} = -S_1\left(\beta_{11} I_1 + \beta_{21} I_2\right) \\ \frac{dE_1}{dt} = S_1\left(\beta_{11} I_1 + \beta_{21} I_2\right) - \sigma_1 E_1 \\ \frac{dI_1}{dt} = \sigma_1 E_1 - \gamma_1 I_1 \\ \frac{dR_1}{dt} = \gamma_1 I_1 \end{cases} \qquad \begin{cases} \frac{dS_2}{dt} = -S_2\left(\beta_{12} I_1 + \beta_{22} I_2\right) \\ \frac{dE_2}{dt} = S_2\left(\beta_{12} I_1 + \beta_{22} I_2\right) - \sigma_2 E_2 \\ \frac{dI_2}{dt} = \sigma_2 E_2 - \gamma_2 I_2 \\ \frac{dR_1}{dt} = \gamma_2 I_2 \end{cases} \tag{1}$$

Here $S_i(t), E_i(t), I_i(t), R_i(t) \in C^1([0, +\infty))$. To fix the ideas: $i = 1$ represents the children age group and $i = 2$ the adult age group with $S_i(t), E_i(t), I_i(t), R_i(t)$ the corresponding susceptible, exposed, infective, and removed individuals of age group i. The following theorem implies that $SEIR2$ gives biologically meaningful solutions for all times t.

Theorem

For every $0 \le S_{i0}, E_{i0}, I_{i0}, R_{i0} \le 1$ $i = 1, 2$ such that $S_{i0} + E_{i0} + I_{i0} + R_{i0} = 1$ for $i = 1, 2$, there exists a unique solution to system $(SEIR2)$ such that $I_i(0) = I_{i0}, E_i(0) = E_{i0}, I_i(0) = I_{i0}, R_i(0) = R_{i0}$, $0 \le S_i(t), E_i(t), I_i(t), R_i(t) \le 1$ for $i = 1, 2$, and $S_i(t) + E_i(t) + I_i(t) + R_i(t) = 1$ for $i = 1, 2$.

Sketch of the Proof By Picard–Lindelöf existence and uniqueness theorem, there is a unique smooth solution local in time. Summing the equations in each system, we deduce that the population is conserved. Since the total population is conserved $S_i(t) + E_i(t) + I_i(t) + R_i(t) = S_{i0} + E_{i0} + I_{i0} + R_{i0} = 1$. Therefore, the solution is global in time. By its equation, S_1 is decreasing. By taking the ratio between $\frac{dS_1(t)}{dt}$ and $\frac{dR_1(t)}{dt}$ and integrating from 0 to $+\infty$, we get by conservation of total population:

$$\frac{S_{1\infty}}{S_{10}} = e^{\left\{-\left[\frac{\beta_{11}}{\gamma_1} R_{1\infty} + \frac{\beta_{21}}{\gamma_2} R_{2\infty}\right]\right\}} \text{ and so } S_{1\infty} \ge S_{10} e^{\left\{-\left[\frac{\beta_{11}}{\gamma_1} R_{1\infty} + \frac{\beta_{21}}{\gamma_2} R_{2\infty}\right]\right\}} > 0.$$

This applies similarly for the second age-group and analogously for the other compartments.

Simulations

In our simulations, we will use the population values for Indiana and the epidemiological parameters in Table 1.

State	Description	Range/Estimate	Base Case
β_{11}	child-to-child	[0.05-2]	0.1
β_{12}	child-to-adult	[0.05-2]	0.5
β_{21}	adult-to-child	[0.05-2]	0.5
β_{22}	adult-to-adult	[0.05-2]	0.5
$1/\sigma_1$	child latent	3	3
$1/\sigma_2$	adult latent	3	3
$1/\gamma_1$	child infectious	4	4
$1/\gamma_2$	adult infections	4	4

Table 1:This table provides the parameter values for our 17 simulations. The β's are the transmission coefficients, whose range are given per day. The latent and infectious periods are in days.

As an example, we report the simulations for Allen County, which is characterized by the following parameters: children population (≤ 17) $n_1 = 97, 101$, adult population $n_2 = 282, 198$ (> 17), and initial conditions for the eight compartments: $S_{10} = 97, 099/n_1, E_{10} = 2/n_1, I_{10} = 0, R_{10} = 0, S_{20} = 282, 195/n_2, E_{20} = 2/n_2, I_{20} = 1/n_2, R_{20} = 0$.

Results

Conditional Gaussian Model. Interestingly, the number of cases per county is roughly 30 times the sum of the student cases in each county. This value is stable across counties. The estimate of the slope coefficient $\hat{\beta}_1 = 29.694$ gives significance of the predictor with p-value $< 2 * 10^{-16}$.

Age-structured model. The most interesting models had the smallest or the greatest proportion of each age group having contracted COVID-19. Extreme cases 1, 3, and 4 depicted optimal outcomes in which less than 5% of either age group have contracted the disease by the end of a 90 days period (Figure 1). The most calamitous outcomes, which were exhibited in extreme cases 6, 8, 9, 11, 12, 13, 14, 15, and 16, showed that more than 99% of both age groups contract the disease within 90 days. Cases 12, 15, and 16 showcased the worst potential scenarios with over 99% of both age groups being exposed to or contracting COVID-19 before day 20.

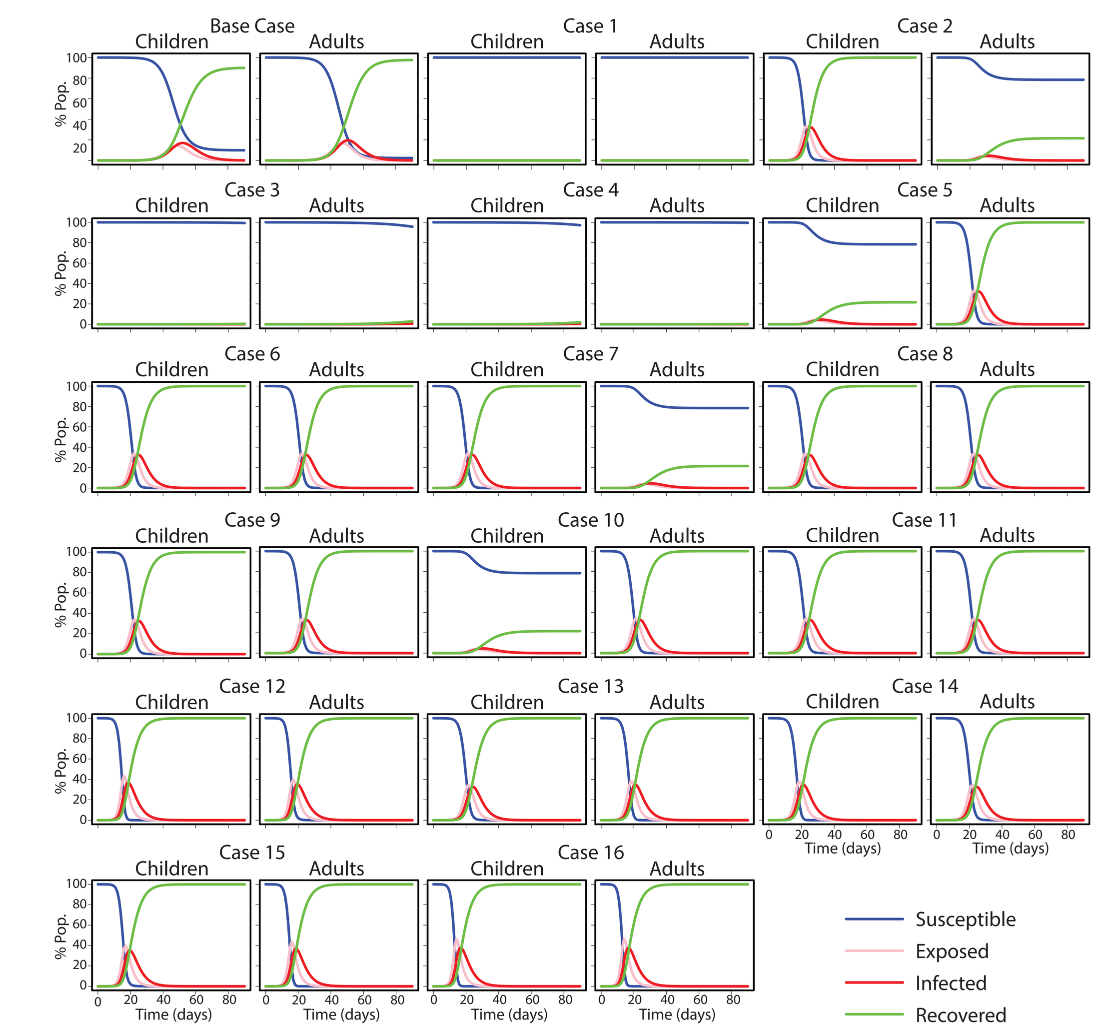

Figure 1:Trajectories of the 8-compartment SEIR models using the parameters of our simulations.

Discussion and Conclusions

Conditional Gaussian Model. The conditional sum of the student cases per county scales linearly with the number of cases of the county. This has speculatively important public policy related consequences, including the possibility of concentrating the testing in schools and using the scaling factor to estimate the incidence of COVID-19 in the full population.

Age Structured Model. The simulations of our models with parameters in line with those of Indiana showed that even if adults keep their contact with other adults to a minimum, transmission from young can present itself to be extremely detrimental to the more at-risk population. This shows that optimal school reopening strategies can potentially benefit not only the school population, but the entire community.

Overall Message. Taken in conjunction, these results underline once more the importance of adopting proper school reopening strategies and how they relate to the diffusion of the coronavirus outside the school environment. The diffusion of the coronavirus among the school population has the potential to not only be a strong determinant of the health of the more at risk population, such as elderly and sick, but also be a proxy for the incidence of COVID-19 in the community.