DATA STORIES ABOUT DISEASE: DATA-CENTERED EPIDEMIOLOGY IN SECONDARY SCHOOLS

By Jan Mokros [1], Jacob Sagrans[1], Penny Noyce[1], and James Nicholson[2]
July 29, 2025
[1]Tumblehome, Inc.
[2] Northumbria University SmartCentre

1. INTRODUCTION

Epidemiology is the study of patterns of health and illness and how these are distributed in a given population. Combining elements of biology and medical science, sociology, environmental science, and statistics, it could be defined as the intersection of public health and data science. Epidemiology is increasingly relevant to modern life. Worldwide travel, human habitations encroaching on wild lands, and climate change are making epidemics ever more frequent. Meanwhile, political leaders in the US make controversial and often unsupported claims about the causes of chronic diseases. Yet epidemiology has virtually no presence in the K–12 curriculum in the US. For example, in the Next Generation Science Standards (NGSS) the words “pandemic,” “epidemic,” and “epidemiology” do not occur at all (NGSS, n.d.). This is a missed opportunity. A thoughtful combination of epidemiology and data science has the potential to provide an interdisciplinary curriculum that young people find highly motivating. Mathematics, biology, and what history teaches us about human behavior can combine with data to tell a coherent story. The topic of diseases and how they spread highlights the nature of scientific discovery while furnishing young people with key skills for the modern workplace such as analyzing information and reasoning from evidence.

Etymologically, “epidemiology” means “the study of that which is upon the people.” Hippocrates (ca. 400 B.C.E., *On Airs, Waters, and Places*) differentiated between epidemic diseases, or those that were visited upon a population from outside, and endemic diseases, which appeared intrinsic within a population. In the early 1800s, epidemiology followed the course of infectious epidemics; only later did epidemiologists begin to investigate the causes and correlations of such chronic diseases as diabetes and heart disease. In relation to data science, infectious disease lends itself to time-series and comparative data, while endemic disease also motivates studies of correlation and multifactorial regression. Our focus in this paper is primarily on infectious diseases and the ways in which time-series data are used in educational settings to study these diseases.

The goal of this paper is to prompt educators, researchers, and policy makers to consider the educational value of teaching data-centered epidemiology: specifically, the value of developing statistical understanding of how to interrogate and interpret large datasets while at the same time developing an understanding of how diseases spread. The work is interdisciplinary, and while it may be hard to fit into a standard curriculum, there are many potential rewards.

Over the past 20 years, several educators have seen the opportunity to engage students with epidemiological data. In this paper, we have sought to gather promising examples of epidemiology projects and curriculum units, especially those in which epidemiology takes full advantage of data science. We begin by discussing efforts to teach epidemiology that do not involve the use of genuine datasets, then briefly describe the rapidly growing interest in teaching data science. We describe an emerging movement called Civic Statistics, which seeks to integrate data science with topics of contemporary social concern. We follow with examples of curricula that organically meld the topics of epidemiology and data science. We conclude this section of the paper by describing technological tools for simulating and modeling the spread of infections or endemic disease. These tools are often easy to use and allow for open-ended student investigations. The final section of the paper offers four ways in which schools and education systems can build on these examples and create more opportunities for young students to explore data science through epidemiology-focused replacement units, enrichment in or outside of school, professional development for teachers, or modification of existing standards to include epidemiology.

2. HISTORY OF THE EPIDEMIOLOGY EDUCATION MOVEMENT AND LESS DATA-CENTRIC APPROACHES TO EPIDEMIOLOGY EDUCATION

Most efforts to teach epidemiology in the 21st century to middle and high school students did not involve much actual data, though the importance of using evidence was emphasized. The Epidemiology Education Movement, begun in the US in the early 2000s, started with a curriculum project titled “Detectives in the Classroom,” which introduced epidemiology to middle and high school students (Epidemiology Education Movement, n.d.-d; Epidemiology Education Movement, n.d.-a). Noteworthy features of the movement included: 1) An emphasis on student investigation of epidemiological problems; 2) A focus on professional development to introduce teachers to epidemiological concepts, and; 3) The development of approximately a dozen “enduring understandings” which, like curriculum standards, provided core beliefs around which teaching materials could be structured (Epidemiology Education Movement, n.d.-c). These understandings emphasize observation, pattern-finding, hypothesis testing, and reasoning from evidence to interpret results. They served as the basis for epidemiology standards in three US states (California, Tennessee, and Georgia—see Epidemiology Education Movement, n.d.-e).

At the time the researchers in the “Movement” were finishing their work, the US Centers for Disease Control and Prevention (CDC) used these enduring understandings to develop a parallel set of

"Epidemiology and Public Health Science Core Competencies" (Cordeira & Cordell, 2015). The CDC work had a greater emphasis on mathematical analysis of empirical data, including mathematical modeling.

Dozens of curricular resources were developed by the "Movement" researchers as well as by collaborators at the CDC itself (Epidemiology Education Movement, n.d.-b; CDC, 2024). Some of the units, such as Detectives in the Classroom, were used in school courses in mathematics, science, and health. Sometimes these same resources were used as preparation for competitions, including the Young Epidemiology Scholars (YES) competition for high school students, which reached approximately 5,000 students and ended in 2011 (YES, n.d.). Another competition, the team-based Science Olympiad, partnered with the CDC to develop background lessons on epidemics to prepare students to compete in its Disease Detectives division (Science Olympiad, 2025). These lessons present pre-formulated data and data visualizations, and ask students to interpret them.

The work of these projects took place largely before free, intuitive to use, accessible online data analysis tools were available to secondary school students, and before the widespread public availability of health datasets. Perhaps as a consequence, the curricular resources developed by these projects did not focus on examining actual data in the ways that epidemiologists do. Students were not, for the most part, asked to query large datasets.

In the past several years, the CDC has developed outreach projects to teach citizens about the importance of vaccination or about monitoring and preventing infectious outbreaks. NERD Academy (CDC, n.d.-b) is an eight-hour enrichment curriculum for middle school students that teaches, through animated videos and discussion, how diseases spread and how public health workers strive to prevent or contain them. Two lessons focus on how surveillance allows for the collection of data and how different kinds of graphs, particularly time-series graphs, tell the story of an outbreak. It is perhaps surprising that this and other CDC offerings (see, for example, the Science Ambassador lesson plans—CDC, 2024) make no use of the wealth of real data collected and reported by the agency. Instead, they use data from a fictional disease to make graphs and develop other visual displays of data. But students do not have opportunities to query the data, filter or group it, or make other "data moves" (Erickson et al., 2019). There is much more that could be done with epidemiological data, as will be seen below in Section 5.

3. THE RISE OF DATA SCIENCE EDUCATION

By 2017, "data science education" for secondary students was in full swing (see, for example, the 2017 Data Science Education Technology Conference, Concord Consortium, 2017). Strong data science programs were emerging in the US, UK, Germany, Israel, the Netherlands, Australia, and New Zealand. Researchers in this field promote the idea that students should be involved in all phases of the data inquiry process, starting from formulating questions and proceeding to descriptive analyses that involve making data displays, filtering data, and finding patterns and

relationships in the data (Israel-Fishelson et al., 2024). An essential part of this work involves examining large, multivariate datasets. However, even by the early 2020s, most teachers were not using these types of datasets with their students and instead worked with relatively small datasets (fewer than 50 cases) with two or three variables (Rosenberg et al., 2022). From the perspective of teaching data science, this was not ideal in that students could not truly investigate patterns in data that were so constrained in size and number of variables. Such small datasets put serious limits on what students can learn about data science.

As an emerging field of study, data science education began to develop teaching and learning guidelines to inform a common pedagogical vision. In the US, the American Statistical Association (ASA) and National Council of Teachers of Mathematics (NCTM) took leading roles by developing the Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education (GAISE) in 2007, which was subsequently revised in 2020 as GAISE II (Franklin et al., 2007; Bargagliotti et al., 2020). The GAISE guidelines provide a roadmap for educators in having students work with data. There are four key elements in this roadmap, the second of which, in the revised GAISE II (2020), is to “collect/consider data.” This is a key change from the 2007 GAISE guidelines, where teachers were encouraged to have students simply “collect” data: now teachers are encouraged to *have students examine and interpret existing datasets* rather than always collect data on their own. A newer set of data science learning progressions, currently being developed by Data Science for Everyone (DS4E), will provide far more detailed learning outcomes for pre-college students (DS4E, 2024). Throughout these progressions, there is a strong focus on students using existing datasets that are meaningful to them. In both the GAISE guidelines and DS4E learning progressions, there is also an emphasis on using data tools that make it feasible to examine large datasets. Without these tools, making meaning of large datasets is an overwhelming task.

At the same time that guidelines/learning progressions in data science education were being initiated, new, accessible, and dynamic public health datasets were emerging from the CDC, Johns Hopkins, and other research institutions (CDC, n.d.-a; Johns Hopkins, 2023). Data about COVID-19 were published at least weekly, and sometimes daily, over time periods that ultimately spanned years. The datasets were certainly large, as they included cases from countries around the world. Equally important, they included multiple variables (number of cases, hospitalizations, deaths, age and gender of patients, country or state in which the infections occurred, and, later on, vaccination rates). Though the public health implications of COVID were grave, the opportunities for learning about data were unparalleled. Not only would students have access to rich, genuine, and timely datasets, but they would also have opportunities to combine their own lived experiences with epidemiological data to better understand a phenomenon that impacted virtually everyone, worldwide (see especially the discussion in sections 5a and 5c below).

4. CIVIC STATISTICS

While there has been much to celebrate in human progress, we still face existential threats, for example, from pandemics and from climate change. The Civic Statistics movement grew out of a recognition that there are systemic difficulties within education in addressing complex contemporary social issues, since they involve multivariate data. In the humanities many teachers lack the necessary quantitative skills, while in STEM many teachers focus on technical skills at the expense of interpretation. Often neglected is evaluation of the myriad consequences of policy decisions. For example, at the outset of a new pandemic, how might mandatory vaccination or a mandate for social distancing affect the spread of disease, economic growth, or mental health of the population? Civic Statistics engages with important social issues by analyzing authentic, authoritative data sets, and visualizing data using innovative technologies to understand the stories the data tell.

ProCivicStat was a collaboration of statistics educators from multiple countries, funded by the EU Erasmus program (ProCivicStat, n.d.). It developed a framework describing facets of what constitutes Civic Statistics. The framework includes components that would be recognizable from many traditional statistics courses, but it has several notable additions (see Ridgway, 2022 for a much fuller elaboration than we have room for here). One of the facets, *Meaning for Society and Policy* is perhaps the heart of Civic Statistics: Social policy deals with strategic plans about allocation of (almost always) scarce resources and other broad actions of governmental bodies or institutional agencies as they try to respond to pressing needs of society. These bodies always operate under the constraints of imperfect information—imperfect in quality, in timeliness, and in relevance. A second facet, *Critical Evaluation and Reflection* is a skill that many statistics education courses aspire to instill. Attacks on evidence-informed policy making and the proliferation of “alternative facts,” misinformation, and disinformation have made this skill absolutely vital in evaluating options for social policy. Introductory statistics courses often teach standard procedures to model data (for example, linear modeling) with little regard to context, underlying assumptions or the quality of the model. Another facet, namely *Models, Patterns and Representations,* emphasizes aspects of modeling used in making judgments about causality and about confounding or mediating variables that may help to explain the patterns seen in the data. The complex nature and many features of social phenomena require sophisticated representations of multivariate data where interactions between variables and non-linear relationships are the norm, so this facet goes well beyond what is covered in most standard statistics courses. Ridgway (2021) provides a useful analysis of the data science skills required when studying a pandemic to develop skills necessary for informed citizenship.

Individual activities would not be expected to cover the full range of facets described above, but a well-designed course in Civic Statistics would provide substantive coverage of them all through exposure to multiple activities. The CIDSEE project (see fuller description in section 5a below) provides considerable opportunities to address *Meaning for Society and Policy* in particular, but is also very strong in the *Critical Evaluation and Reflection* facet to support student thinking.

We are aware of the US National Academy of Education's 2021 report *Educating for Civic Reasoning and Discourse* (C. D. Lee et al., 2021), which concludes that civic engagement includes thinking through "a public issue using rigorous inquiry skills and methods to weigh different points of view and examine available evidence" (p. 1) as well as discussing the challenges of those issues as a way to grow collective and individual understanding. However, the report's recommendations include very little of substance in respect of the statistics aspect of Civic Statistics as described above.

Irgens et al. (2020) argue that Data Science Education (DSE) plays an important role in supporting civic participation by providing context and an environment for students to develop their voices and competencies for engaging in legitimate, authentic action. DSE is grounded in authentic contexts and focuses on critical thinking. These features provide a background for student agency and ownership over the inquiry process and steps towards action (V. R. Lee et al., 2021).

5. PROJECTS THAT MELD DATA SCIENCE AND EPIDEMIOLOGY

As data science has begun to make its way into the K–12 curriculum, curriculum writers and teachers have become more willing to include substantial time for data in units primarily written to teach epidemiology, and epidemiology has more frequently been selected to illustrate aspects of data science. Efforts to examine the data dimensions of epidemiology range from simple descriptive statistics, to time-series data, to examinations of correlation and multiple regression. During the COVID pandemic, as news sources experimented with different ways to represent data —for example, with colored area maps or running averages—so too curriculum developers began to develop and use technology tools to simulate or explore the spread of disease. Certain of these tools will be discussed in greater detail in Section 6 of this paper. Here we give a more general overview of entire curricula.

5A. COVID-INSPIRED DATA SCIENCE EDUCATION THROUGH EPIDEMIOLOGY (CIDSEE)

The start of the COVID pandemic in the spring of 2020 led to an explosion of interest in infectious disease, and several groups started educational efforts to help students understand what was happening around them. One US example is CIDSEE, which stands for COVID-Inspired Data Science Education through Epidemiology (Tumblehome, Inc., n.d.). Developed with National Science Foundation (NSF) funding (DRL-2313212) for students in grades 5–8 in out-of-school clubs, CIDSEE was organized around an adventure novel, *Pandemics!* (Noyce, 2023). The novel leads a pair of middle school students on a journey through time and space to learn from scientists and health workers of the past and present about such epidemics as smallpox, Ebola, measles, and COVID.

Afterschool or summer facilitators, many of whom had little science background, spent four to

six hours preparing to teach the 15-hour curriculum, which featured hands-on activities, roleplays, policy debates, experimenting with NetLogo disease spread simulations, and graphing, including graph sonification. Youth "played" graphs with kazoos, raising or lowering their tone to match the changes in infection rates over time; later, the Common Online Data Analysis Platform (CODAP) allowed them to listen to graphs of infection rates via its built-in Sonify plugin (NetLogo, n.d.; Concord Consortium, n.d.). Investigations included comparing racial disparities in COVID infection rates in different states and the differential uptake of vaccines in different counties.

CIDSEE reached over 1,200 high-need middle schoolers from 2020 to 2025. Assessment showed significant growth in students' STEM engagement, STEM career knowledge, and STEM identity, along with an increased appreciation for the helpfulness of data in the life of the community. A key feature of CIDSEE was youth's access to large amounts (case numbers in the thousands and higher) of real, current data at the county, state, and country level. CIDSEE continues as the abbreviated Data Detectives program, accessible at https://www.tumblehome.org/curriculum.

5B. CONTAGIOUS MATHS – NRICH

NRICH is a mathematics education organization that is part of Cambridge University Mathematics Department, and part of the Mathematics Millenium Project (MMP). NRICH focuses on providing open-access, rich problem-solving activities. Julia Gog, an award-winning professor of mathematical biology who advised the UK government during the pandemic, used one of her awards to produce resources for mathematics students aged 11–16 and for teachers. These resources explore modeling of epidemics and infectious diseases and show how mathematics can change the world for the better.

The materials (NRICH, 2021) are presented in four parts, and each includes teachers' resources to support delivery in the classroom and activities to develop and explore key ideas. In each part, videos introduce either next steps in developing a model or key aspects of the virology or transmission of an infectious disease. Students begin with a physical simulation of disease spread. Once the process is established, students move to a software model that allows them to see exponential growth and the effect of changing the value of R0 on the way the disease spreads.

In later sessions, the model becomes gradually more sophisticated as it takes account of different factors, such as random variations in human behavior and interactions. The simulation allows the disease to be introduced in different settings, such as at a party. Next, variations in physiology and susceptibility to infection are added, including whether individuals have been vaccinated. Simulations created for the project allow users to manipulate different parameters, explore how they affect the spread of the disease, and consider how they match observed behavior of the infection. Together these introduced variations illustrate how various

interventions, such as restrictions on social behavior, can affect disease spread. Some of the visual representations are not immediately easy to relate to the parameters in the model, but the simulations introduce a very impressive array of factors. There are also discussions exploring practical decisions about who should be vaccinated if there is a limited supply, and the ethical issues that arise.

A second version of the materials, developed for use with older students, includes accessible articles on different aspects of contagious diseases and on the mathematics involved in modeling them.

5C. THE INSTITUTE FOR RESEARCH IN SCHOOLS (IRIS)––BIG DATA: COVID-19

IRIS is a UK charity that was set up to provide opportunities for young people to participate in cutting-edge research and to collaborate with leading universities and institutions while still at school. It particularly targets participation in schools serving communities in areas of deprivation, aiming to break down barriers that impact underrepresented young people entering STEM careers. When COVID-19 appeared, the standard IRIS model of expert support could not be followed, so the charity produced a suite of teaching materials, freely available online for teachers to download and use (IRIS, 2020).

The resources give a broad insight into the variety of backgrounds of people working in epidemiology, and the cross-disciplinary nature of the work. There is also a PowerPoint for whole class teaching and discussion that introduces communicable and non-communicable diseases and the elements of public health emergencies. The ambition of the project is very much about the data, but emphasizing that if you want to understand the data you need to understand the context, hence there is also a strong focus on the virology, transmission, and clinical presentation of COVID-19.

The substantive task in the first part is based on a fictional disease, “Exan-21,” with students being asked to be part of a “task force” to advise the government. Students work independently in small groups on seven different questions and then reconvene to each present their conclusions and give advice to the government. They have two documents—information about Exan-21 and its transmission and an additional information booklet that has statistical information about the demographics in the UK—so responding to the questions is fairly data intensive.

Once the pandemic restrictions were lifted, IRIS returned to its normal model with Big Data: COVID-19 (IRIS, 2022). This unit updated previous materials to include seminars and training packages on the virology of COVID-19 and epidemiology analysis using Excel and the statistical programming language R.

The unique aspect of IRIS projects is that students propose their own research questions, using guidance notes and suggested themes to explore. Support is available, especially for the use of R. Working individually or in groups, students submit a poster presentation of their research and have the opportunity to present it at one of the student conferences IRIS holds. These are in the style of professional academic conferences, with a keynote speaker, two sessions of student research presentations plus two poster sessions. Most students present through their poster, but a few present to the full conference of several hundred students, academics, teachers, and scientific researchers. Examples of previous student investigations are one that compared government approaches to lockdowns in the UK and in New Zealand, investigating the effects on the number of COVID deaths (Derby College, 2023). Another compared the effect on recreational behavior in India and in the US in relation to the numbers of new COVID-19 cases being reported (Dartford Grammar School for Boys, 2024). In both investigations the students identified and accessed relevant information from open-source data repositories, and analyzed it to answer their chosen research question.

The “Big Data: COVID-19” project is the most comprehensive attempt we have encountered to engage students with both data science and the science of epidemiology. However, the level of specialist resources and support available to students from IRIS exceeds what would be feasible for large-scale implementation if epidemiology were to be introduced into the curriculum.

6. ACCESSIBLE TECHNOLOGY TOOLS FOR DATA-ENHANCED EPIDEMIOLOGY EDUCATION

In this section, we describe potential uses of technology tools for analyzing epidemiological data or simulating and modeling the spread of infections or endemic disease in educational contexts. By “accessible,” we mean tools that are free, available to run in one’s internet browser or are easily downloadable, and that are suitable for middle school students (ages 10–14) and older. These tools often allow for open-ended student data investigations. Often these tools have simpler, integrated datasets or simulations that teachers or students could start working on at first, but the tools can also have “high ceilings” for more advanced users wanting to explore more complex epidemiological data or create more complex simulations that could be useful to undergraduate or graduate students or even working professionals.

CODAP, developed by the Concord Consortium, a K–12 STEM education nonprofit, provides a free, intuitive online interface for data exploration, analysis, and visualization (visit codap.concord.org to launch the software and view many helpful resources around how to use it) (Concord Consortium, n.d.). Authentic data can be pre-loaded into CODAP. One example can be found in a series of out-of-school Data Clubs modules (Data Clubs, n.d.). Of particular epidemiological interest is the Data Clubs Ticks and Lyme module, which includes a dataset on rates of Lyme disease over time in US states that have deer ticks, combined with various environmental variables such as the percentage of forest cover in each state or the amount of

winter precipitation.

One can upload any CSV file containing data into CODAP for further exploration. One helpful CODAP feature for educators is the ability to create "shared view" links of CODAP documents. One could share a table of epidemiological data via a CODAP link and ask students to make graphs from the data and report on what they discover. This is the approach that the CIDSEE and Data Clubs projects have used in having middle school students work with epidemiological data in informal contexts (the full CIDSEE/Data Detectives program leader guide, which includes all the CODAP shared view links, can be requested at https://www.tumblehome.org/curriculum, and the Data Clubs modules can be requested at https://www.terc.edu/dataclubs/modules/).

NetLogo is a simple, programmable, multi-agent modeling environment created by Uri Wilensky at Northwestern University (NetLogo, n.d.). A disease spread model can be explored alone or can be embedded in CODAP, which allows students to track data generated by the simulations in CODAP tables and graphs. An example of a disease spread simulation from the CIDSEE curriculum with NetLogo embedded in it can be found at https://bit.ly/46jg1aS. This simulation is particularly useful for seeing how outcomes change (such as the surviving population at the end of the simulation) as the percentage of people starting immune increases, simulating the effects of the rollout of vaccinations. Note also that Concord Consortium has developed a CODAP/NetLogo integration guide if one wishes to create their own simulation (see CODAP, 2024).

Another free resource developed by the Concord Consortium is the online **SageModeler** systems modeling software, which can be used to model an epidemic and generate data from the model to explore, using an interface similar to CODAP (SageModeler, n.d.). After launching SageModeler, one can select from an example static epidemic model or two dynamic epidemic models. The main SageModeler website (https://sagemodeler.concord.org/) offers many additional help resources, should one wish to create a more complex epidemic model.

Continuing with epidemiological modeling tools, Brockman (2018) created a website called **Complexity Explorables** with collections of free interactive models that run in one's internet browser, including five epidemiology models. One of these epidemiology models—"I herd you"—allows users to explore herd immunity and how changing factors such as transmission rates and vaccination rates affect the course of the disease. **GLEAMviz** (2022) is another free software tool available to download on Macs, PCs, and Linux systems that offers a more sophisticated modeling environment. It uses real-world data on population sizes/locations to simulate the spread of infectious diseases. GLEAMviz simulations have been used in math camps in Africa as part of the Supporting African Maths Initiatives (SAMI), as discussed in a forthcoming paper (Nicholson & McCusker, 2025).

The physical activities in the maths camps used parameters typical of measles, and of Ebola, and the camp instructors then constructed GLEAMviz simulations of a measles outbreak in Accra

(for use in a camp in Ghana), and of an Ebola outbreak in Monrovia (for use in a camp in Liberia). While it can be used at school age level, the GLEAMviz environment offers almost unlimited scope for exploring more complex models and the interaction with real-time population movement data as an integral part of the process offers the opportunity for it to be used in research (Nicholson & McCusker, 2025, p. 3).

Operation Outbreak has developed a free smartphone app and accompanying guidance that has been used by groups of middle school students through adults to simulate outbreaks involving interpersonal interaction (Operation Outbreak, n.d.). As an "infected" individual (and their phone) comes into contact with a susceptible individual (and their phone), the susceptible individual may become infected as well. After the end of the simulation, data generated by the simulation, such as how many people became infected, become available in the smartphone app for all participants to explore.

Particle People is another free, browser-based disease spread simulation developed by researchers in the UK (Particle People, n.d.). Many parameters can be changed before running a Particle People simulation, such as the probability of a person becoming sick when exposed or the proportion of people isolating when sick. Upon the simulation completing, a graph generates the number of sick, recovered, dead, and unaffected people over the course of the simulation. The developers of the simulation have discussed how they have used it with students aged 8 to 16 (McNeill et al., 2023).

It stands out to us that most of the tools we discovered and discussed above are used for simulating or modeling disease spread, not for exploring real-world epidemiology/public health data, despite the ready availability of such data online, particularly since the start of the COVID pandemic. In order to answer one's own questions about data, and to do descriptive examination of these data, a more general tool such as CODAP is needed, and one often needs to take extra steps such as finding the data of interest, curating it, and importing it into the tool. These tasks can be challenging and time-consuming for teachers, especially if they don't have staff development or IT support.

7. FUTURE DIRECTIONS

Although COVID is receding into the background of our common lives, both epidemiology and data science continue to be vital subjects for children to learn. An understanding of data science allows a citizen to make informed decisions about social and personal questions that matter, and both private employers and governments increasingly seek data scientists to profit from the huge amounts of data being gathered every day. However, epidemiology threatens to fall victim to a cycle of panic and neglect, with interest waning as the most recent pandemic, with its millions of deaths worldwide, fades in memory. It does not have to be so. As demonstrated in the examples above, the study of disease and its spread offers an engaging, relevant field of play as students

learn data skills. Sickness is of interest to all ages, and epidemiology lends itself to the examination of data at different levels of complexity, from simple descriptive statistics at the elementary level, through time-series data and an intuitive understanding of correlating factors at middle school, to more complex methods to account for correlation, causation, modeling and decision-making at the high school level and beyond.

In assembling this paper, we have envisioned four ways to integrate data-centered epidemiology into the K–12 experience of students. The first is the continued development of replacement units to be slotted into the regular school-day curriculum. A replacement unit involves a deliberate decision to substitute one curriculum unit for a related one, usually because the new unit offers more promising pedagogical approaches. With teacher collaboration, an integrated unit of health, science, and mathematics could be scheduled among those courses over two or three weeks to substitute for less engaging content. After all, “Analyzing and Interpreting Data” is one of the science and engineering practices urged on schools in the Next Generation Science Standards, while Continuity and Change is one of the cross-cutting concepts (NGSS, n.d.). A data-focused epidemiology unit is likely to motivate increased interest in STEM and could be addressed at all grade levels. Learning about how to use real data to investigate outbreaks can be very motivating because of its combination of relevance, authenticity, and dynamic nature.

The second path forward, if the formal curriculum remains impervious to change, is the development of enrichment units, which can be taught to a smaller pull-out group of students or, better, a wide range of students in an informal setting. Often, a pullout group for enrichment implies a “Gifted and Talented” group. For example, the CIDSEE curriculum did attract interest from such groups; but afterschool programs serving youth with widely varying backgrounds and skills were also deemed successful by their instructors.

Curriculum for out-of-school learning must be fun and involve movement. Open-ended explorations and games will be especially attractive to students. Epidemiology problems lend themselves well to such competitions as the Science Olympiad. One caveat is that experience suggests epidemiology and data science cannot be taught effectively in drop-in programs: Students need some continuity in order to build their conceptual understanding (Noyce et al., 2025, p. 27).

A third path forward, probably imperative for lasting change, is high-quality professional development. Teachers need instruction in some major tenets of epidemiology. A suggested starting point is the aforementioned “enduring understandings of epidemiology” developed by the Epidemiology Education Movement (n.d.-c). Teachers need to understand R0, surveillance, herd immunity and other concepts. They also need familiarity with at least some of the data and simulation tools presented above. Such familiarity will be best gained through time to use the tools while exploring questions of their own. To take full advantage of large, authentic

compilations of data, teachers will need experience (or help) locating data sets and importing them into tools like CODAP.

Finally, it may be time to consider adding data-based epidemiology to educational standards documents where these are used (for example in the US and New Zealand) and in national curriculum frameworks such as in the UK and some other European countries. Emerging guidelines/learning progressions in data science, such as those under development by DS4E or recently developed by GAISE, could be illustrated through examples from epidemiology. While appetite for rewriting the Next Generation Science Standards remains low, there is considerable interest in adding more elements that relate to evidence-based decision making for personal and community life (Stanford, n.d.; Zucker, 2022). While many state standards documents mirror the NGSS, states show considerable flexibility in updating their standards to meet the needs of a changing world. National curricula have tended to be stagnant for substantial periods of time, until the imposition of substantial changes. The current UK curriculum review is promising evolution rather than revolution, and recognizes that the world is now changing so quickly that updating will need to occur on a more frequent basis than it has in the past.

We hope that this paper will instigate a broader discussion of the advantages of using epidemiology as a motivation for teaching data science, and the potential for epidemiology and data science together to contribute to the education of thoughtful, informed citizens. There is much more to be learned.

FUNDING DETAILS

The work reported on in this paper has been funded by the National Science Foundation (NSF), grant number DRL-2313212. Any opinions, findings, conclusions, or recommendations presented are only those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., and Spangler, D. A. (2020). *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II): A Framework for Statistics and Data Science Education*. National Council of Teachers of Mathematics (NCTM). https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Brockman, D. (2018). Complexity Explorables https://www.complexity-explorables.org/fields/epidemiology/

Centers for Disease Control and Prevention (CDC). (2024). STEM Lesson Plans. Science Ambassador Program. https://www.cdc.gov/science-ambassador/stem-lesson-plans/index.html

CDC. (n.d.-a). COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker/#datatracker-home

CDC. (n.d.-b). NERD Academy. https://www.cdc.gov/nerd-academy/site.html

CODAP. (2024). How do I integrate a NetLogo simulation into CODAP? https://codap.concord.org/faqs/how-do-i-integrate-netlogo-simulation-codap/

Concord Consortium. (n.d.). Common Online Data Analysis Platform (CODAP) [computer software]. https://codap.concord.org/

Concord Consortium. (2017). Data Science Education Technology Conference. February 15-17, 2017. David Brower Center, University of California, Berkeley. https://codap.concord.org/dset/index.html

Cordeira, K. L., & Cordell, R. L. (2015). Epidemiology and Public Health Science: Core Competencies for High School Students. Centers for Disease Control and Prevention (CDC). https://www.cdc.gov/science-ambassador/documents/ephs-competencies.pdf

Data Clubs. (n.d.). Data Clubs Website. https://www.terc.edu/dataclubs/

Data Science 4 Everyone (DS4E). (2024, July 12). The first steps to building a national learning progression for K-12 data science education. https://www.datascience4everyone.org/post/building-a-national-learning-progression-for-k-12-data-science-education

Dartford Grammar School for Boys. (2024). Big Data: Covid – The effect of new daily Covid cases on recreational mobility. https://researchinschools.org/wp-content/uploads/2023/06/L036_BDC_Dartford_Effect-of-New-Daily-Cases-on-Recreational-Mobility-no-surn.pdf

Derby College. (2023). How UK and New Zealand approaches to lockdown affected the number of Covid deaths. https://researchinschools.org/wp-content/uploads/2023/09/L041_BDC_DerbyCollege_different-government-approaches-for-lockdown-in-the-UK-and-New-Zealand_NS.pdf

Epidemiology Education Movement. (n.d.-a). Detectives in the Classroom. https://web.archive.org/web/20240731122913/http://www.epiedmovement.org/detectives.html

Epidemiology Education Movement. (n.d.-b). Develop Curricula. https://web.archive.org/web/20240801071801/http://www.epiedmovement.org/developCurricula.html

Epidemiology Education Movement. (n.d.-c). Enduring Understandings.
https://web.archive.org/web/20220117045142/http://www.epiedmovement.org/BMEU.html

Epidemiology Education Movement. (n.d.-d). Epidemiology Education Movement.
https://web.archive.org/web/20250206171135/https://www.epiedmovement.org/

Epidemiology Education Movement. (n.d.-e). High School Performance Standards for Epidemiology.
https://web.archive.org/web/20240718234212/https://epiedmovement.org/DOCS/StateStandards.pdf

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education, 12(1)*. https://www.doi.org/10.5070/T5121038001

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K–12 Curriculum Framework.* American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaiseprek-12_full.pdf

GLEAMviz. (2022). GLEAMviz [computer software]. https://www.gleamviz.org/
Irgens et al. (2020). Data literacies and social justice: Exploring critical data literacies through sociocultural perspectives. International Conference on Learning Sciences.
https://repository.isls.org/bitstream/1/6665/1/406-413.pdf

IRIS. (2020). It takes a community. https://researchinschools.org/projects/it-takes-a-community/
I
RIS. (2022). Big Data: COVID-19. https://researchinschools.org/projects/big-data-covid/
Israel-Fishelson, R., Moon, P., Tabak, R., & Weintrop, D. (2024). Understanding the data in K-12 data science. Harvard Data Science Review, 6(2). https://doi.org/10.1162/99608f92.4f3ac3da

Johns Hopkins University. (2023). Coronavirus Resource Center. https://coronavirus.jhu.edu/

Lee, C. D., White, G., & Dong, D. (Eds.). (2021). Educating for Civic Reasoning and Discourse. Committee on Civic Reasoning and Discourse. National Academy of Education.
https://naeducation.org/educating-for-civic-reasoning-and-discourse-report/

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. Educational Researcher, 50(9), 664-672.
https://doi.org/10.3102/0013189X211048810

McNeill, G., Sondag, M., Powell, S., Asplin, P., Turkay, C., Moller, F., & Archambault, D. (2023). From asymptomatics to zombies: Visualization-based education of disease modeling for children. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. ACM, New York, Article 839, 17 pages. https://doi.org/10.1145/3544548.3581573

NetLogo. (n.d.). NetLogo [computer software]. https://ccl.northwestern.edu/netlogo/

Next Generation Science Standards (NGSS). (n.d.) https://www.nextgenscience.org/

Nicholson, J., & McCusker, S. (2025). It's in the wind: Opportunity for understanding epidemiological evidence [forthcoming]. Proceedings of the International Association for Statistical Education (IASE) Satellite Conference, "Statistics and Data Science Education in STEAM," September 30th to October 2nd, 2025, Universität Münster, Münster, Germany. https://www.uni-muenster.de/GIMB/en/veranstaltungen/iase/theme.html

Noyce, P. (2023). Pandemics! Tumblehome.

Noyce, P., Martin, L., Mokros, J., & Sagrans, J. (2025). It takes a network: How to scale up an afterschool STEM program. Afterschool Matters (Spring 2025). https://www.niost.org/Afterschool-Matters-Spring-2025/it-takes-a-network

NRICH (2021) Contagious maths: understanding the spread of infectious diseases. https://nrich.maths.org/contagious-maths-understanding-spread-infectious-diseases-0

Operation Outbreak. (n.d.). Operation Outbreak. https://operationoutbreak.org/

Particle People. (n.d.). Particle People. https://contact-viz.cim.warwick.ac.uk/particle-people/

ProCivicStat International Statistical Literacy Project (ISLP). (n.d.) ProCivicStat website. https://iase-web.org/islp/pcs/

Ridgway, J. (2021). COVID and data science: Understanding R0 could change your life. *Teaching Statistics, 43*, S84–S92. https://doi.org/10.1111/test.12273

Ridgway, J. (Ed.). (2022). Statistics for empowerment and social engagement: Teaching civic statistics to develop informed citizens. Springer. https://doi.org/10.1007/978-3-031-20748-8

Rosenberg, J. M., Schultheis, E. H., Kjelvik, M. K., Reedy, A., & Sultana, O. (2022). Big data, big changes? The technologies and sources of data used in science classrooms. *British Journal of Educational Technology, 53*, 1179–1201. https://doi.org/10.1111/bjet.13245

SageModeler. (n.d.). SageModeler [computer software]. https://sagemodeler.concord.org/

Science Olympiad. (2025). Disease Detectives. https://www.soinc.org/disease-detectives-c

SMART Centre. (n.d.). SMART Centre. https://research.northumbria.ac.uk/smartcentre/

Stanford University. (n.d.). Science Education in an Age of Misinformation. https://sciedandmisinfo.stanford.edu/

Tumblehome, Inc. (n.d.). CIDSEE/Data Detectives. https://www.tumblehome.org/projects/datadetectives

Young Epidemiology Scholars Competition (YES). (n.d.) Young Epidemiology Scholars Competition website. http://www.yes-competition.org/yes//index.html

YouTube. (2025, February 10). Disease Models. https://www.youtube.com/watch?v=IK_4FtjJj_0

Zucker, A. (2022, June 17). An important report from Stanford. *Improving Science Education Standards: Science Education for a Vibrant Democracy*. https://improvethengss.org/author/andyzucker/