Syndromic Surveillance in the United Arab Emirates

Magnus Boman
Swedish Institute of Computer Science (SICS), and KTH/ICT/SCS
Kista, Sweden, and
European Institute of Innovation and Technology (EIT ICT Labs), and
Visiting Professor at
Etisalat BT Innovation Center (EBTIC)
Abu Dhabi, UAE
Email: mab@sics.se

Baki Cakici
Stockholm Univ
Dept of Computer & Systems Sciences
Kista, Sweden

Christian Guttmann
Etisalat BT Innovation Center (EBTIC)
Khalifa University of Science, Technology and Research
Abu Dhabi, UAE
Email: christian.guttmann@kustar.ac.ae

Dr. Farida Al Hosani
Manager of Communicable Diseases Dept
Health Authority Abu Dhabi (HAAD)
UAE

Dr. Asma Al Mannaei
Head of Surveillance Section
Health Authority Abu Dhabi (HAAD)
UAE

***Abstract*—Opportunities for innovation in view of three complex problems faced by the UAE health care providers are described. The information dissemination problem faced could be approached by creating new channels for providing the population with public health information. These channels are precisely the ones typically used in so-called syndromic surveillance, including care-related data from communicable disease spread indicators, but also tweets and blog posts, for example. Syndromic surveillance could likewise assist the health authorities in addressing the knowledge elicitation problem: how to get more information on the life style, self care, and prevention among individual citizens. To some extent the prediction problem—how to predict the spread of infectious disease in the future and how to mathematically model social behaviour in the case of various health-threatening scenarios—would also be addressed by syndromic surveillance. Fully employed, the solutions proposed would provide new ICT services enabling preparedness for many forms of communicable disease outbreaks, as well as for natural disasters.**

I. Introduction

In matters of public health, individual patients seeking care for various illnesses constitute the main source of information for the health authorities. Traditional disease surveillance provides clinical data and, in most countries, this is sufficient for gathering adequate population-level data for all but the most extreme situations (e.g., a pandemic or a natural disaster). By contrast, reaching out to healthy individuals is difficult. A healthy individual often lacks motivation for filling out a survey about daily habits and routines that affect their general health condition. This is problematic, since many individuals today leading an unhealthy lifestyle—thus having a risk profile that makes it likely that they will suffer from ill health later in life—need to be informed of the risks.

The region in focus here is the UAE, and Abu Dhabi in particular. In Abu Dhabi, diabetes has the second highest prevalence in the world [1]. Obesity is one indicator in the risk profile of this disease. Diabetes is in turn an indicator in the risk profile of cardiovascular disease (CVD), the leading cause of death in the UAE [2] and a huge global problem [3]. Pro-activity with respect to information about the health risks for the individual for type 2 diabetes—which includes serious symptoms in cases of long progressed disease, such as blindness due to retinal damage—is of utmost important to the region. In Abu Dhabi, the Weqaya Web site (http://www.weqaya.ae/en) is one portal of information to the public about such risks. In spite of obvious benefits to the individual Weqaya user of signing up for an account, such as personalised health recommendations and direct links to Web services for open slots for a doctor's appointment, only a small fraction of the population have thus far registered. Taken together with the high prevalence for chronic diseases related to life style, this again highlights the information dissemination problem in the region.

In some countries (see, e.g., www.influenzanet.eu), the health authorities have experimented with an information-intensive form of public health surveillance called *syndromic surveillance*, where data gathered for non-diagnostic purposes are used to identify signals relevant to public health. This development goes hand in hand with data engineering advances within Information and Communication Technology (ICT) in recent years. Notably, so-called Big Data has moved from a technical obstacle for gaining knowledge, into a huge opportunity to business and governments, worldwide [4].

The diabetes and CVD prevalence in the region merit further investigation into the possibilities for employing syndromic surveillance. A research hypothesis is that the *information dissemination problem* in the UAE could be approached by providing new channels, namely the ones that are typically used in syndromic surveillance, and that this would in turn help increase self care and prevention. These new forms

978-1-4673-1101-4/12/$31.00

of surveillance would also assist the health authorities in the *knowledge elicitation problem*, i.e., how to get more information on the life style and self care of the individuals that are today healthy. Finally, the *prediction problem*, i.e., how to predict the spread of communicable disease in the future and how to mathematically model social behaviour in the case of various health-threatening scenarios, would be addressed. Solving the prediction problem is especially important to communicable disease, since it helps in answering policy questions, such as when and where to distribute vaccine, nurses, or other resources. Ideally, more and better real-time support to first responders can be provided by complementing traditional disease surveillance with syndromic surveillance. Preparedness plans and means to detecting early signals are extremely important in this context, and also assists health care personnel in the so-called race-to-trace [5].

Below, a background including related research is first given. In the following section, the research opportunity and some possible paths for cooperation between medical and ICT experts regarding syndromic surveillance in the UAE are scrutinised. Then, an analysis of current and next steps is made, before the conclusions are stated.

II. Background

Syndromic surveillance is a recent addition to the set of methods collectively described as *disease surveillance*: an epidemiological practice where the activities of a population are monitored for predefined signs. These signs are then interpreted in an attempt to prevent or minimise the spread of the disease.

One of the most comprehensive and influential definitions of *syndromic surveillance* was given by the United States Center for Disease Control and Prevention (CDC):

> Syndromic surveillance for early outbreak detection is an investigational approach where health department staff, assisted by automated data acquisition and generation of statistical signals, monitor disease indicators continually (real-time) or at least daily (near real-time) to detect outbreaks of diseases earlier and more completely than might otherwise be possible with traditional public health methods (e.g., by reportable disease surveillance and telephone consultation). The distinguishing characteristic of syndromic surveillance is the use of indicator data types. [6, p.2]

The unifying property of advanced disease surveillance systems as ICT systems is that they are designed to function without human intervention, performing statistical analyses at regular intervals to discover aberrant signals that match the parameters set by their operators. Recent advances in ICT have made the development and operation of such systems technically feasible, and many systems have been proposed to interpret multiple data sources, including those containing non-health related information, for disease surveillance. The introduction of these systems to the public health infrastructure has in many countries been accompanied by significant criticism regarding the diverting of resources from public health programs to the development of the systems [7], [8], the challenges of investigating the alerts raised by the systems [9], and the claims of rapid detection [10], [11]. Hence, the bounded rationality of introducing syndromic surveillance in any region must be assessed, taking into account the critique already put forth, and taking seriously the lessons learnt.

In the UAE, traditional disease surveillance has already been established. The electronic notification of infectious diseases is implemented in Abu Dhabi, since 2010. In addition, active surveillance of specific communicable diseases of interest is established through routine screening, such as the Visa Screening and Premarital Screening initiatives [12]. The national visa screening is a federal requirement on screening all expatriates above the age of 18 years intending to reside in the country, against specific communicable diseases. All new expatriates coming to work in the UAE are screened, and this is repeated every 2-3 years. About one million applicants are screened every year. Life style and self care questions are now, to a small extent, part of the Premarital Screening questionnaire. In addition, the Health Authority - Abu Dhabi (HAAD) has taken the opportunity to integrate risk factors surveillance within the vital statistics registries.

The motivation behind developing advanced ICT systems for disease surveillance can be partially explained by the observation that epidemiologists tasked with monitoring communicable diseases are expected to maintain an awareness of multiple databases during their daily work. To provide the experts with a rapid overview of all available data, and to equip them with additional information to make decisions, development of ICT systems are proposed [13]. The scope of public health surveillance has expanded over the past few years to include elements beyond diseases as a core requirement for decisions regarding subjective well-being, quality of life (QALE/QALY), spatial data, and more. The increasing demand on data has created an extreme burden on the process of data capture and analysis.

As defined by the CDC, Public Health Surveillance is the ongoing systematic collection, analysis, and interpretation of data (e.g., regarding agent/hazard, risk factor, exposure, health events) essential to the planning, implementation, and evaluation of public health practice. This should be closely integrated with the timely dissemination of these data to those responsible for prevention and control.

Conceptually, disease surveillance systems may be partitioned into collection, analysis, and reporting. The *collection* component contains lists of available data sources, collection strategies for data sources, instructions for formatting the collected data, and storage solutions. The *analysis* component stores a wide variety of computational methods used to extract significant signals from the collected data. The final component, *reporting*, contains the procedures for communicating analysis results to interested parties. The results may be presented in many forms: statistical analysis, incidence plots, heatmaps, or simply as messages advising experts to check a data source for further information.

Data sources most often used for syndromic surveillance, ordered by availability, from earliest to latest, are as follows [11], [14]:

- over-the-counter drug sales
- triage nurse line calls
- work and school absenteeism
- prescription drug sales
- emergency hotline calls
- emergency department visit chief complaints
- laboratory test orders
- ambulatory visit records
- veterinary health records
- hospital admissions and discharges
- laboratory test results
- case reports

In an overview of operational syndromic surveillance systems [15], it was reported that the 52 systems surveyed monitored markedly the following data sources: emergency department visits (84%), outpatient clinic visits (49%), over-the-counter medication sales (44%), calls to poison control centres (37%), and school absenteeism (35%). Another review by [16] examined 56 systems, with a comparable distribution of data source usage. The availability of data sources depends on the local context of the project: jurisdiction of the organisation responsible for the system, diagnoses to be monitored, existing laws regulating data access, and technical concerns such as ensuring sustained connectivity to the data sources. Recent research suggests that additional sources such as Web search queries [17], [18], and Twitter posts [19] can also contain important clues to the spread of disease.

III. The Opportunity for Innovation

At the Etisalat BT Innovation Center (ebtic.org), the iCare: Next Generation Intelligent Collaborative Health Care Technologies research and development programme targets novel ICT solutions to health care issues. Within iCare, the center is already pursuing several goals within collaborative care, in cooperation with HAAD. The iCare programme consists of two parts: CRISISNET and IC^3. The goal of the former is to develop working prototypes for the prevention, management, and mitigation of disasters and catastrophies, taking a micro-meso-macro approach to modelling, simulation, and surveillance. The goal of the latter is to enable intelligent collaborative care of chronic disease. Both of these goals are highly relevant to innovations like employing syndromic surveillance, and could be pursued here first through feasibility studies and prototyping.

Microdata are today widely employed for understanding the spread of infectious disease, both through the new science of computational epidemiology [20] and through the monitoring of individual ICT use. Macrodata, by contrast, are at the heart of traditional models of infectious disease [21], resting on differential equations and statistics for estimating the expected trends of communicable diseases at the population level. The least studied kind of data is the level in-between. Mesodata are neither individual (agent) data, nor population data, but pertain to small groups or clusters of individuals. The best example is the family. In geographically explicit models, where waves of disease can be traced through a landscape, the family resides in the same dwelling, subject to much the same risks and challenges. In Abu Dhabi, such a meso-level family can have up to 200 members. Since it is often hard to pinpoint individuals to an exact address, novel methods for contact networks could be tested. One such method would be to model the social network for information dissemination, obtained from phone call networks. Such studies have been carried out for other purposes, in other countries, such as the U.K. [22]. Its extremely high usage rates of smart phones—a recent survey points to 43.7% [23]—makes the UAE interesting for harvesting data from tweets and blog comments. The use of Twitter is also increasing in the region, with a daily average (Q1, 2011) of more than 200000 Twitter users in the UAE [24]. Blog comments can be harvested and mapped onto social networks [25], e.g., by using the API to the Twingly blog search engine (cf. [26]). The Nationals constitute a very young (and small: less than 20 per cent) part of the population in Abu Dhabi, with a median age of 18; by contrast, the median age for expatriates is 31 [2]. Hence, the wide and quick uptake of social media among Nationals is not surprising.

In view of the knowledge elicitation problem, it is important to remember that the timeliness of a data source is often inversely proportional to its reliability [27]. Sources with immediate availability such as Twitter posts or search queries often contain large amounts of false signals, and usually lack geographic specificity (cf. [28] [29]). By contrast, laboratory test results provide definitive diagnostic information, but they are not available early.

It can also prove a challenge to find the right terms to search for in text-based syndromic surveillance. A term can here be a word, a bag of words, a hashtag, or something more abstract, like a meme. In the case of influenza-like illnesses, for instance, one might wish to harvest search query strings, and words in blog posts or tweets, such as "cough", "sore throat", "fever", and "flu". In the UAE, the harvesting could not be confined to English, either, given the wide variation of nationalities residing in the UAE. The challenge is not only to find terms that provide data, but also terms that do not provide too much of it. In addition, whatever can be done for positioning a term in the geography, e.g., using GPS, IP numbers, or network data, maximises chances of providing useful data. It could, however, also introduce noise in the data or give false indications of regionality with respect to disease spread. All in all, syndromic surveillance is a set of methods to be used with utmost care, and ideally as a complement to more robust (albeit slower) methods for data collection and analysis.

Finally, health care innovations are (in classic economic language [30]) neither about new technologies nor new things, but about new value for current and future patients. When prototyping new methods for solving the three problems described in the introduction, the final evaluation must be done by the end-users of the resulting health care system. Ideally, such an evaluation would also consider end-user appropriation and external innovation [31].

The knowledge elicitation problem can also be addressed by leveraging on existing data collecting efforts. For example, within the aforementioned Weqaya programme, more than 200000 consenting adults in Abu Dhabi were screened for CVD indicators (representing more than 94% of citizens), and in that process their mobile numbers and email addresses were recorded [32]. With privacy issues and the personal integrity of those individuals being respected, some data mining on that data could very well result in important information, such as the spatial distribution of prevalence for CVD. A memorandum of understanding was signed on June 19, 2011, between HAAD and Etisalat concerning the joint development of a number of electronic health services. The Weqaya screening is planned to take place every three years and the data gathered could naturally be used also to address the prediction problem. All in all, Weqaya screenings and scorecards, together with some of the unique demographical features of the region, present a host of possibilities for innovation [33].

IV. Conclusion

Syndromic surveillance was suggested as a means to employ ICT for tackling three complex and important public health problems faced by the health care providers in the UAE: information dissemination, knowledge elicitation, and prediction. The possibilities and problems associated with its employment were detailed, and some indications of what must follow in terms of prototyping and testing were given. This is work in its early stages of specification and review. Future work will include the application of state-of-the-art syndromic surveillance for communicable disease to a number of test cases, and an evaluation of said tests. As in all computational epidemiology, a cross-disciplinary team involving medical experts and ICT experts is required, and the assembly of such a team has commenced.

Acknowledgment

The authors would like to thank the HAAD epidemiologists at the departments of Public Health & Policy, Communicable Diseases, and Public Health Surveillance, for their kind cooperation. The authors also thank Dr Cother Hajat for important technical assistance. MB would like to thank EBTIC for providing him with a fellowship opportunity and a nice working environment.

References

[1] C. Hajat, O. Harrison, and Z. A. Siksek, “Weqaya: A population-wide cardiovascular screening program in Abu Dhabi, United Arab Emirates,” *American Journal of Public Health*, 2011, in press.

[2] “Abu Dhabi health statistics 2010,” HAAD, October 20 2011, available at www.haad.ae/statistics.

[3] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, “Global prevalence of diabetes: estimates for 2000 and projections for 2030,” *Diabetes Care*, vol. 27, no. 5, pp. 1047–1053, 2004.

[4] P. Russom, “Big data analytics,” The Data Warehousing Institute, Tech. Rep., 2011, tdwi best practices report, Fourth Quarter 2011.

[5] M. Boman, A. Ghaffar, F. Liljeros, and M. Stenhem, “Social network visualization as a contract tracing tool,” in *Proc WS on Agent Technology for Disaster Management*, N. Jennings, Ed., pp. 131–133, Hakodate, Japan.

[6] J. W. Buehler, R. S. Hopkins, J. M. Overhage, D. M. Sosin, and V. Tong, “Framework for evaluating public health surveillance systems for early detection of outbreaks,” *MMWR Morbidity and Mortality Weekly Report*, vol. 53, pp. 1–11, 2004.

[7] V. Sidel, R. Gould, and H. Cohen, “Bioterrorism preparedness: cooptation of public health?” *Medicine & Global Survival*, vol. 7, no. 2, pp. 82–89, 2002.

[8] K. C. Dowling and R. I. Lipton, “Bioterrorism preparedness expenditures may compromise public health,” *American Journal of Public Health*, vol. 95, no. 10, 2005.

[9] F. Mostashari, “Syndromic surveillance: a local perspective,” *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, vol. 80, pp. i1–i7, 2003.

[10] A. Reingold, “If syndromic surveillance is the answer, what is the question?” *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, vol. 1, no. 2, pp. 77–81, 2003.

[11] M. Berger, R. Shiau, and J. M. Weintraub, “Review of syndromic surveillance: implications for waterborne disease detection,” *Journal of Epidemiology and Community Health*, vol. 60, no. 6, pp. 543–550, Jun. 2006.

[12] “Communicable diseases bulletin,” HAAD, December 2010.

[13] B. Cakici, “Disease surveillance systems,” Licentiate dissertation, Royal Institute of Technology (KTH), 2011.

[14] S. Babin, S. Magruder, S. Hakre, J. Coberly, and J. S. Lombardo, “Understanding the data: Health indicators in disease surveillance,” in *Disease Surveillance: A Public Health Informatics Approach*, 1st ed., J. S. Lombardo and D. L. Buckeridge, Eds. Wiley-Interscience, apr 2007, ch. 2, pp. 43–90.

[15] J. W. Buehler, A. Sonricker, M. Paladini, P. Soper, and F. Mostashari, “Syndromic surveillance practice in the United States: findings from a survey of state, territorial, and selected local health departments,” *Advances in Disease Surveillance*, vol. 6, no. 3, pp. 1–20, 2008.

[16] H. Chen, D. Zeng, and P. Yan, *Infectious Disease Informatics: Syndromic Surveillance for Public Health and Bio-Defense*, 1st ed. Springer, Dec. 2009.

[17] A. Hulth, G. Rydevik, and A. Linde, “Web queries as a source for syndromic surveillance,” *PloS One*, vol. 4, no. 2, 2009.

[18] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, “Detecting influenza epidemics using search engine query data,” *Nature*, vol. 457, no. 7232, pp. 1012–1014, Feb. 2009.

[19] V. Lampos, T. Bie, and N. Cristianini, “Flu detector - tracking epidemics on twitter,” in *Machine Learning and Knowledge Discovery in Databases*, J. L. Balcázar, F. Bonchi, A. Gionis, and M. Sebag, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, vol. 6323, pp. 599–602.

[20] B. Cakici and M. Boman, “A workflow for software development within computational epidemiology,” *Journal of Computational Science*, vol. 2, no. 3, pp. 216–222, 2011.

[21] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*. Oxford Science Publications, 1992.

[22] N. Eagle, M. Macy, and R. Claxton, “Network diversity and economic development,” *Science*, vol. 328, pp. 1029–1031, 2010.

[23] “Smartphones constitute 43.7% of total cellular handsets in UAE,” Dubai Chronicle, 2011, oct 19.

[24] R. Mourtada and F. Salem, “Civil movements: The impact of facebook and twitter,” *Arab Social Media Report*, vol. 1, no. 2, 2011.

[25] A. Clauset, “Finding local community structure in networks,” *Physical Review E*, vol. 72, p. 026132, 2005.

[26] O. Gornerup and M. Boman, “The blogosphere at a glance—content-based structures made simple,” presented at the 2011 IJCAI WS on Social Web Mining. Available at http://arxiv.org/abs/0909.4416.

[27] D. L. Buckeridge, J. Graham, M. J. O’Connor, M. K. Choy, S. W. Tu, and M. A. Musen, “Knowledge-based bioterrorism surveillance.” *Proceedings of the AMIA Symposium*, pp. 76–80, 2002.

[28] G. V. Steeg, R. Ghosh, and K. Lerman, “What stops social epidemics,” April 2011, available at http://arxiv.org/abs/1102.1985.

[29] A. Clauset, C. Moore, and M. E. J. Newman, “Hierarchical structure and prediction of missing links in networks,” *Nature*, vol. 453, pp. 98–101, 2008.

[30] J. Schumpeter, *The Theory of Economic Development*. Harvard Univ Press, 1934.
[31] K. J. Boudreau and K. R. Lakhani, "How to manage outside innovation," *MIT Sloan Management Review*, vol. 50, no. 4, 2009.
[32] "Weqaya," HAAD WHO presentation, October 27 2011, available at http://www.who.int/tobacco/mhealth/weqaya.pdf.
[33] C. Hajat and O. Harrison, "The Abu Dhabi cardiovascular program: The continuation of Framingham," *Progress in Cardiovascular Diseases*, vol. 53, pp. 28–38, 2010.