Household structure and infectious disease transmission

T. HOUSE* AND M. J. KEELING

Department of Biological Sciences, University of Warwick, Coventry, UK

SUMMARY

One of the central tenets of modern infectious disease epidemiology is that an understanding of heterogeneities, both in host demography and transmission, allows control to be efficiently optimized. Due to the strong interactions present, households are one of the most important heterogeneities to consider, both in terms of predicting epidemic severity and as a target for intervention. We consider these effects in the context of pandemic influenza in Great Britain, and find that there is significant local (ward-level) variation in the basic reproductive ratio, with some regions predicted to suffer 50% faster growth rate of infection than the mean. Childhood vaccination was shown to be highly effective at controlling an epidemic, generally outperforming random vaccination and substantially reducing the variation between regions; only nine out of over 10 000 wards did not obey this rule and these can be identified as demographically atypical regions. Since these benefits of childhood vaccination are a product of correlations between household size and number of dependent children in the household, our results are qualitatively robust for a variety of disease scenarios.

Key words: Household, influenza, modelling, transmission.

INTRODUCTION

For a wide range of directly transmitted infectious diseases, the household plays a pivotal role in transmission due to the greater strength of contacts between individuals sharing living arrangements [1, 2]. We can conceptualize many infections as transmitting readily to household members, but transmitting at a lower rate to individuals in the wider community. This concept led to both early work on quantifying these effects using clinical data [3] and to more recent attempts to incorporate households into models of pandemic influenza in Britain [4–6]. A large body of modelling work explores the spread of infection in populations structured into households, considering both thresholds for large-scale epidemics and optimal deployment of vaccination (see [7–9] and references therein). However, this work has largely been theoretical and has often not sought to relate results to available data or to consider vaccination measures that would be practically achievable. Here we consider household models relevant to the spread of pandemic influenza, and using data from the 2001 census examine the range of geographical heterogeneities in early epidemic behaviour.

Recent concerns over pandemic influenza have prompted a cascade of model development [4, 10–17] with many models acknowledging the role played by

* Author for correspondence: Dr T. House, Department of Biological Sciences, University of Warwick, Coventry, CV4 7AL, UK.
(Email: T.A.House@warwick.ac.uk)

households and structure the population and transmission accordingly [4, 10]. Some of these models are highly complex, and consider the role of transmission in households, schools and workplaces as well as incorporating localized spatial transmission [10, 12]. Here we take a simpler approach and focus exclusively on the implications of strong household transmission together with weaker transmission to the local community. Our model (see online Supplementary information, and [18]) and analysis operate at the scale of wards; there are around 10 000 wards in Great Britain (England, Wales and Scotland) with populations of between 1000 and 35 000 individuals in each according to the 2001 census. The aggregation scale used in this paper is the 2001 Census Standard Table wards – referred to simply as wards throughout – although strictly speaking such Standard Table wards are distinct from (but related to) both other statistical wards, and also electoral wards used in local government. The mathematical model essentially provides a sophisticated and dedicated tool for translating demographic characteristics into epidemiological properties.

We begin by considering household sizes and number of dependent children, both in terms of distributions within Great Britain and in terms of variability between wards. Using our household model, this variability is translated into early expected growth rates of an epidemic allowing us to explore the geographical distribution of this most important epidemiological quantity. Finally, we consider the advantages of prophylactic vaccination targeted towards dependent children compared to vaccination at random or focused towards entire households.

DEMOGRAPHIC PATTERNS IN GREAT BRITAIN

Figure 1 describes the range and distribution of household structure in Great Britain. Households consisting of just one or two individuals dominate, with decreasing numbers of households with larger occupancy (Fig. 1a). We note that the 2001 census does not contain precise information on households containing eight or more occupants, and therefore assume in our model that values of 'eight or more' are exactly eight, which makes little quantitative or qualitative difference to our results compared to any other realistic approach. This variability in household size is clearly important: large households have a greater chance of being infected (as there are more

members to potentially bring infection in from the community) and a higher rate of internal transmission (due to the greater number of contacts). However, quantifying the impact of these features requires the type of detailed mathematical model developed in the next section (and online Supplementary information).

Given the importance of large households it is important to consider their composition in more detail and to determine relationships with other demographic measures. Considering dependent children (Fig. 1b) we find, as expected, that larger households tend to have more dependent children. As such less than 10 % of households of five or more are solely occupied by adults, whereas over 90 % of households of size two have no dependent children. For this we can see that numbers of dependent children and household sizes are closely linked. We observe, as shown in Figure 1c, great geographic diversity in the proportion of dependent children in each ward – while the average is around 23 %, extremes as low as 5 % and as high as 40 % exist. Finally, we observe that the proportion of dependent children within a ward closely correlates with the average household size in that ward (Fig. 1d), although we note that there are several points this figure exhibiting large mean household size but with few dependent children; the demographic features (such as student houseshares) that can lead to this are discussed more fully in the Supplementary information.

It is against the above background of heterogeneous host demography that our mathematical model operates.

Geographic distribution of early growth rates

Our model of household-based transmission is relatively simple and parsimonious, and aims to identify the effects of different household patterns across Great Britain. Two transmission rates are used: transmission between members of the same household and transmission to general members of the local population (ward). Transmission between wards is not included, for model simplicity and transparency of results. While movement between wards and continuous importation of infection from abroad are likely to have a significant impact on the behaviour of pandemic influenza, these operate at a different scale from household-level transmission and so as a first approximation can be ignored. The general spread of infection within the ward community is modelled as frequency-dependent transmission, in accordance
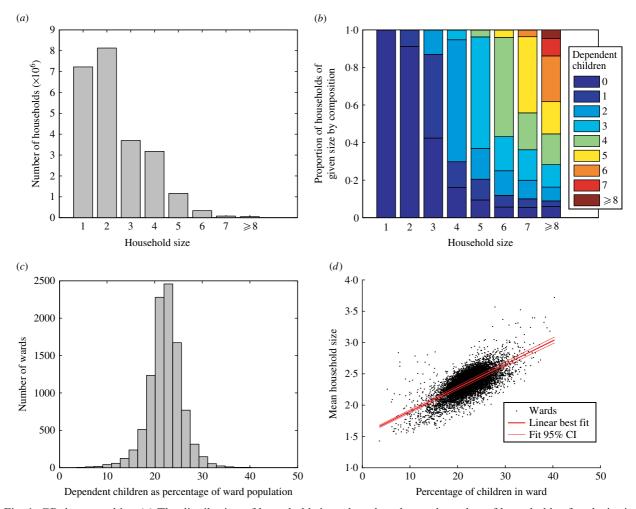
Fig. 1. GB demographics. (*a*) The distribution of household sizes plotted as the total number of households of each size in Great Britain. (*b*) The proportions of households with a given number of dependent children separated into household sizes. (*c*) The distribution of dependent children as percentages of the ward population size. (*d*) The correlation between percentage of dependent children and the mean household size within a ward; dots represent the values for each of the 9976 standard table wards in Great Britain; a simple linear fit and associated confidence intervals are shown in red. (Data source: 2001 Census Commissioned Table [27]. Crown copyright 2003. Crown copyright material is reproduced with the permission of the Controller of HMSO.)

with general modelling of large human populations [19, 20], while transmission within the household is assumed to be density dependent, such that individuals interact in a pairwise manner irrespective of household size. In practice, household transmission probably lies between the extremes of frequency and density dependence [21], but the precise scaling is likely depend on the type of household and ages of occupants. By assuming such density-dependent transmission we are maximizing the degree of heterogeneity – other assumptions produce weaker results but do not effect the qualitative conclusions.

Although our modelling framework can deal with a range of dynamic aspects of infection, here we simply consider the early (asymptotic) growth rate of infection within each ward. In particular, a range of standard theory shows that starting with a low level of infection within the population, and following some initial short-lived transients, the disease incidence and prevalence is predicted to increase exponentially [20, 22, 23]; it is this early exponential growth rate that is of primary interest here. In particular, we seek the 'basic reproductive ratio', r_0, defined such that the early growth of infected cases (I) is given by $I(t) \sim \exp([r_0 - 1]gt)$ where $1/g$ is the average infectious period. We note that this value of r_0 defined in terms of growth rates differs from the R_0 defined in terms of number of secondary cases; although both agree at
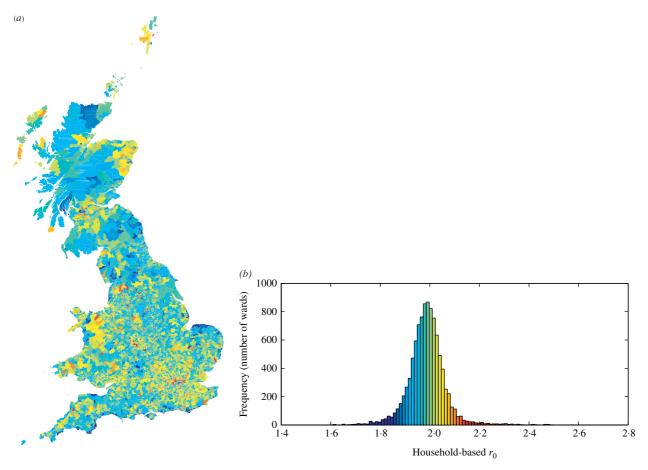
Fig. 2. Distribution of r_0 values in Great Britain, highlighting the predicted variation in early epidemic growth rates. (*a*) Wards are shown in their geographic locations and coloured according to their r_0 value. The histograms in (*b*) utilize the same colour scale and hence provide a reference for the ward-based map. (Electronic information on ward boundaries is census output, which is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland. Source: 2001 Census, Output Area Boundaries [28, 29]. Crown copyright 2003.)

the invasion threshold $r_0 = R_0 = 1$ (see online Supplementary information for a more detailed discussion). We decided to use r_0 as its definition most closely matches observations taken during an epidemic.

While a relatively simple formulation, our model is parameterized to match observables concerning pandemic influenza. Using the national distribution of household sizes, we fix the household and community transmission rates to yield a 40% chance of transmission between any two household members (often called the *secondary attack rate*) and a basic reproductive ratio of approximately 2 for Great Britain as a whole, which is consistent with statistical work in this field [1, 2]. Our qualitative conclusions are robust to the precise choice of parameters.

Figure 2 shows the geographical distribution of basic reproductive ratios (r_0 values) in each ward in

Great Britain. We observe in Figure 2*b* an approximate 25% variation in r_0 between the mean and most extreme wards, which corresponds to a 50% variation in early epidemic growth rates. In general, high growth rates reflect a greater than average abundance of large household sizes and high proportion of dependent children, although the precise relationship is complex and nonlinear. It is clear from both the locations of wards with highest r_0 (Fig. 2*a*), and the discussion in the Supplementary information, that high values of r_0 tend to be associated with the large conurbations of Great Britain, with the areas of highest r_0 having around 20 times the mean population density of Great Britain. Epidemiologically, those wards with the highest r_0 will require the greatest levels of control and therefore may be targeted with high priority during an epidemic; in addition, the fact that these high r_0 wards are generally in urban
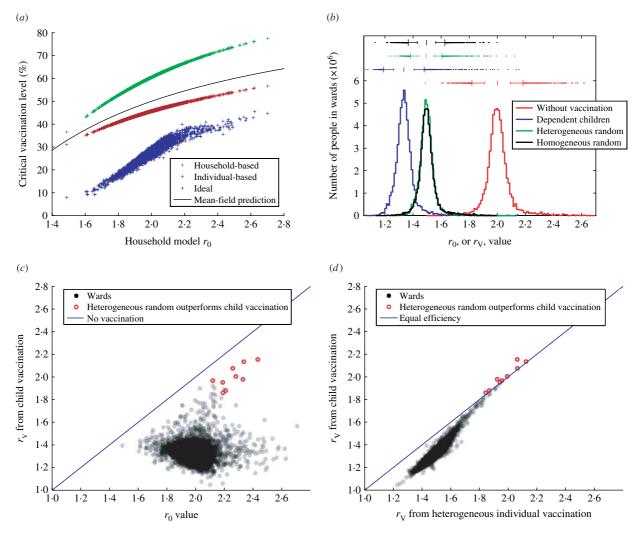
Fig. 3. Effects of vaccination. (*a*) Critical levels of vaccine coverage needed to prevent the spread of infection within a ward are shown for three strategies, along with the prediction from standard models in which there is no population structure. (*b*) The effects on the distribution of ward r_V values of three vaccination strategies. These distributions are calculated at individual level since ward-level results are slightly biased by the trend for less populated wards to have smaller household sizes. The box-whisker plots show the mean, 1 and 2 standard deviations and outliers. (*c*) The ward-level effects of vaccinating dependent children. (*d*) Comparison of the ward-level effects of vaccinating dependent children and heterogeneous random vaccination, in which the same proportion of each ward is vaccinated. The nine exceptional wards in which heterogeneous random vaccination outperforms vaccinating dependent children are highlighted (red circles) in plots (*c*) and (*d*).

areas may mean that pandemic influenza (or any other novel pathogen) is likely to enter such wards early in a national epidemic.

Control by vaccination

Prophylactic vaccination may be a key epidemiological tool in combating any future UK epidemic, either to eliminate completely the risk of a large-scale epidemic or to be used in conjunction with other methods such a social distancing, antivirals or contact tracing [24]. For simplicity, we assume that an

effective vaccine is available. Reducing this efficacy does not change our qualitative results but will make any vaccination strategy less effective. Figure 3*a* considers three methods of targeting the delivery of vaccination within wards, with the results for each ward displayed as a point. The results of our household model agree with previous findings in terms of the critical level of vaccination coverage required to prevent an outbreak [25]. Vaccinating entire households at random (green) is an inefficient means of targeting. This is because effective herd immunity at the household level can be achieved without the need

to vaccinate all household members; in essence vaccine is being wasted on individuals who already have some protection through being in partially vaccinated households. Vaccinating individuals at random (red) is a simpler and better strategy, and is found to outperform the expected vaccination threshold (black line) predicted by simpler unstructured models [19, 20, 26]. The improvement over the prediction from unstructured models is because random vaccination of individuals effectively biases vaccination towards larger households, thereby targeting control at these most epidemiologically important units. However, an ideally targeted strategy [25] – prioritizing individuals in households with the most susceptibles – has even greater benefits with the required level of critical vaccination never exceeding the random-mixing prediction of 50%. We see overall that ideal targeting can reduce by about 40% the amount of vaccine required nationally.

Unfortunately, the optimal method of targeted vaccination is both impractical and unworkable. Therefore we seek an alternative proxy that incorporates insights from the ideal strategy and readily allows vaccination to be targeted towards a proportion of individuals in the larger households. From Figure 1b we predict that vaccinating children biases protection towards the larger households, yet does not waste vaccine immunizing all members; in addition it is likely to be both ethically and socially acceptable. With this in mind, we consider three forms of vaccination at the ward level: (1) vaccination of dependent children (who account for about 23% of the GB population); (2) heterogeneous random vaccination, where individuals are vaccinated at random with the proportion vaccinated equal to the proportion of children within the ward; and (3) homogeneous random vaccination, where individuals are vaccinated at random in every ward, such that the proportion vaccinated nationally matches the proportion of dependent children. Alternatively, we can consider heterogeneous and homogeneous vaccination as randomizations of the vaccination of dependent children; heterogeneous vaccination randomizes the distribution of vaccine within each ward, whereas homogeneous vaccination randomizes the distribution of vaccine over the whole of Great Britain. As such, comparing these three strategies allow us to access the impact of targeting children, both in terms of efficient deployment within a ward and also as a means of proportioning vaccine between wards. Even though all three strategies

ultimately vaccinate the same number of individuals (around 23% of the population), it is clear that targeting has advantages (Fig. 3b). We measure the efficacy of vaccination through r_V (the equivalent of r_0, but after vaccination). Vaccinating dependent children causes a 35% drop in this reproductive ratio (from ≈ 2 to ≈ 1.3), whereas both homogeneous and heterogeneous vaccination only cause a reduction of around 25%.

Comparing homogeneous and heterogeneous vaccination in more detail allows us to assess the impact of targeting wards with the most children, without targeting large families within those wards. The histograms and box-whisker plots of r_V show that the targeting inherent in heterogeneous vaccination offers negligible mean benefit over homogeneous vaccination (Fig. 3b). However, this ward-level targeting does significantly reduce the variability in epidemic growth rates bringing those wards with extremely high growth rates under greater control.

Figure 3(c, d) considers the behaviour at the ward level, with particular focus on r_0 before vaccination and the equivalent measure r_V after a proportion of the population has been vaccinated. In general targeting vaccination towards dependent children not only reduces the average reproductive ratio (r_V) but also significantly reduces much of the variability (Fig. 3c). Wards that originally had high r_0 values due to large average household sizes with many children are now brought much closer to the average. In only nine wards (red circles) out of over 10 000 is heterogeneous random vaccination predicted to outperform vaccination targeted towards children – meaning that at a local as well as a national scale vaccinating children is overwhelmingly effective. The precise socio-demographic characteristics of these nine outliers is explored more fully in the Supplementary material, but all these wards have either large student or older adult households, breaking the general rule that large households are associated with many dependent children.

DISCUSSION

There are a wide range of regional heterogeneities within Great Britain which it may be very important to capture or appreciate if detailed mathematical models are to be effectively used in containment planning. Our results follow the general epidemiological tenet that such heterogeneities can be used to target control measures efficiently. However, ideal

targeting is impractical and socially unacceptable; instead we show that targeting prophylactic vaccination towards dependent children may be an effective (and acceptable) means of targeting intervention towards the largest and therefore most epidemiologically important households, without the disadvantages associated with vaccinating entire households. We have also found an interesting demographic pattern in the small number ($<0.1\%$) of wards that do not obey this rule, which are dominated either by student or older adult socio-demographic categories.

Our model is obviously a simplification of the complex reality of pandemic 'flu transmission in Great Britain; however, our model is sufficiently detailed to highlight the role that household structure can play and the implications of geographic heterogeneities recorded in the 2001 census. The simplifying assumptions that we believe to be most relevant when considering extensions to our work are as follows. First, we assume a compartmental paradigm where individuals are either susceptible, infectious or recovered. In reality, pathogen levels and infectiousness vary during the course of infection and also between individuals. Second, we have assumed that the strength of contacts between individuals within a household and between members of the general population are independent of household size. Finally, we have ignored other geographic diversities in assuming that the general rate of transmission in the population is the same across Great Britain, when in fact it is likely to be higher in areas with higher population density, bigger workplaces and busier transport links, which will probably inflate the variation already observed.

Bearing these limitations in mind, we believe that modelling offers a good tool for understanding socio-demographic patterns and their epidemiological consequences. In particular, our work on household structure offers the robust conclusion that vaccination of children is expected to be an effective approach to control of emergent infectious diseases, since it targets vaccine towards both wards and households with the greatest transmission risk. Furthermore, vaccinating children is likely to be socially acceptable and although not sufficient to prevent an epidemic may help to support other control measures such as social distancing, antimicrobial drugs or quarantine. Finally, we believe that insights from our work can be useful in evaluation and planning of schemes for control of diseases with existing childhood vaccination schemes (such as measles) where the

geographic diversity in epidemiologically relevant quantities that we have considered here may prove important for prioritization of efforts to maintain and increase uptake of vaccine. In this context it is interesting to note that the numbers of GP surgeries and statistical wards in the UK is approximately equal, leading to equivalent levels of geographic diversity.

ACKNOWLEDGEMENTS

DECLARATION OF INTEREST

None.

NOTE

Supplementary material accompanies this paper on the Journal's website (http://journals.cambridge.org).

REFERENCES

1. **Cauchemez S, et al.** A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statistics in Medicine* 2004; **23**: 3469–3487.
2. **Longini IM, et al.** Estimating household and community transmission parameters for influenza. *American Journal of Epidemiology* 1982; **115**: 736–751.
3. **Hope-Simpson RE.** Infectiousness of communicable diseases in the household. *Lancet* 1952; **2**: 549–554.
4. **Wu JT, et al.** Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Medicine* 2006; **3**: 1532–1540.
5. **Dodd PJ, Ferguson NM.** Approximate disease dynamics in household-structured populations. *Journal of the Royal Society Interface* 2007; **4**: 1103–1106.
6. **Fraser C.** Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE* 2007; **2**: 1–12.
7. **Ball F, O'Neill P.** Stochastic epidemic models in structured populations featuring dynamic vaccination and isolation. *Journal of Applied Probability* 2007; **44**: 571–585.
8. **Ball F, Lyne O.** Optimal vaccination schemes for epidemics among a population of households, with application to variola minor in Brazil. *Statistical Methods in Medical Research* 2006; **15**: 481–497.
9. **Ball F, Lyne O.** Optimal vaccination policies for stochastic epidemics among a population of households. *Mathematical Biosciences* 2002; **177–178**: 333–354.

10. **Ferguson NM, et al.** Strategies for mitigating an influenza pandemic. *Nature* 2006; **442**: 448–452.

11. **Viboud C, et al.** Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* 2006; **312**: 447–451.

12. **Germann T, et al.** Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences USA* 2006; **103**: 5935–5940.

13. **Chowell G, Nishiura H, Bettencourt LMA.** Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Society Interface* 2006; **4**: 155–166.

14. **Roberts MG, et al.** A model for the spread and control of pandemic influenza in an isolated geographical region. *Journal of the Royal Society Interface* 2006; **4**: 325–330.

15. **Ferguson N, et al.** Strategies for containing an emerging influenza pandemic in southeast asia. *Nature (London)* 2005; **437**: 209–214.

16. **Gani R, et al.** Potential impact of antiviral drug use during influenza pandemic. *Emerging Infectious Diseases* 2005; **11**: 1355–1362.

17. **Duerr H, et al.** The impact of contact structure on infectious disease control: influenza and antiviral agents. *Epidemiology and Infection* 2007; **135**: 1124–1132.

18. **House T, Keeling MJ.** Deterministic epidemic models with explicit household structure. *Mathematical Biosciences* 2008; **213**: 29–39.

19. **Anderson RM, May RM.** *Infectious Diseases of Humans.* Oxford: Oxford University Press, 1992.

20. **Keeling MJ, Rohani P.** *Modelling Infectious Diseases in Humans and Animals.* Princeton: Princeton University Press, 2007.

21. **Melegaro A, Gay N, Medley G.** Estimating the transmission parameters of pneumococcal carriage in households. *Epidemiology and Infection* 2004; **132**: 433–441.

22. **Diekmann O, Heesterbeek J, Metz J.** On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology* 1990; **28**: 365–382.

23. **Dieckmann O, Heesterbeek JAP.** *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation.* New York: John Wiley & Sons. 2000.

24. **Department of Health.** Pandemic flu: a national framework for responding to an influenza pandemic. Department of Health, London, 2007.

25. **Ball F, Mollison D, Scalia-Tomba G.** Epidemics with two levels of mixing. *Annals of Applied Probability* 1997; **7**: 46–89.

26. **Kermack WO, McKendrick AG.** Contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London, Series A* 1927; **115**: 700–721.

27. **Office for National Statistics.** 2001 census: Commissioned table C0844. ESRC/JISC Census Programme.

28. **Office for National Statistics.** 2001 census: Digitised boundary data (England and Wales) [computer file].

29. **General Register Office for Scotland.** 2001 census: Digitised boundary data (Scotland) [computer file].