

The Surveillance Window Application (SWAP): A Web-Hosted Tool to Facilitate Situational Awareness during Outbreaks

**Alina Deshpande¹, Esteban Abeyta¹, Lauren Castro², Ashlynn Daughton¹, Geoffrey Fairchild¹,
Nicholas Generous¹, Reid Priedhorsky¹, Kirsten Taylor-McCabe³, Nileena Velappan⁴**

1. Defense Systems and Analysis Division, Los Alamos National Laboratory; Los Alamos, NM 87545

2. Department of Integrative Biology, University of Texas; Austin, TX

3. Emerging Threats Program Office, Los Alamos National Laboratory; Los Alamos, NM 87545

4. Bioscience Division, Los Alamos National Laboratory; Los Alamos, NM 87545

deshpande_a@lanl.gov; LA-UR-14-28001

Abstract

The Surveillance Window Application (SWAP) is a prototype tool to help public health professionals confronted with an unfolding disease outbreak, understand the historic context of that situation. It uses the professional's knowledge of the situation to automatically select similar historic outbreaks from the SWAP's library and displays them in a way that helps the professional understand the similarities to and differences from the current situation, along with recommendations for additional data sources that might be useful. Thus, the SWAP facilitates situational awareness by providing a context to otherwise isolated information that a professional is receiving and supports decision making by transforming incoming data into actionable information. In its prototype version, the SWAP is being developed for 11 diseases that affect both animals and humans.

Introduction

Data obtained in isolation has little value to a decision maker during an unfolding outbreak. For example, a public health official, upon receiving reports of a few cases of cholera in their city, can plan a more effective infection control strategy if they understand how a typical outbreak progresses in their area. Because disease control actions are time dependent, understanding where along the timeline of an outbreak one may be is vital for outbreak mitigation, especially if it is within the surveillance window (SW) - the disease-specific window of time when information gathered can be used to assist decision makers in

effectively responding to an impending outbreak (Deshpande et al. 2013).

To provide a frame of reference and context to make effective decisions, we are developing the Surveillance Window Application (SWAP). The SWAP is a tool to help public health professionals who are confronted with an ongoing or potential disease outbreak understand the historic context of that situation, thereby enhancing situational awareness and assisting with decision making during a rapidly unfolding event. The SWAP accomplishes this by soliciting information about the user's current situation and, through an outbreak matching algorithm, matches the user's situation to the most comparable historical outbreaks. The SWAP also provides contextual information about the historical outbreak including what data streams were useful for detecting and monitoring the outbreak, what, if any mitigative actions were taken, and important events during the timeline such as the first declaration of the outbreak or the occurrence of a weather change.

Methods

The development of the SWAP follows five steps for each of the eleven diseases:

1. Construction of SWAP library
2. Identification of input parameters
3. Assignment of weights to input parameters
4. Development of outbreak-matching algorithm
5. Implementation in SWAP

1. Construction of the SWAP library

The SWAP has a database that stores information on case counts, properties, and other disease-specific factors associated with each outbreak within a disease specific library. Each outbreak entry in the database contains a time series of case counts, SW start and end dates, and a list of relevant data streams. We also store additional important factors associated with each outbreak such as risk factors associated with disease spread, mitigations implemented for that outbreak and relevant environmental events.

Inclusion of an outbreak into the SWAP library is based on two specific criteria: more emphasis on recent outbreaks (2000 to present) to provide a frame of reference more reflective of current events and conditions, which would increase the practicality of this app, and availability of reliable data in the open-source literature. To identify and collect time-series data, we use official academic and government reports for open-source data on historical outbreak case-counts and retrospective case-studies. The SWAP library currently includes multiple historical outbreaks for 11 diseases: cholera, dengue, foot-and-mouth disease (FMD), avian influenza, swine influenza, Q fever, chikungunya, norovirus, meningococcal disease, plague and tularemia.

2. Identification of input parameters

In order to determine which information is important to solicit from the user for the outbreak-matching algorithm, the relevant input parameters need to be identified. For every disease there are proximal and distal environmental, economic, and social influences that determine the outbreak's progression (e.g. the introduction of a pathogen in a non-immune population, increased rainfall leading to increased contamination of drinking water, poor political stability that leads to the cessation of disease prevention).

Within a disease portfolio of outbreaks, we evaluate the variation between the curves by the patterns of duration and number of cases at different points in the outbreak: at the beginning of rapid-rise, at the peak, and at the conclusion. We also search for common social, economic, and environmental factors present in outbreaks to see if they could serve as input parameters that help filter out and sort through the outbreaks. These are labeled "properties". Interestingly, common risk factors (e.g poor sanitation), which are shared by an outbreak in general are not good filtering factors (because they are found in all the outbreaks) and thus make poor properties. If a property has retrospective data available for the historical outbreaks, and can be expected to have meaningful data available in the future for a user in an unfolding situation, these factors are used design specific input queries for the SWAP user. For example, following a detailed analysis, outbreak curves collected for cholera could be grouped into endemic and non-endemic sets. Endemic outbreaks tended to be

smaller in scale and occurred at specific times during the year, whereas non-endemic outbreaks were much larger in scale and did not have seasonality. Thus endemicity of cholera in a particular location could be used as a property to filter outbreaks that matched the user's input on location.

Tables 1 and 2 list the properties that were identified to be useful as input parameters for Cholera and FMD respectively. As can be seen, each disease has different influencing properties

Properties included in the SWAP for Cholera	Properties deferred
Case counts over time (per week or total)	Sanitation access
Disease status of location (endemicity)	Case fatality ratio
Population level at risk (village, city, etc.)	Water quality
Regional location	Attack rate
	Age related statistics (infant mortality)
	Meteorological determinants
	Urban versus rural setting
	Surveillance window duration

Table 1: Properties considered for SWAP input parameters (Cholera)

Properties included in the SWAP for FMD	Properties deferred
Location	Control radius
Animal identifier	Response resources
Local density of cattle	Syndrome history
Local density of swine	Local density of sheep
Infected premises per day (Infection rate)	

Table 2: Properties considered for SWAP input parameters (FMD)

3 and 4. Assignment of property weights and development of outbreak matching algorithm

The curve matching algorithms are designed to return a similarity score between the user's input and the historical outbreaks, based on the weighted mean of how properties match. This yields a ranked list of similarity scores for the portfolio of outbreaks in the SWAP library for each disease/syndrome as shown below;

N = Number of attributes

m = match level of property n with input query n

w = weight of property n

$$s(\text{similarity score}) = \frac{\sum_1^N m_n \cdot w_n}{\sum_1^N w_n}$$

Based on the properties selected for each disease, specific functions to score the match level between the user's input to the property query and the historical outbreak property are designed, yielding a result between 0 and 1. These properties help the app sort which outbreaks have similar factors, and give them a higher score. Some match levels may only be a binary 0 or 1, especially for qualitative matches. For example, for the property "disease status" for cholera, an input of "non-endemic", compared to an "endemic" disease status of a specific historical outbreak would yield a match level of 0. Other properties such as "Location" may allow for match levels to be between 0 and 1 as the match level for location decreases as the user input country becomes further away from the historical outbreak location. As all properties are not equally influential in determining the progression of the outbreaks, weights are assigned to each property in the algorithm based on a cross case study of outbreaks and sensitivity analysis, which yields a *ranked* list of influencing properties. The ranks are converted to weights using established swing weighting techniques (e.g. Rank sum method). Table 3 shows an example of weights that were assigned for FMD derived from the ranks for each property.

Property	Weight
Differences in number of cases (y-axis)	0.265
Differences in distance to outbreak curve (x-axis)	0.265
Cases/day (infection rate)	0.19
Density of cattle	0.065
Density of swine	0.065
Animal identifier	0.1
Location	0.05

Table 3: Assignment of weights to FMD input parameters

For a majority of the diseases currently being included in the SWAP, the most indicative property for evaluating the match between the user input and each historical outbreak has been determined to be how closely the number of cases during the time specified by the user matched the number of cases for the same amount of time in the historical outbreak. Therefore the highest weights are often assigned to these properties, with one evaluating the difference between the number of cases for a defined period of time (y-axis difference) and the other evaluating the distance between the user's estimate of time and the curve's actual time for that number of cases (x-axis difference). These properties together also determine the point estimate on the curve in the display of the SWAP.

The similarity score obtained following the application of a disease specific curve matching algorithm for a hypothetical user input looks like the example shown for cholera below (Table 4 and similarity score below).

Match function	Query	Outbreak	Result	Weight
X-Axis (Time) [±]	15 (days)	10	0.66	0.3
Y-Axis (Cases) [±]	100	80	0.8	0.3
Population at risk	city (1)	one province (2)	0.50	0.2
Disease status	endemic	endemic	1.00	0.13
Location	Tanzania	Kenya	0.75	0.07

Table 4: Hypothetical query values following comparison of input answers to property values for a single historical cholera outbreak.

Similarity score

$$S = \frac{0.66 \times 0.3 + 0.8 \times 0.3 + 0.75 \times 0.07 + 1.00 \times 0.13 + 0.50 \times 0.2}{0.3 + 0.3 + 0.2 + 0.13 + 0.07} \approx 0.72$$

Verification of algorithm – Once a disease specific algorithm is created, the library of outbreaks, relevant information and the algorithm are coded into the SWAP and the app is tested through the input of known information about specific outbreaks and evaluating both the similarity score returned by the app as well as the display of outbreak information shown in the output. This is an iterative process to allow refinement of the algorithm and weights assigned to each property, the goal being a similarity score for user input of close to 100% for the outbreak being used for testing.

5. Implementation in SWAP

As shown in Figure 1, the user selects a disease of interest and enters information based on the queries designed from the input parameters. The SWAP calculates the similarity scores for all the historical outbreaks of the selected disease in the SWAP library and returns a ranked list of outbreak scores. When the user clicks on an outbreak with a particular similarity score, the SWAP dynamically displays a time series of case counts on a graph, complete with a highlighted SW, when applicable, and a point estimate of where the user may be located during the timeline of the outbreak. The point estimate is placed along the X-Y axis according to the duration and number of case reports indicated by the user. A table detailing the calculation of the score is displayed in order to facilitate a comparison between the current and historical outbreak. Alongside the generated graph, additional factors that were not included as input parameters about the outbreak are shown. The last piece of information provided to the user is a table that categorizes data streams as either available within or outside of the SW.

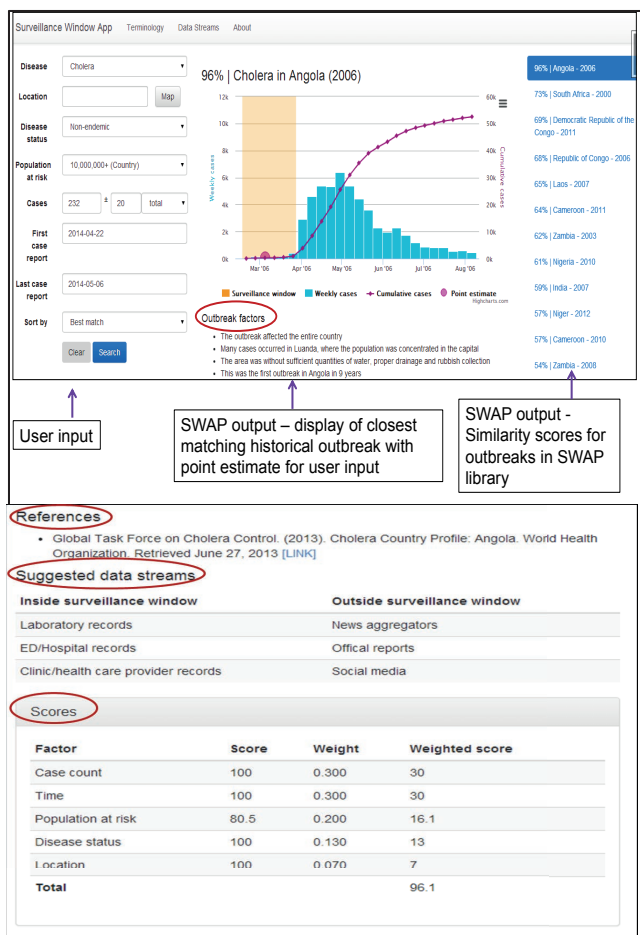


Figure 1: the output of the SWAP following an implementation of the outbreak matching algorithm for FMD.

Conclusion

While still under development and not yet clinically evaluated, the SWAP has the potential to benefit public health officials confronted with an ongoing or potential disease outbreak by assisting them in understanding the historic context of that situation. The SWAP may allow officials to determine if the situation they are receiving notification about is still within the surveillance window. In addition, the tool provides a recommendation for other data streams that the user could access to obtain more information. The SWAP can support rapid decision-making by compiling historical case information, and visually presenting it in a way that quickly provides the user with a context in which to understand their own situation without having to sift through large and unwieldy amounts of data.

To help understand the scenarios for SWAP application, we have developed a SWAP usability document that describes the scope, use cases and personas for which this tool has been designed. The primary use case is situational understanding i.e. what does a typical outbreak in the cur-

rent situation look like? Which historical outbreaks are similar? What is the provenance of information (e.g., article citation, ministry of health website) about a given outbreak, to enable further study? Is the situation still within the surveillance window? Which data streams are useful now? (For example, will looking at sales of OTC medication yield useful information, or has the window of opportunity for this passed?). The SWAP is intended for a diverse set of members of the global infectious disease surveillance community. This would include analysts who need to build and maintain situational awareness regarding a specific region, disease, or outbreak, create recommendations for action and communicate this awareness and/or recommendations to decision makers and physicians. The SWAP can also be used by physicians who may need to understand their individual cases in a wider context or decide whether to report a given case or set of cases to the authorities. Public health officials may find the SWAP useful for building and maintaining local (i.e., state or county) situational awareness regarding any diseases which are relevant, making decisions about local strategy and specific local interventions and justifying these decisions to local politicians and the public.

While the concept of using historical analogies for disease forecasting is not new, there have been no previous attempts to build a decision making application based off of historical data. There are efforts such as Gideon (commercial, <http://www.gideononline.com/>) and Project Tycho (<https://www.tycho.pitt.edu/>) that have data on historical outbreaks collected and digitized; however, a decision support tool such as the SWAP is the first of its kind.

Acknowledgements

This project was funded by the Defense Threat Reductions Agency – Joint Science and Technology Office (Grant # CB10007, DTRA10027-10845)

References

Deshpande, A., Brown, M., Castro, L., et al. 2013. *A Systematic Evaluation of Traditional and Non-traditional Data Streams for Integrated Biosurveillance—Final Report*. LANL LA-UR-13-2289