An Extended SIR Model to Explore the Impact of Syndromic Data Sources on Social Distancing Policy

Jim Duggan,
College of Engineering and Informatics,
National University of Ireland, Galway.

Email: jim.duggan@nuigalway.ie

Abstract

Epidemics such as seasonal influenza are a major worldwide public health concern, and therefore early outbreak detection and outbreak management are prioritized goals of public health professionals. Syndromic surveillance focuses on discovering the earliest possible indicators of a health problem, and therefore much of the focus in on pre-diagnostic data. Information technology has created new opportunities for syndromic surveillance, for example, geographical internet search data can now estimate the probability that a random physician visit was related to an influenza outbreak. However, there are also important challenges in adopting this use of new technology, and the potential harmful side-effects (in terms of public confidence) if the real-time data models are not sufficiently robust. This paper presents an exploratory model that captures the dynamics of information quality, and the potential effect of syndromic information quality on social distancing measures.

Introduction

Epidemiology is often considered the core science of public health, and involves the "study of the distribution and determinants of disease frequency" (Rothman 2002), and is concerned with "formulating strategies for managing established illness, as well as for preventing further cases" (Stewart 2010). Epidemics such as seasonal influenza are a major worldwide public health concern, causing tens of millions of respiratory illness and 250,000 – 500,000 deaths globally each year (WHO 2003). Epidemics outbreaks are exponential growth processes where the doubling time in a fully susceptible population could be as low [0.62-1.25] days (Mathews et al. 2007). Factors driving the spread of disease include the population contact rates, the demographic age structure, the strength or the infection, and the infectious time period. Given the dynamics of transmission (as quantified by the reproduction number

R_0 which measures the average number of infections per infected person), early detection is crucial to isolate infected individuals, prime hospital capacity, ramp up on vaccination, and increase possible hygiene and social distancing measures. This challenge of early outbreak detection and management are prioritised goals of biosurveillance (Toner et al. 2011).

Syndromic surveillance focuses on finding and integrating the earliest possible indicators of a health problem, and therefore much of the focus in on pre-diagnostic data. The use of syndromic data to support public health preparedness is a high priority area for many governments, as are soon-to-be-published Public Health Syndromic Surveillance (PHSS) guidelines for meaningful use[1], which will shape syndromic surveillance measures for infectious diseases into the future. Technology also has a vital role to play in addressing this problem, and the broad objective of public health informatics is to support the activities, programs and needs of those entrusted with assessing and ensuring overall public health (Lombardo and Ross 2007), and, health surveillance - "the ongoing systematic collection, analysis and interpretation of outcome-specific data" - is a key enabler of this goal.

There are many sources of syndromic data (Babin et al. 2007), and these include: data from pharmacy chains on recent medication sales; information from emergency medical services; data from telephone triage hotlines; information on school attendance levels; data from hospital and physicians visits, and indications of test volumes (as distinct from results) from laboratories. Additional syndromic information can include environmental data (e.g. water quality levels), and animal health data. In addition to these established information gathering approaches, recent research has indicated that the internet and social media can provide a valuable source for the observation of illness-related information.

Ginsberg et al. (2009) describe a model, based on geographical search data and a record of influenza-like illness (ILI) visits to physicians, that provides estimates of the probability that a random physician visit was related to an ILI. Their model performed remarkably well, and was able to estimate consistently the current ILI percentage 1–2

[1] http://www.syndromic.org/

weeks ahead of the publication of standard CDC's reports. The value of this model is the potential to reduce the *reporting delay* for influenza information, and therefore more rapidly close the feedback loop for decision makers to take action. Further studies such as Collier et al. (2011) found that micro-blogging services such as Twitter offer the potential to crowd source epidemics in real-time, and that their study adds to evidence supporting a high degree of correlation between pre-diagnostic social media signals and diagnostic influenza case data.

However, while the availability of real-time syndromic information for decision makers in the context of rapidly spreading infections has excellent potential, a challenge also remains to ensure the integrity and quality of the resulting information. For structured information sources such as pharmacies and hospitals (where technically the challenge is to integrate existing information systems) this may be less of a challenge, however, the non-structured sources of information such as Twitter feeds and search data place an additional "burden of proof" on technology providers to find robust information mining algorithms that separate "the signal from the noise." For policy makers to rely less than robust syndromic information sources can lead to a "compliance problem" within the general public, which is highlighted in figure 1.

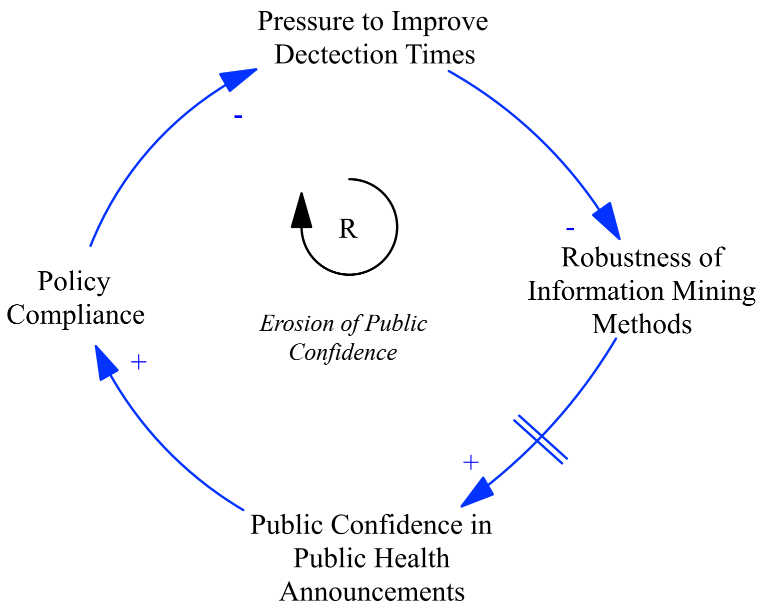

Figure 1: Positive feedback loop highlighting side-effect of less than robust information mining methods

This paper now builds on this idea, and presents a simulation model that models the quality of syndromic information, and its impact on the effectiveness of social distancing measures. The underlying model is the well-document SIR model (Sterman 2000), which can be modeled using individual and agent approaches (Rahmandad and Sterman (2008), Duggan (2008), Duggan (2008b)), and the public confidence sub-model draws on that proposed by Lane and Huseman (2004), where they explored move marketing strategies and the impact of movie quality on release strategies.

The SIR-Syndromic Model

The extended SIR model (see figure 2) has three main elements:

1. The Physical Model

The classic three compartment structure that models "the physics of the virus", and captures the dynamics as the population transitions from Susceptible to Infected, and then becomes Recovered. The key flow equations are summarized in (1) and (2), and the important parameters include:

- contact rate (c), a measure of the population interactions, and measured as *persons/person/day*;
- infectivity (i), a dimensionless measure that captures the ability of the agent of infection to produce disease;
- the recovery delay that measures the average time (exponential distribution) it takes for an individual to recover.

(1)	IR = c * Susceptible * (Infected/N) * infectivity
(2)	RR = Infected / Recovery Delay

In order to facilitate flexibility in scenario generation, the model has special purpose flow equations that allow each stock to be reset at a specific model time to its original value. This allows for the modeling of a new mutated virus, where the recovered population can once again become susceptible. The discrete equation to reset the Susceptible stock to its initial value is shown in (3).

(3)	Reset S = IF THEN ELSE(Time=Model Reset Time,(Init S-Susceptible)/TIME STEP,0)

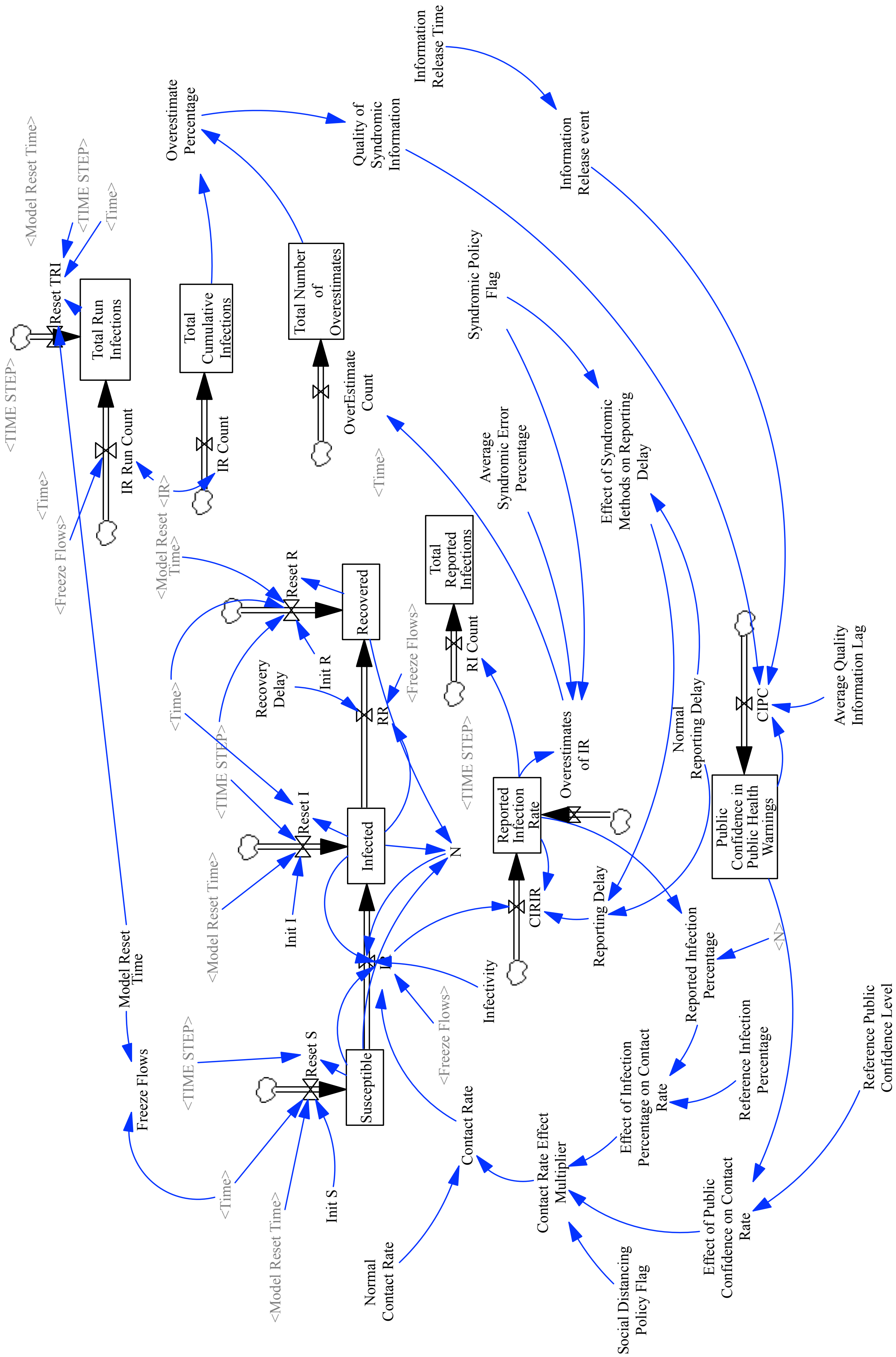

Figure 2: The Extended SIR Model

2. The Social Distancing Feedback Loop

In order to model the impact of higher than normal infection rates, the contact rate can be endogenous (depending on the value of a discrete flag constant), and therefore a public health social distance policy can be modeled. This feedback loop is summarised in figure 3.

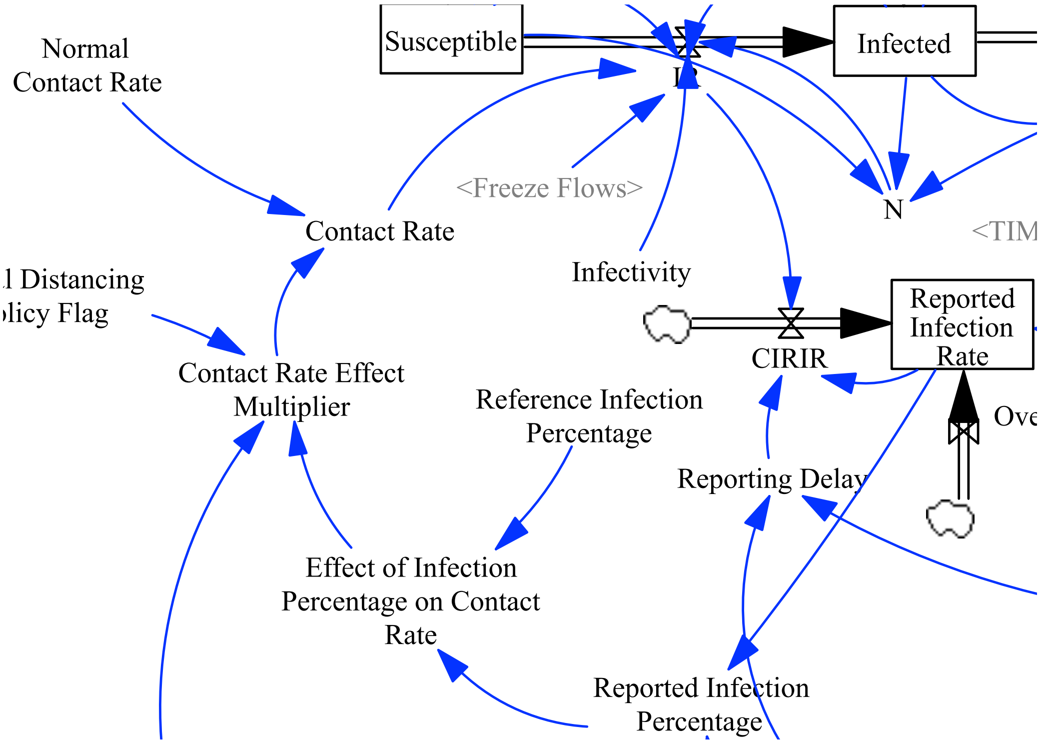

Figure 3: Social Distancing Balancing Loop

For this balancing loop, the *Reported Infection Rate* is a smoothed value of the actual infection rate (1), and this value is governed by the *Reporting Delay.* As we will see in the final part of the model, this reporting delay is important as its value is also influenced by the use of the *Syndromic Policy Flag*. Specifically, if we implement our syndromic policy, the reporting delay will be reduced significantly. The contact rate (4) is a standard effect formulation, which depends on the combined effects of the social distancing policy *and* public confidence in public health announcements.

(4)	Contact Rate = Normal Contact Rate*Contact Rate Effect Multiplier

(5) Contact Rate Effect Multiplier = IF THEN ELSE(Social Distancing Policy Flag=1, Effect of Infection Percentage on Contact Rate*Effect of Public Confidence on Contact Rate,1)

(6) IF THEN ELSE(Reported Infection Percentage<Reference Infection Percentage,1,max(0,1/4*(5-Reported Infection Percentage/Reference Infection Percentage)))

(7) Reported Infection Percentage = Reported Infection Rate/N

(8) Reference Infection Percentage = 0.015

(9) Reporting Delay = Normal Reporting Delay*Effect of Syndromic Methods on Reporting Delay

(10) Effect of Syndromic Methods on Reporting Delay = IF THEN ELSE(Syndromic Policy Flag=1, 1/Normal Reporting Delay,1)

For this model, the effect variable for social distancing (6) is simplified, and has no impact if the *Reported Infection Percentage* is less than the *Reference Infection Percentage,* and otherwise it behaves as a simple linear decreasing function (see figure 4 for sample output from a simulation run). It is assumed that the normal infection percentage is 1.5% of the population (8), and this would be recognised as the base rate of infection for a given population.

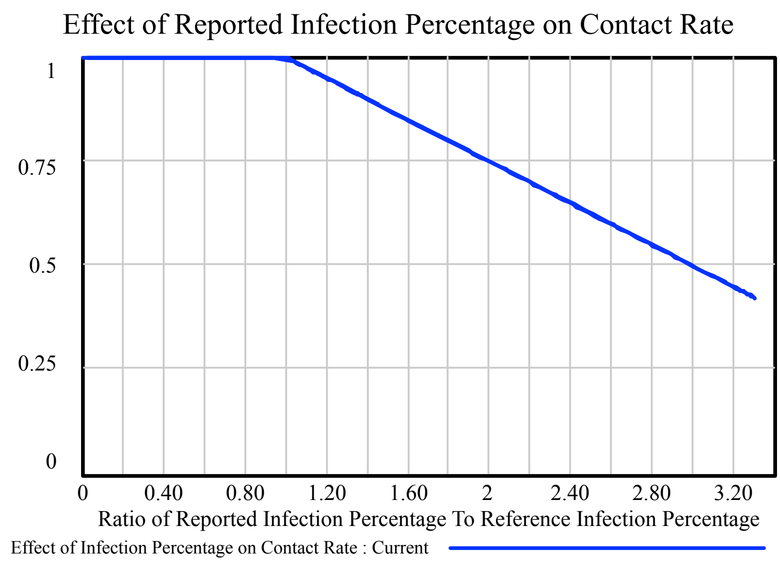

Figure 4: A summary of the effect of infection percentage on contact rate

3. Public Confidence Effect on Social Distancing

The final component of the model captures the impact of public confidence on social distancing measures, which is an important element of equation (5). The idea is that the model contains a true value for the quality of information, but that this value is not readily available to the public (and therefore may not impact on public perceptions.) A snapshot of the important variables are captured in figure 5, and these will now be specified in further detail.

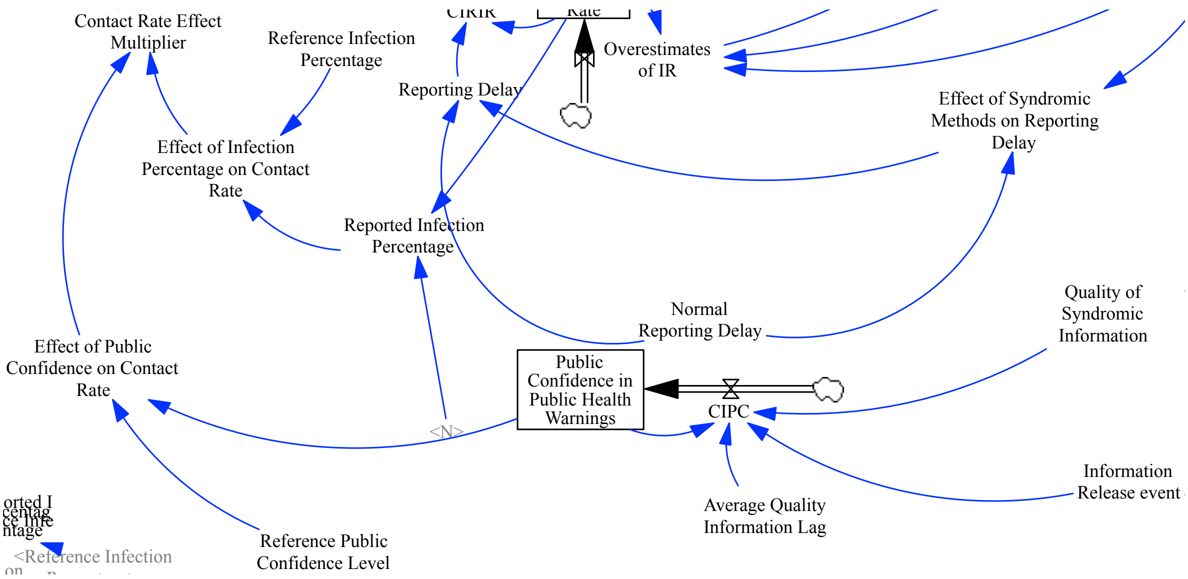

Figure 5: The impact of public confidence on the contact rate

Equation (11) shows the effect variable of public confidence on the contact rate. As with equation (6), this is also a simplified linear function that has no effect when the public confidence is greater than or equal to the threshold level, and otherwise it has an increasing impact (see figure 6) as public confidence falls. The net effect of this is that it nullifies the impact of any social distancing policy, as the level of public compliance drops.

(11)	Effect of Public Confidence on Contact Rate = IF THEN ELSE(Public Confidence in Public Health Warnings>=Reference Public Confidence Level,1,2-(Public Confidence in Public Health Warnings/Reference Public Confidence Level))

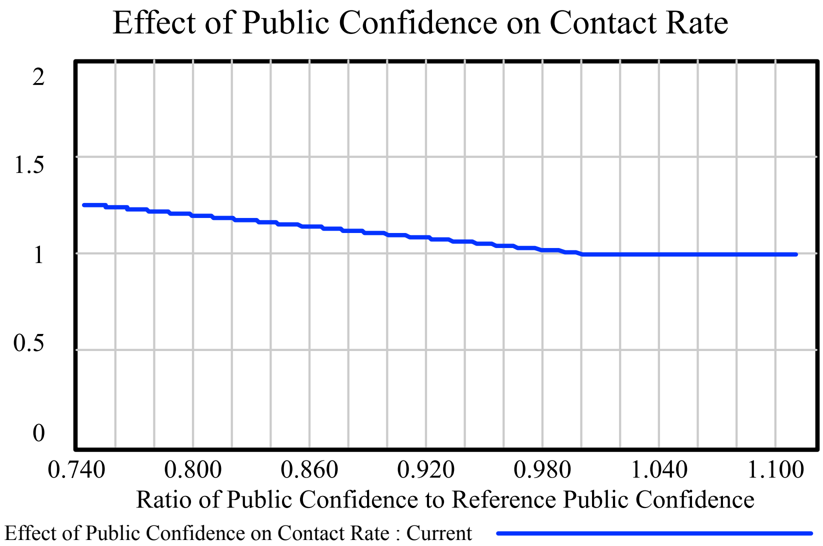

Figure 6: A summary of the multiplier effect of public confidence on contact rate

Public confidence (12) is a stock that is formulated as first order information delay on the variable quality of syndromic information (15), following a given delay constant (14). This change only happens once the information has been released (17). The quality of information is based on a measure of the number of overestimates in the reported infections (as specified in the variable *Average Syndromic Error Percentage*).

(12)	Public Confidence = INTEGRAL(CIPC,100)
(13)	CIPC = IF THEN ELSE(Information Release event=1,(Quality of Syndromic Information-Public Confidence in Public Health Warnings)/Average Quality Information Lag,0)
(14)	Average Quality Information Lag = 10
(15)	Quality of Syndromic Information = (1-Overestimate Percentage) * 100
(16)	Overestimate Percentage = ZIDZ(Total Number of Overestimates,Total Cumulative Infections)
(17)	Information Release event = 0+STEP(1,Information Release Time)
(18)	Information Release Time = 50

Given the model structure, a number of policy scenarios can be run. These include: (1) a reference run (no policies active), (2) social distancing policy only and (3) combined syndromic usage with the social distancing policy.

Initial Experimental Results

While the model is a *proof of concept* – in the sense that it is not calibrated or validated against with a specific real-world structure and data – nonetheless, a number of experiments can be conducted to assess the impact of a syndromic policy on (1) the speed of response to social distancing policy and (2) the long term impact of using lower quality syndromic data in terms of a decrease in public confidence. The simulations capture "two waves" of an epidemic, where the initial conditions are replicated after 50 time units, and for the second wave, the impact of public awareness of syndromic data quality can be assessed. The initial conditions of the model are summarised in table 1, and are based on the values from Sterman (2000, chapter 9, p. 307).

Variable Name	**Variable Type**	**Initial Value**
Susceptible	Stock	9999
Infected	Stock	1
Recovered	Stock	0
Normal Contact Rate	Auxiliary/Constant	6
Infectivity	Auxiliary/Constant	0.25
Recovery Delay	Auxiliary/Constant	2
Average Syndromic Error Percentage	Auxiliary/Constant	0.25
Reporting Delay	Auxiliary/Constant	4
Social Distancing Policy Flag	Auxiliary/Constant	[0|1]
Syndromic Policy Flag	Auxiliary/Constant	[0|1]

Table 1: Initial conditions for experiments

1. Reference Mode Simulation

The reference mode simulation is run for 100 time units, with no policy flags activated. Because of this, the "physics" of the outbreak are modeled, with no feedback loops to modify individual behaviour – for example, a reduction in social distancing. The model reverts to its initial state at time 50, and so the exact same pattern is generated from time 50-100.

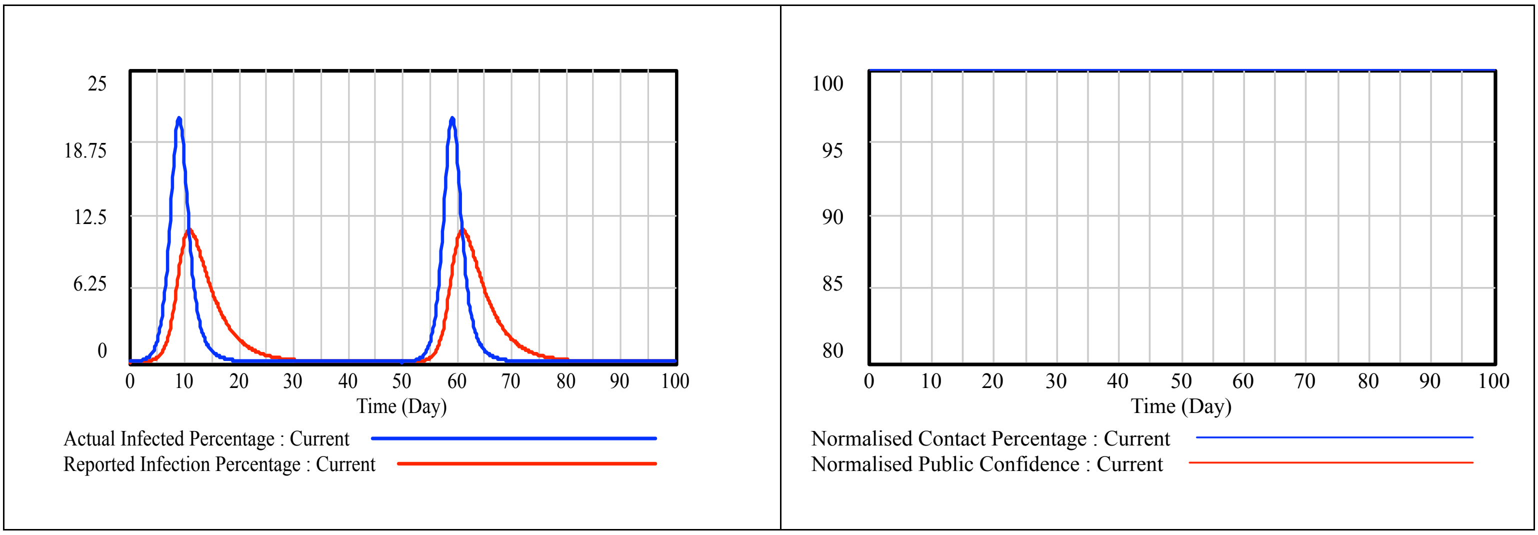

Figure 7: Scenario 1 – no policies active

Figure 7 highlights (LHS) the diffusion of the virus over time, as the actual infected percentage of the population reaches a peak of around 20%, while the contact rates (RHS) and the public confidence remain unchanged.

2. Social Distancing Policy – Social Distancing with no Syndromic Data

For this scenario, social distancing is enabled, and therefore as the reported infection rate rises, the contact rate amongst the population declines. As a result, it is no surprise that the peak infection value falls significantly compared to scenario one. Because the data quality is high (i.e. there is no use of syndromic methods to speed up data collection), public confidence remains unchanged, as we note that public confidence depends on, via an information delay, the overall quality of the syndromic information.

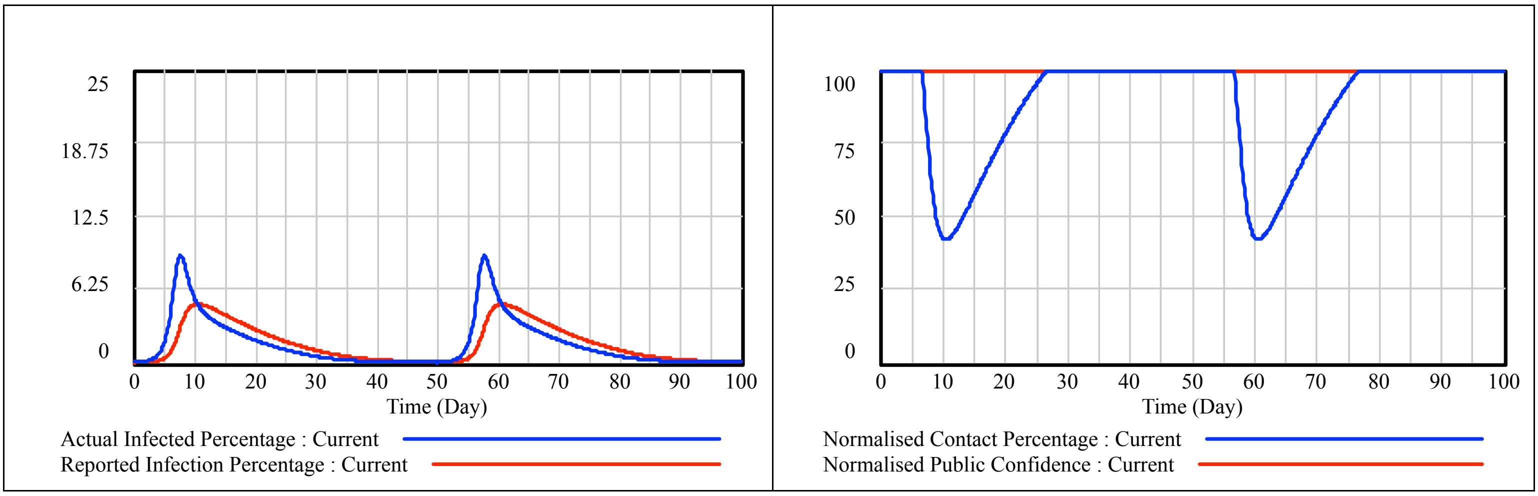

Figure 8: Scenario 2 – social distancing policy active

3. Social Distancing Policy – Syndromic Data and Social Distancing

The final scenario activates syndromic data collection and the social distancing policy. There are a number of observations to make about these results. First, the area under "Actual Infected" and "Report Infected" are different, as the overestimation has added to the reported accumulation (independent of the actual values – note the 2nd flow into Reported Infection Rate as shown in figure 2).

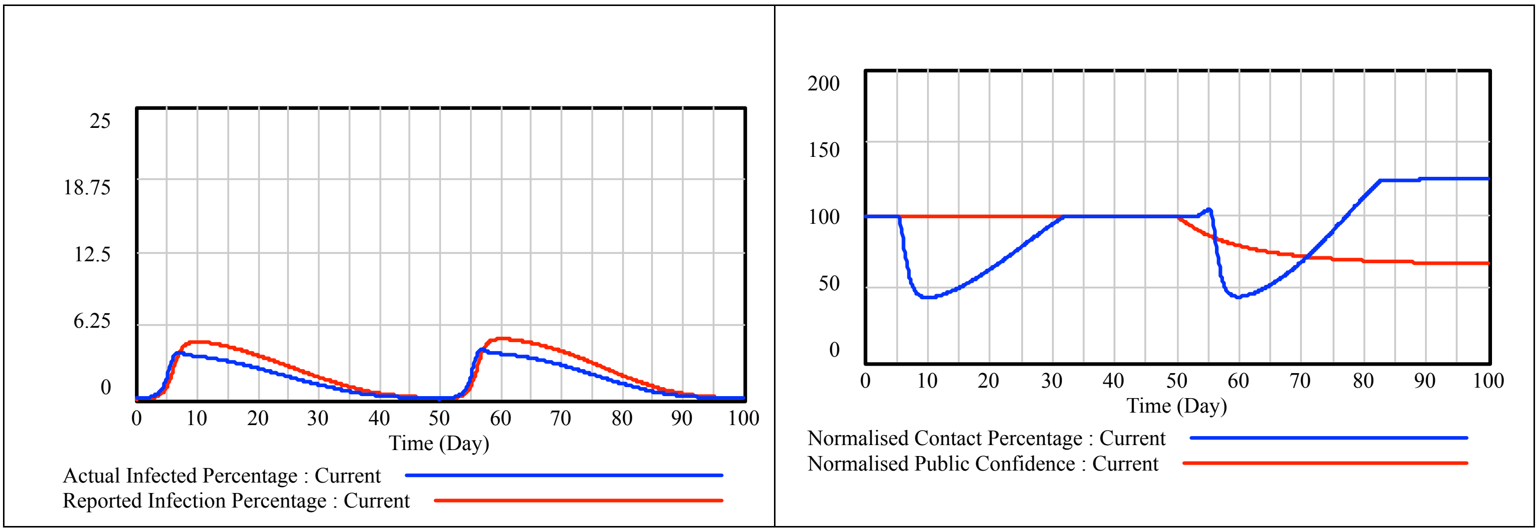

Figure 9: Scenario 3 – social distancing and syndromic data policy active

Also, while the overestimation of reported infections starts early in the model, this does not impact public confidence until such information is “broadcast” at time 50 (see equation 13). Therefore, the undesired side-effect of the syndromic policy only manifests itself in the second outbreak, where we can see the loss in confidence dilutes the social distancing policy and leads to an increase in the contact rate beyond its normal value.

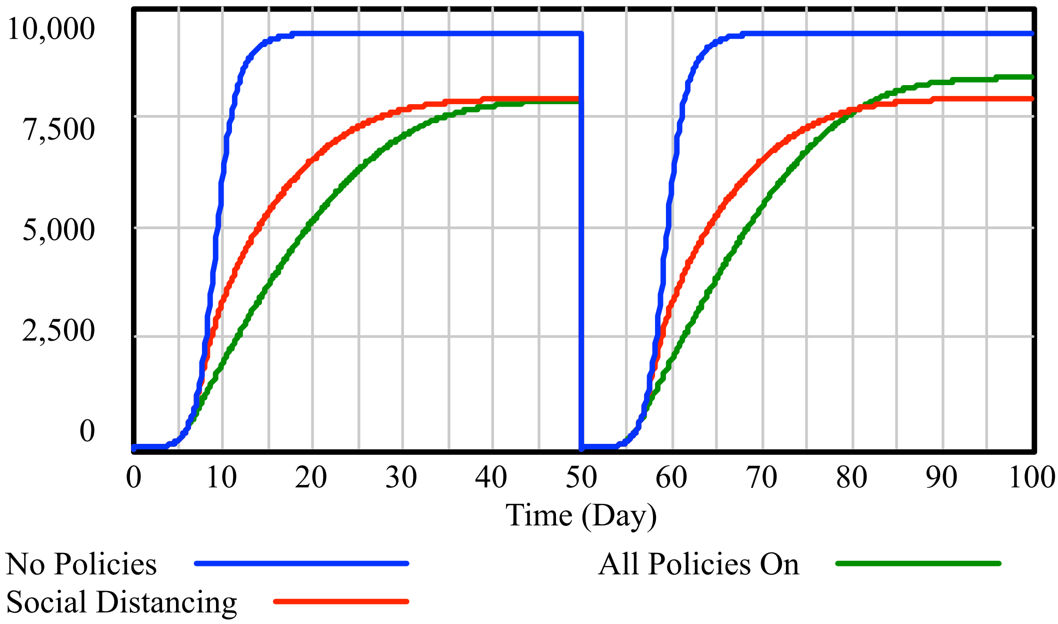

Figure 10: Comparison of Total Infections (Per Outbreak) across the three policies

Finally, figure 10 charts the cumulative total infections for each outbreak, and contracts the final numbers arrived at for each scenario. As expected, scenario one has the largest number of infections, but what is interesting is that for the second outbreak, the interaction of both policies leads to a poorer performance, as public compliance reduces the impact of the social distancing policy. While these results are exploratory, they do provide the means to integrate information systems issues into policy exploration for disease management and control.

Conclusions and Future Work

Syndromic surveillance focuses on finding and integrating the earliest possible indicators of a health problem, and therefore much of the focus in on pre-diagnostic data. The use of information technology to source syndromic information in real time has created new possibilities for public health professionals, with the promise of reducing detection times and therefore enabling early social distancing to combat the exponential processes of disease spread. However, a potential problem with rapidly processed unstructured data is the possibility for under and over-estimation, and the impact this can have on public trust and confidence in public health directives. This paper has presented an exploratory model for capturing the role of information quality on social distancing, and future work will focus on working with public health professionals to gather data to accurately model the impact of trust and confidence on social distancing.

References

Abraham, Thomas. 2011. "Lessons from the pandemic: the need for new tools for risk and outbreak communication." *Emerging Health Threats Journal* 4 (October 17): 1-4. doi:10.3402/ehtj.v4i0.7160. http://eht-journal.net/index.php/ehtj/article/view/7160.

Babin, S., S. MAgruder, S. Hakre, J. Coberly, J.S. Lombardo. 2007. "Understanding the Data: Health Indicators in Disease Surveillance". *Disease Surveillance: A Public Health Informatics Approach,* Edited by Joseph S. Lombardo and David L. Buckeridge, Wiley Interscience.

Collier, Nigel, Nguyen Truong Son, and Ngoc Mai Nguyen. 2011. "OMG U got flu ? Analysis of shared health messages for bio-surveillance." *Journal of Biomedical Semantics* 2 (Suppl 5): S9. doi:10.1186/2041-1480-2-S5-S9.

Duggan, Jim. 2008. "Equation-based policy optimization for agent-oriented system dynamics models". *System Dynamics Review,* Vol. 24. No.1. 97-118.

Duggan, J. 2008. Using System Dynamics and Multiple Objective Optimization to Support Policy Analysis for Complex Systems
Complex Decision Making. H. Qudrat-Ullah, J. M. Spector and P. I. Davidsen, Springer Berlin / Heidelberg. **30:** 59-81.

Ginsberg, Jeremy, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. 2009. "Detecting influenza epidemics using search engine query data." *Nature* 457 (7232): 1012-1014.

Lane, D. C. and Husemann, E. (2004) *Movie marketing strategy formation with system dynamics: towards a multidisciplinary adoption/diffusion theory of cinema-going*. In: Maier, F., (ed.) Komplexität und dynamik als herausforderung für das management. DUV, pp. 179-222. ISBN 3824482185

Mathews, John D, Christopher T Mccaw, Jodie Mcvernon, Emma S Mcbryde, and James M Mccaw. 2007. "A Biological Model for Influenza Transmission" *PLoS ONE* 2 (11). doi:10.1371/journal.pone.0001220.

Rahmandad, Hazhir and John D. Sterman. 2008. "Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models." *Management Science*. vol. 54 no. 5 998-1014.

Rothman, Kenneth J. 2002. *Epidemiology: An Introduction*. Oxford University Press.

Sterman, John D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill Higher Education. Boston, MA.

Stewart, Anthony. 2010. *Basic Statistics and Epidemiology*. Radcliffe, UK.

Toner, Eric S, Jennifer B Nuzzo, Matthew Watson, Crystal Franco, Tara Kirk Sell, Anita Cicero, and Thomas V Inglesby. 2011. "Biosurveillance where it happens: state and local capabilities and needs." *Biosecurity and bioterrorism : biodefense strategy, practice, and science* 9 (4) (December): 321-30.

WHO. 2003. World Health Organization. Influenza fact sheet. (http://www.who.int/mediacentre/factsheets/2003/fs211/en/)