Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Nishant Kishore Date

Flying, Phones and Flu: An evaluation of Keflavik International Airport and its role in the introduction of pandemic H1N1 into Iceland in 2009 using anonymized call records

By

Nishant Kishore

Degree to be awarded: Master of Public Health

Department of Global Epidemiology

_____

Timothy L. Lash, DSc, MPH

Committee Chair

_____

Ymir Vigfusson, PhD

Committee Member

Flying, Phones and Flu: An evaluation of Keflavik International Airport and its role in the introduction of pandemic H1N1 into Iceland in 2009 using anonymized call records

By

Nishant Kishore

Bachelor of Science

The College of William and Mary

2010

Thesis Committee Chair: Timothy L. Lash, DSc, MPH

An abstract of

A thesis submitted to the Faculty of the Rollins School of Public Health of Emory University in partial fulfillment of the requirements for the degree of Master of Public Health in the Global Epidemiology Department

2016

ABSTRACT

Flying, Phones and Flu: An evaluation of Keflavik International Airport and its role in the introduction of pandemic H1N1 into Iceland in 2009 using anonymized call records

By Nishant Kishore

INTRODUCTION: We studied the introduction of pandemic H1N1 to Iceland; an isolated island with centralized national health records, near-ubiquitous mobile phone use, and one likely port of entry: Keflavik International Airport. Using anonymized call detail records linked with health records we evaluated the role that international travelers played in the introduction and propagation of pandemic H1N1 in Iceland by quantifying the association between airport exposure and influenza-like-illness (ILI) diagnosis.

METHODS: This was a nested case-control study comparing odds of exposure to Keflavik International Airport among cases and matched controls producing a longitudinal two-week matched odds ratios (mORs). We further evaluated rates of infection among 1st degree connections of cases compared to their matched controls.

RESULTS: The longitudinal two-week mOR produced for individuals who were exposed to the airport in the 4 days before ILI diagnosis showed an increase in the two-week mOR in the early stages of the epidemic from August 17th until August 31st with a mOR of 2.00 (95% CI: 1.4, 2.9). During the two week period from August 17th through August 31st we calculated the two-week IDR of infection among 1st degree connections to be 14.2 (95% CI: 5.7, 35). The IDR decreased steadily to a threshold IDR of approximately 5 during the epidemic peak.

CONCLUSIONS: We find that there is an association between exposure to Keflavik International Airport and incident ILI diagnoses during the initial stages of the epidemic. Our data show a definitive high rate of transmission earlier in the epidemic. However, even during generalized epidemic in the population, 1st degree connections of individuals diagnosed with an ILI get sick at a rate 5 times higher than the 1st degree connections of their matched healthy controls. Our methods were validated through evaluation of domestic airports as negative controls. Bias analyses showed minimal threats to the validity of our measures of association, assuming the validity of our bias model. Through greater collaboration with both mobile network operators and health officials, the techniques described in this study can be used for hypothesis-driven evaluations of locations and behaviors during an epidemic and their associations with health outcomes.

Flying, Phones and Flu: An evaluation of Keflavik International Airport and its role in the introduction of pandemic H1N1 into Iceland in 2009 using anonymized call records

By

Nishant Kishore

Bachelor of Science

The College of William and Mary

2010

Thesis Committee Chair: Timothy L. Lash, DSc, MPH

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of Master of Public Health in

Global Epidemiology.

2016

ACKNOWLEDGEMENTS

This analysis would not have been possible without the original work by Dr. Guðrún Sigmundsdóttir, Dr. Leon Danon and Dr. Ymir Vigfusson in sourcing, linking and creating the anonymized call detail records dataset.

I cannot express enough thanks to my committee, Dr. Timothy Lash and Dr. Ymir Vigfusson and the epidemiologists who provided invaluable insight and feedback, Dr. Carrie Reed and Dr. Rebecca Mitchell. This thesis could not have been completed without their patience and guidance.

Table of Contents

Introduction

In April of 2009, the United States reported the first cases of H1N1 for that influenza season (1). Over the next year, pandemic H1N1 resulted in 9,000 to 18,000 deaths worldwide (2,3). Within weeks, Argentina, Thailand, Gabon, the Central African Republic, South Africa and Kenya all reported introduction of the H1N1 virus into their populations, with international travelers playing a key role in transmission between continents (4–8). Urban regions showed a higher incidence of cases than rural regions, likely due to greater population densities and their centralized position in travel networks (9). The pandemic then spread outwards from these travel hubs taking advantage of the relatively innocuous initial symptoms of an infectious carrier (5,9,10). A highly connected air travel network facilitated the spread of the disease, with the number of cases rapidly increasing after the initial introduction of the virus in each country (2,11,12). The transmission dynamics of influenza have continued to be a significant point of interest for researchers with emphasis on modes and mechanisms of introduction into new regions (5,11,13).

During the height of the pandemic, interventions to limit the introduction of the virus into new regions focused on screenings at high-volume ports of entry such as airports (11,12,14). Research on transmission dynamics of pandemic H1N1 has primarily focused on two areas: 1) transmission of the virus within planes (11), and 2) the effectiveness of airport screening protocols in identifying cases for quarantine (15–21). Transmission within an airplane was found to be concentrated in the seats located closest to the index case; however, air travel was suggested to be more important as a mode of transport rather than a major point of transmission for the virus. Once passengers landed,

screenings failed to effectively identify cases due to the delay in the appearance of symptoms following infection (17,22,23,16,20,19,18,15). For example, in New Zealand, the screening protocol at Auckland International Airport was found to be only 5.8% sensitive at identifying infected individuals, while other evaluations highlighted the difficulty of identifying asymptomatic and latent passengers (14,16,21).

A focus of post-pandemic network research has been the description, identification and reduction of high-risk nodes and edges in global air travel networks. In an analysis of international airports, a network including only the largest airports in the largest, most-connected and most central cities, accounting for fewer than 10% of all international airports, adequately accounted for the inter-regional disease transmission. Specific global transfer hubs played a larger role in the spread of the pandemic than individual airports (24,25).

Importantly, it is only the fraction of travelers who make disproportionately more journeys than the rest who are of particular importance in disease transmission (26). In situations when these travelers are infected early in a pandemic and the general epidemic growth rate is not high, these travelers can play a pivotal role in the acceleration of international spread of the infection (3,27). Because high-volume airports are central to the propagation of an epidemic, targeted flight cancelations decrease the rate of spread of the epidemic to a greater degree than closing an entire airport, again placing the burden of disease transmission on individuals rather than entire airports (28). While simulated network data and retrospective analysis provide important perspective on transmission of

H1N1 into new populations, a deeper look at international travelers and their role in the introduction of disease into a population using real data is needed.

This study focuses on the introduction of pandemic H1N1 to Iceland; an isolated island with centralized national health records including influenza like illness (ILI) diagnoses, near-ubiquitous mobile phone use, and one likely port of entry: Keflavik International Airport. It leverages a unique opportunity to use anonymized telecommunications call detail records, metadata provided by one of Iceland's largest mobile network operators during the period of the pandemic, linked with health records provided by the Chief Epidemiologist at the Center for Health Security and Communicable Disease Control of the Directorate of Health in Iceland (CHS-CDC).

According to the CHS-CDC, 9,887 cases of influenza-like illness (ILI) diagnosed from July through December of 2009 were considered to be cases caused by pandemic influenza. It was estimated that more than 1 in 10 Icelanders were infected through the course of the pandemic when considering underreporting and asymptomatic cases (29). Social networks play an important role in tracking and predicting the chain of transmission of diseases such as H1N1 by acting as a proxy for physical proximity and thereby the risk of transmission (30–32). In 2009, over 96% percent of the adult population in Iceland reported using a mobile phone (33). Due to the large-scale adoption of cellular technology in Iceland, the telecommunications dataset used in this study serves as a similar proxy while also providing geospatial information on individual movement. In 2009, Icelanders made over a third of all trips out of Keflavik International Airport (33), often using their mobile network while at the airport, thereby facilitating a

granular, yet anonymized, examination of potential travelers at Iceland's largest international port of entry. This study primarily evaluates the role that international travelers played in the introduction and propagation of pandemic H1N1 in Iceland by quantifying the association between international travel and incident H1N1 cases through the course of the epidemic. In doing so, this study also examines the relevance of call detail records to epidemiologic research.

Methods

Study Population and Design

We performed a nested case-control study of Icelanders diagnosed with an ILI between January 2009 and March 2010 and a sample of the source population. The source population consisted of 342,369 distinct phone numbers belonging to Icelanders who owned and used a personal mobile phone operated by the largest mobile phone operator (MNO) in the country, Síminn, during the study period. The numbers include individuals with multiple phones, accounts with more than one MNO and those who may have switched operators several times during the study period creating separate identifiers. The CHS-CDC recorded 9,887 incident ILI cases during the study period, 4,347 of which were among clients of the sample MNO, based on their mobile phone number. No demographic or personal identifiable information, such as age or gender, was linked to this dataset in accordance with privacy standards. In 2009, this record likely contained mainly adults and teenagers, with children too young to own a phone excluded from the dataset.

Datasets – MNO Call Detail Records

The call detail records provide anonymized mobile phone use data from 30% to 40% of the Icelandic population over the course of 18 months, including the 6 months at the peak of pandemic H1N1 from August through December of 2009. The mobile call detail records database included 1,517,276,930 calls, texts and data interactions made by 342,369 clients of the sample MNO from 483 tower locations during the study period. It contained unique anonymous IDs for senders and receivers of the interaction, cell tower GPS coordinates of the customer, timestamp of interaction, type of interaction (incoming

or outgoing call or text message) and length of the interaction. The data were logged automatically and provided directly from the MNO.

Datasets – CHS-CDC ILI Diagnoses

The database of ILI diagnoses contained a record of all individuals in the call detail records database who were diagnosed with an ILI, their date of diagnosis, and their unique anonymized ID, comprising approximately half of all suspected ILI cases. Icelandic physicians were required to enter their diagnoses in electronic patient journals where "*ICD-10 codes for ILI and confirmed influenza were automatically selected and reported within 24 hours via a closed electronic network to the CHS-CDC comprising all healthcare centers and hospitals in Iceland*" (29).

While most demographic information was excluded by the MNO and CHS-CDC, a variety of metadata characteristics of mobile phone use, travel and even activity patterns were calculated from the data. Metrics derived from call-data-records such as calls per day, places visited and closest contacts help approximate social behaviors, travel patterns and friend networks. These were used to build a more robust model by accounting for potential underlying heterogeneity of behaviors and contact networks between individuals in our study (34).

Variables

Six covariates were generated to assist in risk-set sampling and adjustment for mobile phone use and behaviors. Location of a home tower was generated based on identification of the tower through which the majority of the user's interactions were routed between the hours of 7pm and 7am. Average numbers of interactions per week, average number of

individuals called per week, percent of interactions made at night, number of days active per week, and number of geographic locations visited per week, were also calculated as a proxy for social interaction.

We are interested in international travel as our exposure of interest and diagnosis of ILI as determined by the CHS-CDC as our outcome of interest. For our exposure of interest, mobile call detail records are a useful proxy for physical location only during regular mobile phone use. While good predictions could be made for periods of international travel, the lack of sensitivity, or knowing when individuals were traveling, limited the use of these predictions. For example, international travel could be defined as travel to the international airport followed no mobile interaction for at least a 24 hour period among individuals who generally have frequent mobile interactions. However, individuals may not have used their mobile device near the international airport before takeoff or after landing, and absence of mobile interaction does not necessarily indicate international travel. Since the data was collected in 2009, the bulk of the records are from calls and text requiring users to be actively making calls or sending texts for us to signal location and activity. GPRS records, which include automatic mobile data transactions, are present but limited in their use in 2009.

We classified exposure at two levels for the analysis. First, we defined cases as *exposed* (E1) if they had a mobile phone interaction routed through one of six cell towers exclusively serving the Keflavik International Airport in the 4 days before their ILI diagnosis including the day of diagnosis. Controls sampled from the at-risk cohort were similarly evaluated from the date of their matched case's ILI diagnosis. Second, we defined a sub-group of cases as *exposed (E2)* if they had a mobile phone interaction at the

airport, either preceding or following a lack of any mobile phone interaction for at least 48 hours, in the 10 days before an ILI diagnosis. The second definition assumes that lack of mobile phone interaction after or before a physical presence at the international airport indicates air travel outside of Iceland.

We conducted three types of risk set selection for controls. First, we selected randomly from the entire sample population (S1) for individuals who were still at risk for ILI diagnosis at the incidence date of the matched case and had call detail records in the two weeks before and after ILI diagnosis. Second, we matched individuals in the sample population to cases on home tower as the incident case, were still at risk and had call records in the two weeks before and after ILI diagnosis (S2). Third, we matched individuals in the sample population to cases home tower, average calls per week, average active days per week, average number of locations visited per week, average percentage of nocturnal mobile activity per week, average percentage of time spent at home per week, and those who had call records in the two weeks around incidence date (S3). Matching on averages used a one standard deviation caliper. With the second method we reduced the set we randomly selected from and accounted for baseline risk for ILI diagnosis that may have been associated with the individual's residence. With the third method we accounted for baseline residential risk and baseline social risk.

Data Cleaning

We used the Bandicoot (35) framework to generate individual user level metrics from the larger call data record dataset and restricted the sample to individuals for whom we were able to generate home tower locations and who were not considered outliers for the various covariates.

Analysis Plan

We generated the epidemic curve of ILI diagnoses in our sample during the study period and compared the proportion of incident cases per week between the exposed and the unexposed populations. A Kolmogorov-Smirnoff statistic was calculated to evaluate the distance between the cumulative density functions of the epidemic curves of exposed and unexposed cases. This analysis was repeated for all other populated locations in Iceland and a map generated to display the p-value measuring the statistical difference in curves.

We delineated a continuous five month period of evaluation from the start of August until the end of December of 2009, which included 91.7% of cases pertaining to the H1N1 epidemic and evaluated the odds of exposure to that location among cases compared to the odds of exposure to that location among controls. For all odds-ratios (ORs) we evaluated a moving two-week window of time resulting in a longitudinal two-week odds of exposure and 95% confidence interval for each day in our evaluation period. For S1 selection types we calculated a standard OR while for S2 and S3 selection types we calculated a matched OR (mOR). We also conducted a sensitivity analysis by varying the width of the window of interest for the longitudinal two-week OR estimate to a maximum window width of 28 days. Assuming the null hypothesis is true, we would expect an OR of 1 or lower as cases should have a lower odds of exposure to any single location compared to the rest of Iceland. We evaluated any deviation of the longitudinal OR above the null. Based on this evaluation, we identified the two week risk period in August during which we noted the highest odds of exposure among cases. We compared this to a two week period during the peak of the epidemic from October 6th through October 24th, 2009.

Identification of Cases of Interest

Based on the results from the longitudinal two-week mOR, we selected August 17th through August 31st as the period of time when odds of exposure among individuals who were exposed to the airport in the four days before ILI diagnosis were most elevated during the beginning of the epidemic. We selected a period of time from the epidemic peak from October 7th through October 23rd for comparison. We recorded cases, their matched controls and exposure status on both E1 and E2 levels for these time periods.

Positive Controls

We selected the Austurvöllur area in downtown Reykjavík and the Landspítali University Hospital, the largest hospital in Reykjavík, as positive controls. Reykjavík is the capital of Iceland and the home of approximately two-thirds of the population (33). The city is located a half hour drive from Keflavik International Airport. Assuming the validity of our evaluation framework, we expect to see increases in the odds of exposure to these locations among cases early in the epidemic compared to their matched controls as they are both locations where we expect increased risk of transmission early in the epidemic.

Negative Controls

There are several domestic airports in Iceland that provide regular passenger and cargo transport across the country. The airport in Reykjavík provides the largest number of domestic flights followed by Akureyri, Vestmannaeyjar, Egilsstaðir and Ísafjörður. The mobile tower at Reykjavík airport began providing reliable service in 2010 and therefore we recorded no cases as being exposed to this airport during the pandemic. Due to lack of call detail records from towers located at the Reykjavík airport during our evaluation period, we evaluated the next three busiest domestic airports in Iceland with dedicated

mobile towers as negative controls. Assuming the validity of our evaluation framework, we expect to see reduced odds of exposure to these locations among cases early in the epidemic compared to their matched controls. We expect a reduced OR as well as a null OR due to the number of locations being evaluated. Under the null hypothesis with 483 towers, exposure to any individual location should have a null or slightly protective effect compared to exposure to the rest of the tower locations in Iceland. In other words, unless a location is a hotspot for transmission at a specific period of time, odds of exposure to that singular location among individuals who then become sick should be the same or less compared to their matched controls, due to the relatively low disease transmission in that location compared to the rest of Iceland.

Social Network Analysis

From the results of the two-week longitudinal OR estimate of individuals who were at the airport in the four days before ILI diagnosis, we identified cases who were at higher odds of exposure earlier in the epidemic period. We generated 1st degree social networks, defined as call or text contacts in the two months before and after the time of onset, for these cases of interest and evaluated the 10 day rate of infection of these 1st degree connections after infection of the case compared to 1st degree connections of their matched controls. We evaluated this rate ratio against the same measure during the peak of the epidemic. We repeated the same evaluation for individuals who were exposed to the airport in the 7 days before diagnosis and had a 24 hour absence in their call records.

Quantitative Analysis of Systematic Error

We conducted a sensitivity and bias analysis to account for systematic error assuming the validity of the following bias model. Exposure and outcome misclassification were both

evaluated as potential sources of systematic error. Exposure was defined at E1 and E2 levels during our evaluation period. The mobile call detail records did not capture individuals who were at the airport during the evaluation period and did not use their mobile device. It is possible that individuals who were diagnosed with an ILI may have visited the airport more frequently than individuals who were not. However, we believe it is unlikely that these two groups would have drastically different network use behaviors at the airport leading to differential misclassification of the exposure. For individuals who were exposed to the airport in the 4 days before ILI diagnosis, we estimated the sensitivity and specificity of exposure misclassification by comparing measures of air travel of Icelanders within the study population with official statistics of outgoing air travel by Icelanders produced by Statistics Iceland (33). We assumed homogeneity of rates of air travel between the study population and the general Icelandic population. For individuals who were exposed to the airport in the 7 days before ILI diagnosis and had a 24 hour absence in their call record we performed a sensitivity analysis of misclassification to evaluate variations in both sensitivity and specificity from E1 levels as E2 were a subset of E1. Therefore, we expect non-differential misclassification for both levels of exposure with near-perfect specificity and variable sensitivity. We conducted a multidimensional bias analysis for non-differential misclassification of exposure.

We defined the outcome as diagnosis of an ILI. Iceland provides universal healthcare to all citizens, therefore records from levels of care ranging from a hospitals to a pharmacies are centralized. Any Icelander visiting a health care facility was registered with a reason for visit that captured symptoms of an ILI, such as coughing or sneezing, with an ICD-10

or ICPC-2 code. Further evaluation by a physician may have added more ICD-10 and ICPC-2 codes to the record. Not all cases with symptoms of an ILI were evaluated with laboratory testing for H1N1. While a large proportion of the 9,887 cases during the peak of the epidemic are likely due to H1N1, the CHS-CDC estimated that at least a third of Icelanders categorized as having symptoms of ILI would have tested positive for H1N1 (29). Therefore while over 90% of all recorded symptoms of ILI occurred during the epidemic period between August and December of 2009, symptoms of ILI are not necessarily indicative of H1N1.

Individuals presenting with symptoms of an ILI to a physician are not likely to be diagnosed as healthy by the classification system. Likewise, individuals without symptoms of an ILI would not have a reason to visit the doctor and therefore would not likely be categorized as having symptoms of an ILI. While a physician may ask for patient history and diagnose based on recent travel history, any patient showing up with symptoms of an ILI would have been categorized as such. The rates of misclassification of ILI diagnosis are therefore not expected to be differential by exposure status. Sensitivity and specificity of outcome misclassification were estimated using parameters provided by the CHS-CDC. We assumed homogeneity of misclassification rates between the individuals diagnosed with an ILI in the study population and the general Icelandic Population. Accounting for case-control sampling we conducted a multidimensional bias analysis for non-differential misclassification of outcome (36). Since we used a matched design in S2 and S3 selection, we conducted an appropriate bias analysis for matched case control studies (37). Furthermore, we accounted for the sampling fraction of the underlying cohort in control selection in the evaluation of outcome misclassification (36).

Results and Analysis

Data

We extracted 342,369 individual records from the 1.5 billion call, text and data records read into Bandicoot. The analysis excluded 14,265 individuals due to sparse records or duplicated information. Restriction for individuals with home towers excluded 40,580 records. Further restriction for individuals with fewer than 100,000 records, fewer than 1,000 calls per week and fewer than 150 places visited per week resulted in 26 dropped records. This final sample dataset contained 302,021 individuals records and contained 4,122 (88.15%) of recorded ILI cases with call detail records [Table 1].

Epidemic Curves

The general epidemic curve of the H1N1 pandemic showed an initial spike of cases in August of 2009 with a peak in October and a return to baseline levels in December [Figure 1]. The epidemic curves (weekly incidence) comparing case exposure to the airport within 4 days of being diagnosed with an ILI to unexposed cases were generated. Of the 4,347 cases, we classified 117 as exposed. We found nearly four times the expected proportion of cases per week two to six weeks before the initiation of the general epidemic curve in week 39 [Figure 2]. While the count per week is generally small (9 in week 33 and 6 in week 37), these numbers deviate greatly from expectation. To guard against the possibility of random variability in small samples, we generated similar curves using other locations in Iceland as the exposure [Figure 3]. Generally, we found that regardless of the size of the exposed group, the proportions of cases per week followed the expectation [Figure 4]. Notable exceptions include locations in downtown

Reykjavík (Austurvöllur) with significant population movement and rural locations, which had almost all incidences during the peak of the epidemic curve. The Kolmogorov-Smirnov (KS) test conducted on the epidemic curves of the exposure of interest resulted in the 3rd largest distance measure ($D = 0.73$, $p<0.0001$) of all tower locations in Iceland indicating a large difference between the continuous distribution functions of numbers of cases per week who were exposed to the airport versus cases who were not exposed.

Selection

S1 produced a longitudinal OR while S2 and S3 produced a longitudinal two-week mOR. All three selection methods showed similar results in terms of the initial peak in the odds ratio [Figure 5]. We chose the S3 selection method for all subsequent analyses as it performed a matched analysis based on home tower location and call data behaviors accounting for strong potential confounders.

Evaluation of Exposure

Of all cases, 462 were exposed to the airport in the 4 days before ILI diagnosis and 72 were exposed to the airport in the 7 days before ILI diagnosis and had a 24 hour absence in their call records [Table 2]. Evaluation of the second group was limited due to the small number of cases generated. The longitudinal two-week mOR produced for individuals who were exposed to the airport in the 4 days before ILI diagnosis showed an increase over the null in the two-week mOR in the early stages of the epidemic from August 17th until August 31st with a mOR of 2 (95% CI: 1.36, 2.95) [Figure 6; Table 3]. This peak coincides with the initial deviation from expected number of cases in August. Over the entire time period [mOR: 0.82 (0.76, 0.88)] and during the comparison period [mOR: 0.87 (0.77, 0.98)], exposure to the international airport showed a protective effect.

This effect is reproduced at other locations and comparable time periods. The longitudinal two-week mOR produced for individuals who were exposed to the airport in the 7 days before ILI diagnosis and had a 24 hour absence in their call records showed little deviation of the mOR from 1 during the initial period of the epidemic [Figure 7]. We used the first level of exposure for all subsequent analyses with the exception of the evaluation of the social networks of cases as it provided the most robust dataset for further evaluation. We retained the second level of exposure for the social network analysis as it provided an insight into the transmission dynamics of ILI within an important subset of the first case group.

Sensitivity Analysis of continuous OR window

We evaluated the sensitivity of the E1/S3 longitudinal mOR to changes in the evaluation window. Increases in the window of evaluation up to 28 days resulted in narrower confidence intervals, however the spike in the mOR during the initial period of the evaluation period remained [Figure 8]. We used the two-week mOR window for all subsequent analyses. Upon varying this window for other analyses we noted a near-null (mOR = 1) or generally protective association for other individual regions in Iceland.

Positive Controls

We evaluated two positive controls, downtown Reykjavík and a major hospital, defining exposure as presence at a location of interest in the 4 days before ILI diagnosis and matching controls based on home tower location and call data behaviors. Both positive controls detected the expected increase in mOR with significant peaks in mOR early in the evaluation period [Figure 9]. Chronologically, the odds of exposure spiked first at the airport, followed by the major hospital and finally in downtown Reykjavík.

Negative Controls

Reykjavík airport serviced the largest number of domestic flights in Iceland in 2009. However, cell towers in the area did not receive a volume of data comparable to other towers during the evaluation period. Due to the lack of data at this location we evaluated the next three most popular domestic airports as negative controls defining exposure as presence at a location of interest in the 4 days before ILI diagnosis and matching controls based on home tower location and call data behaviors. As expected, all negative controls showed a null or protective mOR in the early stages of the epidemic [Figure 10].

Social Network Analysis

We conducted an analysis on 10-day rate of infection within 1st degree connections among cases and controls, during the initial period of the epidemic and during the epidemic peak, in both exposure levels. During the initial period of the epidemic, 1st degree connections of cases had an infection rate that was 13 times greater than 1st degree connections of controls (95% CI: 6.1, 29) compared to the epidemic peak where 1st degree connections of cases had an infection rate 4 (95% CI: 3.6, 5) times greater than 1st degree connections of controls. Cases who were exposed to the airport in the 4 days before ILI diagnosis had a higher point estimate of 10-day incidence density ratio compared to cases who were not exposed. This difference was noted in both the initial period of the epidemic and the epidemic peak, nevertheless, the confidence intervals of these measures had large overlaps indicating little difference in rate of infection among 1st degree connections of cases who were exposed to their airport and cases who were not in this time period [Table 4]. Cases with E2 exposure had controls with no incident cases of ILI within their 1st degree connections during the initial period of the epidemic.

Quantitative Analysis of Systematic Error

We conducted a multidimensional bias analysis for non-differential misclassification of exposure and outcome. Exposure specificity ranged from 0.95 to 1 while exposure sensitivity ranged from 0.5 to 0.9. We expected to correctly classify nearly all individuals who did not visit the international airport and travel. Analysis of the numbers of individuals in our sample frame who visited Keflavik International Airport matched seasonal changes in Icelandic travel trends recorded for 2010. However, we expect a threat to validity as not all individuals who visited the airport would have used their mobile phones. We ranged outcome specificity from 0.95 to 1 and outcome sensitivity from 0.9 to 0.99. Due to the nationalized, easy-to-access healthcare system and generalized definition of ILI from various ICD and ICPC codes, we expect minimal threat to validity from outcome misclassification alone. The longitudinal 2 week mOR for both analyses showed minor deviations to the confidence intervals of the mOR during the initial period of the epidemic [Figure 11; Figure 12] which worsened in relation to decreasing specificity. Using the point estimate calculated by evaluating the 2 week window from August 15th till September 4th, we saw a similarly minor deviation in mOR as we varied the bias parameters with increased variability due to changes in specificity [Table 5; Table 6].

Conclusions

Primary Findings

Our study evaluates the role that international travelers played in the introduction and propagation of pandemic H1N1 in Iceland. We find that there is an association between exposure to Keflavik International Airport and incident ILI diagnoses during the initial stages of the epidemic [August 24th 2-week mOR: 2 (95% CI: 1.4, 2.9)].

Secondary Findings

As a secondary research objective, we evaluated the rates of infection among 1st degree connections of cases compared to 1st degree connections of controls. We expected the comparative incidence density ratio (IDR) to be high during the initial stages of the epidemic and the data aggregated from the call detail records confirmed our belief. During the two week period from August 17th through August 31st we calculated a two-week IDR of infection among 1st degree connections to be 14.2 (95% CI: 5.7, 35). The IDR decreased steadily to a threshold IDR of approximately 5 during the epidemic peak [Figure 11]. In other words, our data shows that there is a definitive high rate of transmission earlier in the epidemic. However, even during the peak of cases in October when there was a generalized epidemic in the population, 1st degree connections of individuals diagnosed with an ILI get sick at a rate 5 times higher rate than the 1st degree connections of their matched controls.

Negative Controls

All domestic airports evaluated in the study showed a null or protective association to outcome. In contrast, the evaluation with Keflavik International Airport for the sample time period was greater than the null.

Quantitative Analysis of Systematic Error

Based on the conventional result, we inferred that the odds of exposure to the international airport is double among cases when compared to the at-risk cohort controls during the initial period of the epidemic. Through the bias analysis, we noted that the bias parameters for both exposure and outcome had high specificities, but exposure had low sensitivity. Correction of the data using our bias models for exposure and outcome resulted in a bias-adjusted odds ratio of 2.7 assuming a valid bias model. The original inference was strengthened and moved away from the null by adjusting for misclassification as the bias. A multidimensional sensitivity analysis of the bias parameters showed that the magnitude of the bias was related strongly to the specificity of both misclassifications. Due to relatively high specificity in exposure and outcome misclassification, our model showed few threats to the validity of our mOR. The bias-adjusted odds ratio, correcting for misclassification of exposure and outcome, showed that the odds of exposure to Keflavik International Airport among cases is 2.7 times higher than the odds of exposure to Keflavik International Airport among randomly selected at-risk cohort control, assuming the validity of our bias model.

Locations and their Roles in an Epidemic

We note a general protective association for all locations in Iceland when the matched odds ratio is evaluated over the entire data period. This is expected since odds of exposure to a single location compared to odds of exposure to all other locations in Iceland will nearly always be smaller. However, as the window of evaluation shrinks, we see temporally local amplification of odds of exposure to specific regions of interest. The utility of this type of evaluation is especially important in the progression of an epidemic. For example, in Gabon, pandemic H1N1 propagated in urban centers during the early stages of the epidemic before expanding through transport networks to rural areas (7). Such propagation is demonstrated in our data through the evaluation of our positive controls. As there is only one major point of entry into Iceland, we might expect an epidemic to be introduced there first, followed by transmission in areas where sick patients congregate, such as a major hospitals and finally an urban center just before expansion of the epidemic to the general population. As expected, we see clear spikes in 2 week odds of exposure, moving temporally from August 10th through September 14th, moving from the Keflavik International Airport, to Landspítali University Hospital in Reykjavík and finally the Austurvöllur area in downtown Reykjavík [Figure 14].

Limitations

First, call detail records for this study were captured in 2009 and 2010 when the majority of the call detail records consisted of calls and texts rather than mobile data transactions. Therefore, individual records were dependent on users interacting with their mobile device. This led to issues of sparse data and low predictive values in our study. In other words, if a user didn't make a call, send a text or use mobile data for a period of time we

were unable to discern their location or behaviors. In contrast, modern phones generate large records of mobile data transactions and ping frequently for updates regardless of user interaction providing more granular data.

Second, we conducted an evaluation of negative controls using various domestic airports around Iceland. However we lacked data from the towers located at the largest domestic airport in Reykjavík as it was in inconsistent use at the time of the H1N1 epidemic. While we make a strong case for validity using three other domestic airports as negative controls, a similar evaluation using the Reykjavík domestic airport as a negative control would strengthen the validity of our results.

Future Analyses

The results of this study highlight the relevance of call detail records to epidemiologic practice. Since the collection of this data in 2009 the global number of mobile phone subscriptions has risen from 68 to 96.8 per 100 inhabitants (38), the world population has flocked to urban centers (39), and nearly 2 billion new smart phone users have been registered with 6 billion projected by 2020. (40) Modern call detail records include considerably more data transfer information allowing for more robust analyses of location and fewer threats to validity from misclassification. Through greater collaboration with both mobile network operators and health officials, the techniques described in this study can be used for hypothesis-driven evaluations of locations and behaviors during an epidemic, and their associations with health outcomes.

References

1. Ginsberg M, Hopkins J, Maroufi A, et al. Swine Influenza A (H1N1) Infection in Two Children --- Southern California , March--April. *Morb. Mortal. Wkly. Rep.* 2009;58(15):400–402.
2. Dawood F, Jain S, Finelli L, et al. Emergence of a Novel Swine-Origin Influenza A (H1N1) Virus in Humans. *N. Engl. J. Med.* 2009;360(25).
3. Colizza V, Barrat A, Barthelemy M, et al. Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. *PLoS Med.* 2007;4(1):0095–0110.
4. Ali ST, Kadi AS, Ferguson NM. Transmission dynamics of the 2009 influenza A (H1N1) pandemic in India : The impact of holiday-related school closure. *Epidemics*. 2013;5(4):157–163.
5. Archer BN, Timothy G a., Cohen C, et al. Introduction of 2009 pandemic influenza a virus subtype H1N1 into South Africa: Clinical presentation, epidemiology, and transmissibility of the first 100 cases. *J. Infect. Dis.* 2012;206(SUPPL.1).
6. Charoenvisal N, Keawcharoen J, Sreta D, et al. Genetic characterization of Thai swine influenza viruses after the introduction of pandemic H1N1 2009. *Virus Genes*. 2013;47(1):75–85.
7. Lekana-Douki SE, Mouinga-Ondémé A, Nkoghe D, et al. Early introduction and delayed dissemination of pandemic influenza, Gabon. *Emerg. Infect. Dis.* 2013;19(4):644–647.
8. Nakouné E, Tricou V, Manirakiza A, et al. First introduction of pandemic influenza A/H1N1 and detection of respiratory viruses in pediatric patients in Central African Republic. *Virol. J.* [electronic article]. 2013;10:49. (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3598402&tool=pmcentrez&rendertype=abstract)
9. Kucharski AJ, Kwok KO, Wei VWI, et al. The Contribution of Social Behaviour to the Transmission of Influenza A in a Human Population. *PLoS Pathog.* 2014;10(6).
10. Yang W, Lipsitch M, Shaman J. Inference of seasonal and pandemic influenza transmission dynamics. *Proc. Natl. Acad. Sci.* [electronic article]. 2015;112(9):201415012. (http://www.pnas.org/lookup/doi/10.1073/pnas.1415012112)
11. Mangili A, Gendreau MA. Transmission of infections during commercial air travel. *Lancet*. 2005;365(9478):2176–2177.
12. Warren A, Bell M, Budd L. Airports, localities and disease: Representations of global travel during the H1N1 pandemic. *Heal. Place* [electronic article].

2010;16(4):727–735. (http://dx.doi.org/10.1016/j.healthplace.2010.03.004)

13. States U, June O. Introduction and transmission of 2009 pandemic influenza A (H1N1) Virus--Kenya, June-July 2009. *MMWR. Morb. Mortal. Wkly. Rep.* 2009;58(41):1143–1146.

14. Webster CH. Airline operating realities and the global spread of infectious diseases. *Asia. Pac. J. Public Health*. 2010;22(3 Suppl):137S–143S.

15. Seale H, Marich A, Mcanulty J. Airport arrivals screening during pandemic (H1N1) 2009 infl uenza in New South Wales, Australia. 2014;1991(March):2009–2011.

16. Hale MJ, Hoskins RS, Baker MG. Screening for Infl uenza A(H1N1)pdm09, Auckland International Airport, New Zealand. 2012;18(5):866–868.

17. Malone JD, Brigantic R, Muller G a., et al. U.S. airport entry screening in response to pandemic influenza: Modeling and analysis. *Travel Med. Infect. Dis.* [electronic article]. 2009;7(4):181–191. (http://dx.doi.org/10.1016/j.tmaid.2009.02.006)

18. Khan K, Eckhardt R, Brownstein JS, et al. Entry and exit screening of airline travellers during the A (H1N1) 2009 pandemic : a retrospective evaluation. *Bull. World Health Organ.* 2013;91(march):368–376.

19. Nishiura H, Kamiya K. Fever screening during the influenza (H1N1-2009) pandemic at Narita International Airport, Japan. *BMC Infect. Dis.* [electronic article]. 2011;11(1):111. (http://www.biomedcentral.com/1471-2334/11/111)

20. Schlaich C, Sevenich C, Gau Hamburg B. Public Health Measures at the Airport of Hamburg during the Early Phase of Pandemic Infl uenza (H1N1) 2009. 2012;145–153.

21. Fujita M, Sato H, Kaku K, et al. Airport quarantine inspection, follow-up observation, and the prevention of pandemic influenza. *Aviat. Sp. Environ. Med.* 2011;82(8):782–789.

22. Baker MG, Thornley CN, Mills C, et al. Transmission of pandemic A/H1N1 2009 influenza on passenger aircraft: retrospective cohort study. *BMJ*. 2010;340:c2424.

23. Swaan CM, Appels R, Kretzschmar ME, et al. Timeliness of contact tracing among flight passengers for influenza A/H1N1 2009. *BMC Infect. Dis.* 2011;11(1):355.

24. Bobashev G, Morris RJ, Goedecke M. Sampling for global epidemic models and the topology of an International airport network. *PLoS One*. 2008;3(9).

25. Epstein JM, Goedecke DM, Yu F, et al. Controlling pandemic flu: The value of international air travel restrictions. *PLoS One*. 2007;2(5).

26. Hollingsworth TD, Ferguson NM, Anderson RM. Frequent travelers and rate of spread of epidemics. *Emerg. Infect. Dis.* 2007;13(9):1288–1294.

27. Hsu CI, Shih HH. Transmission and control of an emerging influenza pandemic in a small-world airline network. *Accid. Anal. Prev.* 2010;42(1):93–100.

28. Marcelino J, Kaiser M. Critical Paths In A Metapopulation Model Of H1N1-Efficiently Delaying Influenza Spreading Through Fight Cancellation. *PLoS Curr. Influ.* 2012;1–16.

29. Sigmundsdottir G, Gudnason T, Ólafsson Ö, et al. Surveillance of influenza in Iceland during the 2009 pandemic. *Eurosurveillance*. 2010;15(April 2009):3.

30. Cauchemez S, Bhattarai A, Marchbanks TL, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proc. Natl. Acad. Sci. U. S. A.* 2011;108(7):2825–2830.

31. Kwok KO, Cowling BJ, Wei VWI, et al. Social contacts and the locations in which they occur as risk factors for influenza infection Social contacts and the locations in which they occur as risk factors for influenza infection. 2014;

32. Kerckhove K Van, Hens N, Edmunds WJ, et al. The impact of illness on social networks: Implications for transmission and control of influenza. *Am. J. Epidemiol.* 2013;178(11):1655–1662.

33. Statistics Iceland. (http://www.statice.is/)

34. Christakis NA, Fowler JH. Social network sensors for early detection of contagious outbreaks. *PLoS One*. 2010;5(9):1–8.

35. Montjoye Y De, Quoidbach J, Robic F, et al. Predicting Peronality Using Novel Mobile Phone-Based Metrics. *Int. Conf. Soc. Comput. Behav. Model. Predict.* 2013;48–55.

36. Jurek AM, Maldonado G, Greenland S. Adjusting for outcome misclassification: The importance of accounting for case-control sampling and other forms of outcome-related selection. *Ann. Epidemiol.* [electronic article]. 2013;23(3):129–135. (http://dx.doi.org/10.1016/j.annepidem.2012.12.007)

37. Greenland S. The effects of misclassification and matching in case-control studies have been extensively studied in isolation from one another . It has long been recognized , however , that the deleterious effects of misclassification can be aggravated by more recentl. *Am. J. Epidemiol.* 1982;116(2):402–406.

38. Global ICT developments, 2001-2015. 2015.(http://www.itu.int/ict/statistics)

39. World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352). United Nations; 2014 32 p.(http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf)

40. Ericsson. Ericsson Mobility Report: 70 percent of world's population using smartphones by 2020. 2015;1–4. (http://www.ericsson.com/news/1925907)

Tables

Table 1: Call data meta record demographics generated[1] from of Icelandic customers of Síminn from March 2009 till November 2010 linked to an ILI diagnosis database of all Icelandic citizens							
All Records	342639						
	Min	1st Q	Median	Mean	3rd Q	Max	Missing
Number of records per individual	1	14	560	2836	3088	125112439	
Active days per week	1	1.2	3.833	3.755	6.091	7	
number of calls per week	1	1	5.38	9.35	13.47	70472.17	
Number of places visited per week	1	1.353	4.671	8.483	12.571	228.76	
Percent of activity between 7pm - 7am	0.00%	13.64%	22.95%	29.16%	36.71%	100.00%	
Percent of time at home	0.00%	13.00%	24.00%	30.00%	41.00%	100.00%	40580
ILI(+) individuals who generated a record	4177/4347	96.09%					
	Min	1st Q	Median	Mean	3rd Q	Max	Missing
Number of records per individual	1	671	2564	4136	6005	64387	
Active days per week	1	4	5.909	5.121	6.542	6.961	
number of calls per week	1	6.115	11.703	13.07	17.972	94.536	
Number of places visited per week	1	5.339	11.745	13.931	19.734	74.364	
Percent of activity between 7pm - 7am	0.00%	18.25%	25.00%	27.38%	34.65%	100.00%	
Percent of time at home	0.00%	14.47%	23.44%	26.47%	35.47%	100.00%	65
Sample Population[2]	302021	88.15%					
ILI(+) individuals	4112/4347	94.59%					
	Min	1st Q	Median	Mean	3rd Q	Max	Missing
Number of records per individual	1	62	923	2766	3647	99723	
Active days per week	1	1.857	4.516	4.113	6.231	7	
number of calls per week	1	1.885	7	10.153	14.713	584.794	
Number of places visited per week	1	2	6.144	9.457	13.88	139.376	
Percent of activity between 7pm - 7am	0.00%	17.02%	25.22%	33.07%	39.77%	100.00%	
Percent of time at home	0.00%	12.55%	0.24%	30.43%	40.82%	100.00%	

[1] Records generated using Bandicoot (de Montjoye, Y.-A. et.al)
[2] Restricting for individuals with home tower (-40580), < 100,000 records (-3), < 1000 calls per week (-3), < 150 places visited (-20); Including both ILI(+) and ILI(-)

Table 2: Distribution of exposure among cases and their controls during select periods of the H1N1 epidemic in Iceland in 2009

Exposure to the international airport in the 4 days before diagnosis (E1)

	All data[1]	Period of interest[2]	Initial stages of the epidemic[3]	Two-week period of high risk in initial stages[4]	Comparison two-week period in epidemic peak[5]
Cases	462	421	99	29	170
Controls	545	501	102	17	189

Exposure to the International airport and at least 24 hours without call data record activity in the 4 days before diagnosis (E2)

	All data[1]	Period of interest[2]	Initial stages of the epidemic[3]	Two-week period of high risk in initial stages[4]	Comparison two-week period in epidemic peak[5]
Cases	72	68	11	2	24
Controls	95	92	18	7	31

[1] March 12th, 2009 through November 16th, 2009
[2] August 2009 through December 2009
[3] July 7th through September 15th, 2009
[4] August 17th through August 31st, 2009
[5] October 7th through October 23th, 2009
*Diagnosis date for controls was based on the diagnosis of their relative cases
mOR - matched odds ratio

Table 3: Associations between exposures of interest and subsequent ILI diagnosis with controls matched on home tower location and similar call data behaviors

Primary exposures of interest - *mOR [95% CI]*					
	All data[1]	Period of interest[2]	Initial stages of the epidemic[3]	Two-week period of high risk in initial stages[4]	Comparison two-week period in epidemic peak[5]
Keflavik International Airport (E1)	0.82 [0.76 ,0.88]	0.81 [0.75 ,0.87]	0.97 [0.82 ,1.15]	2.00 [1.36 ,2.95]	0.87 [0.77 ,0.98]
Keflavik International Airport + 24hr absence (E2)	0.76 [0.65 ,0.89]	0.74 [0.63 ,0.87]	0.56 [0.37 ,0.85]	0.17 [0.06 ,0.5]	0.79 [0.6 ,1.04]
Negative controls - *ln(mOR) [95% CI]*					
Akureyri Domestic Airport	0.89 [0.78 ,1.02]	0.85 [0.73 ,0.98]	0.46 [0.32 ,0.66]	0.67 [0.35 ,1.28]	1.00 [0.8 ,1.24]
Ísafjörður Domestic Airport	1.13 [0.94 ,1.36]	1.00 [0.82 ,1.21]	1.22 [0.78 ,1.91]	0.67 [0.27 ,1.67]	0.91 [0.67 ,1.23]
Egilsstaðir Domestic Airport	0.83 [0.75 ,0.92]	0.82 [0.74 ,0.91]	0.88 [0.69 ,1.13]	0.51 [0.39 ,0.67]	0.86 [0.73 ,1.01]

[1] March 12th, 2009 through November 16th, 2009
[2] August 2009 through December 2009
[3] July 7th through September 15th, 2009
[4] August 17th through August 31st, 2009
[5] October 7th through October 23th, 2009
mOR - matched odds ratio

Table 4: Two-week rate ratio of infection among 1st degree connections of cases compared to friends of controls in different periods of the epidemic	
Time periods of interest - *ln(Rate Ratio) [95% CI]*	
Two-week period of high risk in Initial Stages[1]	14.16 [5.66, 35.41]
Comparison two-week period in Epidemic Peak[2]	4.41 [3.66, 5.32]

[1] August 17th through August 31st, 2009
[2] October 7th through October 23th, 2009

Table 5: Multidimensional bias analysis of exposure misclassification[1]		
Sensitivity	Specificity	Bias Adjusted OR[2]
0.9	1	2.17
0.8	1	2.50
0.6	1	3.33
0.5	1	10.00
0.9	0.97	2.23
0.8	0.97	2.56
0.6	0.97	10.24
0.5	0.97	Error
0.9	0.95	2.27
0.8	0.95	2.61
0.6	0.95	10.43
0.5	0.95	Error

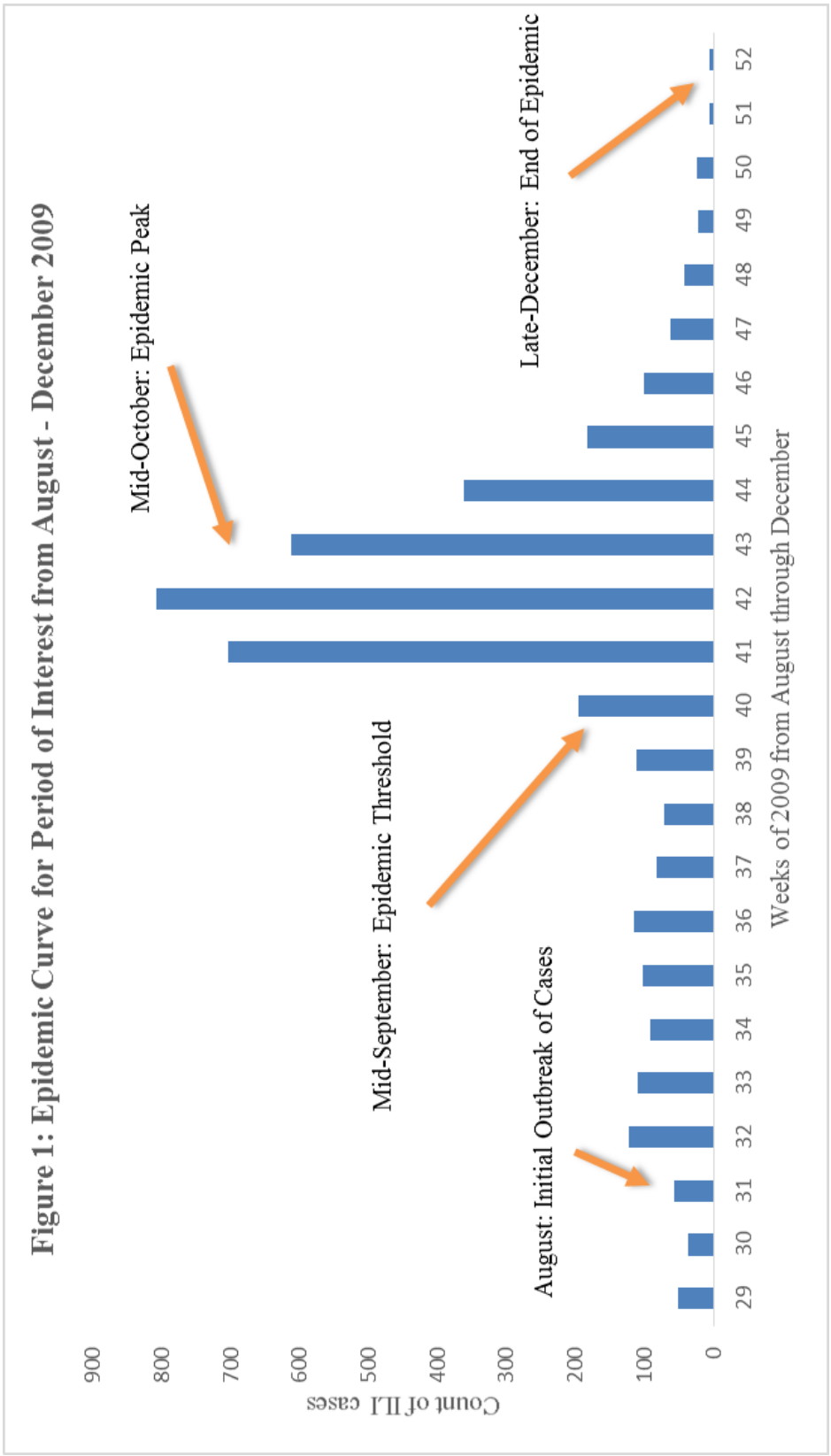
[1]Assuming the validity of our bias model
[2]From August 15th through September 4th

Table 6: Multidimensional bias analysis of outcome misclassification[1]

Sensitivity	Specificity	Bias Adjusted OR[2]
0.99	1	2.02
0.97	1	2.07
0.94	1	2.15
0.9	1	2.29
0.99	0.97	2.05
0.97	0.97	2.10
0.94	0.97	2.18
0.9	0.97	2.32
0.99	0.95	2.08
0.97	0.95	2.12
0.94	0.95	2.21
0.9	0.95	2.35

[1]Assuming the validity of our bias model
[2]From August 15th through September 4th

Figures and Figure Legends



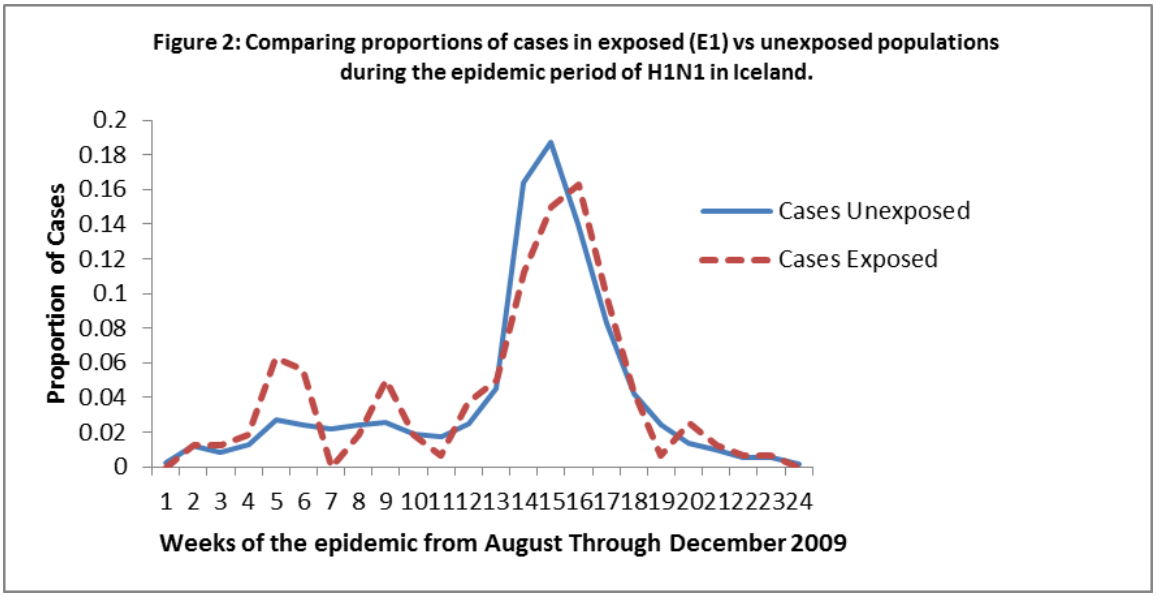
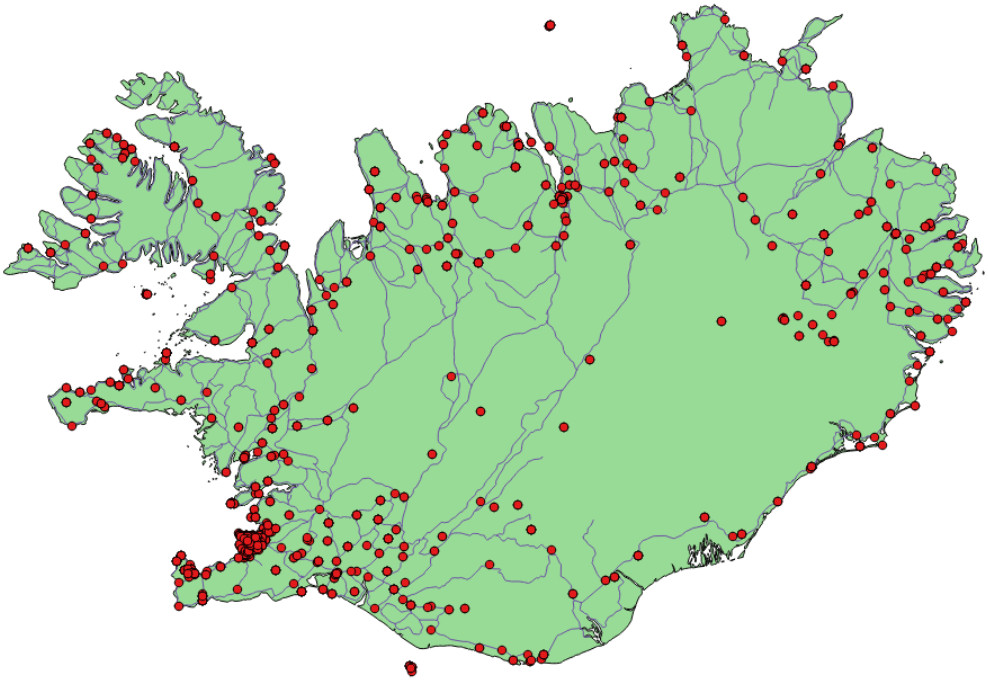


Figure 2: Comparing proportions of cases in exposed (E1) vs unexposed populations during the epidemic period of H1N1 in Iceland. There are 117 exposed cases. While the n at week 33 and week 37 are small, 9 and 7 respectively, they are well more than the cases expected in those week. The KS test is significant at alpha = 0.05. Most importantly this type of curve is only noticed at the airport (E1).

Figure 3: Locations of all towers belonging to study MNO in Iceland



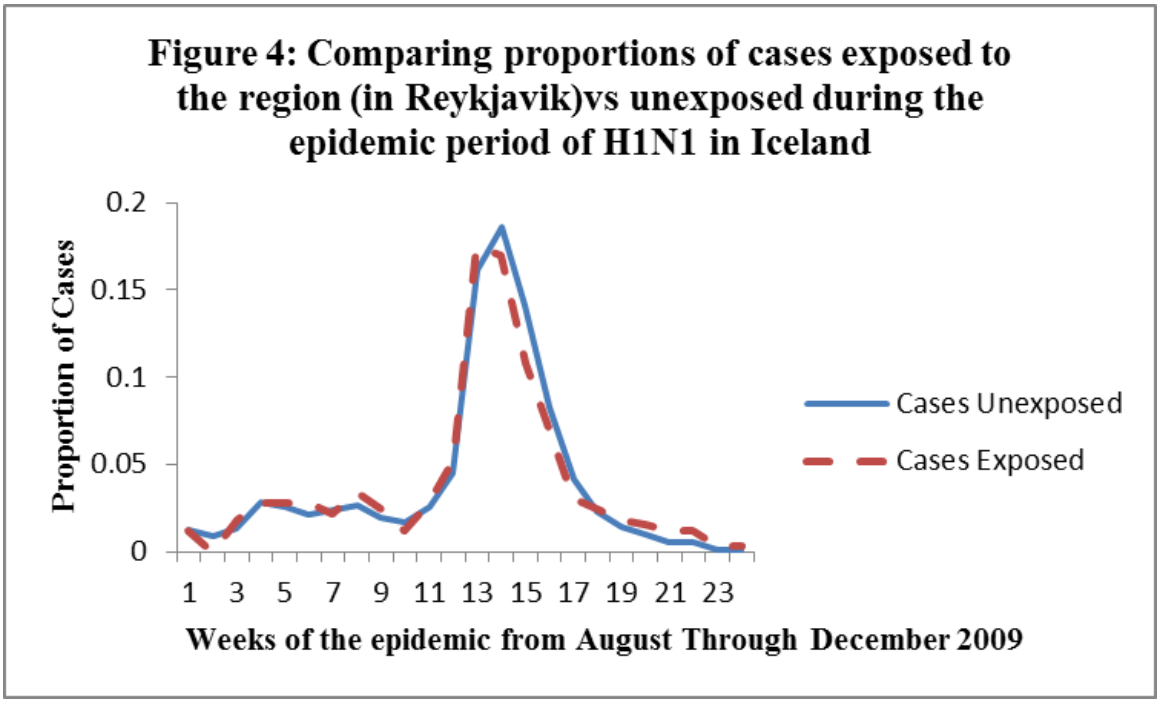


Figure 5: Control Selection Strategies

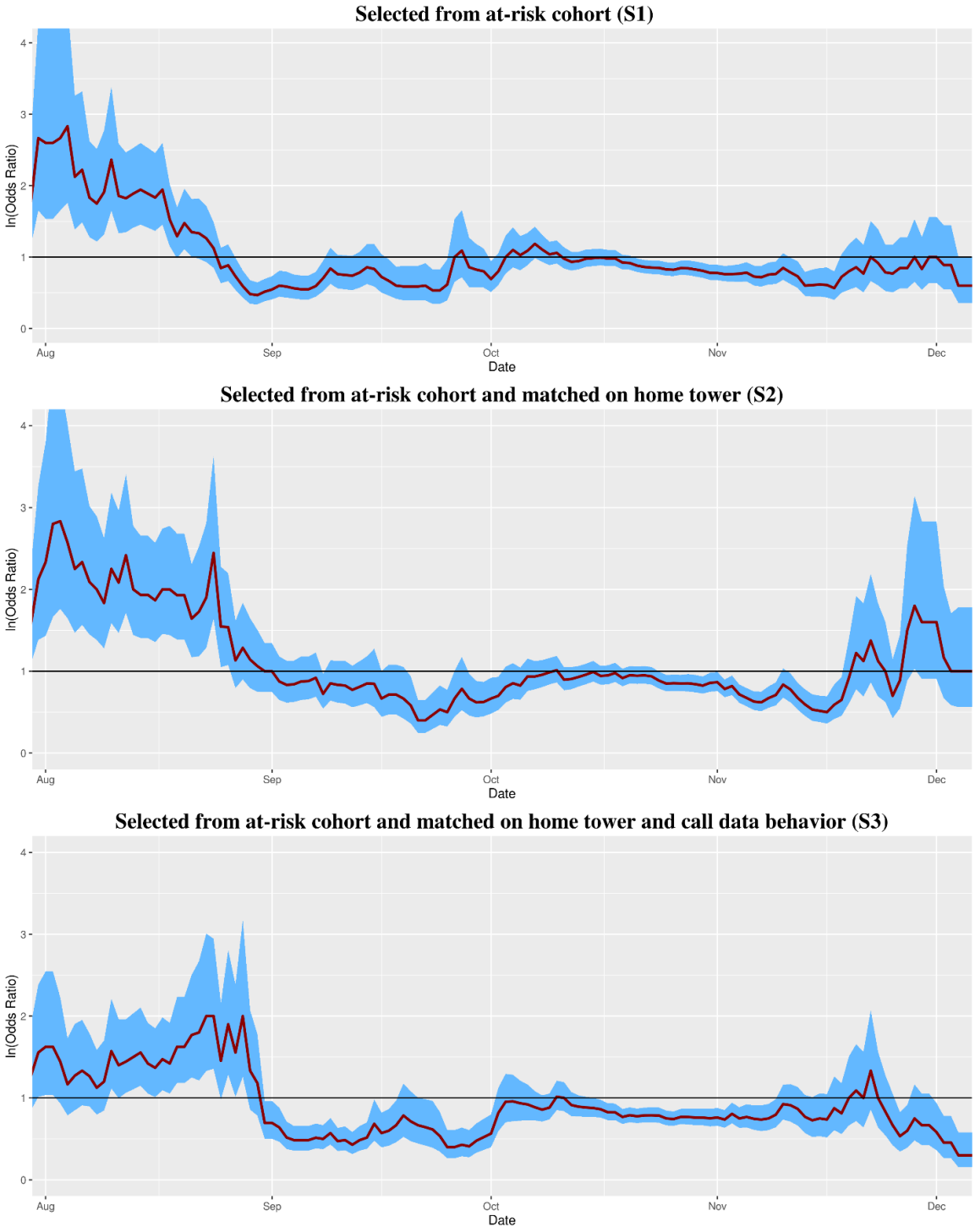


Figure 6: Two-week matched odds ratio (S3) for Keflavik International Airport (E1)

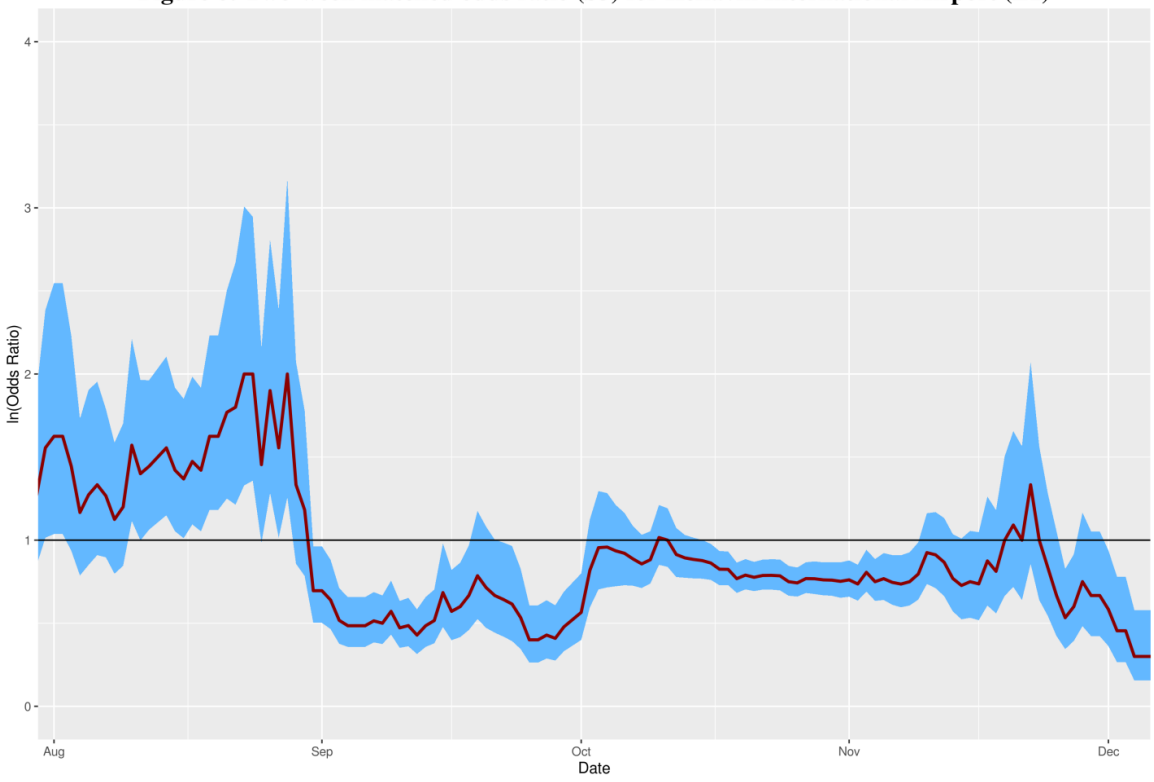


Figure 7: Two-week matched odds ratio (S3) for Keflavik International Airport + 24hr absence (E2)

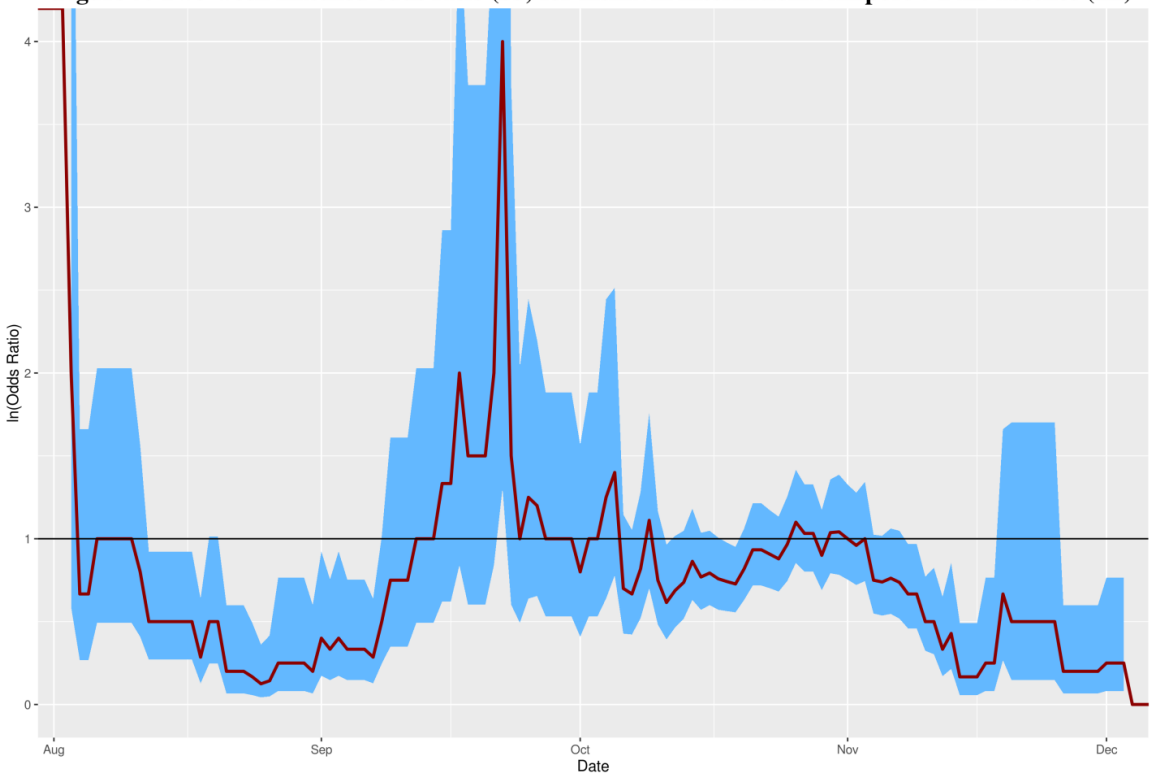


Figure 8: Sensitivity analysis for size of mOR window

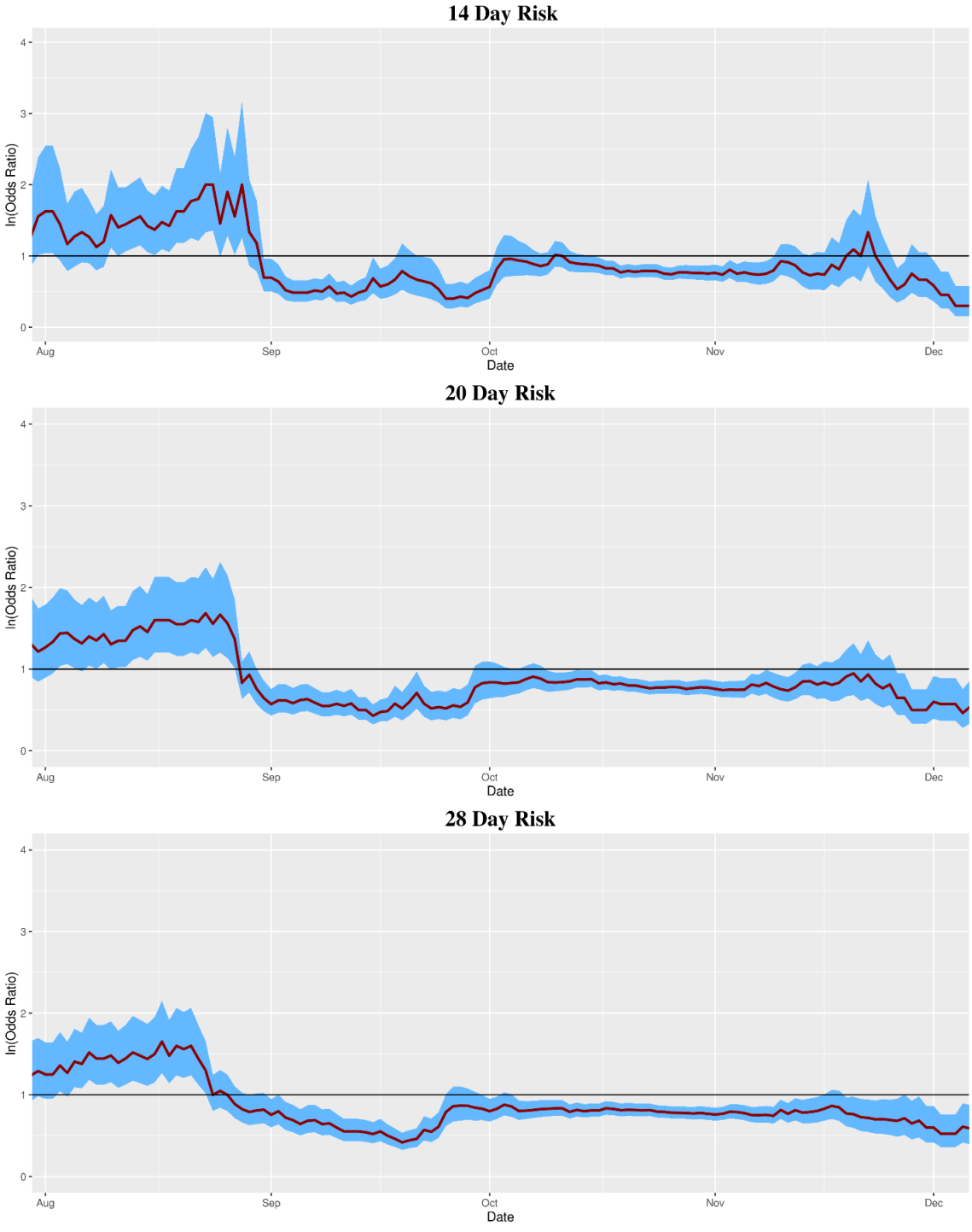


Figure 9: Positive Controls

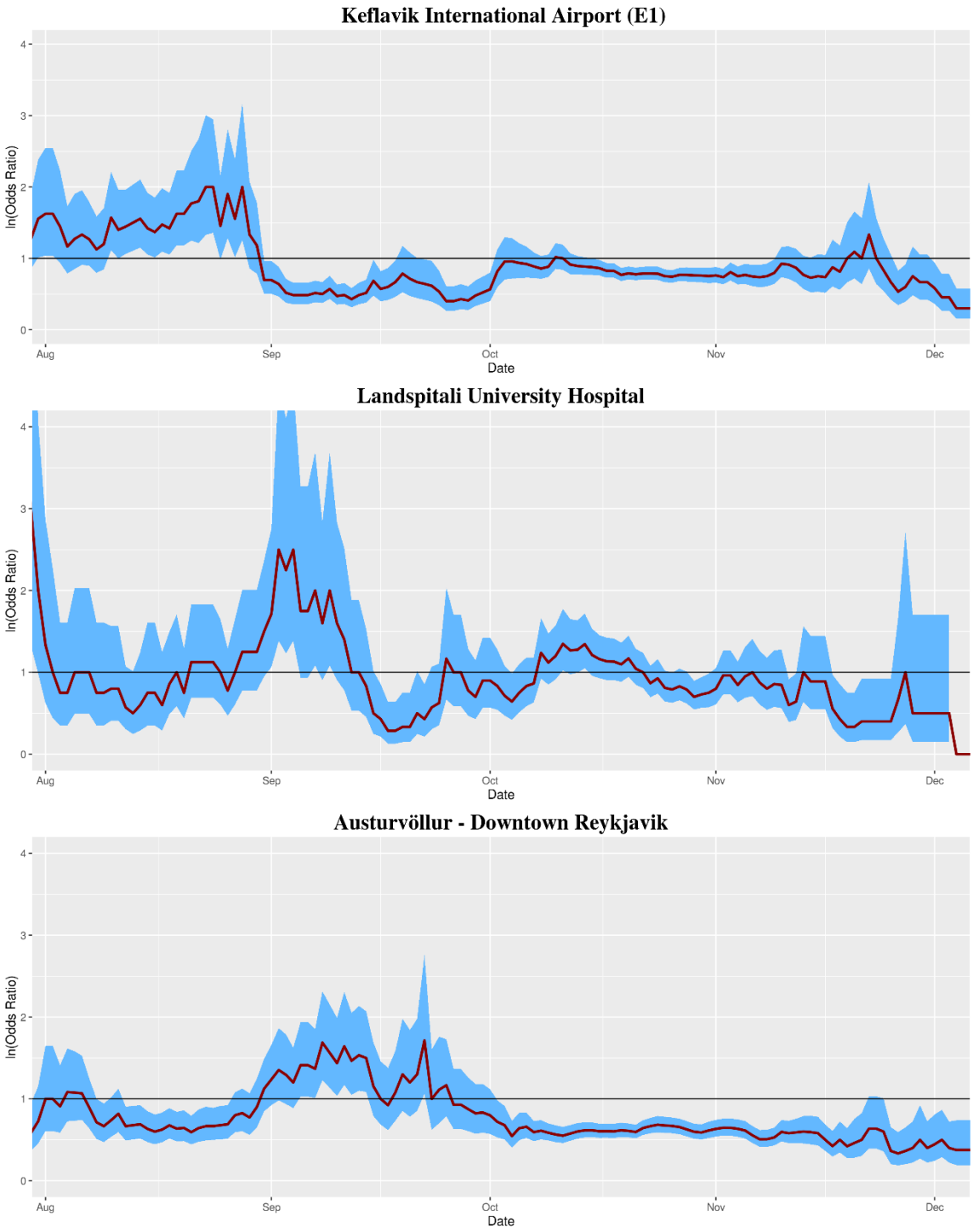


Figure 10: Negative Controls

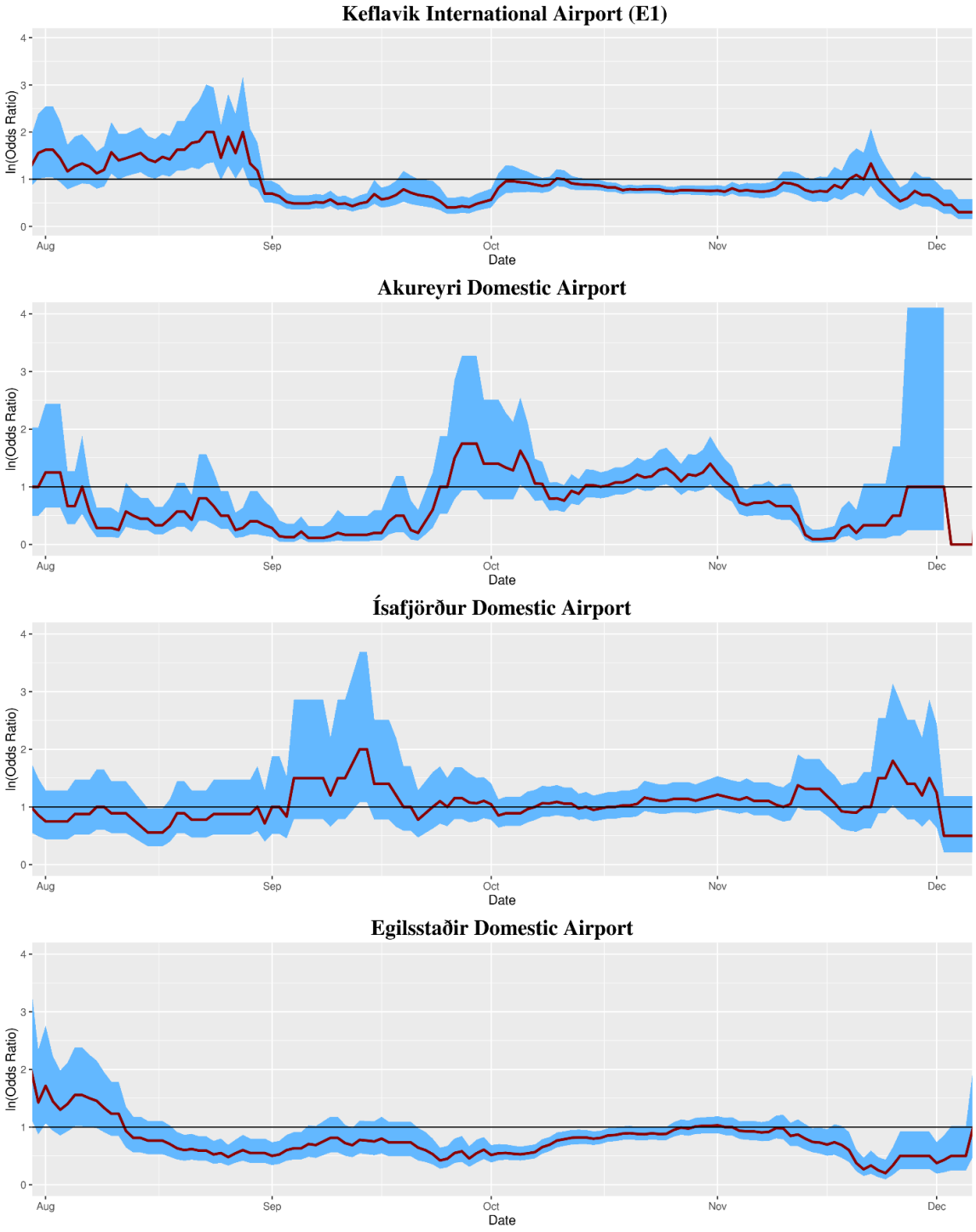


Figure 11: Rate ratio of infection amongst 1st degree connections of cases compared to 1st degree connections of controls

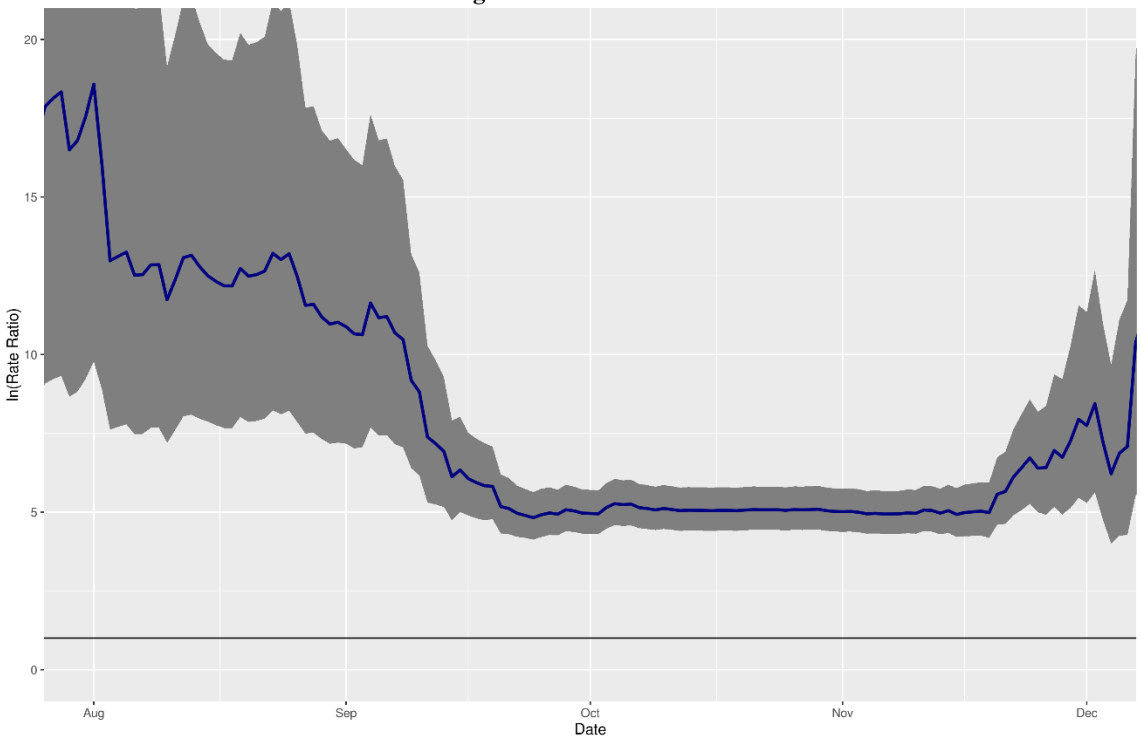


Figure 12: Multidimensional Sensitivity Analysis of Exposure

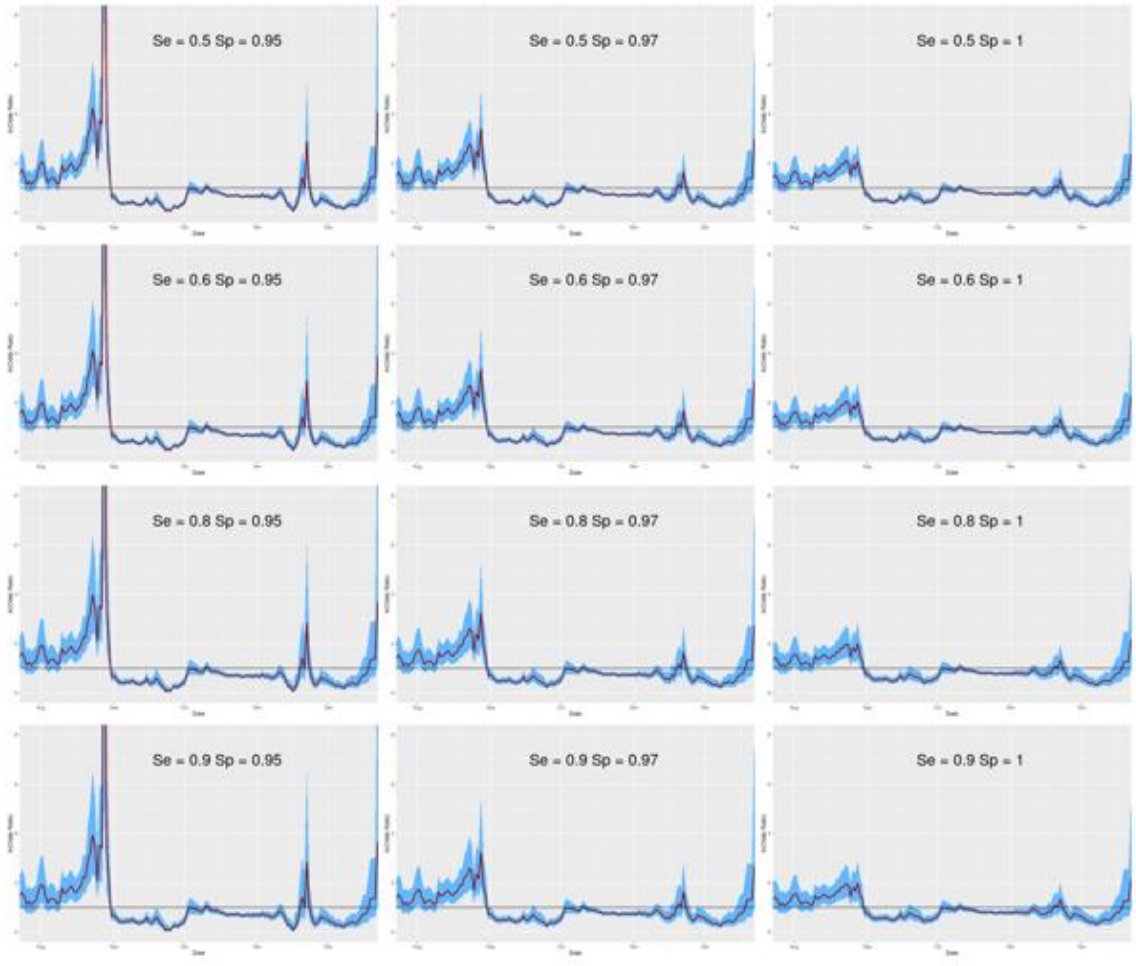


Figure 13: Multidimensional Sensitivity Analysis of Outcome

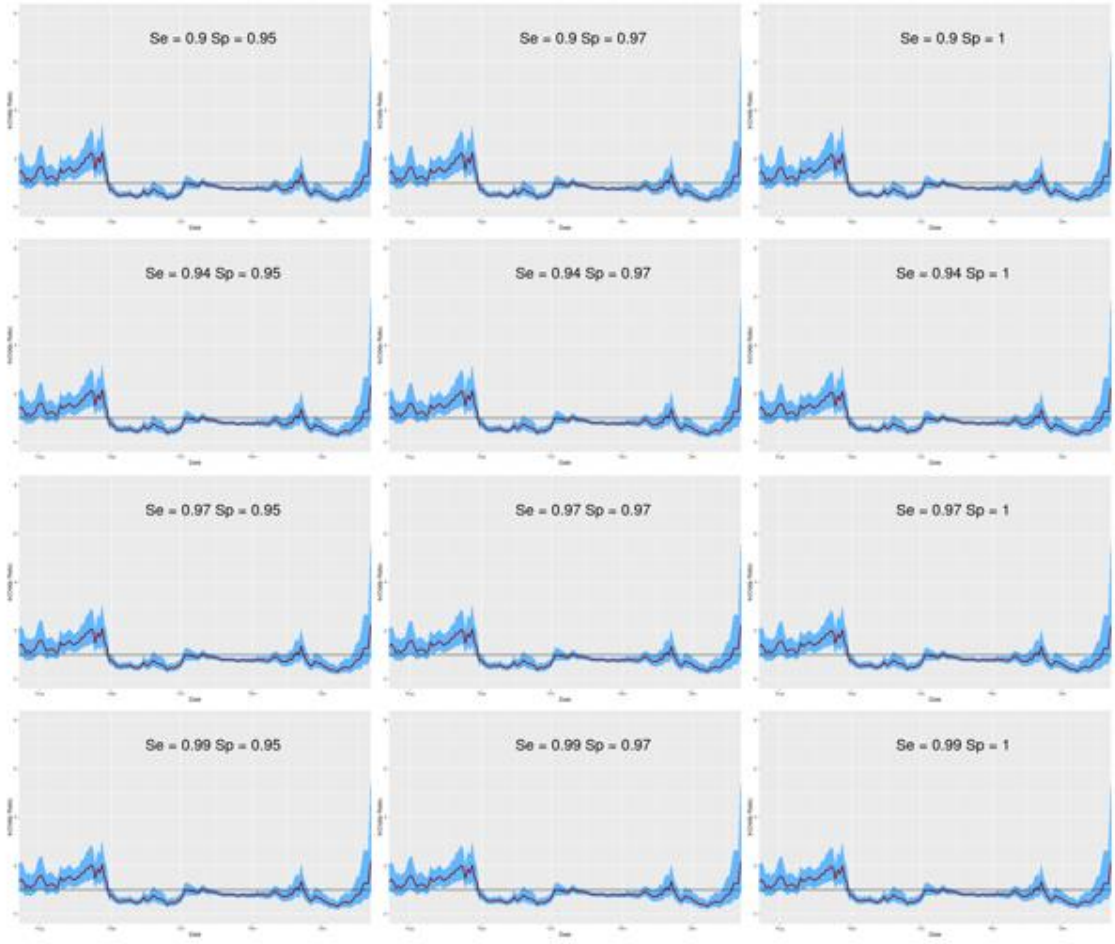


Figure 14: Locations and their roles in epidemics