

Original Paper

# Improving Detection of Disease Re-emergence Using a Web-Based Tool (RED Alert): Design and Case Analysis Study

Nidhi Parikh<sup>1</sup>, PhD; Ashlynn R Daughton<sup>1</sup>, PhD; William Earl Rosenberger<sup>1</sup>, BS; Derek Jacob Aberle<sup>2</sup>, BSc; Maneesha Elizabeth Chitanvis<sup>3</sup>, MPH; Forest Michael Altherr<sup>4</sup>, MPH; Nileena Velappan<sup>5</sup>, MS; Geoffrey Fairchild<sup>1</sup>, PhD; Alina Deshpande<sup>5</sup>, PhD

<sup>1</sup>Information Systems and Modeling Group, Los Alamos National Laboratory, Los Alamos, NM, United States

<sup>2</sup>Neutron Science and Technology Group, Los Alamos National Laboratory, Los Alamos, NM, United States

<sup>3</sup>Weill Cornell Medical, Cornell University, New York, NY, United States

<sup>4</sup>BioFire Diagnostics, LLC, Salt Lake City, UT, United States

<sup>5</sup>Biosecurity and Public Health Group, Los Alamos National Laboratory, Los Alamos, NM, United States

**Corresponding Author:**

Alina Deshpande, PhD

Biosecurity and Public Health Group

Los Alamos National Laboratory

TA-43, Building 1, MS - M888

Los Alamos, NM, 87545

United States

Phone: 1 505 667 9938

Email: [deshpande\\_a@lanl.gov](mailto:deshpande_a@lanl.gov)

## Abstract

**Background:** Currently, the identification of infectious disease re-emergence is performed without describing specific quantitative criteria that can be used to identify re-emergence events consistently. This practice may lead to ineffective mitigation. In addition, identification of factors contributing to local disease re-emergence and assessment of global disease re-emergence require access to data about disease incidence and a large number of factors at the local level for the entire world. This paper presents Re-emerging Disease Alert (RED Alert), a web-based tool designed to help public health officials detect and understand infectious disease re-emergence.

**Objective:** Our objective is to bring together a variety of disease-related data and analytics needed to help public health analysts answer the following 3 primary questions for detecting and understanding disease re-emergence: Is there a potential disease re-emergence at the local (country) level? What are the potential contributing factors for this re-emergence? Is there a potential for global re-emergence?

**Methods:** We collected and cleaned disease-related data (eg, case counts, vaccination rates, and indicators related to disease transmission) from several data sources including the World Health Organization (WHO), Pan American Health Organization (PAHO), World Bank, and Gideon. We combined these data with machine learning and visual analytics into a tool called RED Alert to detect re-emergence for the following 4 diseases: measles, cholera, dengue, and yellow fever. We evaluated the performance of the machine learning models for re-emergence detection and reviewed the output of the tool through a number of case studies.

**Results:** Our supervised learning models were able to identify 82%-90% of the local re-emergence events, although with 18%-31% (except 46% for dengue) false positives. This is consistent with our goal of identifying all possible re-emergences while allowing some false positives. The review of the web-based tool through case studies showed that local re-emergence detection was possible and that the tool provided actionable information about potential factors contributing to the local disease re-emergence and trends in global disease re-emergence.

**Conclusions:** To the best of our knowledge, this is the first tool that focuses specifically on disease re-emergence and addresses the important challenges mentioned above.

(*JMIR Public Health Surveill* 2021;7(1):e24132) doi: [10.2196/24132](https://doi.org/10.2196/24132)

**KEYWORDS**

disease re-emergence; infectious disease; supervised learning; random forest; visual analytics; surveillance

**Introduction**

Infectious diseases remain a leading cause of death, contributing to millions of deaths each year [1]. The current COVID-19 pandemic demonstrates the speed with which an infectious disease can travel from one location to another including new locations, and in turn become a global health threat in today's world of increased travel and globalization. COVID-19 is an infectious disease caused by a newly discovered coronavirus called SARS-CoV-2. In addition to such newly emerging diseases, some diseases that were considered controlled or eliminated are also re-emerging. The past few decades have seen the re-emergence of dengue in Brazil [2], measles in France [3], and yellow fever in Angola [4]. A re-emerging infectious disease is a disease that was a major health problem historically in a location, saw a persistent decline in its incidence, and then saw its incidence increase again. Many factors such as ecological disruptions, changing environment, urbanization and human behaviors, international travel and commerce, and war and civil unrest contribute to the re-emergence of infectious diseases [2,5-8].

Early detection and understanding of disease re-emergence is important for better response and mitigation of these events. However, there are several challenges: The definition of disease re-emergence merely suggests an up-down-up incidence pattern and does not offer any guidance on quantitative measures by which such patterns can consistently identify re-emergence. The current practice of identifying disease re-emergence relies on the knowledge and experience of public health analysts rather than specific criteria, which can lead to inconsistent identification of re-emergence [9]. While high-level factors (such as those mentioned above) contribute to the re-emergence of infectious diseases, it is difficult to identify specific factors contributing to a local disease re-emergence and requires a systematic analysis of a number of factors. Local public health analysts may not have this kind of information readily available. Currently, the recognition and understanding of global disease re-emergence relies on analysis of data about historical outbreaks at the country level around the world [10-13]. Again, such data may not be easily available for the entire world and even if available, retrospective analysis is a time-consuming process. Better methods and data are thus essential to address this challenge.

In the last few years, a number of web-based analytics, tools, and databases have been developed to collect data from multiple sources to monitor disease-related activities [14-16], provide situational awareness [17], or now-cast infectious diseases [18]. While there are currently no tools focused on detecting re-emergence, this presents an opportunity for developing new analytics.

Machine learning algorithms use observation data to identify trends and patterns that can help make better decisions. Supervised algorithms identify patterns from the data that are useful in predicting specific outcomes while unsupervised

algorithms extract trends and patterns from the data without relating them to any outcomes. Both supervised and unsupervised methods are used extensively in public health. Unsupervised machine learning is used to understand spatial dynamics of an epidemic [19], extract meaningful structure in electronic health records [20], and identify subgroups among home health patients with heart failure [21]. Supervised machine learning is used for disease forecasting [22,23], mortality risk score prediction in an elderly population [24], predicting blood pressure based on health behaviors [25], and assessing vaccination sentiments [25,26]. Recently, our team developed supervised machine learning models to detect potential infectious disease re-emergence for 4 infectious diseases: measles, cholera, yellow fever, and dengue [9]. Combining such an algorithm with visual analytics could provide a rapid, easy to use, and easy to interpret tool for detecting potential re-emergence.

Visual analysis is a technique that utilizes interactive visualizations to support analytical reasoning [27]. It can help with investigative analysis and hypothesis generation [28] and is especially useful for analyzing large data sets by reducing the load on working memory, offering cognitive support, and utilizing the power of human perception [29]. Recently, visual analytics are increasingly used to analyze data in public health and health care, including human emergency room and veterinary hospital data [30]; relationships between chronic conditions, demographics, behavioral and mental health, preventative health, overarching conditions [31]; and tracking symptom evolution during disease progression [32]. We have also developed a web-based visual analytic for the investigation of infectious disease outbreaks [17].

This paper details Re-emerging Infectious Disease Alert (RED Alert), a web-based tool [33] that integrates our supervised machine learning models [9] with visual analytics to help detect/warn and understand potential re-emergence at both local and global levels for 4 diseases: measles, cholera, dengue, and yellow fever. The diseases were selected in consultation with subject matter experts (SMEs) at the World Health Organization (WHO) as diseases of concern for re-emergence. These diseases also show diversity in transmission and disease burden, allowing us to show transferability of our approach. RED Alert combines disease-related data and analytics needed to help the public health community answer the following questions for detecting and understanding disease re-emergence: Is there a potential disease re-emergence at the local (country) level? What are the potential contributing factors for this re-emergence? Is there a potential for global re-emergence?

This publication describes the methods used to answer these questions and evaluation of machine learning classifiers to detect disease re-emergence and the tool through case studies.

## Methods

### Data

Historical case count data, together with disease subcategories such as severe dengue and deaths, were obtained from the WHO [34-36], Gideon [37], and the Pan American Health Organization (PAHO) [38]. Population data were obtained from 2 data sets: LandScan [39] and the World Bank population data [40]. Rates for measles-containing vaccine first dose and second dose were obtained from the WHO [41] together with the WHO region membership information for each country [42]. The host, pathogen, and environment represent the traditional epidemiological triad [43] and can provide information about the potential causes of re-emergence. For indicators that can be a proxy for re-emergence causes, public health indicator data were obtained from the World Bank [44] using their application programming interface (API) [45]. Detailed information about these data sources can be found in [Multimedia Appendix 2](#).

### Development of RED Alert

RED Alert was developed for application to 4 primary diseases of concern: cholera, measles, dengue, and yellow fever. The visual analytic was developed to have a web application as a front end to the data and analysis. A web API was developed to be used by any program to access the analysis results and underlying data. The back end was developed as a Django-based application. The front end uses JavaScript to read from these API endpoints and dynamically build the corresponding visualizations.

### Detection of Potential Disease Re-emergence

We integrated previously developed supervised machine learning classifiers to detect potential disease re-emergence for a given location and year [9] into RED Alert. Classifiers are supervised learning algorithms that use a set of labeled data (known observation–class pairs, eg, samples of re-emergence and non-re-emergence events [or outbreaks]) and extract patterns that help predict class (eg, re-emergence or not). These patterns can then be used to map a new observation (eg, outbreak) to a class (eg, re-emergence or not re-emergence).

We used yearly disease data at the country level to train disease-specific classifiers for the 4 diseases: measles, cholera, dengue, and yellow fever. For creating the labeled data set for each disease, the SMEs in our team were given data for 100 countries selected at random (and anonymized), and they labeled each location–year pair as a re-emergence or not. A systematic approach was followed to label the training data. For each disease, SMEs developed a re-emergence schema described in detail by Chitanvis et al [9] that takes into account general disease incidence and trend information (eg, raw incidence, case counts, change in incidence from last few years, or percentile rank) and relevant disease-specific information (eg, vaccination coverage for measles and information on severe dengue cases and death due to dengue) that can help detect potential re-emergence. These factors were organized in a decision tree format to guide the labeling process.

### Selection of the Classifier

We compared 2 classifiers, decision tree and random forest, using scikit-learn, a free machine learning Python library [46]. See tables 1a-b in [9] for features used for training the classifiers and imputation methods for missing data. For both decision tree and random forest, we explored the following parameter values: (1) Split criteria: gini and entropy; (2) The number of minimum samples required at leaf nodes: 1 to 10; and (3) The number of trees for random forest: 20 to 100.

Precision, recall, and  $F_1$  are widely used metrics to evaluate the performance of classification and can be calculated as follows:

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

$$F_1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

As our goal was to identify all potential cases of disease re-emergence while allowing some false positives, we used  $F_2$  to evaluate the performance of the classifiers.  $F_2$  takes into account both precision and recall but recall is given more weightage. It can be calculated as follows:

$$F_2 = 5 \times (\text{Precision} \times \text{Recall}) / ([4 \times \text{Precision}] + \text{Recall})$$

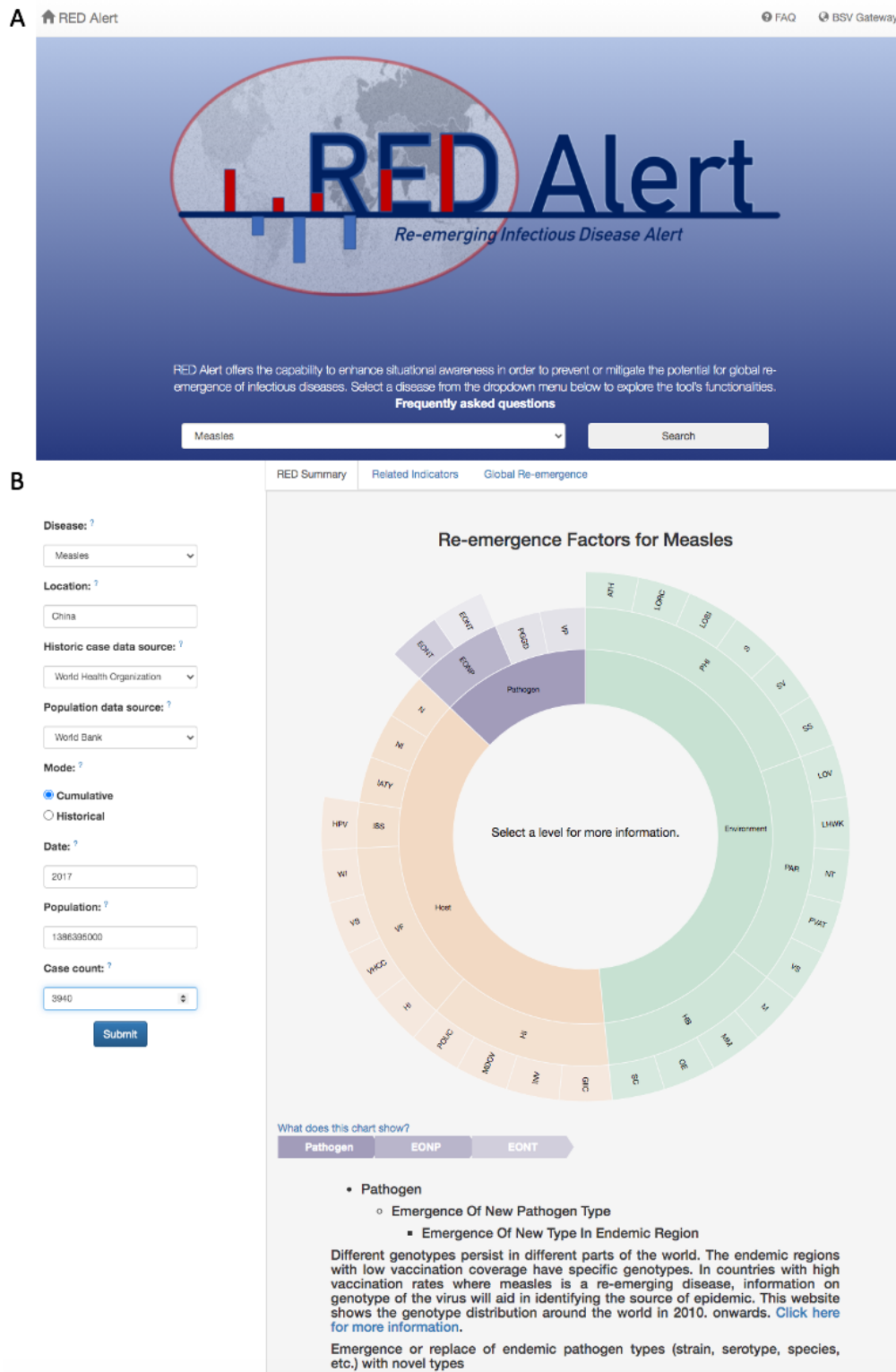
We evaluated classifiers on held-out test data using nested cross-validation [47], where the inner cross-validation is used to choose the optimal parameters, and the outer cross-validation is used to evaluate the performance of the model with the optimal parameters on a held-out data set to test for overfitting or generalization error. Overfitting occurs when the model learns the structure of the given data set instead of the underlying data-generating phenomenon, so it performs well on the given training data set but fails to perform well on additional data or new observations. We used leave-one-out or 1000-fold cross-validation (whichever is lower) for the inner cross-validation and 10-fold cross-validation for the outer cross-validation.

### Identifying Potential Contributing Factors for Re-emergence

We developed a re-emergence causal wheel for each disease in RED Alert; an example can be seen in [Figure 1B](#). The causal wheel was modeled on the epidemiological triad [43]: host, pathogen, and environment. In the causal wheel, these categories were further divided into subcategories based on disease-specific factors that contribute to re-emergence identified from the literature. We thus created multiple rings around the primary inner ring of the epidemiological triad in our visual display for this information in RED Alert. For example, for cholera, the broad category of environment was divided into socioeconomic and natural factors affecting the environment which included natural environment, population density, public health infrastructure, and human behavior. The natural environment was further divided into weather patterns, climate change, and natural disasters. Natural disasters were further divided into floods, typhoon/hurricane, earthquakes, and drought. This causal wheel is displayed on the web application when a user selects

a disease of interest, providing general information about the component causes of re-emergence for the disease. We also added links to detailed information about a component cause to facilitate access.

**Figure 1.** RED Alert’s input form and causal wheel. Panel A shows RED Alert’s first tab, which provides descriptions about the application’s core functionalities and allows user to select the disease of interest. Panel B shows the causal wheel for measles. It includes factors known to have contribute to resurgence of measles in previous scenarios. Factors at the center of circle represent components of epidemiological triad and expanding distance from the center of the circle correspond to increased specificity of factors contributing to component causes. When the user clicks on a component cause, detailed information about the component cause is provided at the bottom of the chart along with the associated references.



These component causes were mapped to 1 or more indicator variables (obtained from the World Bank), which served as the proxy measurement for the corresponding component cause.

Assessment of these component causes and their interactions can help guide effective intervention strategies. We developed a table for visualization of disease-specific indicators that allows

comparison of the values for the user's location and year of interest to the historical range, so that the user can determine which re-emergence cause and indicator might be contributing to country's potential re-emergence.

Component causes and corresponding related indicators for a given disease and location from 2000 to the year of interest are shown in the table. If data were not available for the year of interest, the indicator value for the most recent year when the data are available was displayed. We also identify indicators where the values for the year of interest are outside the 25th and 75th percentile or 10th and 90th percentile, as these indicators show relatively extreme values for the re-emergence year as compared to the historical values for the location of interest and hence may be potential contributing factors for the disease re-emergence. These indicators and components are displayed to the user in a form of table along with the associated value for the year of interest and statistics for historically observed values (eg, median and 25th and 75th percentiles). Indicators with values outside the 10th and 90th percentiles are highlighted in dark red or dark blue colors if they are potential risk or protective factors, respectively. Similarly, indicators with less extreme values (ie, values outside the 25th and 75th percentiles) are highlighted in light red or light blue colors if they are potential risk or protective factors, respectively.

### ***Understanding Potential for Global Disease Re-emergence***

To help identify the potential for global re-emergence, we developed a visual summary in the form of a map showing the spatial distribution of national re-emergence events (identified through the machine learning classifier described above) worldwide within recent history (last 10 years). The map is time enabled, allowing the user to scroll dynamically through the last 10 years of historic data. Re-emergence events in the year selected by the slider are colored in red, whereas re-emergence events identified in previous years are identified by black points. The size of the points represents the number of historic re-emergence events in the last 10 years. Multiple re-emergence events across different countries or continents may suggest potential for global re-emergence and require further investigation by the user.

### ***Additional Visual Analytics***

In addition to features developed to answer the 3 primary objectives described above, we developed visual analytics that could help deepen the understanding of potential contributing factors for the re-emergence and global re-emergence assessment. RED Alert visual analytics were developed to illustrate the relationship between potential contributing factors (eg, sanitation facilities, urbanization, or vaccinated percentage of the population) and disease re-emergence. We also developed visual analytics to compare locations with similar disease incidence (ie, locations with incidence within 50%-150% of user-specified data). These additional analytics were provided in a second tab of the RED Alert output.

To help assess the global disease re-emergence situation, we organized different types of global data in a third tab of RED Alert. This includes information about disease incidence globally

for the last 10 years from the year of interest input by a user and recent reports of disease occurrence on FluTrackers [48], an online disease community bulletin board. We also provided the following questions on this tab that guide users through the data and facilitate hypothesis generation:

- Are the highest 2 quantiles of disease incidence dispersed over multiple continents?
- Has disease incidence intensified, across geographic areas, over time?
- Are the most recent FluTrackers community posts dispersed over multiple continents?

### **Evaluating RED Alert Through Case Studies**

To evaluate the performance of the fully developed RED Alert analytic, we used case studies for each of the 4 diseases (measles, dengue, cholera, and yellow fever). Specific inputs were identified based on the outbreak selected, and we evaluated the output with respect to its utility in addressing the 3 main objectives that the visual analytic was developed for: (1) Can we identify potential disease re-emergence in a country? (2) What might be the contributing factors to re-emergence in that location? and (3) Are there indications of a global re-emergence based on the input situation? Using the same case studies, we also evaluated the utility of the additional visual analytics that guide hypothesis generation and provide actionable information to the user. We identified the scope of use and the type of actionable information that can be obtained from RED Alert by defining specific work roles to also understand the broad utility and diversity of information that can be used.

Case studies were selected from the 2015 to 2019 timeframe to best illustrate every feature of the analytic. One of the primary challenges is the availability of updated global data. As RED Alert is dependent upon the updating cycle of data sources used (World Bank and WHO), it is often difficult to examine all the features using the current year. Complete, global data sets for public health indicators and infectious disease case counts are currently available up to 2017 or 2018. However, we believe this is still a reasonable representation of situations that occurred in 2019/2020 and the near future of about 5 years, as the natural and built environments are not expected to significantly change in such a short timeframe.

## **Results**

### **Detecting Potential Disease Re-emergence**

We selected random forest as the classifier to integrate into RED Alert because it outperformed the decision tree classifier in terms of the  $F_2$  score for the re-emergence class for all diseases. Table 1 shows the performance of random forest classifiers in terms of average and SD of precision, recall,  $F_1$ , and  $F_2$  measures over 10 nested cross-validations. For the specific diseases in RED Alert, the models were able to identify 82%-90% of all potential re-emergence events as potential re-emergence cases. Of all instances classified as potential re-emergence, about 19% to 31% (except 46% for dengue) were false positives. Our models identified most of the country-level re-emergence events identified in the literature while missing a few events that were restricted to smaller geographic areas

and did not contribute enough disease cases to affect disease incidence at the country level. In some cases, our models also identified earlier disease re-emergence events as compared to

the literature, underscoring the utility of our models for early detection and warning.

**Table 1.** Random forest performance over 10 nested cross-validation.<sup>a</sup>

| Measure and class    | Measles<br>Mean (SD) | Cholera<br>Mean (SD) | Dengue<br>Mean (SD) | Yellow fever<br>Mean (SD) |
|----------------------|----------------------|----------------------|---------------------|---------------------------|
| <b>Precision</b>     |                      |                      |                     |                           |
| RED                  | 0.7100 (0.1015)      | 0.8100 (0.1197)      | 0.5411 (0.0436)     | 0.6914 (0.1270)           |
| Not RED              | 0.9925 (0.0057)      | 0.9913 (0.0063)      | 0.9883 (0.0040)     | 0.9964 (0.0036)           |
| <b>Recall</b>        |                      |                      |                     |                           |
| RED                  | 0.9064 (0.0736)      | 0.8236 (0.1267)      | 0.8421 (0.0554)     | 0.8856 (0.1130)           |
| Not RED              | 0.9689 (0.0147)      | 0.9889 (0.0098)      | 0.9480 (0.0117)     | 0.9857 (0.0087)           |
| <b>F<sub>1</sub></b> |                      |                      |                     |                           |
| RED                  | 0.7909 (0.0688)      | 0.8051 (0.0814)      | 0.6567 (0.0439)     | 0.7631 (0.0752)           |
| Not RED              | 0.9805 (0.0075)      | 0.9901 (0.0046)      | 0.9677 (0.0060)     | 0.9910 (0.0037)           |
| <b>F<sub>2</sub></b> |                      |                      |                     |                           |
| RED                  | 0.8546 (0.0601)      | 0.8129 (0.0971)      | 0.7557 (0.0425)     | 0.8278 (0.0781)           |
| Not RED              | 0.9735 (0.0118)      | 0.9893 (0.0074)      | 0.9558 (0.0094)     | 0.9878 (0.0066)           |

<sup>a</sup>RED and not RED represent re-emergence and non-re-emergence classes, respectively.

### Evaluation of RED Alert Through Case Studies

RED Alert features 2 primary modes for users to engage with the application: cumulative and historical analysis. The modes depend on the user's access to data and the user's willingness to upload data into the application. It is important to note that any data the user inputs in the form is not stored by the application at any point. The lowest burden mode to the user is the historical mode. This mode displays all historic data as calculated incidence for the user's defined location. The cumulative mode is of moderate complexity and is the most frequently utilized option in RED Alert. This mode requires that the users know the year they are interested in analyzing as well as the corresponding case counts. For each disease, the analytic provides the most appropriate data source depending on the location. A user selects the cumulative mode if he/she intends to utilize the tool to explore how the data relate to the historic collection of case counts. We describe the results of using RED Alert for a case study for measles. We describe additional case studies for yellow fever, cholera, and dengue in [Multimedia Appendix 1](#). The tool is very rich in information and data, and wherever possible, we have tried to evaluate how the different facets of the analytic could support different types of analysis.

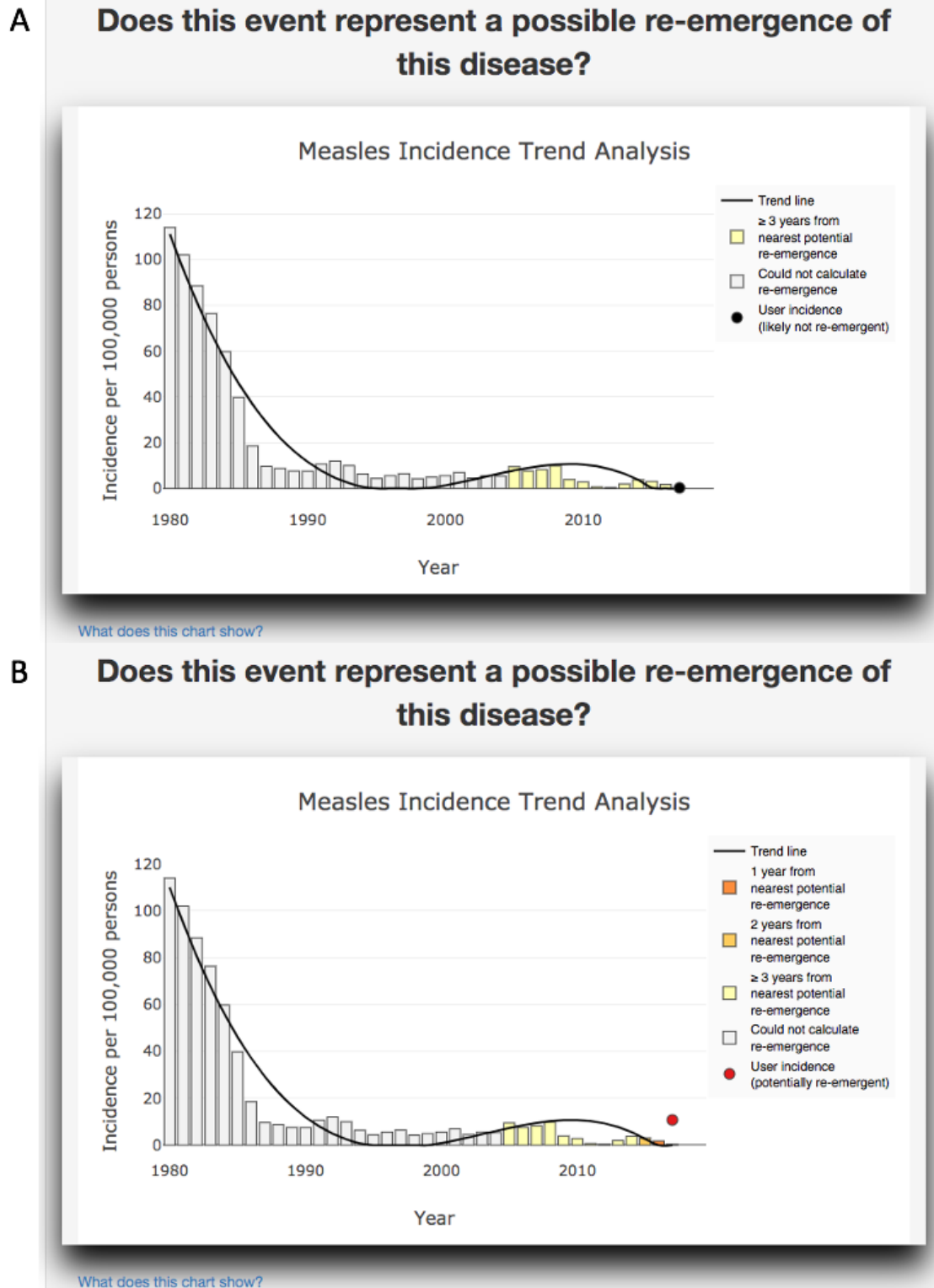
For the measles case study, we specified a public health analyst as the work role and identified the following task for the analyst: Determine the historical profile of measles in China over the

past several decades to review the natural temporal fluctuations in measles, and determine if the reported case count for China in 2017 is indicative of a re-emergence. Following the selection of measles from the drop-down menu on the first tab ([Figure 1A](#)), the first image seen was a sunburst chart ([Figure 1B](#)) that provided the user information on the various causes of re-emergence of measles. The causes were broadly categorized into host, pathogen, and environmental causes, and the user could obtain further detailed information for each of these causes. For example, one of the pathogen-specific factors leading to re-emergence is a new measles type introduced into an endemic country. The following case study inputs were used to generate answers to the 3 main questions used for evaluating the tool: Location—"China," population data source—Default (World Bank), mode—cumulative, year of interest—2017, number of cases—3940.

The output was seen in a tabbed format, with the first tab "RED summary" showing the answers to the 3 primary questions:

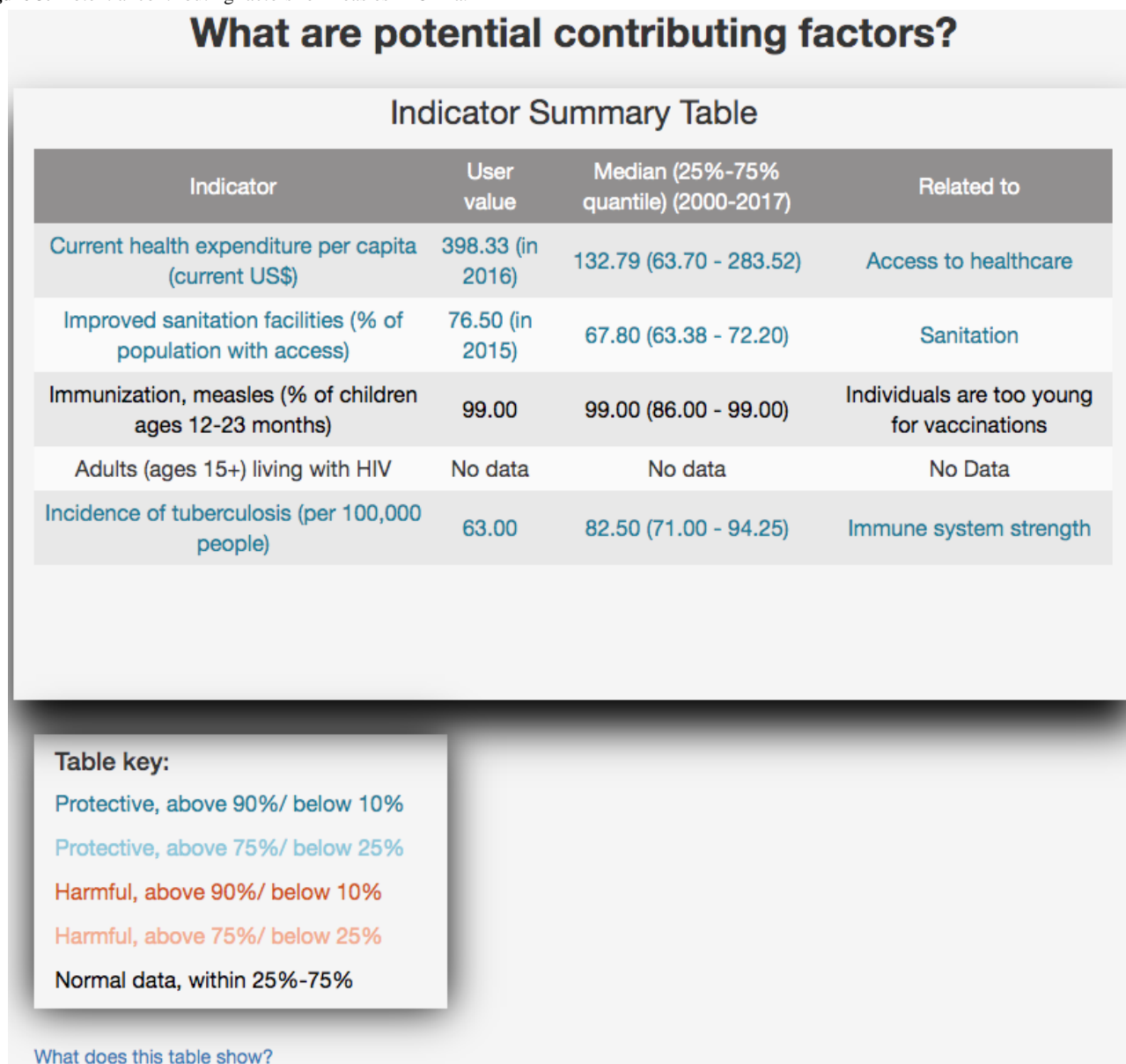
Q1, "Does this event represent a possible re-emergence of this disease?": The time series ([Figure 2A](#)) showed a dip in incidence in 2011 and 2012 followed by a slight rise in cases in 2013 and 2014 and a steady decrease since then. The legend on the chart also indicated that the input data did not reflect a potential re-emergence. When the case count was changed to 150,000, the chart did change ([Figure 2B](#)) and the legend on the chart indicated a potential re-emergence together with a red dot on the chart.

**Figure 2.** Measles incidence trend analysis and re-emergence detection for China. Panel A shows the case corresponding to 3940 measles cases in China in 2017 (as in the case study). The model does not predict re-emergence for this case. Panel B shows the case when the number of measles cases in China are increased to 150,000 in 2017. As this represents a large change in incidence, the model predicts potential re-emergence.



Q2, “What are potential contributing factors?”: A summary table (Figure 3) showed the range of factors that potentially contribute to re-emergence, including the values of public health

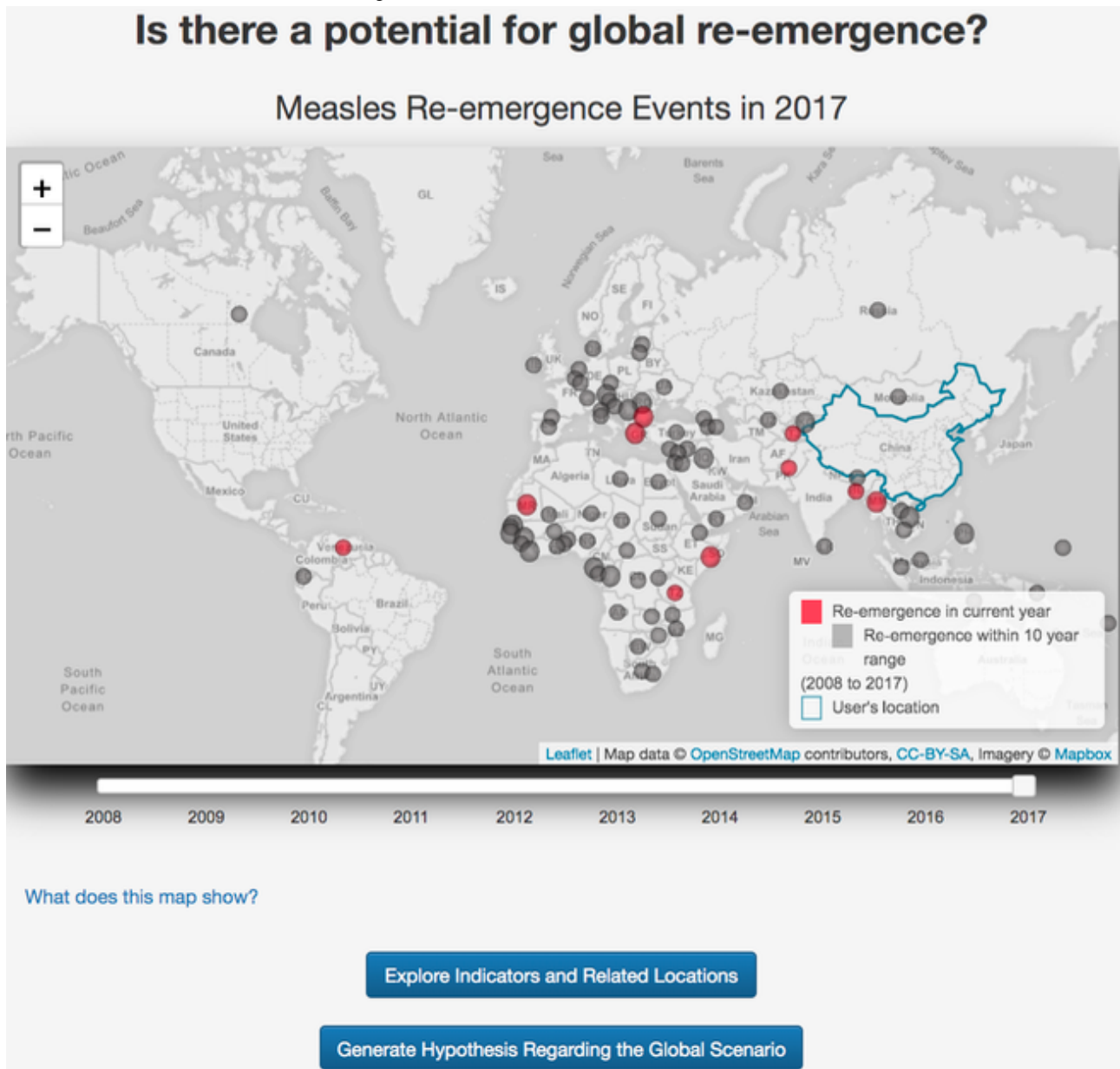
indicators that map to causes of RED for measles, for both the user input year and the median for the recent history (2000 to present). Harmful or protective values were colored red or blue.

**Figure 3.** Potential contributing factors for measles in China.

Q3, “Is there a potential for global re-emergence?”: A dynamic review of the past 10 years from input year (Figure 4) showed that global re-emergence likely began around the 2008-2009 timeframe. Interestingly, most experts identified global re-emergence around the 2011-2012 timeframe, indicating that RED Alert could have provided earlier warning. Within the past

5 years from 2017, several countries showed re-emergence of measles, but the geographic distribution was concentrated in Eastern Europe and Africa. Myanmar and Bangladesh, which border China, experienced potential re-emergence of measles in 2017, but the disease did not travel across the border to elicit a similar disease event in China.



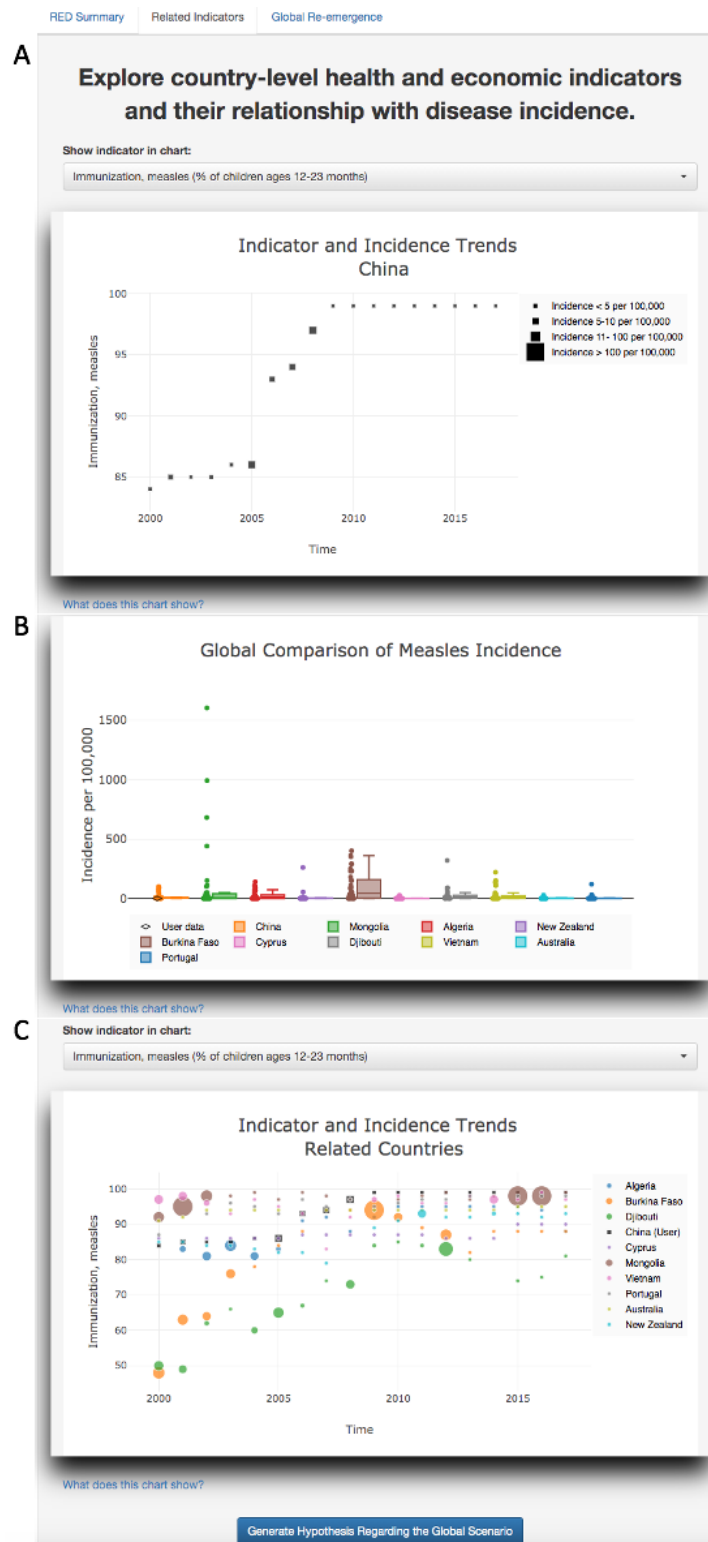
**Figure 4.** Distribution of national measles re-emergence events worldwide.

Thus, RED Alert was able to successfully address the primary objectives for which it was developed, and provide actionable information.

The output of the additional analytics was examined on the “related indicators” tab. Selection of “Immunization, measles (% of children ages 12-23 months)” for the first plot (Figure 5A) on this tab showed that the measles immunization rate exceeded 90% and was maintained above the 90% threshold since 2006, offering a potential reason as to why re-emergence

was not identified in China. The comparative boxplot (Figure 5B) showed the countries that had an incidence between 50% and 150% of China’s incidence in 2017, offering a global context. For example, the chart showed that New Zealand and China had very similar incidence perhaps due to similar vaccination rates. This hypothesis could be validated by the selection of “Immunization, measles (% of children ages 12-23 months)” above the third plot (Figure 5C), which showed incidence rates and vaccination coverage to be similar within the past 5 years for New Zealand and China.

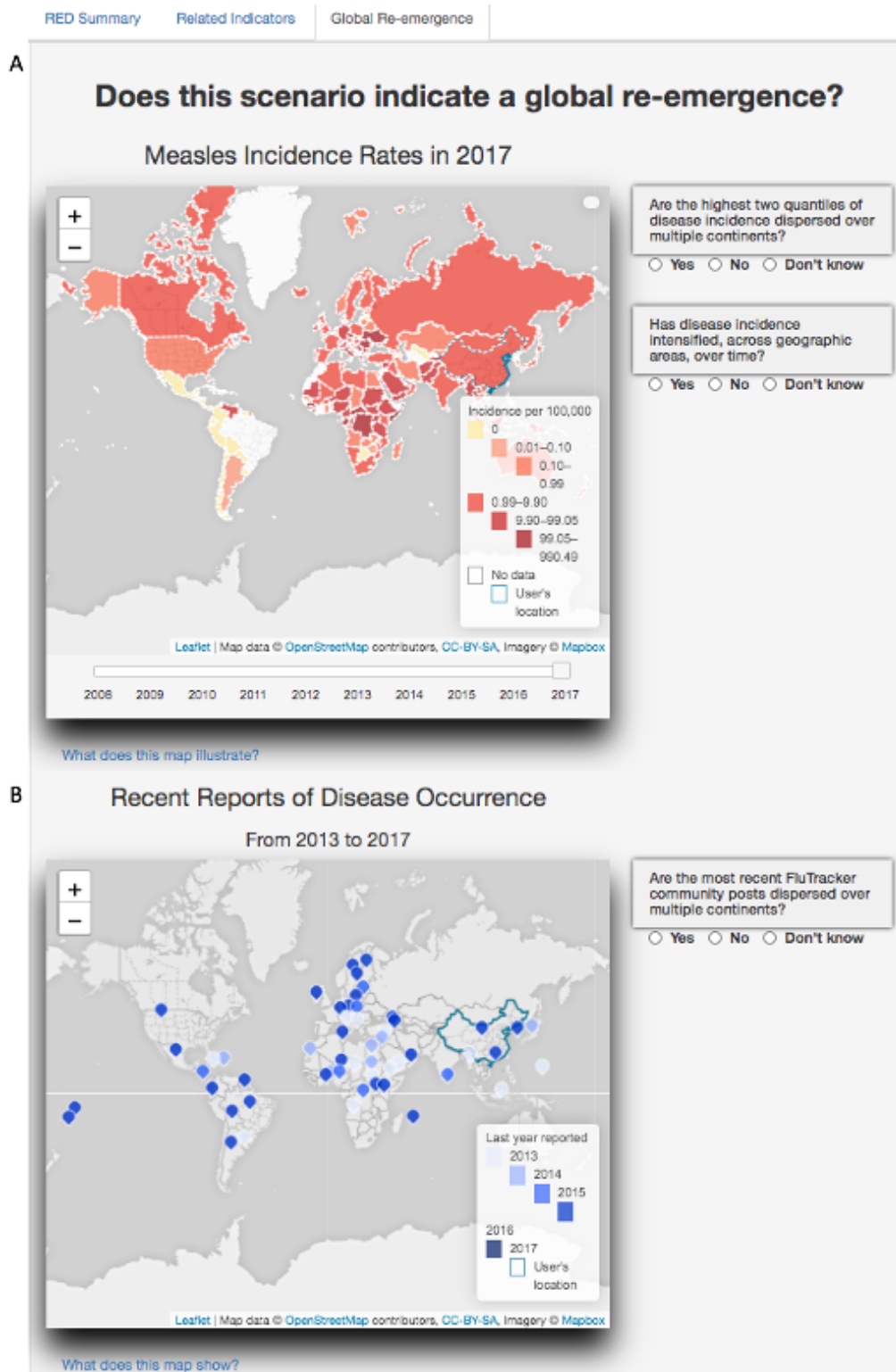
**Figure 5.** RED Alert’s related indicators tab. Panel A shows relationship between measles incidence and measles vaccination for China. Panel B shows global comparison of measles incidence. Panel C shows indicator and measles incidence trend for related countries.



Finally, the utility of visual analytics to understand the global scenario of re-emergence was examined on the “Global Re-emergence” tab. The first global map (Figure 6A) showed the incidence in 2017 and the highest incidence values in Africa. A dynamic review of the past 10 years showed that the incidence was globally higher 5 years before 2017. The second global

map showed that measles had been discussed on the international disease bulletin website FluTrackers across all continents within the past 2 years (Figure 6B). These maps provided a context to the 2017 China situation and indicated that global re-emergence of measles has occurred much earlier.

**Figure 6.** RED Alert’s global re-emergence tab. Panel A shows measles incidence worldwide. Panel B shows recent reports of disease occurrence for measles based on a disease bulletin.



## Discussion

### Principal Findings

In this paper, we presented RED Alert, a web-based tool that can provide early warning and detection of infectious disease re-emergence (not disease emergence). It is designed to help public health analysts detect and understand disease re-emergence at both the local (ie, country level) and the global

scale through contextual data analysis. It uses supervised machine learning models to detect local disease re-emergence and visual data analytics to help identify and explore potential factors contributing to this re-emergence and assess global situation for potential disease re-emergence. Consistent with our goal of identifying all potential cases of disease re-emergence events while allowing some false positives, our supervised learning models were able to classify 82%-90% of

the re-emergence cases, however, with 19% to 31% (except 46% for dengue) false positives. A detailed evaluation of the models used for re-emergence detection is described in [9]. We have also evaluated the utility of the tool through a number of case studies. RED Alert contains all the relevant information to not only provide early warning for potential re-emergence of disease locally and globally, but also offers causes for it. Through the diverse visual presentations and data at their fingertips, RED Alert allows users to verify their hypotheses about local and global re-emergence, and thus facilitates decision making in real-time. A user can access this tool as a one-stop shop for both data and relevant analyses and write a complete report.

While there are a number of online tools for disease surveillance [15,16,18], to the best of our knowledge, this is the first tool that is designed specifically for re-emerging diseases and focuses on detection of potential re-emergence at both local and global level as well as identification of potential contributing factors for the local re-emergence event.

Prior work in disease re-emergence has focused on the contributing factors of re-emergence. In particular, recent work has focused on the tremendous impact of climate change [49,50]. Changes to the climate impact almost every facet of disease transmission from increasing the habitat of disease vectors [51] to increasing the threat of civil unrest and violence [52], which in turn destabilizes infrastructure necessary for resiliency to re-emergence. To complicate this, it is clear that human factors such as urbanization and international travel also impact disease re-emergence [5-8]. However, despite the fact that the literature is clear that there is a complex system at work, the authors have not been able to find any other work in the data fusion or visualization space to allow public health experts to actually interact with the components necessary. Indeed, it is because of this complex milieu that RED Alert is necessary.

Our hope is that RED Alert can provide actionable information to public health analysts and decision makers that can be used for planning purposes. Our tool can provide indications that disease re-emergence may be occurring in a given region (or globally) and also help inform the user of possible contributing factors. This information may be useful in helping better understand the situation, as well as helping determine possible mitigations.

Currently, the tool has data for 4 diseases at the country level and yearly time scale. However, our methodology is applicable

to other diseases, as well as other spatial and temporal scales. In addition, although the current application is designed for use on a laptop or desktop computer, we are currently also developing a mobile app for this tool.

### Limitations

RED Alert is the first tool designed for detecting and understanding disease re-emergence and provides novel analysis. However, it relies on the availability and quality of data, which depends upon the public health infrastructure of the country. Under-reporting is common in biosurveillance systems [53]. While there are some missing data, historical data collected by the tool are relatively complete. By contrast, there is often some delay in reporting case counts data to agencies such as the WHO or PAHO or data collection companies such as Gideon. Similarly, there is also some delay in the estimation and availability of population and related indicators on the World Bank website. This often leads to missing data for many countries for a couple of recent years. To deal with this, we allow users to input recent case counts data and use values from the latest available years for population and related indicators for analysis purpose. We believe this is reasonable, as this information is less likely to change significantly in a short period. However, discrepancies in the data may affect our analysis.

While it is common in machine learning applications for humans to label data, due to the lack of the concrete definition of re-emergence, labeling is a subjective assessment. It may be possible that the SMEs in our team mislabeled data in some cases. Further, due to the lack of a concrete quantitative definition of re-emergence, it is difficult to fully validate our analysis.

### Future Directions

There are many opportunities for future work including adding more diseases to the tool based on their likelihood of re-emerging. Currently, the ability to perform the same analysis at a subnational level is mainly restricted by data availability. We are working on obtaining data at subnational levels for a few diseases and countries and plan to make this functionality available through a mobile app. Re-emergence detection models can also be improved by using other disease-related factors such as weather or climate data for mosquito-borne diseases, as mosquito density depends upon temperature and humidity.

---

### Acknowledgments

This work was supported by the U.S. Department of Energy through Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the US Department of Energy (Contract No. 89233218CNA000001). This work was funded by the Defense Threat Reduction Agency (DTRA) (grant #10027). Dr Ramesh Krishnamurthy, along with other team members of the WHO's Department of Information, Evidence, and Research in the Health Systems and Innovation Cluster, provided subject matter expertise on disease re-emergence as a global phenomenon, as well as detailed insight into the formative stages of the study. Drs Bryan Lewis and James Schlitt, Biocomplexity Institute, University of Virginia, were important contributors to understanding the spatial components of disease re-emergence.

---

## Authors' Contributions

NP, ARD, WR, MC, FA, NV, and GF collected the data. NP, ARD, WR, and GF cleaned and ingested the data. WR and GF designed the back end and ARD, WR, DA, and GF developed the front end. All authors helped design the front end/visualizations. ARD, MC, FA, and NV labeled the data for classification. NP and GF trained the classifiers and integrated them with the tool. AD conceived and led the project. NP and AD wrote the first version of the manuscript and all authors reviewed it.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Additional case studies for cholera, dengue, and yellow fever to show utility of RED Alert in detecting and understanding disease re-emergence.

[[DOCX File , 25478 KB-Multimedia Appendix 1](#)]

## Multimedia Appendix 2

Data Sources for RED Alert.

[[DOCX File , 16 KB-Multimedia Appendix 2](#)]

## References

1. The top 10 causes of death. World Health Organization. URL: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> [accessed 2020-12-17]
2. Teixeira MG, Costa MDCN, Barreto F, Barreto ML. Dengue: twenty-five years since reemergence in Brazil. *Cad Saude Publica* 2009;25 Suppl 1:S7-18 [FREE Full text] [doi: [10.1590/s0102-311x2009001300002](https://doi.org/10.1590/s0102-311x2009001300002)] [Medline: [19287868](https://pubmed.ncbi.nlm.nih.gov/19287868/)]
3. Antona D, Lévy-Bruhl D, Baudon C, Freymuth F, Lamy M, Maine C, et al. Measles elimination efforts and 2008-2011 outbreak, France. *Emerg Infect Dis* 2013 Mar;19(3):357-364 [FREE Full text] [doi: [10.3201/eid1903.121360](https://doi.org/10.3201/eid1903.121360)] [Medline: [23618523](https://pubmed.ncbi.nlm.nih.gov/23618523/)]
4. Woodall J, Yuill T. Why is the yellow fever outbreak in Angola a 'threat to the entire world'? *Int J Infect Dis* 2016 Jul;48:96-97 [FREE Full text] [doi: [10.1016/j.ijid.2016.05.001](https://doi.org/10.1016/j.ijid.2016.05.001)] [Medline: [27163382](https://pubmed.ncbi.nlm.nih.gov/27163382/)]
5. Barrett R, Kuzawa CW, McDade T, Armelagos GJ. Emerging and Re-emerging Infectious Diseases: The Third Epidemiologic Transition. *Annu. Rev. Anthropol* 1998 Oct 21;27(1):247-271. [doi: [10.1146/annurev.anthro.27.1.247](https://doi.org/10.1146/annurev.anthro.27.1.247)]
6. Fauci AS. Emerging and reemerging infectious diseases: the perpetual challenge. *Acad Med* 2005 Dec;80(12):1079-1085. [doi: [10.1097/00001888-200512000-00002](https://doi.org/10.1097/00001888-200512000-00002)] [Medline: [16306276](https://pubmed.ncbi.nlm.nih.gov/16306276/)]
7. Harrus S, Baneth G. Drivers for the emergence and re-emergence of vector-borne protozoal and bacterial diseases. *Int J Parasitol* 2005 Oct;35(11-12):1309-1318. [doi: [10.1016/j.ijpara.2005.06.005](https://doi.org/10.1016/j.ijpara.2005.06.005)] [Medline: [16126213](https://pubmed.ncbi.nlm.nih.gov/16126213/)]
8. Plans P, Torner N, Godoy P, Jané M. Lack of herd immunity against measles in individuals aged <35 years could explain re-emergence of measles in Catalonia (Spain). *Int J Infect Dis* 2014 Jan;18:81-83 [FREE Full text] [doi: [10.1016/j.ijid.2013.09.015](https://doi.org/10.1016/j.ijid.2013.09.015)] [Medline: [24211476](https://pubmed.ncbi.nlm.nih.gov/24211476/)]
9. Chitanvis M, Daughton AR, Altherr F, Parikh N, Fairchild G, Rosenberger W, et al. Development of a Supervised Learning Algorithm for Detection of Potential Disease Reemergence: A Proof of Concept. *Health Secur* 2019;17(4):255-267. [doi: [10.1089/hs.2019.0020](https://doi.org/10.1089/hs.2019.0020)] [Medline: [31433278](https://pubmed.ncbi.nlm.nih.gov/31433278/)]
10. Staples J, Breiman R, Powers A. Chikungunya fever: an epidemiological review of a re-emerging infectious disease. *Clin Infect Dis* 2009 Sep 15;49(6):942-948. [doi: [10.1086/605496](https://doi.org/10.1086/605496)] [Medline: [19663604](https://pubmed.ncbi.nlm.nih.gov/19663604/)]
11. Hartskeerl R, Collares-Pereira M, Ellis W. Emergence, control and re-emerging leptospirosis: dynamics of infection in the changing world. *Clin Microbiol Infect* 2011 Apr;17(4):494-501 [FREE Full text] [doi: [10.1111/j.1469-0691.2011.03474.x](https://doi.org/10.1111/j.1469-0691.2011.03474.x)] [Medline: [21414083](https://pubmed.ncbi.nlm.nih.gov/21414083/)]
12. Seleem M, Boyle S, Sriranganathan N. Brucellosis: a re-emerging zoonosis. *Vet Microbiol* 2010 Jan 27;140(3-4):392-398. [doi: [10.1016/j.vetmic.2009.06.021](https://doi.org/10.1016/j.vetmic.2009.06.021)] [Medline: [19604656](https://pubmed.ncbi.nlm.nih.gov/19604656/)]
13. Gubler DJ. The Global Threat of Emergent/Re-emergent Vector-Borne Diseases. In: Atkinson PW, editor. *Vector Biology, Ecology and Control*. Dordrecht, The Netherlands: Springer; 2010:39-62.
14. Generous N, Fairchild G, Khalsa H, Tasseff B, Arnold J. Epi Archive: automated data collection of notifiable disease data. *OJPHI* 2017;9(1). [doi: [10.5210/ojphi.v9i1.7615](https://doi.org/10.5210/ojphi.v9i1.7615)]
15. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc* 2008;15(2):150-157 [FREE Full text] [doi: [10.1197/jamia.M2544](https://doi.org/10.1197/jamia.M2544)] [Medline: [18096908](https://pubmed.ncbi.nlm.nih.gov/18096908/)]
16. Collier N, Doan S, Kawazoe A, Goodwin RM, Conway M, Tateno Y, et al. BioCaster: detecting public health rumors with a Web-based text mining system. *Bioinformatics* 2008 Dec 15;24(24):2940-2941 [FREE Full text] [doi: [10.1093/bioinformatics/btn534](https://doi.org/10.1093/bioinformatics/btn534)] [Medline: [18922806](https://pubmed.ncbi.nlm.nih.gov/18922806/)]

17. Velappan N, Daughton AR, Fairchild G, Rosenberger WE, Generous N, Chitanvis ME, et al. Analytics for Investigation of Disease Outbreaks: Web-Based Analytics Facilitating Situational Awareness in Unfolding Disease Outbreaks. *JMIR Public Health Surveill* 2019 Feb 25;5(1):e12032 [FREE Full text] [doi: [10.2196/12032](https://doi.org/10.2196/12032)] [Medline: [30801254](https://pubmed.ncbi.nlm.nih.gov/30801254/)]
18. Codeço CT, Cruz OG, Riback TI, Degener CM, Gomes MF, Villela D, et al. InfoDengue: a nowcasting system for the surveillance of dengue fever transmission. *bioRxiv*. 2016. URL: <https://www.biorxiv.org/content/10.1101/046193v1.full.pdf> [accessed 2019-09-30]
19. Wang J, McMichael AJ, Meng B, Becker NG, Han W, Glass K, et al. Spatial dynamics of an epidemic of severe acute respiratory syndrome in an urban area. *Bull World Health Organ* 2006 Dec;84(12):965-968 [FREE Full text] [doi: [10.2471/blt.06.030247](https://doi.org/10.2471/blt.06.030247)] [Medline: [17242832](https://pubmed.ncbi.nlm.nih.gov/17242832/)]
20. Beaulieu-Jones BK, Greene CS. Semi-supervised learning of the electronic health record for phenotype stratification. *J Biomed Inform* 2016 Dec;64:168-178 [FREE Full text] [doi: [10.1016/j.jbi.2016.10.007](https://doi.org/10.1016/j.jbi.2016.10.007)] [Medline: [27744022](https://pubmed.ncbi.nlm.nih.gov/27744022/)]
21. Bose E, Radhakrishnan K. Using Unsupervised Machine Learning to Identify Subgroups Among Home Health Patients With Heart Failure Using Telehealth. *Comput Inform Nurs* 2018 May;36(5):242-248. [doi: [10.1097/CIN.0000000000000423](https://doi.org/10.1097/CIN.0000000000000423)] [Medline: [29494361](https://pubmed.ncbi.nlm.nih.gov/29494361/)]
22. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. *PLoS One* 2013;8(2) [FREE Full text] [doi: [10.1371/journal.pone.0056176](https://doi.org/10.1371/journal.pone.0056176)] [Medline: [23457520](https://pubmed.ncbi.nlm.nih.gov/23457520/)]
23. Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: A case study in China. *PLoS Negl Trop Dis* 2017 Oct;11(10) [FREE Full text] [doi: [10.1371/journal.pntd.0005973](https://doi.org/10.1371/journal.pntd.0005973)] [Medline: [29036169](https://pubmed.ncbi.nlm.nih.gov/29036169/)]
24. Rose S. Mortality risk score prediction in an elderly population using machine learning. *Am J Epidemiol* 2013 Mar 01;177(5):443-452. [doi: [10.1093/aje/kws241](https://doi.org/10.1093/aje/kws241)] [Medline: [23364879](https://pubmed.ncbi.nlm.nih.gov/23364879/)]
25. Chiang P, Dey S. Personalized Effect of Health Behavior on Blood Pressure: Machine Learning Based Prediction and Recommendation. 2018 Presented at: IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom); 2018; Ostrava, Czech Republic p. 1-6. [doi: [10.1109/healthcom.2018.8531109](https://doi.org/10.1109/healthcom.2018.8531109)]
26. Du J, Xu J, Song H, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. *BMC Med Inform Decis Mak* 2017 Jul 05;17(Suppl 2):69 [FREE Full text] [doi: [10.1186/s12911-017-0469-6](https://doi.org/10.1186/s12911-017-0469-6)] [Medline: [28699569](https://pubmed.ncbi.nlm.nih.gov/28699569/)]
27. Thomas J, Cook K. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. United States: National Visualization and Analytics Ctr; 2005.
28. Youn-Ah Kang, Görg C, Stasko J. How Can Visual Analytics Assist Investigative Analysis? Design Implications from an Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 2011 May;17(5):570-583. [doi: [10.1109/TVCG.2010.84](https://doi.org/10.1109/TVCG.2010.84)] [Medline: [20530814](https://pubmed.ncbi.nlm.nih.gov/20530814/)]
29. Simpao AF, Ahumada LM, Rehman MA. Big data and visual analytics in anaesthesia and health care. *Br J Anaesth* 2015 Sep;115(3):350-356 [FREE Full text] [doi: [10.1093/bja/aeu552](https://doi.org/10.1093/bja/aeu552)] [Medline: [25627395](https://pubmed.ncbi.nlm.nih.gov/25627395/)]
30. Maciejewski R, Tyner B, Jang Y, Zheng C, Nehme R, Ebert D, et al. LAHVA: Linked Animal-Human Health Visual Analytics. 2007 Presented at: IEEE Symposium on Visual Analytics Science and Technology; 2007; Sacramento, CA, p. 27-34. [doi: [10.1109/VAST.2007.4388993](https://doi.org/10.1109/VAST.2007.4388993)]
31. Raghupathi W, Raghupathi V. An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach. *Int J Environ Res Public Health* 2018 Mar 01;15(3) [FREE Full text] [doi: [10.3390/ijerph15030431](https://doi.org/10.3390/ijerph15030431)] [Medline: [29494555](https://pubmed.ncbi.nlm.nih.gov/29494555/)]
32. Perer A, Sun J. MatrixFlow: Temporal Network Visual Analytics to Track Symptom Evolution during Disease Progression. *AMIA Annu Symp Proc* 2012:716-725. [Medline: [2330434](https://pubmed.ncbi.nlm.nih.gov/2330434/)]
33. RED Alert, Re-emerging Infectious Disease Alert. URL: <https://redalert.bsvgateway.org/> [accessed 2020-12-20]
34. WHO vaccine-preventable diseases: monitoring system. 2020 global summary.: World Health Organization URL: [http://apps.who.int/immunization\\_monitoring/globalsummary](http://apps.who.int/immunization_monitoring/globalsummary) [accessed 2020-12-19]
35. Global Health Observatory data repository; Number of reported cases Data by country.: World Health Organization URL: <http://apps.who.int/gho/data/node.main.175?lang=en> [accessed 2019-09-29]
36. DengueNet; Welcome to the DengueNet database and geographic information system.: World Health Organization URL: <http://apps.who.int/globalatlas/default.asp> [accessed 2019-09-29]
37. Gideon. Gideon URL: <https://www.gideononline.com/> [accessed 2020-12-20]
38. Dengue and Severe Dengue Cases and Deaths for countries and territories of the Americas.: Pan American Health Organization (PAHO) URL: <https://www.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/dengue-nacional-en/257-dengue-casos-muertes-pais-ano-en.html> [accessed 2019-09-29]
39. LandScan.: Oak Ridge National Laboratory URL: <https://landscan.ornl.gov/> [accessed 2020-12-20]
40. DataBank; Population estimates and projections.: The World Bank URL: <http://databank.worldbank.org/data/reports.aspx?source=population-estimates-and-projections> [accessed 2019-09-29]
41. Measles-containing vaccine.: World Health Organization URL: [http://apps.who.int/immunization\\_monitoring/globalsummary/timeseries/tscoveragemcv1.html](http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tscoveragemcv1.html) [accessed 2019-09-29]
42. Working with the regions.: World Health Organization URL: <https://www.who.int/chp/about/regions/en/> [accessed 2020-12-22]

43. Dicker R, Coronado F, Koo D, Parrish RG. Introduction to Epidemiology. In: Principles of Epidemiology in Public Health Practice. Atlanta, GA: Centers for Disease Control and Prevention; 2006.
44. Indicators.: The World Bank URL: <https://data.worldbank.org/indicator> [accessed 2019-09-29]
45. About the Indicators API Documentation.: The World Bank URL: <https://datahelpdesk.worldbank.org/knowledgebase/articles/889392-apidocumentation> [accessed 2019-09-29]
46. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12:2825-2830.
47. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 2006 Feb 23;7 [FREE Full text] [doi: [10.1186/1471-2105-7-91](https://doi.org/10.1186/1471-2105-7-91)] [Medline: [16504092](https://pubmed.ncbi.nlm.nih.gov/16504092/)]
48. FluTrackers.com, Tracking Infectious Diseases since 2006. URL: <https://flutrackers.com/forum/> [accessed 2019-09-29]
49. Zell R. Global climate change and the emergence/re-emergence of infectious diseases. International Journal of Medical Microbiology Supplements 2004 Apr;293:16-26. [doi: [10.1016/s1433-1128\(04\)80005-6](https://doi.org/10.1016/s1433-1128(04)80005-6)]
50. El-Sayed A, Kamel M. Climatic changes and their role in emergence and re-emergence of diseases. Environ Sci Pollut Res Int 2020 Jun;27(18):22336-22352 [FREE Full text] [doi: [10.1007/s11356-020-08896-w](https://doi.org/10.1007/s11356-020-08896-w)] [Medline: [32347486](https://pubmed.ncbi.nlm.nih.gov/32347486/)]
51. Rochlin I, Ninivaggi DV, Hutchinson ML, Farajollahi A. Climate Change and Range Expansion of the Asian Tiger Mosquito (*Aedes albopictus*) in Northeastern USA: Implications for Public Health Practitioners. PLoS ONE 2013 Apr 2;8(4). [doi: [10.1371/journal.pone.0060874](https://doi.org/10.1371/journal.pone.0060874)]
52. Sofuoğlu E, Ay A. The relationship between climate change and political instability: the case of MENA countries (1985:01–2016:12). Environmental Science and Pollution Research 2020 Feb 8;27:14033-14043. [doi: [10.1007/s11356-020-07937-8](https://doi.org/10.1007/s11356-020-07937-8)]
53. Gibbons CL, Mangen MJ, Plass D, Havelaar AH, Brooke RJ, Kramarz P, et al. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. BMC Public Health 2014 Feb 11;14 [FREE Full text] [doi: [10.1186/1471-2458-14-147](https://doi.org/10.1186/1471-2458-14-147)] [Medline: [24517715](https://pubmed.ncbi.nlm.nih.gov/24517715/)]

## Abbreviations

- API:** application programming interface  
**PAHO:** Pan American Health Organization  
**SME:** subject matter expert  
**WHO:** World Health Organization

*Edited by T Sanchez; submitted 04.09.20; peer-reviewed by H Cousins, H Kadir; comments to author 25.09.20; revised version received 26.10.20; accepted 14.12.20; published 07.01.21*

*Please cite as:*

Parikh N, Daughton AR, Rosenberger WE, Aberle DJ, Chitanvis ME, Altherr FM, Velappan N, Fairchild G, Deshpande A  
*Improving Detection of Disease Re-emergence Using a Web-Based Tool (RED Alert): Design and Case Analysis Study*  
*JMIR Public Health Surveill* 2021;7(1):e24132  
URL: <http://publichealth.jmir.org/2021/1/e24132/>  
doi: [10.2196/24132](https://doi.org/10.2196/24132)  
PMID: [33316766](https://pubmed.ncbi.nlm.nih.gov/33316766/)

©Nidhi Parikh, Ashlynn R Daughton, William Earl Rosenberger, Derek Jacob Aberle, Maneesha Elizabeth Chitanvis, Forest Michael Altherr, Nileena Velappan, Geoffrey Fairchild, Alina Deshpande. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 07.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.