Early Warning Signals of Vaccine Scares

by

A. Demetri Pananos

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Applied Mathematics

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

There exists strong evidence that vaccines are extremely effective in the prevention of pediatric infectious disease, yet despite this evidence, vaccine refusal is still popular amongst some parents. Existing mathematical models of disease and vaccinating dynamics are parsimonious with post scare empirical data, yet lack the ability to predict when a scare will occur. This thesis frames the problem of predicting when a vaccine scare is imminent as a problem in bifurcation theory and critical transitions theory. As a system of differential equations nears a bifurcation point, the system experiences critical slowing down, a loss in resilience to small perturbations from equilibrium. This loss in resilience manifests itself as an increase in the variance and lag-1 autocorrelation of the time series. Using an existing model for vaccinating dynamics, I demonstrate this critical slowing as the system bifurcates from a state of high vaccine coverage to a state of suboptimal vaccine coverage. I also demonstrate that critical slowing can be detected in the population by using the social media site Twitter and Google Trends data. It will be shown that leading to the 2014 measles outbreak in Disneyland, a statistically significant increase in the lag-1 autocorrelation in the time series of tweets with anti-vaccine sentiment is detected. Post outbreak, a statistically significant decrease in the lag-1 autocorrelation is detected, suggesting that population comes sufficiently close to a vaccine scare to elicit outbreaks in those individuals who ceased vaccinating. The results of this thesis provide new tools to monitor vaccine sentiment and maintain vaccine coverage.

Acknowledgements

Special thank you to Chris Bauch for his guidance and supervision. Graduate school would have been very different without you. Thank you to my collaborators for their contributions to any resulting publications. Thank you to Rob Corless for helping shape my mathematical writing. Rob, I still use the Lambert W Function wherever I can.

Thank you to my lab members for sitting with me and reading thousands of tweets to create the training set. Also, thank you for entertaining my mathematical ramblings, as well as enduring my outbursts after a game of chess went south on me.

Last but certainly not least, thank you to all my friends for their endless support. I don't have siblings, but you are the closest thing I will ever have to brothers and sisters.

"Don't be an engineer. You will never get a job."

– My Mother

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Mathematical Epidemiology

Mathematical epidemiology is the application of tools in dynamical systems and applied mathematics to the study of how diseases spread. In this section, I motivate the study of epidemiology through a mathematical lens. I then go on to formulate a mathematical model of epidemics, describing the details of the formulation. I then discuss disease transmissibility and how disease transmissibility can be used to determine the size of the epidemic, demonstrating that in the absence of numerical solvers, the model still provides a wealth of inference.

1.1.1 An Advantage Over Traditional Methods

Mathematical epidemiology differs from most sciences in that it does not lend itself to experimental validation of models. Experiments would be costly, impossible to replicate, and would above all else be completely unethical. On the other hand, computational experiments can be run orders of magnitude faster than any physical experiment, and can be repeated multiple times for a variety of conditions. The inability to perform traditional experiments, coupled with the speed and flexibility of computation, underscores the importance of mathematical and computational models of disease dynamics. The first mathematical model of disease dynamics was introduced by Kermack & McKendrick in their seminal paper *A contribution to the mathematical theory of epidemics* [1].

1.1.2 The SIR Model

Models of disease dynamics are often illustrated using *compartmental diagrams*. These illustrations divide the population into compartments and make assumptions about the nature and time rate of transfer between them. Shown in figure 1.1 is the compartmental

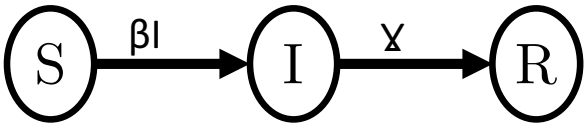


Figure 1.1: The SIR compartmental diagram for Kermack & McKendrick's model. Individuals move from compartment to compartment in the direction of the arrows.

diagram for the Kermack-McKendrick model, known commonly as the SIR model [1]. In the SIR model, individuals are classified as either:

- *Susceptible*: These individuals are presently not infected and do not have immunity to the disease. They may become infected if they come in contact with an infected individual.
- *Infected*: These individuals are presently infected and infectious (i.e they are able to spread the disease through contact with an individual in the susceptible compartment, but not with contact with an individual from the recovered compartment).
- *Recovered*: These individuals have no effect on transmission dynamics. They have left the infected compartment through recovery and are immune to new infections.

Several assumptions about the population and time rate of transfer between compartments are made:

a) The size of the population remains constant throughout the epidemic. Let the size of the population be N, the number of susceptible individuals in the population be S, the number of infected individuals in the population be I, and the number of recovered individuals be R. Therefore, since the population is partitioned into the three compartments, $S + I + R = N$. Often, it will be the case that $N = 1$ so that the size of the compartments are understood to be proportions of the population.

b) An average member of the population makes contact sufficient to transmit infection with βN other members of the population. This assumption is known as *mass action mixing* [2, 3]. The parameter β is called the infectious contact rate (i.e the rate of infection per susceptible an per infective) [3].

c) Members of the infected compartment recover at a rate of γI. Here, γ is the rate of recovery [3].

Assumption (a) asserts that the timescale of the epidemic is much smaller than the timescale for births and deaths. This is especially applicable to epidemics like influenza, which may only last a few months. Assumption (b) implies that the rate of new infections is the product of the probability of an infected individual encountering a susceptible (S/N), the number of new infections per unit time (βN), and the number of infected individuals in the population (I). Thus, the rate of new infections is $(\beta N)(S/N)I = \beta SI$. Assumption (c) requires more mathematical finesse. Consider those members of the population who were all infected at one time, and let $y(t)$ denote the number of these individuals who are still infected after t time units after having been infected. If a proportion γ of these individuals leave in unit time, then

$$y' = -\gamma y \, ,$$

which is a separable differential equation with solution $y(t) = y(0)e^{-\gamma t}$. The function $y(t)$ is the proportion of individuals infected at the same time who remain infectious at time t. Hence, the time to recovery from infection is exponentially distributed. The mean of this distribution is thus $1/\gamma$, and hence the mean length of the infectious period is $1/\gamma$ [2].

1.1.3 SIR as a Differential Equation Model

With the rates of infection and recovery in hand, as well as the flow between compartments from the compartmental diagram, a system of differential equations describing the density as a function of time for each compartment can be created.

$$S' = -\beta SI \,, \tag{1.1}$$

$$I' = \beta SI - \gamma I \,, \tag{1.2}$$

$$R' = \gamma I \,. \tag{1.3}$$

Here, the primes indicate derivatives with respect to time, t. Closed form solutions for S, I, and R as functions of time cannot be found, and so to gain insight into the dynamics as functions of time, numerical methods can be used. See figure 1.2b for numerical solutions to the SIR model. Though no closed form solutions can be found, several inferences can be made directly from the system.

1.1.3.1 The Basic Reproductive Ratio

It has been shown that the expected length of the infectious period is $1/\gamma$, and that βN possible disease transferring contacts are made by an average member of the population. Thus, $\beta N/\gamma$ new infections are produced by an index case in a population sufficiently dense in susceptible individuals. This dimensionless quantity is of utmost importance in mathematical epidemiology, so important that it is given a formal name: the basic reproductive ratio, $\mathcal{R}_0$[1]. The size of $\mathcal{R}_0$ determines if the disease becomes epidemic in the population. Let $N = 1$, and consider the non-dimensionalization of equations (1.1) to (1.3)

[1] $\mathcal{R}_0$ is pronounced *R-nought* and is much less of a mouthful than *the basic reproductive ratio*

by letting $\tau = \gamma t$. The non-dimensionalized system is then

$$S' = -\mathcal{R}_0 SI \,, \tag{1.4}$$

$$I' = \mathcal{R}_0 SI - I \,, \tag{1.5}$$

$$R' = I \,. \tag{1.6}$$

Here, the primes indicate derivatives with respect to the non-dimensional time τ. Fixed point analysis shows that the disease free equilibrium $(S^*, I^*, R^*) = (1, 0, 0)$ is unstable only when $\mathcal{R}_0 > 1$. In other words, a disease will become epidemic in a population sufficiently dense in susceptible when an index case of the disease causes more than one new infection.

1.1.3.2 Solutions in Phase Space

Clever manipulation of equations (1.4) to (1.6) yields solutions in the $S-I$ plane. Note first that the differential equation for R can be decoupled from the system, as R is determined when S and I are known, reducing the dimension of the system from 3×3 to 2×2. By leveraging the chain rule, it can be shown that

$$\frac{dI}{dS} = -1 + \frac{1}{\mathcal{R}_0 S} \,, \tag{1.7}$$

which can be integrated to obtain level curves in the $S - I$ plane [2, 3]

$$I(S) = -S + \frac{1}{\mathcal{R}_0} \ln(S) + C \,. \tag{1.8}$$

Here, $C = I(0) + S(0) - \dfrac{1}{\mathcal{R}_0} \ln(S(0))$. Since the initial conditions must be positive for physical reasons, there are no domain issues with the logarithm in the definition of C. See figure figure 1.2a for level curves of $I(S)$ for varying $\mathcal{R}_0$. With some determination, the final size of the epidemic as a function of $\mathcal{R}_0$ can be obtained. Let S_{fin}, and R_{fin} be the final sizes of the compartments S and R once the epidemic finishes. At the end of

the epidemic, note that $S_{fin} + R_{fin} = 1$. The final size of the epidemic is achieved once $I(S) = 0$. Massaging $I(S) = 0$ into

$$-\mathcal{R}_0 S_{fin} \exp(-\mathcal{R}_0 S_{fin}) = -\mathcal{R}_0 \exp(-\mathcal{R}_0 C) \, .$$

The equation above is now in the form of

$$f(S_{fin}) \exp(f(S_{fin})) = z \, .$$

The inverse of this function is $f(S_{fin}) = W(z)$, where W is the Lambert W function [4]. Applying the Lambert W function yields

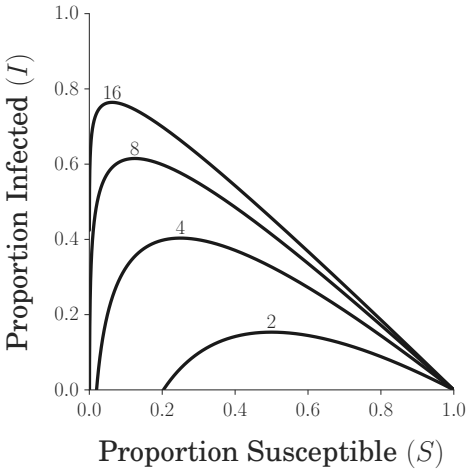
$$-\mathcal{R}_0 S_{fin} = \mathrm{W}\left(-\mathcal{R}_0 \exp(-\mathcal{R}_0 C)\right) \, ,$$

$$S_{fin} = \frac{-\mathrm{W}\left(-\mathcal{R}_0 \exp(-\mathcal{R}_0 C)\right)}{\mathcal{R}_0} \, ,$$

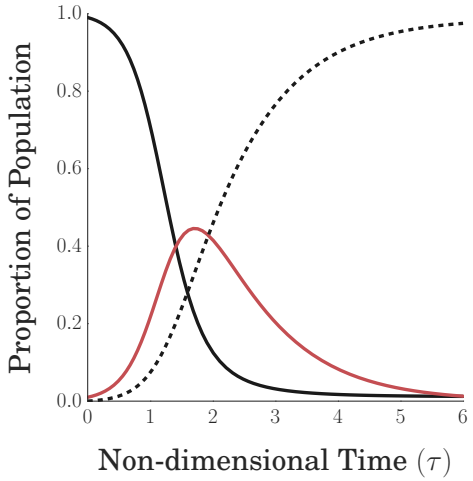
and since $S_{fin} + R_{fin} = 1$,

$$R_{fin} = 1 + \frac{\mathrm{W}\left(-\mathcal{R}_0 \exp(-\mathcal{R}_0 C)\right)}{\mathcal{R}_0} \, .$$

Note that the argument of the Lambert W function in the above expression tends to 0 from below as $\mathcal{R}_0$ becomes arbitrarily large, and that $W(x) \to 0^-$ as $x \to 0^-$ [4]. It is now clear to see that as $\mathcal{R}_0 \to \infty$, $R_{fin} \to 1^-$, indicating that the final size of the epidemic will approach the size of the population as $\mathcal{R}_0$ increases in magnitude. In the absence of closed form expressions for S, I, and R, the model equations allow for inferences to be made about the size of the infection, and the number of expected new infections. The SIR model can be altered so as to account for other epidemiological and biological phenomena, including but not limited to: Exposure to a disease, quarantine, vector-borne diseases, spatial dynamics, and meta-populations. In the following sections, the SIR model will be altered so as to account for births, deaths, and vaccination.



(a) Curves of $I(S)$ for varying $\mathcal{R}_0$. As $\mathcal{R}_0$ increases, each infection causes more secondary infections, thus increasing the final size of the epidemic. These level curves correspond to $C = 1$. The $\mathcal{R}_0$ for each curve is indicated.



(b) Approximate solutions to the SIR model with $\mathcal{R}_0 = 4.5$ obtained via fourth order Runge-Kutta methods as implemented by MATLAB 2015b. In solid black is $S(t)$, in solid red is $I(t)$, in dashed black is $R(t)$. Initial conditions are $S(0) = 0.99$, $I(0) = 0.01$, $R(0) = 0$

Figure 1.2: Evolution of the relative risk as a function of time.

1.2 Vaccination

In this section, I demonstrate the value of vaccination in reducing infectious disease incidence and sometimes eliminating infectious disease altogether. I will motivate an SIR model of endemic infection and investigate how vaccination effects disease prevalence. I will conclude with the phenomena known as herd immunity, demonstrating that herd immunity can not only prevent a disease from becoming endemic, but can prevent an epidemic as well.

1.2.1 An Endemic Model with Vaccination

Endemics differ from epidemics insofar as disease prevalence is sustained through time. In the SIR model, disease prevalence tends to zero as time becomes sufficiently large. In endemic models, the density of infectives at equilibrium will be non-zero and positive. To demonstrate the difference between endemics and epidemics, consider an SIR model with births, deaths, and vaccination. To simplify the analysis, suppose the following:

a) The per capita birth rate is equal to the per capita death rate. This will assumption will allow the population size to remain constant, allowing $S + I + R = 1$.

b) There is no natural immunity to the disease, and no vertical transmission. This means that newborns only enter the S compartment

c) A fraction $0 < p < 1$ of newborns are vaccinated, and thus are moved to the R compartment.

d) The vaccine is perfectly effective. Once vaccinated, immunity is guaranteed, and there is no possibility of vaccine related morbidity.

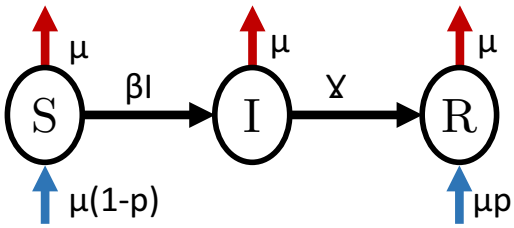


Figure 1.3: The compartmental diagram for the SIR model with births, deaths, and vaccination.

The compartmental diagram for this system is shown in figure 1.3, and the model equations are

$$S' = \mu(1-p) - \beta SI - \mu S\,, \tag{1.9}$$

$$I' = \beta SI - \gamma I - \mu I\,, \tag{1.10}$$

$$R' = \mu p + \gamma I - \mu R\,. \tag{1.11}$$

Here, the primes indicate derivatives with respect to time, t. As before, R is determined when S and I are known, so R can be decoupled from the system. Using arguments from the previous section, it can be shown that $\mathcal{R}_0 = \beta/(\gamma+\mu)$. Fixed point analysis yields two equilibria: A disease free equilibrium $(S^*, I^*) = (1-p, 0)$ stable only when $\mathcal{R}_0(1-p) < 1$ and an endemic equilibrium $(S^*, I^*) = (\,1/\mathcal{R}_0, -\mu(\mathcal{R}_0 p - \mathcal{R}_0 + 1)/(\mathcal{R}_0(\mu+\gamma))\,)$. Using the Routh-Hurwitz Stability Criterion [3], the endemic equilibrium is stable when the following two inequalities are satisfied:

$$\frac{\mu}{\mu+\gamma}\mathcal{R}_0(p-1) < 0\,,$$

$$0 < -\frac{\mu}{\mu+\gamma}(1-\mathcal{R}_0(1-p))\,.$$

Note that since $0 < p < 1$ is a proportion, the first inequality is always satisfied. The second inequality is satisfied when $1 < \mathcal{R}_0(1-p)$, or equivalently, when

$$p < 1 - \frac{1}{\mathcal{R}_0} = p_{crit} \, .$$

1.2.2 Herd Immunity

When $p = p_{crit}$, the disease free equilibrium and the endemic equilibrium collide, and when $p_{crit} < p \leq 1$, the endemic equilibrium leaves the feasible space and destabilizes The collision of equilibra and subsequent exchange of stability happens for $p < 1$, demonstrating that *in order to prevent a disease from becoming endemic, only sufficiently many people need be vaccinated, and not the entire population.* The phenomenon whereby disease spread can be limited if individuals are mostly immune is called *herd immunity* [5, 6]. When herd immunity is achieved, the population is so dense in members immune (or recovered, in the context of the SIR model) to the disease, that an index case has a difficult time finding individuals to infect. Thus, the spread of a disease is curbed by the immunity that the herd provides for those unable or unwilling to vaccinate. See figure figure 1.4 for a visualization of the threshold for herd immunity as a function of the basic reproductive ratio.

Consider now a case where there is no disease present in the population but individuals continue to vaccinate at rate p. The introduction of an index case will cause an epidemic, and if $p < 1 - 1/\mathcal{R}_0$ the disease may remain endemic. Under what circumstances can an epidemic be prevented due to vaccination? Prevention of an epidemic will occur on a short time scale, and so to examine this question, consider the SIR model once more. The differential equation governing the density of infectives is

$$\frac{dI}{dt} = \beta SI - \gamma I \, .$$

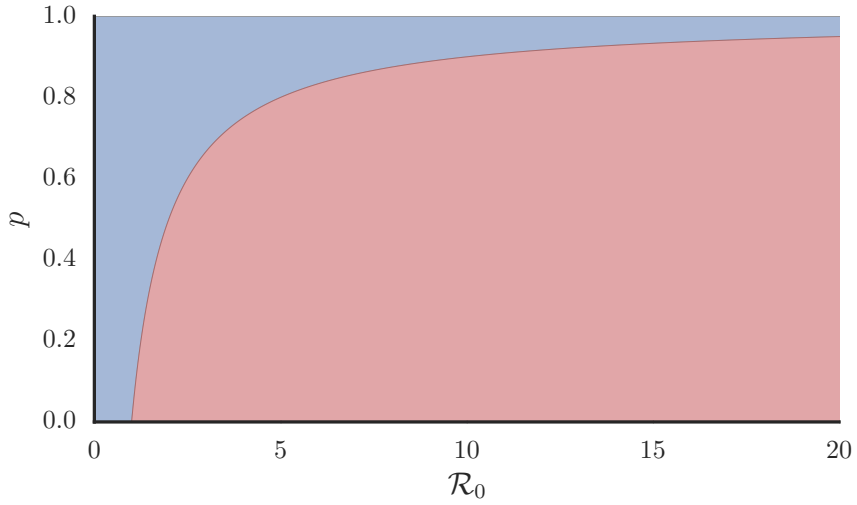


Figure 1.4: In blue is the region in the $\mathcal{R}_0 - p$ plane where herd immunity is achieved. In red is the region where a disease will remain endemic

In order to prevent an epidemic $dI/dt < 0$ for all $0 \leq t$. Equivalently,

$$\beta SI - \gamma I < 0$$

$$I(\mathcal{R}_0 S - 1) < 0$$

$$\mathcal{R}_0 S - 1 < 0$$

$$S < \frac{1}{\mathcal{R}_0}$$

In order for vaccination to prevent an epidemic, the density of susceptible at the time the index case is introduced to the population must be less that $1/\mathcal{R}_0$. Since S is a decreasing function in the SIR model, if $S(0) < 1/\mathcal{R}_0$ then $S(t) < 1/\mathcal{R}_0$ for all $0 < t$. This means that the density of removed individuals must be greater than or equal to $1 - 1/\mathcal{R}_0$, which was shown to be the threshold for herd immunity. Thus, vaccination can prevent a disease from becoming endemic in the population, and can also prevent an epidemic. Vaccination presents tremendous value in disease prevention and public health, but the model given

by equations (1.9) to (1.11) ignores an important aspect of vaccination; risk. In North America, parents are not required to immunize their children They may choose to refuse vaccination for religious or personal reasons, or may choose not to vaccinate because of rumours of vaccine related morbidity. It can be shown through a game theoretical analysis that if a non-zero risk is associated with vaccination, it is impossible for the vaccination rate, p, to meet the required threshold for herd immunity [7]. Thus, a model accounting for vaccinating choices must be developed. Before such a model is put forth, the reasons why a parent may forgo or postpone vaccination must be examined. In the following section, vaccine scares and their motivators are discussed.

1.3 Vaccine Scares

A vaccine scare is a sudden drop in the vaccine coverage due to an increase in risk (actual or perceived) associated with vaccination. Most recently, the Measles, Mumps, and Rubella vaccine (henceforth referred to as the MMR vaccine) has undergone a vaccine scare in the United Kingdom as rumours of an association with autism have surfaced [8, 9]. In this section, I deconstruct vaccine scares, examining how they are caused and how they affect disease incidence. At the end of this section, I present a formalization of vaccinating dynamics via game theory and a differential equation model.

1.3.1 Salient Factors in Determining Vaccinating Habits

Vaccine refusal begins with a lack of confidence in the safety or effectiveness of a vaccine. Though decisions regarding vaccination are multifactorial with high variability from individual to individual, factors affecting vaccination habits can be broadly placed into one of three categories: perceived disease risk, perceived vaccine risk, and social norms.

Vaccines have become victims of their own success. Control over once common pediatric infectious diseases has lead to an unfamiliarity with these diseases, an underestimation of the risks of the disease, and a lack of confidence in the efficacy and safety of vaccines [10, 11, 12, 13]. As seen in section 1.2.2, a population sufficiently dense in immune individuals can eradicate a disease via herd immunity. This poses a paradox between optimal vaccinating habits in order to eradicate a disease and optimal vaccinating habits for the individual. As vaccine coverage climbs, the incentive to vaccinate drops, as the benefits of immunization can be achieved via herd immunity without any of the costs or risks. These acts of self preservation make eradication of a disease difficult under compulsory vaccination [7, 14, 15]. Similarly, an increase in disease prevalence can cause a redoubling effort of vaccination, as

individuals are reminded of the dangers of the disease [7]. Leading up to 1874, Stockholm Sweden saw vaccine coverage drop as low as 40% until a smallpox epidemic shocked the city into widespread vaccination [16]. After rumours surfaced in the 1970s that the whole cell pertussis vaccine caused neurological damage, Wales and Great Britain saw a decrease in vaccine coverage until disease incidence spiked in unvaccinated individuals [14, 17]. Hence, *ceteris paribus*, disease incidence is a salient factor in determining vaccinating habits.

Risks associated with vaccination are also salient factors when determining vaccinating habits. The most notable and visible example is the association of the MMR vaccine with autism and gastrointestinal problems [14, 10, 13]. Parents have noticed a population level increase in autism diagnosis as the number of vaccine administrations and recommended vaccinations has coincidently and simultaneously increased. As parents see diagnosis of infants with autism after vaccination, they form a causal relationship of "vaccines cause autism". However, this causal inference is in fact fallacious, and readers will recognize this rhetoric as *post hoc ergo propter hoc*. There exists a confluence of other factors relating to perceived risks of vaccines: The unpredictability of adverse reactions to vaccinations, waning trust in corporations, public health agencies, and the pharmaceutical-industrial complex (i.e Big Pharma), and a demand for a more "natural" lifestyle all contribute to perceived vaccine risk [10]. Despite these concerns, there exists no evidence of a relationship between autism and the MMR vaccine [18, 19, 20].

Finally, the interactions between individuals, especially peers, has a high impact on vaccinating habits. Vaccine refusal is the exception and not the rule in most developed countries. Polio eradication has been maintained for decades in European countries with voluntary vaccination [21], and vaccine coverage for MCV1 (the first dose of the measles containing vaccine) has been well over 93% for the last ten years in countries like Portugal, Sweden, Germany, and The Netherlands [22]. Though vaccine refusal can be problematic

during times of low disease incidence, local elimination can be maintained, suggesting that perceived disease risk and perceived vaccine risk are not the only salient factors in determining vaccinating habits. The Health Belief Model posits that social factors, specifically social norms, play an important role in determining vaccinating habits [23, 24]. Social norms are behaviours considered "normal" that are enforced by some form of social punishment [25]. With respect to vaccination, punishments may include being type cast with others violating the norm as an irresponsible or a bad parent [26]. It is evident from the literature that childhood vaccination is a social norm. Participants in one particular study thought of childhood vaccination as "normal", "the right thing to do", and "natural [...] you're just supposed to do it" [25]. Hence, pressure from social norms may outweigh any risks or costs of vaccination, causing conscientious objectors to follow the norm and vaccinate themselves or their children

1.3.2 Modelling Vaccine Scares

Previous models of vaccinating dynamics did not account for social norms [14, 7]. These models often predicted oscillations in vaccine coverage, which was not observed in vaccine coverage data. Oraby, Thampi, and Bauch have extended an imitation dynamic model of vaccination to account for social norms [15].

For the imitation dynamic, it is assumed that each individual in the population plays one of two strategies: vaccinating or non-vaccinating. The members sample one another at some constant rate and compare payoffs for their strategy. If two members playing different strategies sample one another, then the member with lower payoff will change strategies with probability proportional to the difference in payoffs [27, 28, 29]. The payoff to vaccinate is given by

$$e_v = -r_v + \delta_0 x \, .$$

Here, r_v is the perceived risk associated with vaccinating, x is the proportion of the population who chooses to vaccinate, and δ_0 is the strength of the social norms. The payoff not to vaccinate is given by

$$e_n = -cI(t) + \delta_0(1 - x) \, .$$

Here, c is the product of cost of infection, reporting probability, and a proportionality constant governing perceived probability of becoming infected. The function $I(t)$ is the density of infective individuals in the population at time t. The difference in payoffs for both strategies is then

$$\Delta E = e_v - e_n = -r_v + cI(t) + \delta_0(2x - 1) \, ,$$

which after some rescaling can be written as

$$\Delta E = -w + I(t) + \delta(2x - 1) \, .$$

Here, $w = r_v/c$ is a rescaled measure of how risky the vaccine is in comparison to infection. Since perceived vaccine risk is dynamic in time, w is usually a function of time (i.e $w = w(t)$). When $\Delta E > 0$ non-vaccinators will switch strategies to become vaccinators, and when $\Delta E < 0$ vaccinators will switch strategies to become non-vaccinators.

The imitation dynamic has vaccinators (x) sampling non-vaccinators ($1 - x$) at some constant rate (k) with the sign of ΔE determining if vaccinators increase or decrease. Hence, the differential equation for the rate change in vaccinators with respect to time can then be written as

$$x' = kx(1 - x)(-w + I(t) + \delta(2x - 1)) \, .$$

Here, the prime indicates a derivative with respect to time, t, and k is the rate at which individuals sample one another. This equation is formally equivalent to those in evolutionary game theory [29], except that the payoffs are functions of time. Using a differential

equation model similar to the one found in section 1.2.1, $I(t)$ can be determined. Thus, the entire model is

$$S' = \mu(1-x) - \beta SI - \mu S \,, \tag{1.12}$$

$$I' = \beta SI - (\gamma + \mu)I \,, \tag{1.13}$$

$$x' = kx(1-x)(-w + I + \delta(2x-1)) \,. \tag{1.14}$$

When the relative risk is constant (i.e $w(t) = m$), model equilibria can be computed. There exist five feasible equlibria. Three of the five have zero disease incidence, corresponding to full susceptibility and no disease incidence

$$\vartheta_1 \coloneqq (S^*, I^*, x^*) = (1, 0, 0) \,,$$

full vaccine coverage with no susceptibility,

$$\vartheta_2 \coloneqq (S^*, I^*, x^*) = (0, 0, 1) \,,$$

and partial vaccine coverage

$$\vartheta_3 \coloneqq (S^*, I^*, x^*) = \left(\frac{\delta - m}{2\delta}, 0, \frac{\delta + m}{2\delta} \right) \,.$$

The remaining two equilibria are endemic equilibria with no vaccine coverage

$$\vartheta_4 \coloneqq (S^*, I^*, x^*) = \left(\frac{1}{\mathcal{R}_0}, \frac{\mu}{\mu + \gamma} \left(1 - \frac{1}{\mathcal{R}_0} \right), 0 \right) \,,$$

and partial vaccine coverage,

$$\vartheta_5 \coloneqq (S^*, I^*, x^*) = \left(\frac{1}{\mathcal{R}_0}, \frac{\mu(m - \delta + (2\delta/\mathcal{R}_0))}{\mu - 2\delta(\mu + \gamma)}, \frac{\mu(1 - 1/\mathcal{R}_0) - (\delta + m)(\mu + \gamma)}{\mu - 2\delta(\mu + \gamma)} \right) \,.$$

Here, $\mathcal{R}_0$ has its familiar definition as $\mathcal{R}_0 = \beta/(\mu + \gamma)$.

1.3.3 Stability & Bifurcations of the Oraby *et al.* Model

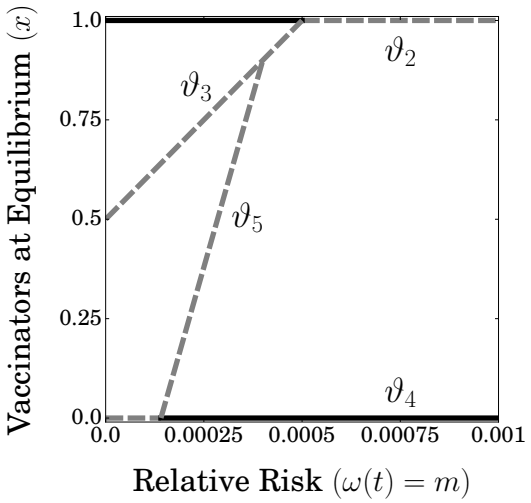
The stability of each equilibria is shown in table 1.1. See figure 1.5 for bifurcation[2] diagrams for $x(t)$.

Equilibrium	**Stability Condition**
ϑ_1	$\mathcal{R}_0 < 1$
ϑ_2	$m < \delta$
ϑ_3	Never stable in feasible space.
ϑ_4	$\frac{\mu}{\mu+\gamma}(1 - 1/\mathcal{R}_0) - \delta < m$
ϑ_5	$2\delta < \frac{\mu}{\mu+\gamma}$

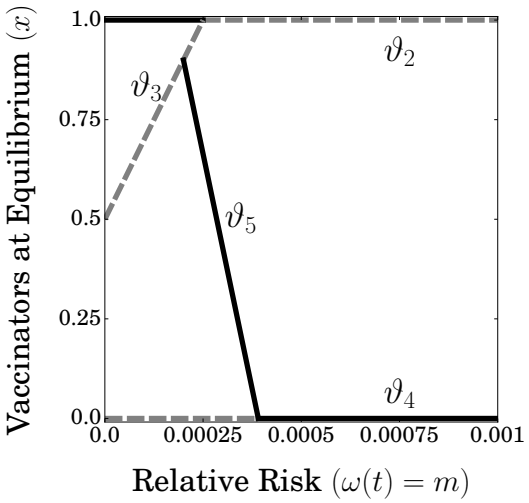
Table 1.1: The stability criteria for the Oraby *et al.* model.

Note that as usual, the disease free equilibrium ϑ_1 is unstable when $1 < \mathcal{R}_0$. The full vaccine coverage equilibrium ϑ_2 is stable when the relative risk of the vaccine, m, is strictly less than the strength of the social norms, δ. When ϑ_3 exists, it is always unstable. This is because when disease incidence is 0, the social norms will force vaccine coverage to 0 or 1. Note that when ϑ_5 is stable, there exists a region of bistability between ϑ_5 and ϑ_2. In this region, partial coverage and full coverage are stable, and the behaviour of the system as $t \to \infty$ will depend on the initial conditions. Endemic infection with no vaccine coverage stabilizes when m becomes sufficiently large, eventually becoming the only stable equilibrium. Physically, when vaccination is exceedingly risky, not even social norms can force parents to accept the risk and vaccinate.

[2]For those unfamiliar with bifurcations, a small review is presented in section 1.5.2



(a) Here $\delta = 5 \times 10^{-4}$, which violates the stability condition under this particular parametrization Hence, ϑ_5 is unstable.



(b) Here, $\delta = 2.5 \times 10^{-4}$ which satisfies the stability condition. Hence, the endemic partial vaccine coverage equilibrium is stable

Figure 1.5: Bifurcation Diagrams of the Oraby *et al* model. Diagrams are for model parameterized for measles: $\mu = 1/50$, $\gamma = 365/13$, $\beta = 281$, $k = 594.95$, and t is measured in years. Stable equilibria in black, unstable equilibria in dashed grey. Not pictured here is the the disease free equilibrium ϑ_1 as it is always unstable when $\mathcal{R}_0 > 1$, which is the case for measles.

In either case, the system demonstrates how a vaccine scare results when the relative risk of vaccination increases. Consider figure 1.5b when $m = 0$. The full coverage equilibrium is the only stable equilibrium (no risk associated with vaccinating always beats a non-zero risk associated with the chance of infection.). As m increases further, the system runs the possibility of dropping off the ϑ_2 branch when $m = \delta$ onto the ϑ_5 branch. This drop into suboptimal vaccine coverage due to an increase in the relative risk of vaccination is a vaccine scare in the context of bifurcation. It can be shown that the maximum of the ϑ_5 branch is

$$1 - \frac{(\gamma + \mu)}{\mu} \frac{1}{\mathcal{R}_0} < p_{crit} \, ,$$

thus explaining why the disease remains endemic on ϑ_5. If m continues to increase, then the vaccine is perceived to be too risky and eventually no parents will vaccinate.

Suppose that m increases to $m = \delta + \epsilon$ where $0 < \epsilon \ll 1$, forcing the system onto the ϑ_5 branch. Decreasing the relative risk to $m = \delta - \epsilon$ will not force the system back onto the full vaccine coverage branch due to bistability of ϑ_5 and ϑ_2. This demonstrates a hidden obstacle to recovering from a vaccine scare. Returning perceived risk to pre-scare levels will not guarantee that a population will achieve herd immunity. The relative risk must be decreased sufficiently so that ϑ_2 is the only stable equilibrium. Physically, this could be expensive, time consuming, or not possible given the available resources. The bistability of this system echoes the importance of maintaining vaccine coverage and preventing an increase in the relative risk of vaccination.

1.4 Machine Learning & Text Classification

With the emergence of "big data", machine learning has become an indispensable tool for deriving insights from a proverbial sea of information. Machine learning has been described as the application of induction algorithms and other algorithms that can be said to "learn" from data [30]. That is to say, the application of those algorithms which may make inferences based on data. The foundations of machine learning lay in computer science and statistics, though applications are far reaching, and can be found in fields as diverse as biology, epidemiology, marketing, and finance [31, 32, 33, 34]. This section is by no means a thorough review of the expansive and ever expanding topic of machine learning[3], but instead will serve as a primer to understand later applications of machine learning in the context of classification. In this section, I discuss: problems machine learning can be applied to, an example of a probabilistic classifier, and common metrics for algorithm performance.

1.4.1 Supervised Learning & Classification

Consider the following problems:

- Predict whether a patient, whom has suffered a heart attack, will suffer a second heart attack based on the patient's clinical measurements, demographic, and diet.
- Predict the selling price of a house based on the house's location, square footage, and number of bedrooms.
- Predict the price of a stock 6 months into the future given the company's performance and other economic indicators.

[3]For if it were, this book would be just one volume of many. Consequently, please excuse the excessive use of the phrases "...not examined here" and "...beyond the scope of this thesis."

In each of these problems, there exists a qualitative/categorical *outcome* measurement (e.g price of a stock, occurrence of a second heart attack, etc.), that is to be predicted based on a set of *features* (e.g company performance, clinical measurements, etc.). There exists a *training set* which includes previous outcome measurements and the associated features for those outcomes.

The problems listed above are all examples of *supervised learning* problems. The availability of an outcome measurement (e.g data on patients who have suffered heart attacks, prices of houses which have been sold in the recent past, etc.) to guide the learning process is what makes these problems supervised learning problems [35, 36]. Problems where no outcomes are previously observed are called *unsupervised learning* problems, and usually focus on relating, or *clustering*, the data to reveal hidden relationships.

The above problems can be categorized into *regression* and *classification* problems. Regression problems seek to predict a continuous outcome (e.g the selling price of a house, the price of a stock, etc.), while classification problems seek to predict a categorical outcome (e.g heart attack or no heart attack) from a finite set of possible classes [35, 36].

Mathematically, there exists a set of observations and outcomes (x_1, y_1), (x_2, y_2), $\ldots$, (x_n, y_n), where $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}^m$. Each outcome is related to its observation by

$$y_i = f(x_i) + \epsilon \, .$$

Here, $\epsilon \sim \mathcal{N}(\mathbf{0}, I_m)$, where I_m is the $m \times m$ identity matrix. Thus, ϵ represents white noise inherent in the observations,and is thus independent of the observations x_i. The function $f : \mathbb{R}^p \to \mathbb{R}^m$ is an unknown function. Since f is unknown, it must be estimated. This is the essence of machine learning. In many problems, a set of features, x, is readily available, but the outcome y is not easily obtained. Using the an estimate of the unknown function, $\hat{f}$, an estimated outcome can be obtained using $\hat{y} = \hat{f}(x)$.

In the following sections, $\mathbf{X}$ will be an $n \times p$ matrix, where the i^{th} row is observation x_i. Denote the outcome matrix (or outcome vector if $m = 1$) as $\mathbf{y}$, where the i^{th} row is y_i. As stated previously, $\mathbf{y}$ will be used to guide the learning process in supervised problems.

1.4.2 Supervised Classification of Natural Language

Text analysis is a major field of machine learning, but the raw text itself can not be fed into learning algorithms, as the algorithms expect numerical feature vectors with fixed size and not raw text documents with variable length [37]. The goal is to somehow map the raw text to a numerical representation so as to obtain an $\mathbf{X}$ capable of being used by learning algorithms. The process by which a document is converted to a numerical feature vector is called *feature extraction* [37].

Let C be a corpus housing documents, which can be as long or as short as needed. Each document is composed of words, separated by punctuation or white space. By counting the number of occurrences of each word in the corpus with respect to each document, a corpus can be turned into numerical matrix, where entry (i, j) is the number of occurrences of word j in document i. This particular strategy is dubbed *bag of words* [37, 38].

Consider this process of feature extraction by means of example. Let D_1 = "My dog has fleas", D_2 = "All dogs go to heaven", D_3 = "No dog in heaven has fleas", and D_4 = "It is a dog eat dog world". Thus, $C = \{D_1, D_2, D_3, D_4\}$. There are 15 unique words in C and 4 documents, so the numerical representation of the corpus, figure 1.6, will be 4×15. The result is a numerical representation of a corpus capable of being fed into various algorithms.

	all	dog	dogs	eat	fleas	go	has	heaven	in	is	it	my	no	to	world
D_1	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0
D_2	1	0	1	0	0	1	0	1	0	0	0	0	0	1	0
D_3	0	1	0	0	1	0	1	1	1	0	0	0	1	0	0
D_4	0	2	0	1	0	0	0	0	0	1	1	0	0	0	1

Figure 1.6: The feature matrix for the corpus C. The rows of the matrix represent a document from the corpus, and the entries indicate the occurrence of each indicated word. Thus, D_1 has one occurrence of the word "dog", one occurrence of the word "fleas", no occurrences of the word "eat", and so on. This approach is named "bag of words" because it is akin to breaking the document into words, and jumbling them in a bag. This particular representation ignores syntactical relationships between words, and so some information is lost.

1.4.3 Classifying with Naive Bayes Classifier

When performing classification of any type, creating a *decision boundary* (a partition of the feature space indicating the separation between elements of different classes) is often the final goal [35, 36]. Some algorithms seek to directly partition $\mathbb{R}^p$ so that elements in a common set belong to a like class, while probabilistic methods may classify by classifying according to whichever class leads to maximum posterior probabilities. In this section, a common probabilistic method is illustrated by means of example with a toy text corpus. This method is known as *The Naive Bayes Classifier*.

1.4.3.1 Naive Bayes Classifier

Given some feature vector, $\mathbf{x} \in \mathbb{R}^p$, the Naive Bayes Classifier assigns probabilities that the feature vector belongs to class C_k for K classes by computing

$$\Pr(C_k \,|\, \mathbf{x}) = \Pr(C_k \,|\, x_1, \ldots, x_p) \,.$$

Using Bayes' Rule, this conditional probability can be decomposed as

$$\Pr(C_k \,|\, \mathbf{x}) = \frac{\Pr(C_k) \Pr(\mathbf{x} \,|\, C_k)}{\Pr(\mathbf{x})} \tag{1.15}$$

The numerator of equation (1.15) is equivalent to a conditional distribution $\Pr(C_k, \mathbf{x})$, thus repeated application of the chain rule allows for the numerator to be written as

$$\Pr(C_k) \cdot \prod_{j=1}^{p} \Pr(x_j \,|\, x_j, x_{j+1}, \ldots, x_p, C_k) \ .$$

The "naive" part of the Naive Bayes Classifier is that the algorithm assumes that feature x_j is independent of x_i for $j \neq i$, and so the numerator collapses to

$$\Pr(C_k) \prod_{j=1}^{p} \Pr(x_j \,|\, C_k) \ .$$

In the context of text classification, $\Pr(x_j \,|\, C_k)$ has easy interpretation. It is the probability of finding word x_j in class C_k. Since the denominator is the same for all C_k, it can be neglected, yielding that

$$\Pr(C_k \,|\, \mathbf{x}) \propto \Pr(C_k) \prod_{j=1}^{p} \Pr(x_j \,|\, C_k) \ . \tag{1.16}$$

The classifier then will predict that a given feature vector $\mathbf{x}$ belongs to the class that maximizes $\Pr(C_k \,|\, \mathbf{x})$. Put mathematically,

$$\hat{y} = \underset{C_k \in \{C_1, \ldots, C_K\}}{\arg\max} \left\{ \Pr(C_k) \prod_{j=1}^{p} \Pr(x_j \,|\, C_k) \right\}$$

1.4.3.2 Classification in Action

Suppose that a corpus contains the following documents about either breakfast or dessert: "Eggs over easy", "Crispy bacon", "Chocolate is scrumptious", "Chocolate cake? Yum!".

$$
\begin{array}{c}
 \\
D_1 \\
D_2 \\
D_3 \\
D_4
\end{array}
\begin{array}{c}
\begin{array}{cccccccccc}
\text{bacon} & \text{cake} & \text{chocolate} & \text{crispy} & \text{easy} & \text{eggs} & \text{is} & \text{over} & \text{scrumptious} & \text{yum}
\end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}.
$$

Figure 1.7: The numerical representation of the breakfast/dessert corpus.

As in section 1.4.2, the feature extraction can be done on the corpus to obtain a numerical representation, shown in section 1.4.3.2. Let $\mathbf{y} = [B\,, B\,, D\,, D]$, where B indicates a document is about breakfast, and D indicates a document is about dessert. Suppose the new document "Eggs over bacon" is to be classified.

The predicted class will be that which maximizes the posterior probability. To determine which class does this, $\Pr(B)$, $\Pr(D)$, and the prior probabilities will need to be calculated. An easy interpretation for determining $\Pr(C_k)$ is either to: determine probability based on prevalence in the training data, or make each class equiprobable. In either case, the example has $\Pr(D) = 0.5 = \Pr(B)$.

The likelihood is calculated by examining the relative frequencies of words in a particular class. Thus,

$$\Pr(x_i \mid C_k) = \frac{\text{count}(x_j, C_k) + 1}{|V| + \sum_{x \in V} \text{count}(x, C_k)}\,.$$

Here, V is the vocabulary, or the set of all words unique words in the corpus, $\text{count}(x_j, C_k)$ is a function that returns how many times x_j occurs in class C_k. Adding 1 in the numerator and $|V|$ in the denominator prevents multiplying by 0 or 0/0 in the case that a word not found in the corpus is found in a new document. This technique is called "Laplace Smoothing".

The document "Eggs over bacon" is now ready to be classified. The calculations are as follow:

$$\Pr(D \mid \mathbf{x}) = \Pr(D) \cdot \prod_{j=1}^{3} \Pr(x_j \mid D) = \frac{1}{2}\left(\frac{1}{16}\right)^3 \approx 1.2 \times 10^{-4}$$

$$\Pr(B \mid \mathbf{x}) = \Pr(B) \cdot \prod_{j=1}^{3} \Pr(x_j \mid B) = \frac{1}{2}\left(\frac{2}{15}\right)^3 \approx 1.2 \times 10^{-3}$$

Thus, $\Pr(D \mid \mathbf{x}) < \Pr(B \mid \mathbf{x})$, and so $\hat{y} = B$.

1.4.4 Performance Metrics

Suppose that some corpus, C, has been mapped to a numerical representation $\mathbf{X}$, through feature extraction. If each document in C has been assigned to a class c_k, then supervised learning can be performed on $\mathbf{X}$ using $\mathbf{y} = [c_{k_1}, \ldots c_{k_n}]$ to guide the learning.

Before using a learning algorithm on new data, the algorithm's performance must first be examined, to ensure that the algorithm is not producing dubious classifications.

1.4.4.1 Estimating Test Error Rate Through a Test/Train Split

No classifier is a perfect classifier, and so a given classifier may produce incorrect classifications. A good classifier is one which produces correct classifications for new data (that is to say, observations not in $\mathbf{X}$). Define the *test error rate* as

$$\mathrm{E}\left[\mathrm{I}(y_0 \neq \hat{y}_0)\right] .$$

Here, I is an indicator function, y_0 is the true class for observation x_0 which is not in $\mathbf{X}$, and $\hat{y}_0$ is the predicted class obtained from the classifier. Hence, a good classifier is one with small test error rate.

Measuring the test error rate can be difficult, because true classifications for a given observation may not be readily available. The *test/train split* approach leverages the availability of elements of $\mathbf{y}$, which are true classifications for observations in $\mathbf{X}$, to estimate the test error rate.

The test/train split approach divides the training data into two unequal parts: an $r \times p$ matrix $\mathbf{X}_{train}$, and an $(n-r) \times p$ matrix $\mathbf{X}_{test}$, where r is chosen[4] so that $\lfloor n/2 \rfloor < r$ [35, 36, 37]. The outcome vector, $\mathbf{y}$, is split accordingly into $\mathbf{y}_{train}$ and $\mathbf{y}_{test}$. The classifier is trained on $\mathbf{X}_{train}$ and $\mathbf{y}_{train}$ and the test error is estimated by using the data in $\mathbf{X}_{test}$ to produce classifications $\hat{\mathbf{y}}_{test}$. Hence, an estimation of the test error rate is

$$\widehat{\text{ter}} = \frac{1}{n-r} \sum_{k=1}^{n-r} \mathrm{I}(y_{test_i} \neq \hat{y}_{test_i}) \, .$$

1.4.4.2 Confusion Matrices

Knowing how a given classier is misclassifying can give important insight into algorithm selection. For instance, the physician in section 1.4.1 may prefer a classifier which more often incorrectly predicts a patient will have a heart attack over one that more often predicts a patient will not have a heart attack. The occurrence of *true/false positives/negatives* can be visualized by a *confusion matrix*.

By means of example, suppose the physician has trained a classifier using the test/train split method, and has obtained the following predicted classifications

$$\hat{\mathbf{y}}_{test} = [c_H \, , c_H \, , c_N \, , c_H \, , c_H \, , c_N \, , c_N \, , c_H \, , c_N \, , c_N] \ ,$$
$$\mathbf{y}_{test} = [c_H \, , c_N \, , c_N \, , c_H \, , c_N \, , c_N \, , c_N \, , c_N \, , c_N \, , c_H] \ .$$

[4] A popular choice is $r = 0.6n$, often called the *sixty-forty split [36].*

		True Label		
		c_A	c_B	Total
Predicted Label	c_A	TP	FN	
	c_B	FP	TN	
	Total			

Figure 1.8: A typical confusion matrix for binary classification. Entries in the matrix indicate the classifier's performance with respect to each class. Totals are calculated by summing along rows and columns.

Here, c_H has been coded to represent the occurrence of a heart attack, and c_N has been coded to represent the absence of a heart attack. The chosen classifier has correctly classified in some instances, and has misclassified in other instances.Using the layout found in figure 1.8, a confusion matrix can be constructed for the physician's classifier.

From the the confusion matrix in figure 1.9, a number of conclusions can be made:

- The classifier correctly predicts the occurrence of a heart attack twice. This measure is called *True Positive* (TP in figure 1.8). Similarly, the classifier correctly predicts a heart attack will not occur four times. This measure is called *True Negative* (TN in figure 1.8) [35, 36].

- Misclassifications are found on the anti diagonal The classifier predicts the occurrence of a heart attack when none occur three times, and predicts no heart attack will occur, when in reality it does, only once. These measures are called *False Negative* and *False Positive* (FN and FP in figure 1.8) respectively [35, 36].

- The *accuracy* of the classifier is the number of times the classifier correctly classifies. In terms of confusion matrix entries, this is $(TP + TN)/n$, where n is the total number of classifications made. The physician's classifier has an accuracy of 6/10, or 60% [35, 36].

		True Label		
		c_H	c_N	Total
Predicted Label	c_H	2	3	5
	c_N	1	4	5
	Total	3	7	10

Figure 1.9: The physician's confusion matrix for $\mathbf{y}_{test}$ and $\hat{\mathbf{y}}_{test}$.

- The classifier's *sensitivity* is the proportion of the time the classifier predicts a heart attack will occur when when a heart attack actually does occur. In terms of the confusion matrix, the sensitivity is the ratio of TP and the sum of the first column. The physician's classifier has a sensitivity of 2/3 or 66.66%. Thus, when a heart attack will occur, the classifier will correctly classify 2/3 of the time [35, 36].

- Similar to sensitivity is *specificity*. A classifier's specificity is the proportion of time the classifier will correctly classify that a heart attack will not occur when no heart attack occurs. The specificity is the ratio of TN and the sum of the second column. The physician's classifier has a specificity of 4/7 or approximately 57%. Thus, when no heart attack occurs, the classifier will correctly classify 57% of the time [35, 36].

1.5 Critical Transitions & Early Warning Signals

Changes in external conditions usually result in a smooth response of a state variable. Such responses are gradual and reversible. For instance, stretching an elastic band and then slowly releasing the band results in a deformation and an eventual restoration of the band's original shape. In other cases, small changes in external conditions result in disproportionally strong changes in the state variable. Such responses may still be continuous and reversible.

There exist systems where minute changes in the external conditions result in catastrophic, discontinuous changes in the state variable, that while theoretically reversible, may be extremely difficult to reverse in practice Such dramatic shifts in a state variable due to changes in external conditions are called *critical transitions* [39] . Mathematically, critical transitions are bifurcations of a dynamical system to a contrasting alternate stable steady state. Such bifurcations are observed in physiological phenomena [39] ,ecological systems [40], and most importantly to this thesis, in vaccine coverage. Since critical transitions may lead to undesirable system states, predicting these transitions has become an active area of interest in the literature. This section discusses critical transitions and bifurcations in general. I review how and when bifurcations occur and discuss their physical interpretation and consequences. I also examine tools for detecting so called *early warning signals*, evidence of an impending critical transition.

1.5.1 A Brief Review of Local Asymptotic Stability

The following proof can be found in many elementary books on dynamical systems. It is a brief review of the conditions for local asymptotic stability. Let $\mathbf{x}(t) \in \mathbb{C}^n$ be a solution

to the differential equation

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) \,. \tag{1.17}$$

Here, $\mathbf{F} : \mathbb{C}^n \to \mathbb{C}^n$ is smooth in $\mathbf{x}$. Suppose that $\mathbf{F}(\mathbf{x}^*) = \mathbf{0}$. Consider now small perturbations $\boldsymbol{\epsilon}(t) \in \mathbb{C}^n \,, \|\boldsymbol{\epsilon}(t)\| \ll 1$ from $\mathbf{x} = \mathbf{x}^*$. Let $\tilde{\mathbf{x}}(t) = \mathbf{x}^* + \boldsymbol{\epsilon}(t)$, and linearize equation (1.17) about $\tilde{\mathbf{x}}(t)$.

$$\begin{aligned}
\dot{\tilde{\mathbf{x}}} &= \dot{\mathbf{x}}^* + \dot{\boldsymbol{\epsilon}}(t) \\
&= \mathbf{F}(\tilde{\mathbf{x}}) \\
&= \mathbf{F}(\mathbf{x}^*) + \mathbf{D}_{\mathbf{F}}(\mathbf{x}^*)\boldsymbol{\epsilon}(t) + \mathcal{O}\big(\|\boldsymbol{\epsilon}(t)\|^2\big) \,,
\end{aligned}$$

which can be simplified to

$$\begin{aligned}
\dot{\boldsymbol{\epsilon}}(t) &= \mathbf{D}_{\mathbf{F}}(\mathbf{x}^*)\boldsymbol{\epsilon}(t) + \mathcal{O}\big(\|\boldsymbol{\epsilon}(t)\|^2\big) \,, \\
\dot{\boldsymbol{\epsilon}}(t) &\approx \mathbf{D}_{\mathbf{F}}(\mathbf{x}^*)\boldsymbol{\epsilon}(t) \,.
\end{aligned} \tag{1.18}$$

Here, $\mathbf{D}_{\mathbf{F}}(\mathbf{x}^*)$ is the Jacobian of $\mathbf{F}$. Equation 1.18 determines how the perturbations change over time. Of particular interest are the conditions on $\mathbf{D}_{\mathbf{F}}$ under which $\mathbf{x}^*$ is a locally asymptotically stable equilibrium (i.e the conditions under which $\|\boldsymbol{\epsilon}(t)\| \to 0$ as $t \to \infty$).

Note that $\mathbf{D}_{\mathbf{F}}(\mathbf{x}^*) \in M^{n\times n}(\mathbb{C})$, and so let $\mathbf{D}_{\mathbf{F}}(\mathbf{x}^*) = A$, a constant matrix. Therefore, equation (1.18) is a linear system of differential equations with constant coefficients

$$\dot{\boldsymbol{\epsilon}}(t) = A\boldsymbol{\epsilon}(t) \,,$$

which has solution

$$\boldsymbol{\epsilon}(t) = e^{At}\boldsymbol{\epsilon}_0 \tag{1.19}$$

Here, $\boldsymbol{\epsilon}(0) = \boldsymbol{\epsilon}_0$. In the case where A is diagonalizable, equation (1.19) can be understood in a simpler way. Let $T \in M^{n\times n}(\mathbb{C})$ be the matrix consisting of column eigenvectors of

A. Then, since the eigenvectors of A are linearly independent, T has an inverse, and $\Lambda = T^{-1}AT$ is a diagonal matrix with eigenvalues of A. It follows that $A = T\Lambda T^{-1}$, and so

$$\boldsymbol{\epsilon}(t) = e^{\left(T\Lambda T^{-1}\right)t}\boldsymbol{\epsilon}_0 \,,$$

which by properties of the matrix exponential is equivalent to

$$\dot{\boldsymbol{\epsilon}}(t) = Te^{\Lambda t}T^{-1}\boldsymbol{\epsilon}_0 \,.$$

Left multiply by T^{-1} and let $\mathbf{z} = T^{-1}\boldsymbol{\epsilon}(t)$. Then, equation (1.19) is equivalent to

$$\mathbf{z}(t) = e^{\Lambda t}\mathbf{z}_0 = \left[z_0(0)e^{\lambda_1 t}, \ldots, z_n(0)e^{\lambda_n t}\right]^T \,. \tag{1.20}$$

Therefore,

$$\begin{aligned} \|\boldsymbol{\epsilon}(t)\|^2 &= \|T\mathbf{z}(t)\|^2 \,, \\ &\leq \|T\|^2 \cdot \|\mathbf{z}(t)\|^2 \,, \\ &= \|T\|^2 \cdot \left(\sum_{i=1}^{n} z_i^2(0)e^{2\lambda_i t}\right) \,. \end{aligned} \tag{1.21}$$

Here, $\|\cdot\|$ denotes the ℓ_2 norm. It is clear to see from equation (1.21) that $\|\boldsymbol{\epsilon}(t)\|$ vanishes as $t \to \infty$ if and only if the real parts of the eigenvalues of A are less than 0. Thus, $\mathbf{x}^*$ is locally asymptotically stable if and only if the real parts of all the eigenvalues of the Jacobian evaluated at $\mathbf{x}^*$ are less than 0.

1.5.2 Bifurcations of a Dynamical System

Differential equations are used to study a wide variety of physical phenomena from financial markets to theoretical physics. Consequently, understanding dynamic regimes under different parameterizations is a crucial part of understanding the phenomena as a whole

since different parameterizations may produce qualitatively different orbits in phase space. A dynamical system is said to undergo bifurcation if the system experiences a qualitative change in its behaviour when a parameter crosses some threshold value [41]. Consider the system

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}; \theta) \, . \tag{1.22}$$

Here, $\theta \in \mathbb{R}$ is a parameter of the system. It can be the case that the eigenvalues of $\mathbf{D_F}(\mathbf{x}^*; \theta)$ are functions of θ (i.e $\lambda_j = \lambda_j(\theta)$). Suppose that when $\theta < \theta_0$, then all eigenvalues have real part less than 0, and that as $\theta \to \theta_0$ the real part of at least one eigenvalue approaches zero from below. If when $\theta_0 < \theta$ the real part of at least one eigenvalue changes sign and becomes positive, then $\mathbf{x}^*$ will destabilize, and the system will undergo a qualitative change in behaviour.

1.5.3 Critical Slowing Down

As the real part of an eigenvalue of $\mathbf{D_F}$ approaches zero from below due to variation in a control parameter θ, the system experiences a loss in resilience to perturbations from equilibrium. This loss in resilience is experienced by all systems approaching bifurcation [**?**, 42] and is known as *critical slowing down* [39, 40, 43]. In the system described by equations (1.12) to (1.14), critical slowing should be observed as $w \to \delta^-$. Indeed, shown in figure 1.10 is a numerical solution to the stochastically forced version of equation (1.14). This numerical experiment confirms that the system experiences critical slowing when equation (1.14) is parameterized close to the bifurcation value. In section 1.5.1, it was shown that near equilibrium

$$\mathbf{z}_j(t) \sim e^{\lambda_j t} \, .$$

With this in mind, it is clear to see why systems experience critical slowing down. Assume that the real part all eigenvalues of $\mathbf{D_F}$ are negative. Since $\mathbf{z}_j(t)$ behaves approximately

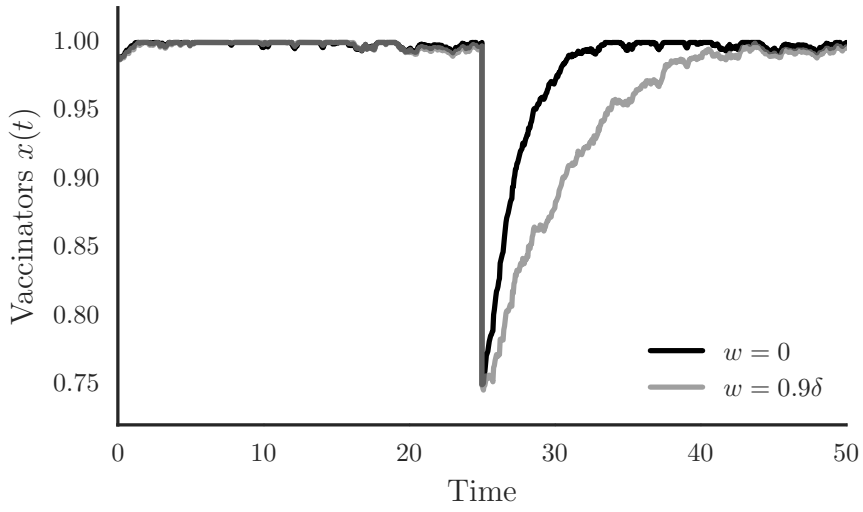


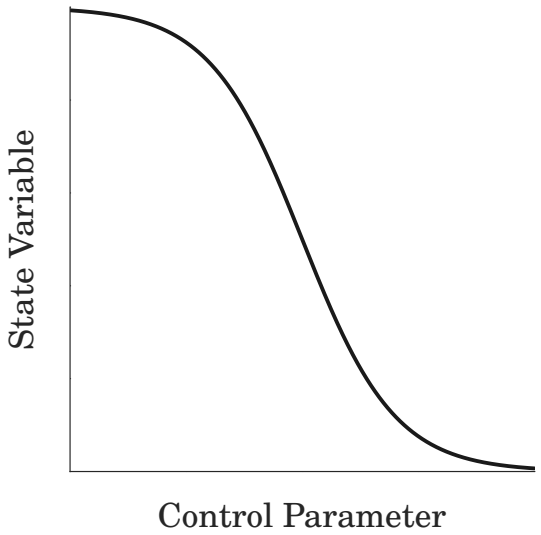
Figure 1.10: Numerical experiments of critical slowing down. Here, $\delta = 5 \times 10^{-4}$,and w is parameterized as indicated. As can be seen, the system parameterized such that w is closer to the bifurcation value returns to equilibrium slower than the system where w is parameterized far from the bifurcation.

exponentially, the orbits of $\mathbf{z}_j(t)$ will return to equilibrium faster if $1 < |\Re(\lambda_j)|$ as compared to when $0 < |\Re(\lambda_j)| \ll 1$. Since near a bifurcation $0 < |\Re(\lambda_j)| \ll 1$ for at least one $1 \leq j \leq n$, the system is expected to experience critical slowing down.

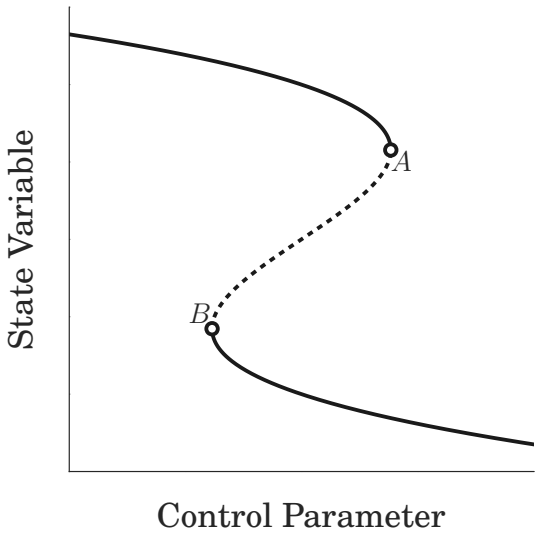
1.5.4 Critical Transitions & Early Warning Signals

In some systems, a small smooth change in a control parameter will result in a smooth change in the state variable. This will be the case if the state variable is continuous in the control parameter. The bifurcation diagram shown in figure 1.11a demonstrates such a system. A small change in the control parameter will result in a small change in the state variable.

Some systems will undergo an abrupt bifurcation to an alternative steady state under



(a) Any change in the state variable can be undone by returning the control parameter to it's original value.



(b) An example of a critical transition. Smooth variation of the control parameter may lead to a discontinuous, irreversible change in the state variable. Such a bifurcation is an example of a critical transition [39]. Stable equilibria shown in solid black, unstable shown in dashed black.

Figure 1.11: Bifurcation diagrams for two different systems. The system on the right will undergo a critical transition. Examples taken from [43]

a smooth change in the control parameter. These bifurcations can be sudden, violent, and may come without any noticeable change in the state variable. Such bifurcations are called *critical transitions* [39, 40]. See figure 1.11b for an example of a system which undergoes a critical transition. If the control parameter is close to, but less than, A, then the system remains on the upper stable branch. Smoothly increasing the control parameter for this system will force the state variable off the top branch onto the bottom branch. If the state variable is the health of an ecosystem, such a bifurcation would send the ecosystem into an unfavourable state. Furthermore, the bifurcation would not be reversible, as simply

changing the value of the control parameter back to its original value would result in the system remaining on the lower branch due to bistability. Hence, to return the system to its original state, the control parameter would have to be moved an area where bistability ceases to exist. Physically, this could difficult, expensive, time consuming, or impossible given the interpretation of the control parameter.

Though the state variable may show no signs of an impending critical transition, there do exist other generic early warning signals of critical transitions [44, 39, 40]. As a consequence of critical slowing, an increase in the lag-1 autocorrelation and variance of the stochastically forced dynamical system should be observed [44, 40, 39, 43, 45, 42]. The increases in these statistics appear in all systems approaching a critical transition, and are said to be early warning signals of an impending critical transition [46, 39].

1.6 Objectives & Direction

Thus far, this thesis has discussed: mathematical modelling of infectious diseases (section 1.1), the effects of vaccination on disease dynamics (section 1.2), salient factors for vaccinating habits (section 1.3), machine learning practices and metrics (section 1.4), and critical transitions (section 1.5).

The remainder of this thesis will examine vaccine scares through the lens of critical transitions theory. It was shown in section 1.3 that the system of differential equations described by equations (1.12) to (1.14) undergoes a discontinuous irreversible bifurcation when $w(t) = \delta$. Bifurcations of this type are called critical transitions. As the the system draws closer to the bifurcation point due to a slow variation in a control parameter, the system becomes less resilient to small perturbations. This results in an increase in the variance and lag-1 autocorrelation in the stochastically forced system. Increases in these statistics are called early warning signals, and may indicate that a critical transition is imminent.

By monitoring the early warning signals as equations (1.12) to (1.14) approach a critical transition, the occurrence of a vaccine scare may be predicted well before it happens. Equations (1.12) to (1.14) will be turned into a system of stochastic differential equations (SDEs) and the early warning signals will be examined as $w(t) \to \delta^-$.

December 2014 saw an epidemic of measles after unvaccinated individuals were exposed to an infected individual [47]. This presents an ideal instance to study the early warning signals (if they exist) leading up to, and immediately after, an epidemic. Temporally dense data is required in order to detect early warning signals, and so for this study, traditional vaccine coverage data is not appropriate as it is calculated annually. Research demonstrates that the sentiment on the social media site Twitter is a good proxy for the

sentiment in the population[48, 49], and that the sentiment on twitter regarding vaccines correlates well with population vaccine coverage [33]. Researchers have investigated how sentiment surrounding the H1N1 vaccine had evolved over time and found that the pro-vaccine sentiment on twitter correlated strongly with CDC vaccine coverage estimates [33]. Tweets are sufficiently temporally dense, and so early warning signals may manifest in tweets made about measles and vaccination. Machine learning algorithms will be used to classify each tweet based on the sentiment the tweet displays toward vaccines. An attempt to extract early warning signals from the time series of tweets will be made.

Vaccines are vital for infectious disease prevention, but vaccine scares can cause renewed disease outbreaks, or even delay global eradication by years. The ability to anticipate which populations are at highest risk of vaccine scares is therefore valuable. Through the methods presented in the following chapter, it could be possible to identify populations at risk for undergoing a vaccine scare. Public health officials may then target their trust building efforts to these populations in hopes of mitigating risk and preventing the scare from occurring

The objective of this thesis is to investigate vaccine scares through the lens of critical transitions. Since early warning signals precede a critical transition, detecting early warning signals of vaccine scares could provide a way to predict when the scare is imminent. Readily available data from the Twitter, along with Google search data, will be analysed in hopes of detecting early warning signals in the population. The chapter that follows is the draft of a manuscript being developed for submission to a peer-reviewed journal. This is followed by a discussion chapter. Because the chapter is formatted as a paper, please note that it includes supplementary figures that appear in an appendix instead of the main text of the paper.

Chapter 2

Early Warning Signals of Vaccine Scares

Manuscript, in preparation

A. Demetri Pananos[1], Thomas Bury[1], Brendan Nyhan[2], Marcel Salathé[3], Chris T. Bauch[1]

1. Department of Applied Mathematics, University of Waterloo.

2. Department of Government, Dartmouth College.

3. School of Life Sciences, École Polytechnique Fédérale de Lausanne.

2.1 Introduction

By 2003, polio had been eradicated from all parts of the world except for northern Nigeria, where unsubstantiated rumours that the oral polio vaccine contained HIV and caused infertility resulted in a vaccination ban [50]. As a result, polio resurged in Nigeria and sparked epidemics in other countries where polio had been eliminated [50, 51]. Vaccine scares have also occurred for the Measles-Mumps-Rubella (MMR) and whole cell pertussis vaccines, causing a return to endemic transmission [12, 19, 50, 51, 52].

The theory of critical transitions and their associated early warning signals may provide the ability to tell which populations are most susceptible to a widespread vaccine scare. A critical transition occurs when a system shifts abruptly from its current state to a strongly contrasting state. These shifts provide little or no obvious warning signs. However, they may exhibit characteristic early warning signals as a consequence of critical slowing down; as the system approaches the critical transition, the dominant eigenvalue becomes smaller, causing the system to lose its ability to recover from disturbances [39, 53, 40, 54]. This loss in resiliency is detectable and manifests as an increase in measures such as the lag-1 autocorrelation and variance of the time series of the state variables [40, 54, 39]. In order to detect these early warning signals, high resolution data is required.

Like the recent polio outbreaks, the 2014-15 Disneyland, California measles outbreak was driven largely by vaccine refusal, often originating in the unfounded perception that the MMR vaccine causes autism and irritable bowel syndrome [12, 19]. It attracted significant national attention in both mainstream media and online social media, causing an enormous spike in both United States geocoded tweets for "MMR" or "measles"(figure 2.1A) and Google searches for "measles"(figure 2.1B). This data provides an opportunity to look for early warning signals in high resolution time series concerning vaccine refusal.

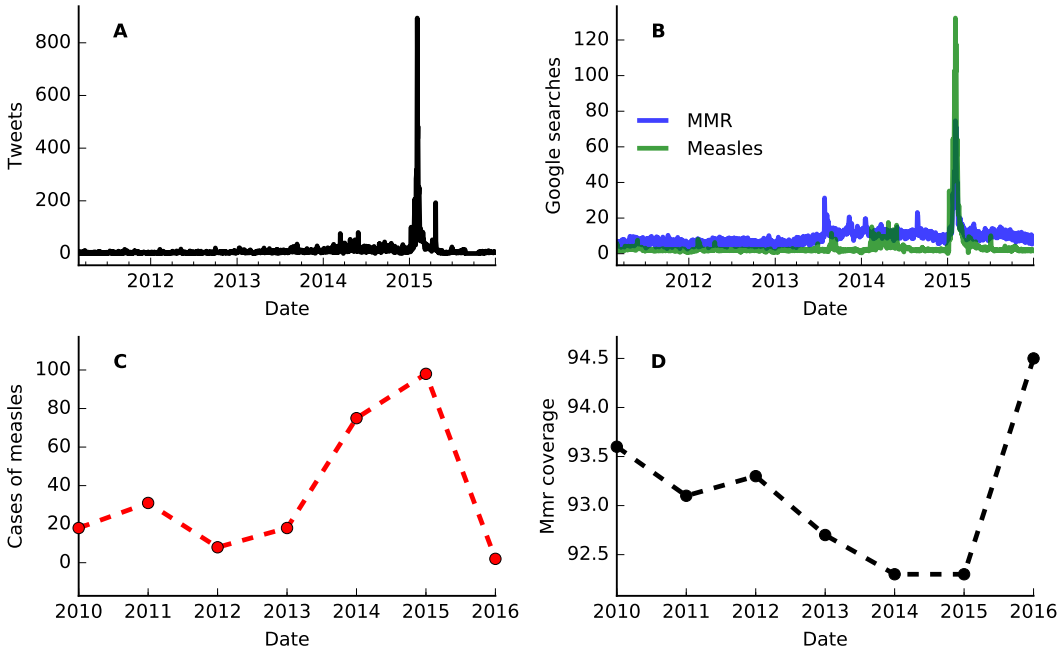


Figure 2.1: **A**: Time series of tweets mentioning "measles" or "MMR". The spike in early 2015 is in reaction to the news of the Disneyland outbreak. **B**: Time series of google searches for "measles" and "MMR". **C**: Cases of Measles in California. It is estimated that the Disneyland outbreak occurred sometime in December 2014, and so many cases are classified as having happed in 2015. **D**: Pictured is kindergarten vaccine coverage for the MMR vaccine over time. A clear decrease in vaccine coverage in the years leading up to the outbreak is observed. As the CDC releases a public health advisory notice in early February 2016, a resurgence of vaccinators can be observed.

The Disneyland outbreak was preceded by significant outbreaks in previous years (figure 2.1C). In fact, the number of US measles cases in 2013 dwarfed the number in the Disneyland outbreak, but the 2013 outbreak was localized in Amish populations and therefore went relatively unnoticed by online social or mainstream media. The MMR vaccine coverage in California kindergartens declined from 93.6% in 2010 to 92.3% statewide in 2013-14, and even lower in certain populations [55]. Following the Disneyland outbreak, the MMR vaccine coverage in California kindergartens rebounded to 94.5% by 2015 [55]

(figure 2.1D) , appearing to exemplify the coupled behaviour-disease dynamics [15].

2.2 Methods

2.2.1 Model

We build on a model of coupled behaviour-disease dynamics [15]. The model posits that members of the population play one of two strategies (vaccinator or non-vaccinator), and that members sample one another at some constant rate. If two individuals playing different strategies sample one another, the individual with lower expected payoff will switch strategies with probability proportional to the difference in payoffs. The payoff to vaccinate is given by

$$e_v = -r_v + \delta_0 x \, . \tag{2.1}$$

Here, $-r_v$ is the perceived risk associated with vaccinating, x is proportion of the population playing the vaccinate strategy, and δ_0 is the effect of injunctive social norms. The payoff not to vaccinate is

$$e_n = -cI(t) + \delta_0(1 - x) \, . \tag{2.2}$$

Here, c is the multiplicative product of the cost of infection, reporting probability, and a proportionality constant governing the perceived probability of becoming infected. The proportion of the population infected with the disease is given by $I(t)$. The differential equation governing the density of vaccinators is then

$$\frac{dx}{dt} = kx(1 - x)(-w + I(t) + \delta(2x - 1)) \tag{2.3}$$

Here, k is a rescaled social learning rate, δ is a rescaled effect of injunctive social norms, and $w = w(t)$ is a rescaled measure of how risky the vaccine is compared with infection. The following mathematical model of coupled behaviour-disease dynamics describes a population where measles transmission occurs due to contact between susceptible and infectious

individuals, and where individuals talk to other individuals and change their vaccine sentiment according to a perceived health gain that depends on perceived vaccine risk, the perceived risk of infection and infection complications, and injunctive social norms in the population:

$$\frac{dS}{dt} = \mu(1-x) - \mu S - \beta SI \tag{2.4}$$

$$\frac{dI}{dy} = \beta SI \ - (\mu + \gamma)I \tag{2.5}$$

$$\frac{dx}{dt} = kx(1-x)(-w + I + \delta(2x-1)) \tag{2.6}$$

Here, S is the proportion of susceptible individuals, I is the proportion of infectious individuals, x is the proportion of individuals who vaccinate their children, μ is the per capita birth and death rate, β is the transmission rate, γ is the rate of recovery from infection. Note that the proportion of recovered individuals, R, is determined when S and I are known since $R = 1 - S - I$, and thus need not appear in the model equations.

The system is studied as it approaches a bifurcation point from high vaccine coverage to low vaccine coverage which happens when $w(t) = \delta$. To study the dynamics leading to the bifurcation, $w(t)$ is written as a piecewise monotonic function, increasing linearly from a baseline value until the bifurcation value is reached.

To study the effects of critical slowing down, the system developed by Oraby *et al.* was converted into a stochastic differential equation with additive noise

$$d\mathbf{y} = \mathbf{F}(t, \mathbf{y}, w, \delta)dt + \boldsymbol{\sigma} d\mathbf{W} \ . \tag{2.7}$$

Here, $d\mathbf{W}$ is a Wiener process, and $\boldsymbol{\sigma} \in \mathbb{R}^3$ is a vector housing the additive noise for the S, I, and x compartments (i.e $\boldsymbol{\sigma} = [\sigma_S, \sigma_I, \sigma_x]^T$) .

2.2.2 Parameterization

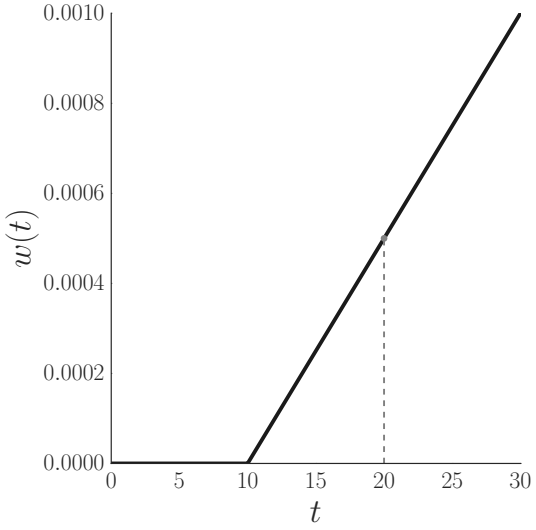
Measles parameter values were taken from a previous investigation of the coupled behaviour-disease model [15], with $\mu = 0.02$ per year, $\gamma = 365/13$ per year, $\mathcal{R}_0 = 16$, and $\beta = \mathcal{R}_0(\mu + \gamma)$ per year. The sampling rate was calibrated to $k = 500$ per year, similar to that found in the previous investigation[15], so as to obtain a rate of time evolution of vaccine uptake similar to that observed in the Disneyland, California outbreak (figure 2.1). The effect of social norms, δ, was set to 5×10^{-4}. This value of the strength of the injunctive norms was selected so as to investigate the worst case scenario; the critical transition would drop from full vaccine coverage to no vaccine coverage. Under this parameterization, the endemic branch is unstable, and so the system will bifurcate from a state of full vaccine coverage to no vaccine coverage.

2.2.3 Simulations

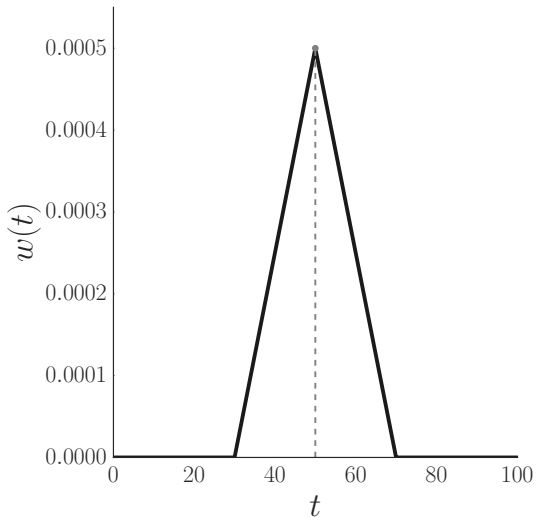
Equation (2.7) was simulated in Matlab 2015b using the SDETools Library by Andrew Horchler[56]. The SDE was solved using the `sde_euler()` function, which implements the Euler-Maruyama Method for numerically solving stochastic differential equations[57]. Stochastic noise may force the solutions out of physical phase space, and so to ensure numerical solutions remained in physical space (i.e in $[0, 1]$), an additional line of code was inserted in line 486. The inserted code read

```
% Force solution to be <=1
if NonNegative
    Yi = min(Yi,1);
end
```

Random seeds were controlled using the `RandSeed` option for the solver to ensure reproducibility. The simulation in figure 2.3B was run for 30 years with a step size of 5×10^{-4}.



(a) Visualization of the evolution of the relative risk in equation (2.8). The grey dashed line indicates where the relative risk reaches the critical transition.



(b) Visualization of the evolution of the relative risk in equation (2.9). The grey dashed line indicates where the relative risk reaches the critical transition.

Figure 2.2: Evolution of the relative risk for figures figure 2.3B and figure 2.3D

The relative risk in figure 2.3D took a functional form of

$$w(t) = \delta H(t-10)\left(\frac{t}{10} - 1\right) \tag{2.8}$$

Here, H is the Heaviside function, as implemented by Matlab 2015b. Under this functional form, the bifurcation value is reached when $t = 20$ (i.e $w(20) = \delta$). The simulation for figure 2.3D was run for 100 years. The relative risk in figure 2.3D took a functional form of

$$w(t) = \delta\Big(\,\mathrm{H}(t-30) - \mathrm{H}(t-70)\Big)\left(1 - \frac{1}{20} \mid t-50 \mid\right) \tag{2.9}$$

Under this parameterization, $w(50) = \delta$, and so $w(t) \leq \delta$ for all time. Additive noise was used in the model to account for case importations. Noise was applied directly to each compartment, with $\sigma = [10^{-2}, 10^{-6}, 10^{-2}]$.

In each simulation, the initial conditions were $S(0) = 0.016$, $I(0) = 0$, and $x(0) = 0.9903$. To obtain these initial conditions, the system was simulated for 100 years with $w(t) = 0$ from a random physically real initial condition. This was done so as to obtain a state of the system without any transient dynamics. The state of the system at $t = 100$ was used to as the initial conditions for further simulations.

Detection of Early Warning Signals

The Kendall tau rank correlation coefficient[58] is used to measure the increase in the lag-1 autocorrelation and variance as the critical transition is approached. A pair of points (x_i, y_i) and (x_j, y_j) are said to be concordant if $x_j < x_i$ and $y_j < y_i$ and are said to be discordant if $x_j < x_i$ and $y_i < y_j$. If $x_j = x_i$ or $y_j = y_i$ then the pair is neither discordant or concordant. The Kendall tau rank correlation coefficient for n points is then

$$\tau = 2\,\frac{\text{Number of Concordant Pairs} - \text{Number of Discordant Pairs}}{n(n-1)}\,.$$

The Kendall tau calculation was performed using Python's `scipy.statistics.kendalltau` function[59, 60]. The lag-k autocorrelation for a stationary signal X_t with mean μ and variance σ^2 is defined[61] as

$$R(k) = \frac{\mathbb{E}\left[(X_t - \mu)(X_{t+k} - \mu)\right]}{\sigma^2}\,.$$

Alternatively, for a discrete array, the autocorrelation at lag k can be written as

$$R(k) = \frac{\sum_{i=1}^{N-k}(X_i - \mu)(X_{i+k} - \mu)}{\sum_{i=1}^{N}(X_i - \mu)^2}$$

Note that $-1 \leq R(k) \leq 1$ for all k and that $R(k) = R(-k)$.

Time series were generated so that an epidemic occurred in the support of $w(t)$. For the detection of the early warning signals, the `earlywarnings` R library [62] was utilized. The function `qda_ews()` was used to detect increases in variance and autocorrelation. The Nadaraya Watson estimator, given by

$$\hat{y}(y) = \frac{\sum_{k=1}^{N} y_i K\left(\frac{x - x_k}{h}\right)}{\sum_{k=1}^{N} K\left(\frac{x - x_k}{h}\right)}$$

was used to smooth the stochastic time series. Here, K is a Gaussian function, and h is the bandwidth for the Gaussian function K. The bandwidth was selected using Silverman's rule of thumb [63], which asserts that the bandwidth should be chosen as

$$h = 0.9N^{-0.2}\min\left\{\hat{\sigma}, \mathrm{IQR}/1.34\right\}$$

so as to minimize the mean integrated square error. Python was used to calculate the rolling statistics. Residuals were converted to `pandas.Series` objects, which have a `rolling` routine, allowing for various functions to be computed for a specified rolling window. Lag-1 autocorrelation was calculated using

```
Res.rolling(window_length).apply( lambda x: pd.Series(x).autocorr(1) )
```

while variance was calculated in a similar fashion, using

```
Res.rolling(window_length).apply(lambda x: pd.Series(x).std() )
```

See [64] for further information on the `pandas` library. The rolling statistics for figure 2.3C were calculated using a window width of 10 years, while the rolling statistics in figure 2.3E were calculated using a window width of 20 years.

Since it is not clear when (if at all) the relative risk is equal to the strength of the injunctive norms in the population, pre and post maximum risk sections are determined by splitting the rolling statistic at its maximum value. Hence, to calculate the pre or post Kendall tau coefficient, the rolling statistic curve was split into two halves at the point where it attains its maximum value, and the Kendall tau coefficient was computed accordingly for each half.

2.2.4 Machine Learning

The Twitter API [65] was used to collect tweets which mentioned keywords related to measles and vaccination. A total of 29,727 tweets from March 2nd, 2011 to May 3rd, 2016 were collected. Tweets were collected if they contained specific key words or key phrases related to measles and if their location was within the continental United States. Keywords/phrases included:

> ((vaccine* OR immuni*) AND (ingredient* OR risk* OR lies OR disease* OR exemption* OR safe* OR unsafe OR killing* OR conspiracy OR scandal* OR whistleblower* OR pharmaceutical OR CDC OR documentary OR truth* OR theory OR health OR infant* OR baby OR babies OR newborn* OR school* OR aluminum OR death* OR dead OR children OR kid* OR child* OR poison* OR toxic OR mercury OR injur* OR harm* OR brain OR paraly* OR scare* OR fear* OR autism OR IBS OR autistic or "irritable bowel syndrome")) OR measles OR mmr OR "andrew wakefield"

Hence, if the above logical was returned as True, the tweet was collected into the database.

The machine learning was done using Python's `sci-kit learn` library [37]. Classification was performed using a supervised learning algorithm, and so a training set of 3667

(roughly split equally among the negative sentiment, positive sentiment, and no sentiment classes) was created. A tweet was classified as "anti-vaccine" if it clearly demonstrated an anti-vaccine sentiment or underestimated the risk associated with infection. Tweets like "Have measles? Eat an orange 3 times a day" and "Diarrhoea kills more people than HIV/AIDS, TB and measles put together" were classified and anti-vaccine. A tweet was classified as "pro-vaccine" if it clearly demonstrated a pro-vaccine sentiment, warned of the dangers of not vaccinating, or warned of the dangers of the disease. Tweets like "That is terrifying: Measles infects 90% of people who come into contact with it" and "oh no..not ever. BS. I bet you wouldn't willingly let your child hang out with a measles patient? " were classified as pro-vaccine. A tweet was classified "neutral" if it was factual or did not display a sentiment. Tweets like "Nationalism is an infantile disease. It is the measles of mankind" and "I feel the measles, it came on strong lmao okay calm down" were classified as neutral.

A support vector classifier with linear kernel, which aims to construct a maximum-margin hyperplane between training examples of different classes [66, 67], was trained on the tweets. A support vector classifier was chosen because it is effective in high dimensional problems, and is robust to over fitting [37]. To estimate prediction accuracy, 3-fold cross validation [36] was used. To improve prediction accuracy, a grid search cross validation [36] was performed to ensure the classifier's hyperparameters were appropriately chosen. Cross validation shows that the classifier predicts classes with 77% accuracy.

2.2.5 Sensitivity Analysis

To ensure results were robust, sensitivity analysis was performed on the coupled behaviour-disease model. For each risk evolution curve (linear and triangular), the system was simulated 500 times with a different random seed. For figures A.2, A.3, A.12 and A.13,

epidemiological parameters, as well as the sampling rate k, were drawn from a triangular distribution ranging from half the baseline value to three halves times the baseline value. The onset of the risk evolution curve was also sampled from a triangular distribution. For the linear case, the onset was drawn from a triangular distribution ranging from 8 to 12 years. Likewise, the onset of the triangular curve was drawn from a triangular distribution ranging from 20 to 40 years. The triangular curve remained symmetric about the $t = 50$ line, attaining a value of δ when $t = 50$.

Each of the 500 realizations was smoothed using R's `ksmooth`, which implements Nadaraya Watson kernel regression with a Gaussian kernel [68]. The bandwidth was chosen using R's implementation of the Silverman rule of thumb [63], `nrd0`. When estimating univariate data with a Gaussian kernel, Silverman's rule of thumb will produce a bandwidth that minimizes the mean integrated squared error. Residuals were obtained by taking the difference of the simulated data and the kernel regression. The rolling statistics were computed on the residuals. In the linear case, a rolling window size of 10 years was used, while a rolling window of length 20 years was used in the triangular case. The element-wise mean was computed for each statistic, and a 95% confidence interval was constructed.

To test the robustness of results over varying window length, 9 different window lengths were chosen with which to compute the rolling statistics. In the linear case, 9 evenly spaced window lengths were chosen ranging from 2 years to 10 years. In the triangular case, 9 evenly spaced window lengths were chosen ranging from 2 years to 20 years. The rolling statistics were computed for each of the 500 simulations, and the element-wise mean was plotted along with a 95% confidence interval.

2.3 Results

In its deterministic version with $\sigma = 0$ and fixed vaccine risk $w(t) = m$, the model described by equations (2.4) to (2.6) has multiple stable equilibria (see section 1.3.2). Of particular interest is the stable equilibrium $(S, I, x) = (0, 0, 1)$ corresponding to a disease-free state of full vaccine coverage. This is stable so long as the strength of the social norms, δ, exceeds the perceived vaccine risk $w(t)$. But, if $w(t)$ exceeds δ due to rising perceived vaccine risk, a transcritical bifurcation occurs: the state of full vaccine coverage loses stability and the system shifts to a state of endemic infection and zero vaccine coverage (figure 2.3A). At other parameter values, a drop to endemic infection and intermediate vaccine coverage is also possible (figure A.1).

Notably, once the system has passed the $w = \delta$ bifurcation, a decrease in the perceived risk w to pre-transition levels is not sufficient to move the system back to a state of full vaccine coverage; this would require that the relative risk be decreased sufficiently so that the full vaccine coverage becomes the only stable equilibrium. This may not be feasible, so the population may remain at a suboptimal vaccine coverage for years. Avoiding such a catastrophic transition is clearly desirable.

Unlike the deterministic version of the model with $\sigma = 0$, real populations are subject to random perturbations (stochasticity). When $\sigma > 0$, and the risk perception parameter $w(t)$ is gradually increased over time, the system again exceeds the transcritical bifurcation and collapses to zero vaccine coverage (figure 2.3B). But before this happens, the system exhibits critical slowing down as the transition is approached: The lag-1 autocorrelation and variance increases as $w \to \delta$ (figure 2.3C and figure A.3). The pattern occurs across a broad range of parameter values (figure A.2) and other window widths (figure A.4).

Applying descriptive statistics to predict future trends in vaccine coverage would work

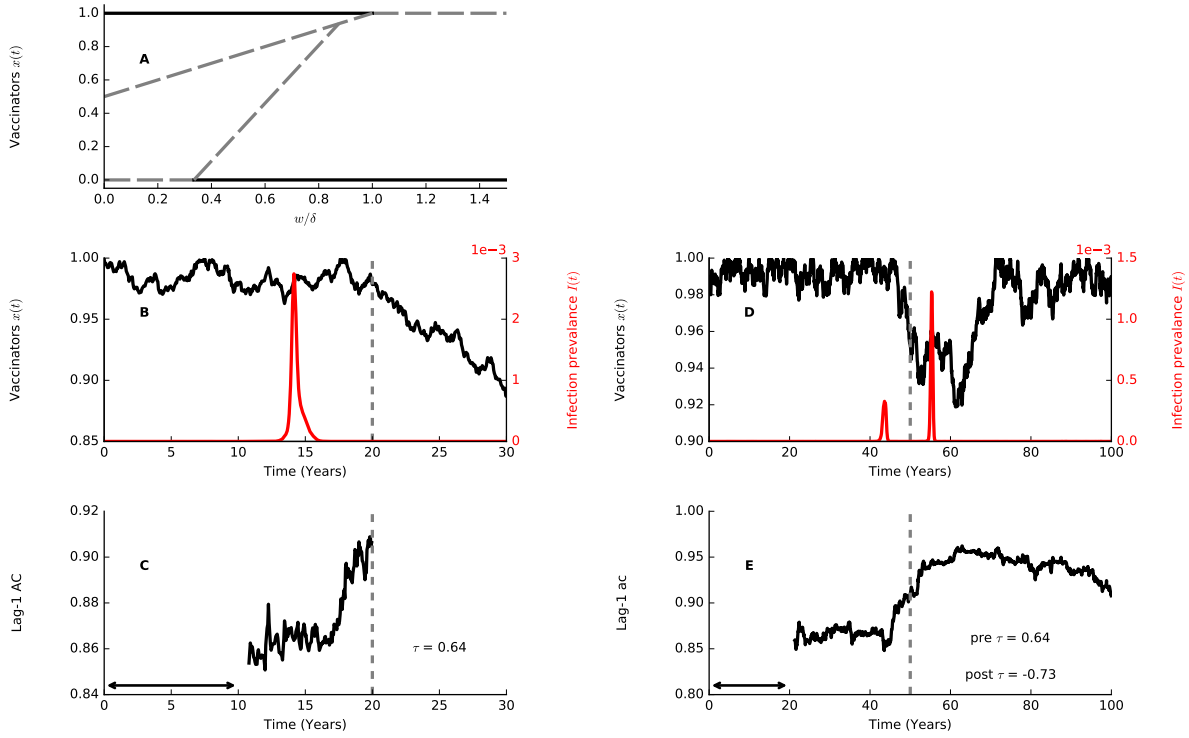


Figure 2.3: **A**: Bifurcation diagram for coupled behaviour-disease model with $\delta = 5 \times 10^{-5}$. Stable equilibira in solid black, unstable equilibria in dashed grey. Under this parameterization, no partial vaccine coverage is possible. **B**: Simulation of the coupled-behaviour disease model. Relative risk w evolves linearly from $t = 10$. Indicated in the grey dashed line is the point where the critical transition occurs. Post critical transition, vaccine coverage drops steeply. **C**: Lag-1 autocorrelation leading up to the critical transition, indicated by the grey dashed line. The length of the rolling window used to calculate the statistic is represented by the black arrow. The Kendall τ rank correlation coefficient measures the increase in the lag-1 autocorrelation. **D**: Simulation of the coupled behaviour-disease model. Risk evolves linearly from $t = 20$ to $t = 50$ at which point $w(50) = \delta$, then decreases linearly to 0. As relative risk increases, vaccine coverage drops allowing for epidemics to occur. As relative risk decreases, vaccine coverage recovers. **E**: Lag-1 autocorrelation for panel **D**. The length of the rolling window is indicated by the black arrow. Lag-1 autocorrelation increases and decreases due to change in the modulus of the dominant eigenvalue. Pre Kendall τ coefficient measures the increase in the autocorrelation before the critical transition is reached, and the post Kendall τ coefficient measures the decrease once the relative risk rebounds from the critical transition. A positive and negative pre and post Kendall τ confirms that the parameter responsible for the critical transition increases towards the threshold and then decreases away from the threshold.

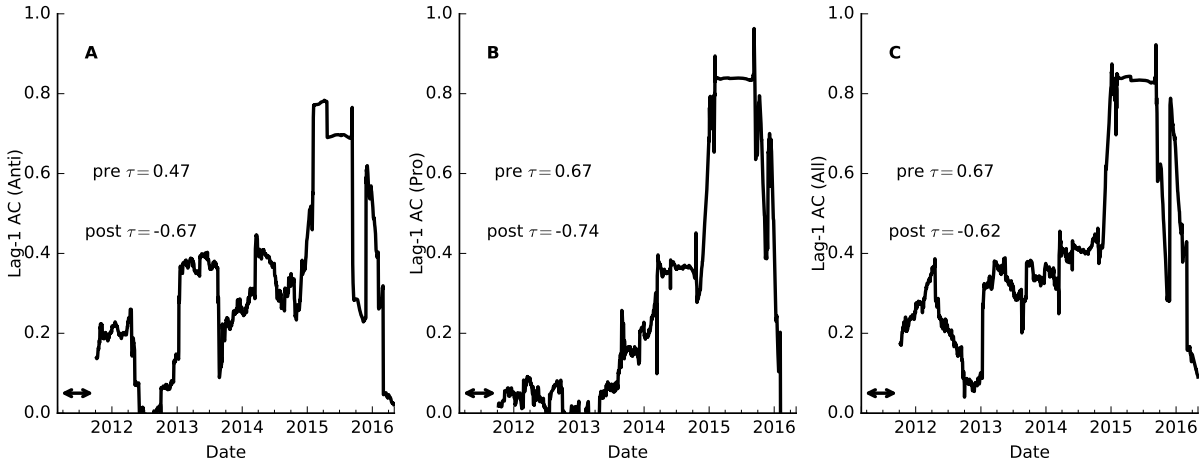


Figure 2.4: Lag-1 autocorrelation of tweets from each sentiment class (anti-vaccine sentiment in **A**, pro-vaccine sentiment in **B**, and all tweets in **C**) from mid 2011 to mid 2016. The length of the rolling window is given by the black arrow.

well away from the critical transition, but would fail close to it: Extrapolating future vaccine coverage based on linear regression to recent vaccine coverage at $w/\delta = 1$ would fail to detect the vaccine scare (figure 2.3B).

The x variable in equation equation (2.6) corresponds to the proportion of the population with pro-vaccine sentiment. Sentiment expressed in tweets is known to be indicative of the sentiment in the general population [49], and machine learning algorithms have been successfully used to evaluate sentiment content for influenza vaccines [33] . Using machine learning algorithms (see Methods), the sentiment of all US geo-coded tweets from March 2$^{\text{nd}}$, 2011 to May 3$^{\text{rd}}$, 2016 were classified as either pro-vaccine, anti-vaccine, or neutral (figure A.5). The sentiment classified tweets show a similar spike due to the Disneyland outbreak.

To help adjust for non stationarity, the time series was smoothed using Nadaraya Watson Kernel Regression with Gaussian kernel. The rolling lag-1 autocorrelation and variance

were performed on the residuals (i.e the difference between the time series and the kernel regression). The lag-1 autocorrelation of sentiment-classified tweets was computed over this time period (see Methods). The lag-1 autocorrelation of the residual increases steadily in the years preceding the Disneyland outbreak as vaccine coverage steadily decreased, for negative tweets (figure 2.4A), positive tweets (figure 2.4B), and all tweets (figure 2.4C). Notably, this increase begins even before the rolling window reaches the large spike in tweets corresponding to reaction on social media to news of the Disneyland outbreak. Hence, the analysis reveals a long-term trend toward increasingly autocorrelated sentiment throughout this time period that is not driven solely by the Disneyland spike. This is robust to rolling window width and also appears when computing variance (figures A.6 to A.11).

Uncoupled systems are simply conceived of as exceeding the critical transition when a parameter is changed sufficiently. However, the Disneyland outbreak resulted in significant media coverage and concern about infections among parents, thereby boosting the vaccine coverage to levels not seen since before 2002. Thus suggests coupled human-and-natural dynamics. As a result, the population appears to pull back from the brink of a critical transition. We observe that the lag-1 autocorrelation first plateaus immediately after the outbreak (as the rolling window moves past the peak), and then declines in the months after the outbreak. We call this effect "critical rebound", and it occurs for negative tweets (figure 2.4A), positive tweets (figure 2.4B), and all tweets (figure 2.4C). The decrease in lag-1 autocorrelation after the outbreak is statistically significant for a broad range of other rolling window widths, as well as for tweets with negative sentiment, positive sentiment, and all tweets (figures A.6 to A.8). The plateau and decline in the lag-1 autocorrelation is also observed in the mathematical model when perceived vaccine risk , w, increase to δ and then falls back down to baseline values (figure 2.3E), and this too is robust to window width (figure A.14) and other parameterizations (figure A.12). The same increase, plateau,

and decrease is observed in the variance (figure A.13). We note that such a decline and fall is not observed in regular pre and post outbreak dynamics away from the critical point, suggesting that the dynamics observed in Disneyland are not explicable in terms of ordinary transient dynamics near a fixed point away from the bifurcation (figure A.16).

The pattern is repeated for Google search data at the state and national levels over the same time period (figure 2.5), and the results are also robust to window widths (figures A.17 to A.24). However, the trend is not as strong for measles searches. There seem to be spurious modes in the lag-1 autocorrelation even at larger window lengths.

The possibility of critical rebound in coupled systems suggests that the understanding of critical transitions and their inevitability a coupled human-and-natural systems should be reframed. A full endogenization of the risk perception parameter in future work would clarify why some populations respond to an outbreak by boosting vaccine coverage to prevent endemic transmission while others, such as Japan and the United Kingdom, went past the critical transition into endemic transmission and widespread vaccine refusal before starting to recover [69].

The mathematical model in equations equations (2.4) to (2.6) is relatively simple and does not capture relevant features of vaccinating behaviour. However, detecting early warning signals only relies on the dominant eigenvalue approaching zero from below, which occurs in a broad range of dynamical systems. Hence, we expect more sophisticated models to show similar results (although exceptions do occur). Additionally, early warning signals cannot predict a particular critical transition in a specific system with complete accuracy due to stochasticity, making accurate prediction of an upcoming climate shift, for instance, difficult. However, for vaccine scares, many replicates in the form of statistics from very many countries are accessible. National and international health authorities could use these methods to rank populations based on their lag-1 autocorrelation to determine which

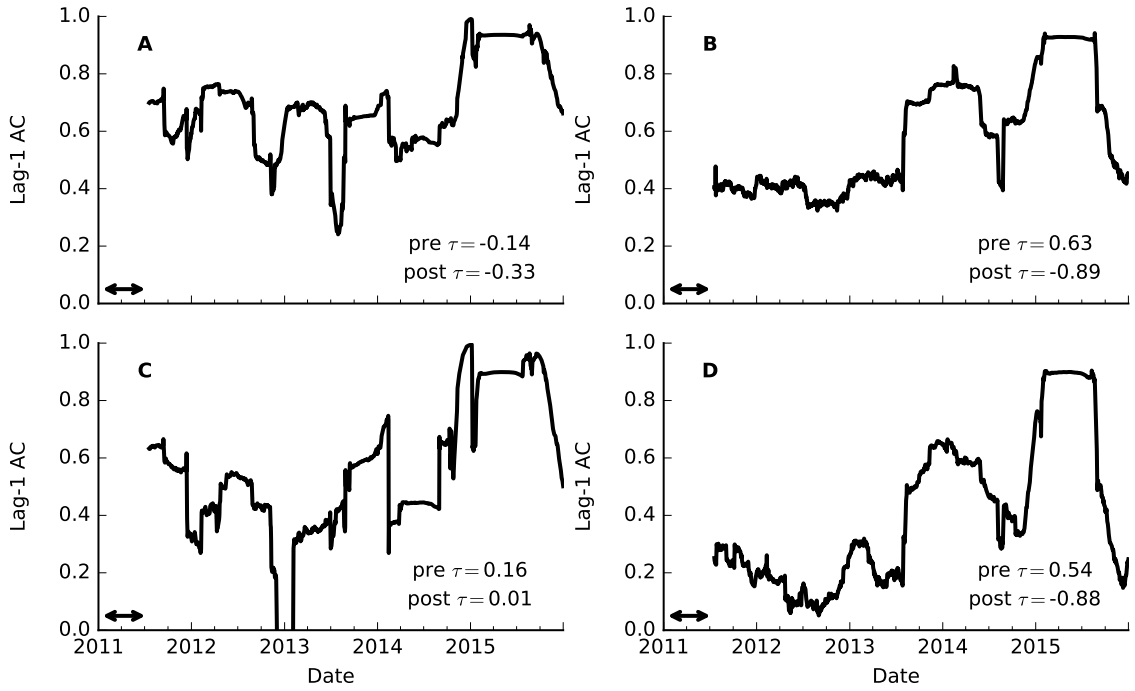


Figure 2.5: Lag-1 Autocorrelation for Google Searches of "MMR" and "measles" at the state and national level. Length of the rolling window given by the black arrow. **A**: National searches for "measles". **B**: National Searches for "MMR". **C**: State searches for "measles". **D**: State searches for "MMR".

countries are most susceptible to an episode of widespread vaccine refusal, and thus in most need of targeted efforts to build trust and disseminate accurate information about vaccines. This may help prevent some vaccine scares and accelerate the global eradication of some vaccine preventable infections.

Chapter 3

Conclusions & Future Work

3.1 Conclusions

When the perceived risk of a vaccine becomes sufficiently high, a vaccine scare may occur which may push vaccine coverage levels below those required to achieve herd immunity. Detecting when a vaccine scare is imminent is therefore valuable. Examining vaccine scares as critical transitions of a coupled behaviour-disease model allows for detection of early warning signals of an impending vaccine scare. As the critical transition is approached, the system experiences critical slowing down. This results in a detectable increase in the variance and lag-1 autocorrelation of the stochastically forced time series.

The 2014 Disneyland measles outbreak presents an opportunity to study these early warning signals through social media. In the years leading up to the Disneyland outbreak, increases in both the lag-1 autocorrelation and variance are detectable in tweets related to measles as well as Google searches for "measles" and "MMR".

These methods present public health officials with tools for monitoring, and possibly intervening in, vaccine scares.

3.2 Limitations

The coupled behaviour-disease model is a simplistic model, and does not capture all of the complex interactions that exist in the population. The model assumes homogeneous mixing between members of the population, which is not realistic for populations of this nature. Furthermore, the process by which members of the population make their vaccinating decisions is extremely simplified and does not capture all factors that one may consider. The size of the Disneyland outbreak is far smaller than what the model predicts, and so other methods (such as network models) may be more appropriate. It is important to note

that critical transitions are quite common, and thusly, early warning signals are relatively generic [70]. There are important exceptions, but these relate more to model mechanisms, and so a more complex model is expected to display early warning signals.

The largest area for improvement is in the machine learning. Since tweets are so small relative to other text corpora (140 characters vs. what could potentially be tens of thousands of characters), the syntactical relationships amongst words becomes increasingly important. Phrases like "*Are vaccines bad?*" and "*Vaccines are bad.*" clearly display different sentiments but use the same words. Since the bag of words model examines only the words used in the sentence, and ignores their order, each of these two phrases are mapped to the same feature vector. This means the classifier will assign both sentences to the same class, leading to an erroneous classification. Attempts to pre-process the data (such as removing common words like "*is*", "*or*", "*and*" etc.) may make classification even more difficult, as tweets like "*Vaccines are not bad!*" may be truncated to "*Vaccines bad*" after pre processing. The sentiment is completely lost from this tweet, making classification prone to error, especially if the remaining words are used commonly between two or more classes. More advanced techniques that incorporate syntactical relationships are needed. Since early warning signals are detectable in all tweets relating to measles and vaccines, is not expected that a change in classification accuracy will drastically change the results of the analysis, so long as the classification is not purposefully erroneous.

The time evolution of the relative risk w presented in the results is ad hoc and represents one possible scenario. In reality, the time evolution may not be symmetric, or may be discontinuous and highly non-linear.

3.3 Future Work

More advanced text analytics and machine learning algorithms will allow for the syntactical relationships of words to be learned, hence leading to better classification accuracy. More sophisticated algorithms have been examined but not yet implemented [71, 72]. One such model that shows promise is Word2Vec [71], which maps words used in the text corpus to a low dimensional vector space. Once mapped, computation is able to be done on the elements in the vector space. This allows for practitioners to ask questions like "Man is to Woman as King is to _____" to which the algorithm will correctly respond with "Queen". This demonstrates a much more in depth "learning" of the language used in the corpora, and may lead to better classification.

The endoginization of the relative risk w would present a step forward. The relative risk is itself dynamic in time and depends many complex interactions such as, but not limited to: The time since the last outbreak, the current disease prevalence, trust in authority, etc. Thus, a differential equation for w must be developed and incorporated into the coupled behaviour-model. A theory of what should be observed in social media may also be beneficial. From the analysis, it seems that activity spikes when an outbreak occurs but else is relatively dormant. If a differential equation for the activity in social media could be developed, investigators may have a better idea of what kinds of trends should be observed in times leading to a vaccine scare. Investigations into slow-fast time scale dynamics may prove fruitful in theoretical applications, as the change in the relative risk w may increase over years while the outbreak happens on a much faster timescale.

Network models would be a beneficial extension. It may be the case that the population is not homogeneously mixed (as is assumed by ODE models), and that those whom share similar views group together. This may cause an echo chamber effect, whereby anti-

vaccinator's and pro-vaccinator's beliefs are reinforced not through injunctive norms but for fear of being rejected from their existing social group. Data from social media would be extremely useful in modelling vaccinating dynamics in this fashion. Examining who follows whom is essentially an examination of a social network whereby individuals are nodes and their follows are edges. On top of the social network is a physical network, which represents which individuals come in sufficient contact with one another to transmit a disease. This suggests that a multilayer network model may be fruitful.

References

[1] William O Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, volume 115, pages 700–721. The Royal Society, 1927.

[2] Linda JS Allen, Fred Brauer, Pauline Van den Driessche, and Jianhong Wu. *Mathematical Epidemiology*. Springer, 2008.

[3] Nicholas Britton. *Essential Mathematical Biology*. Springer Science & Business Media, 2012.

[4] Robert M Corless et al. On the lambertw function. *Advances in Computational mathematics*, 5(1):329–359, 1996.

[5] Community immunity (herd immunity). `http://www.vaccines.gov/basics/protection/`. Accessed: 2016-05-17.

[6] Ray M Merrill. *Introduction to epidemiology*. Jones & Bartlett Publishers, 2015.

[7] Chris T Bauch and David JD Earn. Vaccination and the theory of games. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36):13391–13394, 2004.

[8] Simon Murch. Separating inflammation from speculation in autism. *The Lancet*, 362(9394):1498–1499, 2003.

[9] Peter McIntyre and Julie Leask. Improving uptake of mmr vaccine. *British Medical Journal*, 7647:729, 2008.

[10] Daniel A Salmon et al. Vaccine hesitancy: Causes, consequences, and a call to action. *Vaccine*, 33:D66–D71, 2015.

[11] Heidi J Larson et al. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007–2012. *Vaccine*, 32(19):2150–2159, 2014.

[12] Saad B Omer et al. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. *New England Journal of Medicine*, 360(19):1981–1988, 2009.

[13] Daniel A Salmon et al. Factors associated with refusal of childhood vaccines among parents of school-aged children: a case-control study. *Archives of pediatrics & adolescent medicine*, 159(5):470–476, 2005.

[14] Chris T Bauch. Imitation dynamics predict vaccinating behaviour. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1573):1669–1675, 2005.

[15] Tamer Oraby et al. The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1780):20133172, 2014.

[16] Robert M Wolfe and Lisa K Sharp. Anti-vaccinationists past and present. *BMJ: British Medical Journal*, 325(7361):430, 2002.

[17] Jeffrey P Baker. The pertussis vaccine controversy in Great Britain, 1974–1986. *Vaccine*, 21(25):4003–4010, 2003.

[18] Anjali Jain et al. Autism occurrence by mmr vaccine status among US children with older siblings with and without autism. *JAMA*, 313(15):1534–1540, 2015.

[19] Kreesten Meldgaard Madsen et al. A population-based study of measles, mumps, and rubella vaccination and autism. *New England Journal of Medicine*, 347(19):1477–1482, 2002.

[20] C Paddy Farrington et al. Mmr and autism: further evidence against a causal association. *Vaccine*, 19(27):3632–3635, 2001.

[21] M Haverkate et al. Mandatory and recommended vaccination in the EU, Iceland and Norway: results of the VENICE 2010 survey on the ways of implementing national vaccination programmes. 2012.

[22] World Health Organization. Vaccine coverage time series. `http://www.who.int/immunization/monitoring_surveillance/data/en/`. Accessed: 2016-05-10.

[23] Philip J Smith, Sharon G Humiston, Edgar K Marcuse, Zhen Zhao, Christina G Dorell, Cynthia Howes, and Beth Hibbs. Parental delay or refusal of vaccine doses, childhood vaccination coverage at 24 months of age, and the health belief model. *Public health reports*, pages 135–146, 2011.

[24] Lyndal Bond and Terry Nolan. Making sense of perceptions of risk of diseases and vaccinations: a qualitative study combining models of health beliefs, decision-making and risk perception. *BMC Public Health*, 11(1):1, 2011.

[25] Emily K Brunson. How parents make decisions about their children's vaccinations. *Vaccine*, 31(46):5466–5470, 2013.

[26] Julie Leask et al. What maintains parental support for vaccination when challenged by anti-vaccination messages? a qualitative study. *Vaccine*, 24(49):7238–7245, 2006.

[27] Dirk Helbing. A mathematical model for behavioral changes by pair interactions and its relation to game theory. *Angewandte Sozialforschung*, 18(3):117–132, 1994.

[28] Jonas Björnerstedt, Jörgen W Weibull, et al. Nash equilibrium and evolution by imitation. Technical report, 1994.

[29] Josef Hofbauer and Karl Sigmund. *Evolutionary games and population dynamics*. Cambridge University Press, 1998.

[30] Ron Kohavi and Foster Provost. Glossary of terms. *Machine Learning*, 30(2-3):271–274, 1998.

[31] Adi L Tarca, Vincent J Carey, Xue-wen Chen, Roberto Romero, and Sorin Drăghici. Machine learning and its applications to biology. *PLoS Comput Biol*, 3(6):e116, 2007.

[32] Li-Juan Cao and Francis EH Tay. Support vector machine with adaptive parameters in financial time series forecasting. *Neural Networks, IEEE Transactions on*, 14(6):1506–1518, 2003.

[33] Marcel Salathé and Shashank Khandelwal. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol*, 7(10):e1002199, 2011.

[34] Geng Cui et al. Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science*, 52(4):597–612, 2006.

[35] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The Elements of Statistical Learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.

[36] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*, volume 112. Springer, 2013.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[38] Zellig S Harris. Distributional structure. *Word*, 10(2-3):146–162, 1954.

[39] Marten Scheffer et al. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59, 2009.

[40] V. Dakos et al. Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, 105(38):14308–14312, 2008.

[41] Yuri A Kuznetsov. *Elements of Applied Bifurcation Theory*, volume 112. Springer Science & Business Media, 2013.

[42] C Wissel. A universal law of the characteristic return time near thresholds. *Oecologia*, 65(1):101–107, 1984.

[43] Early Warning Signlas Toolbox, `http://www.early-warning-signals.org/theory/why-should-we-expect-early-warning/`, Accessed: 2016-06-14.

[44] Vasilis Dakos et al. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PloS one*, 7(7):e41010, 2012.

[45] Anthony R Ives. Measuring resilience in stochastic systems. *Ecological Monographs*, 65(2):217–233, 1995.

[46] Marten Scheffer. *Critical transitions in nature and society*. Princeton University Press, 2009.

[47] Measles cases and outbreaks. `http://www.cdc.gov/measles/cases-outbreaks.html`. Accessed: 2016-08-1.

[48] Brendan O'Connor et al. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11(122-129):1–2, 2010.

[49] Saad M Darwish et al. From public polls to tweets: Developing an algorithm for classifying sentiment from twitter based on computing with words.

[50] Ayodele Samuel Jegede. What led to the Nigerian boycott of the polio vaccination campaign? *PLoS Med*, 4(3):e73, 2007.

[51] Mary Jane Friedrich. WHO declares polio an international health emergency. *JAMA*, 311(23):2372–2372, 2014.

[52] Felix Greaves and Liam Donaldson. Measles in the uk: a test of public health competency in a crisis. *BMJ*, 346:f2793, 2013.

[53] Marten Scheffer et al. Catastrophic shifts in ecosystems. *Nature*, 413(6856):591–596, 2001.

[54] Marten Scheffer et al. Anticipating critical transitions. *science*, 338(6105):344–348, 2012.

[55] California Department of Public Health. http://www.cdph.ca.gov/programs/immunize/pages/immunizationlevels.aspx. Accessed: 2016-08-18.

[56] Sdetools: Matlab toolbox for the numerical solution of stochastic differential equations. http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm. Accessed: 2016-08-21.

[57] Peter E Kloeden and Eckhard Platen. Higher-order implicit strong numerical schemes for stochastic differential equations. *Journal of statistical physics*, 66(1-2):283–314, 1992.

[58] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[59] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed 2016-08-21].

[60] William R Knight. A computer method for calculating kendall's tau with ungrouped data. *Journal of the American Statistical Association*, 61(314):436–439, 1966.

[61] Patrick F Dunn. *Measurement and Data Analysis for Engineering and Science*. CRC press, 2014.

[62] Early warning signals toolbox. http://www.early-warning-signals.org/resources/code/. Accessed: 2016-08-12.

[63] Bernard W Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26. CRC press, 1986.

[64] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.

[65] Twitter API. `https://dev.twitter.com/rest/public`. Accessed: 2016-09-19.

[66] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification. 2003.

[67] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152. ACM, 1992.

[68] Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

[69] Harumi Gomi and Hiroshi Takahashi. Why is measles still endemic in Japan? *The Lancet*, 364(9431):328–329, 2004.

[70] et al. Boettiger. Early warning signals: the charted and uncharted territories. *Theoretical Ecology*, 6(3):255–264, 2013.

[71] Tomas Mikolov et al. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[72] Jeffrey Pennington et al. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.

Appendices

Appendix A

Supplementary Figures

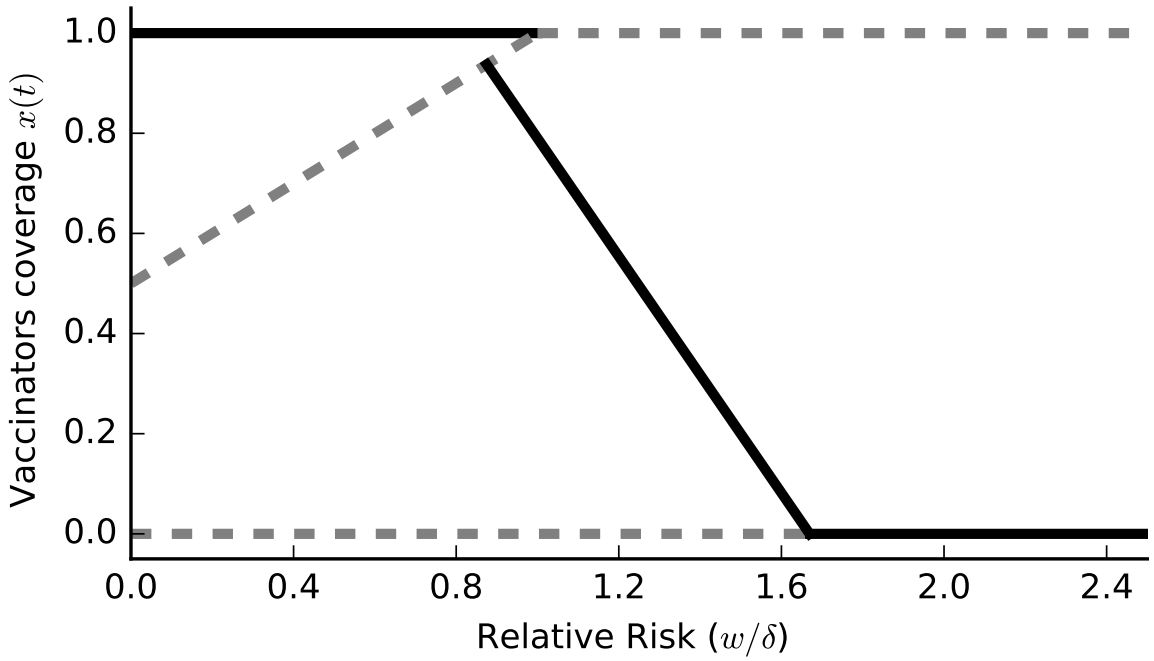


Figure A.1: Alternative bifurcation diagram for the coupled behaviour-disease incidence model represented by equations (2.4) to (2.6). Here, the strength of the injunctive norms δ is 2.5×10^{-4}, allowing for partial vaccine coverage. Solid black lines are stable equilobria. Dashed grey lines are unstable equilibira . Not shown here is the disease free equilibria $(S, I, x) = (1, 0, 0)$, which is always unstable when $\beta/(\mu + \gamma) > 1$, which is the case for measles.

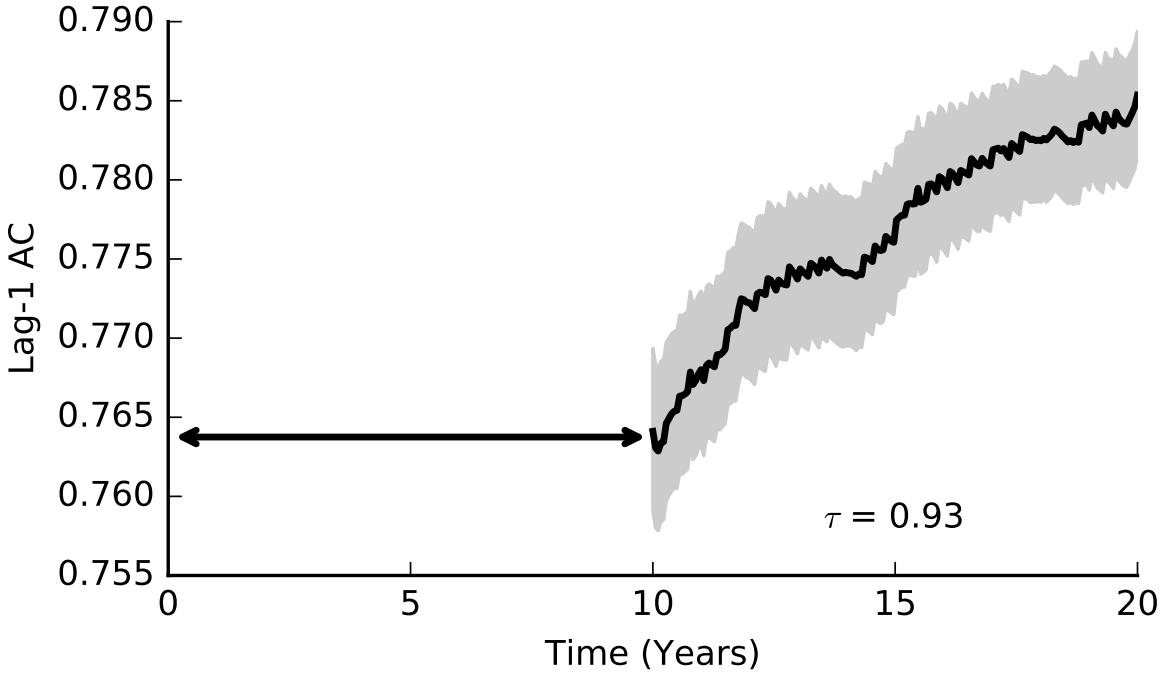


Figure A.2: Sensitivity analysis of lag-1 autocorrelation as critical transition is approached. Here, epidemiological parameters are sampled from a triangular distribution ranging from 50% to 150% of their baseline values. The strength of the injunctive social norms is $\delta = 5 \times 10^{-4}$. The relative risk evolves linearly from $t = 10$. The critical transition is achieved at $t = 20$. Length of the rolling window is indicated by the black arrow.

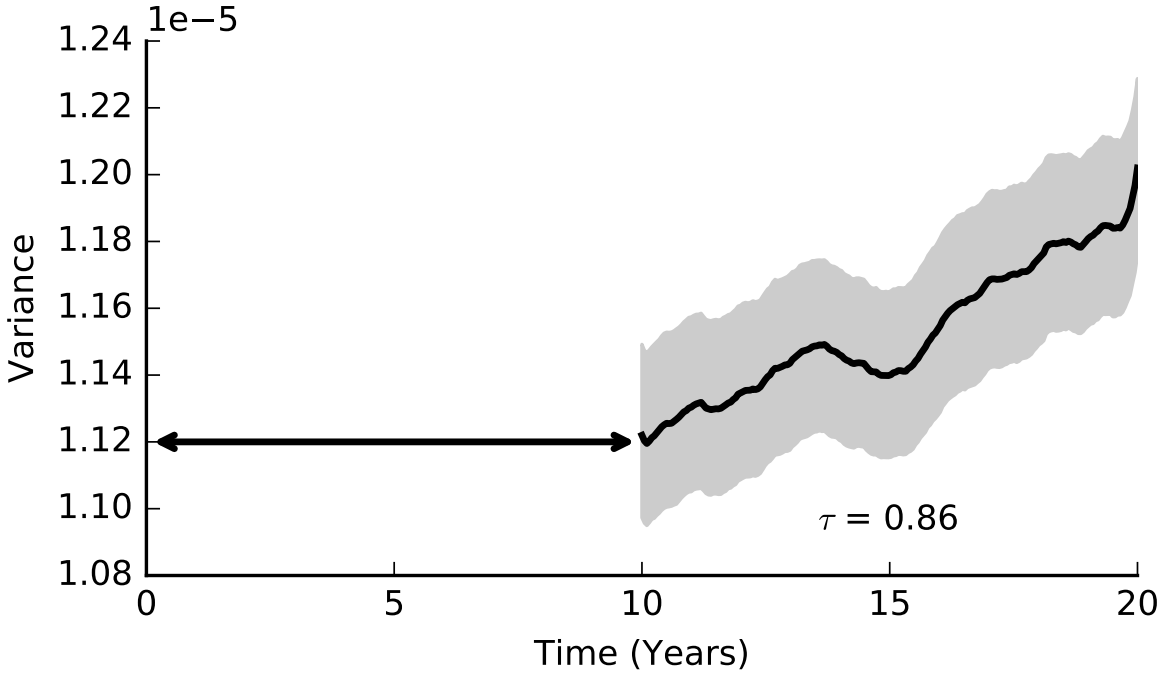


Figure A.3: Sensitivity analysis of variance as critical transition is approached. Here, epidemiological parameters are sampled from a triangular distribution ranging from 50% to 150% of their baseline values. The strength of the injunctive social norms is $\delta = 5 \times 10^{-4}$. The relative risk evolves linearly from $t = 10$. The critical transition is achieved at $t = 20$. Length of the rolling window is indicated by the black arrow.

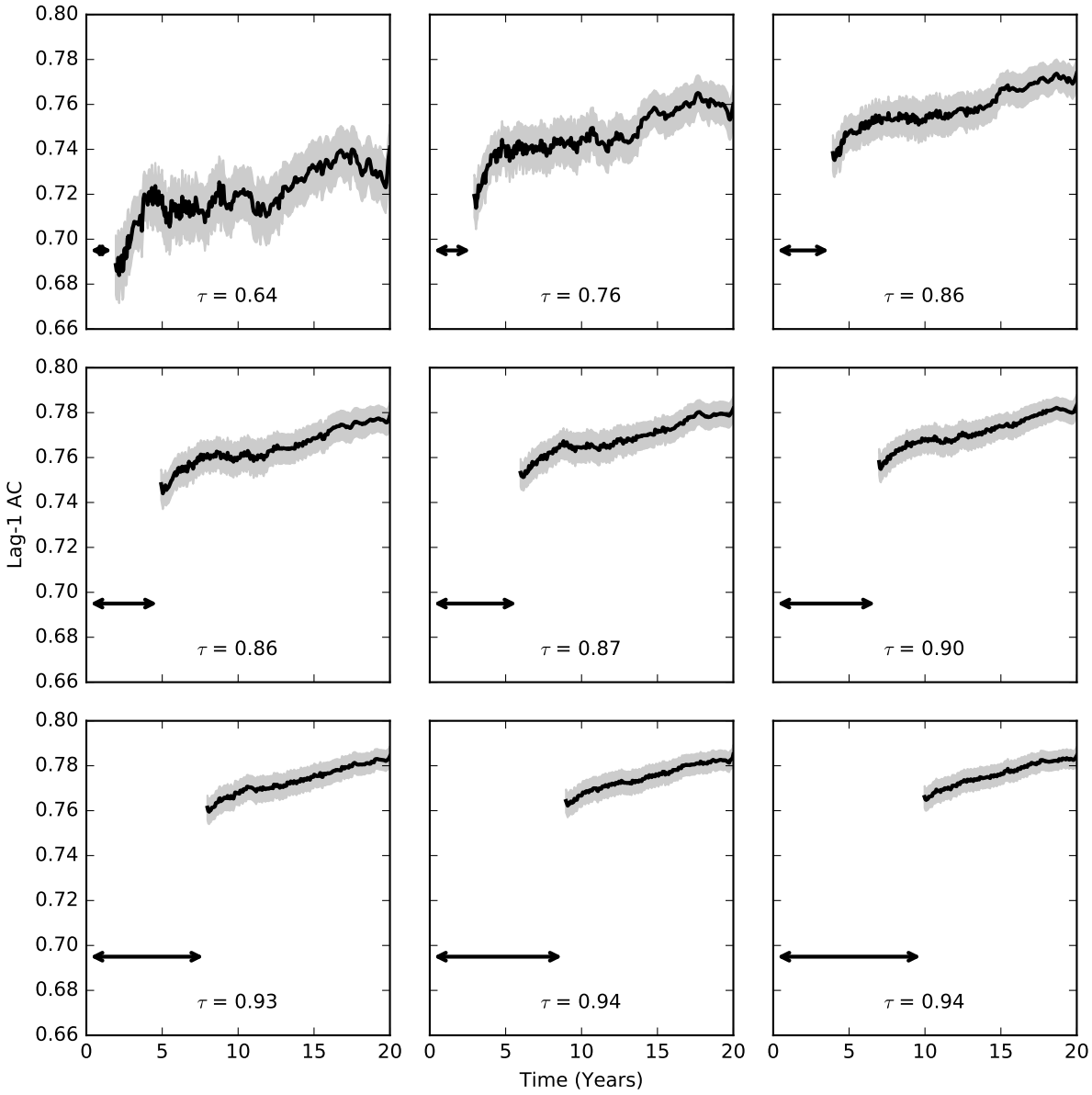


Figure A.4: Sensitivity analysis of the rolling window width for coupled behaviour-disease model. Relative risk evolves linearly from $t = 10$ and undergoes a critical transition at $t = 20$. Length of the rolling window is indicated by the black arrow

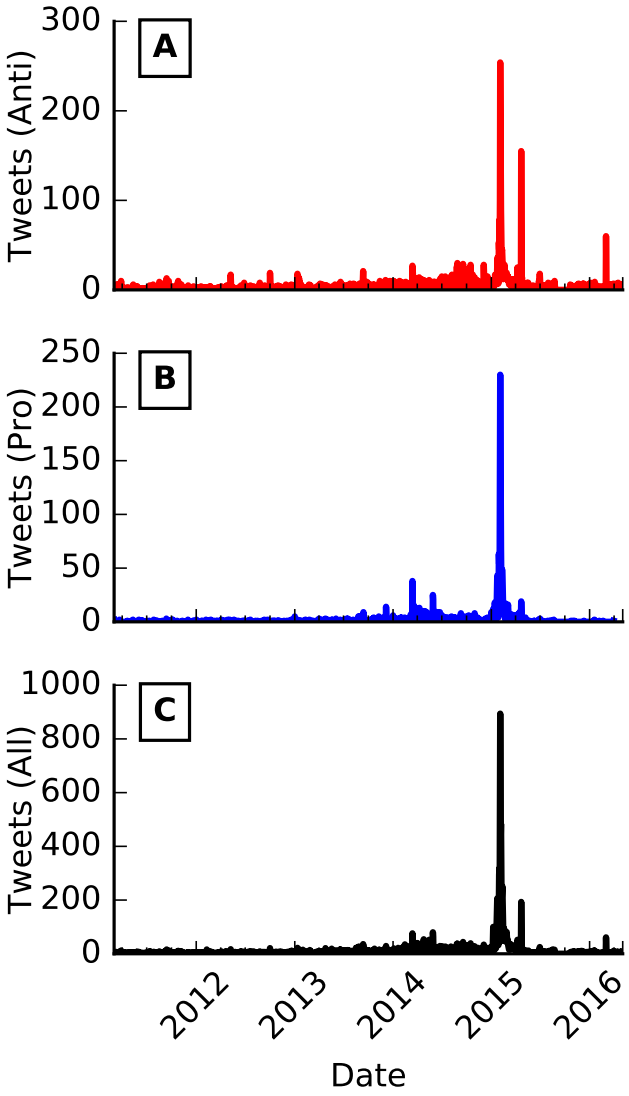


Figure A.5: Time series of sentimented tweets (**A**: Anti-vaccine tweets. **B**: Pro-vaccine tweets. **C**: All tweets). The spike in activity near the beginning of 2015 is in reaction to a CDC public health advisory notice about the Disneyland outbreak.

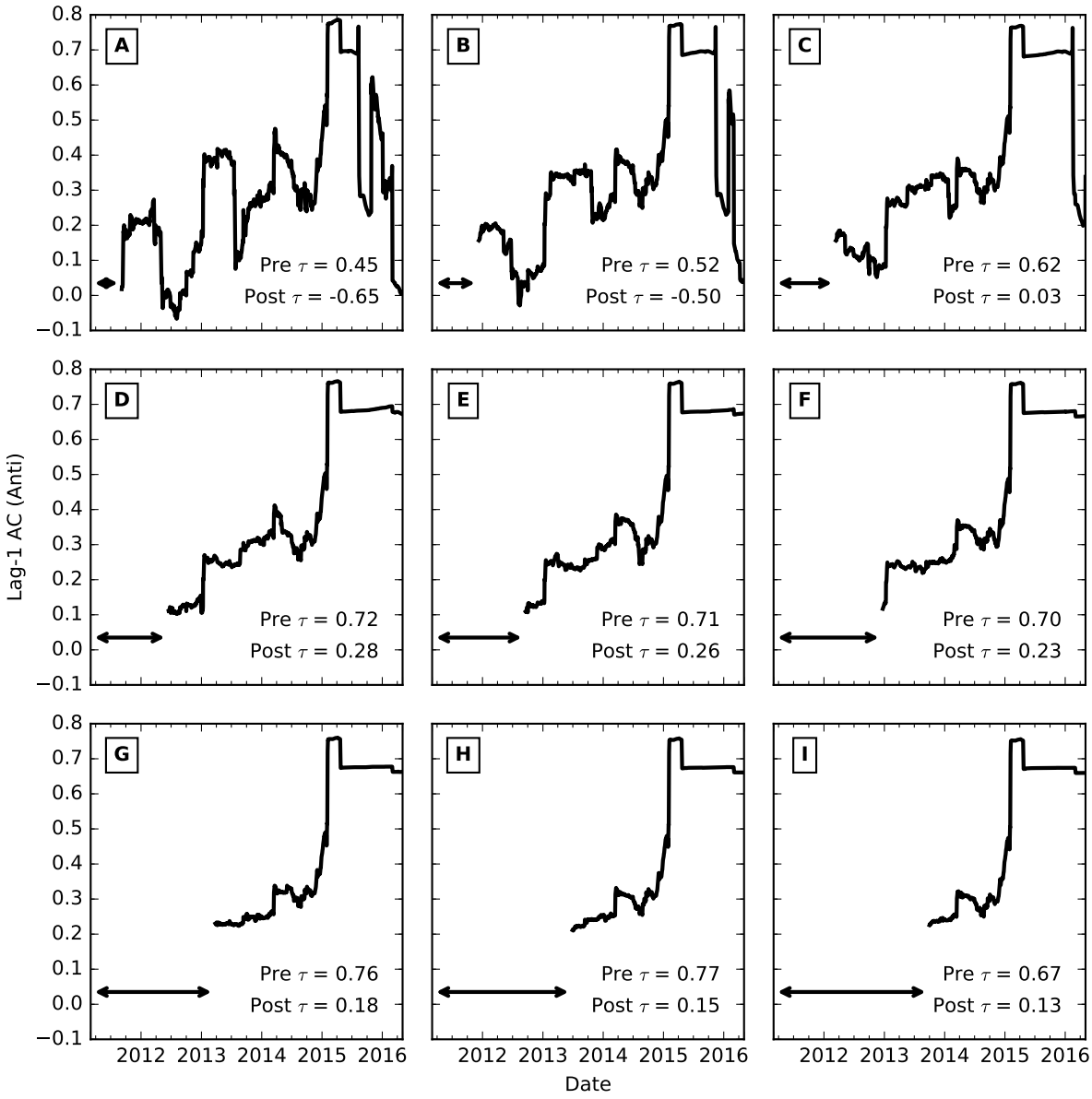


Figure A.6: Sensitivity Analysis of window width and lag-1 autocorrelation for anti sentimented tweets. Length of the moving window is indicated by the black arrow.

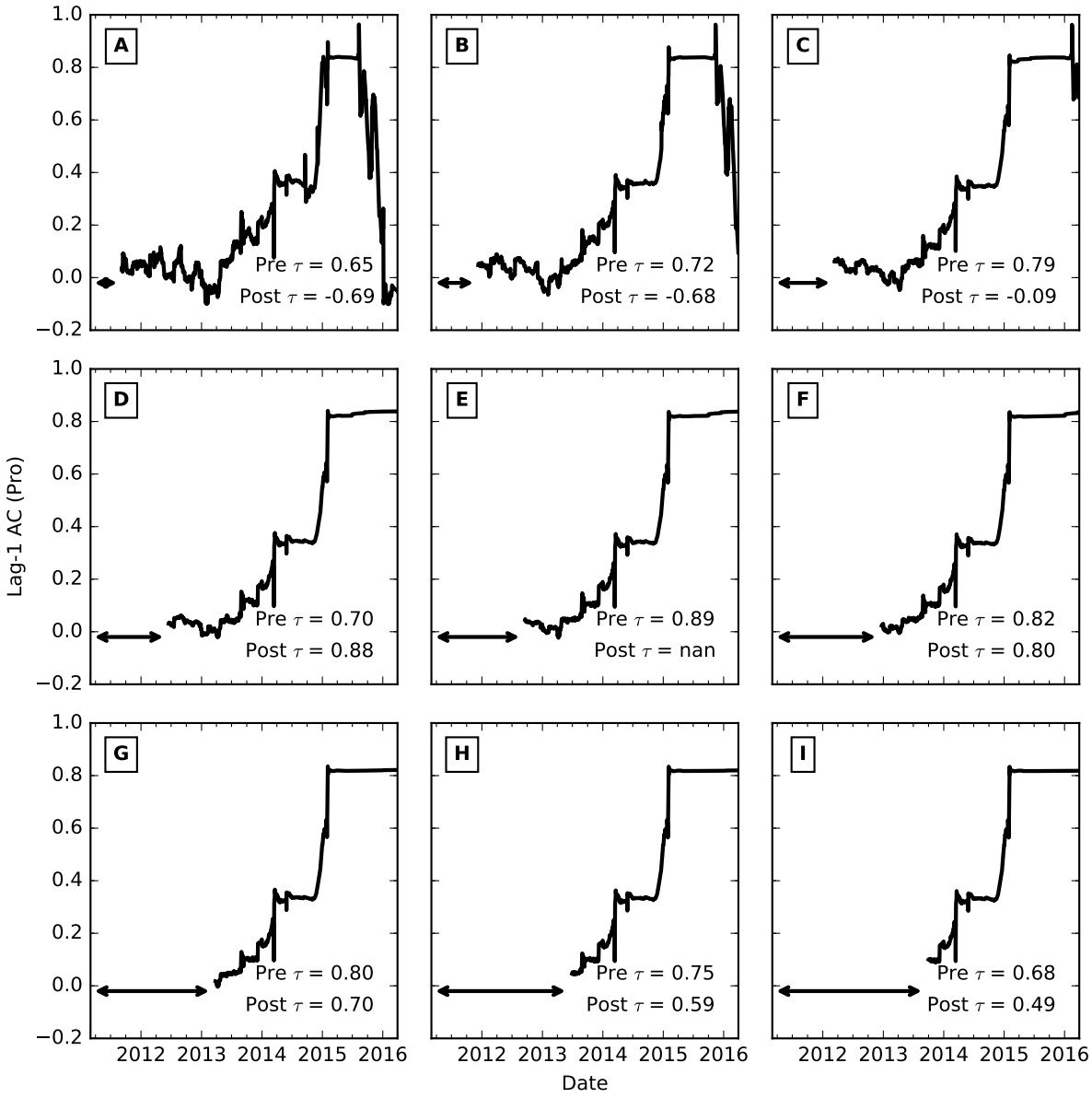


Figure A.7: Sensitivity Analysis of window width and lag-1 autocorrelation for pro sentimented tweets. Length of the moving window is indicated by the black arrow.

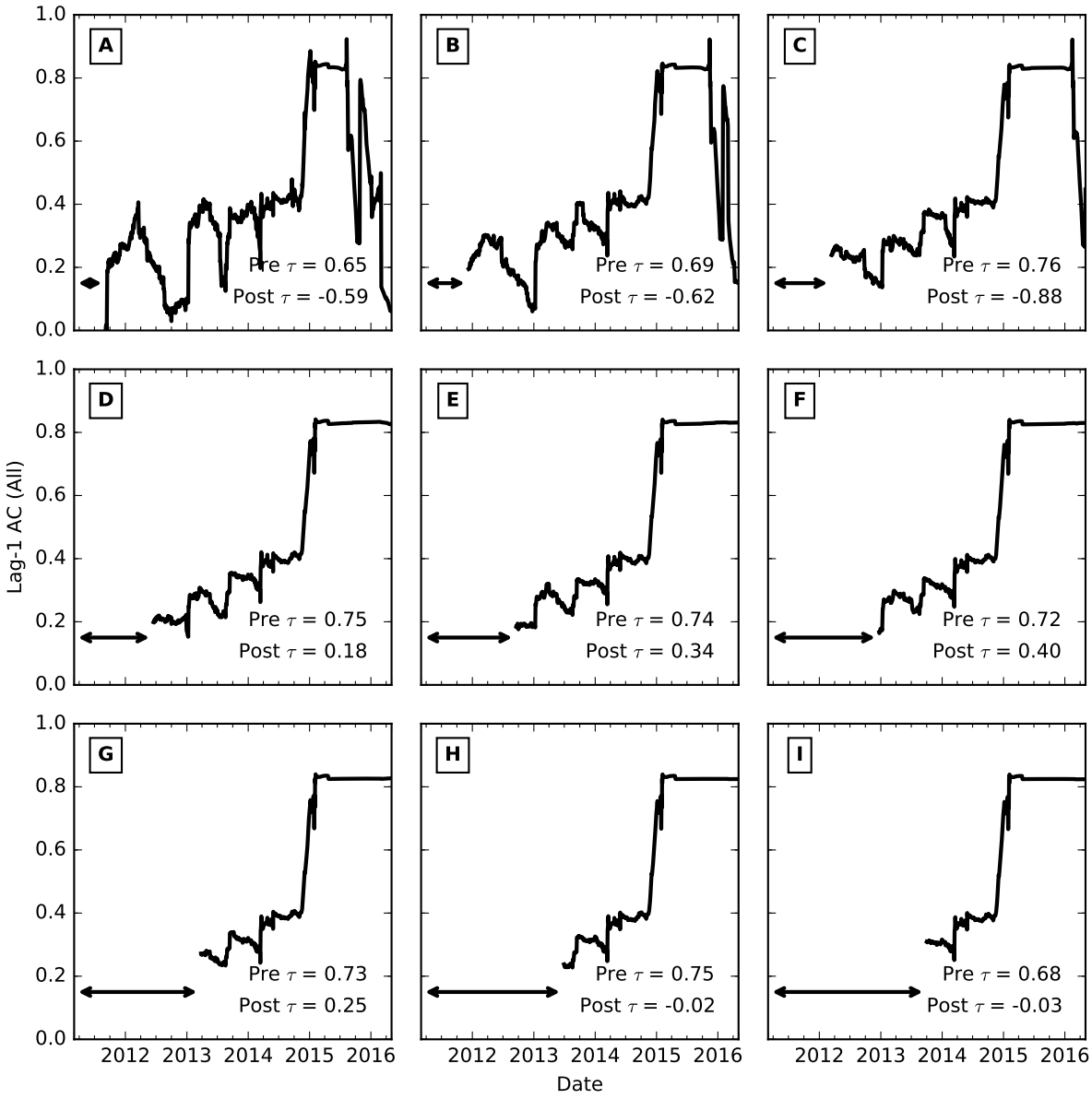


Figure A.8: Sensitivity Analysis of window width and lag-1 autocorrelation for all tweets. Length of the moving window is indicated by the black arrow.

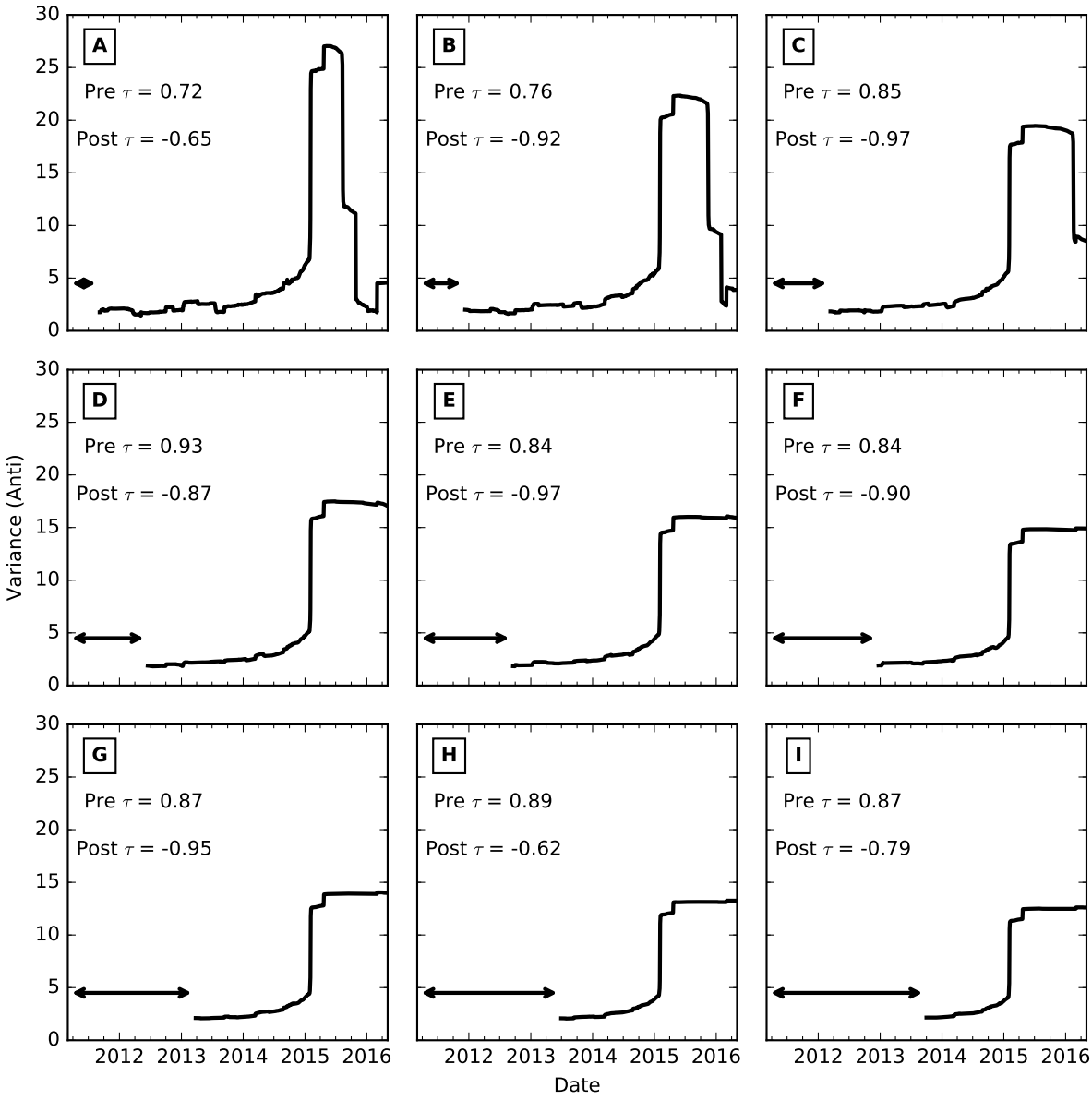


Figure A.9: Sensitivity Analysis of window width and variance for anti sentimented tweets. Length of the moving window is indicated by the black arrow.

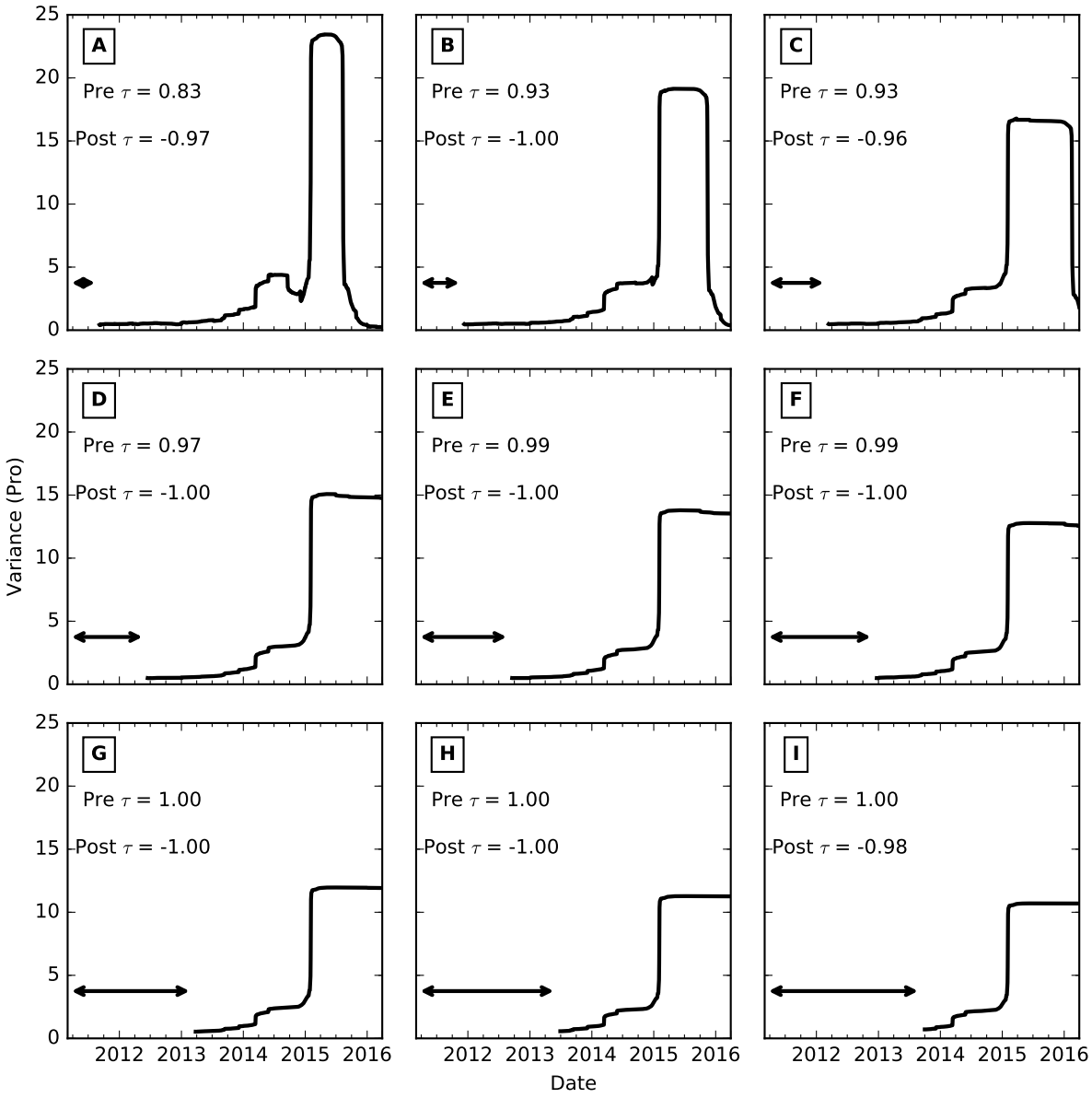


Figure A.10: Sensitivity Analysis of window width and variance for pro sentimented tweets. Length of the moving window is indicated by the black arrow.

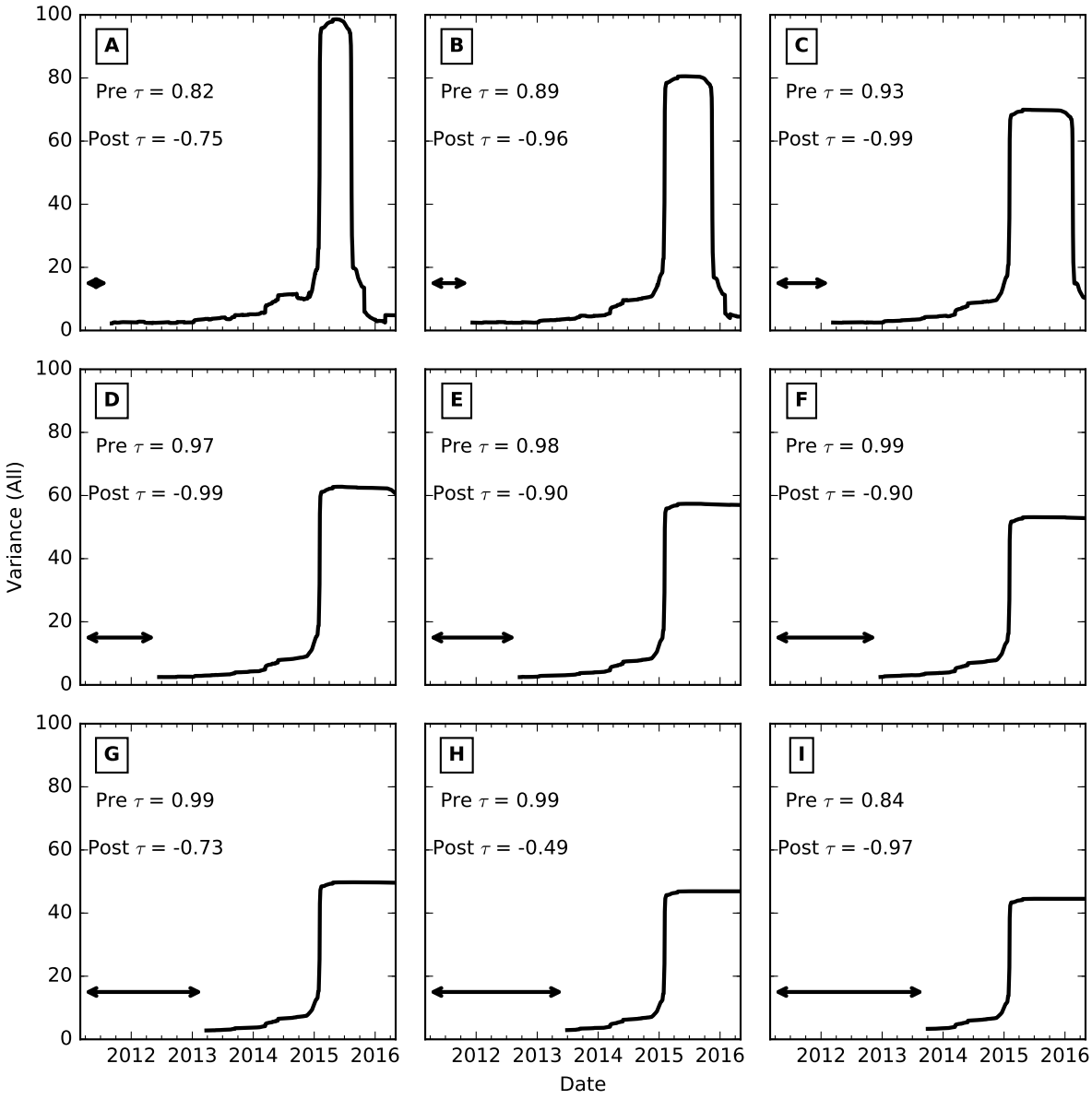


Figure A.11: Sensitivity Analysis of window width and variance for all tweets. Length of the moving window is indicated by the black arrow.

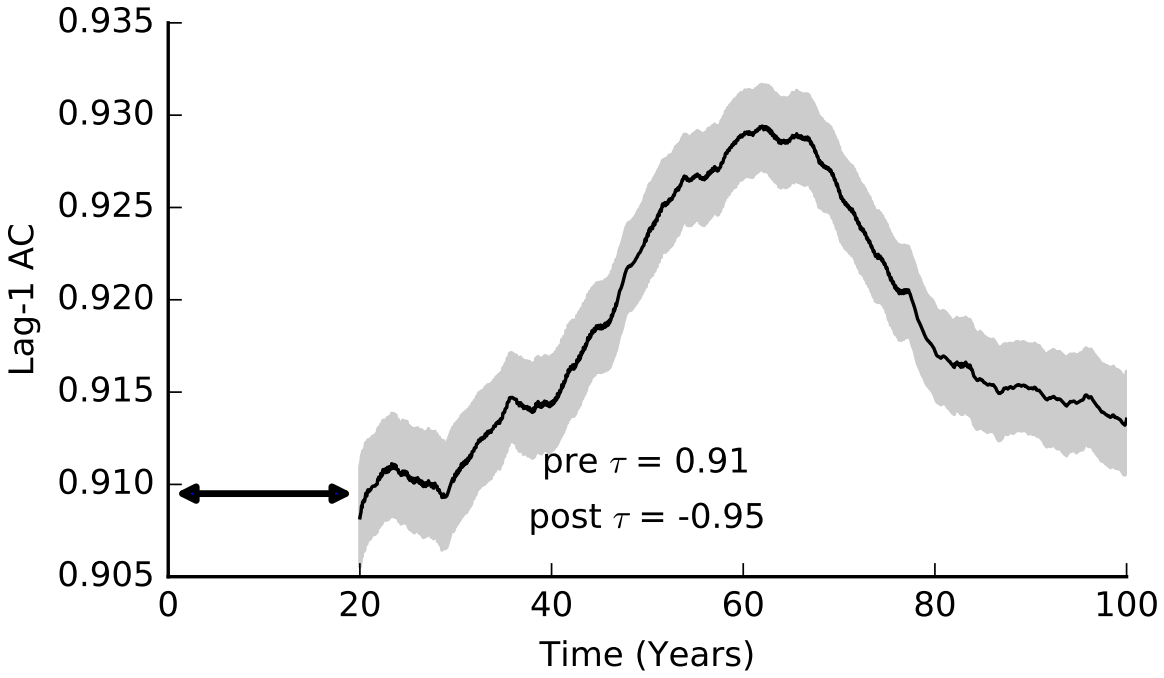


Figure A.12: Sensitivity analysis of lag-1 autocorrelation approaching and rebounding from the critical transition. Epidemiological parameters are sampled from a triangular distribution ranging from 50% to 150% of their baseline values. The relative risk evolves triangularly, reaching a maximum value of δ at $t = 50$. The risk begins to increase at $t = 30$ and returns to the baseline value at $t = 70$. The lag-1 autocorrelation can be seen to increase and decrease as the relative risk increases and decreases.Grey shaded region represents a 95% confidence interval constructed from 500 independent simulations. Length of the rolling window is indicated by the black arrow.

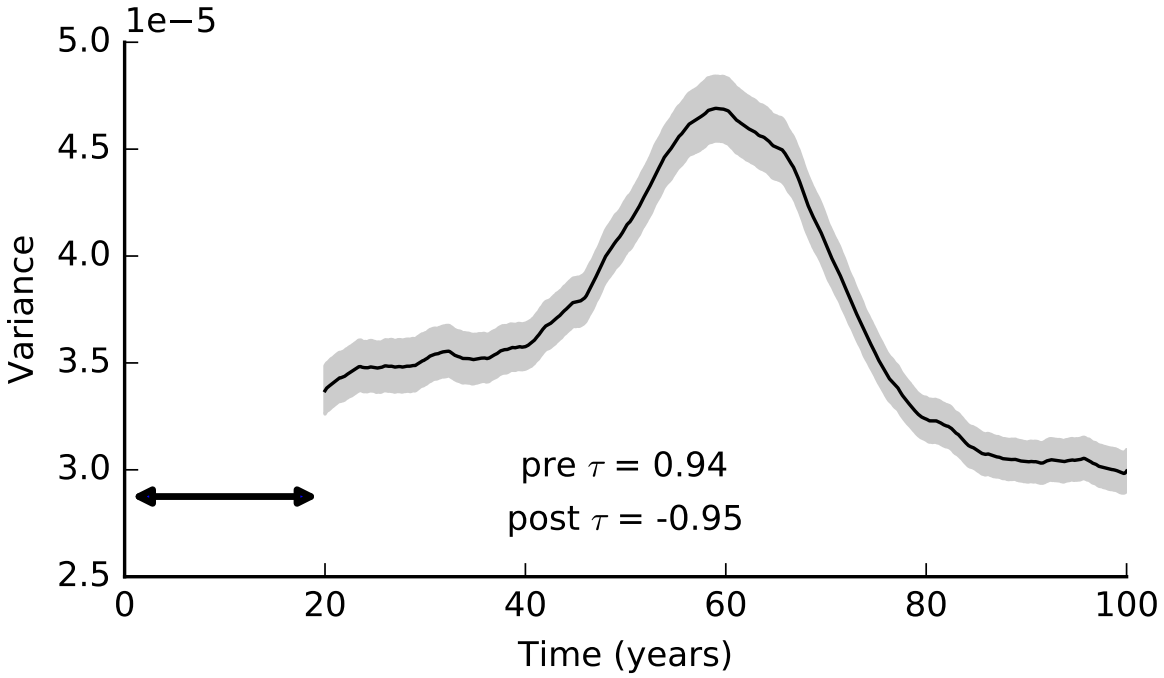


Figure A.13: Sensitivity analysis of variance approaching and rebounding from the critical transition. Epidemiological parameters are sampled from a triangular distribution ranging from 50% to 150% of their baseline values. The relative risk evolves triangularly, reaching a maximum value of δ at $t = 50$. The risk begins to increase at $t = 30$ and returns to the baseline value at $t = 70$. The variance can be seen to increase and decrease as the relative risk increases and decreases. Grey shaded region represents a 95% confidence interval constructed from 500 independent simulations. Length of the rolling window is indicated by the black arrow.

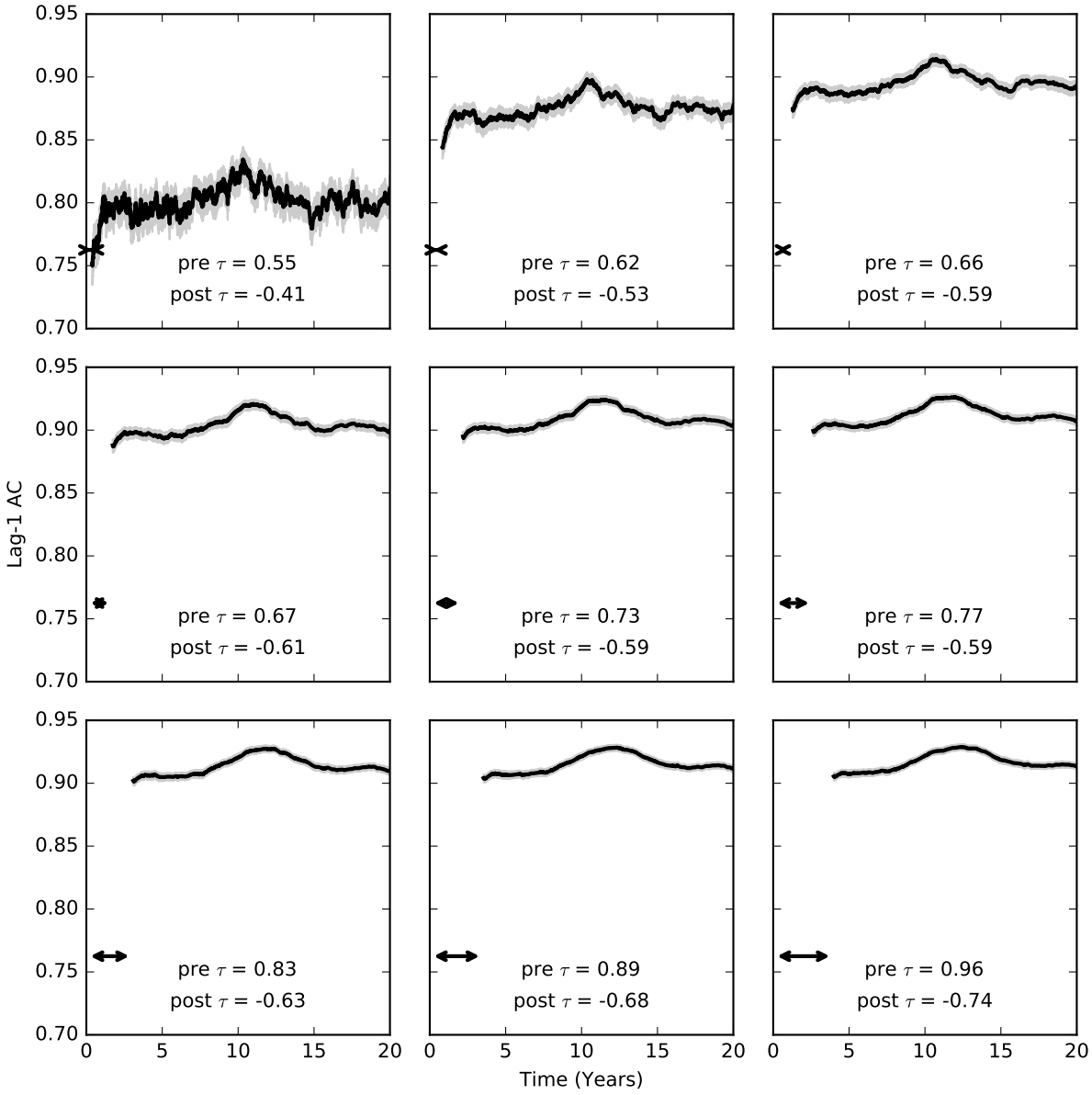


Figure A.14: Sensitivity Analysis of window width and lag-1 autocorrelation approaching and rebounding from the critical transition. Grey shaded regions are 95% confidence intervals constructed from 500 simulations with different random seeds. Epidemiological parameters are fixed in this instance. The length of the rolling window is different in each panel, and is indicated by the black arrow.

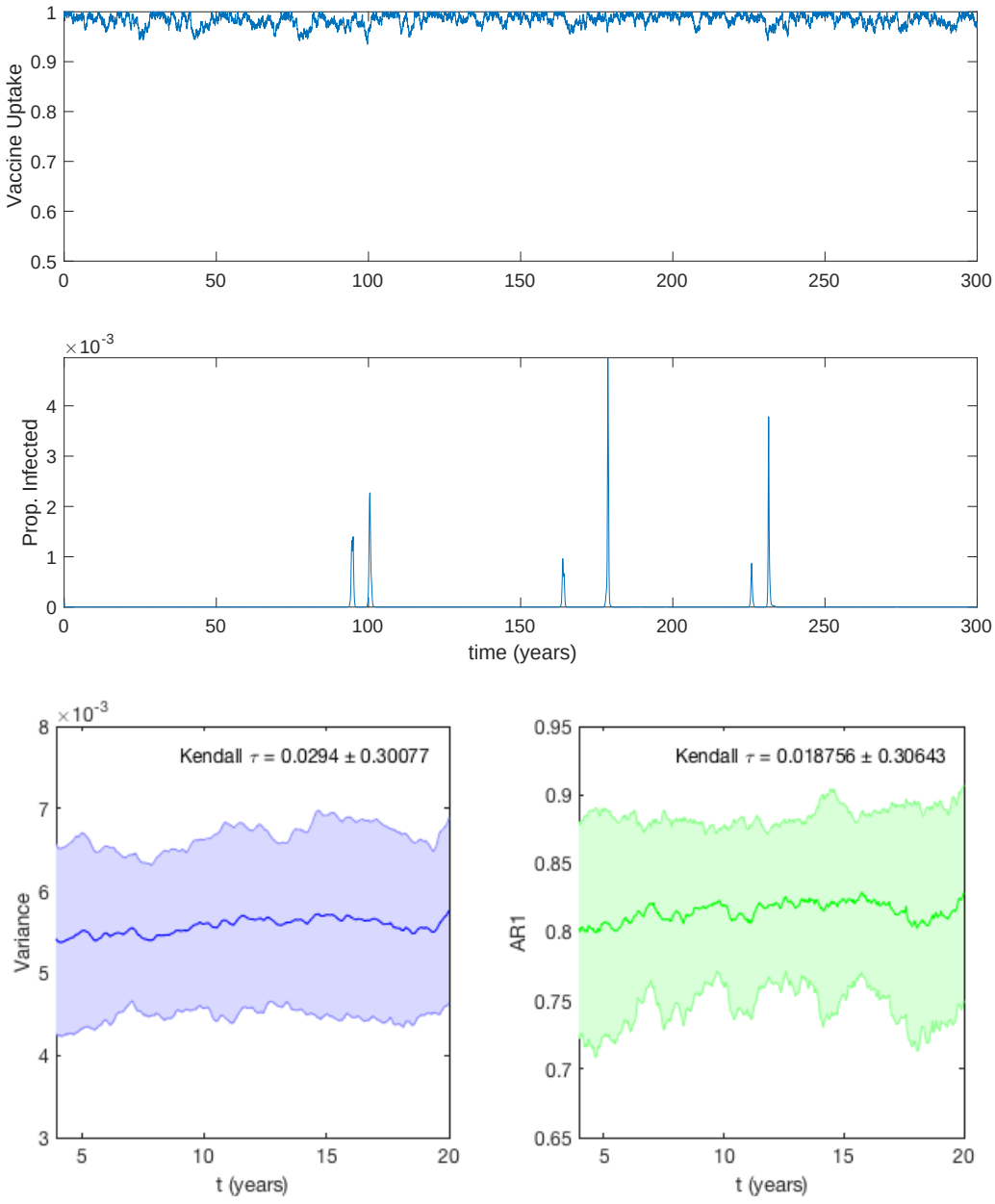


Figure A.15: Transient dynamics of system far from the critical transition. The relative risk is constant throughout the simulation. As can be seen, the variance (bottom left) and lag-1 autocorrelation (bottom right) do not change. This is corroborating evidence that the observed increases in the variance and lag-1 autocorrelation observed in previous figures is due to the variation of the relative risk.

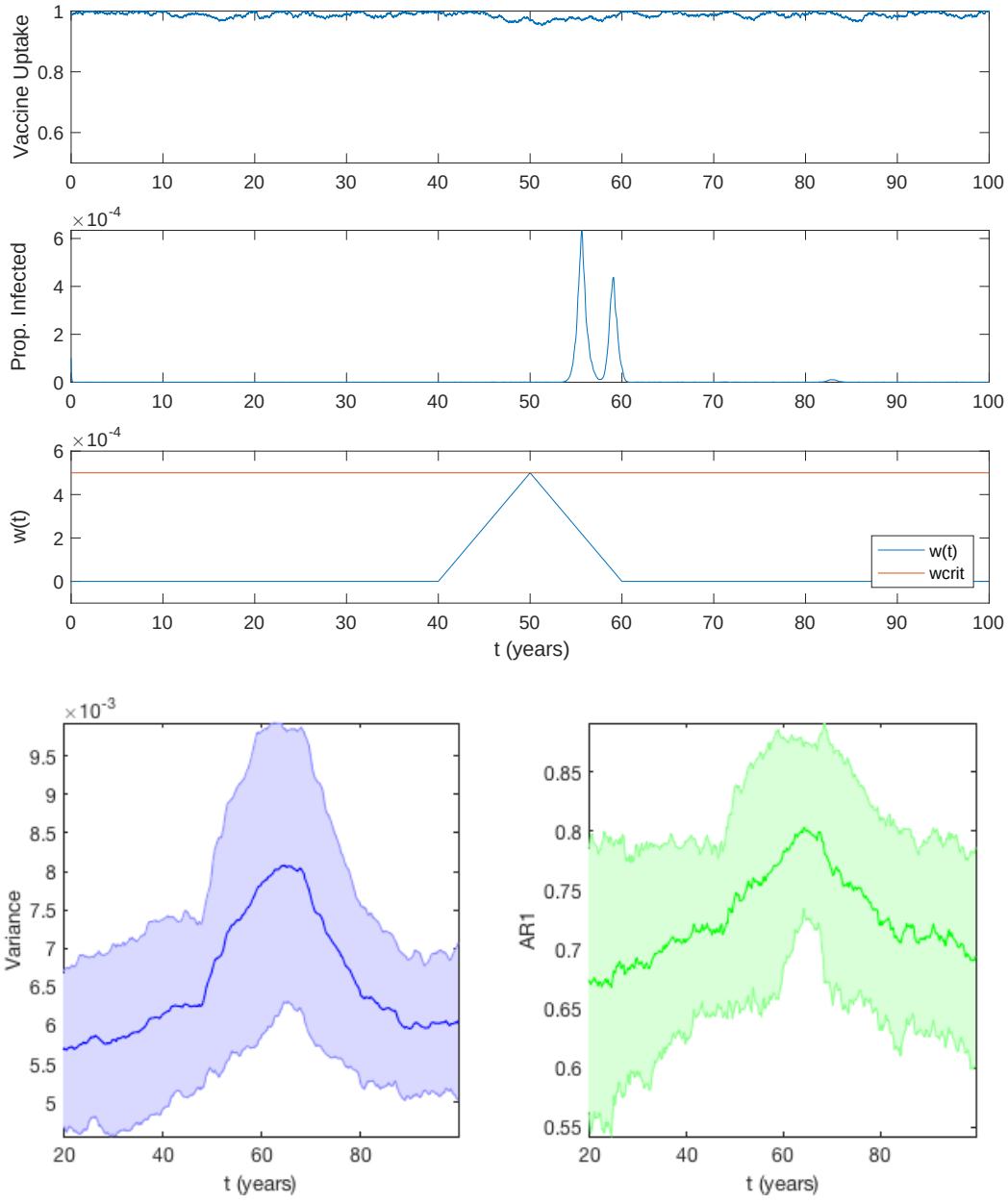


Figure A.16: Dynamics of the coupled behaviour-disease system as w is brought near the critical transition. The top three panels are vaccine coverage $x(t)$, infection prevalence $I(t)$, and relative risk $w(t)$. The red line in panel three is where the critical transition occurs. Variance and lag-1 autocorrelation (bottom left and right panels) increase and decrease as relative risk approaches and rebounds from the critical transition.

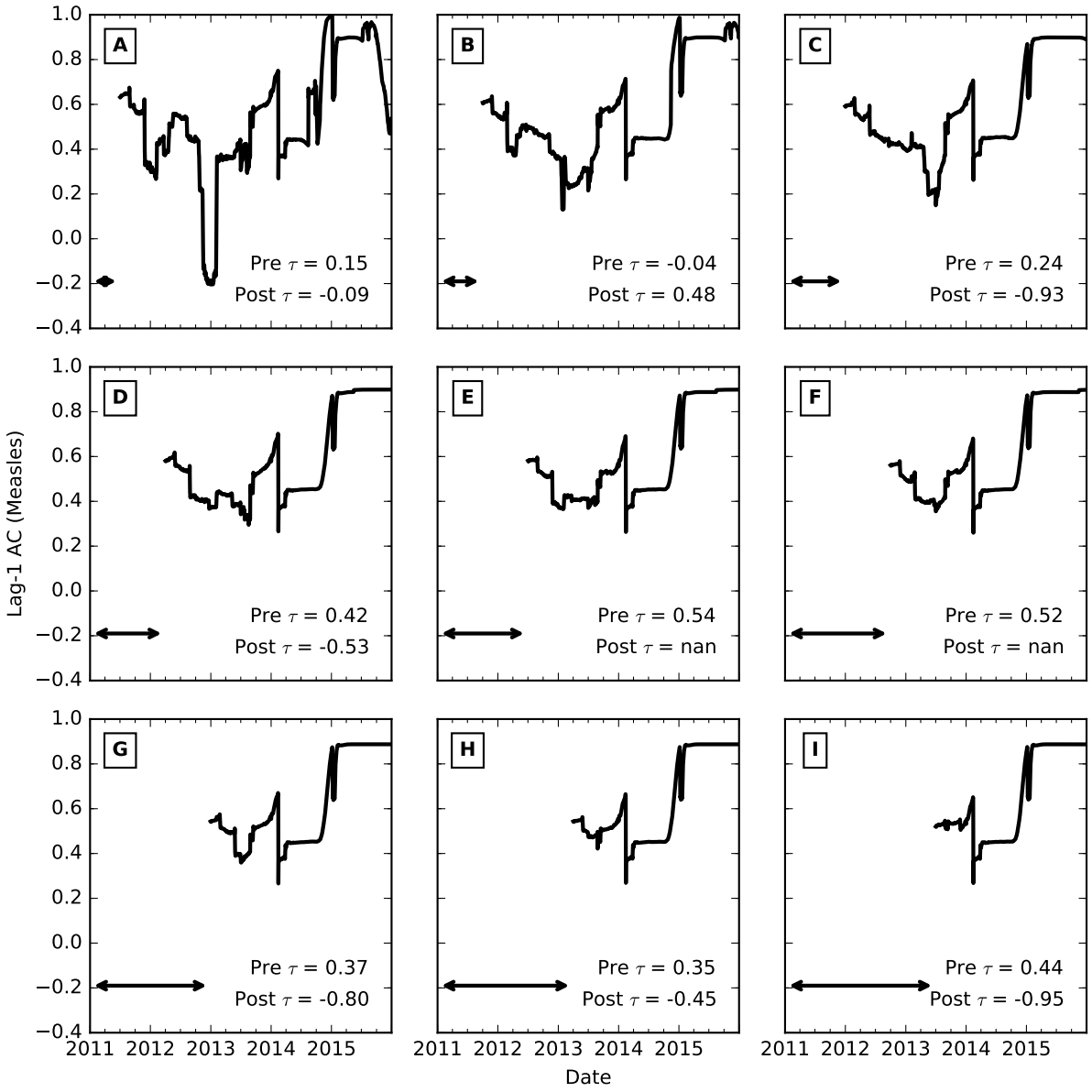


Figure A.17: Sensitivity analysis of window width and lag-1 autocorrelation for "measles" Google searches in California

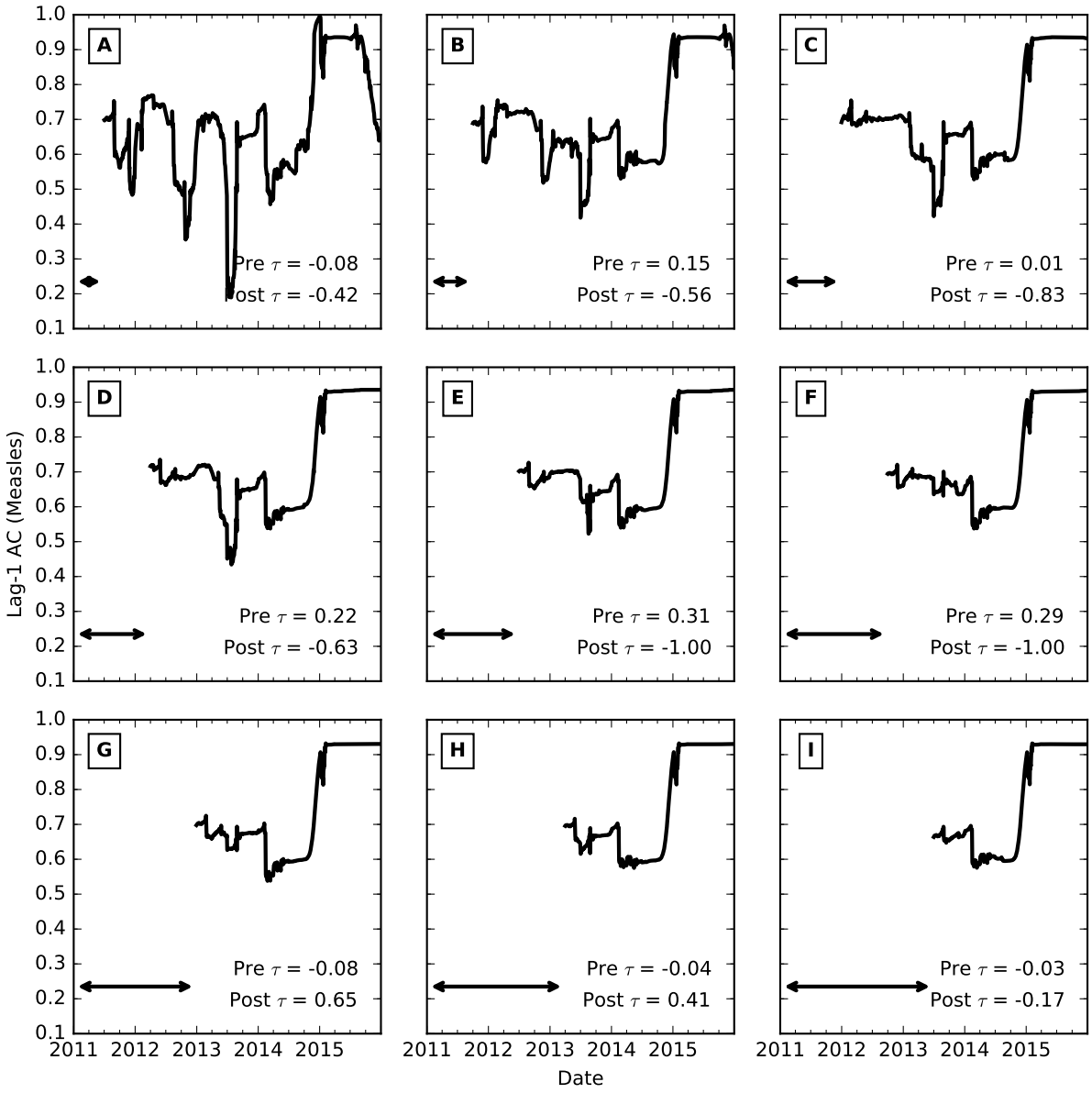


Figure A.18: Sensitivity analysis of window width and lag-1 autocorrelation for "measles" Google searches in The United States

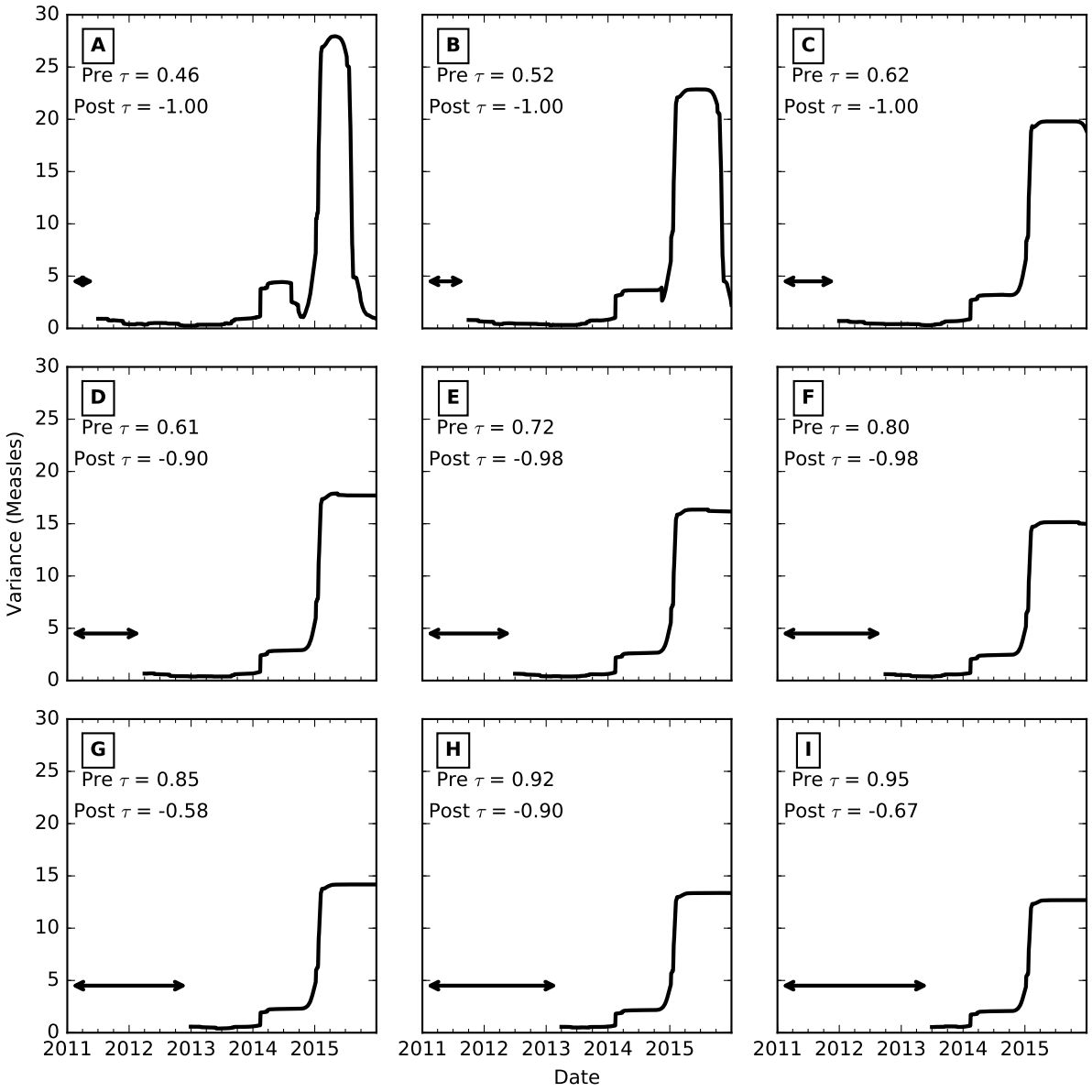


Figure A.19: Sensitivity analysis of window width and variance for "measles" Google searches in California

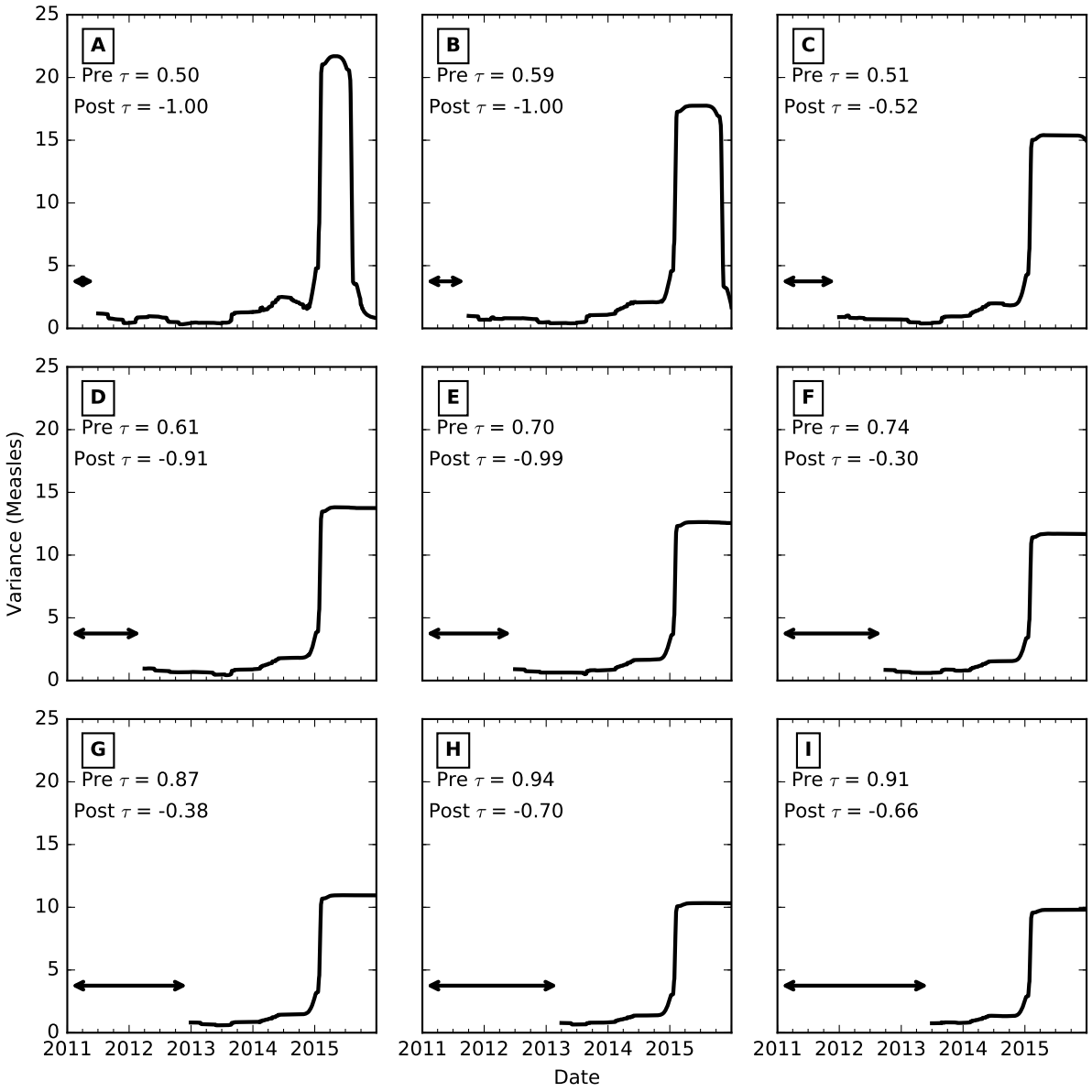


Figure A.20: Sensitivity analysis of window width and variance for "measles" Google searches in The United States

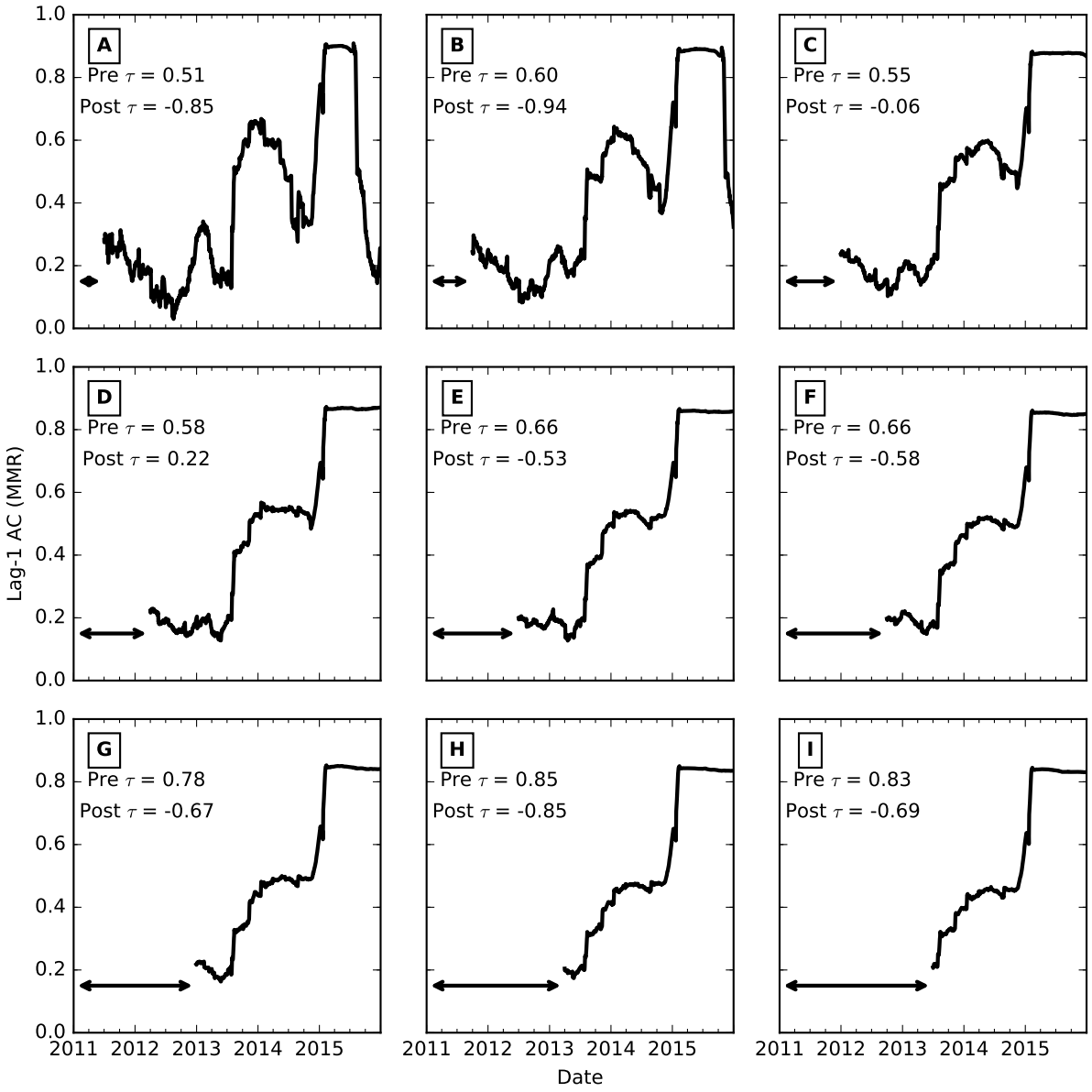


Figure A.21: Sensitivity analysis of window width and lag-1 autocorrelation for "MMR" Google searches in California

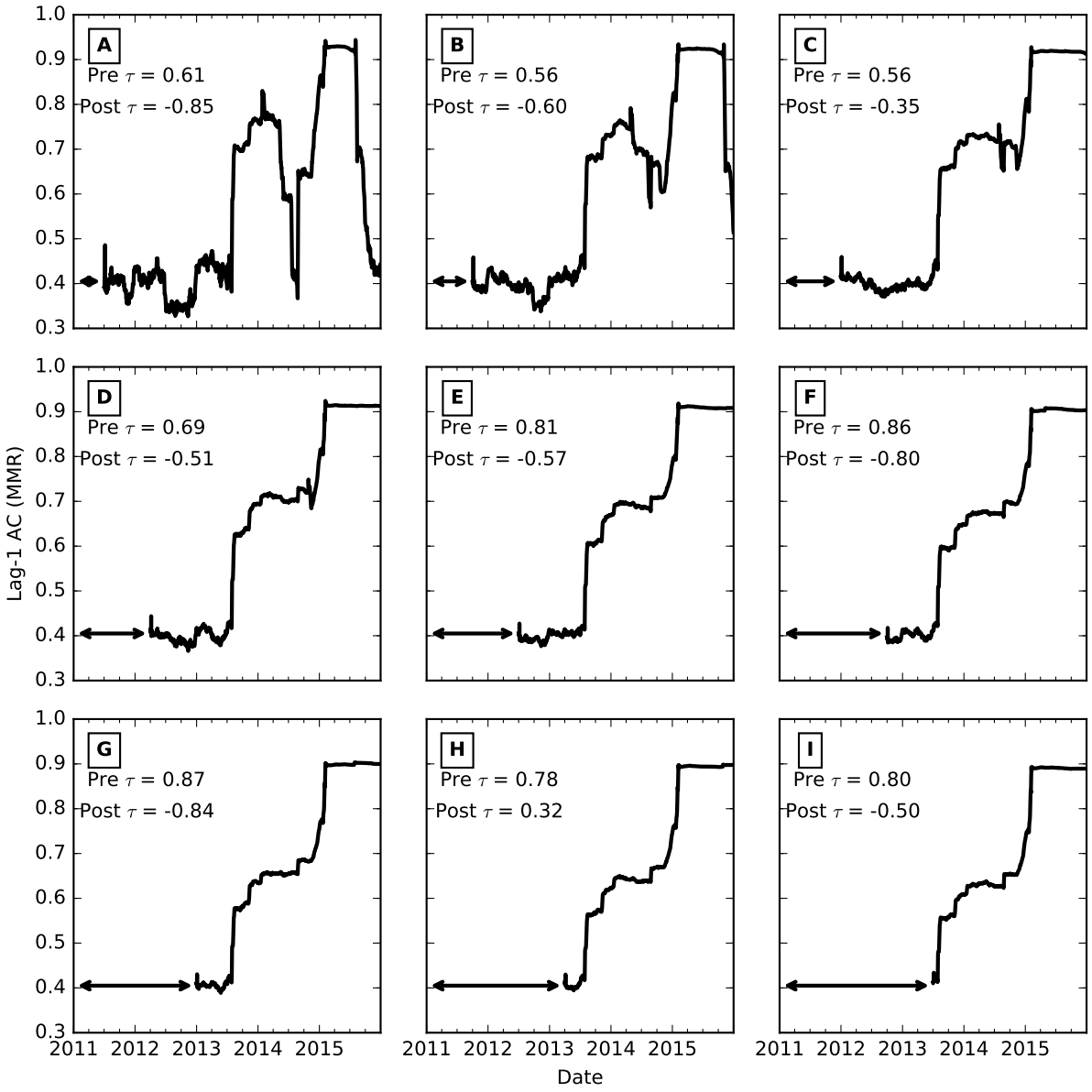


Figure A.22: Sensitivity analysis of window width and lag-1 autocorrelation for "MMR" Google searches in The United States

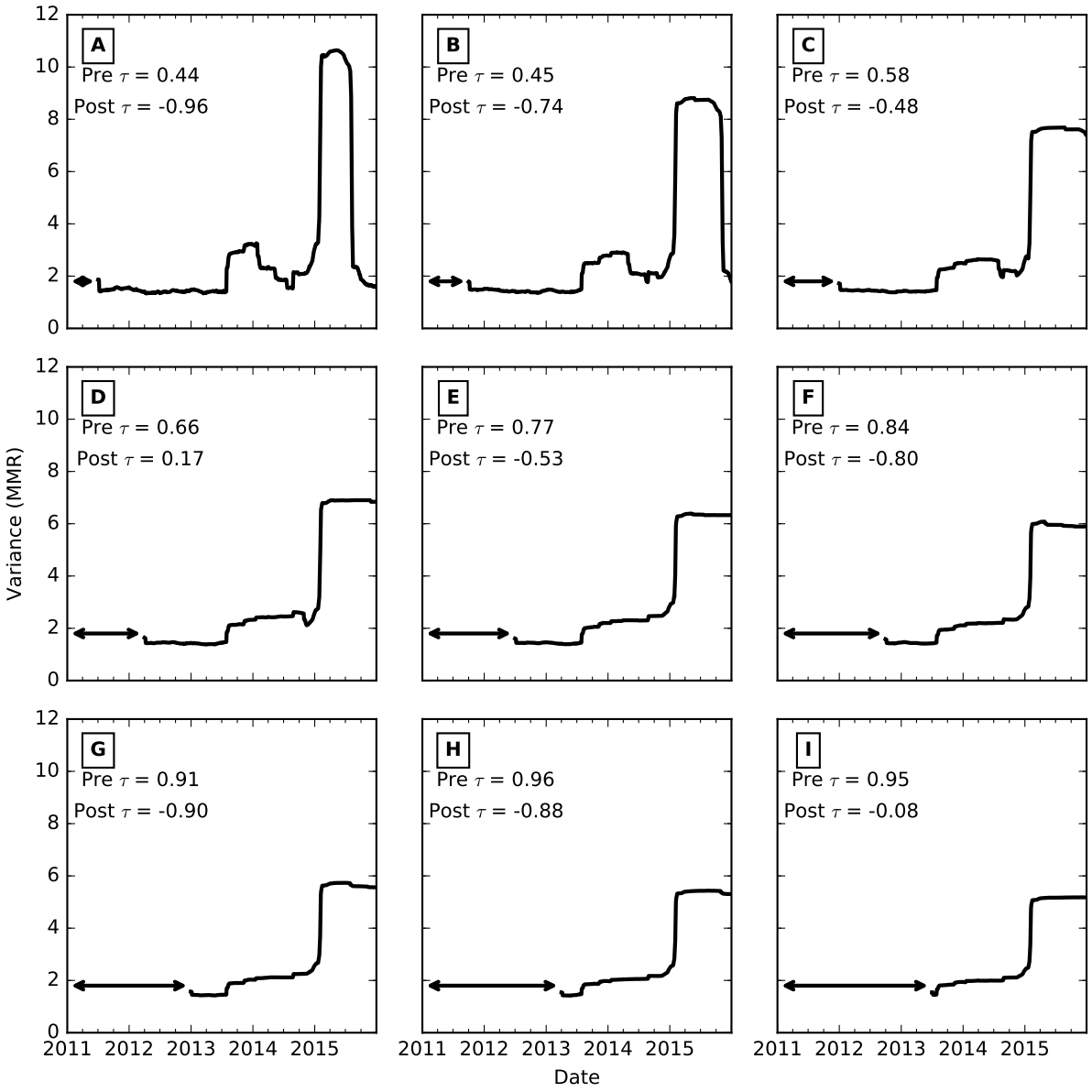


Figure A.23: Sensitivity analysis of window width and variance for "MMR" Google searches in The United States

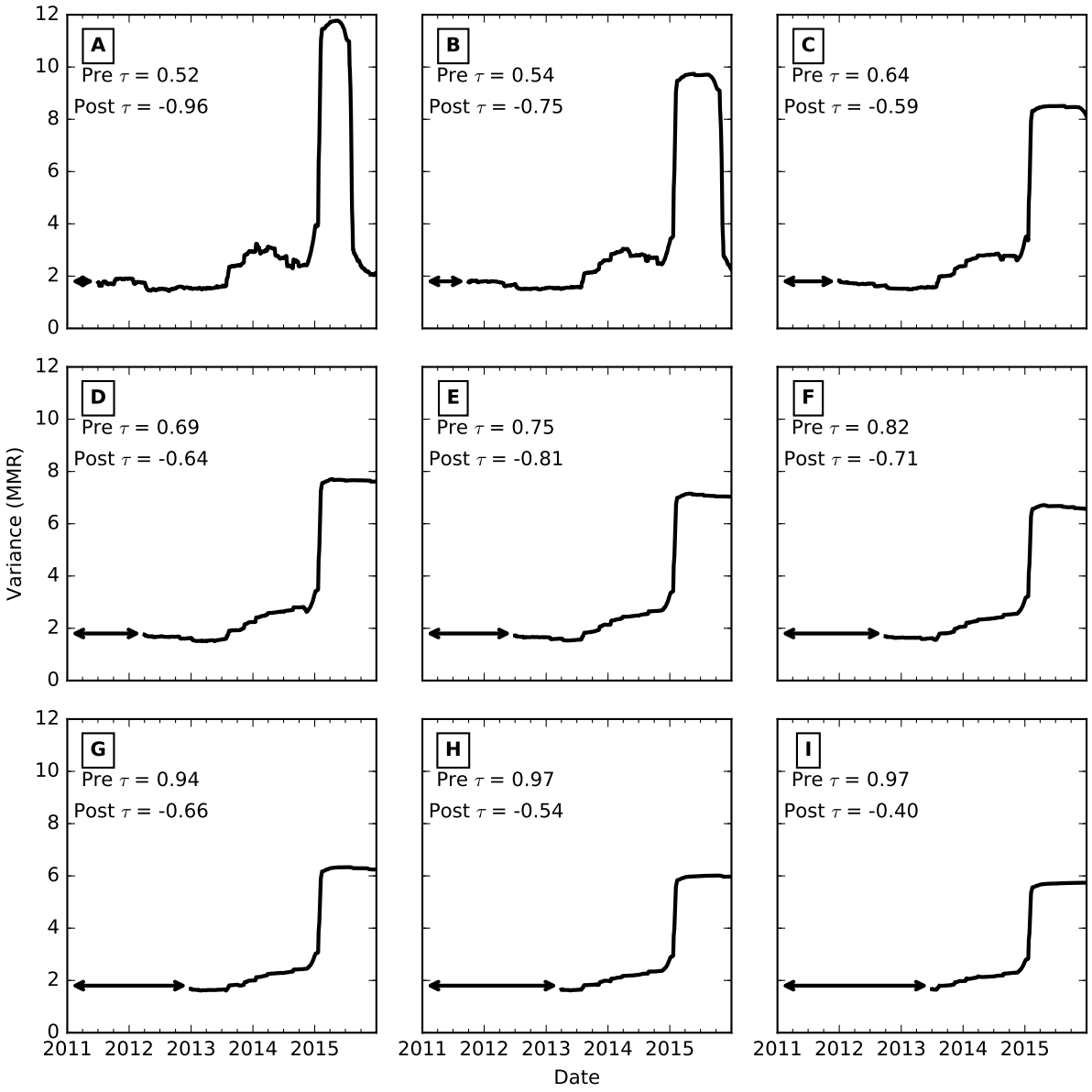


Figure A.24: Sensitivity analysis of window width and variance for "MMR" Google searches in California

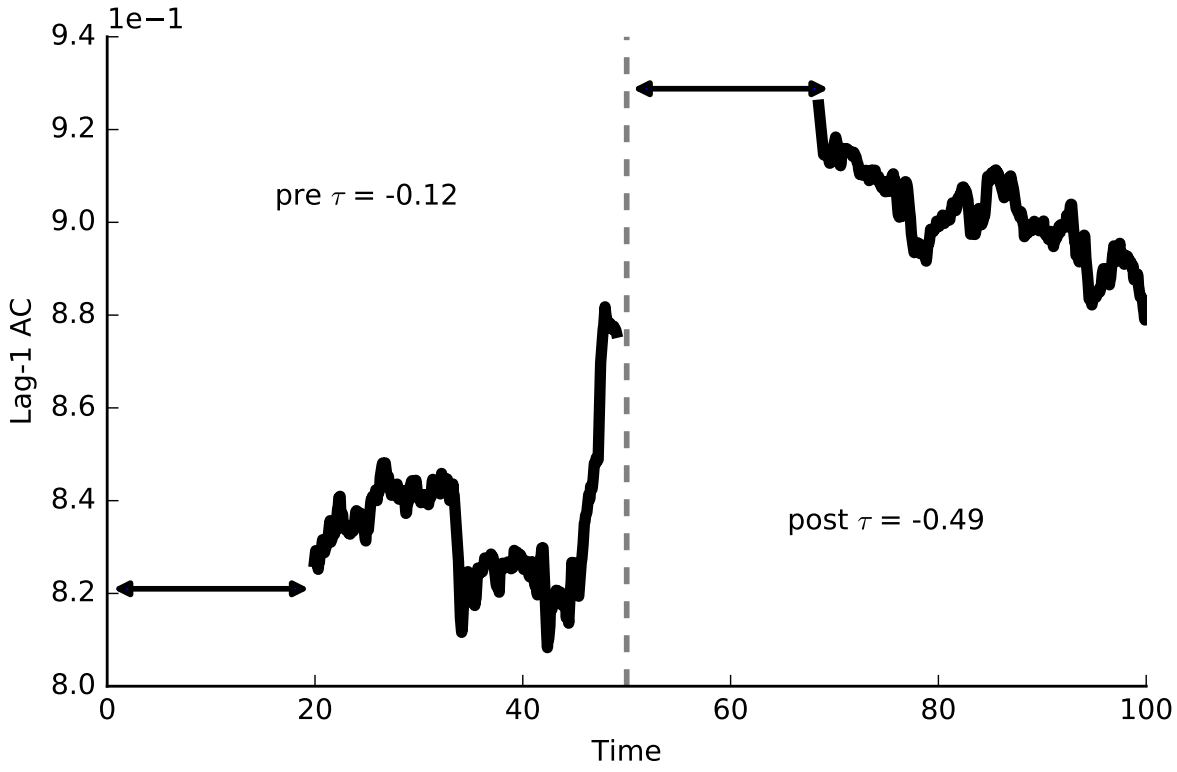


Figure A.25: The lag-1 autocorrelation in the times leading up to, and immediately after, the peak in relative risk for the stochastic coupled behaviour-disease model. The data within a one year radius has been removed and the lag-1 autocorrelation is calculated for each half of the remaining data. Length of the rolling window is 20 years, indicated by the length of the black arrows. Though the pre peak Kendall τ coefficient is negative, there is still a clear increase in the lag-1 autocorrelation as the peak in relative risk draws nearer, which is parsimonious with theory.

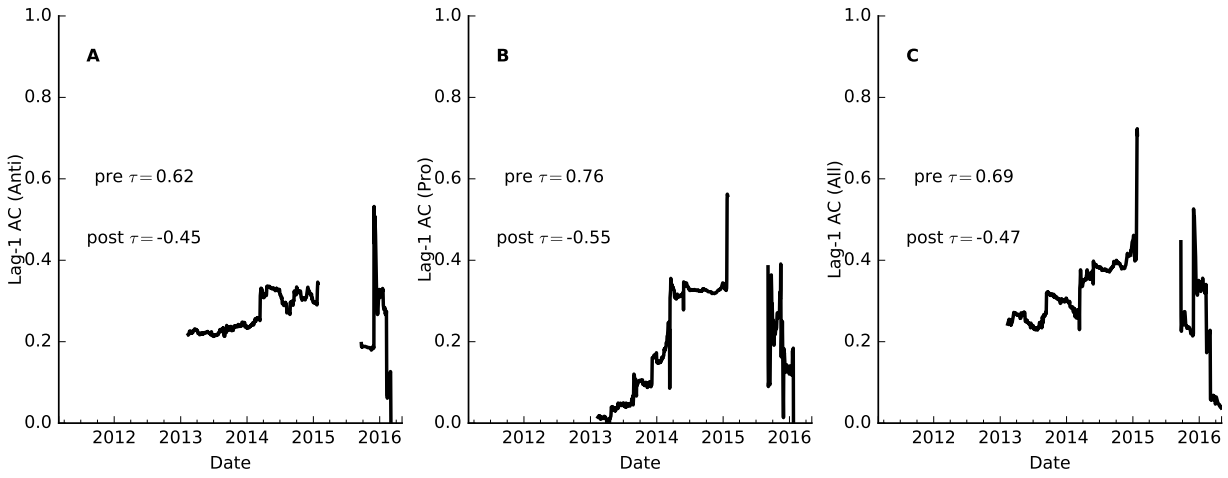


Figure A.26: An analysis of the early warning signals in twitter data for **A** anti-vaccine sentimented tweets, **B** pro-vaccine sentimented tweets, and **C** all tweets. The data within a 5 day radius of the spike in twitter activity is removed and the lag-1 autocorrelation is calculated for each half accordingly. Anti-vaccine, pro-vaccine, and all tweets show a positive Kendall τ leading to the spike in activity and a negative Kendall τ after the activity. This suggests that the relative risk increased up until the CDC reported on the outbreak, and then the coupled nature of the system forced the relative risk to decrease due to disease incidence.

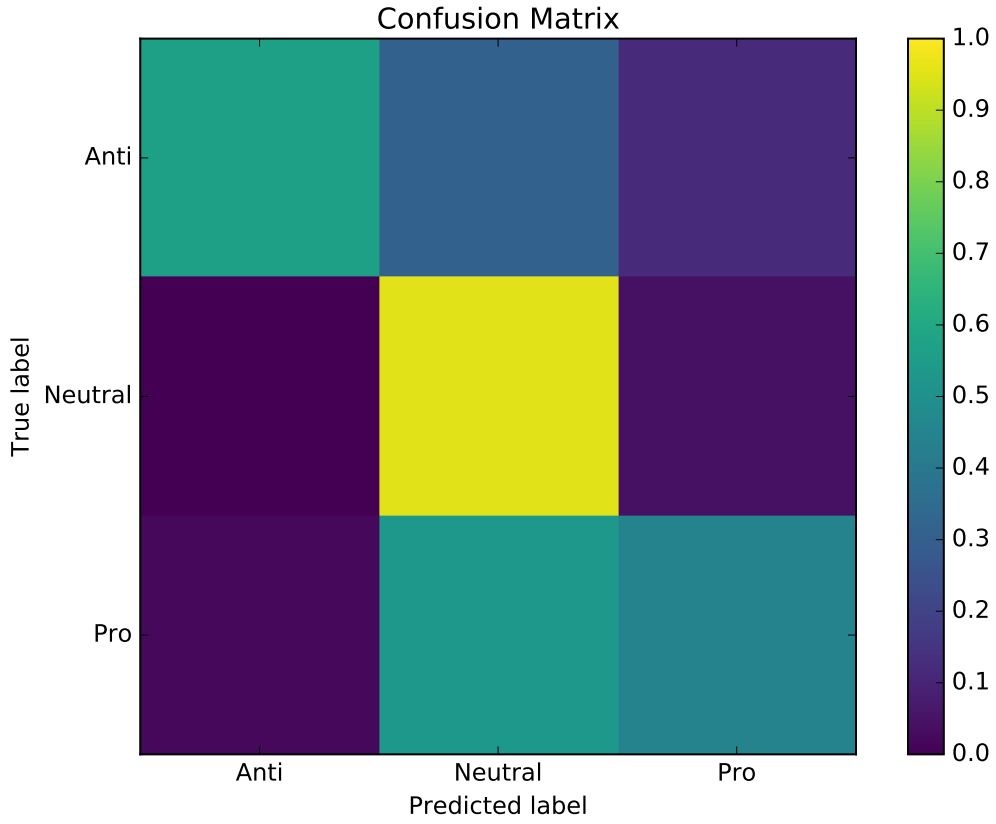


Figure A.27: A visualization of the confusion matrix for the classifier used to classify tweets as either anti-vaccine, pro-vaccine, or neutral. A color closer to yellow shows the classifier performs well. From this visualization, it is clear that the classifier is often confusing pro-vaccine and anti-vaccine tweets as neutral This confusion leads to poorer performance.

Appendix B

Code

B.1 Code for SDE Simulation

```
Disease Parameters

mu = 1/50; R = 16;
gamma = 365/13; b = R*(mu+gamma);

Parameters of interest

k = 500;
delta = 5e-4;
alpha = 1.0; %How far are we coming to the bif point.

Time Parameters
time_step = 5e-4;
T = 100;
t = 0:time_step:T;

w_start = 30;
w_end = 70;
w_mid = (w_start + w_end)/2;
w_spread = (w_end - w_start)/2; %symmetric

w = @(t) (alpha*delta)*( heaviside(t - w_start) - heaviside(t - w_end)).*(1-1/w_spread*(
    abs(t-w_mid)));

system = @(t,y)[mu*(1-y(3))-b*y(1)*y(2)-mu*y(1);... %S

                b*y(1)*y(2)-gamma*y(2)-mu*y(2);... %I

                k*y(3)*(1-y(3))*(-w(t)+y(2)+delta*(2*y(3)-1))]; %x

 IC
 y0 = [0.010 0.0000 0.9900];

Noise Params
sigma = 0.01;
G = @(t,y) sigma*[1;1e-4;1];

Solver Options
```

```
ops = sdeset('SDEType','Ito','ConstGFUN','yes','NonNegative','yes');
1
Solve

y = sde_euler(system,G,t,y0,ops);

y = min(y,1);
I = y(:,2);
x = y(:,3);
%
subplot(3,1,1);
plot(t,I)
axis([-inf, inf, -inf, 5e-3])
subplot(3,1,2);
plot(t,x);
axis([-inf, inf, 0.85, 1])

subplot(3,1,3);
plot(t,w(t));

[vals, idx] = intersect(t,0:5e-2:100);

t = t(idx);
x = x(idx);
I = I(idx);

X = [t' w(t') I x];

csvwrite('triangular_data.csv',X)
```

B.2 Smoothing Data in R

```
setwd('/Users/Demetri/Google Drive/Work for Paper/Figures/')


which = 'tweets'

data_file = paste(which,'csv',sep = '.')

write_me = paste(which,'_smooth',sep = '')

write_me = paste(write_me,'csv', sep = '.')

data = read.csv(data_file)


x = as.matrix(data[,2])
timeindex = 1:dim(x)[1]


bw = round(bw.nrd0(timeindex))
smoother = ksmooth(timeindex,x,kernel = 'normal', bandwidth = bw, range.x = range(
    timeindex), x.points = timeindex)

res = x -smoother$y

residual_matrix = matrix(0,nrow = dim(x)[1],ncol = 1)

residual_matrix[,1] = res


residual_matrix = data.frame(residual_matrix)
colnames(residual_matrix) = c('smoothed')
write.csv(residual_matrix,write_me, row.names = FALSE)
```

B.3 Data Analysis in Python

```
1
2  import pandas as pd
3  import numpy as np
4  from scipy.statistics import kendalltau
5
6  df = pd.read_csv('residual.csv')
7
8
9  #Lag-1 Autcorr
10 lag1autocorrelation = df.residual.rolling(len(df)/2).apply(lambda x: pd.Series(x).autocorr
       (1))
11
12 #Variance
13 variance = df.residual.rolling(len(df)/2).apply(lambda x: pd.Series(x).std() )
```

B.4 Machine Learning in Python

```
#!/Users/Demetri/Library/Enthought/Canopy_64bit/User/bin/python


#Test-Train Split
#Note that Test/Train splits tend to overestimate the error.

#Read in Data
import pandas as pd

#Classifiers and other MLA functions. These tend to be the best/most frequently used
    text classifiers
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import SGDClassifier
from sklearn.svm import SVC, LinearSVC
from sklearn import cross_validation
from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score,
    precision_score, recall_score
from sklearn.grid_search import GridSearchCV
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2

import pickle
import os

#plotsSGD
import matplotlib.pyplot as plt
import numpy as np


#I am going to visualize a confusion matrix. Here is some nice code to do that in MPL

def plot_confusion_matrix(cm, title='Confusion matrix', cmap=plt.cm.viridis):
    plt.figure()
    plt.imshow(cm, interpolation='nearest', cmap=cmap, vmax = 1, vmin = 0)
    plt.title(title)
    plt.colorbar()
```

```
    tick_marks = np.arange(3)
    plt.xticks(tick_marks, 'Anti Neutral Pro'.split(), rotation=0)
    plt.yticks(tick_marks, 'Anti Neutral Pro'.split())
    plt.tight_layout()
    plt.ylabel('True label')
    plt.xlabel('Predicted label')

#Close any old plots
plt.close('all')

#Load the training set
training_set = pd.read_csv('Training Data/New_Training_Data.csv')


#test-train split
X_train,X_test,y_train,y_test = cross_validation.train_test_split(
                                    training_set.Tweet, #Characteristics
                                    training_set.Class, #Classes
                                    test_size= 0.40, #Size of Split
                                    random_state = 1
                                    )

#Now, we will build the pipeline for our classifier
# Training Data --> Vectorizer-->TfIdf-->Feature Selection-->Classifier



tf_idf_vectorizer = TfidfVectorizer()
selector = SelectKBest(chi2)
classifier = SVC(kernel = 'linear')


#Set up the pipe for the pipeline
pipe = [

        ('TF_IDF_V',tf_idf_vectorizer),
        ('features',selector),
        ('classifier',classifier)

        ]
```

```
#Build the pipe
CLF = Pipeline(pipe)


#Parameter Space for GSCV
parameters = {
                  'TF_IDF_V__ngram_range':((1,1),(1,2)),
                  'TF_IDF_V__use_idf':(True,False),
                  'features__k':(100,500,1000,2000,'all'),
                  }




#Perform the GS
grid_search = GridSearchCV( CLF, parameters, n_jobs = -1, verbose = 0,cv = 3, scoring
    = 'accuracy')


#Make the classifier the best estimator
CLF = grid_search.fit(X_train,y_train)




print('\nResults of Grid Search CV\n')
print("Best score: %0.3f +/- %0.3f" % (grid_search.best_score_,2*CLF.grid_scores_[-1][2].
    std()) )
print("Best parameters set:")
best_parameters = CLF.best_estimator_.get_params()
for param_name in sorted(parameters.keys()):
        s = "\t%s: %r" % (param_name, best_parameters[param_name])
        print s


clf = CLF.best_estimator_



y_predicted = clf.predict(X_test)

```

```
115
116
117 print('\nClassifier Report:\n')
118 print(classification_report(y_test, y_predicted))
119
120 cm = confusion_matrix(y_test,y_predicted)
121 cm = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]
122 print('\nClassifier Confusion Matrix\n')
123 print(cm)
124 print('\nPrecision Score\n')
125 print(precision_score(y_test, y_predicted,average = None))
126 print('\nRecall Score\n')
127 print(recall_score(y_test, y_predicted,average = None))
128
129
130 plot_confusion_matrix(cm, title = 'Confusion Matrix')
131 plt.show()
132
133
134
135
136 with open('SVC.txt','w') as f:
137
138     s = pickle.dumps(clf)
139     f.write(s)
```