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

Anonymous authors
Paper under double-blind review

ABSTRACT

In this work, we have studied a Conditional Gaussian model and an SEIR model with age structure to understand the diffusion of the coronavirus in Indiana's secondary school system and its impact to public health. Our analysis revealed that (i) the conditional sum of the school cases per county scales linearly with respect to the cases of the county and that (ii) even if adults keep their contact with other adults low, transmission from young can present itself preponderantly in the adult population within few weeks.

1 INTRODUCTION

COVID-19 emerged in Wuhan (China) at the end of 2019 and was declared a pandemic by the World Health Organization in March 2020 (World Health Organization). With more than 2.5 million deaths worldwide as of late February 2021 (Worldometer), 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 (Viner et al., 2020; Esposito & Principi, 2020). 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 (Landeros et al., 2020; Sheikh et al., 2021). Currently, the situation is highly heterogeneous with even nearby schools adopting alternative strategies (Rafiei & Mello, 2020; Hubbard et al., 2020; Kearney & Childs, 2021). Note that 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 (Hoffman & Miller, 2020; Lee, 2020; Tang et al., 2021; Armitage & Nellums, 2020).

Much about the dynamics of the spread of infectious diseases such as COVID-19 can be analyzed by means of statistical and mathematical models. In this work, 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. 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 (James et al., 2013); 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. An interesting fact emerges from the analysis: 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. 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 (Brauer et al., 2019). The simulations of our models with parameters in line with those of Indiana (Indiana State Department of Health) 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. 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. We plan to extend our study beyond Indiana to all US counties and validate the accuracy and deductions of our models.

The remaining part of the paper is organized as follows. Section 2 is dedicated to the methods, while Section 3 to the results and discussion. Conclusions are presented in Section 4.

2 Materials and Methods

Data is taken from the Indiana Data Hub (Indiana State Department of Health), updated to Dec. 28th, 2020. This dataset includes COVID-19 student cases broken down by school.

2.1 Conditional Gaussian model

In our first 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 manuscript only the result on the relationship between the sum of student cases per county.

2.2 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 theoretical results implies that $SEIR2$ gives biologically meaningful solutions for all times t.

Theorem 2.1. *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 (Ambrosetti, 2012), 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:

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

Similarly for the second age-group and analogously for the other compartments. This completes the proof of the theorem. □

In our simulations, we will use the population values for Indiana (Indiana State Department of Health) and the epidemiological parameters in Table 1. Similar epidemiological parameters have been used in (Landeros et al., 2020; Li et al., 2020; Bar-On et al., 2020):

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 (Landeros et al., 2020; Li et al., 2020; Bar-On et al., 2020). The β's are the transmission coefficients, whose range are given per day. The latent and infectious periods are in days.

The simulations have been performed with *R* software and the package *deSolve*. As an example, we report the simulation for Allen County, which is characterized by the following parameters (Stats Indiana): children population ($\leq$ 17) $n_1 = 97,101$, adult population $n_2 = 282,198$ ($>$ 17), and initial conditions for the eight compartments: $S_{10} = 97,099/n_1, E_{10} = 2/n_1, I_{10} = 0, R_{10} = 0, S_{20} = 282,195/n_2, E_{20} = 2/n_2, I_{20} = 1n_2, R_{20} = 0$.

3 Results and Discussion

3.1 Conditional Gaussian Model

Our analysis provided a very interesting result with the number of cases per county being roughly 30 times the sum of the student cases in each county. The value is stable across counties. The estimate of the slope coefficient $\hat{\beta}_1 = 29.694$ gives significance of the predictor with p-value $< 2 * 10^{-16}$ with test statistic $F_{1,90} = 171.8$.

3.2 Age-structured model

Best Case Scenarios: The models produced that were of most interest were those in which the smallest proportion of each age group and the greatest proportion of each age group 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).

Worst Case Scenarios: The most calamitous outcomes, which were exhibited in extreme cases 6, 8, 9, 11, 12, 13, 14, 15, and 16, showed more than 99% of both age groups contracting 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 the twentieth day. The plots for these three cases show the highest infectious curves, meaning that these cases had both the greatest proportions of both age groups infected at one time and the earliest time for such a large proportion of both groups to be infected, both of which would attest to a great burden on the theoretical population's healthcare system (Figure 1). Adults, particularly those over 65, are considered more likely to have a severe case of COVID-19 than children. Case 12, employed a minimum rate for transmission among adults, though the rate for transmission from children to adults took on a maximum value, which accounted for over 99% of adults having been exposed to or contracted COVID-19 by Day 20. Although this case is strictly hypothetical, it enables the deduction that even if adults keep their contact with other adults to a minimum and follow other proper protocols, transmission from children can present itself to be extremely detrimental to the more at-risk population.

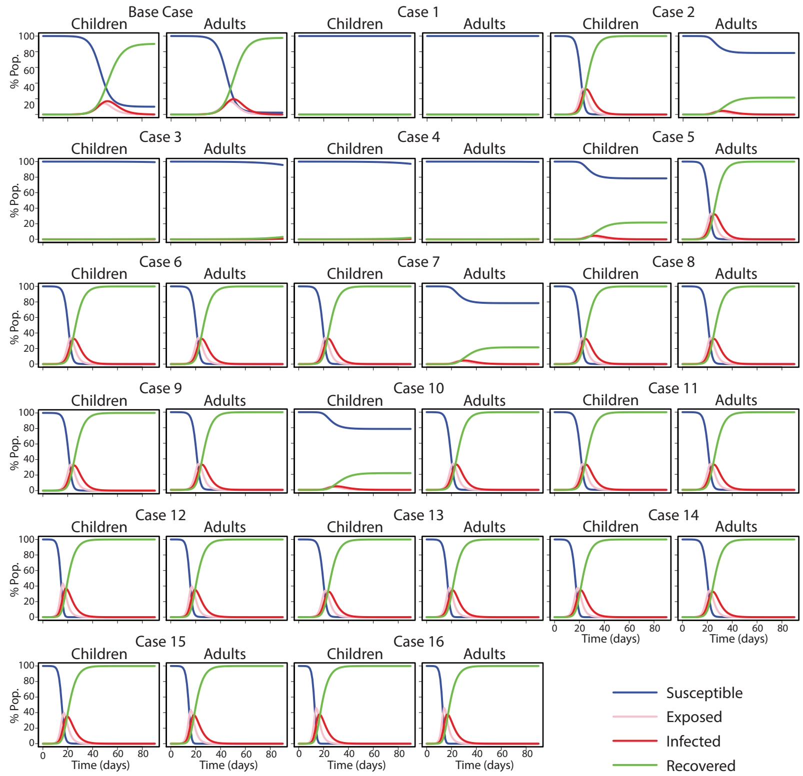

Figure 1: Trajectories of the 8-compartment SEIR models using the parameters of our simulations.

4 CONCLUSIONS

In this paper, we have studied the COVID-19 incidence in the Indiana secondary school system. We considered two separate models: (i) a Conditional Gaussian model, which showed that, in the time period considered, COVID-19 cases of a county and the sum of student cases in the county are related by a stable constant factor and (ii) a compartmental model with age-structure, which showed that even if adults keep their contact with other adults to a minimum, transmission from children can be extremely detrimental to the adult population in a way that in the worst case scenario more than 99% of both age compartments contract the disease within 90 days. These results further confirm that the incidence of coronavirus in schools can affect the health of the more at-risk population, but it can also be representative of the incidence of COVID-19 in the community at large. We plan to extend our study beyond Indiana to all US counties, validating the accuracy and deductions of our models further, and present this extension at the workshop.

REFERENCES

Antonio Ambrosetti. *Appunti sulle equazioni differenziali ordinarie*, volume La Matematica per il 3+2. Springer Verlag Italia, 2012.

Richard Armitage and Laura B. Nellums. Considering inequalities in the school closure response to covid-19. *The Lancet Global Health*, 8(5):e644, 2020.

Yinon M Bar-On, Avi Flamholz, Rob Phillips, and Ron Milo. Science forum: Sars-cov-2 (covid-19) by the numbers. *eLife*, 9:e57309, 2020. doi: 10.7554/eLife.57309.

Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical Models in Epidemiology*, volume 69. Springer-Verlag New York, 2019.

Susanna Esposito and Nicola Principi. School closure during the coronavirus disease 2019 (covid-19) pandemic: An effective intervention at the global level? *JAMA Pediatrics*, 174(10):921–922, 10 2020. doi: 10.1001/jamapediatrics.2020.1892.

Jessica A. Hoffman and Edward A. Miller. Addressing the consequences of school closure due to covid-19 on children's physical and mental well-being. *World Medical & Health Policy*, 12(3): 300–310, 2020. doi: https://doi.org/10.1002/wmh3.365.

Lea Hubbard, Hollie Mackey, and Jonathan A. Supovitz. District response to the covid-19 pandemic. *CPRE Policy Briefs*, 2020. URL https://repository.upenn.edu/cpre_policybriefs/88.

Indiana State Department of Health. Covid-19 cases by school. URL https://www.stats.indiana.edu/profiles/profiles.asp?scope_choice=a&county_changer=18003.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *Introduction to Statistical Learning*. Springer, 2013.

Christopher A. Kearney and Joshua Childs. A multi-tiered systems of support blueprint for re-opening schools following covid-19 shutdown. *Children and Youth Services Review*, 122:105919, 2021. doi: https://doi.org/10.1016/j.childyouth.2020.105919.

Alfonso Landeros, Xiang Ji, Kenneth L. Lange, Timothy C. Stutz, Jason Xu, Mary E. Sehl, and Janet S. Sinsheimer. An examination of school reopening strategies during the sars-cov-2 pandemic. *medRxiv*, pp. 1–16, 2020.

Joyce Lee. Mental health effects of school closures during covid-19. *The Lancet Child and Adolescent Health*, 4(6):421, 2020.

Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science*, 368(6490):489–493, 2020. doi: 10.1126/science.abb3221.

Yasmin Rafiei and Michelle M. Mello. The missing piece — sars-cov-2 testing and school reopening. *New England Journal of Medicine*, 383(23):e126, 2020. doi: 10.1056/NEJMp2028209.

Aziz Sheikh, Asiyah Sheikh, Zakariya Sheikh, and Sangeeta Dhami. Reopening schools after the covid-19 lockdown. *Journal of Global Health*, 10:010376, 2021.

Stats Indiana. Allen county, indiana. URL https://hub.mph.in.gov/dataset/covid-19-cases-by-school.

Suqin Tang, Mi Xiang, Teris Cheung, and Yu-Tao Xiang. Mental health and its correlates among children and adolescents during covid-19 school closure: The importance of parent-child discussion. *Journal of Affective Disorders*, 279:353–360, 2021. doi: https://doi.org/10.1016/j.jad.2020.10.016.

Russell M Viner, Simon J Russell, Helen Croker, Jessica Packer, Joseph Ward, Claire Stansfield, Oliver Mytton, Chris Bonell, and Robert Booy. School closure and management practices during coronavirus outbreaks including covid-19: a rapid systematic review. *The Lancet Child & Adolescent Health*, 4(5):397–404, 2020. doi: https://doi.org/10.1016/S2352-4642(20)30095-X. URL https://www.sciencedirect.com/science/article/pii/S235246422030095X.

World Health Organization. Rolling updates on coronavirus disease (covid-19). URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen.

Worldometer. Covid-19 coronavirus pandemic. URL https://www.worldometers.info/coronavirus/.