World Wide Web and Public Health Intelligence: Papers from the AAAI-14 Workshop

Towards Timely Public Health Decisions to Tackle Seasonal Diseases With Open Government Data

Vandana Srivastava
Freelance Analyst
vandy0105@gmail.com

Biplav Srivastava
IBM Research – India
sbiplav@in.ibm.com

Abstract

Improving public health is a major responsibility of any government, and is of major interest to citizens and scientific communities around the world. Here, one sees two extremes. On one hand, tremendous progress has been made in recent years in the understanding of causes, spread and remedies of common and regularly occurring diseases like Dengue, Malaria and Japanese Encephalistis (JE). On the other hand, public agencies treat these diseases in an ad hoc manner without learning from the experiences of previous years. Specifically, they would get alerted once reported cases have already arisen substantially in the known disease season, reactively initiate a few actions and then document the disease impact (cases, deaths) for that period, only to forget this learning in the next season. However, they miss the opportunity to reduce preventable deaths and sickness, and their corresponding economic impact, which scientific progress could have enabled. The gap is universal but very prominent in developing countries like India.

In this paper, we show that if public agencies provide historical disease impact information openly, it can be analyzed with statistical and machine learning techniques, correlated with best emerging practices in disease control, and simulated in a setting to optimize social benefits to provide timely guidance for new disease seasons and regions. We illustrate using open data for mosquito-borne communicable diseases; published results in public health on efficacy of Dengue control methods and apply it on a simulated typical city for maximal benefits with available resources. The exercise helps us further suggest strategies for new regions that may be anywhere in the world, how data could be better recorded by city agencies and what prevention methods should medical community focus on for wider impact.

Introduction

Disease outbreaks are natural calamities. However, managing their impact is in human control. For example, mosquito-borne diseases affect a large part of the world population at regular intervals creating periods of prominence, which we will refer to as their "*disease season",* and periods of lull.

Improving public health is not only a major responsibility of any government, but also an imperative to contain high economic impact of the alternative. Let us consider just one disease cause by mosquitoes, Dengue. According to a study conducted by World Health Organization (WHO) in 2005-06, the estimated overall cost of a Dengue case was US$ 828 (Sabchareon et al 2012). The research was carried out in eight countries (although Dengue has spread to more than 100 countries worldwide): five in the Americas (Brazil, El Salvador, Guatemala, Panama, Venezuela) and three in Asia (Cambodia, Malaysia, Thailand). Even in this small sample, the average annual number of officially reported cases in the period of 2001–2005 was 532,000 cases, which results in the cost of officially reported Dengue cases as US$ 440 million. Noting that this very conservative estimate ignores not only the under reporting of cases but also the substantial costs associated with Dengue surveillance and vector control programs, one can estimate the global impact of just one disease into billions of dollars annually.

The medical community has taken up this challenge aggressively. Other communities have joined hands with medical community for better understanding of disease spread. Computational epidemiology (Marathe and Vullikanti 2013) is an interdisciplinary area, which develops and uses computer models to understand and control the spatiotemporal diffusion of disease through populations. The models may be descriptive (macro-level) and work with statistical results over large databases or generative (micro-level) which computes disease spread via individual interactions.

However, such advances have not translated to large-scale impact on the ground. Public agencies routinely treat communicable diseases in an ad hoc manner without learning from the experiences of previous years.

Specifically, they would get energized only when reported cases rise abnormally in the *known* disease season (i.e., epidemic has already set in), reactively initiate a few actions and then once the epidemic is over, document the disease impact (cases, deaths) for that period. However, they miss the opportunity to reduce preventable deaths and sickness, and their corresponding economic impact, pro-actively, which scientific progress could have enabled.

In this paper, we attempt to narrow the gap between medical results and ground-level practices since it has received little attention. We do so by exploring what cost-effective strategies could be for tackling a mosquito-borne disease, Dengue. For this, we first set up a social benefit optimization setting for a simulated city and articulate how the recommendation may be used by public agencies there. Then, we analyze the effectiveness of known methods from health community and articulate how the recommendation may be used by public agencies there based on their social context. Although we take examples from India and a specific disease, the implications are global for any public agency tackling seasonal diseases. The exercise will help: (a) public agencies to increase adoption of promising approaches to new regions, (b) computer science researchers to convey how data could be better recorded by city agencies for deeper analysis, and (c) medical researchers to prioritize what prevention methods they should focus on for wider impact.

Simulated Setting for Timely Public Health Decisions

We will use an exemplar, but fictitious, city setting to analyze how diverse means for tackling a disease as reported in medical literature may be productively used for managing public health under practical constraints. The reason to keep the city abstract is because it helps keep the discussion focused on objective considerations of maximizing well being without touching on politically and morally sensitive issues around existence of the constraints in the first place. Our city is SundarPur[1] and it is made up of 10 districts with 10,000 people each. Each district allocates $10,000 per annum to prevent disease. The city has a district-level health administrator per district and then an overall citywide public health administrator.

A disease can be tackled with a set of known **methods**. Each method has a cost and an associated effectiveness rating. Every administrator selects an **approach** to tackle the disease in their district in consultation with others, where an approach is a single method or a combination of methods. We assume that for an administrator to select an approach, she will want a published medical study to be in place documenting how the corresponding method or methods were employed in a controlled environment along with their benefits and costs. Thus, SundarPur may follow different approaches to tackle the disease in its different districts in one season and based on its experience, try alternative approaches in subsequent years.

In this setting, we want to help the administrators decide which approach to take for their district to maximize overall public health within given budget constraints.

Mosquito Borne Diseases in India

We will consider mosquito borne diseases (Wikipedia 2013, APPL2011) in India as our running example in the paper. Among such diseases, Malaria, Dengue and Japanese Encephalitis (JE) are widespread around the world as well.

Indian's national portal, Data.gov.in, provides data about prevalence of different diseases between 2000-2011 (Data.gov.in 2013). Among the 5 included, viz. Acute Diarrheal Diseases, Malaria, Acute Respiratory Infection, Japanese Encephalitis (JE), Viral Hepatitis, two are mosquito borne (Malaria and JE). In absolute terms, for just 2011, there were about 1.2M mosquito related cases (in the two category) and 3.8M (in all 5). Although mosquito-borne diseases, as per this data, leads to small proportion of total cases (3.4% in 2011 and 3.5% over the last 5 years reported: 2007-2011), they significantly and increasingly contribute to the total number of deaths caused (27.6% in 2011 and 20.3% over the last 5 years reported: 2007-2011). Even this is a lower estimate because Dengue is missed out from this dataset. But it is available from another public dataset (NVBDCP 2013a) over 2007-2013, with 2012-2013 being provisional estimates. From these data, we can conclude that mosquito-borne diseases pose an increasing and virulent threat to public health in India.

In terms of controlling these diseases in India, Malaria has received the most attention and this has reflected in its decreasing impact over the past few years (NVBDCP 2013b). In contrast, JE and Dengue have seen increase in the number of cases as well as deaths. JE is confined to select regions (Uttar Pradesh and a few states) but has been entrenched for decades (APPL 2011, APPL 2013). For the rest of the paper, we will focus on Dengue, which is more widespread in India.

Dengue and Its Control

Dengue is a painful and infectious mosquito-borne virus affecting millions of people worldwide every year. It grew rapidly in 1960s and has spread from 9 countries then to

[1] SundarPur means *beautiful city* in Hindi.

more than 110 countries now. In India, it affects all her states.

The prominent, well-accepted, known **methods** to tackle Dengue (WHO 2013, Entogenex 2013) can be summarized as:

1. M1: Public awareness campaigns: to prevent conditions conducive to disease propagation, to improve reporting
2. M2: Chemical Control: Aerosol space spray
3. M3: Biological Control: Use of biocides
4. M4: Distributing equipments: bednets, insecticide-treated curtains
5. M5: Vaccination against the disease

Hence, one can try to control the habitat in which mosquitos grow, prevent their contact or control its impact with vaccines and new drugs. The methods vary in terms of scale of adoption, cost and proven effectiveness. Researchers have been working on Dengue fever vaccines and that is seen as the long-term solution, but the best prevention for now is to reduce mosquito habitat in areas where Dengue fever is common.

Any public agency wanting to control the disease will like to know how cost-effective these methods are individually and in combination? It will have a fixed budget and would want to minimize the cases and deaths immediately as well as better plan for future. With increasing case studies appearing in health community worldwide on how Dengue was controlled in their situations, can the advancements be analyzed to develop prescriptive guidance for new regions?

Towards the same, we will start by defining some commonly used metrics.

- Expense for disease control
  - $/person spent: How much money (in $) is spent for a given method divided by the population of the region. Lower is better.
- Impact of a disease control method
  - Reduction: What is the magnitude of reduction in disease cases due to a method, expressed as a percentage, in a time period (e.g., year, disease season)? Higher is better.
  - Cases/ person: How many reported cases of a disease occurred in a time period divided by the population of the region when a method was adopted? Lower is better.
- Cost-effectiveness:
  - Cases / $: how many cases were reported for a disease per dollar spent on controlling it in a given time period? Lower is better.

Dengue Control Lessons

To get the latest results on Dengue control, we reviewed the medical literature for case studies. They are summarized in Table 1. The cases are spread around the world and use different methods for control. We refine them in terms of the metrics defined earlier (expense, impact and cost-effectiveness). Further, to get an estimation of the effectiveness of each of the control measures in reducing Dengue spread, we looked up published results in which one method or a combination of more were used to determine the cost/person in Dengue prevention.

A1 represents a hybrid set of methods of (Taliberti and Zucchi 2005). In Sao Paolo, Brazil, a study was conducted for the year 2005 to calculate the direct costs of the Dengue fever control. The costs included the following items: human resources, uniforms, field materials, individual protection equipment, spraying equipment (Chemical Control), strategic supplies (insecticides and larvicides – Biological Control), and vehicles. The costs associated with laboratory tests for entomological surveillance and Dengue fever diagnosis were also calculated, as well as costs relating to information and printed materials for educational campaigns (Public Awareness). It was estimated that the total direct costs of the City of São Paulo Dengue Fever Program in 2005 were R$ 21,774,282.82 (US$ 12,486,941.34 considering the US Dollar/ Real exchange rate in December 2009). Of this amount, 59.4% was directed to human resources, 38.3% to epidemic control measures, and 2.2% to capital expenditure. The expense per person was R$ 1.99(US$ 1.14) in 2005.

Approach	Methods used (Mi)	Nature (Region, Population,area, year)	Expense per person	Reduction in number of cases
A1	M1, M2, M3	Sau Paulo, Brazil; 10,927,985; 2005	US $1.14	34%
A2	M3	Puerto Rico; -; 2003	< US$ 2.50	50% (in Dengue transmission)
A3	M2	Songkhla, Thailand; 162,645; 2009	US$ 1.24	
A4	M5	Bang Phae, Thailand; 207,000; - AND Thailand; 4002; 2009 - 2014		0-70%, 30.2%

Table 1: Dengue control case studies from literature

A2 represents the chemical method of (McConnell and Gubler 2003). In 2003, a study about expense effectiveness of larval control programs (Biological Control method) to reduce Dengue transmission in Puerto Rico was carried out. It was calculated that less than 2.0 US$ / person was needed to bring Dengue transmission by 50%.

Table 2: Cost-benefits for different approaches. * represents assumption made to compensate for missing data.

Approach Option	Population P	Amount available for expenditure (in USD) (a)	Expense per person for each method (in USD) (b)	Number of people exposed to the given method in the given amount c = (a)/(b)	Reduction in number of cases for each method (d)	Reduction in number of cases among exposed persons (e) = (c)*(d)	Effectiveness of the method E = (e) / P
O_def	10,000	10,000	0	0	0%	0	0%
O1_A1	10,000	10,000	1.14	8772	34%	2982	30%
O2_A2	10,000	10,000	2.5	4000	50%	2000	20%
O3_A3	10,000	10,000	1.24	8065	10% *	806	8%
O4_A4	10,000	10,000	8*	1250	70%	875	9%

A3 represents the biological method of (Ditsuwan et al 2012). In this study in Songkhla province in Thailand, the authors estimated the expense of standard indoor ultra-low-volume (SID-ULV) space spraying as a chemical control method for adult Dengue vectors. The components that were considered for estimation included recurrent expenses, capital expenses and productivity loss to the impacted population. The last component is irrelevant for our expense calculation. Hence, ignoring productivity loss, the study reports that the average unit expense per 10 houses was USD 35.7. With a per expense of $3.57 and the fact that the study was carried out on 56,353 houses with population of 162,645, giving 2.89 persons per house, we arrive at an expense per person of $1.24.

A4 represents the vaccination method from a few field studies (Longini et al 2013, Sabchareon et al 2012). In one study, researchers tried to find the effects of Sanofi Pasteur Tetravalent Dengue Vaccination in southeast Asia in a single year (Longini et al 2013). The effect was noted for various percentage of vaccination coverage (0%, 30%, 50% and 70% respectively) and the corresponding number of infected cases per 100,000 (4851, 1463, 688, 342 respectively). The research concluded that for each target age-group, vaccination of 70% population, should be effective for control.

In another research, also funded by Sanofi Pasteur, to find the effect of CYD tetravalent Dengue vaccine, a study was carried out on Thai schoolchildren (Sabchareon et al 2012). In this study, healthy school children aged 4-11 years were randomly chosen to receive three injections of Dengue vaccine or control (rabies vaccine or placebo) at months 0, 6, and 12. Participants were observed till month 25. This trial is registered at ClinicalTrial.gov, as NCT00842530 (Sanofi 2012). For the study, 4002 participants were given the vaccine (n= 2669) or became part of the control group (n=1333). 3673 were included in the primary analysis (2452 vaccine, 1221 control). The efficacy of the vaccine was found to be 30·2%. It was observed that Dengue vaccine was tolerated well with no side effects, 2 years after the first dose was administered. Large-scale phase 3 studies in various epidemiological settings are ongoing.

Note that although all the cited studies are in-depth in terms of their scope, they do not guide a well-meaning public agency who wants to control the disease in the upcoming disease season (e.g., monsoon).

Controlling Dengue in a Region

Using the results, we now try to derive prescriptive guidance for Sundarpur. Our goal will be to achieve maximum reduction in the infected cases at the district level with the given amount of USD $10,000 using the given methods. Ideally, one would have liked to know the right approach for their unique situation – a mix of control methods to produce the maximum reduction feasible. Realistically, even deciding which approach to replicate from all those published in literature is daunting since a study may have skipped reporting some information that is an important factor in making go-ahead decision at a new place. Our scaled down objective is to decide which published approach to replicate for maximum reduction in the likely infected cases for Sundarpur.

Table 2 presents the options for the approach the health administrator for a district can take. O_def refers to the default option of not doing anything. If the administrator does not act, the full impact of the disease will be felt without any intervention to reduce it. A variant of this is to spend the available money but not do anything of relevance that will reduce the disease cases. Although it is obvious to see that everyone should do better than O_def, in reality, this is the prevalent option. We want to do better than O_def.

Option O1_A1 is based on (Taliberti and Zucchi 2005) and summarized as A1 in Table 1. It involves a combination of
methods. However, any approach tried elsewhere earlier may or may not be applicable at Sundarpur to the same extent but one would only know about it after one tries. Even if the approach would work here as previously reported, it will cover a subset of the population (see second row, column (c)) but achieve 30% reduction in cases in the available budget.

Similarly, O2_A2 refers to (McConnell and Gubler 2003). and summarized as A2 in Table 1. It will bring down the cases by 20% in the given budget. O3_A3 refers to (Ditsuwan et al 2012) and summarized as A3 in Table 1. Note that literature does not report on the efficacy of the approach but the decision maker must know this information. Rather than dismiss the option altogether, we assume a 10% reduction which will be verified with actual results, if the approach gets chosen.

O4_A4 refers to (Longini et al 2013, Sabchareon et al 2012) and summarized as A4 in Table 1. Here, the literature does not report on the cost of the vaccines but the decision maker must know this information. Rather than dismiss the option altogether, we assume a US $8 cost per person who is administered based on other studies on end-to-end cost of vaccines. This option, when applied judiciously for maximum coverage reported in literature, results in reduction of cases by 9% in the given budget.

Interpreting Options to Make Decisions

We see that the best tactical option for administrators at Sundarpur (at district and the whole city level) is O1_A1 since it brings the maximum reduction. If the administrators are interested to cover the maximum number of people in the given budget, the best method is still O1_A1. If the administrators are interested to show maximum reduction in cases for a pocket of the city (sub-district level which may be more prone to the disease), they may choose O4_A4 but it costs maximum and thus can be perceived as taking resources away from the not-directed areas.

The city administrator can take a strategic view as well. Specifically, she can decide to try all approaches in different districts in a particular year, obtain specific data on cost and effectiveness as applicable in Sundarpur, and then try a more educated decision in subsequent year. A mixed approach will be for the administrator to take only a few options, example top-2 (O1_A1 and O2_A2), and try them in the districts in one year, and based on efficacy, decide the best option for Sundarpur in subsequent year. She may also use the vaccine option only when the disease outbreak is above certain threshold.

Researchers Focus for Creating Better Implementable Options

In the next sections, we look at what researchers can do for creating better options that can be implemented.

Advanced Options Overcoming Data Issues in Disease Control

Until now, we looked at how reported approaches for tacking a disease (Dengue) can be used in a new area. However, we assumed that the approaches are independent since the reported studies were geographically diverse and used a cost-benefit model for reuse. In reality, methods can have correlation (positive or negative) among them and so can approaches over time and space, which may affect their performance in a new area like Sundarpur. There may be many reasons for this but a few common ones are: resistance of disease agent (mosquito in case of Dengue) to methods used, e.g., vaccines, chemicals for fumigation; weather changes impacting disease spread, change in demographics, correlation between impact of methods. If such data is available, it can be used to extend the reuse model.

Further, as previously noted, approaches may not work in the new area as effectively as previously reported but we assumed it so for simplicity. In such a case, if the uncertainty (e.g., of effectiveness) is known, techniques from decision-making under uncertainty can be applied to overcome them. As one example, portfolio optimization in finance looks at maximizing the outcome from a portfolio of stocks that may themselves have uncertain returns. But such techniques need extensive rate of return data for individual stocks. Similarly, if we have data about approach effectiveness for a region, we can create an
optimal portfolio of disease control against uncertainty for that region.

Another class of analysis that can be done to create advanced options is optimization of multiple objectives, e.g., effectiveness, reduction, people coverage. However, for any of them, it is very important for cities to maintain and publish data on what approaches they already tried for disease control in the past and its efficacy, so that they can be improved.

Medical Research for Easy Rollout

In our review of existing work on efficacy of approaches to tackle diseases, we found that researchers sometimes overlook reporting of cost or effectiveness data. Both should be in clear terms since missing either of them makes adoption of their results hard. Further, vaccine as a means to reduce Dengue leads to sharp reduction, but is expensive. Hence, medical community is rightly focusing on ways to reduce this cost. However, they can also partner with data science community (computer science and mathematics) to try mixed approaches that reduce cost and improve overall effectiveness. Option O1_A1 was one such work we found in Dengue and it turned out to be the best tactical option for maximum reduction.

Conclusion and Future Work

In this paper, we attempted to bridge the gap between how diseases are handled by public agencies and the approaches that are suggested by the latest results in heath community. Specifically, we showed that if agencies provide historical disease impact information openly, it can be analyzed with statistical and machine learning techniques, correlated with best emerging practices in disease control, and simulated in a setting to optimize social benefits to provide timely guidance for new disease seasons and regions. We illustrated using open data for mosquito-borne communicable diseases in India, published results in public health on efficacy of Dengue control methods and apply it on a simulated typical city for maximal benefits with available resources.

One can extend this work in many ways. (1) One can expand the study to more approaches, methods and diseases. (2) One can expand the reuse analyses, as previously noted, to consider correlation with past approaches (time), similar geographies (space) and multi-objective criteria. (3) One can partner with actual city administrators to try the identified options in the field. (4). Finally, one can create a linked open data (LOD) portal and provide Sparql end-point to facilitate reuse of options and experience across cities around the world (W3CGLD 2013). We intend to work on them in future.

References

WHO, 2013, Dengue Control. At http://www.who.int/Denguecontrol/research/en/, Accessed 21 June 2013.

Entogenex, 2013, Integrated Mosquito Management. At http://www.entogenex.com/what-is-integrated-mosquito-management.html, Accessed 21 June 2013.

Marathe, M. and Vullikanti, A. K. S. Computational Epidemiology. Communications of the ACM , July 2013, Vol. 56, No. 7

Data.gov.in, 2013. Number of cases and deaths due to diseases (2000-2011). At http://data.gov.in/dataset/number-cases-and-deaths-due-diseases Accessed 21 June 2013.

NVBDCP, 2013a. Dengue cases and deaths in India (2007-2013). At http://nvbdcp.gov.in/den-cd.html Accessed 21 June 2013.

Wikipedia. Mosquito borne diseases, 2013. At http://en.wikipedia.org/wiki/Mosquito-borne_disease Accessed 21 June 2013.

APPL, 2011. Ground Realities in Japanese Encephalitis. At http://appltrust.org/files/GROUND-REALITIES-FINALISED-report-Dec2011.pdf Accessed 28 June 2013.

APPL, 2013. Studies on Japanese Encephalitis. At http://appltrust.org/projects/je-projects.htm Accessed 28 June 2013.

NVBDCP, 2013b. Strategic Action Plan for Malaria Control in India (2007-2012). At http://nvbdcp.gov.in/Round-9/Annexure-2%20%20Strategic%20action%20plan.pdf Accessed 21 June 2013.

McConnell, K.J. and Gubler, D.J. Guidelines cost-effectiveness of larval control programs to reduce dengue transmission in Puerto Rico, Rev Panam Salud Publica. 2003 Jul;14(1):9-16. At http://www.ncbi.nlm.nih.gov/pubmed/12952602, Accessed 28 June 2013.

Taliberti H and Zucchi P. Direct costs of the dengue fever control and prevention program in 2005 in the City of São Paulo, 2005. At http://www.ncbi.nlm.nih.gov/pubmed/20414506/ Accessed 28 June 2013.

Ditsuwan T, Liabsuetrakul T, Ditsuwan V, Thammapalo S. Cost of standard indoor ultra-low-volume space spraying as a method to control adult dengue vectors, Trop Med Int Health. 2012 Jun;17(6):767-74. At www.ncbi.nlm.nih.gov/pubmed/22943302 Accessed 28 June 2013.

Ira Longini, Dennis Chao, M. Elizabeth Halloran, Eben Kenah, Jonathan Sugimoto, Scott Halstead. Statistical and Mathematical Models for the Control of Dengue, 2013. At http://www.isi.it/sites/default/files/3.3-longini.pdf . Accessed 20 June 2013