BMJ Open

# An observational study to assess the effects of social networks on the seasonal influenza vaccine uptake by early career doctors.

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2018-026997 |
| Article Type: | Research |
| Date Submitted by the Author: | 01-Oct-2018 |
| Complete List of Authors: | Edge, Rhiannon; Lancaster University Faculty of Health and Medicine, Lancaster Medical School<br>Keegan, Thomas; Lancaster University, Lancaster Medical School<br>Isba, Rachel; Lancaster University Faculty of Health and Medicine, Lancaster Medical School<br>Diggle, Peter; Lancaster University, Lancaster Medical School; University of Liverpool, Epidemiology and Population Health |
| Keywords: | Social network analysis, influenza vaccination, auto-logistic regression, occupational health |
| | |

# An observational study to assess the effects of social networks on the seasonal influenza vaccine uptake by early career doctors.

R Edge*, T J Keegan, R Isba, P J Diggle

Lancaster Medical School, Furness Building, Lancaster University, Bailrigg, Lancaster, LA1 4YW. Senior Teaching Associate, Senior Lecturer, Head of Department, Professor of Statistics.

* Correspondence to: r.edge@lancaster.ac.uk.

## Abstract

### Background

The Chief Medical Officer for England recommends that healthcare workers have a seasonal influenza vaccination, in an attempt to protect both patients and NHS staff. Despite this, only 55% of healthcare workers are vaccinated. Social networks have been found to affect the behaviour of the individuals within them, thus they may be useful in understanding vaccination habits.

### Methods

Data were collected from a population of early career doctors who self-reported their seasonal influenza vaccination status, along with basic demographic characteristics and information about their social relationships. Social network analysis and statistical modelling were used to assess the vaccination distribution within this network of doctors, and whether the likelihood of an individual receiving the vaccination was associated with their peers' vaccination behaviour.

### Results

Of the 200 eligible early career doctors, 138 (70%) provided complete data, of whom 100 (72%) reported that they had received a seasonal influenza vaccination. Statistical modelling demonstrated that having vaccinated neighbours reduced an individual's likelihood of being vaccinated. Adjusting for year-group and geographical area did not account for this effect.

### Conclusion

This population exhibited higher than expected vaccination coverage levels– providing protection both in the workplace and for vulnerable patients. The modelling approach allowed covariate effects to be incorporated into social network analysis, which gave us a better understanding of the network structure. These techniques have a range of applications in understanding the role of social networks on health behaviours.

**Key words**

Social network analysis, influenza vaccination, auto-logistic regression, occupational health.

**Article Summary**

Because of their occupation, HCWs are at a higher risk of influenza, and the consequences from contracting/transmitting the virus are increased. Social networks are known to affect health behaviours in a range of different settings. Network effects on behaviour are complex, but statistical modelling provides an effective way of assessing behaviour on a real social network in the presence of other measured variables that affect the network structure. Using social network analysis to explore behaviour may become instrumental in defining targeted approaches to improving health in settings where social networks exist.

**Strengths and Limitations**

- This study uses a novel approach to understanding the effects of an individual's social network on their vaccination status.
- Rather than the expected diffusion of behaviour, we observed that having vaccinated neighbours reduced an individual's likelihood of being vaccinated.

- The application of the study findings may be limited because there are many factors that affect influenza vaccination decisions which could not be captured.

## Introduction

Influenza affects millions of people each year - it causes considerable morbidity and is a primary or underlying cause of death for thousands of people worldwide (1). The General Medical Council's (GMC) guidance on Good Medical Practice (2013), advises that healthcare workers (HCWs) in the UK receive immunisation against common serious communicable diseases, such as influenza, in order to protect both patients and colleagues (2). Higher coverage of influenza vaccination within a hospital is believed to reduce patient mortality, staff absences, and potential influenza epidemic size, thus protecting some of those at the greatest risk from influenza (3). Despite this, vaccination rates remain highly variable for HCWs and are below the government target of 75%. Currently around 63% of healthcare workers in England and Wales have a seasonal influenza vaccination (4,5).

There is increasing interest in the effects exerted by social networks on public health (6). A social network is made up of nodes (individuals) connected via ties (relationships) (7). Disease dynamics within a network may be influenced by characteristics such as its density, how the individuals in the network interact, and which individuals are vaccinated against, or susceptible to, the disease. For example, changes in the vaccination status of a few key individuals within a network may have a disproportional impact on disease spread (8). It has been shown that an individual's behaviour may be influenced by their peers – for example, research has found that smokers are more likely to befriend other smokers (9). The grouping of similar individuals within a population, known as homophily, could have a considerable impact on behaviour as well as disease dynamics. For example, if clusters of non-vaccinated

individuals exist within a network, a disease could rapidly spread through these groups, reducing the protective effects exerted by herd immunity.

Healthcare workers' vaccination behaviour may be influenced by the behaviour of their neighbours within their social network. Baron et al suggest that healthcare workers seem to be influenced by their co-workers' vaccination practices (10). In this study, network analysis is used to study the characteristics of a social network of foundation doctors (FDs) - early career doctors in the first two years of postgraduate training in the UK – and related these to the distribution of seasonal influenza vaccination within the same population. This was extended by investigating how the probability of an individual receiving an influenza vaccine was influenced by the behaviour of his/her neighbours in the network.

## Methods

Prospective ethical approval was obtained (15RECNA17) from Lancaster University Research Ethics Committee and the Pennine Acute Hospitals NHS Trust (PAT). Prior to data collection, each participant gave informed consent following a verbal and written explanation of the study. Identifiable data were collected and subsequently anonymised before data entry and analysis, as is accepted practice in SNA studies of this type.

## Participants

Data were collected during January/February 2015. All foundation doctors (FDs) working at the PAT during that period were invited to participate. The foundation training programme at the PAT runs over two years and across four different hospital sites in Greater Manchester, forming two geographically distinct axes, east and west. As

part of their training, FDs are required to attend compulsory weekly teaching sessions. Data collection took place during several of these sessions to optimise response rates.

## Patient and Public Involvement

This study involved early career doctors and no patients were involved. Findings have been presented at the study setting as part of ongoing work; however, it is likely (due to staff turnover) that some participants will not have had access to the findings of this work prior to its publication.

## Data Collection

Each participant completed a paper-based questionnaire. Participants self-reported their seasonal influenza vaccination status for winter 2014/15, alongside basic demographic information.

Participants were then asked how often they had contact with every other person on the foundation training programme using a six-point scale: 0 - I have never met this person; 1 - I recognise this person's name but wouldn't see them regularly; 2 - I occasionally see this person for very short periods of time; 3 - I see this person briefly at irregular intervals; 4 - I see this person on most shifts/4 or more days a week; 5 - I see this person on almost every shift for long time periods/live with them.

The relational data were then transferred into a numerical adjacency matrix, from which a network was constructed. Prior to analysis, the data were dichotomised at level 4, "I see this person on most shifts/4 or more days a week" and above, in line with previous research (8). Where one person declared a relationship with another at this level, this was assumed to be reciprocal. There may be cases in which neither person declared any

relationship, although one was present, this was treated as missing data and excluded. This produced an un-weighted (relationships were binary) and undirected (reciprocal ties were assumed) network.

## Social Network Analysis

The FDs' influenza vaccination status was evaluated as a node attribute on the social network. Individual-level network characteristics, such as a doctor's degree score (the number of ties an individual possesses), were examined along with global measures such as overall network density, and density in different groups within the network (the number of ties throughout the network in relation to the number of individuals within the network).

The assortativity coefficient was calculated to assess whether or not vaccination status showed homophily within the FD population. The assortativity coefficient is a standard network measure originally defined by Newman (11). The coefficient can range from -1 to 1, where -1 suggests negative assortativity (opposites attract) and 1 implies positive assortativity (like attracts like). With the assortativity coefficient we provide a tolerance interval for a random network by calculating the range of assortativity values expected from multiple generated random networks. We generated a reference distribution using permutation. Multiple networks (n=1000) were generated with the same topological structure, but with vaccination status (yes/no) permutated randomly amongst the participants The assortativity value for each was then calculated – this provided the range of assortativity values we would expect under random permutation. Similar techniques are outlined by Barclay et al. (12).

## Auto-logistic Regression

The auto-logistic model, was used to further investigate the effect of an individual's social connections on their influenza vaccination decision (13). This model allows an individual's vaccination behaviour to be modelled as a function of their demographic information and the behaviour of their neighbours in the social network. The specification of the auto-logistic model is given in Equation 1.

$$For,\ Y_i = \begin{bmatrix} 0 : not\ vaccinated \\ 1 : vaccinated \end{bmatrix}$$

[Equation 1]

$$\log\left(\frac{\mathrm{P}(\mathrm{Y_i} = 1|\ \text{all other } \mathrm{Y_i})}{1 - \mathrm{P}(\mathrm{Y_i} = 1|\ \text{all other } \mathrm{Y_i})}\right) = \alpha + \underline{x'}_i\underline{\beta} + \gamma \sum_{\mathrm{j}\sim\mathrm{i}} (\mathrm{Y_j} = 1)$$

*Where j ~ i indicates contact between individuals i and j, α indicates the intercept and* $\underline{x_i}$ *is a vector of covariates associated with individual i.*

The parameters β describe how the covariates affect the likelihood of an individual being vaccinated, whilst the parameter γ describes how this likelihood is modified by the behaviour of the individual's social contacts in the network. To fit the model, we used Monte Carlo likelihood inference (14), using numerical optimisation with initial values of β derived by fitting a standard logistic regression and initial value of γ = 0. The logistic regression model is a sub-model of the auto-logistic model when γ = 0, which was used to give initial parameter estimates for $\alpha$ and β, but not for formal

inference. We repeated the optimisation multiple times, using data simulated from the fitted model, to generate parametric bootstrap confidence intervals.

## Results

One hundred and thirty-eight of the 200 foundation doctors invited to take part provided complete data (sex, year of training, axis, and vaccination status). Amongst respondents, 100 (72%) were vaccinated (Table 1).

| | | Number Vaccinated | Total | Vaccination coverage (%) |
|---|---|---|---|---|
| **Sex** | **Female** | 51 | 68 | 75.00 |
| | **Male** | 49 | 70 | 70.00 |
| **Year** | **1** | 55 | 76 | 72.37 |
| | **2** | 45 | 62 | 72.58 |
| **Axis** | **East** | 47 | 69 | 68.12 |
| | **West** | 53 | 69 | 76.81 |

**Table 1: Seasonal influenza vaccination uptake by the foundation doctors stratified by their demographic factors.**

Figure 1 shows the foundation doctors' social network, along with their influenza vaccination status (n=138). The assortativity coefficient for the entire FD social network was -0.034 with a tolerance interval of (-0.12, 0.10).

*figure 1 here*

**Figure 1. The foundation doctor social network sociogram for those who returned complete data, dichotomised at >= 4 ("I see this person on most shifts/ 4 or more days a week"), and coloured according to individual vaccination status.**

The social network structure of the foundation doctors varied between geographical areas and year-groups (Figure 2). For example, amongst second-year doctors, the network density is higher in the east than in the west axis, with 223 ties amongst the n=31 doctors in the east axis compared with 73 ties amongst the same number in the west axis.

*figure 2 here*

**Figure 2. A sociogram depicting the foundation doctor network (n=138), coloured by sub-groups: year and axis.**

We first fitted the auto-logistic model without covariates. The maximum likelihood surface for auto-logistic model 1 is described by Figure 3. The coefficient for γ, the social network parameter, was -0.122, with 95% confidence interval (-0.197, -0.047), i.e. a repulsion effect – individuals were more likely to act in opposition to the behaviour of their neighbours. Note that this agrees qualitatively with the negative estimate of the assortativity coefficient, but the more efficient model-based approach is more efficient, leading in this instance to a statistically significant departure from γ=0.

We then added covariate effects for year and axis. We also tried to fit a model that included the interaction term between these main effects. However, the data proved to be too sparse for this. In the model with interaction between year and axis, participants are split into four groups, with very small numbers of non-vaccinated and vaccinated individuals within each group, even before adding the complication of the neighbourhood structure (Table 2 and Supplementary). We were able to fit and compare three models: model one – with no covariates; model two – with main effects for year and axis (two binary factors); and model three – a main effect for those in year two in the west axis vs all others (one binary factor). Model three is not nested within model two, thus likelihood ratio comparison between these would be a non-regular problem. Comparison between model one and the other two models suggested that model one was adequate and so this model was selected. Adjusting for covariates did not account for the negative effect of the network term, $\gamma$. In all cases, the confidence intervals for the covariate effects included zero. This suggests that the repulsion effect seen in both the social network analysis and the auto-logistic analysis without covariates (albeit non-significant in the former case) is an artefact of the sub-division of the network according to year group and/or axis.

*figure 3 here*

**Figure 3. Contour plot showing the likelihood surface for auto-logistic model 1.**

**Table 2. Contingency table showing the spread of the foundation data between year, axis and vaccination status.**

| | | **Axis east** | | **Axis west** | |
|---|---|---|---|---|---|
| | | year 1 | year 2 | year 1 | year 2 |
| **vaccinated** | yes | 23 | 24 | 32 | 21 |

| | | | | | |
|---|---|---|---|---|---|
| | no | 15 | 7 | 6 | 10 |

**Table 3. Parameter estimates for the auto-logistic regression models fit using the foundation doctor data.**

| | | Parameter Estimate | Lower CI | Upper CI | MC se |
|---|---|---|---|---|---|
| **Auto-logistic model 1** ***(Excluding covariates)*** | Intercept coefficient | 1.532 | 0.977 | 2.087 | 0.09 |
| | Gamma coefficient | -0.122 | -0.197 | -0.047 | 0.01 |
| **Auto-logistic model 2** | Intercept coefficient | 1.342 | 0.539 | 2.145 | 0.136 |
| | Year coefficient | -0.149 | -1.045 | 0.747 | 0.115 |
| | Axis coefficient | 0.498 | -0.942 | 1.240 | 0.163 |
| | Gamma coefficient | -0.121 | -0.223 | -0.019 | 0.010 |
| **Auto-logistic model 3** | Intercept | 1.701 | 0.937 | 2.465 | 0.007 |
| | Year = 2 & Axis = West | -0.501 | -1.481 | 0.479 | 0.009 |
| | Gamma coefficient | -0.136 | -0.210 | -0.062 | 0.0007 |

# Discussion

After excluding missing data, the foundation doctors' self-reported vaccination coverage of 72% (100 vaccinated out of 138, with possible range 50% - 81% dependent on the vaccination status of non-respondents), was higher than the national average of 55% (15).

The auto-logistic model has allowed us to assess which areas of the population are the less likely to vaccinate, taking into account their social network structure. For example, we hypothesised that year group or axis may affect an individual's likelihood of receiving the vaccination. However, the confidence intervals for all demographic factors in the auto-logistic model included zero. This suggests that the statistically significant network structure of vaccination cannot be accounted for by the demographic

information. Using this statistical modelling approach, has provided a better understanding of the social network structure on vaccination uptake than could be obtained using only the assortativity coefficient, both through its greater statistical efficiency and its ability to investigate whether, and if so to what extent, measured covariates can explain the network structure.

Our analysis of the foundation doctor population using the auto-logistic model without covariates produced a negative estimate of γ, the network parameter, suggesting that having vaccinated neighbours reduces the individual's probability of being vaccinated (Table 3). We observed other differences in the network structure amongst the four sub-groups defined by year and geographical axis. Second year foundation doctors on the west axis of the Trust had a much sparser social network than the other year/axis groups. In sparse social networks the potential for information transfer (behaviour adoption, infectious disease spread, etc) is fundamentally diminished by social distancing (16). However, Shirley et al. suggest that even when network density is equivalent, network topology may still have an effect on diffusion of information (17). The analysis of the FD data suggests that demographic covariates were unable to account for the dis-assortativity of vaccination on the network. However, only a limited number of covariates were available. More research would be needed to identify other factors that may affect the transfer of vaccination attitude amongst friends. Interventions aimed at improving vaccination uptake should be sensitive to the differences between sub-groups within the relevant population and may need to be targeted at specific demographic sub-groups. Network effects on behaviour are complex, but the auto-logistic model provides an effective way of assessing behaviour on a real social network in the presence of other variables that affect individuals' responses.

The repulsion effect on vaccination uptake seen in this population is unusual, as more commonly diffusion of behaviour is observed. However, vaccination is a complex behaviour in which there is a cost to taking the vaccination (pain of injection, perceived side effects, etc.) to be weighed against the benefits of vaccinating. The behaviour of others directly affects the individual – if more people are vaccinated the risk of infection is lower for all (18). Furthermore, it may also be the case that the misperceptions surrounding the influenza vaccination are more commonly discussed than the benefits within this population (19). Better understanding of the role social relationships play in establishing the vaccination behaviour of HCWs in the workplace is necessary to inform vaccination campaigns, whose ultimate goal is to improve occupational health and patient wellbeing.

We have outlined a methodological approach to understanding behaviour in a network. The auto-logistic model could easily be modified in a number of ways, for example to include the network term using the proportion, rather than number, of neighbours who are vaccinated. Furthermore, although the approach has been successful in fitting a parsimonious model to this relatively small dataset, attempts to fit more complex models quickly lead to large standard errors and, consequently, low power to detect more complex network structure. A larger dataset would enable more complex models to be fitted and more precise inferences. In our analysis, we chose to ignore non-respondents because their social contacts were unknown. Future research into this modelling approach should include investigation into the estimation of missing data to allow subjects with partially observed information to be included in the analysis, and to investigate whether non-participation is informative, i.e. non-participants have atypical vaccination behaviour (20)

An inherent limitation of our data is that they were self-reported, and therefore potentially subject to reporter bias. Additionally, making a decision about influenza vaccination is a complex process – many people are neither completely for nor completely against influenza vaccination (21). There may be varying levels of attitudes to vaccination that could be described using an ordinal or continuous scale, rather than as a simple binary variable. Extracting this more nuanced data is a challenge, and requires qualitative methods such as in-depth interviews with participants (21).

Using the auto-logistic modelling approach, we have expanded on the results of the social network analysis. This novel approach to analysing social network data allows us to investigate in more detail the underlying process that has led to an observed network and its vaccination distribution. Quantitative methods that explore social behaviour are likely to become instrumental in defining targeted approaches to improving public health - this study outlines a suitable approach to investigating how an individual's behaviour might be influenced by the behaviour of their neighbours in a network.

Social networks are powerful phenomena that may be harnessed to encourage diffusion of positive health behaviours (21). We have shown that this is particularly relevant in an occupational setting where somewhat artificial social networks are formed with clearly defined boundaries, and knowledge about occupational health is exchanged between workers.

## Acknowledgments

The Authors would like to thank the foundation doctors at the Pennine Acute Hospitals NHS Trust for generously giving up their time to participate in this study and staff who enabled data collection.

## Funding

This work was supported by the Colt Foundation.

## Author Contributions

RE wrote the manuscript with input from all other authors all of which have approved the final version. RE collected the data. PJD devised the statistical methodology, which was implemented by RE and PJD. All authors contributed to the conception of the project. This study is part of a larger programme of work devised by RI.

## Data Sharing

Additional data is available in the supplementary material.

# References

1. WHO . Influenza: Fact Sheet [Internet]. 2014. Available from: http://www.who.int/mediacentre/factsheets/fs211/en/

2. General Medical Council . Good Practice Guidelines [Internet]. 2013 [cited 2016 Aug 18]. Available from: http://www.gmc-uk.org/guidance/good_medical_practice/contents.asp

3. Poland GA, Tosh P, Jacobson RM. Requiring influenza vaccination for health care workers: seven truths we must accept. Vaccine. 2005;23(17):2251–5.

4. Anikeeva O, Braunack-Mayer A, Rogers W. Requiring influenza vaccination for health care workers. Am J Public Health. 2009;99(1):24–9.

5. Public Health England. Seasonal flu vaccine uptake in healthcare workers in England: winter season 2016 to 2017 [Internet]. 2017 [cited 2018 Mar 2]. Available from: https://www.gov.uk/government/statistics/seasonal-flu-vaccine-uptake-in-healthcare-workers-in-england-winter-season-2016-to-2017

6. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008;5(3):e74.

7. Scott J, Carrington PJ. The SAGE handbook of social network analysis. SAGE publications; 2011.

8. Edge R, Heath J, Rowlingson B, Keegan TJ, Isba R. Seasonal Influenza Vaccination amongst Medical Students: A Social Network Analysis Based on a Cross-Sectional Study. PloS One. 2015;10(10):e0140085.

9. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med. 2008;358(21):2249–58.

10. Baron G, De Wals P, Milord F. Vaccination practices of Quebec family physicians. Influenza vaccination status and professional practices for influenza vaccination. Can Fam Physician. 2001;47(11):2261–6.

11. Newman ME. Assortative mixing in networks. Phys Rev Lett. 2002;89(20):208701.

12. Barclay VC, Smieszek T, He J, Cao G, Rainey JJ, Gao H, et al. Positive Network Assortativity of Influenza Vaccination at a High School: Implications for Outbreak Risk and Herd Immunity. PloS One. 2014;9(2):e87042.

13. Besag J. Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Ser B Methodol. 1974;192–236.

14. Geyer CJ. Likelihood inference for spatial point processes. Stoch Geom Likelihood Comput. 1999;80:79–140.

15. Public Health England . Seasonal influenza vaccine uptake amongst frontline HCWs in England: winter season 2013 to 2014 [Internet]. Available from: https://www.gov.uk/government/publications/seasonal-influenza-vaccine-uptake-amongst-frontline-hcws-in-england-winter-season-2013-to-2014

16. Valente TW. Social network thresholds in the diffusion of innovations. Soc Netw. 1996 Jan 1;18(1):69–89.

17. Shirley MDF, Rushton SP. The impacts of network topology on disease spread. Ecol Complex. 2005 Sep;2(3):287–99.

18. Bauch CT, Earn DJ. Vaccination and the theory of games. Proc Natl Acad Sci U S A. 2004;101(36):13391–4.

19. Hofmann F, Ferracin C, Marsh G, Dumas R. Influenza vaccination of healthcare workers: a literature review of attitudes and beliefs. Infection. 2006;34(3):142–7.

20. Rubin DB. Inference and missing data. Biometrika. 1976 Dec 1;63(3):581–92.

21. Edge R, Goodwin D, Isba R, Keegan T. Socialization, Indifference, and Convenience: Exploring the Uptake of Influenza Vaccine Among Medical Students and Early Career Doctors. Qual Health Res. 2017 Jul 24;1049732317720664.

22. Edge R, Goodwin D, Isba R, Keegan T. Socialization, Indifference, and Convenience: Exploring the Uptake of Influenza Vaccine Among Medical Students and Early Career Doctors. Qual Health Res. 2017 Jul 24;1049732317720664.

23. Valente TW. Social networks and health: Models, methods, and applications. Vol. 1. Oxford University Press New York; 2010.

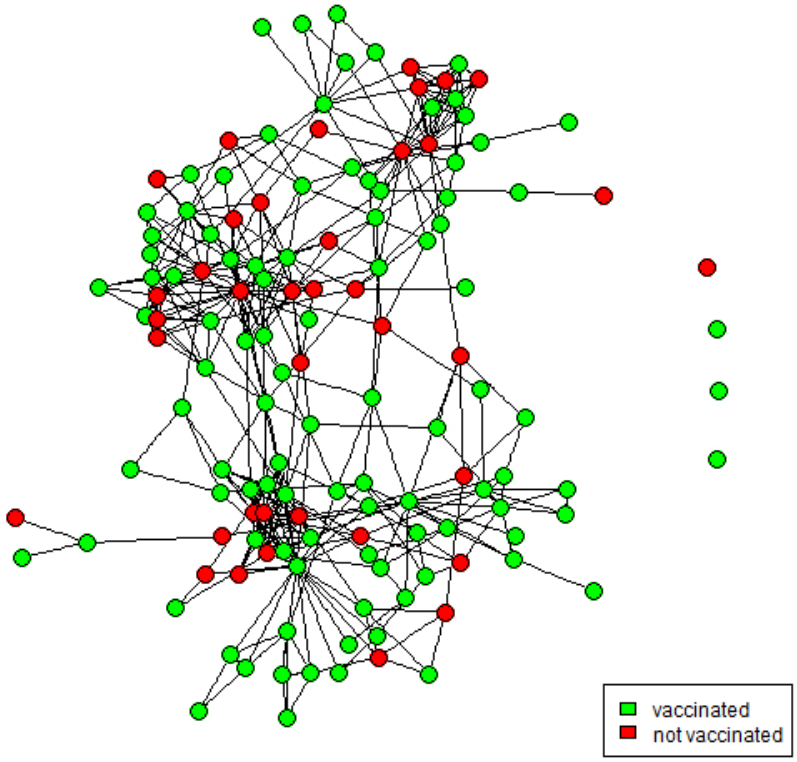



Figure 1. The foundation doctor social network sociogram for those who returned complete data, dichotomised at >= 4 ("I see this person on most shifts/ 4 or more days a week"), and coloured according to individual vaccination status.

211x211mm (72 x 72 DPI)

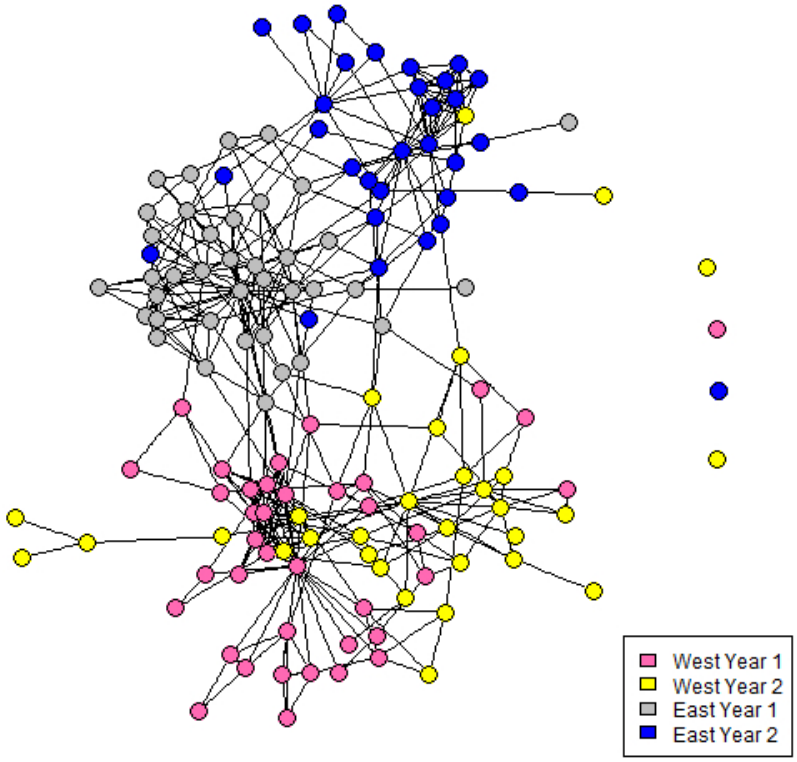



Figure 2. A sociogram depicting the foundation doctor network (n=138), coloured by sub-groups: year and axis.

211x211mm (72 x 72 DPI)

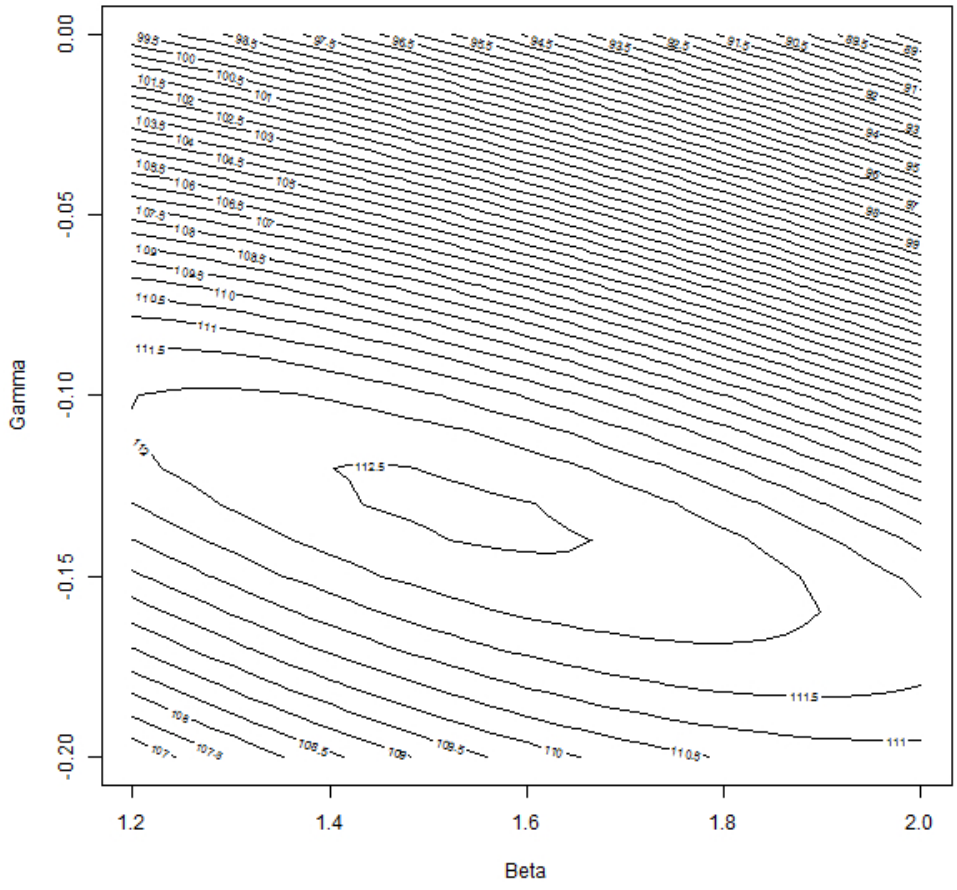



Figure 3. Contour plot showing the likelihood surface for auto-logistic model 1.

211x211mm (72 x 72 DPI)

# Supplementary

# Auto-logistic model parameter estimation

The auto-logistic model incorporates spatial correlation into the logistic model for binary data. The specification is as follows, let Y be our variable of interest, where $Y_i$ $\epsilon(0,1)$ represents the observation at the ith data point for *i = 1…n,* the full conditional distributions are given by:

$$For,\ Y_i = \begin{bmatrix} 0 : not\ vaccinated \\ 1 : vaccinated \end{bmatrix}$$

$$\log\left(\frac{P(Y_i=1|\text{ all other } Y_i)}{1-P(Y_i=1|\text{ all other } Y_i)}\right) = \underline{x}'_i\underline{\beta} + \gamma \sum_{j\sim i}(Y_j = 1).$$

From Besag 1974, we have

$$\frac{P(\underline{y})}{P(\underline{0})} <=> f\left(\underline{y}\right) = c(\theta)g(y)$$

Where $c(\theta)$ is an intractable constant and $g(y)$ is a known function. We can manipulate this to use Geyer's method of Monte-Carlo maximum likelihood (17):

$$f\left(\underline{y}\right) = c(\theta)g(y)$$

$$\int f\left(\underline{y}\right) = \int c(\theta)\ g(y;\theta)$$

$$1 = c(\theta) \int g(y;\theta)dy$$

$$1 = \frac{c(\theta)}{c(\theta_0)} \int g(y;\theta) * \frac{g(y;\theta_0)}{g(y;\theta_0)}\ c(\theta_0)\ dy$$

$$= \frac{c(\theta)}{c(\theta_0)} \int \frac{g(y;\theta)}{g(y;\theta_0)}\ f(y;\theta_0)\ dy$$

$$= \frac{c(\theta)}{c(\theta_0)} * E_0\left[\frac{g(y;\theta)}{g(y;\theta_0)}\right] \quad (1)$$

We can simulate the expectation in equation 1 using a Monte-Carlo approximation to the expectation.

$$\approx \frac{1}{\varphi}\sum_{1}^{\varphi}\frac{g(y;\theta)}{g(y;\theta_0)}$$

Rearranging (1) gives:

$$c(\theta) = \frac{c(\theta_0)}{\widehat{E_0}}$$

Thus, we have:

$$f(y;\underline{\theta}) \approx c(\theta)g(y;\theta) = \frac{c(\theta_0)}{\widehat{E_0}} * g(y;\theta)$$

$$L_{mc}(\theta) = \log\big(f(y;\theta)\big) = \log\big(c(\theta_0)\big) - \log\big(\widehat{E_0}\big) + \log(g(y;\theta))$$

The term: $\log\big(c(\theta_0)\big)$ is a constant, therefore, $\hat{\theta}$ maximises the terms: $-\log\big(\widehat{E_0}\big) + \log\big(g(y;\theta)\big)$.

Maximising this gives a Monte-Carlo approximation to the maximum likelihood estimator (MLE). When n is large maximum likelihood estimators have normal properties:

$$\hat{\theta} \approx N(\theta_0, \frac{1}{n\, I(\theta_0)})$$

where I(θ) is the information matrix.

As we are utilising this methodology with network data, it is necessary to check whether the asymptotic principles hold given our sample size (and network structure) – thus, we used simulation studies to investigate the properties of the model.

**Simulation Experiments**

For each experiment using the auto-logistic model, fixed parameter θ, and the network structure from our foundation doctor social network data, we are able to generate data samples, Y.

To explore the behaviour of the auto-logistic model and our implementation, we firstly generate multiple new response data sets, $Y_i$. We then estimate parameters for these realisations using Monte-Carlo maximum likelihood estimation. This results in a set of estimates for θ. Inference on this set allows us to explore the model's behaviour under different conditions i.e. different values of θ. This scheme is outlined graphically by Figure 1. Monte-Carlo maximum likelihood tends to the true values for θ as n tends to infinity. We have a finite value of n, thus we need to check whether we are providing sensible estimates for θ. This is achieved by comparing estimates of $\hat{\theta}$ to known values of θ.

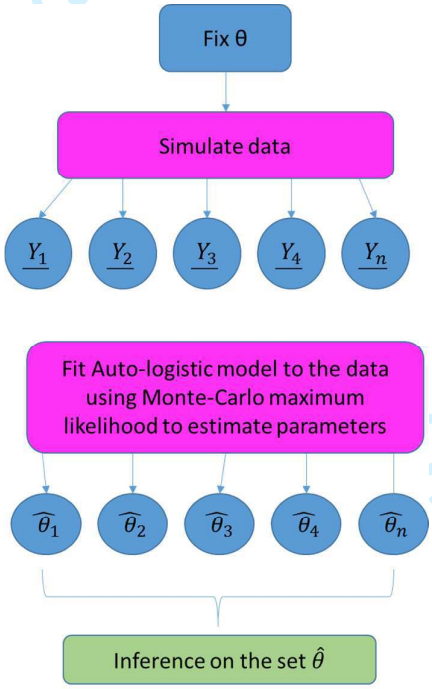


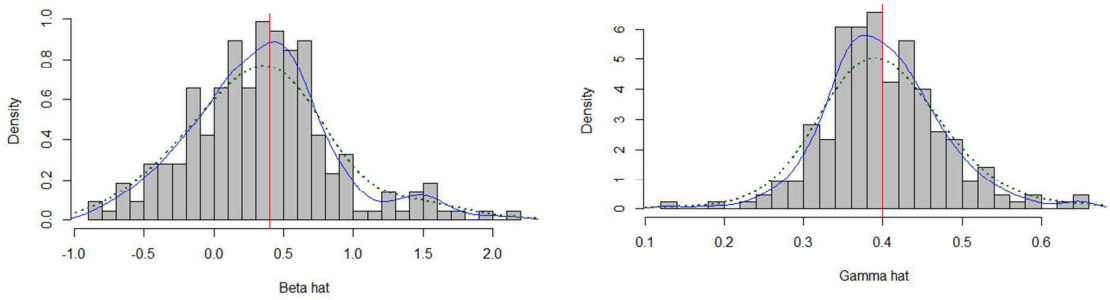
**Figure 1: Regime for the simulation experiments using the auto-logistic model.**

**Test 1: Examine the model implementation.**

We chose, $\theta = \{\beta = 0.3, \gamma = 0.4\}$, to test the model code's functionality.

We repeated the experiment 200 times (i = 200) producing $Y_i$ and $\widehat{\theta_i}$ .

We are returning reasonably accurate and normally distributed results, even with a relatively small number of simulation runs, particularly for our estimates of γ, which is the parameter that estimates the relational effect of neighbours in the network (Figure 2, Table 1). The variance for the β parameter is much larger - this parameter controls for the intercept / marginal probability of vaccination. However, despite the poor model performance due to a very small number of monte-carlo simulations, we are still returning expected parameter estimates. This adds to the body of evidence that our code is acting as expected, although it is obvious from this test that the model must be run for far longer.




| | $\hat{\beta}$ | $\hat{\gamma}$ |
|---|---|---|

**Figures 2a and 2b: The distribution of $\hat{\beta}$ and $\hat{\gamma}$ for test 1**

**Table 1: Descriptive statistics for the first experiment (2dp).**

| | | |
|---|---|---|
| Median | 0.36 | 0.40 |
| Mean | 0.35 | 0.40 |
| Variance | 0.27 | 0.006 |

Figure 3 shows that extreme values for γ, tend to correspond with extreme β values, suggesting that more extreme values are mostly likely associated with random noise in the simulation process (Pearson's correlation coefficient: -0.78).

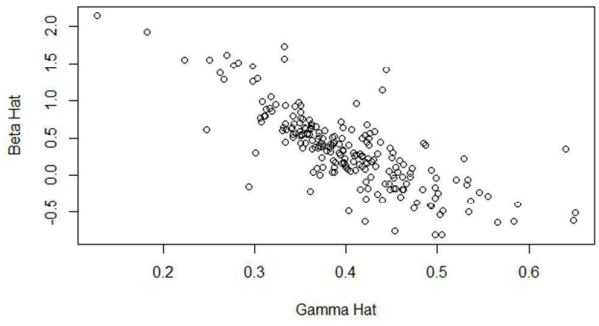



**Figure 3: Scatter plot of estimates for beta against corresponding estimates for gamma.**

## Test 2: Examine the properties of the hessian matrix

The Hessian is a square matrix of second-order partial derivatives. When maximising a likelihood, the covariance matrix of the estimates is (asymptotically) the inverse of the

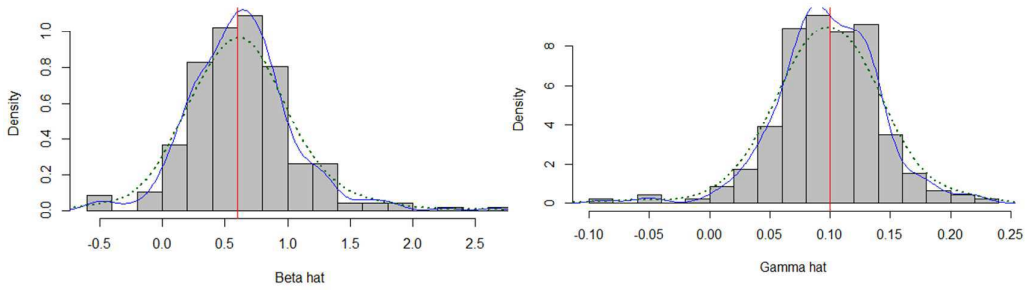
negative of the Hessian. Thus, the standard errors are the square roots of the diagonal elements of the covariance.

$$\boldsymbol{Var(\theta) \approx [I(\theta)]^{-1}}$$

Before using the hessian matrix to derive confidence intervals for our parameter estimates we must confirm, using a simulation study that this principle holds when using our true network (n=138).

We ran this simulation study using β = 0.6, γ = 0.1. The simulation was repeated for 235 runs to generate a dataset of parameter estimates. This was used to generate a parametric bootstrap of the variance. At each iteration, we also saved the hessian matrix, allowing us to derive a dataset of asymptotically estimated variances'.

Figure 4 confirms that again the simulation is producing sensible results for the given values of β and γ - the plots are approximately normally distributed, and returned expected parameter estimates despite a limited number of monte-carlo simulations (see Table 2).



| | Beta hat | Gamma hat |
|---|---|---|
| median | 0.62 | 0.10 |
| mean | 0.63 | 0.10 |
| variance | 0.18 | 0.002 |

**Figure 4: The distribution of $\hat{\beta}$ and $\hat{\gamma}$ for test 2.**

**Table 2: Summary statistics for experiment 3 parameter estimates. The real values are beta = 0.6, gamma= 0.1.**

We computed the asymptotic variance for each run of the simulation using the hessian matrix. This generated a set of values for the asymptotic variance for each run of the simulation, the mean for the asymptotic variance for beta was <0.0001, and the mean asymptotic variance for gamma was <0.0001. This is small compared to the bootstrapped variance, however, generating and storing the hessian matrix for each simulation was costly in computational time.

## Real data application

The auto-logistic model was imputed using initial values derived from maximising the pseudo-likelihood for β (β = 0.9675), and for γ we initially assumed the null hypothesis (that γ = 0). We can see, from Figure 5 that the model performed as expected under these conditions – the peak is equivalent to the pseudo-likelihood.

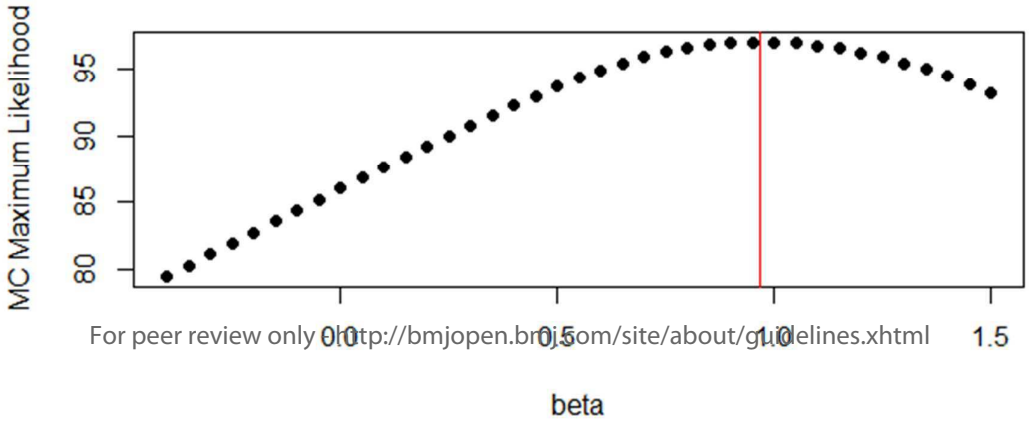
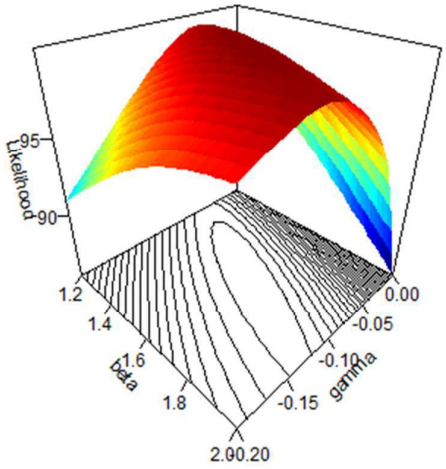


**Figure 5. Auto-logistic model performance when applied to the real data, under the null hypothesis.**

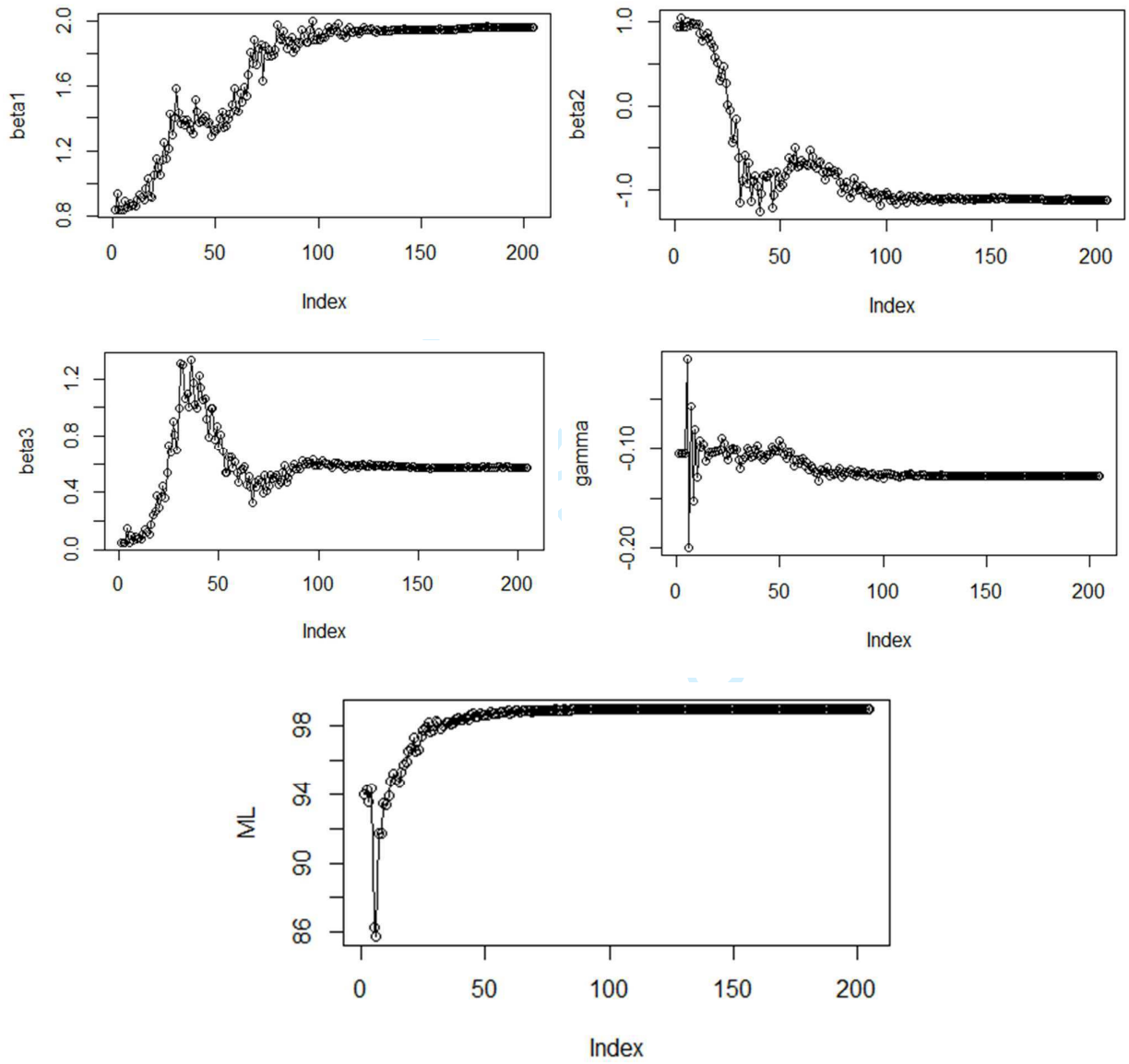
Parameter estimates were chosen by optimisation of the maximum likelihood function. Figure 6, shows a surface plot of the likelihood for the auto-logistic model with parameter estimates for β, and γ.



Following our analysis of the simulation experiments we felt confident to continue with the analysis of the foundation doctor data as described in the manuscript attached. For the real data analysis we generated many more simulations and increased the number of runs in an attempt to reduce the Monte-Carlo error.

Figure panel 7 shows a trace plot for each of the parameters in the second model (with axis and year as covariates). This is an example of the results from one run of the model, which were combined with many more to produce the parameter estimates and the associated monte-carlo simulation error given previously. It is clear that the model mixes and converges appropriately.

**Figure 7. An example of a series of trace plots for each of the parameters in the auto-logistic model with year and axis as covariates, produced from one run of the model.**




When we included an interaction term into the model we experienced poor convergence in which parameter estimates did not stabilise at realistic values, but

rather showed evidence of numerical instability – the parameter estimates tended to very large values (both positive and negative). We applied a range of diagnostic techniques to investigate this phenomenon. To confirm the code was correct in this case, we fixed the interaction parameter at zero, and re-ran the model, the results were as expected, the parameter estimates converged and the traces plots were comparable to those produced by the second model (Figure panel 7). This suggests that the model code was performing as expected. By plotting the parameter estimates against one another, we deduced that the model with an interaction term was not identifiable. We believe that this is due to a sparsity of data (Table 3).

**Table 3. Contingency table of the frequency of foundation doctors by, neighbourhood, year, axis and vaccination status.**

| | | | Axis east | | Axis west | |
|---|---|---|---|---|---|---|
| | | | year 1 | year 2 | year 1 | year 2 |
| **Neighbours vaccinated = 0** | vaccinated | yes | 1 | 1 | 2 | 1 |
| | | no | 0 | 2 | 0 | 1 |
| **Neighbours vaccinated = 1** | vaccinated | yes | 1 | 4 | 2 | 5 |
| | | no | 0 | 0 | 0 | 2 |
| **Neighbours vaccinated = 2** | vaccinated | yes | 4 | 10 | 5 | 2 |
| | | no | 2 | 0 | 0 | 3 |
| **Neighbours vaccinated = 3** | vaccinated | yes | 3 | 4 | 10 | 4 |
| | | no | 2 | 0 | 2 | 2 |
| **Neighbours vaccinated = 4** | vaccinated | yes | 7 | 3 | 6 | 5 |
| | | no | 4 | 3 | 2 | 0 |
| **Neighbours vaccinated = 5** | vaccinated | yes | 3 | 0 | 2 | 1 |
| | | no | 1 | 1 | 0 | 1 |
| **Neighbours vaccinated = 6** | vaccinated | yes | 2 | 0 | 2 | 0 |
| | | no | 4 | 0 | 0 | 0 |
| **Neighbours vaccinated = 7** | vaccinated | yes | 1 | 0 | 1 | 1 |
| | | no | 0 | 0 | 0 | 0 |
| **Neighbours vaccinated = 8** | vaccinated | yes | 1 | 1 | 0 | 1 |
| | | no | 1 | 1 | 0 | 1 |
| **Neighbours vaccinated = 9** | vaccinated | yes | 1 | 0 | 0 | 0 |
| | | no | 3 | 0 | 0 | 0 |
| **Neighbours vaccinated = 10** | vaccinated | yes | 0 | 0 | 0 | 1 |
| | | no | 0 | 0 | 0 | 0 |
| **Neighbours vaccinated = 14** | vaccinated | yes | 0 | 0 | 0 | 0 |
| | | no | 0 | 1 | 0 | 0 |
| **Neighbours vaccinated = 16** | vaccinated | yes | 0 | 0 | 0 | 0 |
| | | no | 1 | 0 | 0 | 0 |

BMJ Open

# An observational study to assess the effects of social networks on the seasonal influenza vaccine uptake by early career doctors.

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2018-026997.R1 |
| Article Type: | Research |
| Date Submitted by the Author: | 13-Jun-2019 |
| Complete List of Authors: | Edge, Rhiannon; Lancaster University Faculty of Health and Medicine, Lancaster Medical School<br>Keegan, Thomas; Lancaster University, Lancaster Medical School<br>Isba, Rachel; Lancaster University Faculty of Health and Medicine, Lancaster Medical School<br>Diggle, Peter; Lancaster University, Lancaster Medical School; University of Liverpool, Epidemiology and Population Health |
| Primary Subject Heading: | Occupational and environmental medicine |
| Secondary Subject Heading: | Epidemiology, Research methods, Public health |
| Keywords: | Social network analysis, influenza vaccination, auto-logistic regression, occupational health |
| | |

# An observational study to assess the effects of social networks on the seasonal influenza vaccine uptake by early career doctors.

R Edge*, T J Keegan, R Isba, P J Diggle

Lancaster Medical School, Furness Building, Lancaster University, Bailrigg, Lancaster, LA1 4YW.

Lecturer, Senior Lecturer, Professor, Professor.

* Correspondence to: r.edge@lancs.ac.uk.

## Abstract

### Objectives

To evaluate the effect of social network influences on seasonal influenza vaccination uptake by healthcare workers.

### Design

Cross-sectional, observational study.

### Setting

A large secondary care NHS Trust, which includes four hospital sites in Greater Manchester.

### Participants

Foundation doctors (FDs) working at the PAT during the study period. Data collection took place during compulsory weekly teaching sessions there were no exclusions. Of the 200 eligible FDs, 138 (70%) provided complete data.

### Primary outcome measures

Self-reported seasonal influenza vaccination status.

### Results

Amongst participants, 100 (72%) reported that they had received a seasonal influenza vaccination. Statistical modelling demonstrated that having a higher proportion of vaccinated neighbours increased an individual's likelihood of being vaccinated. The coefficient for $\gamma$, the social network parameter, was 0.965 (95% confidence interval: 0.248, 1.682; odds: 2.625 (95% confidence interval: 1.281, 5.376)), i.e. a diffusion effect. Adjusting for year group, geographical area, and sex did not account for this effect.

### Conclusions

This population exhibited higher than expected vaccination coverage levels– providing protection both in the workplace and for vulnerable patients. The modelling approach

allowed covariate effects to be incorporated into social network analysis, which gave us a better understanding of the network structure. These techniques have a range of applications in understanding the role of social networks on health behaviours.

**Key words**

Social network analysis, influenza vaccination, auto-logistic regression, occupational health.

## Strengths and limitations

- This study uses a novel auto-logistic regression approach to understanding the effects of an individual's social network on their vaccination status.
- The auto-logistic regression approach to social network analysis provides a unique quantitative framework for comprehensively understanding social behaviours.
- The application of the study findings may be limited because there are many factors that affect influenza vaccination decisions that could not be captured using the data collection methods.
- Data were self-reported, which may have introduced bias.

## Introduction

Influenza affects millions of people each year - it causes considerable morbidity and is a primary or underlying cause of death for thousands of people worldwide (1). The General Medical Council's (GMC) guidance on Good Medical Practice (2013), advises that healthcare workers (HCWs) in the UK receive immunisation against common serious communicable diseases, such as influenza, in order to protect both patients and colleagues (2). Higher coverage of influenza vaccination within a hospital is believed to reduce patient mortality, staff absences, and potential influenza epidemic size, thus protecting some of those at the greatest risk from influenza (3). Despite this, vaccination rates remain highly variable for HCWs and are below the government target of 75%. In 2016/17, around 63% of healthcare workers in England and Wales recieved a seasonal influenza vaccination (4,5).

There is increasing interest in the effects exerted by social networks on public health (6). A social network is made up of nodes (individuals) connected via ties (relationships) (7). Disease dynamics within a network may be influenced by characteristics such as its density, how the individuals in the network interact, and which individuals are vaccinated against, or susceptible to, the disease. For example, changes in the vaccination status of a few key individuals within a network may have a disproportional impact on disease spread (8). It has been shown that an individual's behaviour may be influenced by their peers – for example, research has found that smokers are more likely to befriend other smokers (9). The grouping of similar individuals within a population, known as homophily, could have a considerable impact on behaviour as well as disease dynamics. For example, if clusters of non-vaccinated

individuals exist within a network, a disease could rapidly spread through these groups, reducing the protective effects exerted by herd immunity.

Healthcare workers' vaccination behaviour may be influenced by the behaviour of their neighbours within their social network. Baron et al suggest that healthcare workers seem to be influenced by their co-workers' vaccination practices (10). In this study, network analysis is used to study the characteristics of a social network of foundation doctors (FDs) - early career doctors in the first two years of postgraduate training in the UK – and related these to the distribution of seasonal influenza vaccination within the same population. This was extended by investigating how the probability of an individual receiving an influenza vaccine was influenced by the behaviour of his/her neighbours in the network.

## Methods

Prospective ethical approval was obtained (15RECNA17) from Lancaster University Research Ethics Committee and the Pennine Acute Hospitals NHS Trust (PAT). Prior to data collection, each participant gave informed consent following a verbal and written explanation of the study. Identifiable data were collected and subsequently anonymised before data entry and analysis, in line with accepted practice in SNA studies of this type.

## Participants

Data were collected during January/February 2015. All foundation doctors (FDs) working at the PAT during that period were invited to participate. The foundation training programme at the PAT runs over two years and across four different hospital sites in Greater Manchester, forming two geographically distinct axes, east and west. As

part of their training, FDs are required to attend compulsory weekly teaching sessions. Data collection took place during several of these sessions to optimise response rates.

All participants will have been offered a free seasonal influenza vaccine before the point of data collection. The PAT actively encourages influenza vaccination for its staff, as does the GMC. Staff are given numerous opportunities to have the vaccine, there are often vaccination points established at mutually convenient locations (hospital entrances, cafeterias, etc.) as well as travelling vaccination nurses who offer the vaccine ward-to-ward. We have assumed that all participants have had ample opportunity to vaccinate however, we have not collected data specifically regarding participant's exposure to seasonal influenza vaccination opportunities.

## Patient and Public Involvement

This study involved early career doctors and no patients were involved. Initial findings were presented at the study setting as part of ongoing work; however, it is likely (due to staff turnover) that many participants will not have had access to the findings of this work prior to its publication.

## Data Collection

Each participant completed a paper-based questionnaire. Participants self-reported their seasonal influenza vaccination status for winter 2014/15, alongside basic demographic information.

Participants were then asked how often they had contact with every other person on the foundation training programme using a six-point scale: 0 - I have never met this

person; 1 - I recognise this person's name but wouldn't see them regularly; 2 - I occasionally see this person for very short periods of time; 3 - I see this person briefly at irregular intervals; 4 - I see this person on most shifts/4 or more days a week; 5 - I see this person on almost every shift for long time periods/live with them.

The relational data were then transferred into a numerical adjacency matrix, from which a network was constructed. Prior to analysis, the data were dichotomised at level 4, "I see this person on most shifts/4 or more days a week" and above, in line with previous research (8). Where one person declared a relationship with another at this level, this was assumed to be reciprocal. There may be cases in which neither person declared any relationship, although one was present, this was treated as missing data and excluded. This produced an un-weighted (relationships were binary) and undirected (reciprocal ties were assumed) network.

## Social Network Analysis

The FDs' influenza vaccination status was evaluated as a node attribute on the social network. Individual-level network characteristics, such as a doctor's degree score (the number of ties an individual possesses), were examined along with global measures such as overall network density, and density in different groups within the network (the number of ties throughout the network in relation to the number of individuals within the network).

The assortativity coefficient was calculated to assess whether or not vaccination status showed homophily within the FD population. The assortativity coefficient is a standard network measure originally defined by Newman (11). The coefficient can range from -1 to 1, where -1 suggests negative assortativity (opposites attract) and 1 implies positive assortativity (like attracts like). With the assortativity coefficient we provide a tolerance

interval for a random network by calculating the range of assortativity values expected from multiple generated random networks. We generated a reference distribution using permutation. Multiple networks (n=1000) were generated with the same topological structure, but with vaccination status (yes/no) permutated randomly amongst the participants The assortativity value for each was then calculated – this provided the range of assortativity values we would expect under random permutation. Similar techniques are outlined by Barclay et al. (12).

## Auto-logistic Regression

The auto-logistic model was used to further investigate the effect of an individual's social connections on their influenza vaccination decision (13). This model allows an individual's vaccination behaviour to be modelled as a function of their demographic information and the behaviour of their neighbours in the social network. The specification of the auto-logistic model is given in Equation 1.

$$For,\ Y_i = \begin{bmatrix} 0 : not\ vaccinated \\ 1 : vaccinated \end{bmatrix}$$

[Equation 1]

$$\log\left(\frac{\mathrm{P}(\mathrm{Y_i}=1|\ \text{all other } \mathrm{Y_i})}{1-\mathrm{P}(\mathrm{Y_i}=1|\ \text{all other } \mathrm{Y_i})}\right) = \alpha + \underline{x'_i}\underline{\beta} + \gamma\sum_{\mathrm{j\sim i}}(\mathrm{Y_j}=1)$$

*Where j ~ i indicates contact between individuals i and j, α indicates the intercept and* $\underline{x_i}$ *is a vector of covariates associated with individual i.*

The parameters β describe how the covariates affect the likelihood of an individual being vaccinated, whilst the parameter γ describes how this likelihood is modified by the behaviour of the individual's social contacts in the network.

In the specification above the network effect (γ) is based on the total number of vaccinated neighbours an individual possess, however, this is highly correlated with the number of neighbours an individual possess. Therefore, the model was re-parameterised so that the network effect (γ) was based on the proportion of an individual's neighbours who were vaccinated, and the total number of neighbours an individual possessed was included as covariate information (see Equation 2).

$$For,\ Y_i = \begin{bmatrix} 0 : not\ vaccinated \\ 1 : vaccinated \end{bmatrix}$$

[Equation 2]

$$\log\left(\frac{P(Y_i = 1 | \text{ all other } Y_i)}{1 - P(Y_i = 1 | \text{ all other } Y_i)}\right) = \alpha + \underline{n'}_i\underline{\beta} + \gamma\left(\frac{\sum_{j\sim i}(Y_j = 1)}{n_i}\right)$$

*Where $n_i$ is the number of neighbours in the individual's immediate network. Covariate information was included as additional $\beta$s.*

To fit the model, we used Monte Carlo likelihood inference (14), using numerical optimisation with initial values of β derived by fitting a standard logistic regression and initial value of γ = 0 (additional details in the Supplementary material). The logistic regression model is a sub-model of the auto-logistic model when γ = 0, which was used to give initial parameter estimates for $\alpha$ and β, but not for formal inference. The logistic regression model can be used to make inferences about a response (y) from covariate information (x). However, standard logistic regression techniques are unable to make inferences based on information from responses (y). This is problematic in cases such

as spatial or network data, in which we might hypothesise that responses are correlated, for example, based on some arbitrary measure of distance. The auto-logistic model specified by Besag (1974) and outlined here is an extension of the logistic regression model, and is able to account for information from responses (y) in the right-hand side of the equation.

Confidence intervals for the parameters were generated from standard errors derived from the hessian matrix. Hypothesis testing was performed using a Wald Test. Parameter estimates $\underline{\hat{\theta}} = \{\hat{\alpha}, \hat{\beta}, \hat{\gamma}\}$ were assumed to follow a multivariate normal distribution $\underline{\hat{\theta}} \sim MVN\ (\theta, V)$, where V is the variance-covariance matrix, derived from the hessian matrix. The vector C was defined as a binary vector, used for parameter testing, which gives $\underline{\varphi} \equiv C\underline{\theta}$ and $\underline{\hat{\varphi}} \sim MVN\ (\varphi, CVC')$. A Wald test was performed using a chi-squared distribution.

The auto-logistic model does not assume that an individual $y_i$ is independent of their neighbour's neighbours. In this model, the individual is conditionally independent of their neighbours (by the inclusion of γ). This is also true for the neighbours of the individual (and so on). Therefore, formally, the model accounts for information from indirect contacts through this mechanism – by accounting for neighbours the model implicitly accounts for information passed from indirect contacts through the network.

## Results

One hundred and thirty-eight of the 200 foundation doctors invited to take part provided complete data (sex, year of training, axis, and vaccination status). Amongst respondents, 100 (72%) were vaccinated (Table 1).

| | | Number Vaccinated | Total | Vaccination coverage (%) |
|---|---|---|---|---|
| **Sex** | **Female** | 51 | 68 | 75.00 |
| | **Male** | 49 | 70 | 70.00 |
| **Year** | **1** | 55 | 76 | 72.37 |
| | **2** | 45 | 62 | 72.58 |
| **Axis** | **East** | 47 | 69 | 68.12 |
| | **West** | 53 | 69 | 76.81 |

**Table 1: Seasonal influenza vaccination uptake by the foundation doctors stratified by their demographic factors.**

Figure 1 shows the foundation doctors' social network, along with their influenza vaccination status (n=138). The assortativity coefficient for the entire FD social network was -0.034 with a tolerance interval of (-0.12, 0.10).

*figure 1 here*

**Figure 1. The foundation doctor social network sociogram for those who returned complete data, dichotomised at >= 4 ("I see this person on most shifts/ 4 or more days a week"), and coloured according to individual vaccination status.**

The social network structure of the foundation doctors varied between geographical areas and year-groups (Figure 2). For example, amongst second-year doctors, the network density is higher in the east than in the west axis, with 223 ties amongst the n=31 doctors in the east axis compared with 73 ties amongst the same number in the west axis.

*figure 2 here*

**Figure 2. A sociogram depicting the foundation doctor network (n=138), coloured by sub-groups: year and axis.**

We first fitted the re-parameterised auto-logistic model without covariates (Equation 2). Figure 3 describes the maximum likelihood surface for auto-logistic model 1, and Monte Carlo log-likelihood functions are shown in Figure Panel 4. The coefficient for γ, the social network parameter, was 0.965, with 95% confidence interval (0.248, 1.682), i.e. a diffusion effect – individuals were more likely to act in agreement with the behaviour of their neighbours (Table 2). However, this effect was somewhat altered by the negative effect from total number of neighbours, which was near to statistical significance. The model-based approach is more efficient than the assortativity coefficient, leading in this instance to a statistically significant departure from γ=0. For Model 1, an additional Wald test was conducted for the null hypothesis: $\beta_1 = \gamma = 0$, which returned a chi-squared value of 7.091, and p-value of 0.029.

We then added covariate effects for year, axis, and sex. The maximal model allowed us to perform Wald tests for the inclusion of each covariate (model 2, Table 2). The covariates did not account for the social network effect on likelihood of vaccination.

*figure 3 here*

**Figure 3. Contour plot showing the likelihood surface for auto-logistic model 1.**

**Table 2. Parameter estimates for the auto-logistic regression models fit using the foundation doctor data.**

| Model | | Parameter Estimate | Standard Error (Hessian derived) | Lower CI (Including MCSE) | Upper CI (Including MCSE) | Chi-squared | P-value |
|---|---|---|---|---|---|---|---|
| **Auto-logistic Model 1 (Equation 2) ML: 107.835** | $\alpha$ (Intercept) | 0.984 | 0.409 | 0.180 | 1.788 | 5.679 | 0.017 |
| | $\beta_1$ (Number of neighbours) | -0.105 | 0.062 | -0.227 | 0.017 | 2.862 | 0.091 |
| | $\gamma$ | 0.965 | 0.365 | 0.248 | 1.682 | 7.051 | 0.008 |
| **Auto-logistic Model 2 (Equation 2) ML: 108.702** | $\alpha$ (Intercept) | 0.933 | 0.509 | -0.064 | 1.930 | 3.362 | 0.067 |
| | $\beta_1$ (1 = Year 2) | -0.132 | 0.385 | -0.886 | 0.622 | 0.118 | 0.732 |
| | $\beta_2$ (1 = West) | 0.295 | 0.375 | -0.440 | 1.030 | 0.618 | 0.432 |
| | $\beta_3$ (1 = female) | 0.103 | 0.402 | -0.685 | 0.891 | 0.066 | 0.798 |
| | $\beta_4$ (Number of neighbours) | -0.100 | 0.066 | -0.229 | 0.029 | 2.315 | 0.128 |
| | $\gamma$ | 0.795 | 0.377 | 0.056 | 1.534 | 4.441 | 0.035 |

*figure 4a here**figure 4b here**figure 4c here*

**Figure 4. Monte Carlo log-likelihood functions for Model 1, $\boldsymbol{\theta_0} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = (\mathbf{0.984}, \ -\mathbf{0.105}, \ \mathbf{0.965})$ and 10000 simulations per log-likelihood evaluation.**

# Discussion

After excluding missing data, the foundation doctors' self-reported vaccination coverage of 72% (100 vaccinated out of 138, with possible range 50% - 81% dependent on the vaccination status of non-respondents), was higher than the national average of 55% (15). The statistical analysis suggests that the individual's social network has potential to exert both positive and negative effects on likelihood to vaccinate. The higher the proportion of vaccinated neighbours in an individual's network the more likely they were to be themselves vaccinated.

The auto-logistic model has allowed us to assess which areas of the population are the less likely to vaccinate, taking into account their social network structure. For example, we hypothesised that year group or axis may affect an individual's likelihood of receiving the vaccination. However, the confidence intervals for all demographic factors in the auto-logistic model included zero. This suggests that the effects of network structure on vaccination cannot be accounted for by the demographic information. Using this statistical modelling approach has provided a better understanding of the social network structure on vaccination uptake than could be obtained using only the assortativity coefficient, both through its greater statistical efficiency and its ability to investigate whether, and if so to what extent, measured covariates can explain the network structure.

Our analysis of the foundation doctor population suggests that as the proportion of neighbour's who vaccinate increases, the individual's likelihood of vaccination increases – similar to the usual diffusion of behaviour observed in social networks (16). However, this may be offset if having more neighbours reduces the individual's probability of

being vaccinated – this effect was close to statistical significant and requires further investigation (Table 2). This suggests that social networks may exert both repulsion and diffusion effects on vaccination behaviours. This combination makes social networks vital to understanding vaccination dynamics within a population.

We observed other differences in the network structure amongst the four sub-groups defined by year and geographical axis. Second year foundation doctors on the west axis of the Trust had a much sparser social network than the other year/axis groups. In sparse social networks the potential for information transfer (behaviour adoption, infectious disease spread, etc) is fundamentally diminished by social distancing (16). However, Shirley et al. suggest that even when network density is equivalent, network topology may still have an effect on diffusion of information (17). The analysis of the FD data suggests that demographic covariates were unable to account for the social network effects on vaccination. However, only a limited number of covariates were available. More research would be needed to identify other factors that may affect the transfer of vaccination attitudes amongst friends. Interventions aimed at improving vaccination uptake need to be sensitive to the differences between sub-groups within the relevant population and may need to be targeted at specific demographic sub-groups. Network effects on behaviour are complex, but the auto-logistic model provides an effective way of assessing behaviour on a real social network in the presence of other variables that affect individuals' responses.

Vaccination is a complex behaviour in which there is a cost to taking the vaccination (pain of injection, perceived side effects, etc.) to be weighed against the benefits of vaccinating (prevention of disease), within a social setting in which individuals both conform/dissent with social norms. It may be the case that the misperceptions

surrounding the influenza vaccination are more commonly discussed than the benefits within this population (18). Vaccinated individuals may be more likely to provide a favourable assessment of the vaccination to their peers. This may have an effect on their neighbour's assessment of the costs/benefits associated with receiving the influenza vaccination. Spread of vaccination information through a network is complex - previous work has shown that sharing factual corrections about controversial issues relating to vaccinations may have the counterintuitive result of decreasing intent to vaccinate (19). It is also possible that individual's with a larger network are more exposed to varying influences regarding vaccination, where negative assessments are given greater weight. The behaviour of others directly affects the individual – if more people are vaccinated the risk of infection is lower for all (20). The data presented here was collected from a workplace environment and explores an occupational social network, which may be formed somewhat artificially; in this case, members of the same social group may have dissimilar demographic characteristics. Better understanding of the role social relationships play in establishing the vaccination behaviour of HCWs in the workplace is necessary to inform vaccination campaigns, whose ultimate goal is to improve occupational health and patient wellbeing.

Similarities or dis-similarities in behaviour between social contacts could arise due to an endogenous effect or an exogenous effect (via correlation or causation) – known as the reflection problem (21). The data presented here are cross-sectional; there is no way to explore how the observed behaviour arose. Simulation studies have suggested that the influence of 'stubborn' individuals (those who do not change their vaccination behaviour) on others in a network greatly depends on their proportion within a

population (22). Future work might include longitudinal studies to explore the mechanisms that lead to observed vaccination behaviour in a social network.

We have outlined a novel methodological approach to understanding behaviour in a network. We also fit the auto-logistic regression model as given in Besag's original specification (without the re-parameterisation of γ, equation 1 above), this produced a negative γ term, suggesting that vaccination likelihood was negatively associated with number of vaccinated neighbours, however this is highly correlated with overall neighbourhood size. The model presented above is better suited to exploring diffusion of behaviour as these two elements (overall neighbourhood size and proportion of vaccinated neighbours) are separated. It is clear that there is much potential for the future use of this class of model, but that it may need adjustments (such as those shown here) to suitably address questions of interest when considering social networks. Furthermore, although the approach has been successful in fitting a parsimonious model to this relatively small dataset, attempts to fit more complex models quickly lead to large standard errors and, consequently, low power to detect more complex network structure.

We dichotomised the social network at level 4, "I see this person on most shifts/4 or more days a week" and above. We assumed that this represented a strong relationship due to the high amount of contact – it also provides an unambiguous definition of close contact. However, this simplification of intensity and direction of social ties is a limitation of this work and one commonplace in social network analyses. A larger dataset would enable more complex models to be fitted and more precise inferences.

An inherent limitation of our data is that they were self-reported, and therefore potentially subject to reporter bias. Given the size of the dataset there is no way to

empirically check for responder bias. We have made the assumption that the data collected was a fair representation of the population. We believe this assumption is plausible. Data collection took place during teaching sessions where a large proportion of the population were expected to attend, irrespective of their vaccination status (the teaching session was not related to influenza vaccination) – and there was no benefit/coercion for individuals to respond positively or otherwise. Future research into this modelling approach should include investigation into the estimation of missing data to allow subjects with partially observed information to be included in the analysis, and to investigate whether non-participation is informative, i.e. non-participants have atypical vaccination behaviour (23).

Making a decision about influenza vaccination is a complex process – many people are neither completely for nor completely against influenza vaccination, and this may not be in alignment with their self-reported vaccination status (24). There may be varying levels of attitudes to vaccination that could be described using an ordinal or continuous scale, rather than as a simple binary variable. Extracting this more nuanced data is a challenge, and requires qualitative methods such as in-depth interviews with participants (24).

Using the auto-logistic modelling approach, we have expanded on the results of the social network analysis. This novel approach to analysing social network data allows us to investigate in more detail the underlying process that has led to an observed network and its vaccination distribution. Quantitative methods that explore social behaviour are likely to become instrumental in defining targeted approaches to improving public health - this study outlines a suitable approach to investigating how an individual's behaviour might be influenced by the behaviour of their neighbours in a network.

Social networks are powerful phenomena that may be harnessed to encourage diffusion of positive health behaviours (21). We have shown that this is particularly relevant in an occupational setting where somewhat artificial social networks are formed with clearly defined boundaries, and knowledge about occupational health is exchanged between workers.

## Acknowledgments

The Authors would like to thank the foundation doctors at the Pennine Acute Hospitals NHS Trust for generously giving up their time to participate in this study and staff who enabled data collection. The Authors would like to thank the reviewers and editors at the BMJ Open, their insightful comments led to substantial improvements to the paper.

## Funding

This work was supported by the Colt Foundation.

## Author Contributions

RE wrote the manuscript with input from all other authors all of which have approved the final version. RE collected the data. PJD advised on the statistical methodology, which was implemented by RE. TJK guided the initial data collection and SN analysis. All authors contributed to the conception of the project. This study is part of a larger programme of work devised by RI.

## Data Availability Statement

Data used in this analysis are freely available from Lancaster University PURE repository DOI: 10.17635/lancaster/researchdata/299.

## Competing Interests Statement

None declared.

# References


1. WHO . Influenza: Fact Sheet [Internet]. 2014. Available from: http://www.who.int/mediacentre/factsheets/fs211/en/

2. General Medical Council . Good Practice Guidelines [Internet]. 2013 [cited 2016 Aug 18]. Available from: http://www.gmc-uk.org/guidance/good_medical_practice/contents.asp

3. Poland GA, Tosh P, Jacobson RM. Requiring influenza vaccination for health care workers: seven truths we must accept. Vaccine. 2005;23(17):2251–5.

4. Anikeeva O, Braunack-Mayer A, Rogers W. Requiring influenza vaccination for health care workers. Am J Public Health. 2009;99(1):24–9.

5. Public Health England. Seasonal flu vaccine uptake in healthcare workers in England: winter season 2016 to 2017 [Internet]. 2017 [cited 2018 Mar 2]. Available from: https://www.gov.uk/government/statistics/seasonal-flu-vaccine-uptake-in-healthcare-workers-in-england-winter-season-2016-to-2017

6. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008;5(3):e74.

7. Scott J, Carrington PJ. The SAGE handbook of social network analysis. SAGE publications; 2011.

8. Edge R, Heath J, Rowlingson B, Keegan TJ, Isba R. Seasonal Influenza Vaccination amongst Medical Students: A Social Network Analysis Based on a Cross-Sectional Study. PloS One. 2015;10(10):e0140085.

9. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med. 2008;358(21):2249–58.

10. Baron G, De Wals P, Milord F. Vaccination practices of Quebec family physicians. Influenza vaccination status and professional practices for influenza vaccination. Can Fam Physician. 2001;47(11):2261–6.

11. Newman ME. Assortative mixing in networks. Phys Rev Lett. 2002;89(20):208701.

12. Barclay VC, Smieszek T, He J, Cao G, Rainey JJ, Gao H, et al. Positive Network Assortativity of Influenza Vaccination at a High School: Implications for Outbreak Risk and Herd Immunity. PloS One. 2014;9(2):e87042.

13. Besag J. Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Ser B Methodol. 1974;192–236.

14. Geyer CJ. Likelihood inference for spatial point processes. Stoch Geom Likelihood Comput. 1999;80:79–140.

15. Public Health England . Seasonal influenza vaccine uptake amongst frontline HCWs in England: winter season 2013 to 2014 [Internet]. Available from: https://www.gov.uk/government/publications/seasonal-influenza-vaccine-uptake-amongst-frontline-hcws-in-england-winter-season-2013-to-2014

16. Valente TW. Social network thresholds in the diffusion of innovations. Soc Netw. 1996 Jan 1;18(1):69–89.

17. Shirley MDF, Rushton SP. The impacts of network topology on disease spread. Ecol Complex. 2005 Sep;2(3):287–99.

18. Hofmann F, Ferracin C, Marsh G, Dumas R. Influenza vaccination of healthcare workers: a literature review of attitudes and beliefs. Infection. 2006;34(3):142–7.

19. Nyhan B, Reifler J. Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. Vaccine. 2015 Jan 9;33(3):459–64.

20. Bauch CT, Earn DJ. Vaccination and the theory of games. Proc Natl Acad Sci U S A. 2004;101(36):13391–4.

21. Manski CF. Identification of Endogenous Social Effects: The Reflection Problem. Rev Econ Stud. 1993;60(3):531–42.

22. Fukuda E, Tanimoto J. Effects of stubborn decision-makers on vaccination and disease propagation in social networks. Int J Autom Logist. 2016 Jan 1;2(1–2):78–92.

23. Rubin DB. Inference and missing data. Biometrika. 1976 Dec 1;63(3):581–92.

24. Edge R, Goodwin D, Isba R, Keegan T. Socialization, Indifference, and Convenience: Exploring the Uptake of Influenza Vaccine Among Medical Students and Early Career Doctors. Qual Health Res. 2017 Jul 24;1049732317720664.

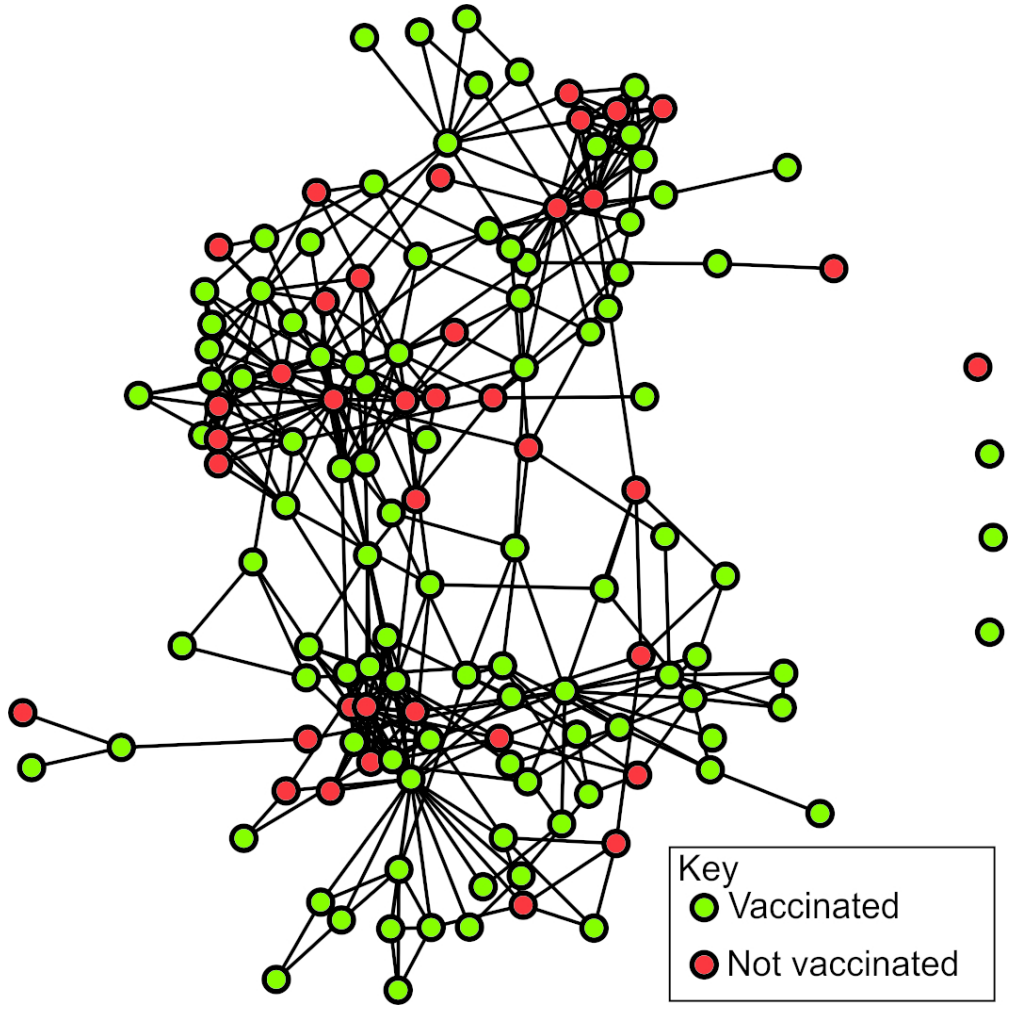



Figure 1. The foundation doctor social network sociogram for those who returned complete data, dichotomised at >= 4 ("I see this person on most shifts/ 4 or more days a week"), and coloured according to individual vaccination status.

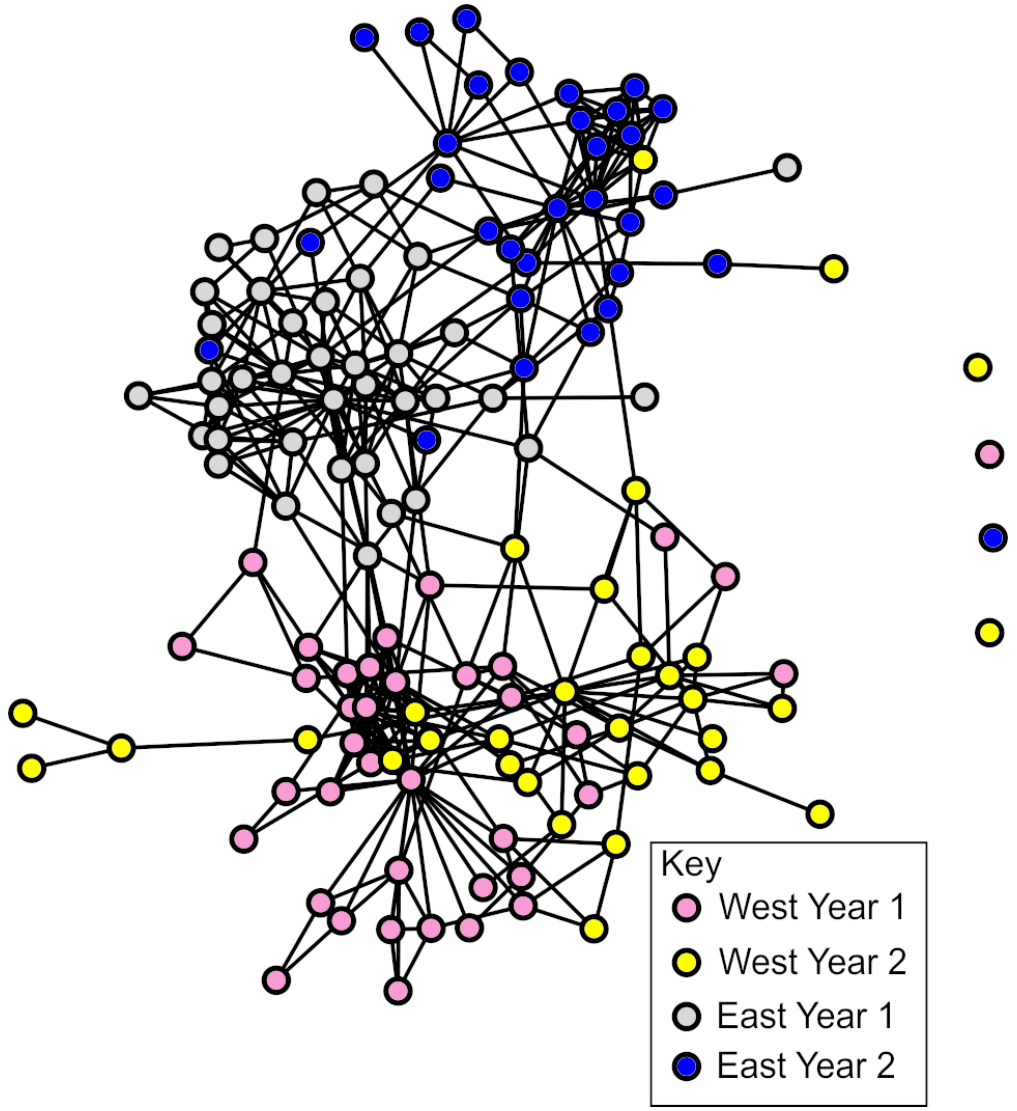



Figure 2. A sociogram depicting the foundation doctor network (n=138), coloured by sub-groups: year and axis.

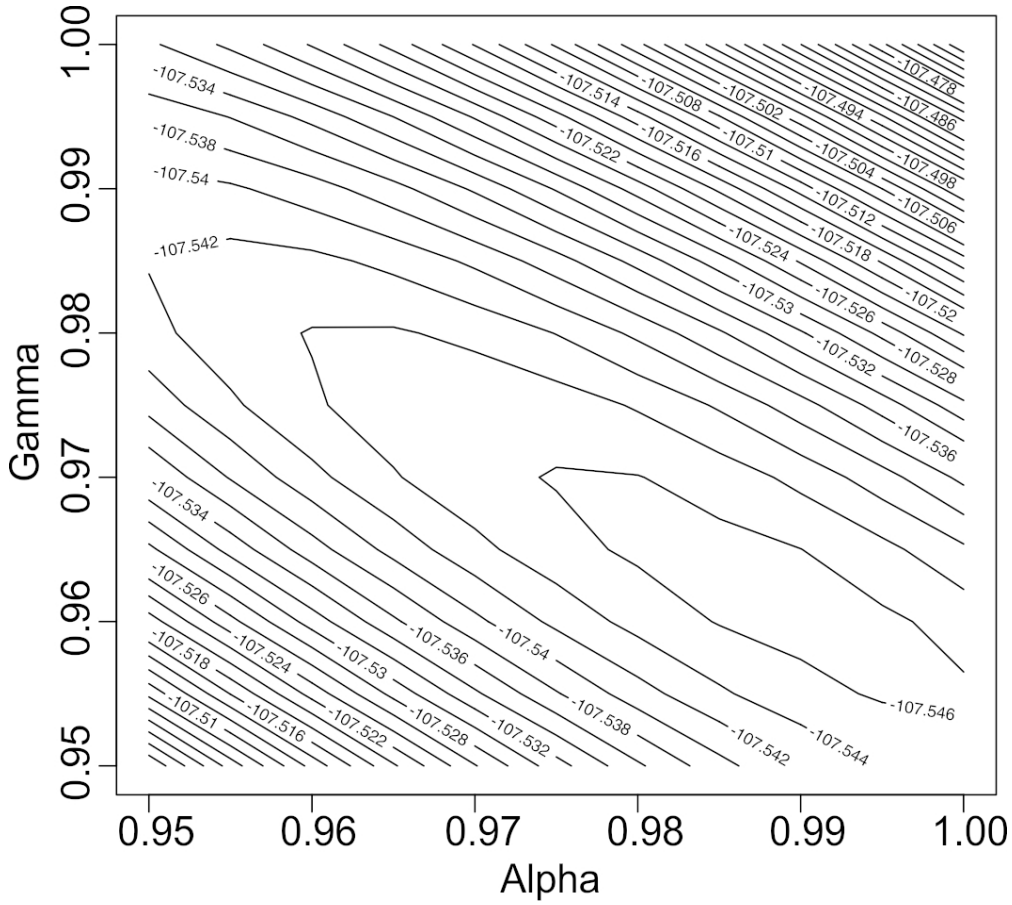



Figure 3. Contour plot showing the likelihood surface for auto-logistic model 1.

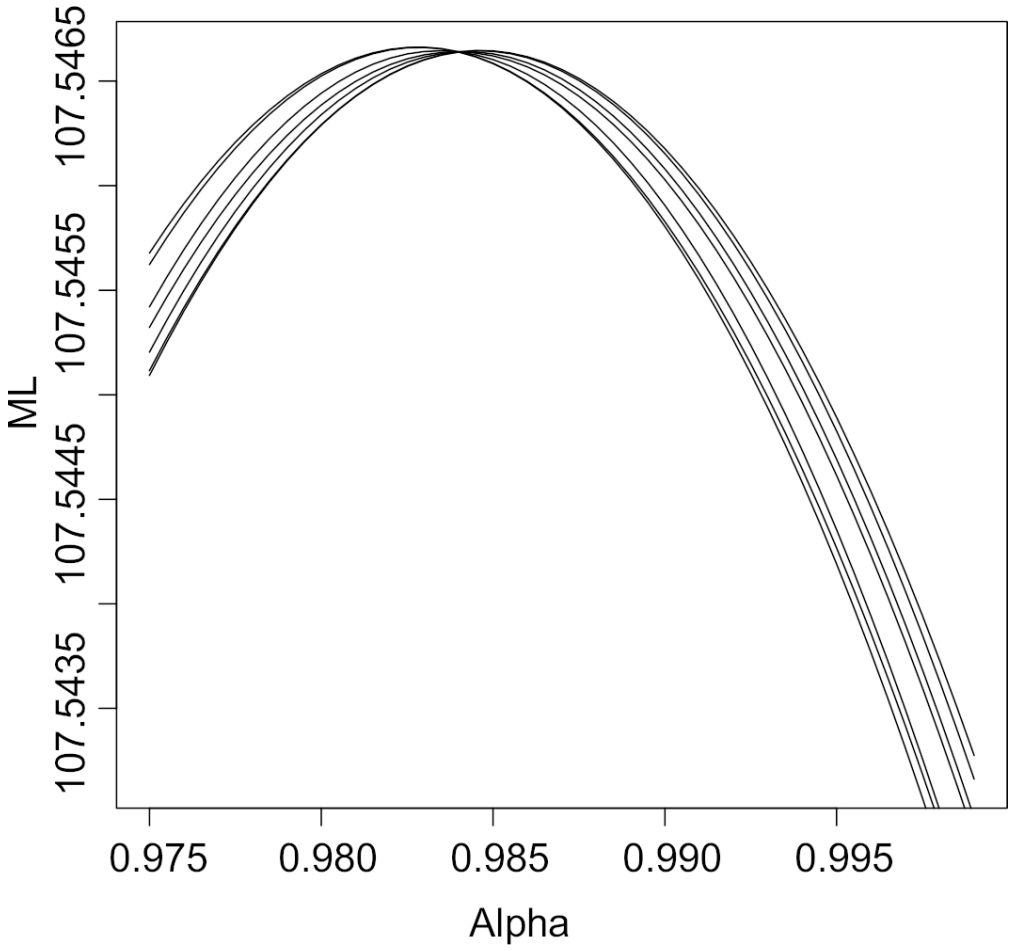



Figure 4. Monte Carlo log-likelihood functions for Model 1, θ_0=(α,β,γ)=(0.984,-0.105,0.965) and 10000 simulations per log-likelihood evaluation (α in figure 4A, β in figure 4B, γ in figure 4C).

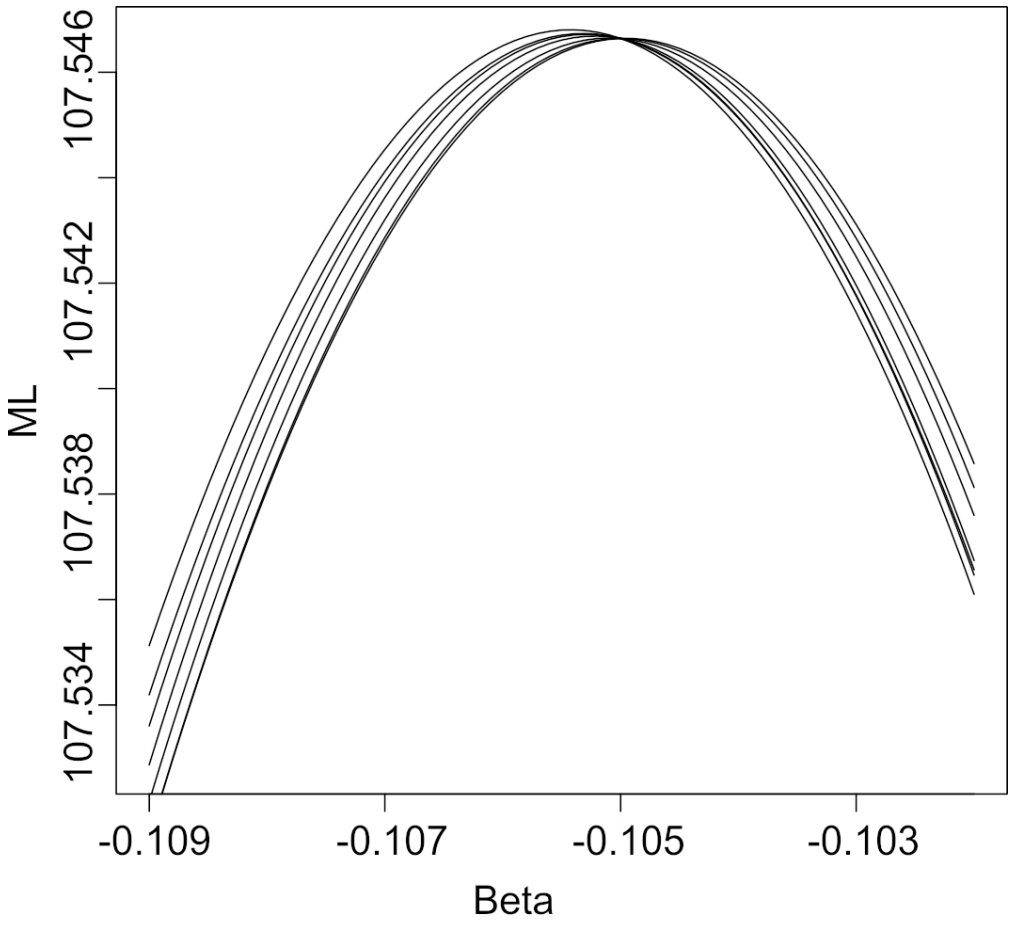



Figure 4. Monte Carlo log-likelihood functions for Model 1, θ_0=(α,β,γ)=(0.984,-0.105,0.965) and 10000 simulations per log-likelihood evaluation.

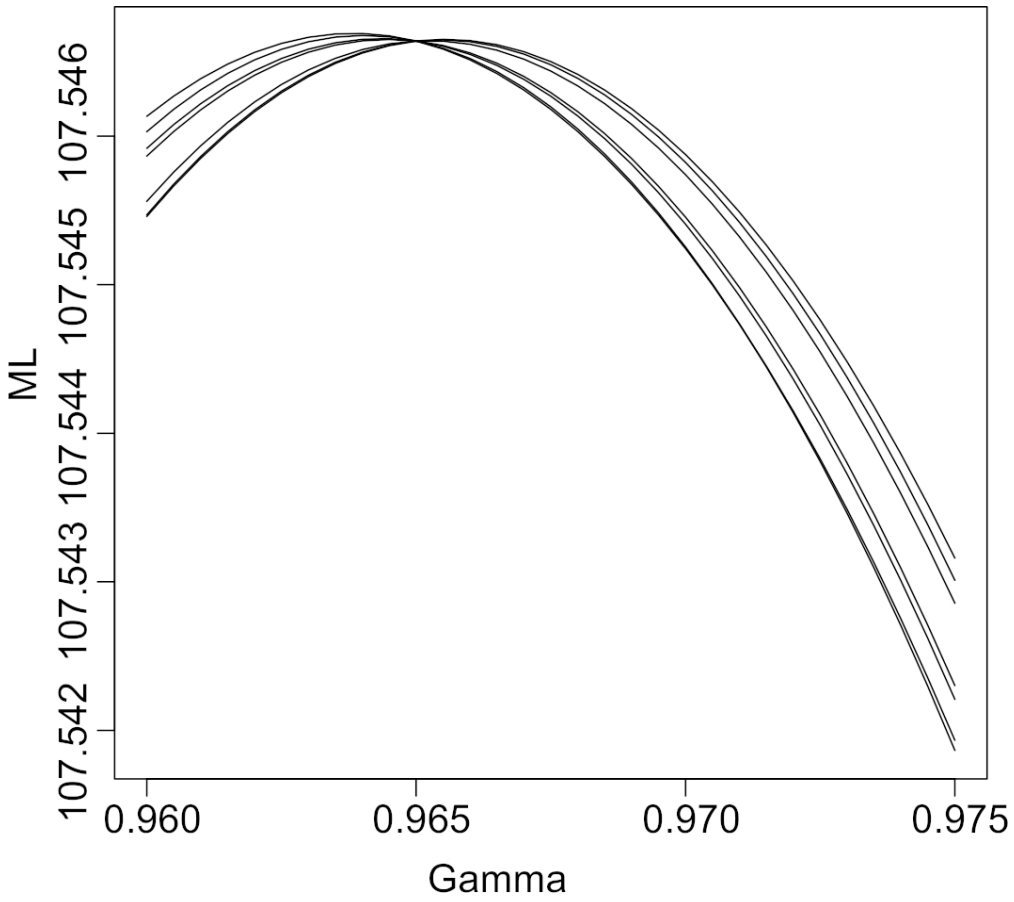



Figure 4. Monte Carlo log-likelihood functions for Model 1, θ_0=(α,β,γ)=(0.984,-0.105,0.965) and 10000 simulations per log-likelihood evaluation.

# Supplementary

# Auto-logistic model parameter estimation

The auto-logistic model incorporates spatial correlation into the logistic model for binary data. The specification is as follows, let Y be our variable of interest, where $Y_i$ $\epsilon(0,1)$ represents the observation at the ith data point for *i = 1…n,* the full conditional distributions are given by:

$$For,\ Y_i = \begin{bmatrix} 0 : not\ vaccinated \\ 1 : vaccinated \end{bmatrix}$$

$$\log\left(\frac{P(Y_i=1|\ \text{all other } Y_i)}{1-P(Y_i=1|\ \text{all other } Y_i)}\right) = \underline{x}'_i\underline{\beta} + \gamma \sum_{j\sim i}(Y_j = 1).$$

From Besag 1974, we have

$$\frac{P(\underline{y})}{P(\underline{0})} <=> f\left(\underline{y}\right) = c(\theta)g(y)$$

Where $c(\theta)$ is an intractable constant and $g(y)$ is a known function. We can manipulate this to use Geyer's method of Monte-Carlo maximum likelihood (17):

$$f\left(\underline{y}\right) = c(\theta)g(y)$$

$$\int f\left(\underline{y}\right) = \int c(\theta)\ g(y;\theta)$$

$$1 = c(\theta) \int g(y;\theta)dy$$

$$1 = \frac{c(\theta)}{c(\theta_0)} \int g(y;\theta) * \frac{g(y;\theta_0)}{g(y;\theta_0)}\ c(\theta_0)\ dy$$

$$= \frac{c(\theta)}{c(\theta_0)} \int \frac{g(y;\theta)}{g(y;\theta_0)}\ f(y;\theta_0)\ dy$$

$$= \frac{c(\theta)}{c(\theta_0)} * E_0\left[\frac{g(y;\theta)}{g(y;\theta_0)}\right] \quad (1)$$

We can simulate the expectation in equation 1 using a Monte-Carlo approximation to the expectation.

$$\approx \frac{1}{\varphi}\sum_{1}^{\varphi}\frac{g(y;\theta)}{g(y;\theta_0)}$$

Rearranging (1) gives:

$$c(\theta) = \frac{c(\theta_0)}{\widehat{E_0}}$$

Thus, we have:

$$f(y;\underline{\theta}) \approx c(\theta)g(y;\theta) = \frac{c(\theta_0)}{\widehat{E_0}} * g(y;\theta)$$

$$L_{mc}(\theta) = \log\big(f(y;\theta)\big) = \log\big(c(\theta_0)\big) - \log\big(\widehat{E_0}\big) + \log(g(y;\theta))$$

The term: $\log\big(c(\theta_0)\big)$ is a constant, therefore, $\hat{\theta}$ maximises the terms: $-\log\big(\widehat{E_0}\big) + \log\big(g(y;\theta)\big)$.

Maximising this gives a Monte-Carlo approximation to the maximum likelihood estimator (MLE). When n is large maximum likelihood estimators have normal properties:

$$\hat{\theta} \approx N(\theta_0, \frac{1}{n\,I(\theta_0)})$$

where I(θ) is the information matrix.

As we are utilising this methodology with network data, it is necessary to check whether the asymptotic principles hold given our sample size (and network structure) – thus, we used simulation studies to investigate the properties of the model.

**Simulation Experiments**

For each experiment using the auto-logistic model, fixed parameter θ, and the network structure from our foundation doctor social network data, we are able to generate data samples, Y.

To explore the behaviour of the auto-logistic model and our implementation, we firstly generate multiple new response data sets, $Y_i$. We then estimate parameters for these realisations using Monte-Carlo maximum likelihood estimation. This results in a set of estimates for θ. Inference on this set allows us to explore the model's behaviour under different conditions i.e. different values of θ. This scheme is outlined graphically by Figure 1. Monte-Carlo maximum likelihood tends to the true values for θ as n tends to infinity. We have a finite value of n, thus we need to check whether we are providing sensible estimates for θ. This is achieved by comparing estimates of $\hat{\theta}$ to known values of θ.

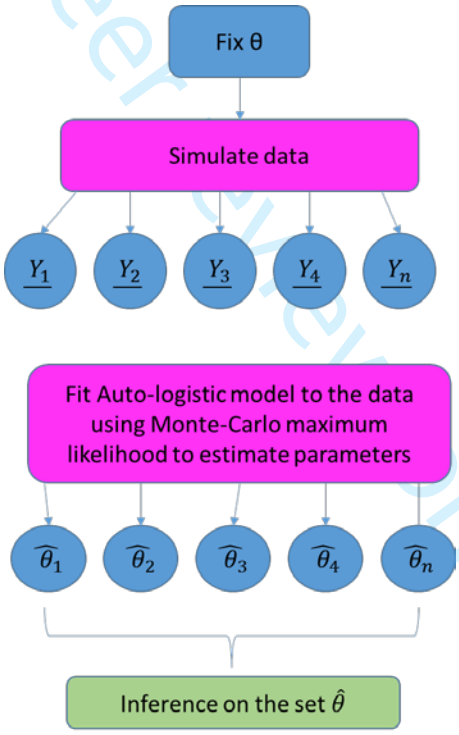



**Figure 1: Regime for the simulation experiments using the auto-logistic model.**

**Table 3. Contingency table of the frequency of foundation doctors by, neighbourhood, year, axis and vaccination status.**

| | | | Axis east | | Axis west | |
|---|---|---|---|---|---|---|
| | | | year 1 | year 2 | year 1 | year 2 |
| **Neighbours vaccinated = 0** | vaccinated | yes | 1 | 1 | 2 | 1 |
| | | no | 0 | 2 | 0 | 1 |
| **Neighbours vaccinated = 1** | vaccinated | yes | 1 | 4 | 2 | 5 |
| | | no | 0 | 0 | 0 | 2 |
| **Neighbours vaccinated = 2** | vaccinated | yes | 4 | 10 | 5 | 2 |
| | | no | 2 | 0 | 0 | 3 |
| **Neighbours vaccinated = 3** | vaccinated | yes | 3 | 4 | 10 | 4 |
| | | no | 2 | 0 | 2 | 2 |
| **Neighbours vaccinated = 4** | vaccinated | yes | 7 | 3 | 6 | 5 |
| | | no | 4 | 3 | 2 | 0 |
| **Neighbours vaccinated = 5** | vaccinated | yes | 3 | 0 | 2 | 1 |
| | | no | 1 | 1 | 0 | 1 |
| **Neighbours vaccinated = 6** | vaccinated | yes | 2 | 0 | 2 | 0 |
| | | no | 4 | 0 | 0 | 0 |
| **Neighbours vaccinated = 7** | vaccinated | yes | 1 | 0 | 1 | 1 |
| | | no | 0 | 0 | 0 | 0 |
| **Neighbours vaccinated = 8** | vaccinated | yes | 1 | 1 | 0 | 1 |
| | | no | 1 | 1 | 0 | 1 |
| **Neighbours vaccinated = 9** | vaccinated | yes | 1 | 0 | 0 | 0 |
| | | no | 3 | 0 | 0 | 0 |
| **Neighbours vaccinated = 10** | vaccinated | yes | 0 | 0 | 0 | 1 |
| | | no | 0 | 0 | 0 | 0 |
| **Neighbours vaccinated = 14** | vaccinated | yes | 0 | 0 | 0 | 0 |
| | | no | 0 | 1 | 0 | 0 |
| **Neighbours vaccinated = 16** | vaccinated | yes | 0 | 0 | 0 | 0 |
| | | no | 1 | 0 | 0 | 0 |

**Table 4. Contingency table showing the spread of the foundation doctor data between year, axis and vaccination status.**

| | | Axis east | | Axis west | |
|---|---|---|---|---|---|
| | | year 1 | year 2 | year 1 | year 2 |
| **vaccinated** | yes | 23 | 24 | 32 | 21 |
| | no | 15 | 7 | 6 | 10 |

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

| | Item No | Recommendation |
|---|---|---|
| **Title and abstract** | 1✓ | (*a*) Indicate the study's design with a commonly used term in the title or the abstract |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found |
| **Introduction** | | |
| Background/rationale | 2✓ | Explain the scientific background and rationale for the investigation being reported |
| Objectives | 3✓ | State specific objectives, including any prespecified hypotheses |
| **Methods** | | |
| Study design | 4✓ | Present key elements of study design early in the paper |
| Setting | 5✓ | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection |
| Participants | 6✓ | (*a*) Give the eligibility criteria, and the sources and methods of selection of participants |
| Variables | 7✓ | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable |
| Data sources/ measurement | 8*✓ | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group |
| Bias | 9✓ | Describe any efforts to address potential sources of bias |
| Study size | 10✓ | Explain how the study size was arrived at |
| Quantitative variables | 11✓ | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why |
| Statistical methods | 12✓ | (*a*) Describe all statistical methods, including those used to control for confounding |
| | | (*b*) Describe any methods used to examine subgroups and interactions |
| | | (*c*) Explain how missing data were addressed |
| | | (*d*) If applicable, describe analytical methods taking account of sampling strategy |
| | | (*e*) Describe any sensitivity analyses |
| **Results** | | |
| Participants | 13*✓ | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed |
| | | (b) Give reasons for non-participation at each stage |
| | | (c) Consider use of a flow diagram |
| Descriptive data | 14*✓ | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders |
| | | (b) Indicate number of participants with missing data for each variable of interest |
| Outcome data | 15*✓ | Report numbers of outcome events or summary measures |
| Main results | 16✓ | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included |
| | | (*b*) Report category boundaries when continuous variables were categorized |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period |
| Other analyses | 17✓ | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses |

| | | |
|---|---|---|
| **Discussion** | | |
| Key results | 18✓ | Summarise key results with reference to study objectives |
| Limitations | 19✓ | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias |
| Interpretation | 20✓ | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence |
| Generalisability | 21✓ | Discuss the generalisability (external validity) of the study results |
| **Other information** | | |
| Funding | 22✓ | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based |

*Give information separately for exposed and unexposed groups.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

# An observational study to assess the effects of social networks on the seasonal influenza vaccine uptake by early career doctors.



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2018-026997.R2 |
| Article Type: | Research |
| Date Submitted by the Author: | 07-Jul-2019 |
| Complete List of Authors: | Edge, Rhiannon; Lancaster University Faculty of Health and Medicine, Lancaster Medical School<br>Keegan, Thomas; Lancaster University, Lancaster Medical School<br>Isba, Rachel; Lancaster University Faculty of Health and Medicine, Lancaster Medical School<br>Diggle, Peter; Lancaster University, Lancaster Medical School; University of Liverpool, Epidemiology and Population Health |
| Primary Subject Heading: | Occupational and environmental medicine |
| Secondary Subject Heading: | Epidemiology, Research methods, Public health |
| Keywords: | Social network analysis, influenza vaccination, auto-logistic regression, occupational health |
| | |

# An observational study to assess the effects of social networks on the seasonal influenza vaccine uptake by early career doctors.

R Edge*, T J Keegan, R Isba, P J Diggle

Lancaster Medical School, Furness Building, Lancaster University, Bailrigg, Lancaster, LA1 4YW.

Lecturer, Senior Lecturer, Professor, Professor.

* Correspondence to: r.edge@lancs.ac.uk.

## Abstract

### Objectives

To evaluate the effect of social network influences on seasonal influenza vaccination uptake by healthcare workers.

### Design

Cross-sectional, observational study.

### Setting

A large secondary care NHS Trust, which includes four hospital sites in Greater Manchester.

### Participants

Foundation doctors (FDs) working at the PAT during the study period. Data collection took place during compulsory weekly teaching sessions there were no exclusions. Of the 200 eligible FDs, 138 (70%) provided complete data.

### Primary outcome measures

Self-reported seasonal influenza vaccination status.

### Results

Amongst participants, 100 (72%) reported that they had received a seasonal influenza vaccination. Statistical modelling demonstrated that having a higher proportion of vaccinated neighbours increased an individual's likelihood of being vaccinated. The coefficient for $\gamma$, the social network parameter, was 0.965 (95% confidence interval: 0.248, 1.682; odds: 2.625 (95% confidence interval: 1.281, 5.376)), i.e. a diffusion effect. Adjusting for year group, geographical area, and sex did not account for this effect.

### Conclusions

This population exhibited higher than expected vaccination coverage levels– providing protection both in the workplace and for vulnerable patients. The modelling approach

allowed covariate effects to be incorporated into social network analysis, which gave us a better understanding of the network structure. These techniques have a range of applications in understanding the role of social networks on health behaviours.

**Key words**

Social network analysis, influenza vaccination, auto-logistic regression, occupational health.

## Strengths and limitations

- This study uses a novel auto-logistic regression approach to understanding the effects of an individual's social network on their vaccination status.
- The auto-logistic regression approach to social network analysis provides a unique quantitative framework for comprehensively understanding social behaviours.
- The application of the study findings may be limited because there are many factors that affect influenza vaccination decisions that could not be captured using the data collection methods.
- Data were self-reported, which may have introduced bias.

## Introduction

Influenza affects millions of people each year - it causes considerable morbidity and is a primary or underlying cause of death for thousands of people worldwide (1). The General Medical Council's (GMC) guidance on Good Medical Practice (2013), advises that healthcare workers (HCWs) in the UK receive immunisation against common serious communicable diseases, such as influenza, in order to protect both patients and colleagues (2). Higher coverage of influenza vaccination within a hospital is believed to reduce patient mortality, staff absences, and potential influenza epidemic size, thus protecting some of those at the greatest risk from influenza (3). Despite this, vaccination rates remain highly variable for HCWs and are below the government target of 75%. In 2016/17, around 63% of healthcare workers in England and Wales recieved a seasonal influenza vaccination (4,5).

There is increasing interest in the effects exerted by social networks on public health (6). A social network is made up of nodes (individuals) connected via ties (relationships) (7). Disease dynamics within a network may be influenced by characteristics such as its density, how the individuals in the network interact, and which individuals are vaccinated against, or susceptible to, the disease. For example, changes in the vaccination status of a few key individuals within a network may have a disproportional impact on disease spread (8). It has been shown that an individual's behaviour may be influenced by their peers – for example, research has found that smokers are more likely to befriend other smokers (9). The grouping of similar individuals within a population, known as homophily, could have a considerable impact on behaviour as well as disease dynamics. For example, if clusters of non-vaccinated

individuals exist within a network, a disease could rapidly spread through these groups, reducing the protective effects exerted by herd immunity.

Healthcare workers' vaccination behaviour may be influenced by the behaviour of their neighbours within their social network. Baron et al suggest that healthcare workers seem to be influenced by their co-workers' vaccination practices (10). In this study, network analysis is used to study the characteristics of a social network of foundation doctors (FDs) - early career doctors in the first two years of postgraduate training in the UK – and related these to the distribution of seasonal influenza vaccination within the same population. This was extended by investigating how the probability of an individual receiving an influenza vaccine was influenced by the behaviour of his/her neighbours in the network.

## Methods

Prospective ethical approval was obtained (15RECNA17) from Lancaster University Research Ethics Committee and the Pennine Acute Hospitals NHS Trust (PAT). Prior to data collection, each participant gave informed consent following a verbal and written explanation of the study. Identifiable data were collected and subsequently anonymised before data entry and analysis, in line with accepted practice in SNA studies of this type.

## Participants

Data were collected during January/February 2015. All foundation doctors (FDs) working at the PAT during that period were invited to participate. The foundation training programme at the PAT runs over two years and across four different hospital sites in Greater Manchester, forming two geographically distinct axes, east and west. As

part of their training, FDs are required to attend compulsory weekly teaching sessions. Data collection took place during several of these sessions to optimise response rates.

All participants will have been offered a free seasonal influenza vaccine before the point of data collection. The PAT actively encourages influenza vaccination for its staff, as does the GMC. Staff are given numerous opportunities to have the vaccine, there are often vaccination points established at mutually convenient locations (hospital entrances, cafeterias, etc.) as well as travelling vaccination nurses who offer the vaccine ward-to-ward. We have assumed that all participants have had ample opportunity to vaccinate however, we have not collected data specifically regarding participant's exposure to seasonal influenza vaccination opportunities.

## Patient and Public Involvement

This study involved early career doctors and no patients were involved. Initial findings were presented at the study setting as part of ongoing work; however, it is likely (due to staff turnover) that many participants will not have had access to the findings of this work prior to its publication.

## Data Collection

Each participant completed a paper-based questionnaire. Participants self-reported their seasonal influenza vaccination status for winter 2014/15, alongside basic demographic information.

Participants were then asked how often they had contact with every other person on the foundation training programme using a six-point scale: 0 - I have never met this

person; 1 - I recognise this person's name but wouldn't see them regularly; 2 - I occasionally see this person for very short periods of time; 3 - I see this person briefly at irregular intervals; 4 - I see this person on most shifts/4 or more days a week; 5 - I see this person on almost every shift for long time periods/live with them.

The relational data were then transferred into a numerical adjacency matrix, from which a network was constructed. Prior to analysis, the data were dichotomised at level 4, "I see this person on most shifts/4 or more days a week" and above, in line with previous research (8). Where one person declared a relationship with another at this level, this was assumed to be reciprocal. There may be cases in which neither person declared any relationship, although one was present, this was treated as missing data and excluded. This produced an un-weighted (relationships were binary) and undirected (reciprocal ties were assumed) network.

## Social Network Analysis

The FDs' influenza vaccination status was evaluated as a node attribute on the social network. Individual-level network characteristics, such as a doctor's degree score (the number of ties an individual possesses), were examined along with global measures such as overall network density, and density in different groups within the network (the number of ties throughout the network in relation to the number of individuals within the network).

The assortativity coefficient was calculated to assess whether or not vaccination status showed homophily within the FD population. The assortativity coefficient is a standard network measure originally defined by Newman (11). The coefficient can range from -1 to 1, where -1 suggests negative assortativity (opposites attract) and 1 implies positive assortativity (like attracts like). With the assortativity coefficient we provide a tolerance

interval for a random network by calculating the range of assortativity values expected from multiple generated random networks. We generated a reference distribution using permutation. Multiple networks (n=1000) were generated with the same topological structure, but with vaccination status (yes/no) permutated randomly amongst the participants The assortativity value for each was then calculated – this provided the range of assortativity values we would expect under random permutation. Similar techniques are outlined by Barclay et al. (12).

## Auto-logistic Regression

The auto-logistic model was used to further investigate the effect of an individual's social connections on their influenza vaccination decision (13). This model allows an individual's vaccination behaviour to be modelled as a function of their demographic information and the behaviour of their neighbours in the social network. The specification of the auto-logistic model is given in Equation 1.

$$For,\ Y_i = \begin{bmatrix} 0 : not\ vaccinated \\ 1 : vaccinated \end{bmatrix}$$

[Equation 1]

$$\log\left(\frac{\mathrm{P}(\mathrm{Y_i}=1|\ \text{all other } \mathrm{Y_i})}{1-\mathrm{P}(\mathrm{Y_i}=1|\ \text{all other } \mathrm{Y_i})}\right) = \alpha + \underline{x'_i}\underline{\beta} + \gamma\sum_{\mathrm{j\sim i}}(\mathrm{Y_j}=1)$$

*Where j ~ i indicates contact between individuals i and j, α indicates the intercept and* $\underline{x_i}$ *is a vector of covariates associated with individual i.*

The parameters β describe how the covariates affect the likelihood of an individual being vaccinated, whilst the parameter γ describes how this likelihood is modified by the behaviour of the individual's social contacts in the network.

In the specification above the network effect (γ) is based on the total number of vaccinated neighbours an individual possess, however, this is highly correlated with the number of neighbours an individual possess. Therefore, the model was re-parameterised so that the network effect (γ) was based on the proportion of an individual's neighbours who were vaccinated, and the total number of neighbours an individual possessed was included as covariate information (see Equation 2).

$$For,\ Y_i = \begin{bmatrix} 0 : not\ vaccinated \\ 1 : vaccinated \end{bmatrix}$$

[Equation 2]

$$\log\left(\frac{P(Y_i = 1 | \text{ all other } Y_i)}{1 - P(Y_i = 1 | \text{ all other } Y_i)}\right) = \alpha + \underline{n'}_i\underline{\beta} + \gamma\left(\frac{\sum_{j\sim i}(Y_j = 1)}{n_i}\right)$$

*Where $n_i$ is the number of neighbours in the individual's immediate network. Covariate information was included as additional $\beta$s.*

To fit the model, we used Monte Carlo likelihood inference (14), using numerical optimisation with initial values of β derived by fitting a standard logistic regression and initial value of γ = 0 (additional details in the Supplementary material). The logistic regression model is a sub-model of the auto-logistic model when γ = 0, which was used to give initial parameter estimates for $\alpha$ and β, but not for formal inference. The logistic regression model can be used to make inferences about a response (y) from covariate information (x). However, standard logistic regression techniques are unable to make inferences based on information from responses (y). This is problematic in cases such

as spatial or network data, in which we might hypothesise that responses are correlated, for example, based on some arbitrary measure of distance. The auto-logistic model specified by Besag (1974) and outlined here is an extension of the logistic regression model, and is able to account for information from responses (y) in the right-hand side of the equation.

Confidence intervals for the parameters were generated from standard errors derived from the hessian matrix. Hypothesis testing was performed using a Wald Test. Parameter estimates $\underline{\hat{\theta}} = \{\hat{\alpha}, \hat{\beta}, \hat{\gamma}\}$ were assumed to follow a multivariate normal distribution $\underline{\hat{\theta}} \sim MVN\ (\theta, V)$, where V is the variance-covariance matrix, derived from the hessian matrix. The vector C was defined as a binary vector, used for parameter testing, which gives $\underline{\varphi} \equiv C\underline{\theta}$ and $\underline{\hat{\varphi}} \sim MVN\ (\varphi, CVC')$. A Wald test was performed using a chi-squared distribution.

The auto-logistic model does not assume that an individual $y_i$ is independent of their neighbour's neighbours. In this model, the individual is conditionally independent of their neighbours (by the inclusion of γ). This is also true for the neighbours of the individual (and so on). Therefore, formally, the model accounts for information from indirect contacts through this mechanism – by accounting for neighbours the model implicitly accounts for information passed from indirect contacts through the network.

## Results

One hundred and thirty-eight of the 200 foundation doctors invited to take part provided complete data (sex, year of training, axis, and vaccination status). Amongst respondents, 100 (72%) were vaccinated (Table 1).

| | | Number Vaccinated | Total | Vaccination coverage (%) |
|---|---|---|---|---|
| **Sex** | **Female** | 51 | 68 | 75.00 |
| | **Male** | 49 | 70 | 70.00 |
| **Year** | **1** | 55 | 76 | 72.37 |
| | **2** | 45 | 62 | 72.58 |
| **Axis** | **East** | 47 | 69 | 68.12 |
| | **West** | 53 | 69 | 76.81 |

**Table 1: Seasonal influenza vaccination uptake by the foundation doctors stratified by their demographic factors.**

Figure 1 shows the foundation doctors' social network, along with their influenza vaccination status (n=138). The assortativity coefficient for the entire FD social network was -0.034 with a tolerance interval of (-0.12, 0.10).

*figure 1 here*

**Figure 1. The foundation doctor social network sociogram for those who returned complete data, dichotomised at >= 4 ("I see this person on most shifts/ 4 or more days a week"), and coloured according to individual vaccination status.**

The social network structure of the foundation doctors varied between geographical areas and year-groups (Figure 2). For example, amongst second-year doctors, the network density is higher in the east than in the west axis, with 223 ties amongst the n=31 doctors in the east axis compared with 73 ties amongst the same number in the west axis.

*figure 2 here*

**Figure 2. A sociogram depicting the foundation doctor network (n=138), coloured by sub-groups: year and axis.**

We first fitted the re-parameterised auto-logistic model without covariates (Equation 2). Figure 3 describes the maximum likelihood surface for auto-logistic model 1, and Monte Carlo log-likelihood functions are shown in Figure Panel 4. The coefficient for γ, the social network parameter, was 0.965, with 95% confidence interval (0.248, 1.682), i.e. a diffusion effect – individuals were more likely to act in agreement with the behaviour of their neighbours (Table 2). However, this effect was somewhat altered by the negative effect from total number of neighbours, which was near to statistical significance. The model-based approach is more efficient than the assortativity coefficient, leading in this instance to a statistically significant departure from γ=0. For Model 1, an additional Wald test was conducted for the null hypothesis: $\beta_1 = \gamma = 0$, which returned a chi-squared value of 7.091, and p-value of 0.029.

We then added covariate effects for year, axis, and sex. The maximal model allowed us to perform Wald tests for the inclusion of each covariate (model 2, Table 2). The covariates did not account for the social network effect on likelihood of vaccination.

*figure 3 here*

**Figure 3. Contour plot showing the likelihood surface for auto-logistic model 1.**

**Table 2. Parameter estimates for the auto-logistic regression models fit using the foundation doctor data.**

| Model | | Parameter Estimate | Standard Error (Hessian derived) | Lower CI (Including MCSE) | Upper CI (Including MCSE) | Chi-squared | P-value |
|---|---|---|---|---|---|---|---|
| **Auto-logistic Model 1 (Equation 2) ML: 107.835** | $\alpha$ (Intercept) | 0.984 | 0.409 | 0.180 | 1.788 | 5.679 | 0.017 |
| | $\beta_1$ (Number of neighbours) | -0.105 | 0.062 | -0.227 | 0.017 | 2.862 | 0.091 |
| | $\gamma$ | 0.965 | 0.365 | 0.248 | 1.682 | 7.051 | 0.008 |
| **Auto-logistic Model 2 (Equation 2) ML: 108.702** | $\alpha$ (Intercept) | 0.933 | 0.509 | -0.064 | 1.930 | 3.362 | 0.067 |
| | $\beta_1$ (1 = Year 2) | -0.132 | 0.385 | -0.886 | 0.622 | 0.118 | 0.732 |
| | $\beta_2$ (1 = West) | 0.295 | 0.375 | -0.440 | 1.030 | 0.618 | 0.432 |
| | $\beta_3$ (1 = female) | 0.103 | 0.402 | -0.685 | 0.891 | 0.066 | 0.798 |
| | $\beta_4$ (Number of neighbours) | -0.100 | 0.066 | -0.229 | 0.029 | 2.315 | 0.128 |
| | $\gamma$ | 0.795 | 0.377 | 0.056 | 1.534 | 4.441 | 0.035 |

*figure 4 here*

**Figure 4. Monte Carlo log-likelihood functions for Model 1, $\boldsymbol{\theta_0} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = (\mathbf{0.984}, \ -\mathbf{0.105}, \ \mathbf{0.965})$ and 10000 simulations per log-likelihood evaluation ($\boldsymbol{\alpha}$ shown in 4A, $\boldsymbol{\beta}$ shown in 4B, $\boldsymbol{\gamma}$ shown in 4C).**

# Discussion

After excluding missing data, the foundation doctors' self-reported vaccination coverage of 72% (100 vaccinated out of 138, with possible range 50% - 81% dependent on the vaccination status of non-respondents), was higher than the national average of 55% (15). The statistical analysis suggests that the individual's social network has potential to exert both positive and negative effects on likelihood to vaccinate. The higher the proportion of vaccinated neighbours in an individual's network the more likely they were to be themselves vaccinated.

The auto-logistic model has allowed us to assess which areas of the population are the less likely to vaccinate, taking into account their social network structure. For example, we hypothesised that year group or axis may affect an individual's likelihood of receiving the vaccination. However, the confidence intervals for all demographic factors in the auto-logistic model included zero. This suggests that the effects of network structure on vaccination cannot be accounted for by the demographic information. Using this statistical modelling approach has provided a better understanding of the social network structure on vaccination uptake than could be obtained using only the assortativity coefficient, both through its greater statistical efficiency and its ability to investigate whether, and if so to what extent, measured covariates can explain the network structure.

Our analysis of the foundation doctor population suggests that as the proportion of neighbour's who vaccinate increases, the individual's likelihood of vaccination increases – similar to the usual diffusion of behaviour observed in social networks (16). However, this may be offset if having more neighbours reduces the individual's probability of

being vaccinated – this effect was close to statistical significant and requires further investigation (Table 2). This suggests that social networks may exert both repulsion and diffusion effects on vaccination behaviours. This combination makes social networks vital to understanding vaccination dynamics within a population.

We observed other differences in the network structure amongst the four sub-groups defined by year and geographical axis. Second year foundation doctors on the west axis of the Trust had a much sparser social network than the other year/axis groups. In sparse social networks the potential for information transfer (behaviour adoption, infectious disease spread, etc) is fundamentally diminished by social distancing (16). However, Shirley et al. suggest that even when network density is equivalent, network topology may still have an effect on diffusion of information (17). The analysis of the FD data suggests that demographic covariates were unable to account for the social network effects on vaccination. However, only a limited number of covariates were available. More research would be needed to identify other factors that may affect the transfer of vaccination attitudes amongst friends. Interventions aimed at improving vaccination uptake need to be sensitive to the differences between sub-groups within the relevant population and may need to be targeted at specific demographic sub-groups. Network effects on behaviour are complex, but the auto-logistic model provides an effective way of assessing behaviour on a real social network in the presence of other variables that affect individuals' responses.

Vaccination is a complex behaviour in which there is a cost to taking the vaccination (pain of injection, perceived side effects, etc.) to be weighed against the benefits of vaccinating (prevention of disease), within a social setting in which individuals both conform/dissent with social norms. It may be the case that the misperceptions

surrounding the influenza vaccination are more commonly discussed than the benefits within this population (18). Vaccinated individuals may be more likely to provide a favourable assessment of the vaccination to their peers. This may have an effect on their neighbour's assessment of the costs/benefits associated with receiving the influenza vaccination. Spread of vaccination information through a network is complex - previous work has shown that sharing factual corrections about controversial issues relating to vaccinations may have the counterintuitive result of decreasing intent to vaccinate (19). It is also possible that individual's with a larger network are more exposed to varying influences regarding vaccination, where negative assessments are given greater weight. The behaviour of others directly affects the individual – if more people are vaccinated the risk of infection is lower for all (20). The data presented here was collected from a workplace environment and explores an occupational social network, which may be formed somewhat artificially; in this case, members of the same social group may have dissimilar demographic characteristics. Better understanding of the role social relationships play in establishing the vaccination behaviour of HCWs in the workplace is necessary to inform vaccination campaigns, whose ultimate goal is to improve occupational health and patient wellbeing.

Similarities or dis-similarities in behaviour between social contacts could arise due to an endogenous effect or an exogenous effect (via correlation or causation) – known as the reflection problem (21). The data presented here are cross-sectional; there is no way to explore how the observed behaviour arose - the direction of the causal relationship between social networks and vaccination status cannot be determined. The casual relationship may be explored using longitudinal data. Simulation studies have suggested that the influence of 'stubborn' individuals (those who do not change their

vaccination behaviour) on others in a network greatly depends on their proportion within a population (22). Future work might include longitudinal studies to explore the mechanisms that lead to observed vaccination behaviour in a social network.

We have outlined a novel methodological approach to understanding behaviour in a network. We also fit the auto-logistic regression model as given in Besag's original specification (without the re-parameterisation of γ, equation 1 above), this produced a negative γ term, suggesting that vaccination likelihood was negatively associated with number of vaccinated neighbours, however this is highly correlated with overall neighbourhood size. The model presented above is better suited to exploring diffusion of behaviour as these two elements (overall neighbourhood size and proportion of vaccinated neighbours) are separated. It is clear that there is much potential for the future use of this class of model, but that it may need adjustments (such as those shown here) to suitably address questions of interest when considering social networks. Furthermore, although the approach has been successful in fitting a parsimonious model to this relatively small dataset, attempts to fit more complex models quickly lead to large standard errors and, consequently, low power to detect more complex network structure.

We dichotomised the social network at level 4, "I see this person on most shifts/4 or more days a week" and above. We assumed that this represented a strong relationship due to the high amount of contact – it also provides an unambiguous definition of close contact. However, this simplification of intensity and direction of social ties is a limitation of this work and one commonplace in social network analyses. A larger dataset would enable more complex models to be fitted and more precise inferences.

An inherent limitation of our data is that they were self-reported, and therefore potentially subject to reporter bias. Given the size of the dataset there is no way to empirically check for responder bias. We have made the assumption that the data collected was a fair representation of the population. We believe this assumption is plausible. Data collection took place during teaching sessions where a large proportion of the population were expected to attend, irrespective of their vaccination status (the teaching session was not related to influenza vaccination) – and there was no benefit/coercion for individuals to respond positively or otherwise. Future research into this modelling approach should include investigation into the estimation of missing data to allow subjects with partially observed information to be included in the analysis, and to investigate whether non-participation is informative, i.e. non-participants have atypical vaccination behaviour (23).

Making a decision about influenza vaccination is a complex process – many people are neither completely for nor completely against influenza vaccination, and this may not be in alignment with their self-reported vaccination status (24). There may be varying levels of attitudes to vaccination that could be described using an ordinal or continuous scale, rather than as a simple binary variable. Extracting this more nuanced data is a challenge, and requires qualitative methods such as in-depth interviews with participants (24).

Using the auto-logistic modelling approach, we have expanded on the results of the social network analysis. This novel approach to analysing social network data allows us to investigate in more detail the underlying process that has led to an observed network and its vaccination distribution. Quantitative methods that explore social behaviour are likely to become instrumental in defining targeted approaches to improving public

health - this study outlines a suitable approach to investigating how an individual's behaviour might be influenced by the behaviour of their neighbours in a network.

Social networks are powerful phenomena that may be harnessed to encourage diffusion of positive health behaviours (21). We have shown that this is particularly relevant in an occupational setting where somewhat artificial social networks are formed with clearly defined boundaries, and knowledge about occupational health is exchanged between workers.

## Acknowledgments

The Authors would like to thank the foundation doctors at the Pennine Acute Hospitals NHS Trust for generously giving up their time to participate in this study and staff who enabled data collection. The Authors would like to thank the reviewers and editors at the BMJ Open, their insightful comments led to substantial improvements to the paper.

## Funding

This work was supported by the Colt Foundation.

## Author Contributions

RE wrote the manuscript with input from all other authors all of which have approved the final version. RE collected the data. PJD advised on the statistical methodology, which was implemented by RE. TJK guided the initial data collection and SN analysis. All authors contributed to the conception of the project. This study is part of a larger programme of work devised by RI.

## Data Availability Statement

Data used in this analysis are freely available from Lancaster University PURE repository DOI: 10.17635/lancaster/researchdata/299.

## Competing Interests Statement

None declared.

# References


1. WHO . Influenza: Fact Sheet [Internet]. 2014. Available from: http://www.who.int/mediacentre/factsheets/fs211/en/

2. General Medical Council . Good Practice Guidelines [Internet]. 2013 [cited 2016 Aug 18]. Available from: http://www.gmc-uk.org/guidance/good_medical_practice/contents.asp

3. Poland GA, Tosh P, Jacobson RM. Requiring influenza vaccination for health care workers: seven truths we must accept. Vaccine. 2005;23(17):2251–5.

4. Anikeeva O, Braunack-Mayer A, Rogers W. Requiring influenza vaccination for health care workers. Am J Public Health. 2009;99(1):24–9.

5. Public Health England. Seasonal flu vaccine uptake in healthcare workers in England: winter season 2016 to 2017 [Internet]. 2017 [cited 2018 Mar 2]. Available from: https://www.gov.uk/government/statistics/seasonal-flu-vaccine-uptake-in-healthcare-workers-in-england-winter-season-2016-to-2017

6. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008;5(3):e74.

7. Scott J, Carrington PJ. The SAGE handbook of social network analysis. SAGE publications; 2011.

8. Edge R, Heath J, Rowlingson B, Keegan TJ, Isba R. Seasonal Influenza Vaccination amongst Medical Students: A Social Network Analysis Based on a Cross-Sectional Study. PloS One. 2015;10(10):e0140085.

9. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med. 2008;358(21):2249–58.

10. Baron G, De Wals P, Milord F. Vaccination practices of Quebec family physicians. Influenza vaccination status and professional practices for influenza vaccination. Can Fam Physician. 2001;47(11):2261–6.

11. Newman ME. Assortative mixing in networks. Phys Rev Lett. 2002;89(20):208701.

12. Barclay VC, Smieszek T, He J, Cao G, Rainey JJ, Gao H, et al. Positive Network Assortativity of Influenza Vaccination at a High School: Implications for Outbreak Risk and Herd Immunity. PloS One. 2014;9(2):e87042.

13. Besag J. Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Ser B Methodol. 1974;192–236.

14. Geyer CJ. Likelihood inference for spatial point processes. Stoch Geom Likelihood Comput. 1999;80:79–140.

15. Public Health England . Seasonal influenza vaccine uptake amongst frontline HCWs in England: winter season 2013 to 2014 [Internet]. Available from: https://www.gov.uk/government/publications/seasonal-influenza-vaccine-uptake-amongst-frontline-hcws-in-england-winter-season-2013-to-2014

16. Valente TW. Social network thresholds in the diffusion of innovations. Soc Netw. 1996 Jan 1;18(1):69–89.

17. Shirley MDF, Rushton SP. The impacts of network topology on disease spread. Ecol Complex. 2005 Sep;2(3):287–99.

18. Hofmann F, Ferracin C, Marsh G, Dumas R. Influenza vaccination of healthcare workers: a literature review of attitudes and beliefs. Infection. 2006;34(3):142–7.

19. Nyhan B, Reifler J. Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. Vaccine. 2015 Jan 9;33(3):459–64.

20. Bauch CT, Earn DJ. Vaccination and the theory of games. Proc Natl Acad Sci U S A. 2004;101(36):13391–4.

21. Manski CF. Identification of Endogenous Social Effects: The Reflection Problem. Rev Econ Stud. 1993;60(3):531–42.

22. Fukuda E, Tanimoto J. Effects of stubborn decision-makers on vaccination and disease propagation in social networks. Int J Autom Logist. 2016 Jan 1;2(1–2):78–92.

23. Rubin DB. Inference and missing data. Biometrika. 1976 Dec 1;63(3):581–92.

24. Edge R, Goodwin D, Isba R, Keegan T. Socialization, Indifference, and Convenience: Exploring the Uptake of Influenza Vaccine Among Medical Students and Early Career Doctors. Qual Health Res. 2017 Jul 24;1049732317720664.

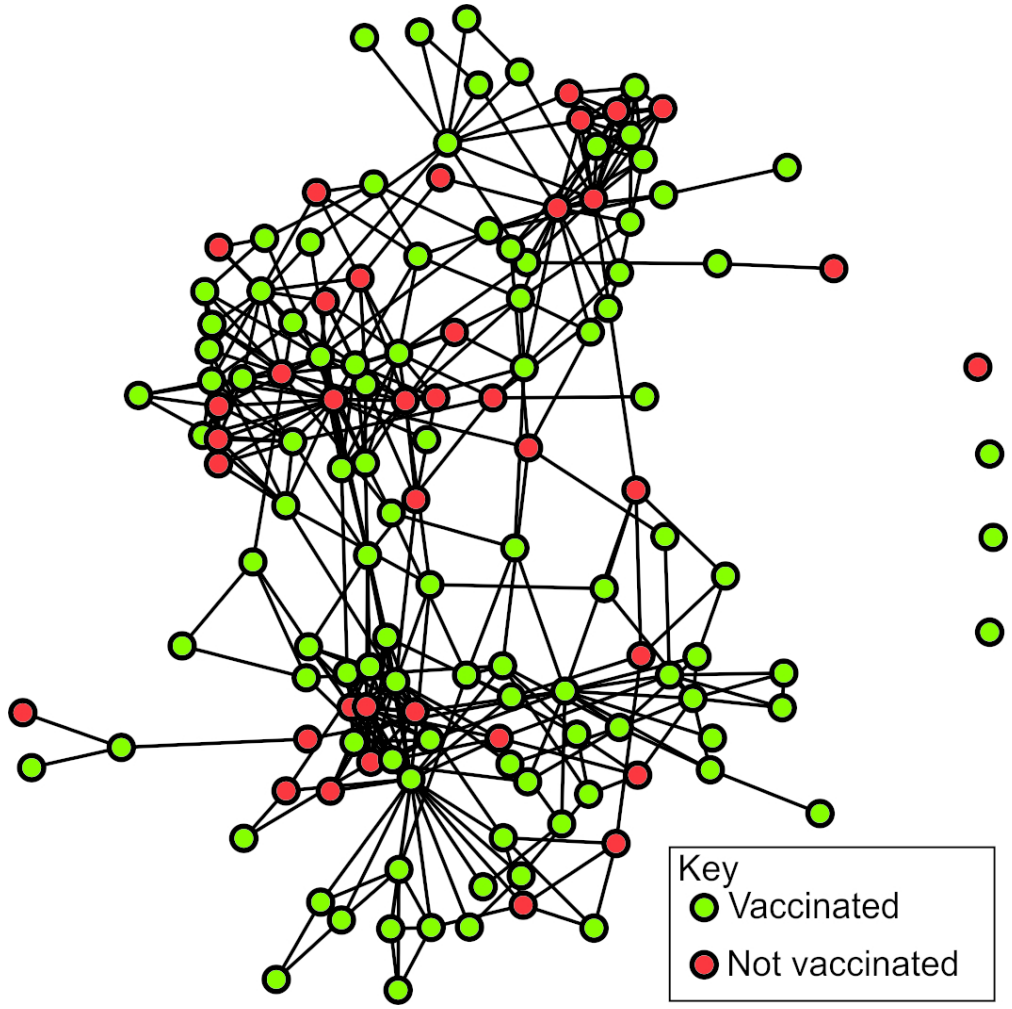



Figure 1. The foundation doctor social network sociogram for those who returned complete data, dichotomised at >= 4 ("I see this person on most shifts/ 4 or more days a week"), and coloured according to individual vaccination status.

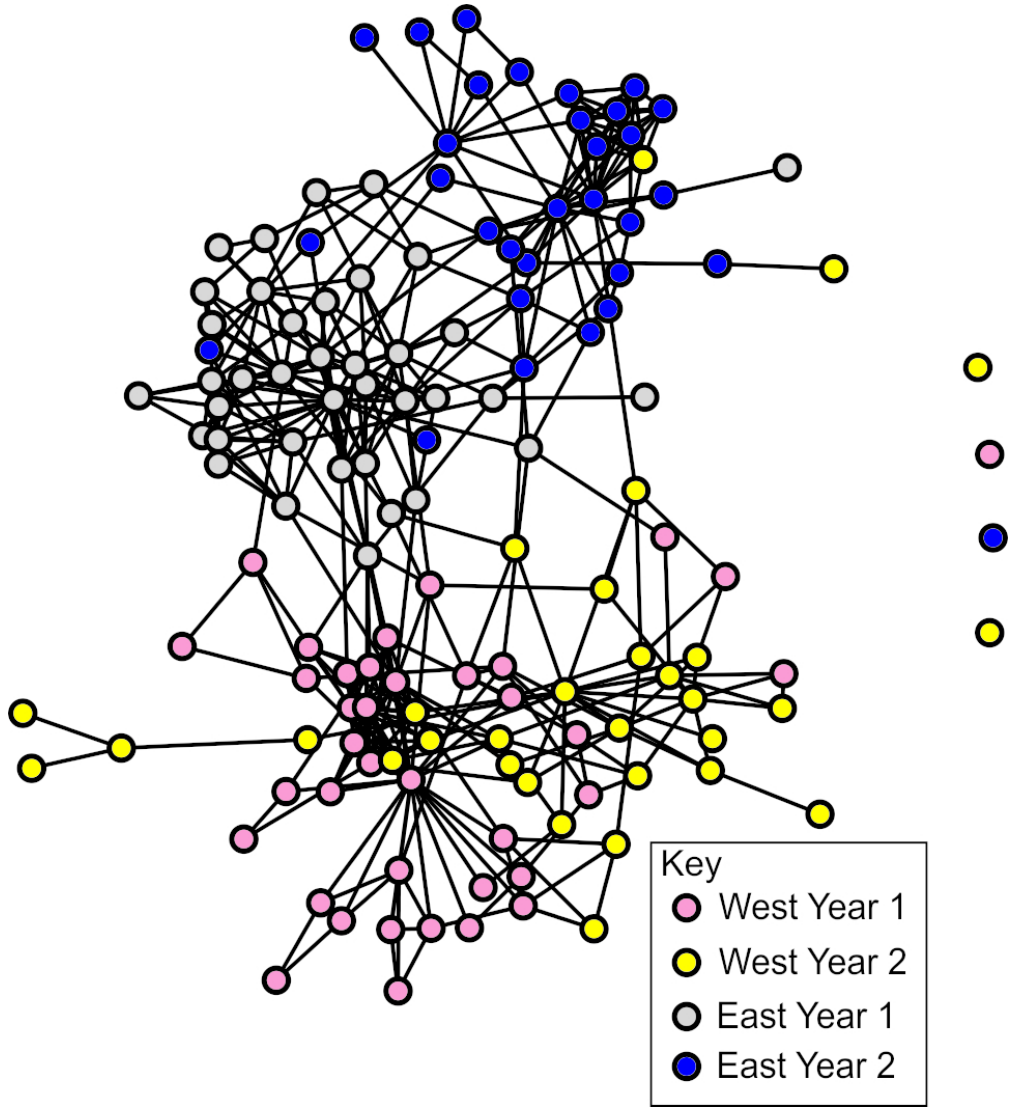



Figure 2. A sociogram depicting the foundation doctor network (n=138), coloured by sub-groups: year and axis.

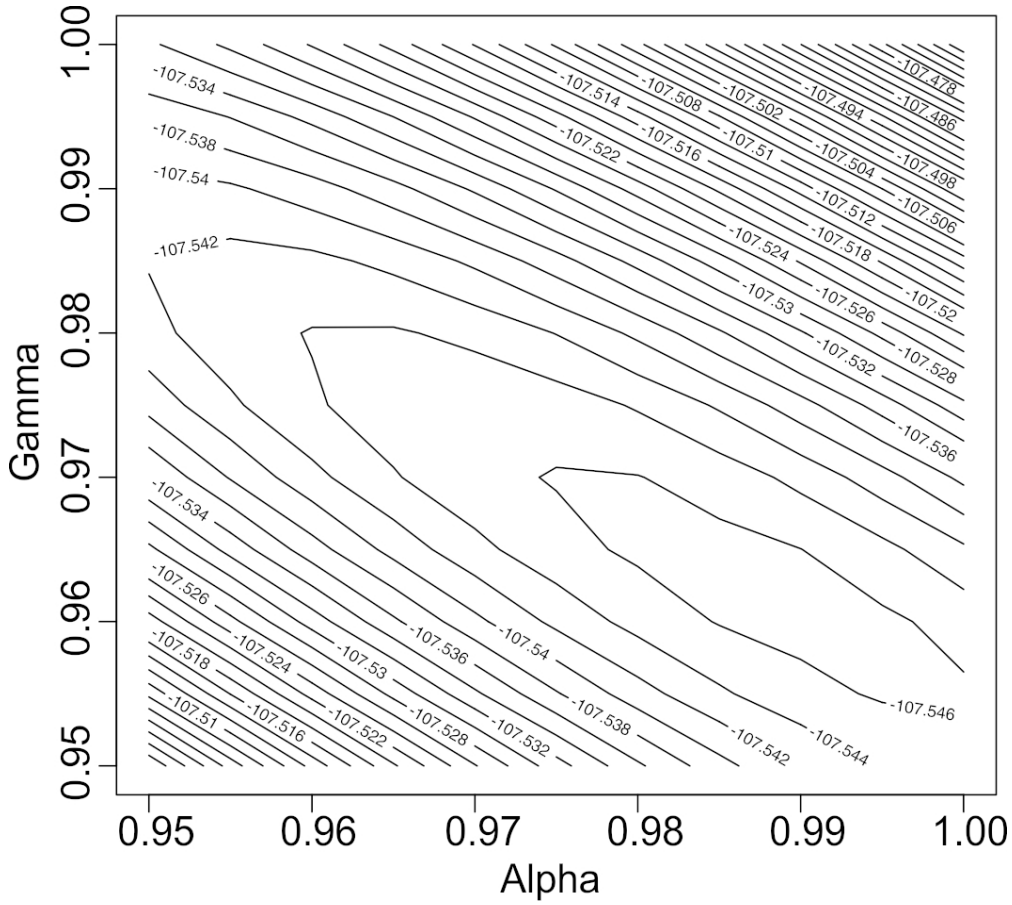



Figure 3. Contour plot showing the likelihood surface for auto-logistic model 1.

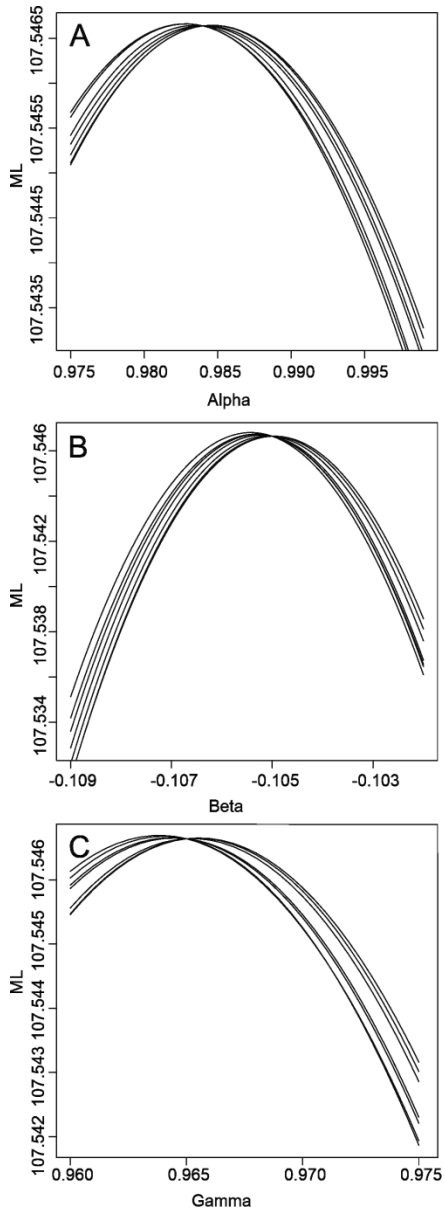



Figure 4. Monte Carlo log-likelihood functions for Model 1, θ_0=(α,β,γ)=(0.984,-0.105,0.965) and 10000 simulations per log-likelihood evaluation (α shown in 4A, β shown in 4B, γ shown in 4C).

# Supplementary

# Auto-logistic model parameter estimation

The auto-logistic model incorporates spatial correlation into the logistic model for binary data. The specification is as follows, let Y be our variable of interest, where Y $\epsilon(0,1)$ represents the observation at the ith data point for *i = 1…n,* the full ***conditional*** distributions are given by:

*For,* $Y_i$ $= \begin{cases} 0: \\ 1: \end{cases}$

$$\frac{(\ -(Y = |\ \ \ \ \ Y)\ ) = \ '\ \ \ + \Sigma\ \ (Y = \ \ \ )}{\begin{matrix} P_{\ldots} \ \ \ \

$$= \quad (\theta) \quad \cdot 0 \, \blacksquare \quad (\theta) \quad \blacksquare \quad (1)$$

We can simulate the expectation in equation 1 using a Monte-Carlo approximation to the expectation.

$$\approx \frac{1}{\blacksquare} \quad (\ ;\ )$$

$$1 \qquad (\theta)$$

Rearranging (1) gives:

$$(\ ) = \frac{(\ 0)}{\blacksquare}$$

Thus, we have:

$$\blacksquare \quad ; \blacksquare \approx (\ )\ (\ ;\ ) = \frac{(\ 0)}{\blacksquare} \qquad *(;)$$

$$0 \qquad \blacksquare$$

$$(\ ) = \log\blacksquare \quad (\ ;\ )\blacksquare = \log\blacksquare \quad (\ )\blacksquare \qquad - \log\blacksquare \quad \blacksquare + \log(\ (\ ;\ ))$$

The term: $\log\blacksquare \quad (\ )\blacksquare$ is a constant, therefore, maximises the terms: $- \log\blacksquare \quad \blacksquare +$ $\log\blacksquare \quad (\ ;\ )\blacksquare$.

Maximising this gives a Monte-Carlo approximation to the maximum likelihood estimator (MLE). When n is large maximum likelihood estimators have normal properties:

$$\blacksquare \qquad 0 \ \frac{1}{\qquad}$$

where I(θ) is the information matrix.
As we are utilising this methodology with network data, it is necessary to check whether the asymptotic principles hold given our sample size (and network structure) – thus, we used simulation studies to investigate the properties of the model.

**Simulation Experiments**

For each experiment using the auto-logistic model, fixed parameter θ, and the network structure from our foundation doctor social network data, we are able to generate data samples, Y.

To explore the behaviour of the auto-logistic model and our implementation, we firstly generate multiple new response data sets, Y . We then estimate parameters for these realisations using Monte-Carlo maximum likelihood estimation. This results in a set of estimates for θ. Inference on this set allows us to explore the model's behaviour under different conditions i.e. different values of θ. This scheme is outlined graphically by Figure 1. Monte-Carlo maximum likelihood tends to the true values for θ as n tends to infinity. We have a finite value of n, thus we need to check whether we are providing sensible estimates for θ. This is achieved by comparing estimates of  to known values of θ.

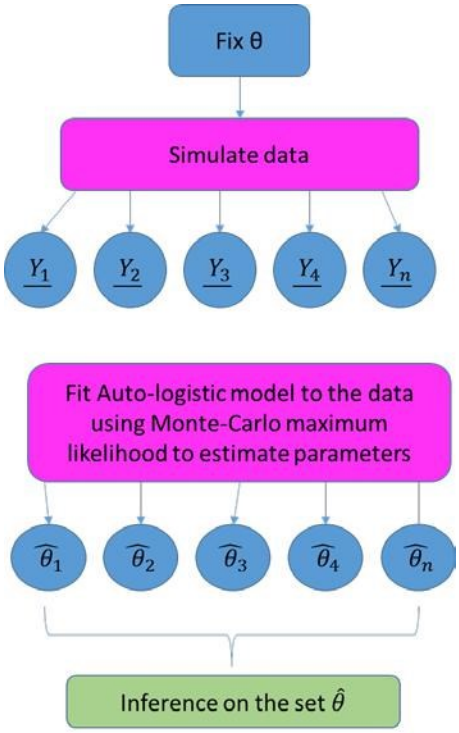



**Figure 1: Regime for the simulation experiments using the auto-logistic model.**

**Table 1. Contingency table of the frequency of foundation doctors by, neighbourhood, year, axis and vaccination status.**

| | | | year | year | year | year |
|---|---|---|---|---|---|---|
| | | | **Axis east** | | **Axis west** | |
| **vaccinated = 0** | vaccinated | yes | 1 | 1 | 2 | 1 |
| **Neighbours** | | no | 0 | 2 | 0 | 1 |
| **Neighbours** | | | 1 | 2 | 1 | 2 |
| **vaccinated = 1** | vaccinated | yes | 1 | 4 | 2 | 5 |
| **Neighbours** | | no | 0 | 0 | 0 | 2 |
| **vaccinated = 2** | vaccinated | yes | 4 | 10 | 5 | 2 |
| **Neighbours** | | no | 2 | 0 | 0 | 3 |
| **vaccinated = 3** | vaccinated | yes | 3 | 4 | 10 | 4 |
| **Neighbours** | | no | 2 | 0 | 2 | 2 |
| **vaccinated = 4** | vaccinated | yes | 7 | 3 | 6 | 5 |
| **Neighbours** | | no | 4 | 3 | 2 | 0 |
| **vaccinated = 5** | vaccinated | yes | 3 | 0 | 2 | 1 |
| **Neighbours** | | no | 1 | 1 | 0 | 1 |
| **vaccinated = 6** | vaccinated | yes | 2 | 0 | 2 | 0 |
| **Neighbours** | | no | 4 | 0 | 0 | 0 |
| **vaccinated = 7** | vaccinated | yes | 1 | 0 | 1 | 1 |
| **Neighbours** | | no | 0 | 0 | 0 | 0 |
| **vaccinated = 8** | vaccinated | yes | 1 | 1 | 0 | 1 |
| **Neighbours** | | no | 1 | 1 | 0 | 1 |
| **vaccinated = 9** | vaccinated | yes | 1 | 0 | 0 | 0 |
| **Neighbours** | | no | 3 | 0 | 0 | 0 |
| **vaccinated = 10** | vaccinated | yes | 0 | 0 | 0 | 1 |
| **Neighbours** | | no | 0 | 0 | 0 | 0 |
| **vaccinated = 14** | vaccinated | yes | 0 | 0 | 0 | 0 |
| **Neighbours** | | no | 0 | 1 | 0 | 0 |
| **vaccinated = 16** | vaccinated | yes | 0 | 0 | 0 | 0 |
| | | no | 1 | 0 | 0 | 0 |

**Table 2. Contingency table showing the spread of the foundation doctor data between year, axis and vaccination status.**

| | | Axis east | | Axis west | |
|---|---|---|---|---|---|
| **vaccinated** | | | | | |
| | | | | | |

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

| | Item No | Recommendation |
|---|---|---|
| **Title and abstract** | 1✓ | (*a*) Indicate the study's design with a commonly used term in the title or the abstract |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found |
| **Introduction** | | |
| Background/rationale | 2✓ | Explain the scientific background and rationale for the investigation being reported |
| Objectives | 3✓ | State specific objectives, including any prespecified hypotheses |
| **Methods** | | |
| Study design | 4✓ | Present key elements of study design early in the paper |
| Setting | 5✓ | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection |
| Participants | 6✓ | (*a*) Give the eligibility criteria, and the sources and methods of selection of participants |
| Variables | 7✓ | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable |
| Data sources/ measurement | 8*✓ | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group |
| Bias | 9✓ | Describe any efforts to address potential sources of bias |
| Study size | 10✓ | Explain how the study size was arrived at |
| Quantitative variables | 11✓ | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why |
| Statistical methods | 12✓ | (*a*) Describe all statistical methods, including those used to control for confounding |
| | | (*b*) Describe any methods used to examine subgroups and interactions |
| | | (*c*) Explain how missing data were addressed |
| | | (*d*) If applicable, describe analytical methods taking account of sampling strategy |
| | | (*e*) Describe any sensitivity analyses |
| **Results** | | |
| Participants | 13*✓ | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed |
| | | (b) Give reasons for non-participation at each stage |
| | | (c) Consider use of a flow diagram |
| Descriptive data | 14*✓ | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders |
| | | (b) Indicate number of participants with missing data for each variable of interest |
| Outcome data | 15*✓ | Report numbers of outcome events or summary measures |
| Main results | 16✓ | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included |
| | | (*b*) Report category boundaries when continuous variables were categorized |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period |
| Other analyses | 17✓ | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses |

| | | |
|---|---|---|
| **Discussion** | | |
| Key results | 18✓ | Summarise key results with reference to study objectives |
| Limitations | 19✓ | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias |
| Interpretation | 20✓ | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence |
| Generalisability | 21✓ | Discuss the generalisability (external validity) of the study results |
| **Other information** | | |
| Funding | 22✓ | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based |

*Give information separately for exposed and unexposed groups.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.