Planning for monitoring the introduction and effectiveness of new vaccines using real-word data and geospatial visualization: An example using rotavirus vaccines with potential application to SARS-CoV-2

T. Christopher Mast [a,*], David Heyman [b], Erik Dasbach [a], Craig Roberts [a], Michelle G. Goveia [a], Lyn Finelli [a]

[a] Merck & Co., Inc., Kenilworth, NJ, USA
[b] Axis Maps, Hewitt, TX, USA

ARTICLE INFO

ABSTRACT

Background: Infectious diseases continue to cause significant impact on human health. Vaccines are instrumental in preventing infectious diseases and mitigating pandemics and epidemics. SARS-CoV-2 is the most recent example of an urgent pandemic that requires the development of vaccines. This study combined real-world data and geospatial visualization techniques to demonstrate methods to monitor and communicate the uptake and impact of existing and new vaccines.

Methods: Observational data of existing pediatric rotavirus vaccines were used as an example. A large US national insurance claims database was accessed to build an analytic dataset for a 20-year period (1996–2017). For each week and multiple geographic scales, animated spatial and non-spatial visualization techniques were applied to demonstrate changes in seasonal rotavirus epidemic curves and population-based disease rates before, during, and after vaccine introduction in 2006. The geographic scales included national, state, county and zip code tabulation areas. An online web-based digital atlas was built to display either continuous or snapshot visualizations of disease patterns, vaccine uptake, and improved health outcomes after vaccination (http://www.mapvaccines.com).

Results: Over 17 million zip code-weeks of data were available for analysis. The animations show geospatial patterns of rotavirus-related medical encounter rates peaking every year from November – February prior to vaccine availability in 2006. Visualizations showed increasing vaccination coverage rates at all geographic scales over time. Declines in medical encounter rates accelerated as vaccination coverage rapidly increased after 2010. The data maps also identified geographic hotspots with low vaccination rates and persistent disease rates.

Conclusion: This project developed novel web-based methods to communicate location and time-based vaccine uptake and the related reduction in medical visits due to viral infection. Future applications of the visualization could be used by health agencies to monitor known or novel disease patterns over time in conjunction with close assessment of current and future vaccine utilization.

© 2021 Merck Sharp & Dohme Corp. and the Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Infectious diseases continue to have a significant impact on human health. Recent vaccine development has focused on the prevention of pathogens such as human papillomavirus, respiratory syncytial virus, pneumococcal disease and cytomegalovirus [1]. As of January 2021, a global pandemic caused by the novel coronavirus (SARS-CoV-2) [2] has resulted in approximately 92 million reported cases and over 1.9 million deaths [3,4]. In the United States, approximately 22.7 million cases and 379,000 deaths have been reported [5]; the number of hospitalized cases is over 131,000 [6]; and the short-term prediction of new COVID-19 hospitalizations per day ranges from 9,200 to 23,000 [7]. The pandemic continues to be a dynamic public health emergency that emphasizes the importance of vaccine development and delivery.

After a vaccine is approved by regulatory agencies, the capacity to rapidly obtain post-licensure real-world data on vaccine effectiveness, coupled with monitoring of vaccine distribution is critical for assessing the successful implementation of vaccination campaigns. In addition, these data provide information on the value of the vaccination program.

* Corresponding author at: Center for Observational and Real-World Evidence, Merck & Co., Inc., Mailstop WP37A-150, 770 Sumneytown Pike, West Point, PA 19486, USA.
E-mail address: christopher.mast@merck.com (T.C. Mast).

Early in the COVID-19 pandemic, nonpharmaceutical interventions (NPIs) were introduced to mitigate the impact of COVID-19 and control transmission [8]. A recent meta-analysis of over 172 studies suggested that, when used optimally, NPIs such as social distancing are effective at reducing infections but not preventing all infections [9]. Epidemic and pandemic planning [8] is now intensely focused on the collaborative development of antiviral medications and vaccines [10] integral to mitigation.

While data visualization techniques have previously assessed disease patterns, there are no known examples of large real-world data visualizations that use geospatial animation to monitor the introduction and impact of new medical interventions. Specifically, we are not aware of existing visualizations that demonstrate temporal trends for both the utilization of new vaccines and the associated impact on health outcomes at the national, state, county and zip code level. Geographic information systems (GIS), when designed and developed appropriately, can be important tools for public health decision-makers – particularly when operating under conditions of change and uncertainty [11]. Published reports have visualized the geographic location of increasing COVID-19 case counts over time [4,12]; however, these visualizations have not linked case counts to medical interventions. Prior visualization work has illustrated the impact of several routine childhood vaccines on associated diseases but lacked precise geographic data on vaccine administration or uptake [13].

We demonstrate the application of real-world data, epidemiological methods, and geospatial visualization to monitor the uptake and health impact of a new vaccine. This study used the example of a licensed pediatric vaccine that was developed to prevent rotavirus gastroenteritis. Before rotavirus vaccines became available in 2006, rotavirus gastroenteritis was estimated to cause over 500,000 deaths in children under 5 years of age globally every year and 50,000 annual hospitalizations in the US. The introduction of a rotavirus vaccine in 2006 resulted in sustained reductions in gastroenteritis disease rates and associated hospitalizations in the US [14,15].

Geospatial visualization of new vaccine uptake dynamics was previously developed by our group as a demonstration model [16]. The current project extends the previous work by describing how to visualize the impact of a vaccination program on health outcomes by time and geographic location. The use-case mapped hotspots where there were low vaccination rates and higher disease rates. Rotavirus vaccines were used as a real-world example of monitoring the uptake and impact of a new vaccine. It will be useful to have visualizations that track the uptake of new vaccines (e.g., COVID-19 vaccine) and rates of infectious disease.

The specific objectives of the project were to use a known vaccine preventable disease to: (1) develop an animated visualization of the hospitalizations and medical encounters of rotavirus gastroenteritis trends over time and by geographic units before and after vaccine introduction in the US; (2) use rotavirus vaccines as an example to show the rate and geographic pattern of uptake after US licensure using geospatial visualization techniques; (3) develop an animated visualization showing the association of rotavirus vaccine uptake and rotavirus health outcomes over time and geography; and (4) develop and deploy an interactive web-based digital atlas to demonstrate and disseminate study outcomes.

2. Methods

2.1. Design

The study design was primarily descriptive. The project goal was to develop novel methods for geospatial visualization that could be deployed as an internet accessible tool to describe changing health outcomes as a function of a new vaccine introduction.

2.2. Methodology components

There were four components required to visualize vaccine impact. First, the study required development of an analytic dataset that captured historical patterns of disease before and after a vaccine was available. Second, the database required analysis parameters relevant to the new vaccine (e.g., vaccination status by dose and age, geographic location of vaccine administration). Third, the analytic dataset required relevant health outcomes that could be linked, using discrete geographic and time-based units, to vaccine administration so that a geospatial visualization could be developed. Lastly, customized spatial and non-spatial interactive graphics were designed and implemented in order to visualize the vaccine and disease data by combining multiple visualization techniques and datasets across time and space.

2.3. Database source and overall epidemiological approach

The source of the analytic file is the IBM MarketScan® commercial database that consists of medical and drug data from employers and health plans for over 203 million individuals annually, encompassing employees, their spouses and dependents who are covered by employer-sponsored private health insurance in the US. Medical claims were linked to vaccination administrative claims and person-level enrollment information. This database is one of largest in the US with health-related information and was best suited to develop a visualization demonstration of national data with detailed geospatial and spatio-temporal information.

To develop the overall analysis file, several variables were constructed. The detailed variables and underlying coding definitions are described in Appendix A. The methods are summarized in the following sections.

2.4. Vaccination data

There are currently two rotavirus vaccines administered in the US. Vaccine status was defined as uptake of either RV5 (RotaTeq®, Rotavirus Vaccine, Live Oral Pentavalent, Merck & Co., Inc., Kenilworth, NJ, USA) or RV1 (Rotarix®, Rotavirus Vaccine, Live, Oral, GlaxoSmithKline Biologicals, Rixensart, Belgium). Because the vaccines are to be completed by a certain age, vaccination status represents the percent of children under eight months of age (the maximum Advisory Committee on Immunization Practices (ACIP) recommended vaccination age) in each zip code and by week with at least one code for administration of RV5 or RV1. As noted in previous work [15], the database was queried for weekly Current Procedural Terminology (CPT) codes for RV5, which was licensed in 2006, and CPT codes for RV1, which was licensed in 2008. Uptake was not assigned to individuals but rather defined as the percent of the analytic denominator, which refers to the number of children under eight months of age in each geographic unit (i.e., national, state, country, zip code) for each week.

Data were assessed for children with continuous enrollment in the insurance database during the specific time periods that were being assessed (e.g., 6 weeks of age through 8 months of age for vaccination coverage). If the child was not continuously enrolled during that time period, they were not included in the analysis. Stop and start dates were calculated to assess continuous enrollment during the analysis period.

2.5. Medical outcomes related to disease

Using previously developed methods [15], the database contained outcome data on actual medical encounters (hospitalizations, emergency department visits, or outpatient visits) using International Classification of Diseases (ICD) 9/10 codes for acute gastroenteritis (AGE) and rotavirus gastroenteritis (RGE) from February 1, 1996 to December 31, 2017. To focus this demonstration project, only RGE was assessed in this study. Because the primary burden of rotavirus-related disease on the healthcare system is among children < 2 years of age, the analytic file for this project was based on all children < 24 months of age in the database. The overall time period provides a pattern of RGE before and after vaccine introduction in 2006. In alignment with vaccine completion and known patterns of rotavirus gastroenteritis disease burden, health outcomes were based on children between 8 and 24 months of age with continuous enrollment for that time period. Over the time period of 1996–2017, the database contains a median of 316,623 births per year. The outcome was expressed as a population-based rate of medical encounters per 1,000 population of children eligible for the vaccine. As with vaccine uptake, outcomes are not assigned to an individual child but rather are defined as the percent of the analysis denominator in geographic unit that had a medical encounter in any given week. The monthly rate of rotavirus medical encounters is calculated as the sum of the weekly rate of medical encounters due to rotavirus gastroenteritis. The monthly vaccination rate is the mean of the weekly vaccination rate among members of the study group.

2.6. Geographic data

The vaccination and outcome data were assigned to four geographic levels: 5-digit zip code, county, state and national. For the analysis, each geographic unit was newly assessed for each subject at the beginning of each epidemiologic week.

2.7. Hotspot map to identify areas requiring further interventions

This county-based national map was developed as a potential public health tool to allow health officials to locate geographic areas that were identified as hotspots. We defined this outcome by visualizing counties that had both: 1) low vaccination rates defined as below the national average of 70% in 2015–17 and 2) rotavirus-related medical encounter rates 5 times higher than the national average (1 per 1,000 eligible population) during the same period.

2.8. Overall outcomes

An animated visualization tool was developed to show patterns of rotavirus gastroenteritis medical encounters over time and by distinct geographic units in the time periods before and after rotavirus vaccine introduction in the US. Visualizations also showed the variance of health outcomes by percent of vaccine uptake among the eligible population in each geographic analysis unit.

2.9. Visualization and programming methods

The analytic datafile was structured for visualization. Subsequently, data were animated by using a rolling average of outcomes to reduce excessive variation in low population areas. The animation was driven by weekly snapshots of data from 2006 to 2017 for vaccination uptake rates and 1996 to 2017 for pre- and post-vaccine health outcome trends. Maps to demonstrate vaccination were built using the Data-Driven Documents JavaScript (D3) visualization library, which renders states and counties as Scalable

Vector Graphics (SVG) or HTML Canvas with dynamic styling and interactivity controlled by JavaScript. Since there are many more zip codes than counties, the display of national 5-digit zip code data on webpages was best suited by the use of HTML canvas graphics and using a compressed data format utilizing the date at which variables change instead of all weekly values. The database and geographic data are served from a PostGIS database via an ExpressJS Application Programming Interface (API). PostGIS is a spatial database extension to the Postgres Structured Query Language (SQL) that supports geographic objects and location queries. This application allows users to view the various visualizations using interfaces such as play buttons to start animations and data probes to reveal additional information about single entities. The final website was optimized for use on desktop computer browsers, tablets and mobile devices. The project is hosted on publicly available servers.

2.10. Privacy considerations

The study involved retrospective review of administrative claims data. Individual records were not provided to the research team and no child's identity or medical records were disclosed for the purposes of this study. Because geographic data was derived from the residential zip codes of children in the database, only the location data with no patient identifiers was provided in the analytic dataset received for analysis.

3. Results

3.1. Database

The analytic file variables required for visualizations were extracted from the main database as described in the methods section. The smallest geographic unit was the 5-digit zip code and the shortest time interval was the week over the 21-year period. The extracted data included over 40,000 zip codes in the continental US. Because the time span for the project ranged from 1996 through 2017, there were over 1,100 weekly time points. The final analysis dataset comprised disease and vaccination data for all available geographic and time datapoints. This resulted in over 44 million rows of data to drive the visualizations for each geographic and weekly interval.

3.2. Visualization maps of vaccination uptake

The visualizations show increasing vaccine uptake rates using national, state, country and zip code level "live" maps for every week from 2006 to 2017. Fig. 1a is a static snapshot of rotavirus vaccine uptake at the zip code level, mid- 2011, which is five years after first vaccine introduction. Fig. 1b is a static image of a visualization showing vaccine uptake at the county level and the inserted map shows that specific and detailed data can be displayed for any county when the user selects the county by using the device pointer. After vaccine introduction, the visualizations show areas, such as Alabama, that have rapid rates of vaccine utilization after licensure. In this paper, the figures are shown as static images. In the figure captions, readers are referred to the relevant website link.

Vaccine uptake was the first variable assessed in the initial phase of this study [16] and was the first use of 5-digit zip code geography to be visualized on the website. However, the main limiting factor of the zip code level data was that the underlying dataset had very small sample sizes for certain zip codes that resulted in unstable continuous rates over time. In addition, the large size of the dataset at high geographic resolution resulted in slow website load times of over 3 seconds, depending on internet connection
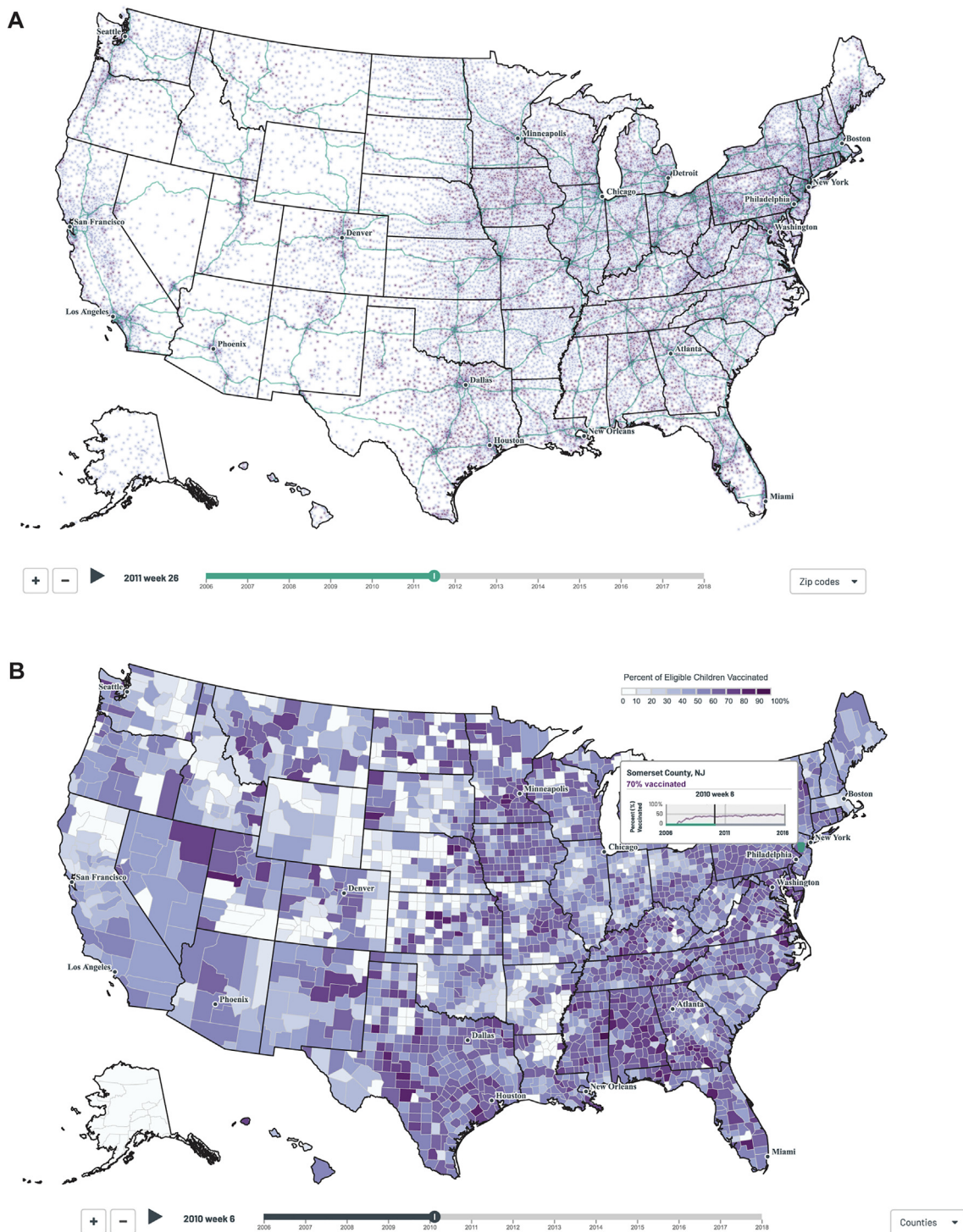
Fig. 1. A. A static image from the rotavirus vaccine uptake animated visualization at the zip code level. In this figure darker purple areas show higher vaccine rates. The timeline at the bottom shows that this point in the visualization was in mid 2011 which is five years after vaccine introduction. [http://www.mapvaccines.com/atlas/vaccination]. B. A static image from the rotavirus vaccine uptake visualization at the county level. In this figure, darker purple areas indicate higher vaccination rates. The timeline at the bottom shows that this point in the visualization was at week 6 of 2010. Also shown is the inserted map of a specific and detailed data probe with a line graph which can be displayed for any county when activated by mouseover.

speed of the user. This was mitigated by expressing vaccination coverage rates as a binary outcome of coverage above or below a threshold of 75% in each zip code. Although this method reduced the ability to assess continuous data of higher and lower coverage levels in each zip code, the overall national vaccination uptake trends remained evident. Subsequent visualizations that demonstrated the impact of vaccination rates on disease were therefore displayed at the national, state and county level.

3.3. Animated heat maps: national disease rates before and after vaccine introduction

Fig. 2a is a temporal gridded heatmap that organizes the data displays by year and month. Rotavirus medical encounter rates are higher from February to April as previously reported [17]. There is a clear decrease in rotavirus-related medical encounters after the rapid uptake of RV5 between 2006 and 2008. The bottom
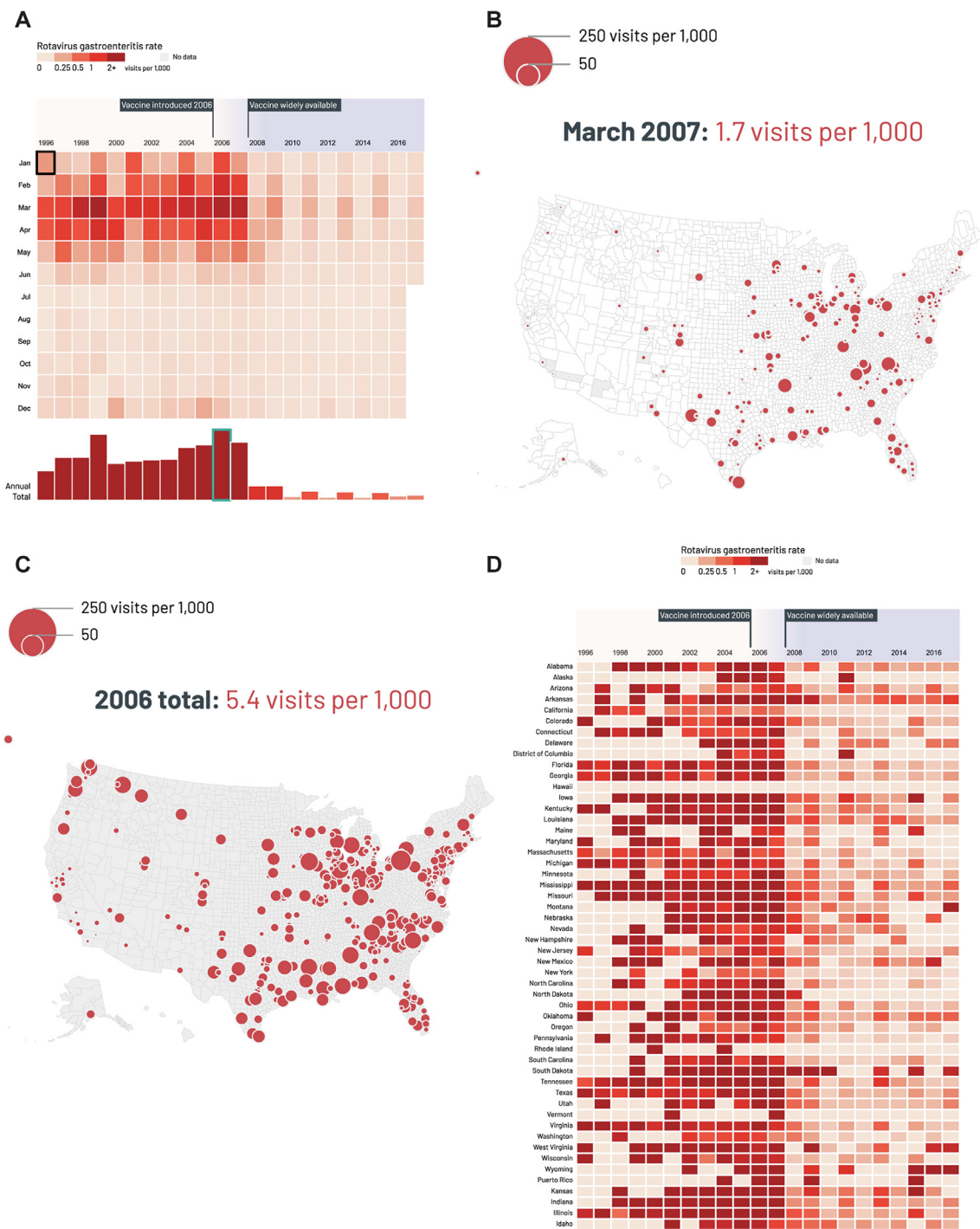
Fig. 2. A. This time-based heatmap is organized by year and month is shown in the box graph [http://www.mapvaccines.com/atlas/heatmap]. From top to bottom, the visualization shows the motion of seasonal trends of increasing rotavirus rates during the epidemic season from January until May (shown by darker red on the left). From left to right, declining rates (light red) after vaccination are over the annual periods. Each box can be clicked on to show data for a specific month and year. The bottom bar graph shows the total national rate by year and highlights the vaccine impact after 2008. B. When clicking on a box in the heatmap, the heatmap for a specific month and year can be displayed for each box as a county proportional symbol map. C. Each bar showing the annual total can be activated. When clicking on the annual bar, the bar is activated to show a national map of disease for each yearly time point. D. A location-based gridded heat map that organizes the data by state and year and employs a static vaccination time-point to demonstrate declining rates of rotavirus-related medical encounters in each state after rotavirus vaccine introduction (see text for details). [http://www.mapvaccines.com/atlas/state_chart]

bar graph shows the total national rate by year and highlights the vaccine impact after 2008. The web site allows a user to mouse-over the grid cell to view the national rotavirus rate for any specific month (Fig. 2b) or for the entire year (Fig. 2c). The website allows the same time-based heatmap at the state level (not shown). Fig. 2d is a location-based gridded heat map, based on a previous publication [13], that organizes the graphic by time and state rather than latitude and longitude to demonstrate declining rates of rotavirus-related medical encounters in each state after rotavirus vaccine introduction. On a mouseover, each grid cell displays temporal patterns of rotavirus rates for each state with a reference for the chosen month.

3.4. Animated line maps: increased vaccination rates and decreased rates of medical encounters by geographic areas and over time

Fig. 3 is a snapshot from the animated line graph of US national data from January 1996 through June 2017. The graph displays the distinct annual epidemic peaks of rotavirus related medical encounters at the national level over a 20-year period. After rotavirus vaccines were introduced in 2006, the motion line graph shows the rapid uptake of RV5 between 2006 and 2008 and then subsequent flattening of seasonal rotavirus disease as vaccine uptake continues to increase after 2008. Fig. 4a shows the same analysis but utilizes a cartogram map of equally sized US states and shows increasing vaccination utilization is associated with declining rotavirus medical encounters over time. Each state can be clicked on to show details for the state; Fig. 4b shows the example of Michigan state. The website also allows exploration at the county level to demonstrate rotavirus disease and vaccination uptake graphs during the entire period 1996–2017. However, the underlying dataset begins to show sparsity in some counties and visual suppression of counties was employed when data were limited.

3.5. Hotspot map to identify areas requiring further interventions

Fig. 5 shows potential geographic hotpots by county in the time period 2015–2017. This period corresponded to overall high rates of vaccination coverage at the national level. The hotspot map highlights counties with both low vaccination rates and the associated higher rates of rotavirus medical encounters at that time. Counties mapped in red have vaccination rates below the national average (70%) and rotavirus-related medical encounters rates 5 times higher than national average (expressed as 1 per 1,000 eligible children). The selection of these cut points was a result of an iterative qualitative process to identify potential hotspots. After

several iterations, a vaccination rate below the national average (70%) and a rotavirus medical encounter rate 5 times higher than the national average (1 per 1,000) was the most interpretable demonstration of this type of visualization. This map was developed as a potential public health tool to allow health officials to locate areas that may need focused attention on vaccination coverage.

3.6. Website of all visualizations

A publicly available digital atlas website was deployed and designed to sequentially walk a user through the visualizations that were developed and described in this article. The homepage is located at http://www.mapvaccines.com.

4. Discussion

This is the first web-based analytic platform to simultaneously visualize the uptake and health impact of a vaccine using detailed geospatial motion-based mapping techniques applied to a national dataset. This visualization showed the rapid vaccination uptake at all levels followed by a reduction in rotavirus gastroenteritis-related medical encounters over time in every state and county where robust data were available. The study also highlighted the practical use of such visualizations to monitor geographic hotspots of rotavirus-related medical encounters in areas where vaccination coverage was suboptimal.

The data visualization of multiple geospatial mapping and non-spatial data graphics allowed the development of a vaccine specific health atlas. The website allows the different visualizations to collectively provide a full picture of vaccine uptake and health impacts using various data perspectives.

The outcomes of this project suggest several future practical applications. First, the framework could be applied to national pre-
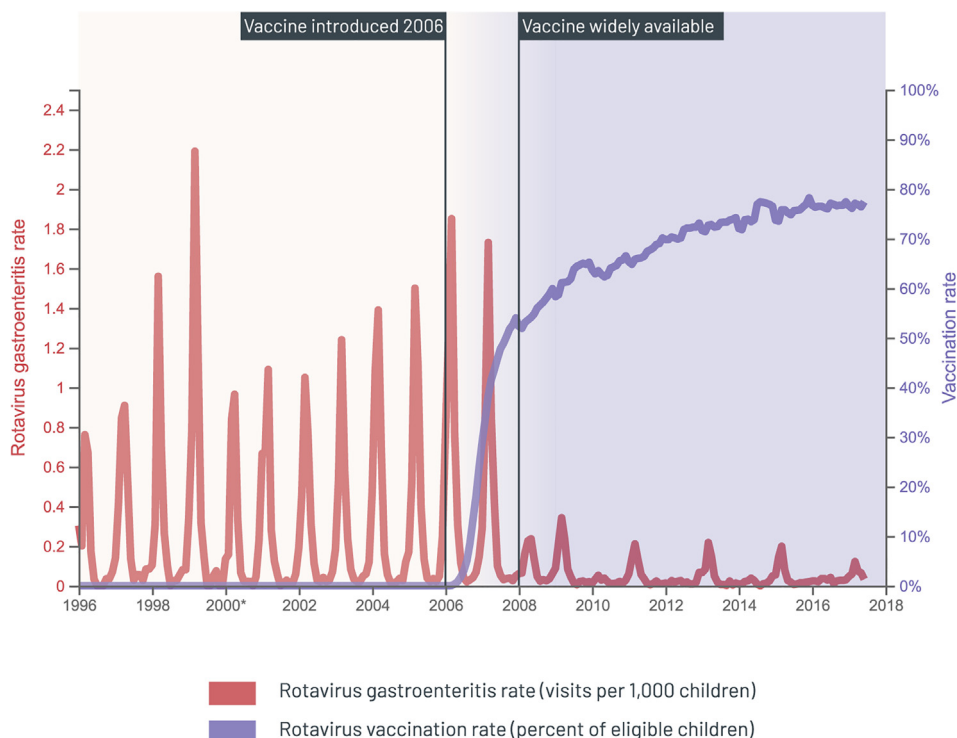

Fig. 3. Example of the animated line graph of US national data from January 1996 through June 2017. The visualization shows the rapid uptake of the vaccine between 2006 and 2008 (blue) and the subsequent dramatic flattening of seasonal rotavirus medical encounters (red) as vaccine uptake continues to increase after 2008. [http://www.mapvaccines.com/atlas/line]
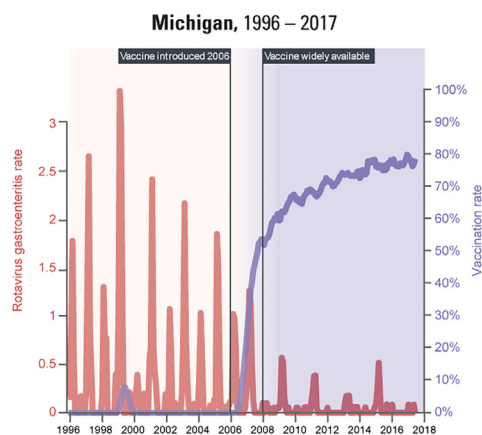
A

B

Fig. 4. A. This static image shows the visualized line cartogram of vaccination uptake and impact for every state in the US. Hovering over a state runs the visualization for that state. Medical encounter rates are charted in red against the vaccination rates in purple. In every state, increasing vaccine uptake over time reduces medical encounters [http://www.mapvaccines.com/atlas/state_multies]. B. This static image shows the visualized line map of vaccination uptake and impact at the state level. This image appears when hovering over a specific state in the map shown in Figure A. Figure B shows the example of Michigan.

paredness planning for monitoring the introduction and health impact of future vaccines, including COVID-19 vaccines. Second, declining coverage in existing programs could be monitored. For example, this tool could be used to monitor unexpected events such as the COVID-19 pandemic disruption of routine vaccination administration [18]. As shown in Fig. 5, our study demonstrated the ability to scan a national database to identify hotspots, i.e.,

locations with low vaccination rates and emergence of vaccine-preventable diseases at the state or county level.

The further development and application of the methods used in this project assumes the existence of well-resourced databases and analytic platforms that include geographically detailed and time-based longitudinal data. Both private and public insurance claims databases and/or electronic medical record (EMR) databases
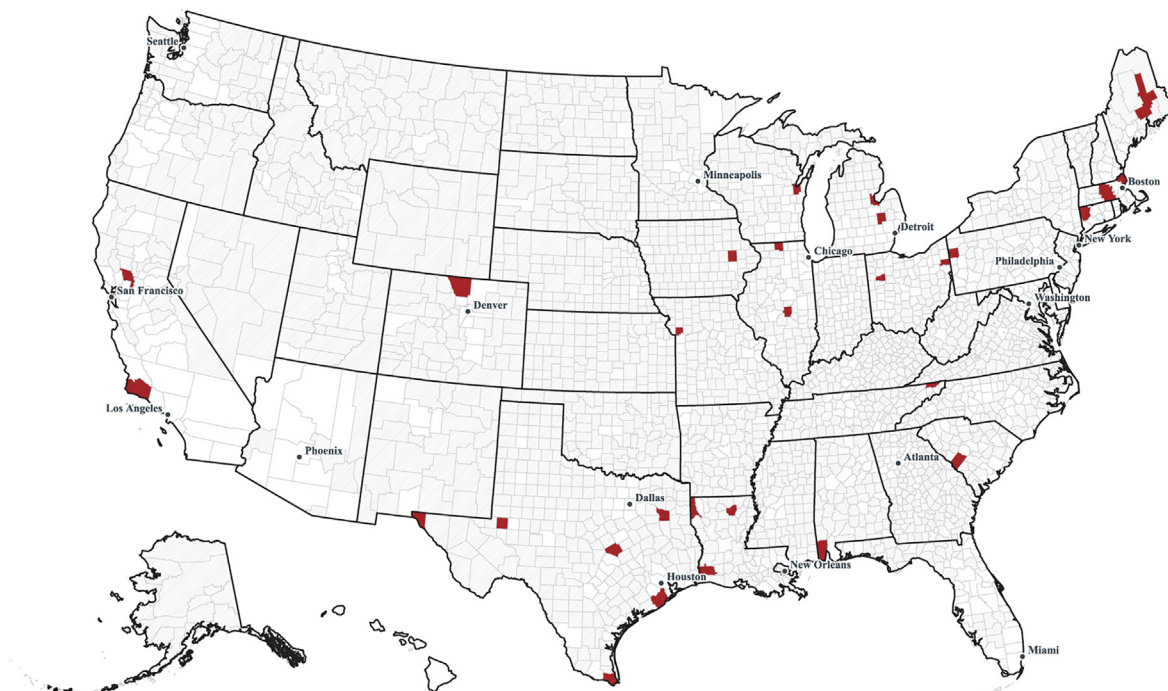
Fig. 5. Potential Hotspots: Counties mapped in dark red have both low vaccination rates and high rotavirus rates in a single month 2015 – 2017 (see text). [http://www.mapvaccines.com/atlas/bivariate]

that can link vaccine exposure and medical outcomes have demonstrated high levels of precision and validity in evaluating vaccine safety and vaccine effectiveness [15,19–24]. Our study identified the need for large and extensively linked databases in order to avoid data sparsity gaps when assessing detailed geographic and time-based data. One challenge is that health data are often not reported with geographic precision primarily due to privacy concerns. Yet, the tension between data privacy concerns and the urgent need to monitor public health [25] is an important policy consideration.

This study suggests some areas of focus for the planning of COVID-19 vaccination programs. Current reporting of COVID-19 cases with daily updates at the national and county level [26] demonstrate that health data can be collected, analyzed and rapidly disseminated [12,27]. Although extremely valuable for the geographic tracking of disease over time, the various currently available case count data maps are constrained by a lack of consistent denominator data. These maps are also not currently designed for direct linkage, at the individual level, to vaccination data. Individual level data are required to monitor the impact of future vaccines on health outcomes. Success of this analytic approach relies on rapid and consistent reporting of vaccine administration data during the pandemic period at the zip code level so that uptake can be monitored. At a minimum, states could plan to capture COVID-19 vaccination in registries with linkage to current case mapping. Such enhancements at the state and county level, along with population-based denominator data, could allow the application of the visualization methods described in this project to support the development of 'situational awareness', which the CDC defines as the timely monitoring of the spread and intensity of COVID-19 disease to allow efficient targeting of public health measures such as vaccination [28]. The CDC suggests that surveillance data can be updated daily or weekly to create an ongoing, accurate understanding of impacted regions, affected populations, trends over time, and viral characteristics. Several states have existing policies to support timely entry of vaccination administration data into registries and therefore it seems reasonable to suggest that the rolling average of vaccination rates could be refreshed at daily or weekly intervals and linked to COVID-19 outcomes for timely visualizations.

Ideally, standardized linkage of vaccination data – accounting for the availability of multiple vaccines - and disease data at the state-level in a common data model [29] would allow geographic comparisons, particularly if both public and privately insured data could be combined. In consideration of the time and effort required to develop and implement these visualizations in a rapid manner for COVID-19, planning for the coordination of various data sources, linkage of vaccination and outcomes, analytic capacity, and detailed geospatial reporting should be initiated as quickly as possible.

5. Limitations

The use of insurance claims data for health care research is associated with certain limitations. The study used a commercial database of a privately insured population and may not reflect characteristics of vaccination uptake among those who are not privately insured. For example, the data used in this study would not include vaccines purchased by the US Vaccines for Children (VFC) program, which currently purchases approximately 50% of all pediatric vaccines in the US. However, there is well-established evidence that rotavirus and gastroenteritis outcomes are significantly decreased in association with increased vaccination rates in privately insured populations [14,15]. Therefore, even without access to publicly insured populations, the study objective of demonstrating an innovative geospatial visualization of health outcomes in association with vaccine uptake was achieved and consistent with previous research. Additionally, limited available data for certain small units of time or geography could misrepresent levels of vaccine uptake or health outcomes at the inspection of such granular data levels. Finally, privacy concerns of protected

health information required data suppression when there were very few cases for a geographic unit. To mitigate these potential biases, rolling averages were used to balance some of the variability and zip code data was not used to assess health outcomes due to data sparsity. Laboratory confirmation of RGE diagnosis was not available for this study, although it has been reported that 98% of patients with an RGE ICD diagnosis code tested positive for rotavirus. In addition, because rotavirus is not routinely tested for in clinical practice, absolute rates of RGE using ICD codes are likely to underestimate the actual rates of RGE; however, rates of RGE identified using ICD-9 codes have been shown to accurately reflect rate trends over time [30].

A methodological limitation is that the visualizations were not based on data that directly linked vaccination to outcomes for specific subjects. For COVID-19, the visualizations of changes in geographic disease trends over time after a vaccine is introduced may be influenced by an ecologic fallacy if there are confounders such as non-vaccine interventions, local health policies or individual risk factors that affect disease patterns independent of local vaccination rates. Other factors that may vary geographically include vaccine hesitancy, vaccine supply availability and vaccine administration logistics. As previously discussed, developing large-linked datasets with geographic information would ameliorate this potential challenge when using visualizations techniques.

Despite these limitations, this study was able to successfully implement a visualized analysis of vaccine impact using a national medical claims database. The resulting visualization patterns were consistent with previously published studies that indicated a sustained and substantial decrease in medical encounters in association with vaccination. When assessing the introduction of new COVID-19 vaccines, no single analysis or data source will provide the complete picture of the health system in the US and thus multiple approaches will be required.

6. Conclusion

This study developed novel web-based methods to visualize and communicate location and time-based vaccine uptake and the related reduction in medical encounters. With a view towards public health preparedness, this approach could be proactively developed to monitor the utilization and impact of new vaccines developed to address novel emerging epidemics caused by infectious agents such as SARS-CoV-2.

CRediT authorship contribution statement

T. Christopher Mast: Conceptualization, Supervision, Data curation, Methodology, Visualization, Writing - original draft, Writing - review & editing. **David Heyman:** Methodology, Visualization, Data curation, Software, Writing - review & editing. **Erik Dasbach:** Writing - review & editing. **Craig Roberts:** Writing - review & editing. **Michelle G. Goveia:** Visualization, Writing - review & editing. **Lyn Finelli:** Writing - review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: TC Mast, E Dasbach, C Roberts, M Goveia and L Finelli are current employees of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA, and may hold equity interest in Merck & Co., Inc., Kenilworth, NJ, USA. D Heyman led the data visualization programming and was compensated for activities related to execution of the study.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jvacx.2021.100084.

References

[1] Adis International Ltd. Adis R&D insight database. Adis Int Springer Sci 2020; 2020. http://adisinsight.springer.com/ [accessed August 14, 2020].
[2] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020;323:1061–9. https://doi.org/10.1001/jama.2020.1585.
[3] World Health Organization. Coronavirus disease (COVID-19) Pandemic. https://www.who.int/emergencies/diseases/novel-coronavirus-2019. [accessed January 13, 2021].
[4] Johns Hopkins University. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html [accessed January 13, 2021].
[5] Centers for Disease Control and Prevention (CDC). CDC COVID Data Tracker https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html [accessed January 13, 2021].
[6] The Atlantic Monthly Group. The COVID tracking project (Historical Data). https://www.covidtracking.com/data/national [accessed January 13, 2021].
[7] US Centers for Disease Control and Prevention. COVID-19 Forecasts: hospitalizations. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/hospitalizations-forecasts.html [accessed January 13, 2021].
[8] CDC. Interim pre-pandemic planning guidance: community strategy for pandemic influenza mitigation in the United States – Early, Targeted, Layered Use of Nonpharmaceutical Interventions; 2007, available at https://www.cdc.gov/flu/pandemic-resources/pdf/community_mitigation-sm.pdf. [accessed June 3, 2020].
[9] Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020;395:1973–87. https://doi.org/10.1016/S0140-6736(20)31142-9.
[10] Collins FS, Stoffels P. Accelerating COVID-19 therapeutic interventions and vaccines (ACTIV): an unprecedented partnership for unprecedented times. JAMA – J Am Med Assoc 2020;323:2455–7. https://doi.org/10.1001/jama.2020.8920.
[11] Joyce K. "To me it's just another tool to help understand the evidence": public health decision-makers' perceptions of the value of geographical information systems (GIS). Health Place 2009;15:831–40. https://doi.org/10.1016/j.healthplace.2009.01.004.
[12] The New York Times. Coronavirus in the U.S.: latest map and case count. https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html [accessed November 20, 2020].
[13] DeBold T, Friedman D. Battling infectious diseases in the 20th century: the impact of vaccines – WSJ.com. Wall Str J 2015:1–7 [accessed June 3, 2020].
[14] Payne DC, Selvarangan R, Azimi PH, Boom JA, Englund JA, Staat MA, et al. Long-term consistency in rotavirus vaccine protection: RV5 and RV1 vaccine effectiveness in US children, 2012–2013. Clin Infect Dis 2015;61:1792–9. https://doi.org/10.1093/cid/civ872.
[15] Wang FT, Mast TC, Glass RJ, Loughlin J, Seeger JD. Effectiveness of the pentavalent rotavirus vaccine in preventing gastroenteritis in the United States. Pediatrics 2010;125. https://doi.org/10.1542/peds.2009-1246.
[16] Mast TC. Geospatial Visualization of New Vaccine Adoption in the US, 2006–2015, Abstract 403, 33rd International Conference on Pharmacoepidemiology & Therapeutic Risk Management, Montreal, Canada. August 26-30, 2017. Pharmacoepidemiol Drug Safety 2017;26(S2). https://doi.org/10.1002/pds.4274.
[17] Mast TC, Walter EB, Bulotsky M, Khawaja SS, Distefano DJ, Sandquist MK, et al. Burden of childhood rotavirus disease on health systems in the united states: results from active surveillance before rotavirus vaccine introduction. Pediatr Infect Dis J 2010;29. https://doi.org/10.1097/INF.0b013e3181ca7e2e.

[18] Bramer CA, Kimmins LM, Swanson R, Kuo J, Vranesich P, Jacques-Carroll LA, et al. Decline in child vaccination coverage during the COVID-19 pandemic — Michigan care improvement registry, May 2016–May 2020. MMWR Morb Mortal Wkly Rep 2020;69:630–1. , https://doi.org/10.15585/mmwr.mm6920e1.

[19] Loughlin J, Mast TC, Doherty MC, Wang FT, Wong J, Seeger JD. Postmarketing evaluation of the short-term safety of the pentavalent rotavirus vaccine. Pediatr Infect Dis J 2012;31:292–6. https://doi.org/10.1097/INF.0b013e3182421390.

[20] Klein NP, Hansen J, Chao C, Velicer C, Emery M, Slezak J, et al. Safety of quadrivalent human papillomavirus vaccine administered routinely to females. Arch Pediatr Adolesc Med 2012;166:1140–8. https://doi.org/10.1001/archpediatrics.2012.1451.

[21] Davis RL, Black S, Shinefield H, Mahoney L, Zavitkovsky A, Lewis E, et al. Post-marketing evaluation of the short term safety of COMVAX®. Vaccine 2004;22:536–43. https://doi.org/10.1016/j.vaccine.2003.06.001.

[22] Baxter R, Ray P, Tran TN, Black S, Shinefield HR, Coplan PM, et al. Long-term effectiveness of varicella vaccine: a 14-year, prospective cohort study. Pediatrics 2013;131:e1389–96. https://doi.org/10.1542/peds.2012-3303.

[23] Nobel T, Rajupet S, Sigel K, Oliver K. Using Veterans Affairs Medical Center (VAMC) data to identify missed opportunities for HPV vaccination. Hum Vaccines Immunother 2019;15:1878–83. https://doi.org/10.1080/21645515.2018.1559684.

[24] Langan SM, Smeeth L, Margolis DJ, Thomas SL. Herpes zoster vaccine effectiveness against incident herpes zoster and post-herpetic neuralgia in an older US population: a cohort study. PLoS Med 2013;10. https://doi.org/10.1371/journal.pmed.1001420.

[25] Song X, Waitman LR, Hu Y, Luo B, Li F, Liu M. The impact of medical big data anonymization on early acute kidney injury risk prediction. AMIA Jt Summits Transl Sci Proceedings AMIA Jt Summits Transl Sci 2020;2020:617–25.

[26] North Carolina Department of Health and Human Services. COVID-19 North Carolina Dashboard. https://covid19.ncdhhs.gov/dashboard [accessed June 4, 2020].

[27] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20:533–4. https://doi.org/10.1016/S1473-3099(20)30120-1.

[28] US Department of Health and Human Services. CDC activities and initiatives supporting the COVID-19 response and the president's plan for opening America up again; May 2020. https://www.cdc.gov/coronavirus/2019-ncov/downloads/php/CDC-Activities-Initiatives-for-COVID-19-Response.pdf.

[29] Han A, Isaacson A, Muennig P. The promise of big data for precision population health management in the US. Public Health 2020;185:110–6. https://doi.org/10.1016/j.puhe.2020.04.040.

[30] Mast TC, Wang FT, Su S, Seeger JD. Evidence of herd immunity and sustained impact of rotavirus vaccination on the reduction of rotavirus-related medical encounters among infants from 2006 through 2011 in the United States. Pediatr Infect Dis J 2015;34:615–20. https://doi.org/10.1097/INF.0000000000000702.