Horby, Peter William (2012) Avian, inter-pandemic, and pandemic influenza in Vietnam. PhD (research paper style) thesis, London School of Hygiene & Tropical Medicine. DOI: https://doi.org/10.17037/PUBS.01236302

Downloaded from: http://researchonline.lshtm.ac.uk/1236302/

DOI: 10.17037/PUBS.01236302

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: Creative Commons Attribution Non-commercial http://creativecommons.org/licenses/by-nc/3.0/

Avian, inter-pandemic, and pandemic influenza in Vietnam

Peter William Horby

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Department of Infectious Disease Epidemiology

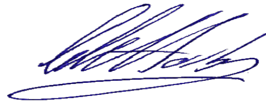
A thesis submitted to the University of London for the degree of Doctor of Philosophy

March 2012

Declaration by candidate

I have read and understood the School's definition of plagiarism and cheating given in the Research Degrees Handbook. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

I have read and understood the School's definition and policy on the use of third parties (either paid or unpaid) who have contributed to the preparation of this thesis by providing copy editing and, or, proof reading services. I declare that no changes to the intellectual content or substance of this thesis were made as a result of this advice, and, that I have fully acknowledged all such contributions

Signed: 

Date: 14th February 2012

Full Name: **Peter William Horby**

Abstract

The burden and behaviour of influenza in Southeast Asia is poorly characterised, leading to uncertainty about the importance of influenza as a local health problem and the role of Southeast Asia in the global epidemiology of influenza. Prospective community-based studies have provided fundamental insights into the epidemiology of influenza in temperate regions; therefore a household-based cohort study was established with the aim of determining the intensity and characteristics of influenza transmission in a semi-rural tropical setting. The primary results of the cohort study are presented, along with the results of a survey of social contact patterns in the cohort and a mathematical model of the spread of pandemic influenza A/H1N1/2009 in Vietnam that utilises data from the cohort.

Highly pathogenic avian influenza A/H5N1 remains endemic in poultry in parts of Southeast Asia and continues to infect humans. Marked familial clustering of human H5N1 cases has led to speculation that susceptibility to H5N1 infection may have a host genetic component. The epidemiological data that led to the hypothesis of genetic susceptibility to H5N1 is summarised, whilst the evidence for a role of host genetics in susceptibility to influenza in general is systematically reviewed. A genome-wide case-control genetic association study was conducted in Vietnam and Thailand to test the hypothesis of genetic susceptibility to H5N1 infection, and the results are presented.

This work provides new data and understanding of the patterns and determinants of inter-pandemic, pandemic, and avian influenza epidemiology. The cohort study has added to the body of knowledge that is accruing on the burden and epidemiology of influenza in the tropics by providing community level data that were previously absent. The genetics study has provided the first direct evidence of genetic loci associated with susceptibility to H5N1 and opens new avenues of research to test these findings and their relevance to the pathogenesis of H5N1 and other types of influenza.

Acknowledgements

Foremost I would like to thank my supervisor Neal Alexander for his wise guidance, his thought provoking questions, and his readiness to help solve problems at any time of night or day with good humour and patience. Also my co-supervisors Punam Mantagni and Branwen Hennig.

The research owes much to the hard work of many people, but particular thanks are due to my friend and colleague Annette Fox for so ably managing the extensive laboratory work that was required and for the long nights and weekends she spent in the lab; to my colleagues Marcel Wolbers and Maciej Boni for their statistical and modelling expertise; to all my friends at the Oxford University Clinical Research Unit in Hanoi for their friendship and dedication; and to Jeremy Farrar for his mentorship.

None of this would have been possible without my colleagues at the National Institute for Hygiene and Epidemiology in Hanoi who guided me through the intricacies and subtleties of conducting field research in Vietnam. I had enormous fun with them and learnt so much. Thanks are especially due to Le Quynh Mai, Pham Quang Thai, Le Thi Thu Yen, Tran Nhu Duong, and Nguyen Tran Hien.

Thanks also to Felix and Hana, whose delightful "Sorry to interrupt you Daddy" was always the highlight of a long day at the desk, especially if they brought a Collywobble.

My final thanks must go to Mandy for her enduring love, unfailing support, and endless capacity to listen to my minor trials and tribulations: "thank-you darling and Happy Valentines Day!".

February 14th 2012

Contents

Page

List of Figures **7**

List of Tables **9**

Preface **11**

1 General introduction **13**

1.1 Influenza viruses . . . 13
1.2 Influenza disease in humans . . . 14
1.3 Influenza epidemiology in humans . . . 15
1.4 Influenza in East and Southeast Asia . . . 16
1.5 Influenza in Vietnam . . . 17
1.6 Vietnam . . . 18

2 Background to household cohort study **19**

2.1 Longitudinal community studies . . . 19
2.2 Social contact patterns . . . 20
2.3 Inter-pandemic influenza in East and Southeast Asia . . . 21
2.4 Inter-pandemic influenza in Vietnam . . . 22
2.5 Objectives of household cohort study . . . 22
2.6 Candidate's role . . . 22

3 Contribution of research papers*(cohort study)* **24**

3.1 Outline of research papers . . . 24

4 Research paper 1 **26**

5 Research paper 2 **71**

6 Research paper 3 **79**

7 Discussion *(cohort study)* **101**

7.1 Contribution to knowledge of influenza epidemiology 101

7.2 Further research directions 103

8 Background to host genetics study **104**

8.1 Host genetics and infectious diseases 104

8.2 Host genetics and influenza 104

8.3 Objective of host genetic study 105

8.4 Candidate's role 106

9 Contribution of research papers *(genetics study)* **107**

9.1 Outline of research papers 107

10 Research paper 4 **109**

11 Research paper 5 **119**

12 Research paper 6 **150**

13 Discussion *(genetics study)* **174**

13.1 Contribution to knowledge on host genetic susceptibility to H5N1 174

13.2 Further research directions 175

14 Concluding remarks **177**

References **180**

A Supplementary research paper 1 **195**

B R code **204**

List of Figures

Research paper 1 - Figure 1: Number of study participants included in assessment of influenza illness and infection status by influenza season, Ha Nam, Vietnam, 2007-2010. 27

Research paper 1 - Figure 2: Timeline of ILI cases, RT-PCR confirmed influenza illnesses, and cross sectional bleeds, Ha Nam, Vietnam, 2007-2010. 27

Research paper 1 - Figure 3: Risk of influenza infection by season, influenza sub-type, and age group, Ha Nam, Vietnam, 2007-2010. 27

Research paper 1 - Web Figure 1: Age distribution of the cohort compared to provincial and national rural populations, Ha Nam, Vietnam, 2007-2010. 27

Research paper 1 - Web Figure 2: Frequency of bleeding amongst cohort participants under ILI surveillance, by age and gender. Ha Nam, Vietnam, 2007-2010. 27

Research paper 1 - Web Figure 3: Influenza infection rates in Ha Nam 2007-2010, compared to historic household cohort studies. 27

Research paper 1 - Web Figure 4: Risk of influenza infection by season, influenza sub-type, gender, and age group. Ha Nam, Vietnam, 2007-2010. . 27

Research paper 2 - Figure 1: Household sizes (A) and number of reported contacts per person per day (B). 72

Research paper 2 - Figure 2: Contacts by location, duration and frequency. . 72

Research paper 2 - Figure 3: The location, duration and frequency of contacts. 72

Research paper 2 - Figure 4: Contact intensity matrices for all contacts (A) and for physical contacts only (B). 72

Research paper 2 - Figure 5: The predicted effect on R0 of immunizing individuals or households. 72

Research paper 3 - Figure 1: The range of possible epidemics in Vietnam. (A) The range of possible epidemics in Vietnam. (B) Median number of new cases by day, with day zero corresponding to the epidemic peak. . . . 80

Research paper 3 - Figure 2: Geographic spread of swine-origin influenza A (H1N1) in Vietnam. 80

Research paper 3 - Figure 3: Timing of provincial epidemic peaks based on the distance from the nearest airport to the capital city. 80

Research paper 3 - Figure 4: Geographic timeline of chicken exposure during an influenza epidemic in Vietnam. 80

Research paper 3 - Figure 5: Result sensitivity relative to the R0-value as it would have been measured in Ho Chi Minh City. 80

Research paper 4 - Figure 1: Family pedigree showing three H5N1 affected individuals, with infections separated by 2 years. 110

Research paper 4 - Figure 2: Proportion of cases occurring in household clusters by probability of infection for different household sizes. (a) All data. (b) Enlargement of left-hand corner of panel (a). 110

Research paper 5 - Figure: Identification and screening of articles for inclusion in systematic review. 120

Research paper 6 - Figure 1: Quantile-quantile plot of the joint analysis of Vietnam and Thailand data. 151

Research paper 6 - Figure 2: Manhattan plot of the joint analysis of Vietnam and Thailand data. 151

Research paper 6 - Figure S1: Analysis of genetic ancestry for H5N1 cases and controls against HapMap reference collections. 151

Research paper 6 - Figure S2: Illumina genotype cluster plots for SNPs rs4849124 and rs7560562 in Vietnam and Thailand H5N1 sample collections. 151

List of Tables

Research paper 1 - Table 1: Characteristics of Participants and Households at Recruitment, Ha Nam, Vietnam, 2007-2010. 27

Research paper 1 - Table 2: ILI Episodes and Influenza Virus Detections by Season and Age Group, Ha Nam, Vietnam, 2007-2010. 27

Research paper 1 - Table 3: Unadjusted and Standardized Risks of Influenza Infection and Influenza Illness by Season, Persons Aged 5 Years, Ha Nam, Vietnam, 2007-2010. 27

Research paper 1 - Table 4: Risk Factors for Influenza Infection, Aggregated Over Influenza Subtypes and Seasons, Ha Nam, Vietnam, 2007-2010. 27

Research paper 1 - Web Table 1: Source data for Web Figure 3. 27

Research paper 2 - Table 1: Number of recorded contacts per participant per day by characteristics, and relative number of contacts from weighted GEE analysis. 72

Research paper 3 - Table 1: Median, quartile and minimum - maximum values for selected outputs of one year of model simulation. 80

Research paper 4 - Table 1: Number of confirmed H5N1 cases and clusters by country. 110

Research paper 5 - Table 1: Key studies of heritability or genetic susceptibility in mice 120

Research paper 5 - Table 2: Key studies of familial aggregation, heritability, or genetic susceptibility in humans 120

Research paper 6 - Table 1: Results of Single Nucleotide Polymorphism analysis 151

Research paper 6 - Table S1: Minor allele frequencies for SNPs rs7560562 and rs4849124 in Vietnamese H5N1 cases (N = 45), Vietnamese unrelated adult controls (N = 178), and Vietnamese cord blood controls (N = 2,018). 151

Research paper 6 - Table S2: Two-locus association test for TRPM8 rs7560562 and IL-1α/IL-1β rs4849124. 151

Preface

This thesis is divided into two major sections. The first deals with the results of a prospective community cohort that was established to provide data on the epidemiology of inter-pandemic influenza in Vietnam. The cohort was running when pandemic influenza A/H1N1/2009 emerged and therefore also provides data on the epidemiology of pandemic influenza in Vietnam. The second section deals with work to test the hypothesis that host genetic factors play an important role in susceptibility to highly pathogenic influenza A/H5N1. Since the epidemiology of H5N1 is quite different from inter-pandemic and pandemic influenza, each section begins with a separate description of the scientific background to the work and the research questions.

In accordance with the LSHTM Research Degrees Regulations, much of the thesis is presented as a series of published or accepted research manuscripts. The research papers are supplemented by additional material to explain my contribution, to provide additional methodological details, and to describe how the thesis represents a coherent body of new scientific material. The presentation of the work as a series of stand-alone manuscripts inevitably results in some repetition of background information. Differing editorial conventions between journals may result in inconsistencies in terminology and formatting.

I conceived of all the presented research, drafted all the research protocols and data collection instruments, submitted all protocols for ethical approval, implemented the studies, and supervised all the field work and data collection. I either analysed data personally or was closely involved in any statistical or mathematical analysis that was led by other individuals. The laboratory work was conducted by others but I was involved in the planning and design of all laboratory analyses and the interpretation of results. Where colleagues have contributed to laboratory or statistical aspects of the work this is clearly acknowledged.

Research by foreign academic individuals and institutes within Vietnam must be conducted in partnership with a local 'competent authority', which for health research is an institute under the jurisdiction of the Ministry of Health. The work presented in this thesis was conducted under the umbrella of a project agreement between the National Institute of Hygiene and Epidemiology (NIHE) and the Oxford University Clinical Research Unit in Hanoi. The

research was approved by the Institutional Review Boards of NIHE, the Vietnam Ministry of Health, and the Oxford University Tropical Research Ethics Committee. The host genetics work in Thailand was conducted in partnership with the Ministry of Public Health.

The term 'inter-pandemic influenza' is used to distinguish influenza subtypes that normally circulate in humans from avian influenza viruses and from the pandemic influenza A/H1N1 strain that emerged in 2009. The term 'seasonal influenza' is not used since seasonality is less marked in the tropics and pandemic influenza A/H1N1/2009 also exhibits seasonality. For brevity the term 'H5N1' is used to refer to highly pathogenic avian influenza A/H5N1 unless otherwise stated.

Chapter 1

General introduction

1.1 Influenza viruses

Influenza viruses are enveloped RNA viruses of the family *Orthomyxoviridae* with a segmented negative-sense, single-stranded genome. They can be categorised serologically and genetically into three types (A,B, and C). Influenza B and C are predominantly human pathogens whilst type A naturally infects a wide range of birds and mammals. Influenza A is the most important type since it regularly causes large epidemics and, when a new subtype emerges to which the population are immunologically naïve, can cause a global outbreak (pandemic). Type A is categorised into subtypes based on the antigenic characteristics of two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA). There are currently 16 identified HA (H 1-16) and 9 NA (NA 1-9) antigenic variants, giving a total of 144 possible HA-NA combinations. Only 3 HA subtypes (H1, H3, and H2) and 2 NA subtypes (N1 and N2) are known to have caused sustained transmission in human populations. The majority of HA and NA combinations have been identified in aquatic birds, which are the primary natural reservoir of influenza A viruses, and which are thought to be the source of progenitor viruses or gene segments of pandemic influenza A strains (Alexander, 2007). Zoonotic, non-human-adapted, influenza A viruses, such as highly pathogenic avian influenza (HPAI) A/H5N1, occasionally infect humans.

Host range restriction of influenza A viruses is multi-factorial but one key determinant is HA receptor binding specificity (Kuiken et al., 2006). Influenza viruses infect cells through the binding of HA to the terminal sialic acid molecules of glycoproteins and glycolipids expressed on host cell membranes, and the subsequent fusion of viral and cell membranes (Gamblin and Skehel, 2010; Wiley and Skehel, 1987). HA is synthesised as a precursor protein HA0 that must be cleaved by host proteases into HA1 and HA2 for membrane fusion to occur. Whilst the HA of human influenza viruses preferentially bind sialic acid with an α2-6 linkage to galactose residues, which are abundant in the upper respiratory tract of humans, avian viruses bind more readily with an α2-3 linkage, which are found in the intestinal tract of birds and the lower respiratory tract of humans (Kuchipudi et al., 2009; Shinya et al., 2006). Swine viruses are able to bind sialic acid in both types of linkage (Gamblin and Skehel, 2010; Glaser et al., 2007; Matrosovich et al., 2006; Skehel and Wiley, 2000). The receptor binding and infection process is however almost certainly more complex and nuanced than a simple dichotomy between human-α2-6-upper airway and avian-α2-3-lower airway, and the specific identify of the cell-surface

molecules to which HA binds is not yet known (Matrosovich et al., 2006; Nicholls et al., 2008; Peiris et al., 2009; Shelton et al., 2011; Viswanathan et al., 2010).

Influenza viruses have a high potential for genetic and antigenic diversity. This is the result of two intrinsic characteristics. First, influenza viruses have no system of 'proof-reading' during replication of viral RNA by the viral polymerase, leading to high rates of uncorrected transcription errors. This potential for diversity is actualised most strongly in the HA as a result of immune-mediated selection pressures. The binding site of the HA protein is the main target for neutralising antibodies and this region is therefore under intense immune-mediated selection pressure, resulting in the acquisition and retention of amino acid substitutions which favour escape from immunity. The result is regular epidemics caused by antigenically distinct (drifted) strains, which are well visualised using the technique of antigenic cartography (Smith et al., 2004). The second characteristic is the segmented nature of the influenza genome, which allows for reassortment of gene segments when cells are concurrently infected with more than one influenza strain. Reassortment can result in the generation of novel strains to which there is little population immunity, causing a pandemic.

1.2 Influenza disease in humans

Influenza is one of the commonest infections of humans, with a recent meta-analysis estimating there were 20 million cases of influenza-associated acute lower respiratory tract infection and 1 million severe cases in children younger than 5 years in 2008 (Nair et al., 2011). In the US between 1976 and 2007, influenza-attributable mortality was estimated at between 1.4 to 16.7 deaths per 100,000 persons, with around 90% occurring in people $\geq$ 65 years of age (CDC, 2010). Similar results have been found elsewhere (Cohen et al., 2010; Nielsen et al., 2011; Nunes et al., 2011; Yang et al., 2011). Estimating the health burden of influenza is however difficult since influenza may not be recognised as the immediate cause of death even though it precipitated the illness event, and influenza burden is highly heterogenous, with an epidemic profile that may go relatively unnoticed or cause a pandemic with tens of millions of deaths.

The clinical syndrome of influenza is classically associated with upper respiratory tract symptoms accompanied by fever, headache and myalgia; although mild and subclinical infection is common. Children generally have the highest clinical attack rate since they are immunologically naïve. Although antibodies directed against the binding site of the HA are the primary mechanism for neutralising immunity, a wide range of other host responses influence the probability of infection and the severity of illness e.g. antibodies to NA, antibodies to conserved regions of the HA protein, innate immune responses, cell-mediated immunity, and HA glycosylation. There is particular interest in the potential of inducing immunity to a range of subtypes (heterosubtypic immunity) through cell-

mediated or humoral immunity directed against conserved epitopes, such as the fusion-peptide of the HA2 molecule (Corti et al., 2011; McMichael et al., 1983).

Severe disease is most frequently seen in the very young, the elderly, and people with chronic medical conditions, and, as with several other infections, pregnancy is also a risk factor for severe influenza-associated disease (Jamieson et al., 2006). Whilst such host factors are important in determining disease severity, viral factors are equally important. The determinants of the high virulence of the 1918 H1N1 and H5N1 viruses is not completely understood but is clearly polygenic, with polymorphisms in the HA, NS1, and the polymerase complex genes all playing a role (Basler and Aguilar, 2008; Fukuyama and Kawaoka, 2011). However, one key determinant of virulence is tissue tropism, which is co-determined by receptor binding affinity and the specificity of host proteases involved in post-translational cleavage of HA0. The α2-3 linked sialidases found in the intestinal tract of birds, to which avian influenza viruses have a high binding affinity, are also found more commonly in the terminal bronchioles and alveoli of humans (Shinya et al., 2006; van Riel et al., 2006). Whilst human-type receptor affinity seems to be associated with attenuation of virulence and efficient human-to-human transmission (van Riel et al., 2010; Watanabe et al., 2011), avian-type receptor affinity is seen in H5N1 and has been identified in some viruses causing severe 1918 H1N1 and 2009 H1N1 disease (Chutinimitkul et al., 2010; Sheng et al., 2011; Watanabe et al., 2011). Human-adapted viruses and low-pathogenic avian viruses posses a single arginine at the site of cleavage of HA0 into HA1 and HA2. The presence of multiple basic amino acids[1] at this site is a feature of highly pathogenic avian influenza viruses and permits the cleavage of HA0 by intracellular enzymes present in wide range of tissues, resulting in extensive virus replication and systemic infection (Perdue et al., 1997). This feature is however not alone sufficient to cause severe disease in humans since a highly pathogenic H7N7 virus with multiple basic amino acids at the cleavage site caused generally mild illness in humans in the Netherlands in 2003 (Fouchier et al., 2004).

1.3 Influenza epidemiology in humans

Influenza epidemiology is characterised by annual winter epidemics at higher latitudes, with periodic influenza A pandemics that may escape seasonal patterns (Shaman et al., 2011; Tamerius et al., 2011). The attack rate varies from 10-40% per season depending on the immunological attributes of the population, the antigenic novelty of the virus, and the transmission intensity (which may vary with climate and contact behaviours). Transmission is through infectious respiratory secretions but the relative contribution of large respiratory droplets versus small airborne particles remains unresolved, and perhaps differs with climate (Brankston et al., 2007; Lowen and Palese, 2009; Tellier, 2009). The

[1] Histidine, Lysine, and Arginine

explosive nature of influenza epidemics in temperate areas is largely due to the short serial interval[1] rather than its intrinsic infectiousness (Boelle et al., 2011; Cowling et al., 2009; Truscott et al., 2012).

Annual influenza epidemics and periodic pandemics are a testament to the adaptive competence of influenza, yet whilst the molecular processes underlying influenza evolution are fairly well described, the population and environmental dynamics driving global influenza evolution are less well understood. A central question remains "when and where does most antigenic drift occur?"(Nelson et al., 2007). The answer is important since it may offer opportunities to either predict evolution or use strategic surveillance to detect emerging new strains. Antigenic drift has not been observed within one locality and season (Lavenu et al., 2006), and several authors have proposed a source-sink model, where East and Southeast Asia acts as a regional source of drifted viruses that seed annual seasonal epidemics in temperate regions (Rambaut et al., 2008; Russell et al., 2008). This source-sink model has however recently been challenged by Bahl et al. (2011) and the role of the tropics in general and of Southeast Asia in particular in the global epidemiology of influenza remains unclear.

In temperate regions the annual pattern of influenza epidemics, with alternation of dominant subtypes/antigenic variants, has been attributed to a combination of homosubtypic immunity lasting 3-8 years, short lived heterotypic immunity, and a seasonal forcing component (Ferguson et al., 2003; Truscott et al., 2012). The seasonal forcing component has been variously postulated to be contact behaviours, absolute humidity, and Vitamin D concentrations (Cannell et al., 2008; Shaman et al., 2011). Attempts have not been made to model the parameters that determine patterns of influenza in the tropics, since patterns have not yet been well delineated.

Avian and swine influenza viruses occasionally infect humans, causing disease of variable severity, and although limited person-to-person transmission of both H5N1 and H7N7 has occurred it has not been sustained. Recent publicised but unpublished work in ferret models show that H5N1 can readily adapt to become transmissible by the aerosol route whilst retaining virulence, so there is no reason to believe that highly pathogenic animal influenza viruses are unable to successfully adapt to humans. Indeed, the pandemic viruses of 1918, 1957, 1968, and 2009 all possessed novel gene segments thought to be of animal origin (Guan et al., 2010).

1.4 Influenza in East and Southeast Asia

East and Southeast Asia has been of interest to the influenza community for some time as a possible epicentre for the generation of novel influenza viruses (Shortridge and Stuart-Harris, 1982). However, three relatively recent events reignited international interest in

[1] The time interval between successive cases in the chain of transmission

influenza in tropical areas of East and Southeast Asia: Severe Acute Respiratory Syndrome highlighted the potential for new zoonotic pathogens to emerge and spread rapidly from Asia to affect the rest of the world; the epizootic of H5N1 that began in 2003 resulted in unprecedented numbers of human H5N1 cases with concerns of person to person spread of a highly virulent pathogen; and genetic and antigenic analyses of the HA suggested that influenza H3N2 virus strains causing seasonal outbreaks in temperate areas may originate from a transmission network within East and Southeast Asia (Rambaut et al., 2008; Russell et al., 2008). These events have led to a perception that the control of avian and human influenza in East and Southeast Asia is a 'global public health good' (Lee et al., 2007).

Vietnam is viewed as an important stakeholder in improved global understanding of avian and human influenza since it possesses a large backyard and commercial poultry industry in which H5N1 is endemic and has a large human population living in close proximity to one another and to poultry and pigs. It is also one of only three countries to have reported more than 100 human H5N1 cases and is closely linked to Southern China, where avian influenza viruses circulate widely in wild and domestic birds. The surveillance and control of inter-pandemic influenza in Vietnam, although of less direct international interest, is viewed as a mechanism for establishing the infrastructure, expertise, and systems for control of more globally threatening influenza viruses.

1.5 Influenza in Vietnam

Prior to 2004 influenza was a low priority for the Ministry of Health (MoH) and research community in Vietnam, with national surveillance limited to routine monthly reporting of influenza-like illness (ILI) cases by administrative area, and one small study of virologically confirmed influenza in Hanoi (Nguyen et al., 2007). Influenza vaccination was not available in the public or private sector. This changed with the emergence in 2003 of H5N1 and its transmission to humans. Although it was known prior to 2003 that H5N1 was present in poultry in Vietnam, this fact and its significance had escaped local health and veterinary authorities (Jadhao et al., 2009; Nguyen et al., 2005).

The re-emergence of H5N1 in Asia in 2003 was first detected in Vietnam following the investigation of a cluster of unexplained deaths from respiratory illness in children and this resulted in enormous international pressure on Vietnam to investigate and control H5N1 in poultry and prevent human infections (Tran et al., 2004). As a result, Vietnam has seen considerable investment in influenza surveillance, with the establishment in 2005 of a World Health Organisation (WHO) designated National Influenza Centre (NIC) at the National Institute of Hygiene and Epidemiology (NIHE) and the development of a national influenza surveillance program in 2005 (Nguyen et al., 2009). Vietnam has also invested in influenza vaccine development, with the generation of H5N1 vaccine seed strains and the development of influenza vaccine pilot lot production capabilities at three sites under

a cooperative agreement with United States Department of Health and Human Services (Hoa et al., 2011).

1.6 Vietnam

Vietnam has a land area that is about 30% greater than that of the United Kingdom and is an elongated country with a length of 1,650 km that straddles different climate zones. It lies within the Tropic of Cancer, and whilst the climate is constantly hot in the southern region, with temperatures rarely dropping below 20°C, the northern region has four distinct seasons, with winter temperatures occasionally falling as low as 10°C in Hanoi. It has a long land border, with China to the north and Laos and Cambodia to the east. It is the third most populous country in Southeast Asia after Indonesia and the Philippines, with an estimated population of 85.7 million (GSO, 2010). The population density is 259 persons/km^2, slightly higher than that of the UK. 30% of the population live in urban areas and the population is concentrated in the agriculturally productive and industrial zones of the Red River Delta in the north and the Mekong River Delta in the south. The age distribution of the population is: 25% aged under 15 years, 68% aged 15-64 years, and 7% aged 65 years or more (GSO, 2010).

Vietnam has experienced sustained economic growth in the last two decades (average annual growth 7.1%) and recently transitioned from a low income to a lower-middle income country. The per capita gross national income in 2010 was estimated to be around 1,000 US$, and around 60% of the population have unskilled occupations or work in agriculture, fisheries, and forestry. Around 13% of the population live on less than 1.25 US$ per day (2008 Asian Development Bank estimate). Vietnam has good health and development indicators relative to its gross domestic product.

Chapter 2

Background to household cohort study

2.1 Longitudinal community studies

Influenza infection is often mild and cases self-presenting to health service providers are not representative of the true incidence and distribution of infection. Community-based studies are therefore necessary to fully understand the epidemiology of influenza and since influenza epidemiology varies by season, these studies need to be longitudinal (Monto, 1994). Such studies are important since the impact of public heath measures such as immunisation, social distancing, and chemoprophylaxis, on the incidence of influenza are dependent on the dynamics of influenza transmission (Ferguson et al., 2006). With the direct and indirect benefits of influenza control measures targeted at specific population subgroups being contingent on the burden of infection and disease in that group, and the contribution of that group to onward transmission of infection.

Seminal community-based studies of influenza transmission were conducted in the United Kingdom and the United States from the 1920's to the early 1980's, and are the source of much of our current understanding of the epidemiology of influenza. These studies have been summarised by Fox (1974) and Monto (1994), and have consistently demonstrated that influenza is a common cause of respiratory illness in the studied communities. The Cleveland Family Study reported seroconversion rates[1] of 15-25% during epidemic seasons (Jordan et al., 1958); the New York Virus Watch study reported seroconversion rates of 12% for influenza A and 3% for influenza B (Hall et al., 1971); the Seattle Virus Watch study reported seasonal infection rates of up to 31% (Fox et al., 1982, 1972); the Tecumseh study reported influenza seroconversion rates of 16.7% for influenza A and 7.6% for B (Monto and Kioumehr, 1975); and the Houston Family study found H3N2 infection rates of 33%, H1N1 of 15%, and B infection rates of 20-24% (Frank et al., 1983, 1985). These studies have also provided data on the relative pathogenicity of influenza types and subtypes, with a tendency for H3N2 to cause more severe disease and more infections in adults compared to H1N1 and B (Fox et al., 1982; Hope-Simpson, 1984; Monto et al., 1985; Monto and Sullivan, 1993). However, the data also demonstrate that "there is no such thing as a typical year" (Monto, 2008).

[1] The denominator for rates varies between studies and are summarised in the table on page 58

These community studies have also demonstrated that there are high rates of serological evidence of infection without corresponding disease (Hall et al., 1973; Jordan et al., 1958; Monto, 1994); that, depending on the season and subtype, pre-school or school-age children have the highest rates of infection, (Fox et al., 1982; Glezen and Couch, 1978; Hall et al., 1973; Hope-Simpson, 1984; Jordan et al., 1958; Lidwell and Sommerville, 1951; Monto and Kioumehr, 1975; Monto et al., 1985); that mothers have higher attack rates than fathers (Badger et al., 1953b; Hall et al., 1971; Jordan et al., 1958; Monto and Sullivan, 1993); and that school-age children play an important role in introducing infection into families (Badger et al., 1953a; Fox et al., 1982; Lidwell and Sommerville, 1951; Monto and Kioumehr, 1975; Philip et al., 1961). More recent studies have enrolled households of index influenza cases presenting to health care facilities, confirming the importance of children in transmitting infection within households, and refining estimates of the household secondary attack rate and serial interval (Boelle et al., 2011; Cauchemez et al., 2004; Cowling et al., 2010; Viboud et al., 2004). The findings from these observational studies have been used to guide influenza control interventions in both field studies (Hurwitz et al., 2000; Monto et al., 1970; Reichert et al., 2001) and computer simulations (Ferguson et al., 2006; Glass and Barnes, 2007), and provide the rationale for the use of school closure and childhood immunisation to reduce the overall incidence of influenza in the community (Cauchemez et al., 2009; Reichert et al., 2001).

2.2 Social contact patterns

Rates of contact between individuals of different ages are thought to be a key determinant of influenza transmission patterns and are a core component of mathematical models to predict epidemic dynamics and the impact of control measures (Wallinga et al., 1999, 2006). Such models are increasingly being used to guide decisions about the application of control measures (Keeling and Danon, 2009). Social contact patterns relevant to the spread of respiratory-transmitted infections such as influenza are dependent on many factors, such as household size, family structures, child care arrangements, working patterns, commuting patterns, and schooling. Since these factors are likely to vary with socio-economic, cultural, and climatic conditions, it is possible that social contact patterns may also vary. As with the community studies of influenza transmission, empirical studies of social contact patterns relevant to influenza transmission have been conducted almost exclusively in Europe and the US (Edmunds et al., 1997; Glass and Glass, 2008; Mossong et al., 2008).

2.3 Inter-pandemic influenza in East and Southeast Asia

Whilst it has long been suggested that China may be an epicentre for influenza evolution (Shortridge and Stuart-Harris, 1982), influenza associated morbidity and mortality had been assumed to be unimportant in the tropics and subtropics. This assumption began to be challenged by studies in the late 1990's and early 2000's. Simmerman and Uyeki (2008) have summarised the English language literature on the burden of influenza in East and Southeast Asia from 1980 to 2006, and although the data are somewhat limited, it appears that the burden of influenza in high income countries of East and Southeast Asia (Hong Kong, Singapore, Thailand, and Japan) is similar to Europe and North America; with 11-26% of outpatients with ILI having laboratory confirmed influenza infection, and influenza being identified in 6-14% of pneumonia admissions (Simmerman and Uyeki, 2008). Studies published since the review by Simmerman and Uyeki add to the evidence that influenza causes a substantial burden of disease in Southeast Asia (Brooks et al., 2010; Clague et al., 2006; Hanshaoworakul et al., 2009; Lee and Fidler, 2007; Lee et al., 2009; Leo et al., 2009; Nair et al., 2011; Simmerman et al., 2009; Viboud et al., 2006; Yang et al., 2011). In fact, influenza associated hospitalisation rates in young children may be higher than in temperate developed countries (Chiu et al., 2002; Nair et al., 2011; Simmerman et al., 2009; Wong et al., 2006). Nevertheless, large gaps remain in our understanding of the burden of influenza amongst the enormous population of poor people in East and Southeast Asia (e.g. China, Indonesia, and the Philippines).

The periodicity of inter-pandemic influenza in Southeast Asia is also gradually being elucidated, with data showing that although epidemics may be less pronounced, less synchronous between countries, and less predictable in tropical and subtropical regions compared to temperate regions, seasonal influenza epidemics clearly do occur, usually associated with the rainy season (Moura, 2010; Nelson et al., 2007; Russell et al., 2008; Shek and Lee, 2003; Tamerius et al., 2011; Viboud et al., 2006), with some areas experiencing more than one transmission period per year (Chew et al., 1998; Lee et al., 2009; Simmerman et al., 2009; Wong et al., 2006; Yang et al., 2011). The less distinct epidemic patterns probably contribute to the earlier misperception that influenza is not a significant health issue in the tropics and subtropics. Parts of the tropics and subtropics experience extended viral circulation compared to temperate areas, but it has not been shown that there is continuous year-round circulation of any single influenza subtype in any one country (Bahl et al., 2011; Russell et al., 2008).

Despite the value of the longitudinal community studies mentioned above, very few comparable studies have been undertaken in Southeast Asia. Since population densities, family structures, behaviours, material conditions, health and climate are different in Southeast Asia compared to the USA or the UK, the epidemiology of influenza transmission may also differ in important ways. Only one household-based study of influenza in Asia has been published (Riley et al., 2011). This study analysed paired serology from

770 individuals aged 3 - 103 years in Hong Kong to make estimates of the 2009 pandemic H1N1 infection rate and to identify risk factors for infection. The study found age-specific infection rates similar to those reported elsewhere, and identified an increased risk of infection in adults who shared their household with a child (Riley et al., 2011). The 2009 pandemic H1N1 is not however characteristic of inter-pandemic influenza, and Hong Kong is not characteristic of Southeast Asia.

2.4 Inter-pandemic influenza in Vietnam

Influenza vaccine development is being encouraged and pursued in Vietnam (Hoa et al., 2011; Perdue and Bright, 2011), yet little is known about the epidemiology of influenza in Vietnam other than the periodicity of strains detected through sentinel surveillance (Li et al., 2008; Nguyen et al., 2009, 2007). Vital statistics data are too unreliable and fragmentary to make estimates of excess mortality from influenza-associated illness (Rao et al., 2010) and although ILI is one of 24 legally notifiable disease syndromes in Vietnam, there is no laboratory confirmation of aetiology and the data are not available disaggregated by age. If inter-pandemic and pre-pandemic influenza vaccination is to be considered in Vietnam it is possible that a school age programme may be an efficient strategy since childhood vaccination may have a community wide protective effect (Hurwitz et al., 2000; Monto et al., 1970; Reichert et al., 2001) and Vietnam has no systems established for identifying and vaccinating high-risk adults. However data from Vietnam to guide the design and implementation of immunisation and other influenza control policies are lacking.

2.5 Objectives of household cohort study

A household-based cohort was established with the aim of estimating the incidence of ILI and laboratory confirmed influenza infection in the community; describing age and gender specific clinical and sub-clinical attack rates; identifying risk factors for influenza infection; and describing contact patterns within the cohort in order to provide local parameters to guide influenza control programmes and to improve understanding of the dynamics of influenza transmission in tropical Southeast Asia.

2.6 Candidate's role

I conceived the study, acquired funding (Principal Applicant Wellcome Trust grant WT081613/Z/06/Z), wrote the protocol and case record forms, and prepared all the paperwork for ethical approval in the UK and Vietnam. I was the Principal Investigator of the study and directly supervised all aspects of study implementation in Vietnam, including field staff appointment and training, the preparation of Standard Operating Procedures,

data management, and finances. Laboratory assays were conducted by trained laboratory personnel under the supervision of our Unit immunologist Dr Annette Fox.

Chapter 3

Contribution of research papers *(cohort study)*

This chapter briefly introduces the research papers compiled in this thesis that arose from the cohort study and describes how they form an original contribution. Three papers are included, of which I am the first author on two and senior author on one. I am the corresponding author on all three papers.

3.1 Outline of research papers

3.1.1 Research paper 1

Research paper 1 is currently in press as Horby et al. (2012) and describes the main results of the cohort study, focusing on the estimated incidence of influenza infection and illness by season, subtype and age. The paper also presents an analysis of risk factors for influenza infection. This paper is only the second published longitudinal, household-based study of influenza infection and disease from a tropical country; the other being the paper by Riley et al. (2011) that looked only at pandemic H1N1. Research paper 1 is therefore the first publication to assesses the incidence and risk factors for both inter-pandemic and pandemic influenza in a community in the tropics.

3.1.2 Research paper 2

Research paper 2 was published as Horby et al. (2011) and describes the results of the social contact and mobility survey that was conducted in the cohort in order to provide data for estimating influenza transmission probabilities by age and gender. This paper represents the first published data from either a developing country or a tropical country on social contact patterns relevant to the transmission of influenza. As such the work represents an important contribution to the literature on social contact patterns and infectious disease transmission. It is also the first publication to examine contact patterns in a household structured design.

3.1.3 Research paper 3

Research paper 3 was published as Boni et al. (2009) and describes the results of a mathematical modelling exercise to predict the burden and spread of pandemic H1N1 in Vietnam that was parametised using the social contact and mobility data presented in research paper 2, and other data assembled by the authors. This paper demonstrates the practical application of the data collected through the cohort study to predict the epidemiology of influenza in Vietnam, and is the first published mathematical model of influenza epidemiology in Vietnam.

Chapter 4

Research paper 1

Title: The epidemiology of inter-pandemic and pandemic influenza in Vietnam, 2007-2010: the Ha Nam household cohort study.

Author(s): Peter Horby, Le Quynh Mai, Annette Fox, Pham Quang Thai, Nguyen Thi Thu Yen, Le Thi Thanh, Nguyen Le Khanh Hang, Tran Nhu Duong, Dang Dinh Thoang, Jeremy Farrar, Marcel Wolbers, Nguyen Tran Hien.

Journal/Publisher: American Journal of Epidemiology

Type of publication: Original contribution

Stage of publication: In press

Academic peer-reviewed: Yes

Copyright: Open access publication charges have been paid.

Candidate's role: All the data presented in this paper were collected through the community cohort study of which I was the principal investigator. I prepared the dataset and wrote the first draft of the analysis plan, which was then further developed in discussion with a biostatistican Marcel Wolbers. Marcel Wolbers wrote and implemented the R-code for the statistical analysis. I reviewed and commented on all the statistical analysis outputs. I wrote the first and all subsequent drafts of the manuscript, submitted the manuscript for publication, and responded to all reviewers comments.

Candidate's signature:

Supervisor or senior author's signature to confirm Candidates role:

Title: The epidemiology of inter-pandemic and pandemic influenza in Vietnam, 2007-2010: the Ha Nam household cohort study.

Authors

Peter Horby, Le Quynh Mai, Annette Fox, Pham Quang Thai, Nguyen Thi Thu Yen, Le Thi Thanh, Nguyen Le Khanh Hang, Tran Nhu Duong, Dang Dinh Thoang, Jeremy Farrar, Marcel Wolbers, Nguyen Tran Hien.

Corresponding author: Peter Horby

BSc, MBBS, MSc, DTM&H, MRCP, FFPH

Oxford University Clinical Research Unit,

National Hospital of Tropical Diseases, 78 Giai Phong Street, Dong Da District, Hanoi, Vietnam

Fax: + 84 4 3576 4319

Phone: + 84 91 233 8567

Email: peter.horby@gmail.com

ABSTRACT

Prospective community-based studies have provided fundamental insights into the epidemiology of influenza in temperate regions but few comparable studies have been undertaken in the tropics. The authors conducted prospective influenza surveillance and intermittent seroprevalence surveys in a household-based cohort in Vietnam between December 2007 and April 2010, providing 1,793 person-seasons of influenza surveillance. Age and gender standardized estimates of the risk of acquiring any influenza infection per season in persons aged ≥5 years were 21.1% (95% confidence interval: 17.4, 24.7) in season 1, 26.4% (95% confidence interval: 22.6, 30.2) in season 2, and 17.0% (95% confidence interval: 13.6, 20.4) in season 3. Some individuals experienced multiple episodes of infection in the same season with different influenza types/subtypes (n=27), or re-infection with the same subtype in different seasons (n=22). The highest risk of influenza infection was in the age group 5-9 years, where the risk of influenza infection per season was 41.8%. Although the highest infection risk was in school-aged children, there were important heterogeneities in the age of infection by sub-type and season that may influence the impact of school closure and childhood vaccination on influenza transmission in tropical areas such as Vietnam.

Key words: Influenza, human; epidemiology; disease transmission, infectious; tropical climate; prevention and control.

INTRODUCTION

Until relatively recently influenza had been conceptualized as a developed country problem, with little consideration given to the frequency and burden of influenza in low-income and tropical countries. This all changed with the widespread reemergence of highly pathogenic avian influenza A/H5N1 in 2004, and further intensified with analyses suggesting that Southeast Asia may be the region where H3N2 influenza viruses undergo evolution before subsequent spread to the Northern and Southern hemispheres (1,2). As a result, influenza surveillance and control in Southeast Asia has come to be perceived as a global public health good, with substantial investment in influenza surveillance, anti-viral stockpiling, vaccine development, and epidemic preparedness (3,4).

Despite this interest in influenza in Southeast Asia, very little data are available on influenza transmission at the community level. In order to plan responses to both seasonal and pandemic influenza and the optimal application of interventions such as vaccination, antiviral prophylaxis, or school closure, it is first necessary to have a detailed understanding of how influenza is transmitted within the community (5). A key source of information on the transmission behavior of influenza has been community-based studies with follow up over several years. These studies have provided important insights into the epidemiology of respiratory infections and crucial information for the design of public health interventions. They have demonstrated that pre-school and school age children have the highest rate of respiratory illnesses (6–9); that mothers have higher attack rates than fathers (6,10); that children play an important role in introducing infection into families (9,11–13); and that there are high rates of serological evidence of infection without corresponding disease (5,14,15). However, those studies largely took place between the 1940's and early 1980's in the U.S., and very few comparable community or household-based studies have been undertaken in the tropics. Population densities, family structures, behaviors, mobility, material conditions,

health and climate are different in Southeast Asia compared to the U.S., so the epidemiology of influenza may also differ in important ways. Studies of influenza in Southeast Asia have largely assessed the incidence of clinical illness at health-care facilities, or analysis of surveillance and health care utilization data (16–20). Such studies are important in quantifying the clinical burden of influenza but cannot provide a full understanding of the epidemiology and transmission of influenza. Recent studies of pandemic influenza A/H1N1 have mostly relied on single cross-sectional serology to infer infection rates, but this is a less robust method of identifying recent infection than the detection of increases in antibody titers in paired sera (21). The authors therefore established a household-based cohort with the objective of quantifying the burden of influenza infection in a semi-rural community of northern Vietnam and to gain insights into the epidemiology of influenza in the tropics.

MATERIALS AND METHODS

Study design and setting

A full description of the materials and methods are provided in the Web Appendix and only a brief description is provided here. In 2007 a prospective, household-based community cohort was established in Thanh Ha Commune, Thanh Liem District, Ha Nam Province, Vietnam. The primary sampling unit of study was the household and all households in the Commune were eligible for inclusion in the study. Households were randomly selected from a list of all households using a random number table. If a randomly selected household declined to participate, the next nearest household was approached until a household was successfully recruited. All permanent residents in the household were eligible for inclusion and were requested to participate.

Blood sampling

Participants aged 5 years and older (at time of sampling) were asked to provide blood at recruitment and at three further time points. Recruitment blood samples were drawn between 1st-7th December 2007 (bleed 1). Subsequent bleeds took place between 9th-15th December 2008 (bleed 2), 2nd-4th June 2009 (bleed 3), and on the 3rd April 2010 (bleed 4). The bleeding time points were not decided *a priori* but were chosen when national influenza surveillance data indicated that influenza circulation was minimal. The four sets of samples provided three sets of paired sera.

Influenza-like illness surveillance

Trained hamlet Health Workers undertook weekly active surveillance of each participating household for episodes of influenza-like illness (ILI) and for changes in household composition. ILI was defined as 'as an illness with oral temperature of 38°C or

more and either a cough or a sore throat'. Any participant reporting an ILI was asked to provide a nose swab and a throat swab, and complete a 10-day symptom diary.

Definition of exposure and outcome variables

For the purpose of analysis, an influenza 'season' was defined as the period between consecutive bleeds, and an influenza 'transmission period' was defined as the period when influenza was known to be circulating on the basis of RT-PCR confirmed clinical cases.

'Influenza infection' was defined as either the detection of influenza RNA in a swab sample by reverse transcription polymerase chain reaction (RT-PCR) or a four fold or greater rise in HI antibody titre in paired sera, with the second titre at least 1:40. If paired sera were not available, a single high titre of at least 1:160 for seasonal influenza, or a titre of ≥ 1:80 in someone aged under 40 years for pandemic influenza H1N1, was also considered to indicate recent 'influenza infection'.

'Influenza illness' was defined as the detection of influenza-specific RNA in a swab by RT-PCR and the reporting of an ILI, or serological evidence of recent influenza infection (see above) plus an ILI episode occurring during a known period of transmission of the relevant influenza subtype. For linking serological evidence of recent influenza A infection to specific ILI episodes, influenza 'transmission periods' were defined based on periods of detection of RT-PCR confirmed influenza.

Laboratory methods

Detection of influenza viruses was performed on all nasal- and throat-swab specimens using RT-PCR. Influenza hemagglutination inhibition (HI) assays were performed using standard methods. Samples that were negative by HI assay in the lowest dilution (1:10) were assigned a titre of 1:5 for the purposes of computing seroconversion.

Statistical methods

Absolute observed risks of ILI (for subjects under ILI surveillance) and of influenza infection (for subjects under influenza infection surveillance) were calculated per season. Participants were considered under ILI surveillance for a particular season if they were under weekly ILI surveillance throughout the influenza transmission period and were considered under influenza infection surveillance if they additionally contributed a post-season blood sample.

Survey analysis methodology was used to derive risk estimates and associated 95% confidence intervals standardized to the age and gender structure of the Vietnamese rural population based on the 2009 Population and Housing Census. This provides valid inference accounting for effects of the survey design, which was based on cluster sampling by household, and biases in the provision of blood samples. As children under 5 years of age were not asked to give blood samples, standardization for influenza risks was to the census population aged ≥5 years. Standardization was implemented by raking, i.e. post-stratification on the target age and gender distribution in turn until convergence (22).

Seven potential risk factors for influenza infection were pre-defined. To assess these factors, data were pooled over all three seasons and the overall risk of an influenza infection was modeled with a logistic mixed effects model depending on the season, a random household effect (to account for potential clustering within households), a random subject effect (to account for potential within-subject correlation between seasons) and the respective risk factors.

All analyses were performed with the statistical software R 2.10.1 (R Foundation for Statistical Computing, Vienna, Austria) and the companion R packages survey 3.22-3 (for survey sampling) and lme4 0.999375-35 (for mixed models) (23).

RESULTS

940 individuals in 270 households were recruited from a study base of 2127 enumerated households. The household refusal rate was approximately 10% but a record was not kept of the number of refusals or reasons for refusal. The baseline characteristics of the 940 individuals and 270 households are shown in table 1. None of the participants had ever received influenza vaccination. The age distribution of the cohort was significantly different from that of both Ha Nam province and the national rural population (chi-square tests; both $P<0.001$). This was largely due to an over-representation of 10-19 year olds and an under-representation of 20-34 year olds in the cohort (Web Figure 1). The household size distribution of the cohort matched well with that of the Red River Delta rural population (chi-square goodness of fit test: $P=0.86$).

The cohort was studied for three consecutive influenza seasons from December 2007 through April 2010. Data on age was absent from 11 of the original cohort of 940 so these were excluded from all further analysis, and three children were born into participating households during the study, to give a total cohort size of 932 people (Figure 1). Some participants were absent from the study site during periods of influenza transmission and were therefore excluded from analysis for the relevant season. Figure 1 shows the number of participants included in each season's analysis; a total of 1,793 person-seasons of influenza surveillance were available. The completeness of bleeds varied by age and gender, the highest provision of blood being females in bleed one (85%), and the lowest being males in bleed two (55%) (Web Figure 2).

The temporal relationship between periods of ILI and RT-PCR confirmed influenza activity and the bleeding time points are shown in Figure 2. Three clear peaks of influenza A activity were detected: summer 2008 (influenza transmission period 1: 01/07/2008-30/09/2008), spring 2009 (influenza transmission period 2: 01/04/2009-05/06/2009), and

autumn 2009 (influenza transmission period 3: 01/09/2009-31/12/2009). Clear peaks in influenza B activity were not seen. Co-circulation of influenza B, H1N1, and H3N2 was detected in summer 2008 and in spring 2009.

Table 2 shows the number of reported ILI episodes detected in participants under ILI surveillance, the age and gender standardized ILI risk per season, and the number of influenza RT-PCR positive swabs in participants reporting an ILI. The standardized risk of ILI per season ranged from 14.1% in season 1 to 4.9% in season 3, and the maximum risk of RT-PCR confirmed influenza illness occurred amongst 5-19 year olds in season 3, where 5% of this age-group were affected. The influenza A virus strains detected in the cohort during the three seasons were: Season 1 - A/H1N1/Brisbane/59/2007-like, A/H3N2/Brisbane/10/2007-like; Season 2 - A/H1N1/Brisbane/59/2007-like, A/H3N2/Perth/16/2009-like; Season 3 - A/H1N1/California/7/2009-like. There was co-circulation of both influenza B Yamagata lineage and Victoria lineage in seasons 1 and 2, but with a predominance of Yamagata lineage in season 1 and Victoria lineage in season 2. In seasons 1 and 2 the overall rate of successful detection of influenza viruses from respiratory swabs was 18.4%, with the detection rate being greatest in children aged 0-4 years (50%) and declining with age, to 8.9% in those aged 40 years and over.

Unadjusted and standardized estimates of influenza infection and influenza illness rates per season are shown in Table 3. This analysis is restricted to those participants who were under ILI surveillance and also provided at least an end-of-season blood sample (figure 1). Standardized estimates of the risk of acquiring any influenza infection per season in persons aged >=5 years were 21.1% in season 1, 26.4% in season 2, and 17% in season 3. H3N2 infection was more common in season 2 compared to season 1 following a change in the circulating virus strain from A/H3N2/Brisbane/10/2007-like to A/H3N2/Perth/16/2009-like.

427 participants could be assessed for influenza infection over all three seasons, of which 242 (56.7%) showed evidence of at least one acute influenza infection over the whole study period. After adjustment for the household-based sampling design and standardization to the age and gender structure of the Vietnam rural population, the estimated risk of any influenza infection in people aged >=5 years over the entire three season period was 55.4% (95% CI 49.6-61.2%). In all seasons the estimated influenza illness risks were substantially lower than infection risks. The percentage of identified influenza infections in which an influenza illness was also detected was: H1N1 14% (13/95); H3N2 11% (11/97); B 15% (21/137); H1N1/2009 16% (17/109).

Multiple episodes of influenza infection by different influenza types/subtypes in the same season (multiple infections) were identified in 27 individuals; 23 had evidence of infection by two types/subtypes, and 4 had evidence of infection by all 3 types/subtypes (see footnotes Table 3). Re-infection with the same influenza type/subtype in season 2 as in season 1 was detected in 8 participants (H3N2=1 older adult [≥40 years]; seasonal H1N1=2 children aged 5-14 years; B=4 children aged 10-14 and 1 older adult). In addition, 14 participants who had been infected with seasonal H1N1 in season 1 (1 child aged 5-9 and 3 adults) or season 2 (6 children aged 5-14 and 4 adults) were also infected with pandemic H1N1 in season 3.

Influenza infection risk varied by age most clearly for season 3 when pandemic influenza H1N1 first circulated in the cohort and infected a large proportion of children and young adults (Figure 3). Age patterns in infection risk were less marked for inter-pandemic strains. The highest infection risk for H3N2 (seasons 1 and 2) and H1N1 (season 1 only) was in children aged 5-9 years. In season 2 the highest risk of H1N1 infection was in people aged 10-19 years. To assess the significance of the apparent age-dependent peaks of H3N2 infections in season 2 we applied the same methodology as we did for table 4 (without a random effect for patient because there is only one patient record per season). In season 2,

significantly higher H3N2 risks compared to the age group 20-40 years were observed in ages <10 years (OR 3.47; 95% CI 1.37, 8.79; *P*-value=0.009) and 10-20 years (OR 2.3; 95% CI 1.02, 5.17; *P*-value=0.043). The second peak for ages >=40 years is borderline significant with OR 2.12 (95% CI 0.99, 4.54; *P*-value=0.052). Web Figure 3 shows the proportion of participants with influenza infection per season by age group and type/subtype compared to previously publish household-based cohort studies.

Risk factors for influenza infection were explored in univariate and multivariate analysis (Table 4). Age was significantly associated with the risk of influenza infection in both univariate and multivariate analysis. This association was also observed for inter-pandemic H1N1, H3N2, and pandemic H1N1 when analyzed separately, but not for influenza B (see footnote Table 4). The highest risk of influenza infection was in the age group 5-9 years, where the observed absolute risk of influenza infection per season was 41.8%. The lowest infection risk was in the 20-39 year age group. There was no observed gender effect (Table 4, Web Figure 4) and no other covariates were significantly associated with influenza infection risk in either univariate or multivariate analysis.

DISCUSSION

This study is one of the first to prospectively quantify the incidence of influenza infection in the same individuals over multiple seasons in a tropical setting. It demonstrates that influenza infection is common, with an estimated risk of influenza infection in a single season of between 17% and 26%, and around 57% of people experiencing at least one acute influenza infection over a three-year period. These estimates are minimum estimates since we used the HI assay, which is less sensitive than the microneutralization assay, and conservative definitions of laboratory evidence of influenza infection (24). Although varying study designs, laboratory methods, data availability, and differing periods of influenza emergence and re-emergence confound direct comparison with earlier family studies in temperate settings, the levels of infection we identified are similar, as shown in Web Figure 3 (9,10,14,15,25–29). Although the rates we observed are generally in the lower range of those reported from other household studies, most of these previous studies recruited only households with infants or young children and did not standardize the results to the general population structure. For example, 5-14 year olds constituted 47% of the Cleveland family study during the H2N2 pandemic (14). Also, we were not able to obtain blood samples from children aged less than five years old, who are expected to have high rates of infection. The 17% infection rate for pandemic influenza H1N1 in our study is similar to contemporary seroepidemiology reports from other areas (30–34).

As found in other longitudinal studies, multiple infections in the same season with different influenza types/subtypes, and re-infection with the same subtype in consecutive seasons do occur and, although more common in children, can occur at any age (8,15,35–38).

Between 11% and 16% of influenza infections resulted in an illness that was detected by weekly active ILI surveillance. Our figures of the proportion of infections that cause

clinically detected illness are lower than estimates obtained by Monto et al, where at least 15-25 per cent of H3N2 infections and 19-34 per cent of type B infections resulted in clinical disease (39). Whilst this may be a real effect, perhaps influenced by the slightly greater proportion of participants aged >=40 years in our study compared to the Tecumseh study, it may also represent a reporting bias, with a greater propensity for participants to report illnesses in Tecumseh in the 1970's than in our study site.

The data reveal clear variations in the risk of influenza by age and subtype. As observed elsewhere, the 2009 H1N1 pandemic resulted in very high infection rates in young children that dropped sharply with age. The high rates of H1N1/2009 in 5-29 year olds, which exceed those seen for any other subtype and season, may be explained by the immunological naivety of this age group to this antigenic HA variant. The low pandemic H1N1 infection rates in older adults indicate that long-lived and cross-protective immunity against H1N1 may be induced either by repeated infection or by infection with an antigenically related virus (40,41). Similar long-lived protection was observed when H1N1 re-emerged in 1977 after an absence of 20 years (26,39).

H3N2 infections in season 2 (when a drifted variant circulated) peaked in school-aged children, with a second peak in older adults. A recent cross-sectional sero-prevalence study from Canada also found a second peak in H3N2 titers in people aged over 60 years (42), and there is ample evidence that adults experience higher rates of H3N2 infection and reinfection compared to other influenza types/subtypes (28,43–47). These serological measures of risk also translate into clinical illness, with H3N2 more commonly causing clinical illness in adults in the community and in institutional care compared to other influenza viruses (48–52). We observed a fairly constant risk of influenza B infection across the whole age range. This contrasts to some earlier studies in temperate countries where influenza B risk peaked in pre-school or school-age children (15,26,27,29). Whilst this pattern may be due to the absence of

an influenza B epidemic during the study period, the influenza B infection rates were moderately high and the age distribution may therefore be the consequence of prolonged circulation of influenza B viruses without the opportunity for a build-up of a large cohort of susceptible children.

One possible explanation for the more even age-distribution of the risk of inter-pandemic influenza infection in our study compared to some historic studies in temperate climates is that there is less intense seasonal and school-based forcing in the tropics, resulting in less intense school based transmission and greater community based transmission. Multiple epidemics per year and more prolonged virus circulation may also limit the size of the pool of susceptible children. In this respect it is relevant that, in contrast to community studies conducted in temperate settings, we did not identify an increased risk of influenza infection in women compared to men, nor an association between caring for children or the presence of a school aged child in the house and the risk of influenza in adults. This may have important implications for the impact of school closure and childhood vaccination on the transmission of inter-pandemic influenza in tropical areas such as Vietnam. Whilst school-closure was reported to be effective in reducing transmission of H1N1/2009 in Hong Kong, our study shows that the age distribution of H1N1/2009 infection was not characteristic of inter-pandemic influenza in the tropics (53).

Longitudinal studies such as we have presented here, which follow individuals of all ages over multiple seasons with serial serology, provide not only the most robust estimates of true influenza infection incidence, they also provide information that is critical for understanding influenza epidemiology in the tropics and for planning effective influenza control strategies.

Authors' affiliations

Oxford University Clinical Research Unit and Wellcome Trust Major Overseas Programme, Vietnam (Peter Horby, Annette Fox, Jeremy Farrar, Marel Wolbers); National Institute of Hygiene and Epidemiology, Hanoi, Vietnam (Le Quynh Mai, Pham Quang Thai, Nguyen Thi Thu Yen, Le Thi Thanh, Nguyen Le Khanh Hang, Tran Nhu Duong, Nguyen Tran Hien); Ha Nam Centre for preventive Medicine, Ha Nam, Vietnam (Dang Dinh Thoang).

Financial Disclosures

This work was supported by the Wellcome Trust UK (grants 081613/Z/06/Z and 077078/Z/05/Z). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We are grateful to the community of An Hoa Commune for agreeing to participate in this study and for providing their time. We would like to thank the hamlet health workers who conducted the interviews and surveillance. We also wish to thank the Ministry of Health of Vietnam for their continuing support of the research collaboration between the Oxford University Clinical Research Unit and the National Institute for Hygiene and Epidemiology. We are grateful to Dr. Neal Alexander for providing advice on study design and data analysis.

Competing interests

The authors have declared that no competing interests exist.

REFERENCES

1. Rambaut A, Pybus OG, Nelson MI, et al. The genomic and epidemiological dynamics of human influenza A virus. *Nature*. 2008;453(7195):615-619.
2. Russell CA, Jones TC, Barr IG, et al. The global circulation of seasonal influenza A (H3N2) viruses. *Science*. 2008;320(5874):340-346.
3. Hoa LK, Hiep LV, Be LV. Development of pandemic influenza vaccine production capacity in Viet Nam. *Vaccine*. 29 Suppl 1:A34-6.
4. Hanvoravongchai P, Adisasmito W, Chau PN, et al. Pandemic influenza preparedness and health systems challenges in Asia: results from rapid analyses in 6 Asian countries. *BMC Public Health*. 10:322.
5. Monto AS. Studies of the community and family: acute respiratory illness and infection. *Epidemiologic reviews*. 1994;16(2):351-73.
6. Badger GF, Dingle JH, Feller AE, et al. A study of illness in a group of Cleveland families. II. Incidence of the common respiratory diseases. *Am J Hyg*. 1953;58(1):174-178.
7. Hope-Simpson RE, Higgins PG. A respiratory virus study in Great Britain: review and evaluation. *Prog Med Virol*. 1969;11:354-407.
8. Fox JP, Hall CE, Cooney MK, et al. The Seattle virus watch. II. Objectives, study population and its observation, data processing and summary of illnesses. *Am J Epidemiol*. 1972;96(4):286-305.
9. Monto AS, Kioumehr F. The Tecumseh Study of Respiratory Illness. IX. Occurrence of influenza in the community, 1966-1971. *Am J Epidemiol*. 1975;102(6):553-563.
10. Hall CE, Brandt CD, Frothingham TE, et al. The virus watch program: a continuing surveillance of viral infections in metropolitan New York families. IX. A comparison of infections with several respiratory pathogens in New York and New Orleans families. *Am J Epidemiol*. 1971;94(4):367-385.
11. Badger GF, Dingle JH, Feller AE, et al. A study of illness in a group of Cleveland families. IV. The spread of respiratory infections within the home. *Am J Hyg*. 1953;58(2):174-178.
12. Cauchemez S, Carrat F, Viboud C, et al. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat Med*. 2004;23(22):3469-3487.
13. Viboud C, Boelle PY, Cauchemez S, et al. Risk factors of influenza transmission in households. *Br J Gen Pract*. 2004;54(506):684-689.

14. Jordan W. S. J, Denny F. W. J, Badger GF, et al. A study of illness in a group of Cleveland families. XVII. The occurrence of Asian influenza. *Am J Hyg*. 1958;68(2):190-212.
15. Hall CE, Cooney MK, Fox JP. The Seattle virus watch. IV. Comparative epidemiologic observations of infections with influenza A and B viruses, 1965-1969, in families with young children. *Am J Epidemiol*. 1973;98(5):365-380.
16. Wong CM, Yang L, Chan KPKHP, et al. Influenza-associated hospitalization in a subtropical city. *PLoS Med*. 2006;3(4):e121.
17. Simmerman JM, Uyeki M. The burden of influenza in East and South-East Asia : a review of the English language literature. *Influenza Other Respi Viruses*. 2008;2(3):81-92.
18. Hasegawa G, Kyaw Y, Danjuan L, et al. Influenza virus infections in Yangon, Myanmar. *J Clin Virol*. 2006;37(3):233-234.
19. Beckett CG, Kosasih H, Ma'roef C, et al. Influenza surveillance in Indonesia: 1999-2003. *Clin Infect Dis*. 2004;39(4):443-449.
20. Clague B, Chamany S, Burapat C, et al. A household survey to assess the burden of influenza in rural Thailand. *Southeast Asian J Trop Med Public Health*. 2006;37(3):488-493.
21. Lee VJ, Chen MI, Yap J, et al. Comparability of Different Methods for Estimating Influenza Infection Rates Over a Single Epidemic Wave. *Am J Epidemiol*. 2011;174(4):468-78
22. Lumley T. Complex Surveys: A Guide to Analysis Using R. Wiley; 2010.
23. R Development Core Team: R: A Language and Environment for Statistical Computing. (http://www.r-project.org/). (Accessed 26 January 2012).
24. Miller E, Hoschler K, Hardelid P, et al. Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *Lancet*. 2010;375(9720):1100-8.
25. Jordan W. S. J, Badger GF, Dingle JH. A study of illness in a group of Cleveland families. XVI. The epidemiology of influenza, 1948-1953. *Am J Hyg*. 1958;68(2):169-189.
26. Fox JP, Hall CE, Cooney MK, et al. Influenzavirus infections in Seattle families, 1975-1979. I. Study design, methods and the occurrence of infections by time and age. *Am J Epidemiol*. 1982;116(2):212-27.

27. Monto AS, Sullivan KM. Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiol Infect*. 1993;110(1):145-160.
28. Frank AL, Taber LH, Wells JM. Comparison of infection rates and severity of illness for influenza A subtypes H1N1 and H3N2. *J Infect Dis*. 1985;151(1):73-80.
29. Frank AL, Taber LH, Glezen WP, et al. Influenza B virus infections in the community and the family. The epidemics of 1976-1977 and 1979-1980 in Houston, Texas. *Am J Epidemiol*. 1983;118(3):313-25.
30. Miller E, Hoschler K, Hardelid P, et al. Incidence of 2009 pandemic infl uenza A H1N1 infection in England : a cross-sectional serological study. *Health Technology Assessment*. 2010;6736(09):1-9.
31. Wu JT, Ma ESK, Lee CK, et al. The infection attack rate and severity of 2009 pandemic influenza (H1N1) in Hong Kong. *Clin Infect Dis*. 2010;51(10):1184-91.
32. Chen MIC, Lee VJM, Lim W-Y, et al. 2009 influenza A(H1N1) seroconversion rates and risk factors among distinct adult cohorts in Singapore. *JAMA*. 2010;303(14):1383-91.
33. Deng Y, Pang XH, Yang P, et al. Serological survey of 2009 H1N1 influenza in residents of Beijing, China. *Epidemiol Infect*. 139(1):52-58.
34. Riley S, Kwok KO, Wu KM, et al. Epidemiological Characteristics of 2009 (H1N1) Pandemic Influenza Based on Paired Sera from a Longitudinal Community Cohort Study. *PLoS Med*. 8(6):e1000442.
35. Frank AL, Taber LH. Variation in frequency of natural reinfection with influenza A viruses. *J Med Virol*. 1983;12(1):17-23.
36. Frank AL, Taber LH, Wells JM. Individuals infected with two subtypes of influenza A virus in the same season. *J Infect Dis*. 1983;147(1):120-4.
37. Frank AL, Taber LH, Porter CM. Influenza B virus reinfection. *Am J Epidemiol*. 1987;125(4):576-586.
38. Sonoguchi T, Sakoh M, Kunita N, et al. Reinfection with influenza A (H2N2, H3N2, and H1N1) viruses in soldiers and students in Japan. *J Infect Dis*. 1986;153(1):33-40.
39. Monto AS. Tecumseh study of illness XIII. Influenza infection and disease, 1976-1981. *Am J Epidemiol*. 1985;121(6).
40. Duvvuri VR, Moghadas SM, Guo H, et al. Highly conserved cross-reactive CD4+ T-cell HA-epitopes of seasonal and the 2009 pandemic influenza viruses. *Influenza Other Resp Viruses*. 4(5):249-258.

41. Greenbaum J a, Kotturi MF, Kim Y, et al. Pre-existing immunity against swine-origin H1N1 influenza viruses in the general human population. *PNAS*. 2009;106(48):20365-70.
42. Skowronski DM, Hottes TS, McElhaney JE, et al. Immuno-epidemiologic correlates of pandemic H1N1 surveillance observations: higher antibody and lower cell-mediated immune responses with advanced age. *J Infect Dis*. 203(2):158-167.
43. Sohier R, Henry M. Epidemiological data on Hong Kong influenza in France. *Bull World Health Organ*. 1969;41(3):402-404.
44. Hope-Simpson RE. First outbreak of Hong Kong influenza in a general practice population in Great Britain. A field and laboratory study. *Br Med J*. 1970;3(5714):74-77.
45. Zdanov VM, Antonova IV. The Hong Kong influenza virus epidemic in the USSR. *Bull World Health Organ*. 1969;41(3):381-386.
46. Willers H, Hopken W. Epidemiology of influenza in Lower Saxony during the period 1968-1978 with particular emphasis on subtypes A(H3N2) and A(H1N1) in winter 1977-78. *Med Microbiol Immunol*. 1979;167(1):21-27.
47. Davis LE, Caldwell GG, Lynch RE, et al. Hong Kong influenza: the epidemiologic features of a high school family study analyzed and compared with a similar study during the 1957 Asian influenza epidemic. *Am J Epidemiol*. 1970;92(4):240-247.
48. Hope-Simpson RE. Age and secular distributions of virus-proven influenza patients in successive epidemics 1961-1976 in Cirencester: epidemiological significance discussed. *J Hyg (Lond)*. 1984;92(3):303-336.
49. Camacho A, Ballesteros S, Graham AL, et al. Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. *Proc Biol Sci*. 2011;278(1725):3635-43.
50. Glezen WP, Keitel WA, Taber LH, et al. Age distribution of patients with medically-attended illnesses caused by sequential variants of influenza A/H1N1: comparison to age-specific infection rates, 1978-1989. *Am J Epidemiol*. 1991;133(3):296-304.
51. Khiabanian H, Farrell GM, St George K, et al. Differences in patient age distribution between influenza A subtypes. *PLoS ONE*. 2009;4(8):e6832.
52. Lee BE, Mukhi SN, Drews SJ. Association between patient age and influenza A subtype during influenza outbreaks. *Infect Control Hosp Epidemiol*. 31(5):535-537.
53. Wu JT, Cowling BJ, Lau EH, et al. School closure and mitigation of pandemic (H1N1) 2009, Hong Kong. *Emerg Infect Dis*. 16(3):538-541.

Table 1 Characteristics of Participants and Households at Recruitment, Ha Nam, Vietnam, 2007-2010.

Category	Subcategory	x/n (%)
Age in years	0 to <5	83/929 (8.9%)
	5 to <10	70/929 (7.5%)
	10 to <20	209/929 (22.5%)
	20 to <40	246/929 (26.5%)
	40 to <60	241/929 (25.9%)
	≥60	80/929 (8.6%)
Female gender - yes		508/932 (54.5%)
Chronic diseases - yes		5/869 (0.6%)[a]
Adults (age ≥18; N=592):		
Caring for children at home or at work	Never	284/569 (49.9%)
	Sometimes	100/569 (17.6%)
	Most days	185/569 (32.5%)
Smoking – yes		107/560 (19.1%)
Cigarettes smoked per day	<= 5	49/103 (47.6%)
	6-10	45/103 (43.7%)
	11-20	9/103 (8.7%)
Households (N=270):		
Household size	1 person	28/270 (10.4%)
	2 people	41/270 (15.2%)
	3 people	65/270 (24.1%)
	4 people	74/270 (27.4%)
	5 people	42/270 (15.6%)
	>5 people	20/270 (7.4%)
Home crowding (>2 people/room) – yes		46/237 (19.4%)
School age children (age 5 to <18) in household - yes		156/264 (59.1%)

[a] 2 chronic lung diseases, 2 chronic heart diseases, and one chronic liver disease.

Table 2 ILI Episodes and Influenza Virus Detections by Season and Age Group, Ha Nam, Vietnam, 2007-2010.

		Age group at the beginning of the season (years)				
Season	Variable	0 to <5	5 to <20	20 to <40	≥40	All ages
1	Number of participants included	84	273	240	319	916
	Number reporting an ILI episode – x (%)	4 (4.8)	42 (15.4)	31 (12.9)	57 (17.9)	134 (14.6)
	Standardized ILI risk/season [a] - % (95% CI)	5.3 (0.3,10.2)	15.5 (10.5, 20.4)	11.7 (7.9, 15.4)	17.8 (13.5, 22.1)	14.1 (11.3, 16.8)
	Influenza A H1N1 PCR virus detections – x (%)	1 (1.2)	6 (2.2)	2 (0.8)	1 (0.3)	10 (1.1)
	Influenza A H3N2 PCR virus detections – x (%)	0 (0.0)	1 (0.4)	1 (0.4)	1 (0.3)	3 (0.3)
	Influenza B virus PCR detections – x (%)	1 (1.2)	4 (1.5)	1 (0.4)	3 (0.9)	9 (1.0)
2	Number of participants included	59	284	226	326	895
	Number reporting an ILI episode – x (%)	0 (0.0)	16 (5.6)	13 (5.8)	22 (6.7)	51 (5.7)
	Standardized ILI risk/season [a] - % (95% CI)	0.0	5.7 (2.9, 8.5)	4.6 (2.1, 7.1)	6.9 (3.8, 9.9)	5.2 (3.5, 6.9)
	Influenza A H1N1 PCR virus detections – x (%)	0 (0.0)	4 (1.4)	0 (0.0)	0 (0.0)	4 (0.4)
	Influenza A H3N2 PCR virus detections – x (%)	0 (0.0)	3 (1.1)	2 (0.9)	2 (0.6)	7 (0.8)
	Influenza B virus PCR detections – x (%)	0 (0.0)	1 (0.4)	0 (0.0)	0 (0.0)	1 (0.1)
3	Number of participants included	54	279	225	329	887
	Number reporting an ILI episode – x (%)[b]	3 (5.6)	21 (7.5)	11 (4.9	10 (3.0)	45 (5.1)
	Standardized ILI risk/season [a] - % (95% CI)	5.8 (0.0, 12.2)	7.5 (4.3, 10.6)	4.4 (1.4, 7.4)	2.7 (1.0, 4.3)	4.9 (3.1, 6.6)
	Pandemic influenza A virus detections – x (%)	0 (0.0)	14 (5.0)	7 (3.1)	3 (0.9)	24 (2.7)

a Standardized to age and gender distribution of Vietnamese national rural population (2009 Population and Housing Census).

b 5 subjects reported 2 ILI episodes during the season.

Table 3 Unadjusted and Standardized Risks of Influenza Infection and Influenza Illness by Season, Persons Aged ≥ 5 Years, Ha Nam, Vietnam, 2007-2010.

Season	Influenza type/subtype	Sero-conversions	Single high titres	Positive RT-PCR [a]	Observed influenza infections x/n (%)	Standardized [b] influenza infection risk % (95% CI)	Observed influenza illnesses x/n (%)	Standardized [b] influenza illness risk % (95% CI)
1 (n=555; 7 with single titres)	Any [c]	116	0	17 (4)	120/555 (21.6%)	21.1 (17.4, 24.7)	28/555 (5.0%)	4.5 (2.8, 6.3)
	H1N1	36	0	8 (4)	40/555 (7.2%)	7.4 (5.0, 9.8)	8/555 (1.4%)	1.3 (0.3, 2.3)
	H3N2	13	0	3 (0)	13/555 (2.3%)	2.3 (0.8, 3.8)	3/555 (0.5%)	0.6 (0.0, 1.3)
	B	69	0	6 (1)	70/555 (12.6%)	12.0 (8.8, 15.2)	17/555 (3.1%)	2.6 (1.3, 4.0)
2 (n=640; 139 with single titres)	Any [d]	152	23	12 (3)	178/640 (27.8%)	26.4 (22.6, 30.2)	17/640 (2.7%)	2.4 (1.2, 3.7)
	H1N1	46	7	4 (2)	55/640 (8.6%)	8.3 (6.1, 10.5)	5/640 (0.8%)	0.8 (0.1, 1.6)
	H3N2	71	12	7 (1)	84/640 (13.1%)	11.8 (9.0, 14.6)	8/640 (1.2%)	1.0 (0.2, 1.9)
	B	59	7	1 (1)	67/640 (10.5%)	10.2 (7.7, 12.7)	4/640 (0.6%)	0.6 (0.0, 1.1)
3 (n=598; 58 with single titres)	H1N1	98	6	18 (5)	109/598 (18.2%)	17.0 (13.6, 20.4)	17/598 (2.8%)	2.6 (1.3, 3.9)

a Numbers in brackets refer to PCR-positive samples without documented seroconversion or single high titre.
b Standardized to age and gender distribution of Vietnamese national rural population aged >=5 years (2009 Population and Housing Census).
c One subject had both influenza A H1N1 and influenza A H3N2 conversion; one subject had both influenza A H1N1 and influenza B conversion; one subject was H1N1 PCR+ (but no seroconversion) and had influenza B seroconversion.
d 4 subjects were infected by all 3 influenza subtypes, 5 by H1N1 and H3N2, 7 by H1N1 and B, 8 by H3N2 and B.

Table 4 Risk Factors for Influenza Infection, Aggregated Over Influenza Subtypes and Seasons, Ha Nam, Vietnam, 2007-2010.

		Aggregated [a] observed absolute influenza infection risk per season	Univariate association		Multivariate association [b]	
Covariate	Category	x/n (%)	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Age (years)	5 to <10	79/189 (41.8%)	3.65 (2.50, 5.34)	<0.001[c]	4.15 (2.41, 7.13)	<0.001
	10 to <20	120/408 (29.4%)	2.11 (1.52, 2.91)		2.34 (1.44, 3.79)	
	20 to <40	88/513 (17.2%)	- (baseline)		-	
	≥40	120/683 (17.6%)	1.06 (0.77, 1.45)		1.10 (0.77, 1.58)	
Gender[d]	Male	164/745 (22.0%)	-		-	
	Female	243/1048 (23.2%)	1.07 (0.85, 1.35)	0.55	1.06 (0.81, 1.38)	0.66
Household size	1-2	31/185 (16.8%)	0.64 (0.39, 1.04)	0.25	0.76 (0.43, 1.33)	0.81
	3	83/350 (23.7%)	-		-	
	4	138/590 (23.7%)	1.00 (0.71, 1.42)		0.98 (0.69, 1.40)	
	≥ 5	155/668 (23.2%)	0.98 (0.70, 1.39)		0.98 (0.61, 1.56)	
Home crowding (>2 people/room)	No	250/1132 (22.1%)	-		-	
	Yes	119/507 (23.5%)	1.10 (0.84, 1.44)	0.50	0.90 (0.60, 1.35)	0.62
Caring for children at work or home (for adults, age≥18)	No	84/549 (15.3%)	-		-	0.18
	Sometimes	36/201 (17.9%)	1.23 (0.77, 1.95)		1.40 (0.89, 2.21)	

	Most days	88/435 (20.2%)	1.40 (0.98, 2.01)	0.18	1.37 (0.93, 2.00)	
School age children (age 5 to <18) in household (for adults, age≥18)	No	83/462 (18.0%)	-		-	
	Yes	131/767 (17.1%)	0.94 (0.67, 1.32)	0.72	0.87 (0.61, 1.26)	0.47
Smoking (for adults, age≥18)	No	172/970 (17.7%)	-		-	
	Yes	32/200 (16.0%)	0.87 (0.56, 1.34)	0.53	0.89 (0.54, 1.45)	0.63

OR=Odds ratio, *P*-value corresponds to a Wald-type test for the significance of the whole factor.
a Aggregated over all seasons and influenza subtypes (but only pandemic flu assessed in season 3), x=# of influenza infection in subgroup, n=# of person-seasons observed.
b Adjusted for all other covariates in the model. Covariates that are reported in adults only for the univariate associations (e.g. caring for children at work) were included as indicator variables with value 0 for children.
c Also significant for H1N1 alone (p<0.001), H3N2 alone (p=0.02), pandemic H1N1 alone (p<0.001), but not influenza B (p=0.33).
d Univariate analysis of gender for adults only is also non-significant.

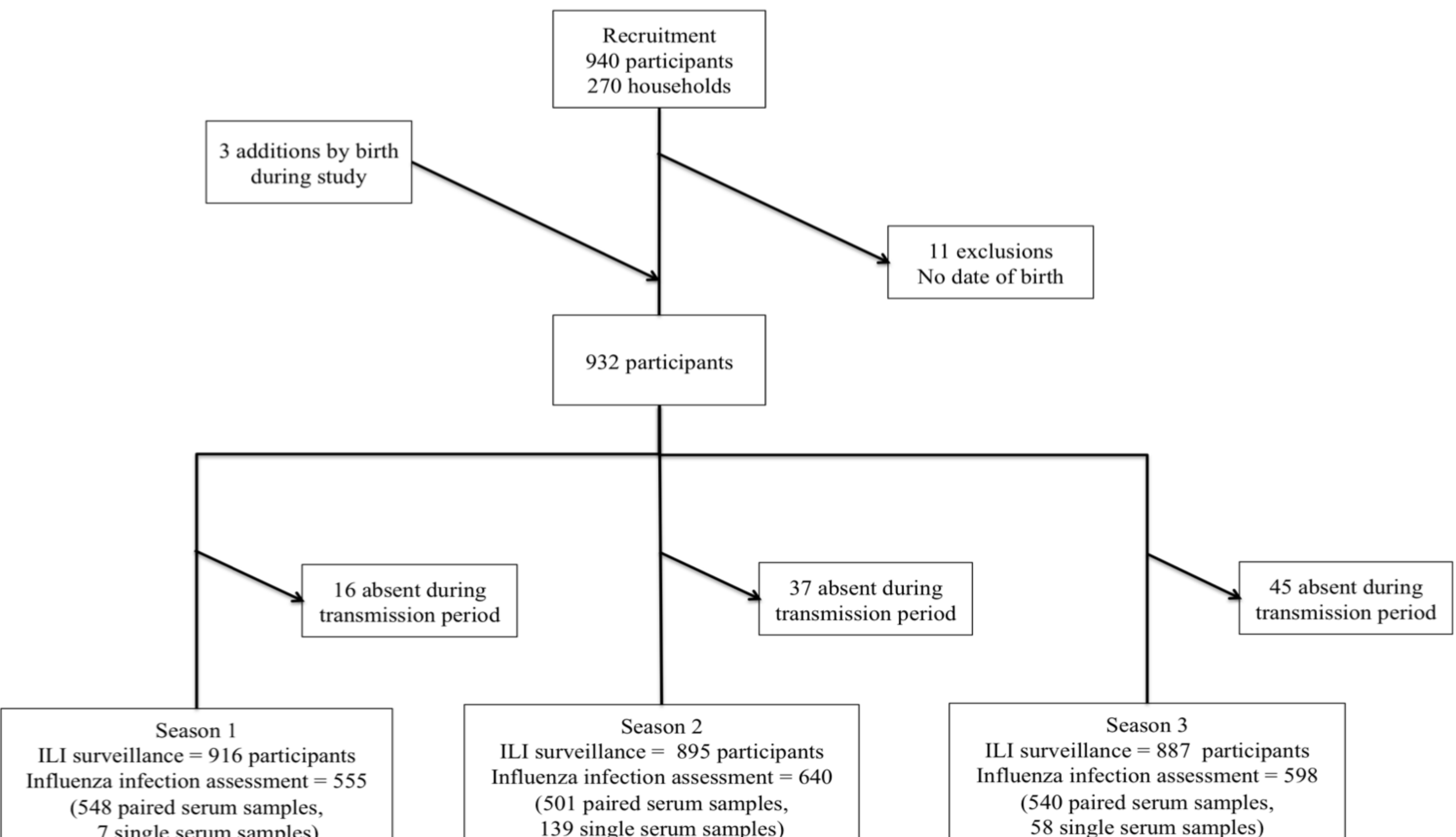


Figure 1. Number of study participants included in assessment of influenza illness and infection status by influenza season.

ILI = Influenza-Like-Illness

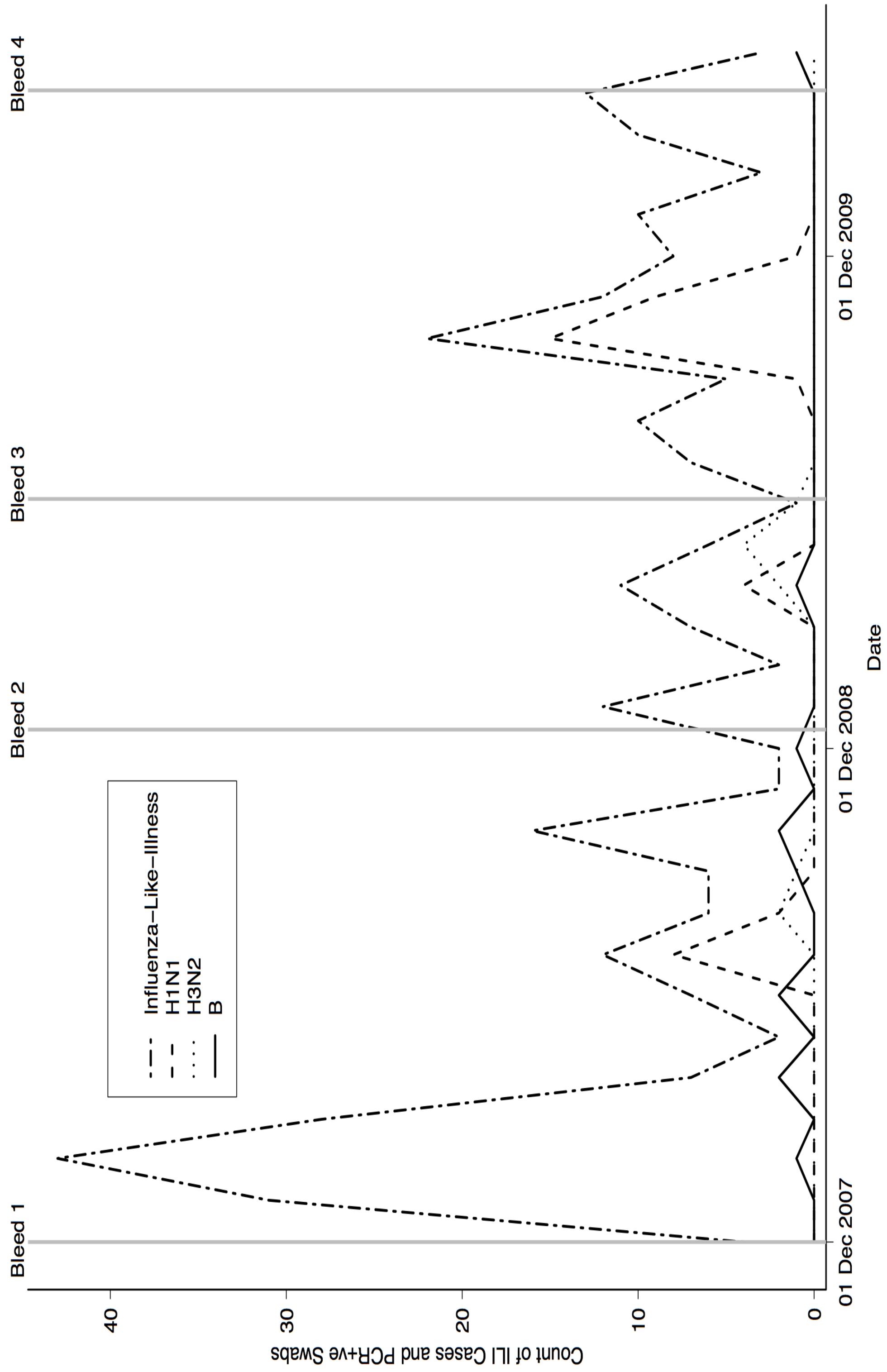


Figure 2. Timeline of ILI cases, RT-PCR confirmed influenza illnesses, and cross sectional bleeds.

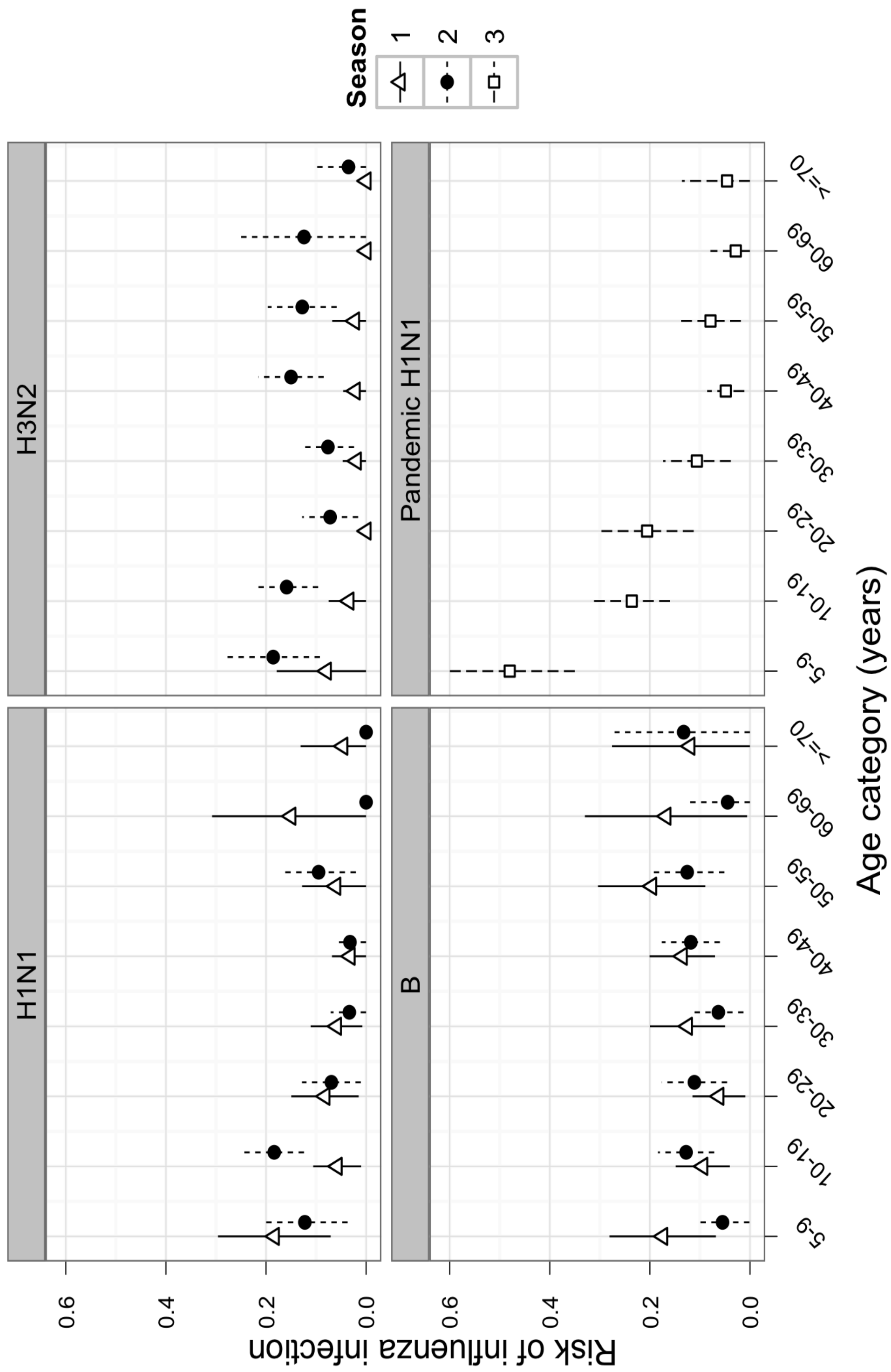
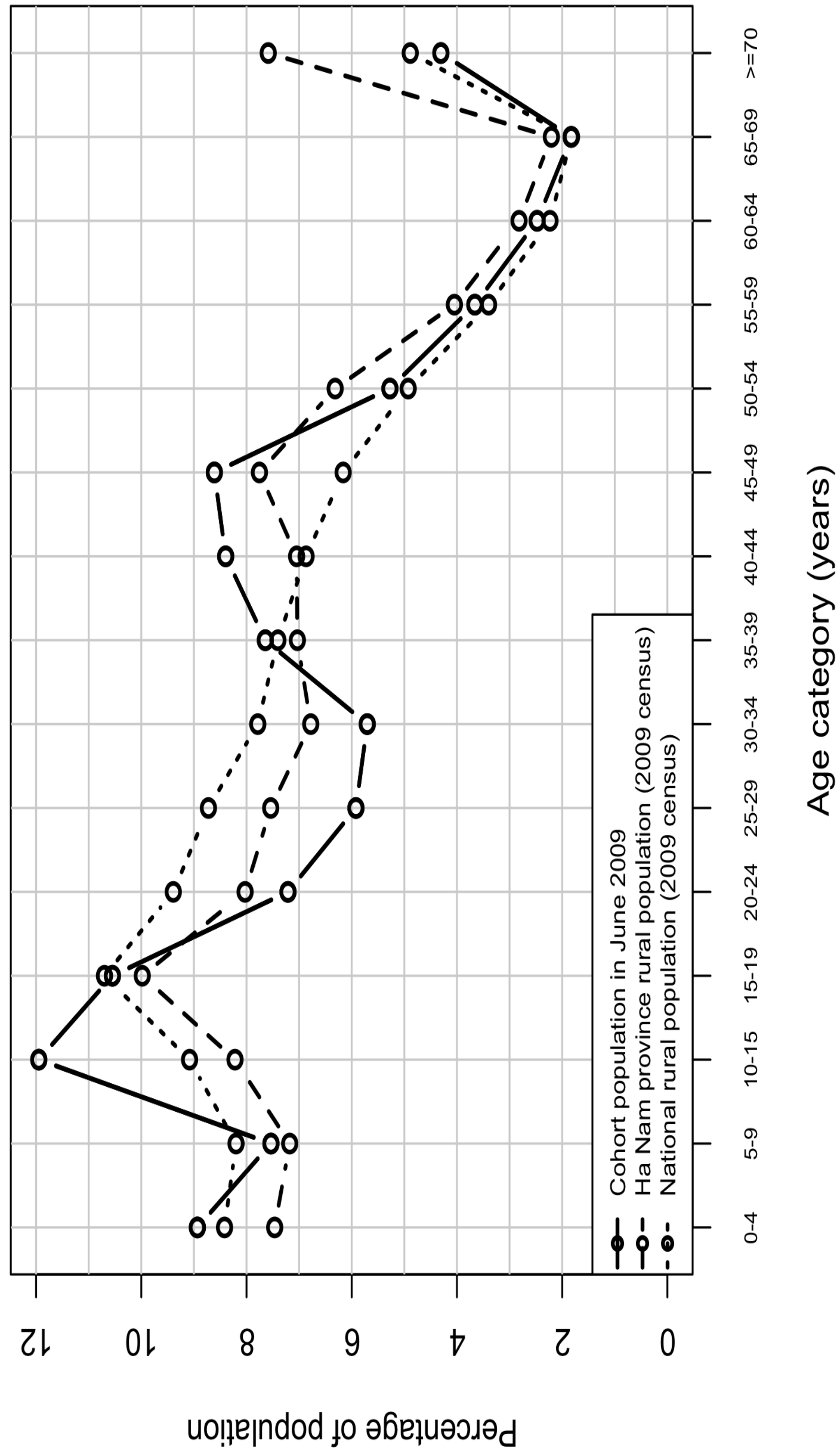


Figure 3. Risk of influenza infection by season, influenza sub-type, and age group.

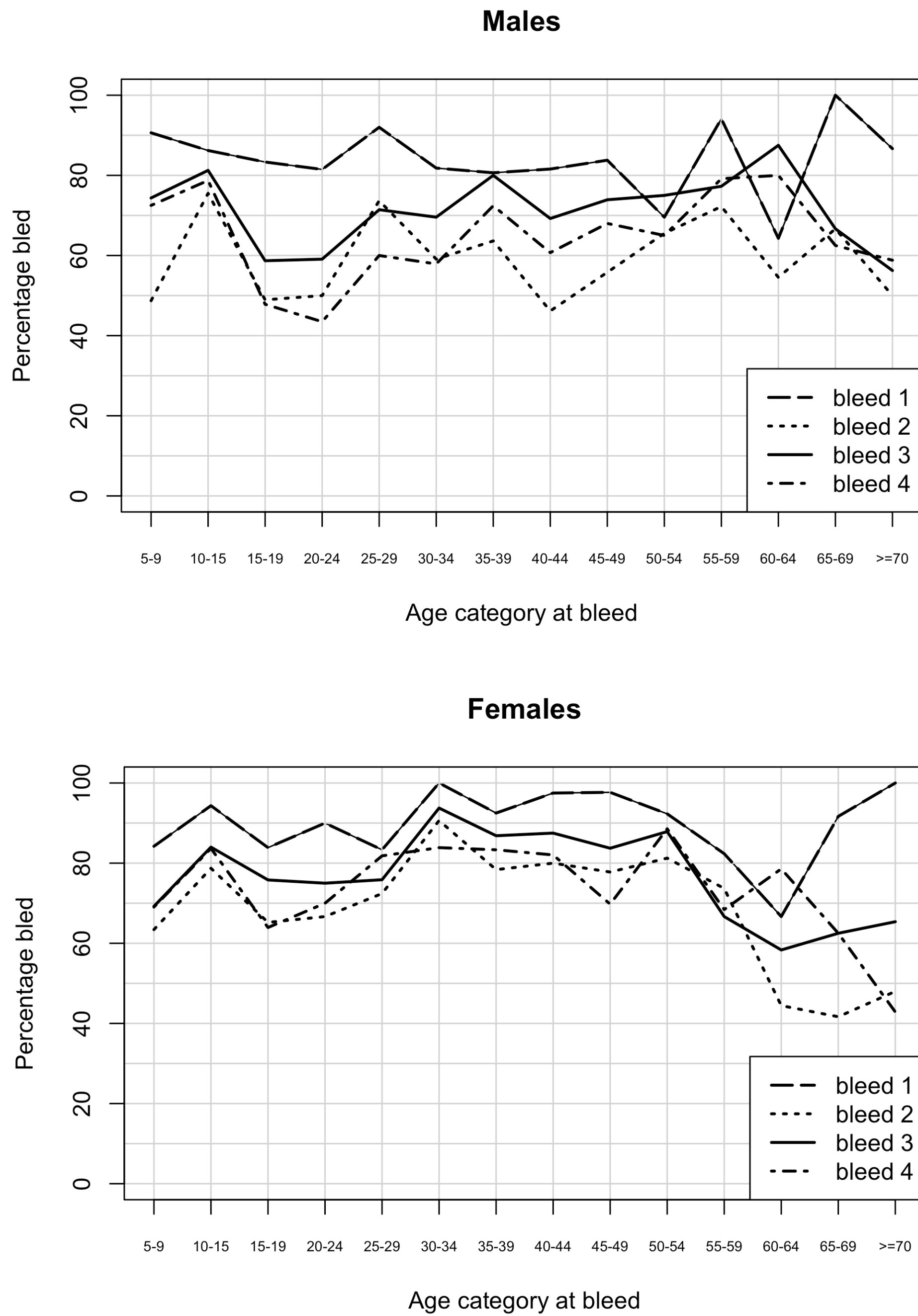
Adjusted for clustered design and standardised to age and gender distribution of Vietnam national rural population aged $\geq$ 5years.

Web Appendix

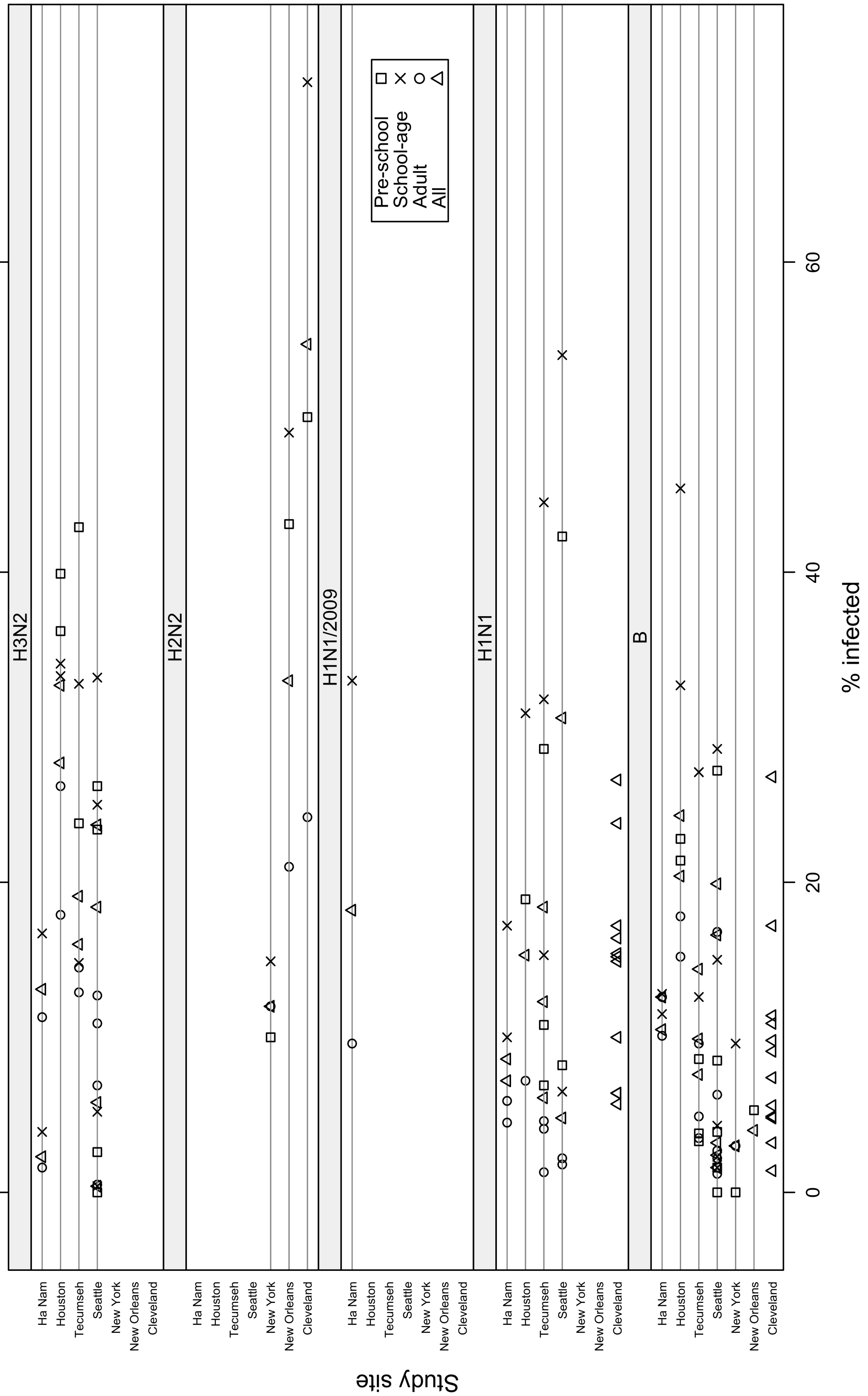


Web Figure 1. Age distribution of the cohort compared to provincial and national rural populations.

Age distribution of the cohort in June, 2009, compared to the age distribution of Ha Nam province and the national rural population as ascertained in the 2009 Population and Housing Census.



Web Figure 2. Frequency of bleeding amongst cohort participants under ILI surveillance, by age and gender. Ha Nam, Vietnam, 2007-2010.



Web Figure 3. Influenza infection rates in Ha Nam 2007-2010, compared to historic household cohort studies.

Web Table 1. Source data for Web Figure 3.

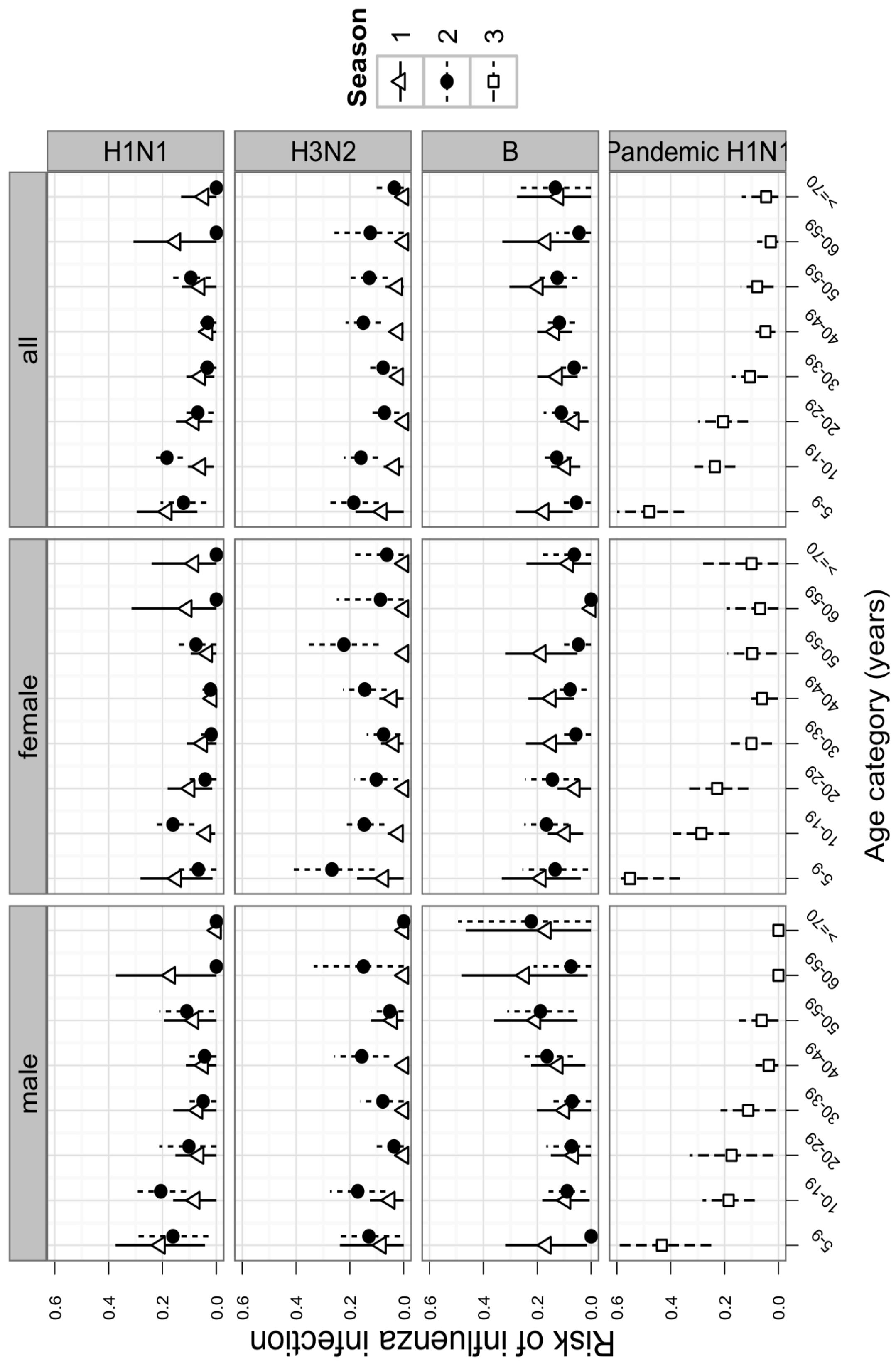
						Proportion infected by age group*				
Reference	**Study site**	**Year(s)**	**Type / subtype**	**Definition of infection**	**Outcome measure**	**Pre-school**	**School -age**	**Adult**	**All**	**Notes**
(4)	Cleveland	Fall 1947 - Spring 1948	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				15.2	
(4)	Cleveland	Spring 1948 - Fall 1948	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				17.2	
(4)	Cleveland	Fall 1948 - Spring 1949	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				16.4	
(4)	Cleveland	Spring 1949 - Fall 1949	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				15.4	
(4)	Cleveland	Fall 1949 - Spring 1950	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				14.9	
(4)	Cleveland	Spring 1950 - Fall 1950	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				6.4	
(4)	Cleveland	Fall 1950 - Spring 1951	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				23.8	
(4)	Cleveland	Spring 1951 - Fall 1951	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				6.4	
(4)	Cleveland	Fall 1951 - Spring 1952	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				10	
(4)	Cleveland	Spring 1952 - Fall 1952	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				5.7	
(4)	Cleveland	Fall 1952 - Spring 1953	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				26.6	
(4)	Cleveland	Spring 1953 - Fall 1953	H1N1	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				6.4	
(4)	Cleveland	Fall 1947 - Spring 1948	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				3.2	

(4)	Cleveland	Spring 1948 - Fall 1948	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				9.8	
(4)	Cleveland	Fall 1948 - Spring 1949	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				7.4	
(4)	Cleveland	Spring 1949 - Fall 1949	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				9.1	
(4)	Cleveland	Fall 1949 - Spring 1950	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				17.2	
(4)	Cleveland	Spring 1950 - Fall 1950	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				10.9	
(4)	Cleveland	Fall 1950 - Spring 1951	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				5.6	
(4)	Cleveland	Spring 1951 - Fall 1951	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				11.4	
(4)	Cleveland	Fall 1951 - Spring 1952	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				26.8	
(4)	Cleveland	Spring 1952 - Fall 1952	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				4.8	
(4)	Cleveland	Fall 1952 - Spring 1953	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				4.9	
(4)	Cleveland	Spring 1953 - Fall 1953	B	A 4-fold or greater rise in HI titer	% of persons with two samples taken at approximately six month intervals.				1.4	
(5)	Cleveland	1956-1957	H2N2	A 4-fold or greater rise in HI titer by CF or HI test, or both.	% of persons with 2 serum samples at approximately six month intervals, except for infants followed from birth whose first serum specimen at 12 to 18 months of age was taken to reflect their infection experience since birth.	50	71.6	24.2	54.7	Pandemic year
(6)	New York	1961-1965	H2N2	Seroconversion to positive or a 4-fold or greater rise in CF antibody titer	% of persons with 2 serum samples at approximately six month intervals (except infants)	10	14.9	12	12	

(6)	New Orleans	1956-1959	H2N2	Seroconversion to positive or a 4-fold or greater rise in CF antibody titer	% of persons with 2 serum samples at approximately six month intervals (except infants)	43.1	49	21	33	Pandemic year
(6)	New York	1961-1965	B	Seroconversion to positive or a 4-fold or greater rise in CF antibody titer	% of persons with 2 serum samples at approximately six month intervals (except infants)	0	9.6	3	3	
(6)	New Orleans	1956-1959	B	Seroconversion to positive or a 4-fold or greater rise in CF antibody titer	% of persons with 2 serum samples at approximately six month intervals (except infants)	5.3			4	
(7)	Seattle	1965-1969	A	Seroconversion to positive or a 4-fold or greater rise in CF antibody titer	Per 100-person years	25.5	18	14.1	18.6	0-<6 years and 6-19 years
(7)	Seattle	1965-1969	B	Seroconversion to positive or a 4-fold or greater rise in CF antibody titer	Per 100-person years	27.2	15	16.8	19.9	0-<6 years and 6-19 years
(8)	Seattle	1975-1976	H3N2	Virus isolation or 4-fold or greater rise in CF or HI titer	% per season	23.4	25	10.9	18.4	
(8)	Seattle	1976-1977	H3N2	Virus isolation or 4-fold or greater rise in CF or HI titer	% per season	2.6	5.2	6.9	5.8	
(8)	Seattle	1977-1978	H3N2	Virus isolation or 4-fold or greater rise in CF or HI titer	% per season	26.2	33.2	12.7	23.7	
(8)	Seattle	1978-1979	H3N2	Virus isolation or 4-fold or greater rise in CF or HI titer	% per season	0	0.4	0.5	0.4	
(8)	Seattle	1975-1976	B	Virus isolation or 4-fold or greater rise in CF or HI titer	% per season	8.5	28.6	6.3	16.6	
(8)	Seattle	1976-1977	B	Virus isolation or 4-fold or greater rise in CF or HI titer	% per season	3.9	1.6	1.2	1.6	
(8)	Seattle	1977-1978	B	Virus isolation or 4-fold or greater rise in CF or HI titer	% per season	1.6	4.3	2.2	3.2	
(8)	Seattle	1978-1979	B	Virus isolation or 4-fold or greater rise in CF or HI titer	% per season	0	2.3	2.7	2.4	
(8)	Seattle	1977-1978	H1N1	Virus isolation or 4-fold or greater rise in CF or HI titer	% per season	8.2	6.5	2.2	4.8	
(8)	Seattle	1978-1979	H1N1	Virus isolation or 4-fold or greater rise in CF or HI titer	% per season	42.3	54	1.8	30.6	Pandemic year
(9)	Tecumseh	1966-1971	A	A 4-fold or greater rises in HI titer or 1:16 titre in person with previously undetectable titre.	% per surveillance year	17.7	18.5	15	16.7	
(9)	Tecumseh	1966-1971	B	A 4-fold or greater rises in HI titer or 1:16 titre in person with previously undetectable titre.	% per surveillance year	3.3	12.6	4.9	7.6	
(10)	Tecumseh	1977-1978	H3N2	A 4-fold or greater rise in HI	% per outbreak period	42.9	32.8	12.9	19.1	Outbreak year

				titer or virus isolation						
(10)	Tecumseh	1980-1981	H3N2	A 4-fold or greater rise in HI titer or virus isolation	% per outbreak period	23.8	14.8	14.5	16	Outbreak year
(10)	Tecumseh	1977-1978	H1N1	A 4-fold or greater rise in HI titer or virus isolation	% per outbreak period	6.9	31.8	4.6	12.3	Outbreak year
(10)	Tecumseh	1978-1979	H1N1	A 4-fold or greater rise in HI titer or virus isolation	% per outbreak period	28.6	44.5	4.1	18.4	Pandemic year
(10)	Tecumseh	1980-1981	H1N1	A 4-fold or greater rise in HI titer or virus isolation	% per outbreak period	10.8	15.3	1.3	6.1	Outbreak year
(10)	Tecumseh	1976-1977	B	A 4-fold or greater rise in HI titer or virus isolation	% per outbreak period	3.8	27.1	3.5	9.9	Outbreak year
(10)	Tecumseh	1979-1980	B	A 4-fold or greater rise in HI titer or virus isolation	% per outbreak period	8.6	27.1	9.6	14.4	Outbreak year
(11)	Houston	1976	H3N2	A 4-fold or greater rise in HI titer or virus isolation	% per between January 1975-April 1976.	36.2	33.3	17.9	27.7	Outbreak year
(12)	Houston	1976-1977	B	A 4-fold or greater rise in HI or microneutralization titer, or virus isolation.	% per epidemiologic year	22.8	45.4	17.8	24.3	Outbreak year
(12)	Houston	1979-1980	B	A 4-fold or greater rise in HI or microneutralization titer, or virus isolation.	% per epidemiologic year	21.4	32.7	15.2	20.4	Outbreak year
(13)	Houston	1977-1979 & 1980-1981	H1N1	A 4-fold or greater rise in HI or microneutralization titer, or virus isolation.	Per 100-person years	18.9	30.9	7.2	15.3	Outbreak years / 0-<6 years and 6-17 years
(13)	Houston	1977-1979 & 1980-1981	H3N2	A 4-fold or greater rise in HI or microneutralization titer, or virus isolation.	Per 100-person years	39.9	34.1	26.2	32.7	Outbreak years / 0-<6 years and 6-17 years
Horby	Ha Nam	2007-2008	H1N1	See main paper	Unadjusted % per season. See main paper		10	5.9	7.2	
Horby	Ha Nam	2007-2008	H3N2	See main paper	Unadjusted % per season. See main paper		3.9	1.6	2.3	
Horby	Ha Nam	2007-2008	B	See main paper	Unadjusted % per season. See main paper		12.8	12.6	12.6	
Horby	Ha Nam	2008-2009	H1N1	See main paper	Unadjusted % per season. See main paper		17.2	4.5	8.6	
Horby	Ha Nam	2008-2009	H3N2	See main paper	Unadjusted % per season. See main paper		16.7	11.3	13.1	
Horby	Ha Nam	2008-2009	B	See main paper	Unadjusted % per season. See main paper		11.5	10.1	10.5	
Horby	Ha Nam	2009-2010	H1N1/2009	See main paper	Unadjusted % per season. See main paper		33	9.6	18.2	Pandemic year

* Unless otherwise stated pre-school = 0-4 years; school-age = 5-19 years; adult ≥ 20 years
HI = Hemagglutination Inhibition assay. CF = Complement Fixation assay



Web Figure 4. Risk of influenza infection by season, influenza sub-type, gender, and age group. Ha Nam, Vietnam, 2007-2010.

Full materials and methods

In 2007 a prospective, household-based community cohort was established in Thanh Ha Commune, Thanh Liem District, Ha Nam Province, Vietnam. Vietnam is a lower middle-income country that has achieved rapid economic growth since the economic reforms of the late 1980's. It has a population of 85.8 million (2009 census), making it the third most populous country in Southeast Asia and 13th in the world. Vietnam has a high population density (259 persons/km^2), with 70% of the population living in rural areas, and good health indicators for its level of development. Ha Nam Province is situated in the Northern Red River Delta of Vietnam, the most densely populated area of Vietnam (930 persons/km^2), about 60km south of the capital city Hanoi. At a latitude of 20.502034 decimal degrees and longitude 105.928642, Thanh Liem District has a tropical climate with an average monthly median temperature of 24.2°C, minimum monthly median 14.2°C, and a maximum monthly median of 33°C (2007-2008). The Province was selected on the basis of the availability of trained staff, the travelling distance from Hanoi, the prior circulation of influenza A/H5N1, and the quality of relationships with the implementing institute, the National Institute of Hygiene and Epidemiology (NIHE). Members of the Provincial Preventive Medicine Centre selected the study site following discussions with various sites about the willingness of the community to participate in the research. Thanh Ha Commune is a semi rural community with a population of 7,663 (2007), making a living mostly through mixed agriculture and small-scale production (e.g. embroidery). The Commune has a Health Centre and is divided administratively into seven hamlets, each with one or more hamlet health workers. A community consultation meeting was held to explain the purpose of the study to community members, elected representatives of the community, and representatives of community organizations.

The primary sampling unit of study was the household and all households in the Commune were eligible for inclusion in the study. A list of all households in the Commune was compiled from the local Government population register and was the source document for the selection of households for inclusion in the study. Households were randomly selected from the household list using a random number table. If a randomly selected household declined to participate, the next nearest

household was approached until a household was successfully recruited. All permanent residents in the household were eligible for inclusion and were requested to participate. All potential participants were given information on the purpose of the study, the associated risks and benefits, and were required to provide written informed consent before inclusion in the study. Parents or legal guardians provided written consent for participants aged less than 18 years.

Baseline variables

Households were recruited and baseline information collected during November and December 2007. Trained hamlet health workers (HHW) conducted face-to-face interviews with all participants. Individual participants provided information on date of birth; gender; ethnicity; occupation; contact with children at work or home; the number of children in the school and class (for participants of school age only); the presence of chronic disease; frequency of travel outside of the Commune, District, Province, and Country; influenza vaccination history; and smoking behavior. The nominated 'household head' provided information on the number of people living in the house; the familial relationship between household members; the number of rooms in the house; and the ownership of household assets.

Blood sampling

Participants aged 5 years and older (at time of sampling) were asked to provide blood at recruitment and at three further time points. Recruitment blood samples were drawn between 1st-7th December 2007 (bleed 1). Subsequent bleeds took place between 9th-15th December 2008 (bleed 2), 2nd-4th June 2009 (bleed 3), and on the 3rd April 2010 (bleed 4). The bleeding time points were not decided *a priori* but were chosen when national influenza surveillance data indicated that influenza circulation was minimal. The four sets of samples provided three sets of paired sera. Sodium heparin blood collection tubes were used for bleeds 1-3 in order that peripheral blood mononuclear cells (PBMCs) could be extracted for a sub-study on T-cell responses in influenza; sodium heparin tubes provided plasma for determining haemagglutination inhibiting (HI) antibody titres. DNA was extracted from the cell pellet of the heparin blood samples for a sub-study of host genetic influences on influenza infection. Bleed 4 used clot-activator serum tubes, which provide serum for determining HI antibody titres.

Influenza-like illness surveillance

Trained HHWs undertook weekly active surveillance of each participating household for episodes of influenza-like illness (ILI) and for changes in household composition. Participants were also encouraged to actively report any episode of ILI as soon as possible directly to the HHW. ILI was defined as 'as an illness with oral temperature of 38°C or more and either a cough or a sore throat'. All participating households were provided with an alcohol-in-glass clinical thermometer and informed of the definition of an ILI used in the study. Any participant reporting an ILI was asked to attend the Commune Health Centre where a trained member of the health centre staff would take a nose swab and a throat swab for storage in viral transport media at 2-4 °C pending transfer to the laboratory for testing. Synthetic tipped swabs with plastic shafts were used and placed in 3 ml of transport media (DMEM with 2% v/v BSA, 0.3% v/v NaHCO3 and antibiotics). Participants whom reported an ILI were asked to complete a 10-day symptom diary.

Definition of exposure and outcome variables

For the purpose of analysis, an influenza 'season' was defined as the period between consecutive bleeds, and an influenza 'transmission period' was defined as the period when influenza was known to be circulating on the basis of RT-PCR confirmed clinical cases.

'Influenza infection' was defined as either the detection of influenza RNA in a swab sample by reverse transcription polymerase chain reaction (RT-PCR) or a four fold or greater rise in HI antibody titre in paired sera, with the second titre at least 1:40. If paired sera were not available, a single high titre of at least 1:160 for seasonal influenza, or a titre of ≥ 1:80 in someone aged under 40 years for pandemic influenza H1N1, was also considered to indicate recent 'influenza infection'.

'Influenza illness' was defined as the detection of influenza-specific RNA in a swab by RT-PCR and the reporting of an ILI, or serological evidence of recent influenza infection (see above) plus an ILI episode occurring during a known period of transmission of the relevant influenza subtype. For linking serological evidence of recent influenza A infection to specific ILI episodes, the following influenza A 'transmission periods' were defined: 01/07/2008-30/09/2008 (influenza transmission

period 1), 01/04/2009-05/06/2009 (influenza transmission period 2); and 01/09/2009-31/12/2009 (influenza transmission period 3). Influenza B circulated throughout 2008 and a 'transmission period' could not be defined, therefore serological evidence of recent influenza B infection was putatively linked to any ILI episode that was not attributable to influenza A infection.

Laboratory methods - reverse transcription polymerase chain reaction (RT-PCR) assay

Detection of influenza viruses in nasal- and throat-swab specimens was performed using either conventional or real-time RT-PCR. The real-time assay was performed according to the U.S. CDC/WHO protocols (CDC reference no. I-007-05, Accessed November 30, 2009, at http://www.who.int/csr/resources/publications/swineflu/CDCRealtimeRTPCR_SwineH1Assay-2009_20090430.pdf.). Conventional RT-PCR assays for H1N1/2009 were performed according to WHO Protocols using primers M30F2/08 and M264R3/08 for influenza A matrix and NIID-swH1 Conv-F1 and NIID-swH1 Conv-R1 for H1N1/2009 (WHO information for laboratory diagnosis of pandemic H1N1/2009 virus in humans – revised. 23 November 2009 (http://www3.ha.org.hk/idctc/document/swineflu/WHO_Diagnostic_RecommendationsH1N1_20090521.pdf). Conventional RT-PCR for seasonal influenza strains was preformed using one-step reactions with primers for influenza A matrix (as above); H3N2 (forward AAGCATTCCYAATGACAAACC, reverse ATTGCRCCRAATATGCCTCTAGT); H1N1 (forward AGGCAAATGGAAATCTAATAGCGC, reverse CCATTGGTGCATTTGAGGTGATG); and influenza B (forward TCCTCAACTCACTCTTCGAGCG, reverse CGGTGCTCT TGACCAAATTGG).

Laboratory methods - hemagglutination inhibition (HI) assay

Influenza hemagglutination inhibition (HI) assays were performed according to standard protocols. Virus stocks used as antigens were cultured from swabs from select study participants with positive RT-PCR assays for each subtype in each season, except for season 3 where the WHO reference strain A/H1N1/California/7/2009-like was used. They were either propagated in the

allantoic cavities of 10-day-old embryonated hen's eggs or in MDCK cells. Virus concentrations were determined by haemagglutination (HA) assay titration with appropriate erythrocytes at 0.5% (v/v) and used at titres of 1:8. Each virus was initially tested in HA with erythrocytes from chickens, guinea pigs and turkeys, and erythrocytes from chicken were selected for 2008 H1N1 (2008), and from turkey for H3N2 and H1N1/2009. Participant and reference serum or plasma was treated with receptor destroying enzyme (Denka Seiken, Japan), heat inactivated then adsorbed against packed appropriate erythrocytes. HI assays were performed in U-bottom 96-well microtitre plates with 0.5% v/v appropriate erythrocytes. Cell controls and positive controls containing WHO reference sera for each strain were included with each batch of sera tested and two sera controls were included for each participant. Paired sera were tested together in the same assay run.

Serum/plasma samples were tested at an initial dilution of 1:10 and then at two-fold serial dilutions to a maximum dilution of 1:1280. Results were accepted if sera and cell controls provided the correct non-agglutinated pattern and if positive controls were within two-fold of anticipated/historical titres. Samples that were negative by HI assay in the lowest dilution (1:10) were assigned a titre of 1:5 for the purposes of computing seroconversion.

Study size

The Tecumseh study of respiratory illness in the community estimated influenza virus associated illness occurred at a rate of around 220 per 1000 population per year with an additional 50 to 100 asymptomatic but serologically confirmed infections (Monto & Sullivan, 1993). Assuming an incidence rate of influenza infection of around 20% per influenza season and ignoring potential household clustering of influenza illness, a total of 1000 recruited subjects would lead to a two-sided 95% confidence interval for the incidence with a precision (width) of 5%.

Handling of quantitative variables in the analysis

The age of participants at the start of each influenza season was calculated from their date of birth. For analysis and presentation of data on ILI episodes and RT-PCR confirmed influenza infection, age was grouped into four categories to ensure sufficient outcome events in each category: 0-<5 years, 5-<20 years, 20-<40 years,

≥40 years. The same categories (except for <5 years olds which were not asked to provide blood) were used for the analysis of data on risk factors for influenza infection. For graphical presentation of serological outcomes, we used a finer age resolution with the following categories: 0-<5, 5-<10, 10-<20, 20-<30, 30-<40, 40-<50, 50-<60, 60-<70, ≥70. Home crowding was defined as being present if there were more than 2 people per room.

Statistical methods

Absolute observed risks of ILI (for subjects under ILI surveillance) and of influenza infection (overall and for influenza subtypes, for subjects under influenza infection surveillance) were calculated per season. Participants were considered under ILI surveillance for a particular season if they were under weekly ILI surveillance throughout the influenza transmission period and they were considered under influenza infection surveillance if they additionally contributed a post-season blood sample. Absolute risks per season were preferred to rates (events per person time) as the incidence of influenza varies strongly over time. For example, while the time from bleed 1 to bleed 2 (season 1) was one year and the time from bleed 2 to bleed 3 (season 2) only 6 months, both seasons contained a full transmission period of both influenza H1N1 and H3N2.

Survey analysis methodology was used to derive risk estimates and associated 95% confidence intervals in the full population and in age subgroups. This provides valid inference accounting for effects of the survey design, which was based on cluster sampling by household. The inclusion of subjects for assessment of influenza infection required blood samples and the willingness to provide blood differed by age and gender. To correct for this sampling bias, and to provide results that can be generalized to the broader population, the influenza risk estimates were standardized to the age and gender structure of the Vietnamese rural population based on the 2009 Population and Housing Census. As children under 5 years of age were not asked to give blood samples, standardization for influenza risks was to the census population aged ≥5 years. Standardization was implemented by raking, i.e. post-stratification on the target age and gender distribution in turn until convergence (Lumley, 2010).

Seven potential risk factors for influenza infection were pre-defined. To assess these factors, data were pooled over all three seasons and the overall risk of an influenza infection was modeled with a logistic mixed effects model depending on the season, a random household effect (to account for potential clustering within households), a random subject effect (to account for potential within-subject correlation between seasons) and the respective risk factors. The analysis was repeated for each influenza subtype separately.

All analyses were performed with the statistical software R 2.10.1 (R Foundation for Statistical Computing, Vienna, Austria) and the companion R packages survey 3.22-3 (for survey sampling) and lme4 0.999375-35 (for mixed models) ("R Development Core Team: R: A Language and Environ," n.d.).

Missing data and loss to follow-up

Participants were excluded from all analyses if data on age or sex were missing. Participants were excluded from analysis of a particular influenza season if they were absent from the study site for a period of one week or more during the influenza transmission period; this included absence due to death, permanent out-migration, or temporary absence.

References

1. Monto AS, Sullivan KM. Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiol Infect*. 1993;110(1):145-160.
2. Lumley T. Complex Surveys: A Guide to Analysis Using R. Wiley; 2010.
3. R Development Core Team: R: A Language and Environment for Statistical Computing. (http://www.r-project.org/). (Accessed 26 January 2012).
4. Jordan W. S. J, Badger GF, Dingle JH. A study of illness in a group of Cleveland families. XVI. The epidemiology of influenza, 1948-1953. *Am J Hyg*. 1958;68(2):169-189.
5. Jordan W. S. J, Denny F. W. J, Badger GF, et al. A study of illness in a group of Cleveland families. XVII. The occurrence of Asian influenza. *Am J Hyg*. 1958;68(2):190-212.
6. Hall CE, Brandt CD, Frothingham TE, et al. The virus watch program: a continuing surveillance of viral infections in metropolitan New York families. IX. A comparison of infections with several respiratory pathogens in New York and New Orleans families. *Am J Epidemiol*. 1971;94(4):367-85.

7. Hall CE, Cooney MK, Fox JP. The Seattle virus watch. IV. Comparative epidemiologic observations of infections with influenza A and B viruses, 1965-1969, in families with young children. *Am J Epidemiol*. 1973;98(5):365-380.

8. Fox JP, Hall CE, Cooney MK, et al. Influenza virus infections in Seattle families, 1975-1979. I. Study design, methods and the occurrence of infections by time and age. *Am J Epidemiol*. 1982;116(2):212-27.

9. Monto AS, Kioumehr F. The Tecumseh Study of Respiratory Illness: IX. Occurrence of Influenza in the Community, 1966-1971. *Am J Epidemiol.* 1975;102(6):553-563.

10. Monto a S, Sullivan KM. Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiology and Infection*. 1993;110(1):145-60.

11. Taber LH, Paredes a, Glezen WP, et al. Infection with influenza A/Victoria virus in Houston families, 1976. *The Journal of Hygiene*. 1981;86(3):303-13.

12. Frank AL, Taber LH, Glezen WP, et al. Influenza B virus infections in the community and the family. The epidemics of 1976-1977 and 1979-1980 in Houston, Texas. *Am J Epidemiol*.1983;118(3):313-25.

13. Frank AL, Taber LH, Wells JM. Comparison of infection rates and severity of illness for influenza A subtypes H1N1 and H3N2. *J Infect Dis*. 1985;151(1):73-80.

Chapter 5

Research paper 2

Title: Social Contact Patterns in Vietnam and Implications for the Control of Infectious Diseases

Author(s): Peter Horby, Pham Quang Thai, Niel Hens, Nguyen Thi Thu Yen, Le Quynh Mai, Dang Dinh Thoang, Nguyen Manh Linh, Nguyen Thu Huong, Neal Alexander, W. John Edmunds, Tran Nhu Duong, Annette Fox, Nguyen Tran Hien.

Journal/Publisher: PLoS ONE, Public Library of Science

Type of publication: Research article

Stage of publication: Published

Academic peer-reviewed: Yes.

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Candidate's role: I conceived and designed the study. I prepared the dataset for analysis and developed the analysis plan in conjunction with Niel Hens, a statistician with experience of analysing social contact data. Niel Hens wrote and implemented the R-script. I read and commented on the R-script. I wrote the first and all subsequent drafts of the manuscript, submitted the manuscript for publication, and responded to all reviewers comments.

Candidate's signature:

Supervisor or senior author's signature to confirm Candidates role:

OPEN ACCESS Freely available online

PLoS one

Social Contact Patterns in Vietnam and Implications for the Control of Infectious Diseases

Peter Horby[1,2]*, Pham Quang Thai[3], Niel Hens[4,5], Nguyen Thi Thu Yen[3], Le Quynh Mai[3], Dang Dinh Thoang[6], Nguyen Manh Linh[3], Nguyen Thu Huong[3], Neal Alexander[7], W. John Edmunds[7], Tran Nhu Duong[3], Annette Fox[1,2], Nguyen Tran Hien[3]

1 Oxford University Clinical Research Unit, Hanoi, Vietnam, **2** Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom, **3** National Institute of Hygiene and Epidemiology, Hanoi, Vietnam, **4** I-Biostat, Hasselt University, Diepenbeek, Belgium, **5** Centre for Health Economics Research and Modeling Infectious Diseases, Vaccine and Infectious Disease Institute, University of Antwerp, Antwerp, Belgium, **6** Ha Nam Centre for Preventive Medicine, Ha Nam, Vietnam, **7** London School of Hygiene and Epidemiology, London, United Kingdom

Abstract

Background: The spread of infectious diseases from person to person is determined by the frequency and nature of contacts between infected and susceptible members of the population. Although there is a long history of using mathematical models to understand these transmission dynamics, there are still remarkably little empirical data on contact behaviors with which to parameterize these models. Even starker is the almost complete absence of data from developing countries. We sought to address this knowledge gap by conducting a household based social contact diary in rural Vietnam.

Methods and Findings: A diary based survey of social contact patterns was conducted in a household-structured community cohort in North Vietnam in 2007. We used generalized estimating equations to model the number of contacts while taking into account the household sampling design, and used weighting to balance the household size and age distribution towards the Vietnamese population. We recorded 6675 contacts from 865 participants in 264 different households and found that mixing patterns were assortative by age but were more homogenous than observed in a recent European study. We also observed that physical contacts were more concentrated in the home setting in Vietnam than in Europe but the overall level of physical contact was lower. A model of individual versus household vaccination strategies revealed no difference between strategies in the impact on R_0.

Conclusions and Significance: This work is the first to estimate contact patterns relevant to the spread of infections transmitted from person to person by non-sexual routes in a developing country setting. The results show interesting similarities and differences from European data and demonstrate the importance of context specific data.

Citation: Horby P, Thai PQ, Hens N, Yen NTT, Mai LQ, et al. (2011) Social Contact Patterns in Vietnam and Implications for the Control of Infectious Diseases. PLoS ONE 6(2): e16965. doi:10.1371/journal.pone.0016965

Editor: Cesar Munayco, Dirección General de Epidemiología, Peru

Received December 2, 2010; **Accepted** January 10, 2011; **Published** February 14, 2011

Copyright: © 2011 Horby et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Wellcome Trust UK (grants 081613/Z/06/Z and 077078/Z/05/Z). NH gratefully acknowledges financial support from "SIMID", a strategic basic research project funded by the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), project number 060081 and by the IAP research network number P6/03 of the Belgian Government (Belgian Science Policy). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: peter.horby@gmail.com

Introduction

Mathematical models of infectious disease transmission have become indispensible tools for understanding epidemic processes and for providing policy makers with an evidence base for decisions when empirical data is limited. The success of mathematical models in informing critical decisions to protect human and animal health has been demonstrated for many diseases including pandemic influenza, SARS, foot and mouth disease, and new variant CJD [1]. Infections directly transmitted from person to person by the respiratory route have been of special interest for modeling because of their ability to spread quickly and affect large numbers of people.

The validity of mathematical models, and therefore the effectiveness of policies based on these models, is dependent on the robustness of the parameters entered in to the model [1,2]. A key parameter in infectious disease models is the probability of contact between an infectious source and a susceptible individual. For infections transmitted from person to person various assumptions are required to simplify the range of human relations into tractable mathematical models. Earlier assumptions of homogenous mixing, where everyone in the population has an equal probability of contact, have been replaced by more realistic frameworks where the probability of contact varies between groups, most often defined by age. The extent to which individuals preferentially mix with people of the same age (assortativeness) is a key heterogeneity that is now routinely included in models and attempts have also been made to further represent the underlying structure of contact patterns by partitioning the population into household and workplace compartments [3,4,5].

Understanding and incorporating the key elements of population contact structures into models is important since it improves

the predictive accuracy of the model and also permits investigation of the effect of interventions targeted at specific settings, such as schools, workplaces or homes [5,6]. Indeed, family size and composition have been associated with both social contact frequency and the risk of infection with influenza and other respiratory pathogens [7,8,9].

Seroepidemiological studies have been used to infer contact patterns relevant to the transmission of infections and a number of surveys have been conducted to directly measure social contacts [10,11,12,13,14]. The self reported social contact data derived from such surveys have been shown to better predict the observed patterns of respiratory infections than other representations of contact probabilities, such as homogenous or proportionate mixing [12,14,15,16]. The frequency and nature of social contacts are however determined by demographic factors, the living and working environment, socio-cultural norms and individual lifestyle choices; all of which vary by place and time. A study of eight European countries found that contact patterns were very similar but little is known about differences in social contact behaviors across more diverse socio-cultural environments [13].

The vast majority of social contact surveys have been conducted in developed western countries yet the majority of the world's population live in less developed countries where family structures, socio-cultural norms, population mobility and the home and work environment may differ in important ways from Europe. Developing countries are also more often sites for the emergence of infectious diseases and in an increasingly connected world, localized outbreaks can rapidly 'go global' with devastating health and economic impacts. There is therefore a need to determine social contact patterns in developing country settings, so that the benefits of mathematical modeling can be extended to these higher risk and more vulnerable populations [17].

To address this knowledge gap we have used a social contact diary approach to estimate the frequency and nature of social contacts in a semi-rural community of Vietnam. Since the household is a fundamental unit for the transmission of many infections and household characteristics clearly influence transmission risks, we employed a household-based survey design.

Methods

Study area and population

Vietnam has a population of 85.8 million people, making it the 3rd most populous country in Southeast Asia (after Indonesia and the Philippines) and the 13th most populous nation in the world. 70% of the population lives in rural areas. The Red River Delta in the north and the Mekong River Delta in the South together comprise 43% of the population and the Red River Delta is the most densely populated area, with 930 people per km^2 [18]. Data on the national distribution of household sizes and the population age structure was obtained from the Vietnam General Statistics Office (GSO; http://www.gso.gov.vn).

Survey population

In 2007 a household-based cohort was established in a semi-rural community in the Red River Delta of North Vietnam. Households were randomly selected from a list of all households in the commune (the third administrative level) using a random number table. If a selected household declined to participate the nearest neighbor was approached for participation.

Survey methods

A paper-based questionnaire was developed based on an earlier European study but adapted to the local context [13]. With the assistance of a trained interviewer, subjects recorded the details of each contact made on the day preceding the interview. In order to improve recall, subjects were informed of the day on which they would be interviewed in advance. The same definition of a contact was used as the European study, which was: either skin-to-skin contact (a physical contact), or a two-way conversation with three or more words in the physical presence of another person but no skin-to-skin contact (a nonphysical contact). One entry was made for each person contacted during the diary day, which was defined as starting at 5 a.m. on the morning of the day assigned and ending at 5 a.m. the next morning. If an individual was contacted multiple times during the day, the individual was recorded only once but the total time spent with that person during the day was entered. Information was recorded on the age and gender of each contact, the location and duration of the contact, whether skin-to-skin contact had occurred, and how often the interviewee normally had contact with the individual. The diary is available in the Supporting Information (text S1).

Every member of each participating household was requested to complete the contact diary. Participants completed the questionnaire with the assistance of trained village health workers during face-to-face interviews. For children aged 10 years or less, the diary was completed with the assistance of the child's parent or guardian. Data were double entered into an Access database.

Data analysis

We used generalized estimating equations (GEE) to model the number of contacts participants in age-category I make with persons in age-category J while taking into account the correlation introduced by sampling households. GEEs use working correlation matrices to take the correlation into account and provide unbiased estimates even if the working correlation matrix is misspecified, albeit at the potential loss of efficiency. We used an independence working correlation matrix to take into account clustering within households and as a result of using the GEE approach the correlation between the number of contacts from the same participant over different age-categories is also taken into account. Sampling weights are calculated using Vietnamese census data to balance the contribution over the different days of the week and to balance the household size and age distribution towards the Vietnamese population. Matrices of the relative intensity of contact between age groups were estimated using weighted GEE and were made reciprocal (i.e. the relative frequency of 0–5 years old subjects having contact with 0–5 year olds is the same) by averaging across the two cells. Reciprocal, balanced matrices are needed for next generation matrices in mathematical models of disease transmission. The use of a weighted GEE approach allows population level inferences to be made from the sample dataset.

In order to model the effect of individual or household targeted immunization strategies we mimicked the immunization process of individuals or households by setting their corresponding contacts to 0 for all age-categories. The basic reproduction number R_0 can be calculated as the dominant eigenvalue of the next generation operator [19] which can be calculated as the dominant eigenvalue of the matrix $\boldsymbol{ND\beta}$ where $\boldsymbol{N}$ is a vector of age-group specific population sizes, D is the mean infectious period and $\boldsymbol{\beta}$ is the per capita transmission rate. Under the social contact hypothesis, Wallinga et al. 2006 assumed $\boldsymbol{\beta} = q\boldsymbol{C}$ where q is a proportionality factor and $\boldsymbol{C}$ is the per capita contact matrix. The relative reduction in R_0 when immunizing from p = 0% up to 30% of the population can then be calculated as the ratio of dominant eigenvalues of $\boldsymbol{NC_p}$ and $\boldsymbol{NC}$, respectively [20]. Here $\boldsymbol{C_p}$ is the matrix of per capita contact rates between the different age-groups as estimated using the GEE when immunizing a proportion p of

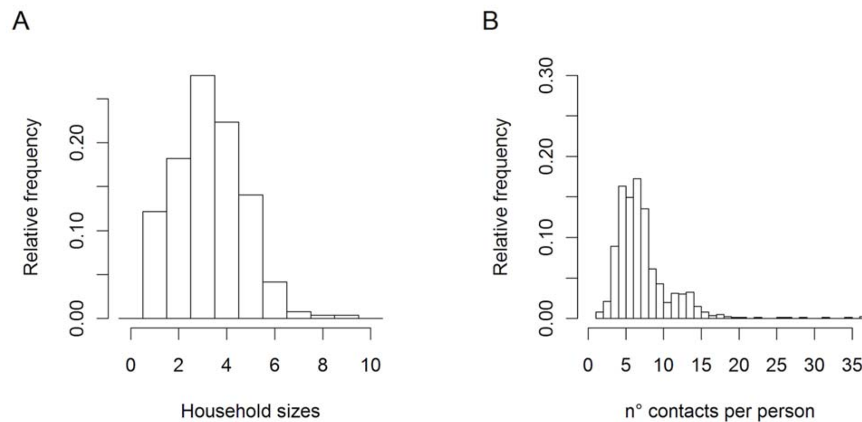


Figure 1. Household sizes (A) and number of reported contacts per person per day (B).
doi:10.1371/journal.pone.0016965.g001

Table 1. Number of recorded contacts per participant per day by characteristics, and relative number of contacts from weighted GEE analysis.

Category	Covariate	Number of participants	Mean (SD) of Number of Reported Contacts	Relative Number of Contacts (95% Confidence interval)
Age of participant	0–4	74	5.47 (2.17)	1.00
	5–9	66	6.74 (3.84)	1.23(1.10–1.37)
	10–14	95	7.91 (5.65)	1.09(0.96–1.25)
	15–19	94	7.67 (3.47)	1.30(1.08–1.56)
	20–29	110	7.02 (2.68)	1.17(0.93–1.46)
	30–39	120	8.02 (3.21)	1.33(1.13–1.58)
	40–49	157	8.65 (4.44)	1.29(1.07–1.55)
	50–59	76	8.71 (3.51)	1.44(1.19–1.75)
	60+	73	8.21 (3.18)	1.31(1.02–1.68)
Sex of participant	Female	471	7.74 (3.78)	1.00
	Male	389	7.67 (3.97)	1.01(0.94–1.08)
	Missing Value	5	9.00 (3.08)	1.77(1.54–2.02)
Household Size	1	32	8.59 (3.40)	1.00
	2	96	7.89 (3.48)	0.94(0.79–1.12)
	3	219	8.01 (4.35)	1.06(0.88–1.26)
	4	236	7.30 (4.35)	1.02(0.84–1.24)
	5	185	7.72 (3.24)	1.16(0.94–1.44)
	6+	97	7.60 (2.86)	1.03(0.84–1.26)
Day of the week	Monday	8	7.75 (2.66)	1.00
	Tuesday	148	8.92(4.50)	1.17(0.92–1.49)
	Wednesday	302	7.83 (3.24)	0.96(0.79–1.15)
	Thursday	181	7.20 (4.21)	0.93(0.76–1.14)
	Friday	134	7.15 (4.04)	0.97(0.81–1.17)
	Saturday	30	6.82 (2.90)	0.93(0.79–1.08)
	Sunday	26	7.19 (2.62)	1.05(0.92–1.18)
	Missing Value	6	12.00 (6.36)	1.52(0.90–2.55)

Dispersion parameter alpha = 0.79 (0.33,1.24); alpha = 0 would correspond to no overdispersion.
NA indicating missing values.
doi:10.1371/journal.pone.0016965.t001

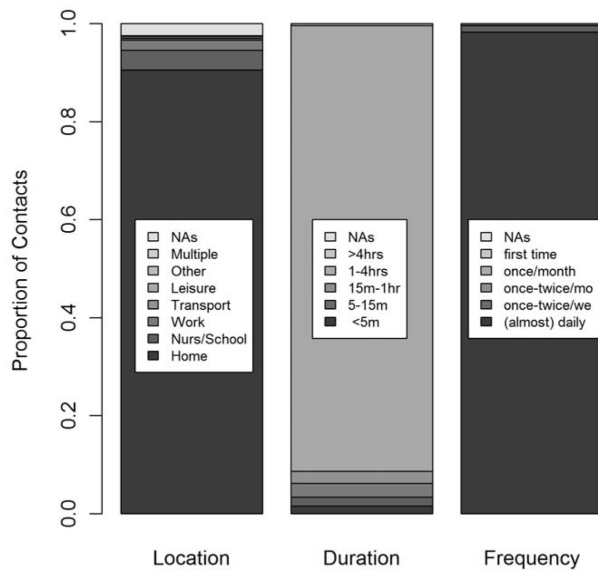


Figure 2. Contacts by location, duration and frequency. The figures are based on a WGEE with weights based on household size, days of the week and age.
doi:10.1371/journal.pone.0016965.g002

the population by either randomly selecting individuals or households and putting their contacts to 0 for all age-categories. $\boldsymbol{C}$ is the matrix of per capita contact rates without immunization. Statistical analysis was conducted in R 2.9.0 (The R Foundation for Statistical Computing).

Results

Participant characteristics, number of contacts and associated covariates

We recorded 6675 contacts from 865 participants in 264 different households. The mean age of respondents was 32 years (range 0–90) and 55% were female. The mode household size was 3 persons and the mean number of different people contacted per respondent per day was 7.7 (sd 3.9) indicating the need to use a count model that allows for overdispersion (i.e. the exhibited variability exceeds what is expected using a Poisson model, where the variability equals the mean. Note that the WGEE approach in addition to the mean parameter uses a dispersion parameter to allow for overdispersion) (figure 1). In a weighted GEE analysis we observed no association between the total number of recorded contacts and household size or gender. The number of reported contacts was found to be smaller for infants aged 0–4 years as compared to older participants, among which no difference was observed (table 1). This demonstrates, at an aggregate level, rather homogenous frequencies of social contacts across ages, genders and days of the week.

Nature, duration, location and frequency of contacts

In the weighted GEE analysis just over 81% of all contacts lasted more than four hours whilst contacts of shorter duration (<5 minutes; 5–15 minutes; 15 minutes to 1 hour; 1–4 hours) contributed between 4–5% of contacts each. Most reported contacts (93%) were with people that the respondent reported meeting daily or almost daily, with only one reported contact with an individual that the respondent had never met before. The most common reported location where contact occurred was the home (85%), followed by school (5%) and work place (4%) (figure 2).

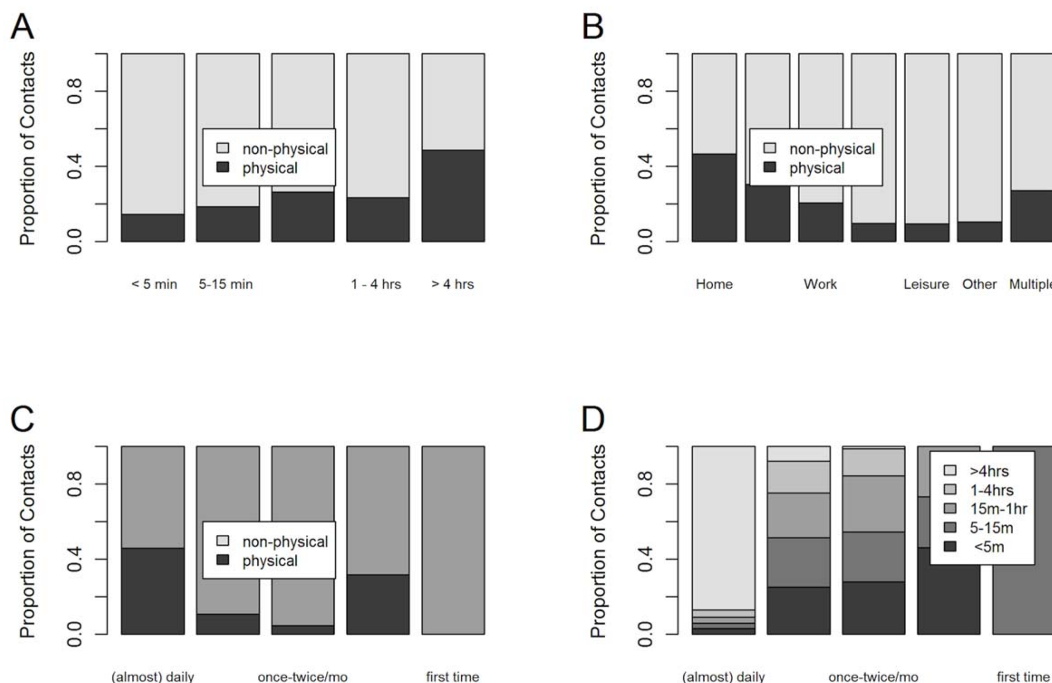


Figure 3. The location, duration and frequency of contacts. The proportion of contacts that were physical or non-physical by duration (panel A), location (panel B) and frequency of contact (panel C). The duration of contact by frequency of contact (panel D). The figures are based on a WGEE with weights based on household size and days of the week.
doi:10.1371/journal.pone.0016965.g003

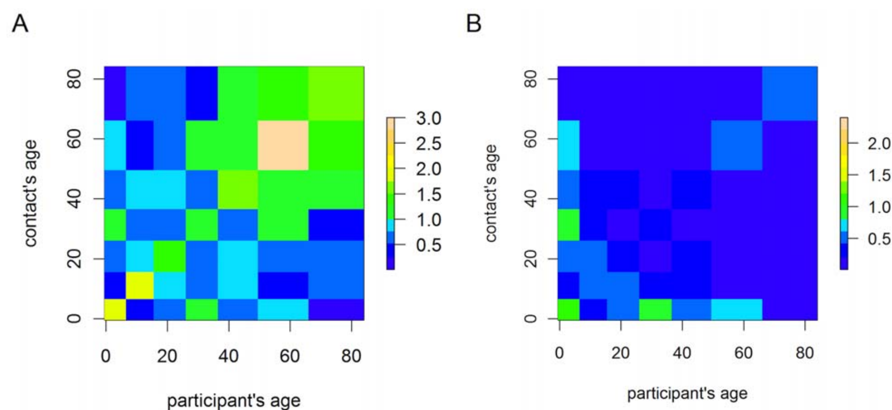


Figure 4. Contact intensity matrices for all contacts (A) and for physical contacts only (B). Yellow indicates high contact rates and blue low contact rates, relative to the mean contact intensity.
doi:10.1371/journal.pone.0016965.g004

Forty four percent of all reported contacts involved physical contact. Physical contact was most common in the home setting, where 91% of all physical contacts occurred. Physical contact was also more common when the duration of contact was long and when the subject had contact with that person on an almost daily basis (figure 3). 91% of physical contacts were with people with whom the respondent spent more than four hours during the day and 93% of physical contacts were with people who the respondent usually contacted daily or almost daily. In total, 85% of all physical contacts were in the home for more than four hours with people the respondent meet daily or almost daily.

Age related social mixing patterns

The weighted GEE-model was used to estimate the intensity of contacts between age groups for all participants (figure 4). The matrix shows that contact intensity for all contacts tends to be highest in the diagonal, demonstrating an assortative mixing pattern where the greatest contact is between individuals of a similar age group. However, a wide area of moderate intensity contact is also apparent for adults aged 26 to 65 years, indicating rather homogenous mixing amongst working age adults. Two secondary areas of moderate intensity contact are also apparent between the 20–65 year age group and children aged 0–5 years. This probably represents contact between parent and their children and, grandparents and their grand children. Physical contacts are most intense amongst children aged 0–5, both within that age group and with young adults, as shown in the right hand panel of figure 4.

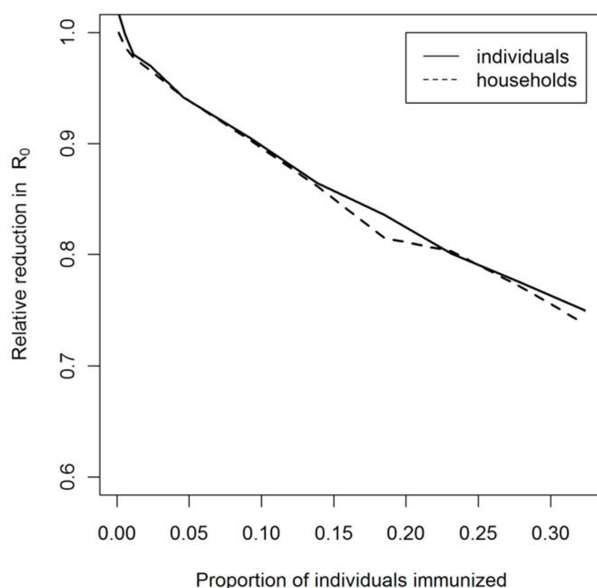


Figure 5. The predicted effect on R_0 of immunizing individuals or households. The figure shows the predicted effect on R_0 immunizing a random selection of individuals (solid line) versus a random selection of households (broken line).
doi:10.1371/journal.pone.0016965.g005

Comparison of immunization strategies

Assuming that infection is transmitted through the recorded contact behaviors and that there is full susceptibility to infection, modeling of the potential impact of individual versus household targeted immunization strategies revealed no difference in the predicted effect for a given level of vaccine coverage (figure 5).

Discussion

The successful spread of an infectious disease that is transmitted from person to person is dependent on many factors, but key amongst these are the susceptibility of the population, and the frequency and assortativeness of contacts that effectively transmit infection. Quantifying these parameters is critical for estimating the impact of such infections, for designing and targeting preventive interventions, and for modelling their impact [1]. Whilst much work has been conducted on defining these parameters for sexually transmitted infections, less has been done on contact behaviours relevant to the transmission of respiratory infections; and what has been done has been conducted exclusively in developed countries [10,11,12,13]. Here we report the first data from a developing country on social contacts relevant to the spread of respiratory pathogens.

Using the same definition of a contact and comparable methodology to a large European study, we have identified both similarities and potentially important differences in our study site in Vietnam [13]. Similarities with the European data include significant over dispersion in the distribution of contacts and no gender differences in reported contact frequency. As observed in Europe, we too found a peak in contact frequency in school age children, but in contrast to the European data, we also observed a second peak in adults aged 40–60 years. Another similarity with

the European study was that prolonged and frequent contacts, and contacts occurring at home were much more likely to be physical in nature. However, there were important differences in the total number of contacts, and the duration and intimacy of contacts.

Overall we recorded a mean of 7.7 contacts per participant per day versus 13.4 in the study by *Mossong et al.* The lower number of daily contacts we recorded may be a feature of the particular community studied or may reflect a recall bias introduced by the retrospective nature of our study design compared to the prospective design of the European study. Over 80% of contacts that occurred on a daily basis in the Vietnam study were more than four hours, compared to only around 45% in the study by Mossong *et al* [13] Physical contact was more common in the European study, with 75% of home contacts being physical compared to around 45% in our study, and over 60% of daily contacts being physical compared to around 40% in our study. The importance of these differences to disease patterns depends on the relative importance of duration of contact versus intimacy of contact on the probability of successful transmission.

In general the contact patterns in our study were more homogenous than that reported elsewhere. We saw smaller differences between age groups in contact frequency and no significant differences between household sizes. We saw similar patterns of age dependent mixing to those reported by Mossong *et al*, with pronounced assortative mixing seen as a high intensity diagonal, signals of parent-child mixing, and a 'plateau' of mixing of adults with one another. We also observed no significant differences in contact frequency by day of the week, whereas significantly more contacts in Europe were recorded on weekdays compared to weekends. This is may be because weekends are not generally observed as a special rest period in rural Vietnam to the extent they are in Europe. We also saw fewer contacts in 'leisure' settings (1% vs 16%), which may reflect true differences in the amount of time devoted to leisure, cultural differences in the conceptual separation between work, family and leisure activities, or limitations of the survey method in distinguishing leisure from other activities. Surprisingly, only one contact was reported with a person that the respondent had never met before. Whilst the studied community is rural, it is within ten kilometres of a small town, so cannot be considered remote.

Although we used weights to make inferences about contact behaviours in the general population of Vietnam, the reliability of such a generalization is limited by the fact that the study was conducted in only one setting and at only one time point. It is possible that contact behaviours may vary significantly between rural and urban areas and by season. Future studies will be needed to further define such heterogeneities.

The added value of our data compared to previous published work is two-fold. We are the first group to report on contact behaviours relevant to the spread of respiratory infections from a developing country, and we are the first to report household structured contact diaries of this nature. These novel features of our data can provide valuable insights into the spread of directly transmitted infections in a rural developing country setting and the potential impact of individual versus household targeted control strategies. Although we found no difference in the estimated impact on R_0 between individual- and household-targeted immunisation strategies, the model assumed that all recorded contacts were equally important in the transmission of infection, whereas it is likely that the risk of successful transmission is heterogeneous and varies with different intensity and duration of contacts.

The spread of directly transmitted infections is dependent on at least four unknown parameters: the susceptibility of the population; the frequency of contacts; the assortativeness of contacts; and the type of contact that transmits infection. The susceptibility of the population can in part be measured by serological and other surveillance data, and this study has gone some way to answering the second two unknowns. The fourth unknown, the types of contact that transmit respiratory infections and their relative importance, is however harder to answer. There has been a vigorous debate over the relative importance of aerosols versus large droplets in the transmission of influenza, and even suggestions that the predominant route may vary between climatic regions [21,22,23,24]. It is a critical question since models that assume all social contacts provide an equal opportunity for infection may result in incorrect conclusions [2,25]. As an adjunct to physico-mechanical explorations of the transmission of respiratory infections, a valuable supplementary approach is to explore associations between the frequency, intensity and duration of contacts and the measured risk of transmission. This has been done to some extent by comparing seroepidemiological data with contact patterns at an aggregated, population level, but might also be done at an individual level [15]. Multi-country studies that incorporate biomarkers of infection will help to further define spatial and temporal heterogeneities in contact behaviours and the relevance of particular contact profiles to infection risk.

Supporting Information

Text S1 Contact diary.
(DOC)

Acknowledgments

We are grateful to the community of An Hoa Commune for agreeing to participate in this study and for providing their time. We would like to thank the village health workers who conducted the interviews. We also wish to thank the Ministry of Health of Vietnam for their continuing support of the research collaboration between the Oxford University Clinical Research Unit and the National Institute for Hygiene and Epidemiology.

Author Contributions

Conceived and designed the experiments: PH. Performed the experiments: PQT PH NTTY LQM DDT NML NTH TND AF NTH. Analyzed the data: PH NH PQT. Contributed reagents/materials/analysis tools: NA WJE. Wrote the paper: PH NH.

References

1. Grassly NC, Fraser C (2008) Mathematical models of infectious disease transmission. Nat Rev Microbiol 6: 477–487.
2. Duerr HP, Schwehm M, Leary CC, De Vlas SJ, Eichner M (2007) The impact of contact structure on infectious disease control: influenza and antiviral agents. Epidemiol Infect 135: 1124–1132.
3. Ball F, Neal P (2002) A general model for stochastic SIR epidemics with two levels of mixing. Math Biosci 180: 73–102.
4. Ball F, Neal P (2008) Network epidemic models with two levels of mixing. Math Biosci 212: 69–87.
5. Pellis L, Ferguson NM, Fraser C (2009) Threshold parameters for a model of epidemic spread among households and workplaces. J R Soc Interface 6: 979–987.
6. Becker NG, Dietz K (1995) The effect of household distribution on transmission and control of highly infectious diseases. Math Biosci 127: 207–219.
7. Berglund B (1967) Respiratory syncytial virus infections in families. A study of family members of children hospitalized for acute respiratory disease. Acta Paediatr Scand 56: 395–404.
8. McCaw JM, Forbes K, Nathan PM, Pattison PE, Robins GL, et al. Comparison of three methods for ascertainment of contact information relevant to respiratory pathogen transmission in encounter networks. BMC Infect Dis 10: 166.
9. Monto AS, Cavallaro JJ (1971) The Tecumseh study of respiratory illness. II. Patterns of occurrence of infection with respiratory pathogens, 1965–1969. Am J Epidemiol 94: 280–289.

10. Edmunds WJ, O'Callaghan CJ, Nokes DJ (1997) Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. Proc Biol Sci 264: 949–957.
11. Edmunds WJ, Kafatos G, Wallinga J, Mossong JR (2006) Mixing patterns and the spread of close-contact infectious diseases. Emerg Themes Epidemiol 3: 10.
12. Wallinga J, Teunis P, Kretzschmar M (2006) Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. Am J Epidemiol 164: 936–944.
13. Mossong J, Hens N, Jit M, Beutels P, Auranen K, et al. (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 5: e74.
14. Wallinga J, Edmunds WJ, Kretzschmar M (1999) Perspective: human contact patterns and the spread of airborne infectious diseases. Trends Microbiol 7: 372–377.
15. Ogunjimi B, Hens N, Goeyvaerts N, Aerts M, Van Damme P, et al. (2009) Using empirical social contact data to model person to person infectious disease transmission: an illustration for varicella. Math Biosci 218: 80–87.
16. Goeyvaerts N, Hens N, Aerts M, Beutels P. Model structure analysis to estimate basic immunological processes and maternal risk for parvovirus B19. Biostatistics.
17. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, et al. (2008) Global trends in emerging infectious diseases. Nature 451: 990–993.
18. GSO (2010) The 2009 Vietnam Population and Housing Census. Implementation and Preliminary Results. Hanoi: Vietnam General Statistics Office.
19. Diekmann O, Heesterbeek J (2000) Mathematical Methodology of Infectious Diseases: Model Building, Analysis and Interpretation John Wiley & Sons Ltd.
20. Hens N, Ayele GM, Goeyvaerts N, Aerts M, Mossong J, et al. (2009) Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. BMC Infect Dis 9: 187.
21. Lowen A, Palese P (2009) Transmission of influenza virus in temperate zones is predominantly by aerosol, in the tropics by contact: a hypothesis. PLoS Curr. RRN1002.
22. Tellier R (2009) Aerosol transmission of influenza A virus: a review of new studies. J R Soc Interface 6 Suppl 6: S783–790.
23. Bouvier NM, Lowen AC, Palese P (2008) Oseltamivir-resistant influenza A viruses are transmitted efficiently among guinea pigs by direct contact but not by aerosol. J Virol 82: 10052–10058.
24. Brankston G, Gitterman L, Hirji Z, Lemieux C, Gardam M (2007) Transmission of influenza A in human beings. Lancet Infect Dis 7: 257–265.
25. Smieszek T (2009) A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread. Theor Biol Med Model 6: 25.

Chapter 6

Research paper 3

Title: Modelling the progression of pandemic influenza A (H1N1) in Vietnam and the opportunities for reassortment with other influenza viruses

Author(s): Maciej F Boni, Bui Huu Manh, Pham Quang Thai, Jeremy Farrar, Tran Tinh Hien, Nguyen Tran Hien, Nguyen Van Kinh, and **Peter Horby**.

Journal/Publisher: BMC Medicine, BioMed Central Ltd

Type of publication: Research article

Stage of publication: Published

Academic peer-reviewed: Yes

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Candidate's role: I conceived of the study and contributed equally to the study design and interpretation with the first author Maciej Boni. The mathematical model used data collected through the community cohort study of which I am the principal investigator. Other routinely available data were compiled by other authors under my supervision. The mathematical model was written and implemented by the first author with input from me to the model structure, parameter estimates, and interpretation. I co-wrote the first draft of the manuscript with the first author and we jointly responded to reviewers comments.

Candidate's signature:

Supervisor or senior author's signature to confirm Candidates role:

BMC Medicine

BioMed Central

Research article

Open Access

Modelling the progression of pandemic influenza A (H1N1) in Vietnam and the opportunities for reassortment with other influenza viruses

Maciej F Boni[1,2,4], Bui Huu Manh[1], Pham Quang Thai[3], Jeremy Farrar[1,4,7], Tran Tinh Hien[5,7], Nguyen Tran Hien[3], Nguyen Van Kinh[6,7] and Peter Horby*[1,4,7]

Address: [1]Oxford University Clinical Research Unit, Vietnam, [2]MRC Centre for Genomics and Global Health, University of Oxford, Oxford, UK, [3]National Institute of Hygiene and Epidemiology, Hanoi, Vietnam, [4]Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK, [5]Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam, [6]National Institute for Infectious and Tropical Diseases, Hanoi, Vietnam and [7]South East Asia Infectious Disease Clinical Research Network, Vietnam

Email: Maciej F Boni - mboni@oucru.org; Bui Huu Manh - bhmanh73@gmail.com; Pham Quang Thai - pqthai@nihe.org.vn; Jeremy Farrar - jfarrar@oucru.org; Tran Tinh Hien - hientt@oucru.org; Nguyen Tran Hien - nthiennihe@vnn.vn; Nguyen Van Kinh - kinhvaac@yahoo.com; Peter Horby* - peter.horby@gmail.com

* Corresponding author

Published: 3 September 2009

BMC Medicine 2009, **7**:43 doi:10.1186/1741-7015-7-43

Received: 19 June 2009

Accepted: 3 September 2009

This article is available from: http://www.biomedcentral.com/1741-7015/7/43

© 2009 Boni et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: A novel variant of influenza A (H1N1) is causing a pandemic and, although the illness is usually mild, there are concerns that its virulence could change through reassortment with other influenza viruses. This is of greater concern in parts of Southeast Asia, where the population density is high, influenza is less seasonal, human-animal contact is common and avian influenza is still endemic.

Methods: We developed an age- and spatially-structured mathematical model in order to estimate the potential impact of pandemic H1N1 in Vietnam and the opportunities for reassortment with animal influenza viruses. The model tracks human infection among domestic animal owners and non-owners and also estimates the numbers of animals may be exposed to infected humans.

Results: In the absence of effective interventions, the model predicts that the introduction of pandemic H1N1 will result in an epidemic that spreads to half of Vietnam's provinces within 57 days (interquartile range (IQR): 45-86.5) and peaks 81 days after introduction (IQR: 62.5-121 days). For the current published range of the 2009 H1N1 influenza's basic reproductive number (1.2-3.1), we estimate a median of 410,000 cases among swine owners (IQR: 220,000-670,000) with 460,000 exposed swine (IQR: 260,000-740,000), 350,000 cases among chicken owners (IQR: 170,000-630,000) with 3.7 million exposed chickens (IQR: 1.9 M-6.4 M), and 51,000 cases among duck owners (IQR: 24,000 - 96,000), with 1.2 million exposed ducks (IQR: 0.6 M-2.1 M). The median number of overall human infections in Vietnam for this range of the basic reproductive number is 6.4 million (IQR: 4.4 M-8.0 M).

Conclusion: It is likely that, in the absence of effective interventions, the introduction of a novel H1N1 into a densely populated country such as Vietnam will result in a widespread epidemic. A large epidemic in a country with intense human-animal interaction and continued co-circulation of other seasonal and avian viruses would provide substantial opportunities for H1N1 to acquire new genes.

Background

In early 2009 a novel influenza A (H1N1) variant emerged which spread globally causing the first influenza pandemic in over 40 years. The dynamics and impact of this pandemic are difficult to predict, especially since the world has changed significantly in 40 years - the global population has almost doubled, more people live in cities, people travel more frequently and over longer distances. These facts will undoubtedly influence the global pattern of this pandemic, just as geographical heterogeneities will result in different local patterns [1]. More than 60% of the world's population live in low-income and lower-middle income countries, and yet, at the time of writing only, about 10% of confirmed cases have occurred in these areas [2,3]. In densely populated low-income countries, where public health systems, health care services and drug availability are all stretched, influenza H1N1 is likely to be almost impossible to contain resulting in a greater number of cases occurring in more vulnerable populations resulting in a less benign epidemic.

Even more worrying, almost 60% of the world's human population and over 50% of the world's poultry population live in Asia, where highly pathogenic avian influenza (HPAI) maintains a foothold and seasonal influenza transmission is complex [4]. Previous pandemics have demonstrated the potential consequences of reassortment between human and animal influenza viruses. It is possible, therefore, that the new H1N1 - itself a reassortant of swine viruses that had previously reassorted with human and avian influenza - may follow a similar pattern [5,6]. The new H1N1 variant has already shown that it can be transmitted from humans to pigs, and we know that the H5N1 subtype is capable of infecting humans and of successfully reassorting with human seasonal influenza viruses under *in vitro* and *in vivo* experimental conditions [7,8]. As the population is increasing and standards of living are improving there has been an increase in livestock production and thus there is probably more contact between animals and humans than before. These contacts offer opportunities for reassortment, and, if a novel virus with the transmissibility of H1N1 and even a fraction of the virulence of H5N1 were to emerge, the consequences would be devastating.

In order to explore the potential impact of influenza A (H1N1) on a densely populated low-income country, we developed a mathematical model showing how an influenza A (H1N1) epidemic might progress in Vietnam. We used this model to estimate the frequency of contact between H1N1 infected humans and domestic animals in an attempt to quantify the opportunities for reassortment between H1N1 and animal influenza viruses.

Methods

Mathematical model

We developed an age-structured gravity model - where migration rates among sub-populations are balanced such that there are no changes in the sizes of the sub-populations - based on traditional susceptible exposed infectious recovered (SEIR) equations with stochastic migration and hospitalization processes [9]. The model has geographical resolution to the province level in Vietnam (64 provinces in 2007) and tracks infection and mixing in seven age groups. The incubation period was set at 1.4 days and the infectious stage was separated into four stages to mimic an infectious period that is Γ-distributed with a coefficient of variation equal to 0.5. Mixing and infection among hosts (humans) in the model occurred at the province level and depended on the contact rates among the seven age groups, age-specific susceptibilities, province-specific age distribution and population density. The basic reproductive number, denoted by R_0, is calculated via a next generation matrix assuming at most one cross-province migration event during a single infection [10]. The R_0 value described in the text and figures is for Ho Chi Minh City and assumes that there is no migration from the city (see supplementary materials, additional file 1, for detail on the different R_0 values that can be computed for this model). The results are presented for a single case introduced in Ho Chi Minh City, as this is where the first case was confirmed on 31 May 2009. Infection among animal populations is not modelled. Model equations and details of computing the basic reproduction ratio are presented in the supplementary materials (Additional file 1).

Data sources

We used seven age groups: 0-5 years, 6-15, 16-25, 26-34, 35-49, 50-64 and 65+. Provincial level data on resident population by age class, number of public and private hospital beds, number of households, and number of households raising pigs, chickens, and ducks were derived from the General Statistics Office of Vietnam. The age-class specific daily probability of migration between provinces was derived from a 2007 community survey conducted in northern Vietnam [unpublished data, P Horby]. This gave a mean estimate of 1.35% of the population moving between the provinces each day. This was used as the lower end of the modelled range, as it is known that populations closer to urban areas will have much higher rates of movement. The number of major and minor roads crossing provincial borders was determined from 1:250,000 road maps and were used to obtain a relative measure of interprovincial traffic. Internal migration by air travel was estimated using publicly available flight data from all airlines operating domestically in Vietnam. The known daily travel by air and the unknown daily travel by road were combined to form a scalable migration network

between the provinces of Vietnam where between 1.35% and 5.00% of the population moved between provinces on a daily basis.

Transmission and natural history parameters

Age-dependent mixing was included in the model by creating a contact matrix for seven age groups, using data from a survey of social contact patterns conducted in 2007 among 865 members of a community in one semi-rural district of northern Vietnam. Since both epidemiological and serological data are suggestive of age-dependent susceptibility to H1N1 infection, an age-dependent susceptibility term was also included [11,12]. This was derived using data on the age distribution of cases in the USA and data on age-dependent contact frequency from a European study [13]. We assumed no effect of season on the transmission of infection or on contact patterns, as influenza seasonality in Vietnam is not well understood, even in the northern and more temperate part of the country (unpublished data, PQ Thai).

Since reliable data on the natural history of infection with H1N1 were not available at the time of writing, we applied parameters previously estimated for seasonal influenza. We applied an incubation period with a mean of 1.4 days [14]. The mean of the Γ-distributed infectious period was varied between 3.8 days and 5.5 days [15]. The age-class specific relative probability of hospitalization was derived from data of the proportion of H1N1 cases hospitalized in Mexico and the USA. The overall hospitalization rate was varied between 0.5% and 1.5% of all cases, since reported rates of 5%-6% are likely to be biased by over-ascertainment of severe cases compared to mild cases. Hospitalization time was set at 5 days [16].

Sensitivity analysis

Sensitivity was tested by varying the basic reproduction ratio (1.2 - 3.1), the duration of infection (3.8 - 5.5 days), the individual probability of cross-province migration (1.35% - 5.00% daily probability), the relative amount of traffic on large roads compared to small roads (one to two times), and the overall expected hospitalization in the population (0.5% - 1.5%). One thousand parameter combinations were sampled using Latin hypercube sampling, and sensitivity results are reported for these parameter samples [17]. The key parameter for this sensitivity analysis is R_0, the basic reproductive number. For influenza this is traditionally estimated between 1 and 3 [18-20] and the ranges reported so far for novel H1N1 have been 1.2, 1.4 to 1.6, 2.0 to 2.6, and <2.2 to 3.1 [11,21,22]. For the upper band of our tested range, we used the highest estimate (R_0 = 3.1) as opposed to the highest of the upper band 95% confidence interval (R_0 = 3.5).

Full details of data sources, parameter estimation and model specification are available in the supplementary materials (Additional file 1).

Results

Epidemic curve and geographic spread

Introducing a single infected case in Ho Chi Minh City, and simulating the epidemic for one year (over 1000 randomly sampled parameter sets), resulted in a median 6.4 million infections (IQR: 4.4 million - 8.0 million). In the absence of any intervention, the epidemic would reach half of Vietnam's provinces in 57 days (IQR: 45-86.5), and would peak after 81 days (IQR: 62.5-121). Seventy-seven percent of all cases and 67% of all hospitalizations occur in the 6-34 year age group. Table 1 shows the range of outputs for the model simulations.

The epidemic was dominated by the peaks in Hanoi and Ho Chi Minh City (Figure 1), Vietnam's most densely populated metropolitan areas. Both of these provinces are at least twice as densely populated as any other province in Vietnam. The interval between the 100-case point in Ho Chi Minh City and 100-case point in Hanoi is estimated to be about 29 days (IQR: 23-43), but might be doubled or tripled if a sustained social distancing campaign were able to reduce all contacts by 50%. After the Hanoi wave passes, the epidemic is expected to tail off slowly as the disease spreads to less densely populated rural areas. Figure 2 shows the geographic progression of the median epidemic in Vietnam; Figure 3 shows the median epidemic peak times for all the provinces, indicating an approximate 1-month delay between peaks in the southern provinces and peaks in the northern provinces.

The epidemic in Vietnam is predicted to cause 58,000 hospitalizations (IQR: 39,000-75,000). The health care system would be severely stretched but is unlikely to be overwhelmed, except in the case of a high-R0 epidemic or increased virulence. Vietnam currently has a stockpile of approximately 1.1 million oseltamivir treatment courses (10 75 mg tablets) and sufficient powder to formulate another 900,000 treatment courses. This should be adequate for treatment of severe cases but for not mild cases or for prophylaxis of contacts during a widespread epidemic.

Contacts between infected humans and domestic animals

Because of the slow dispersion of the epidemic into rural areas, the peak exposure of domestic pigs, ducks and chickens to infected humans occurs during the later phases of the epidemic. Figure 1A shows the estimated number of exposures of domestic animals to infected humans; the highest exposure will be among domestic chickens and the exposure of all domestic animals will

Table 1: Median, quartile and minimum - maximum values for selected outputs of one year of model simulation.

Model output	Minimum	Lower quartile	Median	Upper quartile	Maximum
Time to reach 20-case point (days)	9.0	12.0	14.0	19.0	39.0
Time to reach 100-case point (days)	13.0	18.0	22.0	32.0	71.0
Time for 32 provinces to be affected (days)	34.0	45.0	57.0	86.5	314.0
Time for 48 provinces to be affected (days)	41.0	55.0	71.5	112.0	> 1 year
Epidemic peak point (days)	45.0	62.5	80.8	121.0	not reached
Final epidemic size (number of cases)	103,885	4,432,247	6,377,555	8,021,328	9,796,738
Cumulative number hospitalized	594	38,832	58,165	74,935	104.976
Average number of new cases per day over 2-week peak period	1916	88,453	174,804	245,209	326,260
Average number of new hospitalizations per day over 2-week peak period	9	779	1,564	2,238	3508
Cumulative number of cases in swine owners	1940	224,208	410,276	671,703	1,159,291
Cumulative number of cases in chicken owners	630	172,731	351,243	632,316	1,174,682
Cumulative number of cases in duck owners	160	23,732	51,131	95,790	182,520
Number of days Hanoi hospitals running at > 150% bed capacity	0	0	0	14	20
Number of days HCMC hospitals running at > 150% bed capacity	0	0	0	0	11
Time from 100-case point in HCMC to 100-case point in Hanoi	12	23	29	43	156
Number of rural cases	16,105	963,791	1,540,008	2,184,359	3,220,171
Number of urban cases	87,780	3,440,732	4,844,258	5,831,431	6,668,559
Number of exposed pigs	3,053	259,328	462,633	737,443	1,239,324
Number of exposed chickens	4,612	1,890,957	3,745,045	6,417,106	11,419,922
Number of exposed ducks	2,319	580,132	1,176,129	2,072,495	3,708,187
Number of infections by age group					
0 to 5 years	7,611	366,741	543,858	685,656	832,651
6 to 15 years	35,370	1,211,114	1,617,223	2,031,725	2,609,789
16 to 25 years	30,075	1,312,009	1,837,835	2,212,661	2,599,191
26 to 34 years	22,277	998,079	1,450,723	1,784,407	2,096,515
35 to 49 years	6,067	363,076	627,659	870,774	1,134,289
50 to 64 years	1,920	130,147	232,582	329,305	435,306

Table 1: Median, quartile and minimum - maximum values for selected outputs of one year of model simulation. *(Continued)*

65 years and over	565	38,492	71,658	106,032	147,498
Hospitalizations by age group					
0 to 5 years	71	4,493	6,766	8,624	12,144
6 to 15 years	113	5,453	7,467	9,481	13,831
16 to 25 years	119	8,128	11,393	14,091	18,991
26 to 34 years	198	13,580	19,932	24,989	33,959
35 to 49 years	81	5,540	9,424	13,345	19,910
50 to 64 years	10	1,284	2,279	3,310	5,262
65 years and over	2	393	721	1,079	1,792

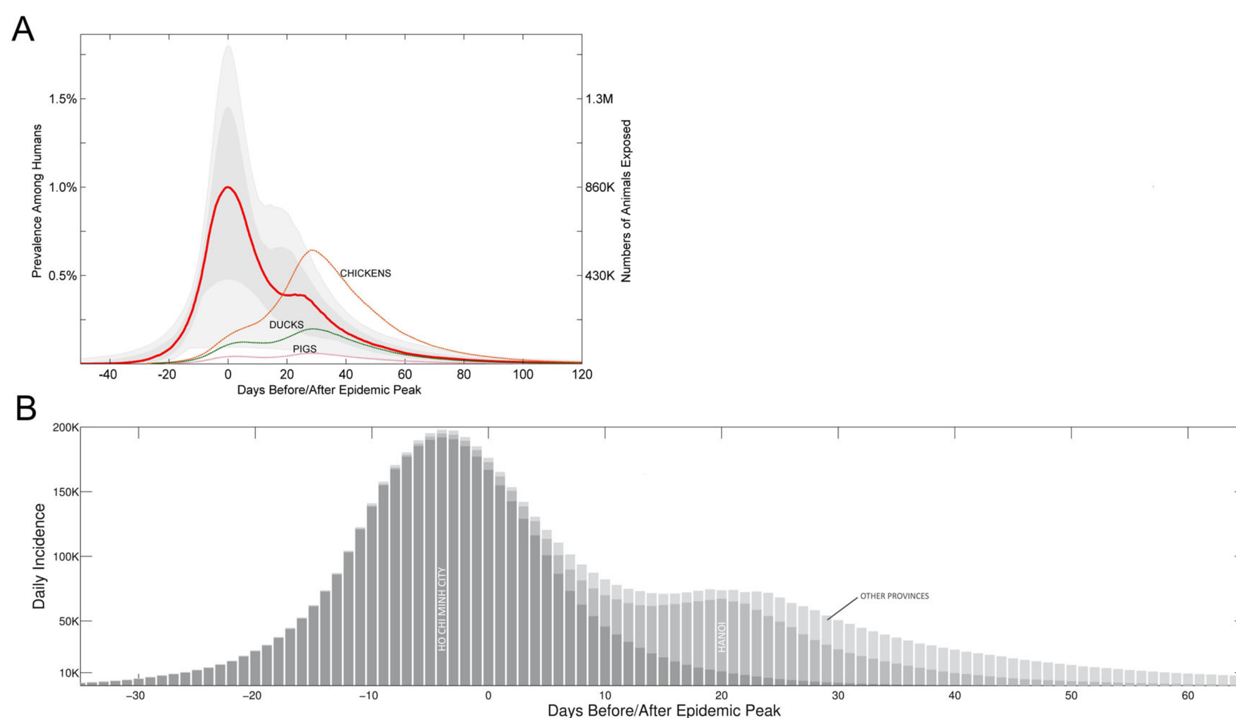


Figure 1

(A) The range of possible epidemics in Vietnam. The graph summarizes 500 simulated epidemics and resets their peaks to day zero so they can be compared on the same time axis. The red line shows the median number of infected persons. The medium gray region shows the interquartile range. The light gray region shows the 95% confidence interval based on the parameter ranges chosen via Latin hypercube sampling. The confidence band width is primarily determined by R_0. The three dotted lines show the median number of exposed animals during the epidemic. **(B) Median number of new cases by day, with day zero corresponding to the epidemic peak**. Stacked bar graph has dark gray bars for Ho Chi Minh City, medium gray bars for Hanoi and light gray bars for the remaining 62 provinces.

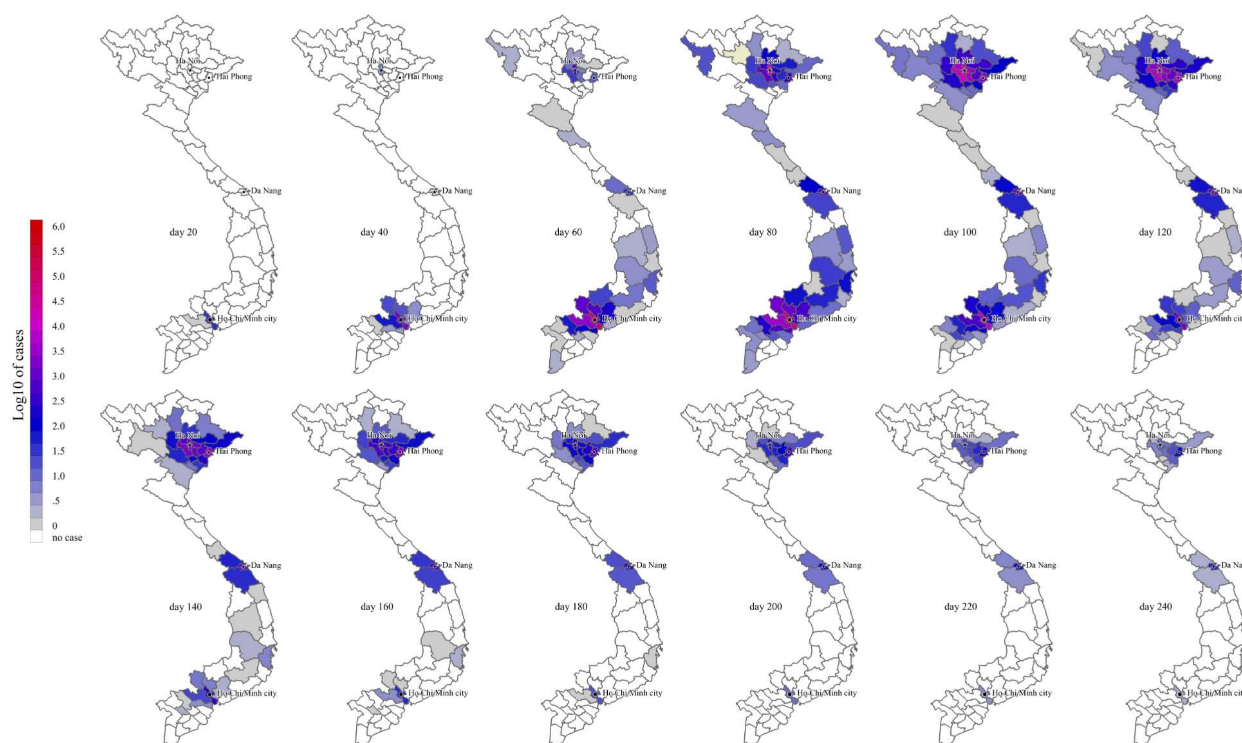


Figure 2
Geographic spread of swine-origin influenza A (H1N1) in Vietnam. Case numbers in each province are medians from 1000 model simulations. See additional file 2 for corresponding animation.

peak roughly 1 month after the peak in Ho Chi Minh City and shortly after the epidemic peak in Hanoi. Note that the tail phase of the epidemic wanes slowly and that a significant number of chickens, ducks and pigs remain exposed for up to 2 months after the human epidemic has peaked in Hanoi (Figure 4).

In total, the epidemic simulations estimate a median 410,000 cases among swine owners (IQR: 220,000 - 670,000) with 460,000 exposed swine (IQR: 260,000-740,000), a median 350,000 cases among chicken owners (IQR: 170,000-630,000) with 3.7 million exposed chickens (IQR: 1.9 M-6.4 M), and a median 51,000 cases among duck owners (IQR: 24,000 - 96,000), with 1.2 million exposed ducks (IQR: 0.6 M-2.1 M).

Effect of public health interventions
By restricting contacts in the 6-15 age group, school closures were modelled but showed little effect on the progression of the epidemic. Even a comprehensive strategy of restricting all contacts within this age group would only delay the epidemic peak by a few days and result in no fewer cases. Any realistic restriction of flights between Ho Chi Minh City and Hanoi (< 2 weeks) had little or no effect on geographic spread or the total number of cases. Monitoring incoming international flights and multiple introductions was not modelled.

Sensitivity analysis
Like all epidemic models, the highest sensitivity is to R_0. All severity indices of the epidemic - total number of cases, peak incidence and total number of hospitalizations - rise steadily with the R_0 value, or, in general, with the transmissibility of the virus (top panels, Figure 5). The most important feature of the model is that with increasing R_0 the epidemic becomes more rural. An increase in the predicted transmissibility of novel H1N1 in Vietnam results not only in more infections, but in a higher proportion of infection among rural populations and among those raising pigs, ducks and chickens domestically (bottom panels, Figure 5). The model is not very sensitive to the other parameters tested: the duration of the infection, the amount of migration between the provinces, the hospitalization rate or the relative amount of traffic on large roads versus small roads.

Discussion
The first cases of H1N1 were detected in Vietnam on 31 May 2009 and by mid-July there were more than 100 confirmed cases with probable community transmission that

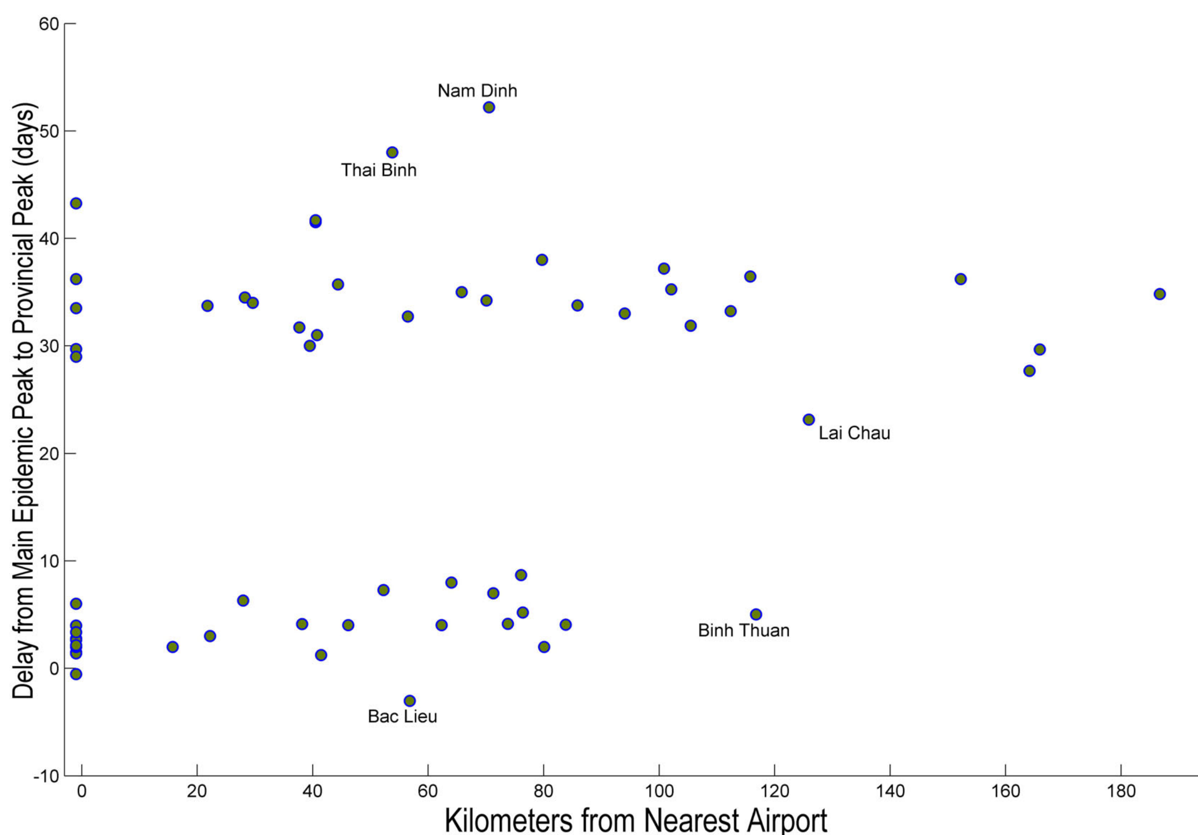


Figure 3
Timing of provincial epidemic peaks based on the distance from the nearest airport to the capital city. The model does not take sub-provincial population structure into account, and the epidemic's progression is determined primarily by south-to-north movement rather than distance to the airport network. Binh Thuan has an early peak because it lies in a densely populated part of southern Vietnam. The Lai Chau peak, as estimated by our model, probably occurs too early. Lai Chau is remote and sparsely populated, but its adjacency to the Dien Bien Phu airport causes the model to predict an early epidemic peak.

was most likely the initial budding of the coming nationwide epidemic. We have used a mathematical model to explore how the epidemic might progress in the absence of interventions and have estimated the number of pigs, ducks and chickens that might be exposed to infected humans during the epidemic. Employing mathematical modelling for such a forecasting exercise comes with many caveats. Of these, the most important are that real individuals are heterogeneous in behaviour and transmission, that human behaviour can change as a result of the severity of the epidemic and that the spatial dimensions of transmission have many nested levels that may or may not alter the progression of the epidemic on a larger scale [23,24]. We used a 'patch model' with coarse province-level spatial resolution for simplicity of model development and rapid computation; the model results should, therefore, be viewed as rough estimates of the epidemic's impact in Vietnam on a year-long time scale.

The most important caveat in our analysis is that the true basic reproductive number is not known; we used a conservative estimate, between 1.2 and 3.1, based on early measurements taken in Mexico, USA and Japan, and we stress that the R_0 for Vietnam may be higher than these estimates. For an R_0 value of 4.0, our model predicted a total of 13.3 million cases among humans; for an R_0 value of 5.0, 16.6 million cases were predicted. Unfortunately, the uncertainty in Vietnam's R_0 will not be resolved until we analyse the progression of cases from the first wave of this pandemic.

Although the model predicts substantially more cases than have so far been reported from other H1N1 affected countries, the clinical illness is predominantly mild and, therefore, reported H1N1 cases to date reflect only a small proportion of the total number of cases. Our modelled epidemic affects a median of 7.4% of the population

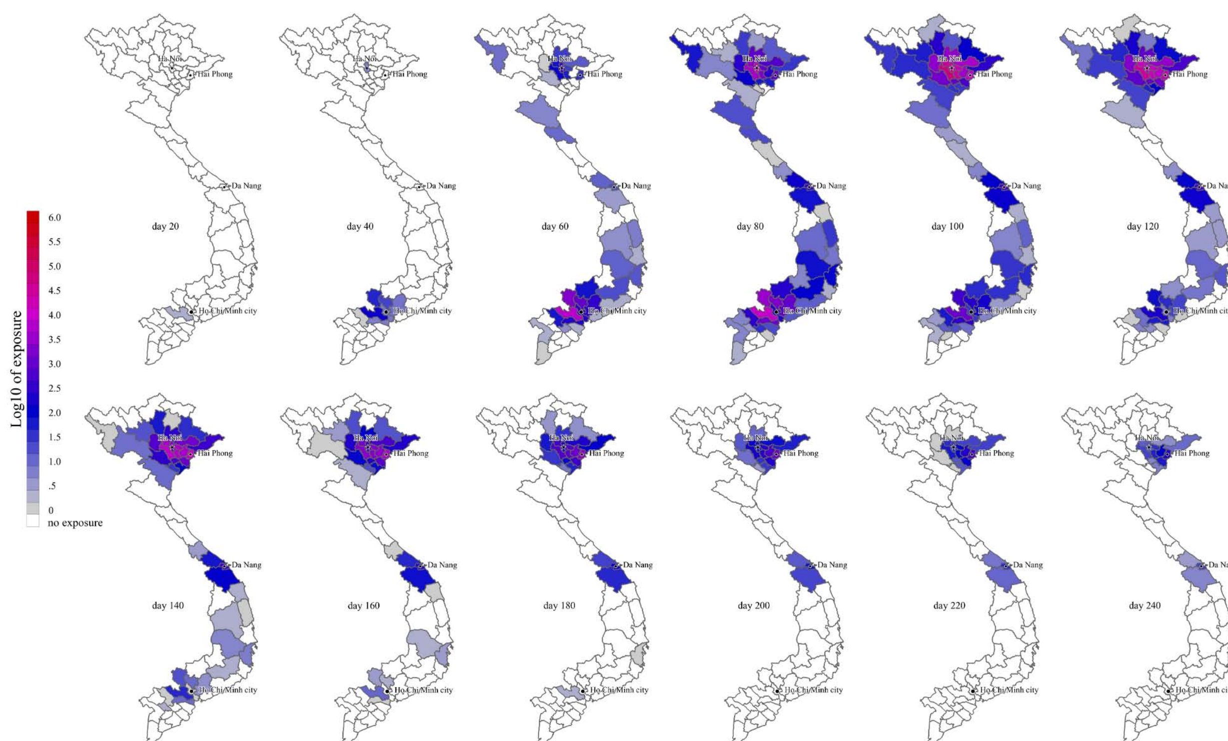


Figure 4

Geographic timeline of chicken exposure during an influenza epidemic in Vietnam. The numbers of chicken exposures are medians from 1000 models simulations. Duck and pig exposures were highly correlated with chicken exposures, geographically and temporally. Note that because of rounding and fractional cases, some sparsely-populated provinces may have a median of 0.2 human cases (rounded down to zero) and 0.8 chicken exposures (rounded up to 1). See additional file 3 for corresponding animation.

(IQR: 5.2-9.3%). This rate is below the figures for previous pandemics and might be considered too low given the high transmissibility of this virus in some settings and the expectation that most of the population would have no immunity [12]. Due to the low probability of hospitalization, it is unlikely that the health sector as a whole will be over-whelmed in the scenario outlined in this model. However, there is considerable variation in reported hospitalization rates for H1N1 and the estimate of 1% that we have used is considerably lower than the maximum of 6% [25]. As elsewhere, the number of intensive care beds is limited in Vietnam and occupancy is routinely at maximum; therefore intensive care capacity is likely to be easily overwhelmed. Also, although Vietnam has impressive health indicators for its economic status - the population may have vulnerabilities, such as under-nutrition in children, which might result in a greater number of severe cases than observed elsewhere.

Containment does appear to have been temporarily successful in some countries (Mexico and Japan) but not in others (Australia and the USA). The reasons for these differences are undoubtedly complex, but successful case detection, isolation and treatment, quarantine and chemoprophylaxis of contacts, and social distancing measures, may all have an effect on the results. In our model, school closures did not make a substantial difference to the epidemic progression, although substantial decreases in contact frequency across all age groups would delay the time course of the epidemic. School outbreaks have been a major feature in the early stages of this pandemic, and it is possible that our model underestimates the role of the range of contacts and susceptibility of school-age children on the epidemic dynamics. School closures did seem to be effective in Kobe, Japan, during 11-24 May 2009, but this may have reflected the low number of overall infections in Japan at that time (between four and 345 confirmed cases) [26]. In the UK, a plateau in consultation rates appears to have coincided with the closure of schools for the annual summer holidays [27]. Previous work suggests that school closure can modify peak attack rates and may result in a modest reduction of the final number of cases,

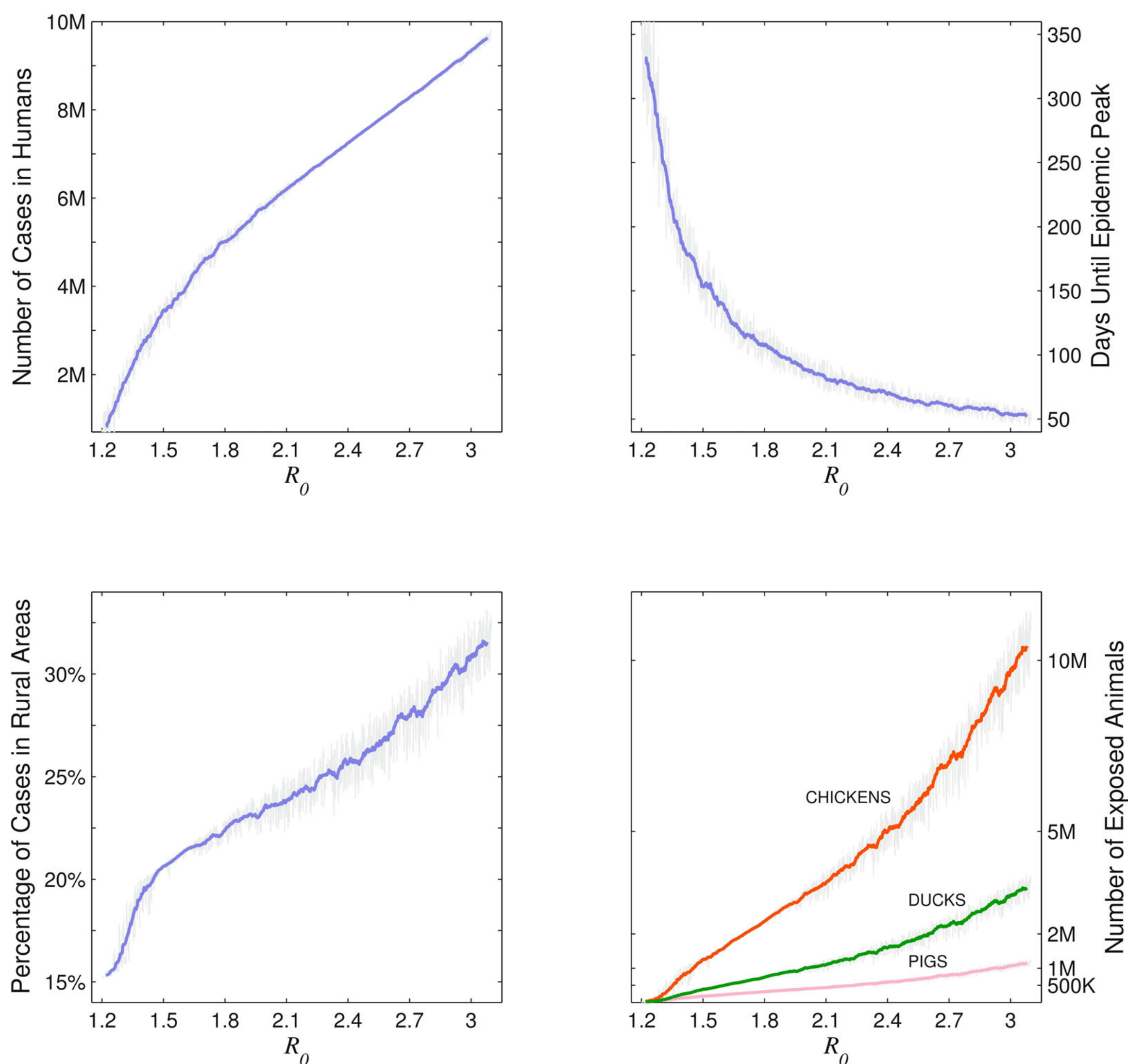


Figure 5
Result sensitivity relative to the R_0-value as it would have been measured in Ho Chi Minh City. Light gray lines show the variation in a particular epidemiological indicator as a function of R_0. The other coloured lines are moving averages over nearby R_0-values. The top two panels show the size of the epidemic and the time taken for it to peak, which always have a predictable relationship to R_0. The bottom two panels show how animal exposure increases and how the epidemic becomes more rural as R_0 increases. Note that with higher R_0, not only does the risk to domestic animal owners increase but the relative risk of an owner to a non-owner also increases (not shown).

but empiric data is still required on the effectiveness of school closure on reducing the number of transmissions [28-31]. Climate and other seasonally variable factors may also have acted to limit transmission in temperate regions [32,33]. Seasonal factors are likely to have less influence in tropical regions where the seasonality of influenza transmission is much less marked [4].

In the absence of effective interventions, we predict a large amount of contact between infected humans and animals that might harbour other influenza viruses, including HPAI. In fact, we believe our model probably underestimates the amount of contact between infected human and animals for three reasons. First, we divided the total number of human cases by the number of people per household in order to derive an estimate of the number of households with an infectious case. We did this to avoid over counting animals that were exposed to multiple infected individuals in the same household, but this is a very conservative correction. Second, domestic animal production is concentrated close to urban centres, where population densities are higher than average. Third, we did not model contacts which occurring in live poultry markets or commercial farms.

The danger of human-animal contact lies in the opportunity for reassortment among different influenza subtypes. It is well known that influenza reassorts in humans [34], that pigs play an important role in reassortment of human/avian/swine influenza viruses [35-37] and that the history of avian influenza viruses includes multiple reassortment events [38,39]. However, very little is known about the potential of human influenza viruses to jump to animals, since most studies to date have focused on animal influenza activity and the risk it poses to humans [40-42]. Pandemic H1N1 has already been detected in swine and, since poultry and swine populations in Asia may harbour many different subtypes of influenza (at least H4, H5, H6, H7, H9, H11, H12), the generation of a new subtype through a reassortment event is a real possibility [43,44] [personal communication, Ken Inui].

Although these opportunities for genetic reassortment are not unique, the current influenza landscape contains worrying features. Widespread epidemics of novel H1N1 are likely in tropical countries where HPAI is endemic and seasonal influenza transmission is complex and sustained, without the seasonal bottlenecks that characterize transmission in temperate regions [4,33]. The overall diversity of influenza viruses in southeastern Asia ensures that an epidemic of the novel H1N1 will create many opportunities for co-infection with other subtypes circulating in the region. Genetic and antigenic data suggest that Asia is a key source of influenza viruses that cause seasonal outbreaks in the northern and southern hemispheres [45]. This region, therefore, possesses the conditions necessary for the genesis and dissemination of new influenza variants [33,45]. Finally, the introduction of H1N1 into southeastern Asia creates an optimal evolutionarily environment for the virus, where re-assortment is neither too frequent nor too rare [46]. This means the virus receives the benefits of limited reassortment (a genetic novelty) but not the penalty of high levels of reassortment (the breaking apart of beneficial gene combinations).

Our model provides a rough picture of what might happen in Vietnam, but it includes many assumptions, uncertainties and un-modelled heterogeneities which require that the results be interpreted with caution. Although changes in human demography and migration over the past 40 years may make a pandemic more difficult to control, the same period has seen massive advances in technology and communication that allow us to monitor and predict this pandemic as never before. Mathematical models are one tool, but a criticism of these models is that the predictions are not subsequently tested against real outbreak data [47]. Our model development has coincided with the arrival of H1N1 in Vietnam and we are planning to track the progression of the outbreak in Vietnam in an attempt at real-time model validation and diagnostics.

Abbreviations
IQR: interquartile range; HPAI: highly pathogenic avian influenza; SEIR: susceptible exposed infections recovered.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
PH and MFB conceived the study and designed the model structure. PH, BHM, PQT and MFB collated the data and model parameters. MFB wrote the model code and ran the model and sensitivity analysis. BHM prepared all maps and video sequences. TTH, NTH, NVK and JF provided data and advised on the model design. PH and MFB wrote the first draft of the paper. All the authors reviewed and edited drafts of the manuscript and approved the final version. PH, BHM, MFB contributed equally.

Additional material

Additional file 1
Supplementary materials*. Describes details of model construction, data sources, parameter estimation, R_0 calculation, and sensitivity analysis.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1741-7015-7-43-S1.pdf]

Additional file 2

Geographic spread of swine-origin influenza A (H1N1) in Vietnam. *Animated GIF file that shows the full day-by-day epidemic shown in Figure 2.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1741-7015-7-43-S2.gif]

Additional file 3

Geographic timeline of chicken exposure during an influenza epidemic in Vietnam. *Animated GIF file that shows the full day-by-day exposure of chickens to human influenza infections shown in Figure 4.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1741-7015-7-43-S3.gif]

Acknowledgements

We are grateful to the Ministry of Health of the Socialist Republic of Vietnam for their continued support for our work and to the staff of the National Institute for Infectious and Tropical Diseases, the National Institute of Hygiene and Epidemiology, and the Hospital for Tropical Diseases for their dedication to high quality research into infectious diseases in Vietnam. This work was supported by the Wellcome Trust UK (grants 081613/Z/06/Z and 077078/Z/05/Z) and the South East Asia Infectious Disease Clinical Research Network (N01-A0-50042). Model simulations were run at the computing facilities of the Wellcome Trust Sanger Institute. We also thank Ms Ho Thi Nhan for gathering the domestic flight information for Vietnam. MFB is funded by a UK Medical Research Council grant G0600718 to Dominic Kwiatkowski.

References

1. Miller MA, Viboud C, Balinska M, Simonsen L: **The signature features of influenza pandemics - implications for policy.** *N Engl J Med* 2009, **360**:2595-2598.
2. World_Bank: **World Development Indicators database.** 2009.
3. **Influenza A(H1N1) - update 58** [http://www.who.int/csr/don/2009_07_06/en/index.html]. accessed July 6, 2009
4. Viboud C, Alonso WJ, Simonsen L: **Influenza in tropical regions.** *PLoS Med* 2006, **3(4)**:e89.
5. Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V, Okomo-Adhiambo A, Gubareva L, Barnes J, Smith CB, Emery SL, Hillman MJ, Rivailler P, Smagala J, de Graaf M, Burke DF, Fouchier RAM, Pappas C, Alpuche-Aranda CM, Lopez-Gatell H, Olivera H, Lopez I, Myers CA, Faix D, Blair PJ, Yu C, Keene KM, Dotson PD Jr, Boxrud D, Sambol AR, Abid SH, St George K, Bannerman T, Moore AL, Stringer DJ, Blevins P, Demmler-Harrison GJ, Ginsberg M, Kriner P, Waterman S, Smole S, Guevara HF, Belongia EA, Clark PA, Beatrice ST, Donis R, Katz J, Finelli L, Bridges CB, Shaw M, Jernigan DB, Uyeki TM, Smith DJ, Klimov AI, Cox NJ: **Antigenic and Genetic Characteristics of Swine-Origin 2009 A(H1N1) Influenza Viruses Circulating in Humans.** *Science* 2009, **325(5937)**:197-201.
6. Trifonov V, Khiabanian H, Rabadan R: **Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus.** *N Engl J Med* 2009, **361**:115-119.
7. Jackson S, Van Hoeven N, Chen LM, Maines TR, Cox NJ, Katz JM, Donis RO: **Reassortment between avian H5N1 and human H3N2 influenza viruses in ferrets: a public health risk assessment.** *J Virol* 2009, **83(16)**:8131-40.
8. **An Alberta swine herd investigated for H1N1 flu virus** [http://www.inspection.gc.ca/english/corpaffr/newcom/2009/20090502e.shtml]
9. Heesterbeek ODaJAP: **Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation.** Chichester: John Wiley & Sons Ltd; 2000.
10. Diekmann O, Heesterbeek JA, Metz JA: **On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations.** *J Math Biol* 1990, **28(4)**:365-382.
11. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, Griffin J, Baggaley RF, Jenkins HE, Lyons EJ, Jombart T, Hinsley WR, Grassly NC, Balloux F, Ghani AC, Ferguson NM, Rambaut A, Pybus OG, Lopez-Gatell H, Alpuche-Aranda CM, Bojorquez Chapela I, Palacios Zavala E, Espejo Guevara DM, Checchi F, Garcia E, Hugonnet S, Roth C, The WHO Rapid Pandemic Assessment Collaboration: **Pandemic Potential of a strain of influenza A (H1N1): early findings.** *Science* 2009, **324(5934)**:1557-1561.
12. **Serum cross-reactive antibody response to a novel influenza A (H1N1) virus after vaccination with seasonal influenza vaccine.** *MMWR Morb Mortal Wkly Rep* 2009, **58(19)**:521-524.
13. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ: **Social contacts and mixing patterns relevant to the spread of infectious diseases.** *PLoS Med* 2008, **5(3)**:e74.
14. Lessler J, Reich NG, Brookmeyer R, Perl TM, Nelson KE, Cummings DA: **Incubation periods of acute respiratory viral infections: a systematic review.** *Lancet Infect Dis* 2009, **9(5)**:291-300.
15. Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boelle PY: **A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data.** *Stat Med* 2004, **23(22)**:3469-3487.
16. **Hospitalized patients with novel influenza A (H1N1) virus infection - California, April-May, 2009.** *MMWR Morb Mortal Wkly Rep* 2009, **58(19)**:536-541.
17. Blower SM, Dowlatabadi H: **Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example.** *Int Statist Rev* 1994, **62(2)**:229-243.
18. Chowell G, Ammon CE, Hengartner NW, Hyman JM: **Estimating the reproduction number from the initial phase of the Spanish flu pandemic waves in Geneva, Switzerland.** *Math Biosci Eng* 2007, **4(3)**:457-470.
19. Chowell G, Miller MA, Viboud C: **Seasonal influenza in the United States, France, and Australia: transmission and prospects for control.** *Epidemiol Infect* 2008, **136(6)**:852-864.
20. Mills CE, Robins JM, Lipsitch M: **Transmissibility of 1918 pandemic influenza.** *Nature* 2004, **432(7019)**:904-906.
21. Nishiura H, Castillo-Chavez C, Safan M, Chowell G: **Transmission potential of the new influenza A(H1N1) virus and its age-specificity in Japan.** *Euro Surveill* 2009, **14(22)**:.
22. Boelle PY, Bernillon P, Desenclos JC: **A preliminary estimation of the reproduction ratio for new influenza A(H1N1) from the outbreak in Mexico, March-April 2009.** *Euro Surveill* 2009, **14(19)**:.
23. Riley S: **Large-scale spatial-transmission models of infectious disease.** *Science* 2007, **316(5829)**:1298-1301.
24. Watts DJ, Muhamad R, Medina DC, Dodds PS: **Multiscale, resurgent epidemics in a hierarchical metapopulation model.** *Proc Natl Acad Sci USA* 2005, **102(32)**:11157-11162.
25. **Considerations for assessing the severity of an influenza pandemic.** *Wkly Epidemiol Rec* 2009, **84(22)**:197-202.
26. **Human infection with new influenza A (H1N1) virus: clinical observations from a school-associated outbreak in Kobe, Japan, May 2009.** *Wkly Epidemiol Rec* 2009, **84(24)**:237-244.
27. **Weekly pandemic flu update - 30 July 2009** [http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1248940838384?p=1231252394302]
28. Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS: **Strategies for mitigating an influenza pandemic.** *Nature* 2006, **442(7101)**:448-452.
29. Glass K, Barnes B: **How much would closing schools reduce transmission during an influenza pandemic?** *Epidemiology* 2007, **18(5)**:623-628.
30. Cauchemez S, Valleron AJ, Boelle PY, Flahault A, Ferguson NM: **Estimating the impact of school closure on influenza transmission from Sentinel data.** *Nature* 2008, **452(7188)**:750-754.
31. Cauchemez S, Ferguson NM, Wachtel C, Tegnell A, Saour G, Duncan B, Nicoll A: **Closure of schools during an influenza pandemic.** *Lancet Infect Dis* 2009, **9(8)**:473-481.

32. Shaman J, Kohn M: **Absolute humidity modulates influenza survival, transmission, and seasonality.** *Proc Natl Acad Sci USA* 2009, **106(9):**3243-3248.
33. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC: **The genomic and epidemiological dynamics of human influenza A virus.** *Nature* 2008, **453(7195):**615-619.
34. Holmes EC, Ghedin E, Miller N, Taylor J, Bao Y, St George K, Grenfell BT, Salzberg SL, Fraser CM, Lipman DJ, Taubenberger JK: **Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses.** *PLoS Biol* 2005, **3(9):**e300.
35. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y: **Evolution and ecology of influenza A viruses.** *Microbiol Rev* 1992, **56(1):**152-179.
36. Brown IH, Harris PA, McCauley JW, Alexander DJ: **Multiple genetic reassortment of avian and human influenza A viruses in European pigs, resulting in the emergence of an H1N2 virus of novel genotype.** *J Gen Virol* 1998, **79(Pt 12):**2947-2955.
37. Zhou NN, Senne DA, Landgraf JS, Swenson SL, Erickson G, Rossow K, Liu L, Yoon K, Krauss S, Webster RG: **Genetic reassortment of avian, swine, and human influenza A viruses in American pigs.** *J Virol* 1999, **73(10):**8851-8856.
38. Chen H, Deng G, Li Z, Tian G, Li Y, Jiao P, Zhang L, Liu Z, Webster RG, Yu K: **The evolution of H5N1 influenza viruses in ducks in southern China.** *Proc Natl Acad Sci USA* 2004, **101(28):**10452-10457.
39. Vijaykrishna D, Bahl J, Riley S, Duan L, Zhang JX, Chen H, Peiris JS, Smith GJ, Guan Y: **Evolutionary dynamics and emergence of panzootic H5N1 influenza viruses.** *PLoS Pathog* 2008, **4(9):**e1000161.
40. Pfeiffer DU, Minh PQ, Martin V, Epprecht M, Otte MJ: **An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data.** *Vet J* 2007, **174(2):**302-309.
41. Gilbert M, Xiao X, Pfeiffer DU, Epprecht M, Boles S, Czarnecki C, Chaitaweesub P, Kalpravidh W, Minh PQ, Otte MJ, Martin V, Slingenbergh J: **Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia.** *Proc Natl Acad Sci USA* 2008, **105(12):**4769-4774.
42. Minh PQ, Schauer B, Stevenson M, Jones G, Morris RS, Noble A: **Association between human cases and poultry outbreaks of highly pathogenic avian influenza in Vietnam from 2003 to 2007: a nationwide study.** *Transbound Emerg Dis* 2009 in press.
43. Palese P: **Influenza: old and new threats.** *Nat Med* 2004, **10(12 Suppl):**S82-87.
44. Jadhao SJ, Nguyen DC, Uyeki TM, Shaw M, Maines T, Rowe T, Smith C, Huynh LP, Nghiem HK, Nguyen DH, Nguyen HK, Nguyen HH, Hoang LT, Nguyen T, Phuong LS, Klimov A, Tumpey TM, Cox NJ, Donis RO, Matsuoka Y, Katz JM: **Genetic analysis of avian influenza A viruses isolated from domestic waterfowl in live-bird markets of Hanoi, Vietnam, preceding fatal H5N1 human infections in 2004.** *Arch Virol* 2009, **154(8):**1249-61.
45. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, Gust ID, Hampson AW, Hay AJ, Hurt AC, de Jong JC, Kelso A, Klimov AI, Kageyama T, Komadina N, Lapedes AS, Lin YP, Mosterin A, Obuchi M, Odagiri T, Osterhaus AD, Rimmelzwaan GF, Shaw MW, Skepner E, Stohr K, Tashiro M, Fouchier RA, Smith DJ: **The global circulation of seasonal influenza A (H3N2) viruses.** *Science* 2008, **320(5874):**340-346.
46. Kimura M: **A model of a genetic linkage system which leads to closer linkage by natural selection.** *Evolution* 1956, **10(3):**278-287.
47. Smith DJ: **Predictability and preparedness in influenza control.** *Science* 2006, **312(5772):**392-394.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1741-7015/7/43/prepub

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

Supplementary material

Spatial structure

The boundaries of the 64 Provinces of Vietnam in 2009 were obtained in GIS format from the Global Mapping Project [http://www.iscgm.org/cgi-bin/fswiki/wiki.cgi]. All disease dynamics and relevant population characteristics were aggregated at the province level.

Age structure

Seven age classes were used: 0-5 years, 6-15, 16-25, 26-34, 35-49, 50-64, 65+. These age classes were chosen since they correspond to the age classes used in a community survey of contact and travel frequency.

Population

The estimated resident population, rural and urban, in 2007 by Province was obtained from the Government Statistics Office of Vietnam (GSO). The 2007 population estimates are compiled from statistical returns to GSO from Provincial and District Statistical Offices. The estimated population by seven age classes was derived by applying the age structure from the 1999 population census (originally 18 age classes) to the 2007 figures. Province specific population density was estimated using 2007 population data and provincial land area calculated in GIS after excluding forested areas (as classified by 2005 vegetation map from Ministry of Agriculture and Rural Development).

Age dependent mixing

Age class specific contact frequency was derived from a survey of social contact patterns conducted in 2007 in 865 members of a community in one semi-rural district of north Vietnam. A contact was defined as:

Either

- a two-way conversation with three or more words in the physical presence of another person

Or

- physical skin-to-skin contact (for example a handshake, hug, kiss or contact sports).

Participants recorded every contact made during one day, the age of the contact, and the duration and location of each contact. If a person was contacted more than once in a day, the contact was recorded only once but the total time spent with that contact over the entire day was recorded. A contact intensity matrix by seven age classes was constructed by adjusting the daily contact frequency data by the size of each age class. The matrix was corrected for reciprocity, i.e. where the contact frequency between age classes was not symmetrical the mean of the two values was used for both. The data were normalized so that the all age-class specific rates were relative to the maximum rate of 1.

The contact rate matrix is:

AGE CLASS	0 TO 5	6 TO 15	16 TO 25	26 TO 34	35 TO 49	50 TO 64	>=65
0 TO 5	.5797	.1726	.2000	.4804	.1873	.1760	.0489
6 TO 15	.1726	.8722	.1986	.2331	.2495	.1193	.1427
16 TO 25	.2000	.1986	.6889	.2782	.3162	.3344	.1630
26 TO 34	.4804	.2331	.2782	.6777	.3713	.4990	.3141
35 TO 49	.1873	.2495	.3162	.3713	.4836	.4593	.3504
50 TO 64	.1760	.1193	.3344	.4990	.4593	1.000	.5606
>=65	.0489	.1427	.1630	.3141	.3504	.5606	.9916

and it represents relative contact rates corrected for population size, meaning these are the relative contact rates one would observe if each of the age classes had the same number of individuals. The true contact patterns are different for each province since the age distribution in each province is somewhat different. This pattern is similar to one calculated from census data from Portland, Oregon (USA), in that the main areas of contact intensity are (1) within age-groups and (2) peaking in children and older adults [1].

Note that these contact patterns differ from the European data presented in Mossong et al, the main difference being the high intensity of contacts in the 50+ age groups in Vietnam [2]. The next-generation matrix computed for these data is different for each province in Vietnam (since age structure varies by province), and it is generally different than the Dutch next-generation matrix presented in Wallinga et al [3]. According to these next-generation matrices, infection patterns in Vietnam would primarily be driven by the 6-15 age group, while infection in the Netherlands would be evenly driven by 6- to 39-year-olds.

Internal migration by land

Age class specific frequency of travel outside of the Province of residence was estimated from a questionnaire based survey of travel patterns conducted in 2007 in 865 members of a community in one semi-rural district of north Vietnam.

A 64 x 64 adjacency matrix was compiled for all 64 Provinces in Vietnam that recorded which Provinces share a common border. For each pair of adjacent Provinces, the number of major and minor roads crossing the common border was determined from 1:250.000 road maps. Key national highways were classified as major roads and other national roads, provincial roads, and bridges were classified as minor roads. Major roads were given a value of 1 and minor road value was varied between 0.5 and 1.0 in our sensitivity analysis; this was deemed a reasonable range of variation as we did not have any other measure of the amount of traffic on large roads relative to small roads. These values were entered into a 64 x 64

connectivity matrix. A special direct connection was created between Hanoi and Hai Phong (the major port in north Vietnam) in the model, since a large amount of road traffic travels between these two Provinces along one major road but the two Provinces share no adjacent border, so connectivity was thought to be under-estimated in the model.

Province and age-class specific frequency of travel overland outside of the Province of residence was estimated by adjusting the frequency of travel survey data by the relative connectivity of each Province.

Internal migration by waterways and railways was ignored in the model since GSO data on the volume of traffic by type of transport indicated that waterway and railway travel together contributed less than 10% of all passenger volume in 2007.

Internal migration by air

Internal migration by air travel was estimated using publicly available data on the frequency of flights and aircraft type between domestic airports in Vietnam. It was assumed that all flights were full since data on the number of passengers was not available. These data are publicly available for Vietnam Airlines (http://www.vietnamair.com.vn/wps/portal/vn/site/flight_info/time_table), JetStar (http://www.jetstar.com/vn/vi/cheap-flights/standard.aspx), and Indochina Airlines (http://www.indochinaairlines.vn/FliSchedule.aspx).

Disease natural history

Data on the natural history of novel H1N1 were derived from published surveillance, epidemiological and clinical data and from a model describing the pandemic potential of H1N1 [4]. See table S1 for parameter values used:

Table S1

Parameter	Source	Value
Incubation period	[5]	Mean 1.4 (exponential distribution mimicked via standard linear term in differential equations)
Infectious period	[6]	Mean varied between 3.8 days and 5.5 days (gamma distribution mimicked by four infectious compartments in SEIR model)
Basic reproduction ratio*	1.4-1.6 (genetic analysis 1.2) [4] Less than 2.2-3.1 [7] Japan 2.3 (2.0-2.6) [8] School outbreak in the U.S. 2.69 (2.20-3.2) ProMED-mail Influenza A (H1N1) - worldwide (51): 20090529.1999	Range 1.2 – 3.1

* number of secondary cases created by a single infectious case introduced into a fully susceptible population

The overall hospitalization rate was set at 1% and the distribution of these hospitalizations across the age-classes was derived from data of the proportion of H1N1 cases hospitalized in Mexico and the U.S by age [9]. In the model the overall hospitalization rate was varied between 0.5% and 1.5% of all cases, since reported rates of 5-6% are likely to be biased by over-ascertainment of severe cases compared to mild cases. (Table S2).

Table S2

Age class	Probability of a case being hospitalized
Age 0-5	0.013
Age 6-15	0.005
Age 16-25	0.006
Age 26-34	0.014
Age 35-49	0.016
Age 50-64	0.010
Age 65+	0.011

Age dependent susceptibility to infection

Fraser et al found that the model that best fit the available data included both an age-dependent contact intensity parameter and age dependent susceptibility to infection [4]. Studies of age specific serological reactivity against the novel H1N1 virus are consistent with an age-dependent susceptibility to infection or disease [10]. Since age-dependent contact intensity is represented in our model by the contact matrix described on page 2 of this supplement, we sought to estimate the contribution of age-dependent susceptibility. To do this we used data on the age distribution of cases in the US and data on age dependent contact frequency from a European study [2, 9]. The relative frequency of contact by age class was factored out of the relative distribution of H1N1 cases by age in the US in order to derive an estimate of susceptibility by age class, independent of contact behaviour (Table S3).

Table S3

Age class	Relative susceptibility to infection
Age 0-5	0.77
Age 6-15	1.00
Age 16-25	0.82
Age 26-34	0.59
Age 35-49	0.19
Age 50-64	0.17
Age 65+	0.10

Seasonality

We assumed no seasonal affect on transmissibility of H1N1 or on contact patterns since the seasonality of human influenza in Vietnam is not predictable. Although clear peaks in influenza activity are observed, several peaks are observed each year and the timing of these peaks is not predictable and can occur throughout the year.

Health care capacity

Data on the number of public and private hospital beds, and the number of doctors and nurses by province in 2007 was obtained from the GSO. Number of ventilators by Province in 2007 was obtained from the Ministry of Health, Health Statistics YearBook 2007. The number of immediately available doses of oseltamivir in Vietnam in May 2009 was obtained from the Ministry of Health (not used in model).

Domestic pig, chicken and duck raising

Data on the number of households raising pigs, chickens and ducks by province were obtained from the GSO 2006 Rural, Agricultural and Fishery census. The denominator for estimating the proportion of

households raising pigs, chickens and ducks by province (the total number of households) was obtained from the 1999 Population and Housing census. The number of households in 1999 by province was inflated proportional to the increase in Province specific population observed between 1999 and 2007. Since two new Provinces were created in 2004 by splitting two single Provinces, the number of households in 1999 in each of the two new Provinces was estimated by dividing the number of households in each original Province in 1999 proportionately to the population in the new and old Provinces in 2007.

Contact between infectious humans and domestic pigs, chickens and ducks

To estimate the number of domestic pigs, chickens and ducks exposed to an H1N1 infected human we divided the total number of infected persons in each province by the average household size for each province to give a conservative estimate of the number of infected households. This is a conservative estimate since it assumes all human H1N1 cases are clustered by household. The estimated number of infected households was then multiplied by the proportion of households raising pigs, chickens and ducks and the average number of pigs, chickens and ducks present in households that raise these animals.

Interventions

We explored the potential impact of school closure by introducing a relative reduction in contact frequency among children in the age class 6 and 15 years. We explored the potential impact of broader social distancing measures by reducing contact frequency across all age classes.

Mathematical Model

An SEIR-model (Susceptible-Exposed-Infectious-Recovered) with a four-stage infectious period was used to model the core infection dynamics in each province. The model equations are

$$\begin{aligned}
\dot{S}_{kl} &= -\Lambda_{kl} \cdot \frac{S_{kl}}{N_l} \\
\dot{E}_{kl} &= \Lambda_{kl} \cdot \frac{S_{kl}}{N_l} - eE_{kl} \\
\dot{I}_{kl,1} &= eE_{kl} - 4\nu I_{kl,1} \\
\dot{I}_{kl,2} &= 4\nu I_{kl,1} - 4\nu I_{kl,2} \\
\dot{I}_{kl,3} &= 4\nu I_{kl,2} - 4\nu I_{kl,3} \\
\dot{I}_{kl,4} &= 4\nu I_{kl,3} - 4\nu I_{kl,4}
\end{aligned}$$

where Λ_{kl} represents the force of infection on age class k in province l. Λ_{kl} is defined by

$$\Lambda_{kl} \;=\; \beta_k d_l \sum_{i=1}^{A} z_{ik}(\tau_1 I_{il,1} + \tau_2 I_{il,2} + \tau_3 I_{il,3} + \tau_4 I_{il,4})$$

where A=7 is the total number of age classes and the τ-variables allow for stage-specific infectivities. The variable S_{kl} represents the number of susceptible individuals in age class k currently in province l. E_{kl} represents exposed individuals and $I_{kl,s}$ represents infected individuals in stage s (out of a total of four) of their infection.

The parameters τ_k were all set to one since we could not find good information on shedding and infection duration at the time the model analysis was run. The parameters β_k represent age-specific susceptibility and can be found in Table S3. Parameters d_l represent province-specific population density and were computed as outlined on page 1 of this supplement. Parameters z_{ik} are mixing rates between age class i and age class k (from contact rate matrix on page 2). The parameter v is the recovery rate ($1/v$ is the duration of infection), and $1/e$ is the length of exposure before a host becomes infectious.

Migration and hospitalization were integrated stochastically into the above differential-equations model. Once a day, discrete individuals could move from one province to another according to the migration matrix outlined earlier; migration probabilities were balanced between provinces so the system behaved like a gravity model. In addition, once a day, discrete individuals in stage 4 of their infection could be hospitalized based on their age-dependent probability of hospitalization and the availability of hospital beds in their province. Hospitalization time was five days. Numbers of hospitalized and migrated individuals were drawn from a Poisson distribution.

The basic reproduction ratio R_0 for the entire model can be calculated with standard next-generation matrix methods (Diekmann and Heesterbeek, 2000). Assuming at most one cross-province migration event during a single infection, a 448 by 448 matrix can be built showing the expected number of infections generated by an individual in age class k and location l in individuals of age class k' and in location l. If m is the probability of migrating from location l to any other location l' during the course of an infection of length $1/v$, then

$$L \;=\; (1 - m) \times (1/v) \;+\; m \times (1/2v)$$

is the duration of time spent infectious in location l. Then,

$$P_T \times L \times \beta_{k'} \times d_l \times z_{kk'} \times N_{kl}$$

is the number of infections generated at location l in age class k', by an infectious individual in location l of age class k. N_{kl} is the number of individuals in age class k at location l, and P_T is a proportionality constant that depends on the unknown probability of transmission given contact. A similar calculation can be done

when the two locations are not equal, and a complete next-generation matrix can be filled in. This assumes only one migration event during a course of infection, but this is a fair approximation since daily migration probabilities are small and infections are short. Note that some other parts of this computation are necessarily approximate since part of the model occurs in continuous time with fractional individuals, while another part (migration, hospitalization) occurs at discrete time points and concerns whole individuals.

The R_0 value described in the figures and text is for a hypothetical Ho Chi Minh City with no cross-province emigration, and model results are presented for a single case introduced in Ho Chi Minh City. The value $R_{0,ng}$ calculated for the model via the next-generation matrix (as above) will usually be 5% to 15% lower than this $R_{0,HCMC}$ computed for Ho Chi Minh City with no migration. $R_{0,HCMC}$ is presented in the paper since (1) we do not know the true rate of migration, (2) sick individuals may not migrate in reality, and (3) true migration probably takes place under a "residency model" (where individuals reside in one place, migrate, then return) rather than under a gravity model.

It is known that transmissibility varies from person to person and contact rates vary from community to community; thus, choosing a single R_0 value for an entire population can prove difficult. As a reference point, we use the R_0 value for a hypothetical Ho Chi Minh City with spatially uniform population density, spatially uniform interpersonal contact behavior, and no emigration. The R_0 values for the other provinces are smaller by some amount, depending on each province's population density and age structure.

Sensitivity analysis

The varied parameters with ranges were: $1.2 < R_0 < 3.1$; 3.8 days < duration of infection < 5.5 days; 1.35% < daily probability of migration < 5.00%; 0.5 < traffic on small road relative to large road < 1.0; 0.5% < probability of hospitalization < 1.5%. 1000 parameter sets were drawn randomly from this range using Latin hypercube sampling and medians and quartile ranges are presented from these 1000 runs [11].

Daily probability of migration was derived from the Ha Nam survey – which gave a mean estimate of 1.35% of people moving province each day; this was used as the lower end of the modeled range.

References

1. Del Valle SY, Hyman JM, Hethcote HW, Eubank SG: Mixing patterns between age groups in social networks. *Social Networks* 2007, 29:539-554.
2. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J *et al*: Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 2008, 5(3):e74.
3. Wallinga J, Teunis P, Kretzschmar M: Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am J Epidemiol* 2006, 164(10):936-944.
4. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, Griffin J, Baggaley RF, Jenkins HE, Lyons EJ *et al*: Pandemic Potential of a Strain of Influenza A (H1N1) : Early Findings. *Science* 2009.
5. Lessler J, Reich NG, Brookmeyer R, Perl TM, Nelson KE, Cummings DA: Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect Dis* 2009, 9(5):291-300.
6. Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boelle PY: A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat Med* 2004, 23(22):3469-3487.
7. Boelle PY, Bernillon P, Desenclos JC: A preliminary estimation of the reproduction ratio for new influenza A(H1N1) from the outbreak in Mexico, March-April 2009. *Euro Surveill* 2009, 14(19).
8. Nishiura H, Castillo-Chavez C, Safan M, Chowell G: Transmission potential of the new influenza A(H1N1) virus and its age-specificity in Japan. *Euro Surveill* 2009, 14(22).
9. Update: novel influenza A (H1N1) virus infections - worldwide, May 6, 2009. *MMWR Morb Mortal Wkly Rep* 2009, 58(17):453-458.
10. Serum cross-reactive antibody response to a novel influenza A (H1N1) virus after vaccination with seasonal influenza vaccine. *MMWR Morb Mortal Wkly Rep* 2009, 58(19):521-524.
11. Blower SM, Dowlatabadi H: Sensitivity and Uncertainty Analysis of Complex Models of Disease Transmission: An HIV Model, as an Example. *International Statistical Review* 1994, 62(2):229-243.

Chapter 7

Discussion *(cohort study)*

7.1 Contribution to knowledge of influenza epidemiology

The three presented papers all arose from the cohort study and provide new insights into the epidemiology of influenza in Vietnam that have direct relevance to influenza control activities in Vietnam. The papers themselves describe these insights in detail but in summary I have shown that influenza infection is as common in Vietnam as in temperate regions, and that although school-age children have the highest infection rates, there are differences in age specific attack rates, risk factors, and contact patterns compared to temperate regions which may attenuate the impact of control measures targeted at school-age children and schools. Specifically, there appears to be less marked concentration of infection risk in school age children, and hypotheses for this pattern are put forward in research paper 1. The data on influenza infection rates in the community when combined with the contact survey data presented in research paper 2, provide a resource for developing evidence-based influenza control strategies in Vietnam. Papers 2 and 3 represent such exercises, with paper 2 modelling the potential effect of different immunisation strategies, whilst paper 3 models the spatial and temporal transmission of influenza and the predicted impact of social distancing measures and internal flight restrictions. The work was conducted in partnership with NIHE, which has a key role in developing Vietnam's influenza control policies. The Director of NIHE is the Vice-Chairman of the National Committee for the Prevention and Control of Human, Avian, and Pandemic Influenza, and as such the work has directly fed-into policy-making forums in Vietnam.

The results of this research also contribute to an improved understanding of the epidemiological processes driving global influenza patterns. Several authors have highlighted the need for better data from Southeast Asia in order to test the source-sink hypothesis, which postulates that East and Southeast Asia acts as a global source of new influenza virus variants (Nelson et al., 2007; Rambaut et al., 2008; Russell et al., 2008). Bahl et al. (2011) have recently reported lower levels of relative genetic diversity of H3 HA in Southeast Asia compared to temperate regions, which would seem to run counter to the hypothesis that Southeast Asia is a reservoir of new viruses, since a source habitat is expected to maintain higher levels of genetic diversity as a consequence of a larger population size (Bedford et al., 2011). Low levels of diversity in Southeast Asia despite the probable absence of

strong seasonal genetic bottle-necks[1] suggests either that influenza transmission is low or that there is strong selection pressure (Bedford et al., 2011). Bahl et al. (2011) made an assumption that the strength of natural selection does not vary between populations, leading to the conclusion that the observed patterns suggest "influenza transmission in Hong Kong and Southeast Asia was not as extensive as that observed in temperate regions". However, we observed substantial rates of infection in the Ha Nam cohort, and an alternative model is that influenza transmission is as extensive in Southeast Asia as in temperate regions but the pattern of transmission and resulting selection dynamics differ. The lack of synchronisation of epidemics, the broader period of virus circulation, the less intense seasonal forcing, and the occurrence of multiple epidemics per year could theoretically result in widespread infection that results in a strong and constant immune-mediated selection pressure that constrains genetic diversity to a relatively small number of 'fit' virus lineages (Bedford et al., 2011). Work in Brazil based on influenza-associated mortality patterns has identified a low reproduction number[2] for influenza mortality (mean reproduction number 1.03, 95% confidence interval: 1.02, 1.04), which the authors hypothesise may be the result of high levels of population immunity (Chowell et al., 2010). This would be consistent with a model of widespread influenza circulation resulting in high levels of partially protective immunity. The findings of Bahl et al. (2011) suggest that even if southeast Asia is a reservoir of influenza viruses, temperate regions are not a 'black-hole sink' from which new viruses never emerge, but are a 'reciprocal sink' that transfers viruses back to the source (Sokurenko et al., 2006). As such it may be that influenza evolution is occurring both in temperate and tropical regions but is qualitatively different due to differences in the pattern of selection pressures. Further work is clearly needed to fully understand the drivers and dynamics of global influenza evolution, including comparative studies of immune-mediated selection pressures in temperate and tropical areas.

Some limitations of the work have been presented in the discussion section of the papers but there are two key changes I would make if I were to conduct the cohort study again. First, I would request permission to take capillary blood samples from children under the age of five years in order to provide infection incidence estimates across the entire age range. This would also provide data on the rate of loss of maternal antibodies and acquisition of new antibodies in the first years of life (a measure of the force of infection) and data on which to estimate the potential impact of immunisation of pregnant women on preventing infection in infants (Ortiz et al., 2011). Second, I would stagger the serological surveys throughout the year, rather than conduct periodic cross-sectional surveys, in order to be better able to examine temporal patterns in infection risk and to look for evidence of year-round influenza transmission. For the contact diary study, I would ask participants to complete the diary prospectively rather than retrospectively in order to reduce recall bias, and I would collect data on the exact age of contacts rather than the age range,

[1] A collapse in virus population size followed by the survival of a restricted number of lineages
[2] The average number of secondary cases generated by a primary case

which restricted the analyses we were able to conduct. We have subsequently conducted a new study of contact patterns in 1000 rural and 1000 urban residents in Vietnam using prospective diary completion.

7.2 Further research directions

The research papers included in this thesis represent only part of the output of the cohort study. Ongoing analysis is examining the relationship of prior HI titres and influenza infection experience on the risk of subsequent influenza infection and illness. Peripheral blood mononuclear cells were collected in the first and second bleeds and have been used to identify T-cell responses to peptides of influenza H5, H3, and H1 hemagglutinin, N1 and N2 neuraminidase, and the internal proteins of H3N2 in an interferon-γ enzyme-linked immunospot T assay in collaboration with the MRC Human Immunology Unit at the University of Oxford. The data on T-cell responses to H5 HA peptides have already been published (Powell et al., 2012) (Annex A) and work is ongoing to explore the relationship of pre-existing T-cell responses to influenza peptides to the risk of developing influenza illness. A large number of respiratory swab samples were negative for influenza, and testing for other viral respiratory pathogens is planned. The influenza PCR positive samples will also be used to study intra-host and intra-community pathogen genetic diversity.

Collaboration is ongoing with the MRC Centre for Outbreak Analysis and Modelling at Imperial College to use advanced statistical techniques to partition the observed influenza infection risk into the probability of acquiring infection in the home or in the community, and to examine heterogeneity in infectiousness (super-spreaders) (Cauchemez et al., 2009). Work is also ongoing to integrate the social contact data with the data on influenza outcomes to assess the risk of influenza in different age groups after controlling for both contact behaviours and prior infection history (Goeyvaerts et al., 2010). This will provide the first individual level (as opposed to ecological[1]) test of the 'social contact hypothesis', which argues that for pathogens spread by the respiratory route, the age distribution of conversational contacts is directly proportional to the age distribution of infection risk (Wallinga et al., 2006). We will also directly test the hypothesis that intimate physical contact, rather than casual face-to-face contact, best explains the transmission patterns of diseases such as influenza (Melegaro et al., 2011).

[1] An 'ecological' analysis looks at exposure-disease associations at the population level only, without linking exposures to outcomes at the individual level

Chapter 8

Background to host genetics study

8.1 Host genetics and infectious diseases

It has long been recognised, going back at least 60 years to the malaria-sickle haemoglobin gene hypothesis, that the co-evolutionary struggle with microbes has shaped the genetic architecture of humans (Piel et al., 2010). Indeed, immune response genes are amongst the most common human genes and, as exemplified by the human leukocyte antigen and KIR[1] loci, are also the most variable because of their need to recognise the shifting diversity of pathogen antigens (Blackwell et al., 2009; Middleton and Gonzelez, 2010).

Studies of adoptees have demonstrated a strong inherited component to the risk of death from a range of infections (Sorensen et al., 1988) and the emergence of affordable genome-wide sequencing of up to one million single nucleotide polymorphisms (SNPs) has resulted in an explosion of hypothesis-free genetic association studies that have identified hundreds of gene-disease associations, although many associations have not been replicated (Ioannidis, 2007; Manolio et al., 2008). Where robust associations with common polymorphisms (minor allele frequency >1%) have been identified, they usually have a small effect size and individually explain relatively little of the heritability (Cirulli and Goldstein, 2010). However, genetic polymorphisms have been identified that have an important health benefit, mostly in the area of pharmacogenomics (Hudson, 2011), and which have provided important insights into disease pathogenesis, such as the role of *CCR5* in HIV (Manolio et al., 2008; Vannberg et al., 2011). Meanwhile the role of rare variants (minor allele frequency ≤1%) is less explored because current SNP arrays cover only common polymorphisms, but this is likely to change with the advent of affordable exome[2], rather than SNP, sequencing (Cirulli and Goldstein, 2010).

8.2 Host genetics and influenza

Until recently there was almost no consideration of the role of host genetics in susceptibility to influenza but the re-emergence and global spread of highly pathogenic H5N1 in 2004 sparked interest in the host determinants of severe influenza. H5N1 is one of the

[1] Killer-cell Immunoglobulin-like Receptor

[2] The part of the genome that is translated into proteins

most lethal infections of humans, with mortality approaching 60% in clinically detected cases, and the determinants of this high virulence are of enormous scientific and public health interest given the demonstrated ability of avian derived influenza A virus gene segments to successfully cross the species barrier. The scientific literature and other sources of information on genetic susceptibility to influenza in general and H5N1 in particular are summarised in research papers four and five and will not be repeated here.

The hypothesis that host genetic factors play an important role in susceptibility to H5N1 infection arose from my own experiences of investigating the outbreak of human H5N1 cases in Vietnam. In December 2003 I was the medical officer in charge of communicable disease outbreaks at the WHO office in Vietnam. I was invited by the Director of the National Paediatric Hospital to investigate a series of unexplained deaths in children from a severe respiratory infection. The initial fear was that these cases may represent a recurrence of SARS, which had affected Hanoi earlier the same year. The clinical and epidemiological picture was not suggestive of SARS and I arranged for respiratory specimens to be sent to Hong Kong for testing for human and avian influenza viruses. The samples were positive for HPAI H5N1 and represented the first identified human cases in the epizootic of HPAI H5N1 that began in 2003 (WHO, 2005).

This very first series of cases included two family clusters. One cluster comprised a mother and daughter, who both died of fulminant H5N1 disease, and the virus cultured from the daughter (A/Vietnam/1194/04) has subsequently been widely used as a prototypical human clade 1 HPAI H5N1 virus and from which a clade 1 candidate vaccine virus has been developed. Also, this mother and daughter were residents of the Commune that was selected for the cohort study described in Chapters 3-7, hence the exploration of H5N1 antibodies and T-cell responses in the cohort. The second cluster comprised two young siblings who both died of a respiratory illness, one of whom was a confirmed H5N1 case. Subsequently I encountered numerous family clusters of H5N1 infection; far in excess of what I had observed with any other infectious disease. The highly unusual epidemiology led me to hypothesise that there is strong familial susceptibility to H5N1 infection or disease. The observations underlying this hypothesis are summarised in research paper 4.

8.3 Objective of host genetic study

Despite the considerable challenges of undertaking a study of host genetic susceptibility to such a rare disease with such a high case fatality ratio, if susceptibility were conferred by a rare, highly penetrant, single locus genotype that greatly increased the risk of infection or disease, this may be detectable even with a small sample size. The literature on genetic susceptibility to influenza is systematically reviewed in research paper 5, and whilst a large number of candidate genes have been proposed, there are no genetic studies in humans that implicate any of these putative loci. Therefore a decision was taken to

undertake an hypothesis free genome-wide approach. The objective of the host genetic study was to determine if any SNPs are associated with H5N1 disease using a genome-wide association (GWA) study design. Besides its direct relevance for H5N1, the study can also be conceptualised as an example of an extreme-trait sequencing design, where a subset of individuals with an extreme phenotype (susceptibility to avian influenza) may provide insights into genetic factors acting in the wider population but which are difficult to detect as the phenotype (severe inter-pandemic influenza) is less distinct and causality is more multifactorial (Cirulli and Goldstein, 2010).

8.4 Candidate's role

I conceived the study, secured funding (Wellcome Trust UK grants 081613/Z/06/Z and 077078/Z/05/Z, and the South East Asia Infectious Disease Clinical Research Network), wrote the protocol and case record forms, and prepared all the paperwork for ethical approval in the U.K., Vietnam and Thailand. I was the Principal Investigator of the study and supervised all aspects of study implementation in Vietnam and Thailand. DNA extraction was conducted by laboratory personnel in Vietnam and Thailand, and was shipped to the Genome Institute of Singapore for genotyping. Statistical analysis was undertaken by Dr Chiea C. Khor of the Genome Institute of Singapore.

Chapter 9

Contribution of research papers *(genetics study)*

In this chapter I will briefly introduce the research papers that are compiled in this thesis on the host genetic aspects and describe how they demonstrate my contribution to knowledge in the field. Three papers are included and I am the first and corresponding author on all three.

9.1 Outline of research papers

9.1.1 Research paper 4

This manuscript was published as Horby et al. (2010) and summarises and interprets the epidemiological data that led to the formulation of the hypothesis addressed in research paper 6, that host genetic factors have a strong influence on susceptibility to H5N1 infection. Other authors had speculated on the possibility of important host genetic factors influencing susceptibility to H5N1 based on the unusual level of family clustering but none had formally assembled the evidence. The manuscript also contains a re-analysis I conducted of the expected level of clustering (using data from the cohort study on household sizes), which was first presented by Pitzer et al. (2007) as evidence that the observed level of clustering was not suggestive of genetic susceptibility. The R-code I wrote for the reanalysis presented in figure 2 of research paper 4 is reproduced in Appendix B.

9.1.2 Research paper 5

Research paper 5 is in press as Horby, Nhu, Dunstan, and Baillie (Horby et al.) and is a systematic literature review conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2009), which summarises the evidence on the role of host genetics in susceptibility to influenza. This extends the work of research paper 4 by including all types of influenza affecting humans and systematically searching and summarising the published literature.

9.1.3 Research paper 6

Research paper 6 presents the results of the genome-wide case-control association study conducted in Vietnam and Thailand. Although a brief paper, it represents a large amount of work undertaken over several years to assemble DNA samples from H5N1 cases. This paper is only the third published study of human genetics and influenza, the other studies having looked at ≈ 50000 SNPs in severe H1N1/2009 associated pneumonia (Zuniga et al., 2011) and eight polymorphisms in the tumor necrosis factor and mannose-binding lectin genes in fatal influenza cases (Ferdinands et al., 2011). Research paper 6 presents the first direct evidence of genetic loci associated with susceptibility to H5N1 infection.

Chapter 10

Research paper 4

Title: What is the evidence of a role for host genetics in susceptibility to influenza A/H5N1?

Author(s): P. Horby, H. Sudoyo, V. Viprakasit, A. Fox, P. Q. Thai, H. YU, S. Davial, M. Hibberd, S. J. Dunstan, Y. Monteerarat, J. J. Farrar, S. Marzuki, N. T. Hien.

Journal/Publisher: Epidemiology and Infection, Cambridge University Press

Type of publication: Research article

Stage of publication: Published

Academic peer-reviewed: Yes.

Copyright: The online version of this article is published within an Open Access environment under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Candidate's role: I conceived the work and compiled all the data included in the manuscript. I wrote and implemented the R script for figure 2 that simulates the proportion of cases that would be expected to be clustered under various assumptions of household size and the probability of infection if exposed to H5N1 (annex 2). I wrote the first and all subsequent drafts of the manuscript, submitted the manuscript for publication, and responded to all reviewers comments.

Candidate's signature:

Supervisor or senior author's signature to confirm Candidates role:

Epidemiol. Infect. (2010), **138**, 1550–1558.
© Cambridge University Press 2010. The online version of this article is published within an Open Access environment subject to the conditions of the Creative Commons Attribution-NonCommercial-ShareAlike licence <http://creativecommons.org/licenses/by-nc-sa/2.5/>. The written permission of Cambridge University Press must be obtained for commercial re-use.
doi:10.1017/S0950268810000518

What is the evidence of a role for host genetics in susceptibility to influenza A/H5N1?

P. HORBY[1,2]*, H. SUDOYO[3], V. VIPRAKASIT[4], A. FOX[1,2], P. Q. THAI[5], H. YU[6], S. DAVILA[7], M. HIBBERD[7], S. J. DUNSTAN[1,2], Y. MONTEERARAT[4], J. J. FARRAR[1,2], S. MARZUKI[3] AND N. T. HIEN[5]

[1] *Oxford University Clinical Research Unit, Vietnam*
[2] *Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, UK*
[3] *Eijkman Institute for Molecular Biology, Jakarta, Indonesia*
[4] *Department of Pediatrics, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok*
[5] *National Institute for Hygiene and Epidemiology, Hanoi, Vietnam*
[6] *Chinese Centre for Disease Control and Prevention, Beijing, People's Republic of China*
[7] *Genome Institute of Singapore, Genome Building, Singapore*

(*Accepted 10 February 2010; first published online 18 March 2010*)

SUMMARY

The apparent family clustering of avian influenza A/H5N1 has led several groups to postulate the existence of a host genetic influence on susceptibility to A/H5N1, yet the role of host factors on the risk of A/H5N1 disease has received remarkably little attention compared to the efforts focused on viral factors. We examined the epidemiological patterns of human A/H5N1 cases, their possible explanations, and the plausibility of a host genetic effect on susceptibility to A/H5N1 infection. The preponderance of familial clustering of cases and the relative lack of non-familial clusters, the occurrence of related cases separated by time and place, and the paucity of cases in some highly exposed groups such as poultry cullers, are consistent with a host genetic effect. Animal models support the biological plausibility of genetic susceptibility to A/H5N1. Although the evidence is circumstantial, host genetic factors are a parsimonious explanation for the unusual epidemiology of human A/H5N1 cases and warrant further investigation.

Key words: Influenza A virus, genetic susceptibility, H5N1 subtype.

INTRODUCTION

Human cases of influenza A subtype H5N1 appear to cluster in families, a pattern which has led several authors to comment that host genetics may play an important role in susceptibility to A/H5N1 infection or disease [1–5]. This has potentially far-reaching implications, since the identification and subsequent characterization of genetic factors that have a strong influence on susceptibility to A/H5N1 disease would highlight key virus–host interactions necessary or contributory to infection or disease. Elucidating these key interactions has the potential to catalyse advances in areas such as the prediction of viral pathogenicity and the development of new or improved preventive and therapeutic interventions, which may be of

* Author for correspondence: Dr P. Horby, Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, Oxford OX3 7LJ, UK.
(Email: peter.horby@gmail.com)

Table 1. *Number of confirmed H5N1 cases and clusters by country*

Country	Total laboratory-confirmed cases*	No. of clusters†	*n*/*N* (%) of confirmed cases occurring in clusters
Azerbaijan	8	2	6/8 (75)
Bangladesh	1	0	0/1 (0)
Cambodia	8	1	1/8 (12)
China, mainland	38	4	4/38 (10)
Djibouti	1	0	0/1 (0)
Egypt	88	4	9/88 (10)
China, Hong Kong	20	2	4/20 (20)
Indonesia	141	18	36/141 (25)
Iraq	3	1	2/3 (67)
Laos PDR	2	0	0/2 (0)
Myanmar	1	0	0/1 (0)
Nigeria	1	1	1/1 (100)
Pakistan	3	1	3/3 (100)
Thailand	25	3	5/25 (20)
Turkey	12	2	6/12 (50)
Vietnam	111	13	26/111 (23)
Total (all countries)	463	52	103/463 (22)

* As of 25 November 2009.
† A cluster is defined as at least two cases of clinically compatible illness with at least one case with laboratory-confirmed H5N1.

relevance not only to zoonotic influenza but also to seasonal and pandemic influenza.

Since it re-emerged in 2003 A/H5N1 has received enormous attention, including the allocation of substantial financial resources for vaccine development and pandemic preparedness. Yet the reasons for its scarcity in humans, its poor ability to transmit between people, the clustering of cases and the risk factors for infection remain elusive; as does our ability to predict the likelihood that A/H5N1 may become a pandemic virus. Most research has focused on the viruses; through genotypic and phenotypic analysis and animal experiments using modified viruses, but the other half of the equation, the host, has been relatively neglected. Since it is epidemiological patterns that have stimulated consideration of host genetic factors, an important first step is to review whether the epidemiological patterns are consistent with a host genetic influence. Currently only two publications have explicitly examined the potential role of host genetics and human A/H5N1 infection. Pitzer *et al.* have looked at whether the observed clustering could be explained by chance alone [6]. Trammell & Toth have reviewed possible biological mechanisms of host susceptibility to influenza, using mostly data from murine models [7]. We examine the epidemiological patterns of human A/H5N1 cases, their possible explanations, and review the evidence for a role for host genetics in susceptibility to influenza A/H5N1.

THE CASE IN FAVOUR OF A ROLE FOR HOST GENETICS

Familial aggregation of cases

Between 1 January 1997 and 25 November 2009 a total of 36 clusters of two or more laboratory-confirmed cases of A/H5N1 have been reported, with at least an additional 16 clusters of one confirmed case plus at least one probable case [3, 4, 8–11] (Table 1). These 52 clusters account for 22% (103/463) of all laboratory-confirmed cases and only six of the 103 cases occurring in clusters did not have a genetic relationship to another case in the cluster. Although there is no data on the familial aggregation of other zoonosis for comparison, this degree of family clustering has surprised many people, especially since A/H5N1 is considered to only rarely transmit from person to person. Since familial aggregation is a hallmark of genetically determined diseases, genetic susceptibility to A/H5N1 infection is one hypothesis

that might explain the familial aggregation. Unfortunately the apparent increased risk in relatives of affected cases compared to background risk has not been quantified and the large cluster in Karo, Indonesia was a missed opportunity to estimate the familial relative risk by comparing the risk in related and unrelated contacts of infected individuals. However, what we do know is that this cluster involved eight cases (seven laboratory-confirmed) in a single extended family residing in four households [12]. Nine family members slept in the same room as the primary case while the case was symptomatic and three of these nine (33%) developed A/H5N1 infection [13]. It is perhaps surprising that there were no unrelated cases despite multiple opportunities for infection of non-related contacts, including unprotected healthcare workers, and onset dates that stretched over a period of 3 weeks [14].

The relative absence of non-familial aggregation of cases

If all members of a community affected by A/H5N1 outbreaks in poultry are at equal risk then it would be more likely to observe pairs of cases of unrelated community members than to see household clusters [6]. Yet of the 103 confirmed cases occurring in 52 clusters, only six cases occurring in four clusters were not genetically related to any other case in the cluster [one husband and wife pair (Vietnam 2005); one healthcare worker (Vietnam 2005); one neighbour (Azerbaijan 2006); two children (Egypt 2009)] [11]. This pattern is important since it suggests either large differences in risk between families within affected communities, or large biases in the detection and reporting of family-based clusters compared to unrelated case clusters.

Related but unassociated cases

At least two incidents have occurred where genetically related individuals developed confirmed or probable A/H5N1 disease independently of one another.

In August 2004 a 25-year-old women from Hau Giang Province, southern Vietnam died from laboratory-confirmed A/H5N1. Both the 19-year-old brother of this case and their 23-year-old cousin died of severe pneumonia within a week of the confirmed case; specimens from these two cases were not tested for A/H5N1. The brother lived with the confirmed case but the cousin lived in a non-adjacent commune and investigations revealed that there had been no contact whatsoever between the cousin (and her immediate family) and the siblings (and their immediate family) in the week prior to the earliest onset of illness and the deaths. Local authorities concluded that there was no likelihood of a common point source of infection or of any other means of transmission of A/H5N1 between the cousin and the sibling cases. Therefore, the disease in the cousin seems to have occurred independently from the sibling cases.

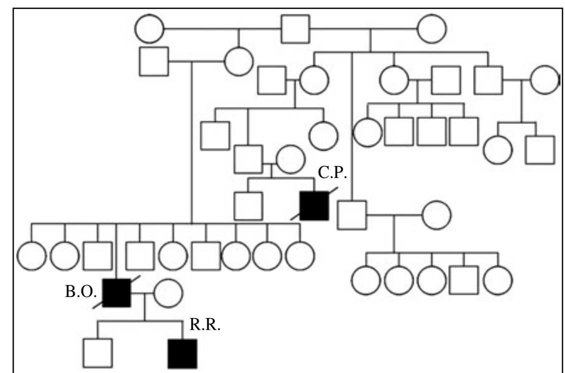


Fig. 1. Family pedigree showing three H5N1 affected individuals, with infections separated by 2 years.

In Thailand three related individuals suffered A/H5N1 infection during two different waves of the outbreak. The first case was a boy (C.P.) who died during the first wave of outbreaks in late 2003–2004 [15]. The mother of case C.P. also died of a respiratory illness at the same time as her son, but samples were not available for testing for A/H5N1 [8]. The other two cases, a father and son (B.O and R.R.), were infected in the 2005 outbreak [16]. Their family pedigree is shown in Figure 1. C.P. lived in the same province but a different district to B.O. and R.R.

Given the scarcity of A/H5N1 disease, these incidents of related but apparently unconnected cases seem an improbable misfortune, unless relatives have an increased risk of A/H5N1 infection compared to the general population.

Exposure and risk are not well correlated

Although data from three case-control studies show that contact with dead or dying poultry is a significant risk factor for A/H5N1 infection, the proportion of cases that can be attributed to this factor is not high [17–19]. About 25% of all confirmed clinical cases of H5N1 infection cannot recall any recent poultry exposure before illness onset and in many other cases

the reported exposure to infected poultry is tenuous [17, 20–22]. The largest case-control study published so far found that only 28% of all cases could be attributed to preparing or cooking sick poultry [17]. The same study found no differences between affected and unaffected households in other poultry-handling practices, hygiene behaviour or in other putative risk factors such as the use of poultry fertilizer. This absence of obvious risky practices in many affected individuals and families juxtaposes starkly with the almost complete absence of clinical cases in groups who are known to have engaged in theoretically very high-risk behaviours, i.e. culling infected poultry flocks without personal protective equipment.

From 2003 to 31 January 2010, 49 countries have reported over 6660 outbreaks of highly pathogenic avian influenza A/H5N1 in domestic poultry or wildlife to the World Organization for Animal Health and several hundred million poultry have died or been culled. These figures are a minimum, since only a proportion of all outbreaks are detected and reported. The number of people exposed to A/H5N1 as a result of reported and unreported outbreaks is not known but we do know that exposure to poultry is very common in many of the worst affected countries. One population-based study of more than 45 000 people in an A/H5N1-affected community in Vietnam found that 25·9% (11 755) lived in households where poultry were sick or had died [23]. A community survey in Cambodia of 155 poultry-raising households in an A/H5N1-affected area identified poultry deaths in 102 households (66%), and 42 households (27%) were considered likely to have experienced an outbreak of A/H5N1 [19]. A larger survey in Cambodia estimated that most of the rural population has frequent contact with poultry and 52% regularly have a potentially high-risk exposure [1]. Therefore it likely that very large numbers of people, possibly millions, have been exposed to A/H5N1 since 2003 yet only 471 human cases have been reported globally over the same period. It can be safely assumed that these numbers, like poultry outbreaks, are a minimum as the clinical presentation is non-specific and few sites possess the capabilities to diagnose A/H5N1. Although a survey in two affected villages in Cambodia found serological evidence of subclinical A/H5N1 infection in seven (1%) out of 674 subjects [24], evidence from active surveillance and serological surveys of populations known to be exposed to A/H5N1 generally indicates that large numbers of cases are not being missed [19, 21, 25–33]. While the sensitivity and reproducibility of serological assays for A/H5N1 infection is variable, many serological studies have used the gold standard of microneutralization assay with Western blot confirmation and therefore provide the best estimate currently available of A/H5N1 infection prevalence [34, 35]. The apparent low incidence of infection following exposure to sick poultry and the low risk in some intensely exposed groups indicates a substantial species barrier, but a barrier that seems to be much weaker in a small number of individuals and families [36].

Person-to-person transmission

Families live together in intimate contact and person-to-person transmission has been convincingly put forward as an explanation for two family clusters [37, 38] and an additional five reports have stated that it could not be ruled out in at least seven families [3, 4, 39–41]. The evidence for person-to-person transmission outside of the family is mixed. In the investigation of the 1997 Hong Kong cases, seropositive healthcare workers were identified, but none have been found in subsequent studies [28, 31, 42] and, as previously mentioned, non-familial clusters are rare. Person-to-person transmission of A/H5N1 clearly can occur but what is perhaps most interesting is the presence of limited intra-familial person-to-person transmission risk but its possible absence in other settings. This could be explained by the special intimacy of familial relationships but alternatively it could be an indicator of a genetic influence on risk; i.e. family members are at increased risk of person-to-person transmission because of a shared genetic susceptibility to infection from any source.

Biological plausibility

Certain animal species are more susceptible to H5N1 than other species and possible factors determining the host-range restriction of avian influenza viruses have been reviewed elsewhere [36, 43, 44]. However, within-species differences also exist and in-bred mice strains exhibit large differences in their susceptibility to influenza infection [7, 45–47]. Indeed, differences in susceptibility of mouse strains to influenza infection, followed by mapping of the mouse disease loci and identification of the region on the human chromosome has led to the identification of the human *Mx* genes involved in response to viral infections [48–50].

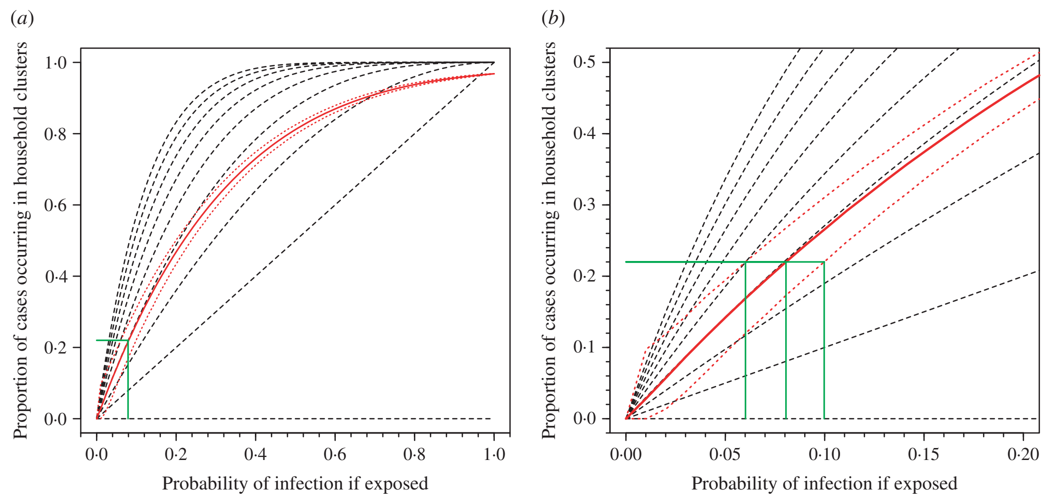


Fig. 2. Proportion of cases occurring in household clusters by probability of infection for different household sizes. (*a*) All data. (*b*) Enlargement of left-hand corner of panel (*a*). The broken black lines represent the modelled data for household sizes ranging from nine persons (top) to one person (bottom). The solid red line is the median estimate of the modelled data applying the observed range of household sizes in a Vietnamese cohort. Red dotted lines represent 95% prediction intervals for 10 000 simulations. The solid green lines show the probability of infection compatible with the observed clustering of about 0·22 for median estimate and 95% prediction intervals of model.

Trammell & Toth have reviewed possible biological mechanisms of host susceptibility to influenza in the mouse model and similar host genetic influences on susceptibility to infection or disease might also exist in humans [7]. In fact a recent study demonstrated genetic susceptibility to A/H5N1 in mice and called for studies of genetic susceptibility in humans [47]. However, to date no human genetic studies of susceptibility to influenza have been conducted, other than two genealogical studies, one of which identified a heritable predisposition to death from influenza [51, 52].

Candidate host genes that may contribute to severe influenza illness can be proposed *a priori* from known virus–host interactions critical to infection, replication and pathogenesis [53]. Alternatively, gene-expression profiling using microarrays may provide insights into genes associated with severe disease [54]. The role of cell surface sialic acid receptors in the determination of host specificity of influenza viruses is well documented and therefore the genes encoding these receptors and their associated glycan modifications are potential candidates [55–58]. Cytokine dysregulation has been shown to be a feature of A/H5N1 infection in clinical and animal studies [59–61] and various aspects of innate immunity including collectin-like mannose-binding lectin, toll-like receptors (TLR 3, 7, 8), cytokines, chemokines, and interferon-inducible proteins such as MxA are also plausible candidates [62–68]. Interestingly, susceptibility to other viral respiratory pathogens has been traced to genes of the innate immune system [69–73].

THE CASE AGAINST

Chance

Pitzer *et al.* [6] have argued that the observed pattern of clustering of A/H5N1 cases can be explained by chance and does not provide evidence for a genetic effect. However, the application of similar methods using real data on household structures in Vietnam show that observing 22% of cases occurring in household clusters is consistent with a risk of infection following exposure of around 8% (95% prediction interval 6–10%) (Fig. 2). If the true risk of infection following exposure were 8%, the current 500 cases would be the result of exposure of only 6250 people globally over the past 5 years. This is an implausibly low number. The number of people exposed is certainly orders of magnitude higher, the risk of infection following exposure substantially lower than 8% and, therefore, the observation of 22% household clustering unlikely to be a probable outcome unless other factors are in play. Even if the risk of infection following exposure were 8%, and 22% of cases occurring in household clusters were

simply a numerical result of this high risk, in any single affected community we would expect to observe around three sporadic cases for every case occurring in a household cluster (75% *vs*. 25%). As described above, we actually see few non-familial community clusters. The model of Pitzer *et al*. also only considered exposure of entire nuclear families with every family member at equal risk. This negates the influence of single or couple households, and occupational exposures outside of the home and therefore over-emphasizes the probability of familial clusters.

Bias

A key premise is that the observed clustering is a true reflection of the real situation – not simply an apparent pattern caused by biases. It is probable that multiple cases of severe pneumonia in healthy children or adults clustered in time and space are more likely than sporadic cases to be perceived as abnormal and therefore reported to the authorities. It is also true that following a first case which was severely ill or fatal, a second case in a family will rapidly seek medical attention; and indeed several of the reported clusters consist of a first fatality which is clinically suspicious of H5N1 and a second laboratory-confirmed case. So the observed level of clustering could be an artifact of differential ascertainment of clustered *vs*. sporadic cases. While this bias is bound to be operating to some extent, the question is whether it fully explains the clustering. Moreover, it might be expected that this ascertainment bias would apply similarly to community clusters of genetically unrelated cases as to family clusters.

Confounding

Familial risk does not necessarily mean genetic risk; families share their homes, food and behaviours with one another and therefore shared 'high-risk' exposures must be a strong contender to explain family clusters. Indeed, the paucity of community clusters of genetically unrelated human H5N1 cases has been suggested to be a reflection of risky behaviours which are unique to the affected families [6]. As noted above, unusual or risky practices have been identified but it has not been possible to attribute many cases to these behaviours since the behaviours are widespread in the community yet absent in many cases. It is certainly possible that behavioural factors partially or completely explain the epidemiological patterns but these have yet to be identified.

The scarcity of human cases despite widespread exposure clearly demonstrates a substantial barrier to humans acquiring infection, and much work has focused on unravelling the genetic and functional characteristics of the viruses which would explain these barriers [74]. There are clear differences between virus strains in their ability to infect and cause severe disease in animal models but the viruses isolated from human cases occurring in family clusters have not been found to be substantially different from viruses causing sporadic human cases or poultry outbreaks [75]. The family clusters have occurred in 11 different countries as a result of infections with five different clades (0, 1, 2·1, 2·2, 2·3). So while virus factors are certainly critical in limiting the transmission of influenza A/H5N1 from animals to humans, current data do not allow us to attribute family clustering to viral factors.

CONCLUSION

The routinely available data on the epidemiology of human cases of A/H5N1 show some unusual patterns to which host genetic susceptibility offers a parsimonious explanation. Of course, this does not mean it is true; but it is both epidemiologically and biologically plausible and worthy of serious investigation. The importance of host genetics in infectious diseases is increasingly being recognized and explored [76–78] and the relationship is usually a complex interaction between the pathogen, environmental influences and a range of innate and adaptive host factors. This poses difficulties for attempts to detect genetic influences on susceptibility to H5N1, since very large sample sizes are needed to detect complex or weak effects, yet the total number of cases is very small. However, a host genetics association study may potentially be informative if high-risk genotypes are present. A more powerful strategy would be a genome-wide linkage study in affected families, which could interrogate the whole genome without assuming any prior hypotheses on plausible candidate genes. Given the importance of understanding the key virus–host interactions underlying severe human influenza, a search for such genetic factors in A/H5N1 is worthwhile. However, the scarcity and widespread distribution of human case means that international collaboration is essential to conduct studies of genetic susceptibility to A/H5N1 disease.

ACKNOWLEDGEMENTS

The authors thank the following for financial support: Wellcome Trust UK; Genome Institute of Singapore, Agency for Science Technology and Research (A*STAR); National Institutes of Health, South East Asia Infectious Diseases Clinical Research Network. Thanks are also due to Neal Alexander for help with Figure 2.

DECLARATION OF INTEREST

None.

REFERENCES

1. **van Kerkhove MD, *et al.*** Frequency and patterns of contact with domestic poultry and potential risk of H5N1 transmission to humans living in rural Cambodia. *Influenza and Other Respiratory Viruses* 2008; **2**: 155–163.
2. **Nguyen TH, Farrar J, Horby P.** Person-to-person transmission of influenza A (H5N1). *Lancet* 2008; **371**: 1392–1394.
3. **Sedyaningsih ER, *et al.*** Epidemiology of cases of H5N1 virus infection in Indonesia, July 2005-June 2006. *Journal of Infectious Diseases* 2007; **196**: 522–527.
4. **Kandun IN, *et al.*** Three Indonesian clusters of H5N1 virus infection in 2005. *New England Journal of Medicine* 2006; **355**: 2186–2194.
5. **Policy document.** Pandemic influenza: science to policy. Policy Document 36/06. London: The Royal Society, 2006.
6. **Pitzer VE, *et al.*** Little evidence for genetic susceptibility to influenza A (H5N1) from family clustering data. *Emerging Infectious Diseases* 2007; **13**: 1074–1076.
7. **Trammell RA, Toth LA.** Genetic susceptibility and resistance to influenza infection and disease in humans and mice. *Expert Review of Molecular Diagnostics* 2008; **8**: 515–529.
8. **Olsen SJ, *et al.*** Family clustering of avian influenza A (H5N1). *Emerging Infectious Diseases* 2005; **11**: 1799–1801.
9. **Oner AF, *et al.*** Avian influenza A (H5N1) infection in eastern Turkey in 2006. *New England Journal of Medicine* 2006; **355**: 2179–2185.
10. **Peiris JS, *et al.*** Re-emergence of fatal human influenza A subtype H5N1 disease. *Lancet* 2004; **363**: 617–619.
11. **WHO.** Avian influenza – situation in Eygpt – update 10, 2009 (updated 8 April 2009) (http://www.who.int/csr/don/2009_04_08a/en/index.html). Accessed 25 November 2009.
12. **Yang Y, *et al.*** Detecting human-to-human transmission of avian influenza A (H5N1). *Emerging Infectious Diseases* 2007; **13**: 1348–1353.
13. **WHO.** Avian influenza – situation in Indonesia – update 16, 2006 (http://www.who.int/csr/don/2006_05_31/en/). Accessed 9 May 2009.
14. **ECDC.** Investigation of the family cluster of human H5N1 cases in North Sumatra, Indonesia, 2006.
15. **Chokephaibulkit K, *et al.*** A child with avian influenza A (H5N1) infection. *Pediatric Infectious Disease Journal* 2005; **24**: 162–166.
16. **Uiprasertkul M, *et al.*** Apoptosis and pathogenesis of avian influenza A (H5N1) virus in humans. *Emerging Infectious Diseases* 2007; **13**: 708–712.
17. **Dinh PN, *et al.*** Risk factors for human infection with avian influenza A H5N1, Vietnam, 2004. *Emerging Infectious Diseases* 2006; **12**: 1841–1847.
18. **Areechokchai D, *et al.*** Investigation of avian influenza (H5N1) outbreak in humans–Thailand, 2004. *Morbidity and Mortality Weekly Reports* 2006; **55** (Suppl. 1): 3–6.
19. **Vong S, *et al.*** Low frequency of poultry-to-human H5NI virus transmission, southern Cambodia, 2005. *Emerging Infectious Diseases* 2006; **12**: 1542–1547.
20. **Abdel-Ghafar AN, *et al.*** Update on avian influenza A (H5N1) virus infection in humans. *New England Journal of Medicine* 2008; **358**: 261–273.
21. **Beigel JH, *et al.*** Avian influenza A (H5N1) infection in humans. *New England Journal of Medicine* 2005; **353**: 1374–1385.
22. **Yu H, *et al.*** Human influenza A (H5N1) cases, urban areas of People's Republic of China, 2005–2006. *Emerging Infectious Diseases* 2007; **13**: 1061–1064.
23. **Thorson A, *et al.*** Is exposure to sick or dead poultry associated with flulike illness?: a population-based study from a rural area in Vietnam with outbreaks of highly pathogenic avian influenza. *Archives of Internal Medicine* 2006; **166**: 119–123.
24. **Vong S, *et al.*** Risk factors associated with subclinical human infection with avian influenza A (H5N1) virus – Cambodia, 2006. *Journal of Infectious Diseases* 2009; **199**: 1744–1752.
25. **Wang M, Fu CX, Zheng BJ.** Antibodies against H5 and H9 avian influenza among poultry workers in China. *New England Journal of Medicine* 2009; **360**: 2583–2584.
26. **Santhia K, *et al.*** Avian influenza A H5N1 infections in Bali Province, Indonesia: a behavioral, virological and seroepidemiological study. *Influenza and Other Respiratory Viruses* 2009; **3**: 81–89.
27. **Apisarnthanarak A, *et al.*** Atypical avian influenza (H5N1). *Emerging Infectious Diseases* 2004; **10**: 1321–1324.
28. **Schultsz C, *et al.*** Avian influenza H5N1 and healthcare workers. *Emerging Infectious Diseases* 2005; **11**: 1158–1159.
29. **Ortiz JR, *et al.*** Lack of evidence of avian-to-human transmission of avian influenza A (H5N1) virus among poultry workers, Kano, Nigeria, 2006. *Journal of Infectious Diseases* 2007; **196**: 1685–1691.
30. **Hinjoy S, *et al.*** Low frequency of infection with avian influenza virus (H5N1) among poultry farmers, Thailand, 2004. *Emerging Infectious Diseases* 2008; **14**: 499–501.

31. **Liem NT, Lim W.** Lack of H5N1 avian influenza transmission to hospital employees, Hanoi, 2004. *Emerging Infectious Diseases* 2005; **11**: 210–215.
32. **Schultsz C, *et al.*** Prevalence of Antibodies against Avian Influenza A (H5N1) Virus among Cullers and Poultry Workers in Ho Chi Minh City, 2005. *PLoS ONE* 2009; **4**(11): e7948.
33. **Dejpichai R, *et al.*** Seroprevalence of antibodies to avian influenza virus A (H5N1) among residents of villages with human cases, Thailand, 2005. *Emerging Infectious Diseases* 2009; **15**: 756–60.
34. **Rowe T, *et al.*** Detection of antibody to avian influenza A (H5N1) virus in human serum by using a combination of serologic assays. *Journal of Clinical Microbiology* 1999; **37**: 937–943.
35. **Stephenson I, *et al.*** Reproducibility of serologic assays for influenza virus A (H5N1). *Emerging Infectious Diseases* 2009; **15**: 1252–1259.
36. **Kuiken T, *et al.*** Host species barriers to influenza virus infections. *Science* 2006; **312**: 394–397.
37. **Ungchusak K, *et al.*** Probable person-to-person transmission of avian influenza A (H5N1). *New England Journal of Medicine* 2005; **352**: 333–340.
38. **Wang H, *et al.*** Probable limited person-to-person transmission of highly pathogenic avian influenza A (H5N1) virus in China. *Lancet* 2008; **371**: 1427–1434.
39. **Katz JM, *et al.*** Antibody response in individuals infected with avian influenza A (H5N1) viruses and detection of anti-H5 antibody among household and social contacts. *Journal of Infectious Diseases* 1999; **180**: 1763–1770.
40. **Tran TH, *et al.*** Avian influenza A (H5N1) in 10 patients in Vietnam. *New England Journal of Medicine* 2004; **350**: 1179–1188.
41. **Gilsdorf A, *et al.*** Two clusters of human infection with influenza A/H5N1 virus in the Republic of Azerbaijan, February–March 2006. *Eurosurveillance* 2006; **11**(5): 122–126.
42. **Apisarnthanarak A, *et al.*** Seroprevalence of anti-H5 antibody among Thai health care workers after exposure to avian influenza (H5N1) in a tertiary care center. *Clinical Infectious Diseases* 2005; **40**: e16–18.
43. **Baigent SJ, McCauley JW.** Influenza type A in humans, mammals and birds: determinants of virus virulence, host-range and interspecies transmission. *BioEssays* 2003; **25**: 657–671.
44. **Klempner MS, Shapiro DS.** Crossing the species barrier – one small step to man, one giant leap to mankind. *New England Journal of Medicine* 2004; **350**: 1171–1172.
45. **Srivastava B, *et al.*** Host genetic background strongly influences the response to influenza a virus infections. *PLoS ONE* 2009; **4**: e4857.
46. **Salomon R, *et al.*** Mx1 gene protects mice against the highly lethal human H5N1 influenza virus. *Cell Cycle* 2007; **6**: 2417–2421.
47. **Boon AC, *et al.*** Host genetic variation affects resistance to infection with a highly pathogenic H5N1 influenza A virus in mice. *Journal of Virology* 2009; **83**: 10417–10426.
48. **Staeheli P, *et al.*** Influenza virus-susceptible mice carry Mx genes with a large deletion or a nonsense mutation. *Molecular and Cellular Biology* 1988; **8**: 4518–4523.
49. **Haller O, Stertz S, Kochs G.** The Mx GTPase family of interferon-induced antiviral proteins. *Microbes and Infection* 2007; **9**: 1636–1643.
50. **Haller O, *et al.*** Genetically determined, interferon-dependent resistance to influenza virus in mice. *Journal of Experimental Medicine* 1979; **149**: 601–612.
51. **Albright FS, *et al.*** Evidence for a heritable predisposition to death due to influenza. *Journal of Infectious Diseases* 2008; **197**: 18–24.
52. **Gottfredsson M, *et al.*** Lessons from the past: familial aggregation analysis of fatal pandemic influenza (Spanish flu) in Iceland in 1918. *Proceedings of the National Academy of Sciences USA* 2008; **105**: 1303–1308.
53. **Zhang L, *et al.*** Systems-based candidate genes for human response to influenza infection. *Infection, Genetics and Evolution* 2009; **9**: 1148–1157.
54. **Pennings JL, Kimman TG, Janssen R.** Identification of a common gene expression response in different lung inflammatory diseases in rodents and macaques. *PLoS ONE* 2008; **3**: e2596.
55. **Gagneux P, *et al.*** Human-specific regulation of alpha 2-6-linked sialic acids. *Journal of Biological Chemistry* 2003; **278**: 48245–48250.
56. **Matrosovich MN, *et al.*** Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proceedings of the National Academy of Sciences USA* 2004; **101**: 4620–4624.
57. **Shinya K, *et al.*** Avian flu: influenza virus receptors in the human airway. *Nature* 2006; **440**: 435–436.
58. **van Riel D, *et al.*** H5N1 virus attachment to lower respiratory tract. *Science* 2006; **312**: 399.
59. **de Jong MD, *et al.*** Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nature Medicine* 2006; **12**: 1203–1207.
60. **Xu T, *et al.*** Acute respiratory distress syndrome induced by avian influenza A (H5N1) virus in mice. *American Journal of Respiratory and Critical Care Medicine* 2006; **174**: 1011–1017.
61. **Maines TR, *et al.*** Pathogenesis of emerging avian influenza viruses in mammals and the host innate immune response. *Immunological Reviews* 2008; **225**: 68–84.
62. **Tecle T, *et al.*** Inhibition of influenza viral neuraminidase activity by collectins. *Archives of Virology* 2007; **152**: 1731–1742.
63. **Tumpey TM, *et al.*** The Mx1 gene protects mice against the pandemic 1918 and highly lethal human H5N1 influenza viruses. *Journal of Virology* 2007; **81**: 10818–10821.
64. **Szretter KJ, *et al.*** Role of host cytokine responses in the pathogenesis of avian H5N1 influenza viruses in mice. *Journal of Virology* 2007; **81**: 2736–2744.
65. **Kash JC, *et al.*** Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature* 2006; **443**: 578–581.

66. **Dittmann J, *et al.*** Influenza A virus strains differ in sensitivity to the antiviral action of Mx-GTPase. *Journal of Virology* 2008; **82**: 3624–3631.
67. **Hidaka F, *et al.*** A missense mutation of the Toll-like receptor 3 gene in a patient with influenza-associated encephalopathy. *Clinical Immunology* 2006; **119**: 188–194.
68. **Le Goffic R, *et al.*** Cutting edge: influenza A virus activates TLR3-dependent inflammatory and RIG-I-dependent antiviral responses in human lung epithelial cells. *Journal of Immunology* 2007; **178**: 3368–3372.
69. **Janssen R, *et al.*** Genetic susceptibility to respiratory syncytial virus bronchiolitis is predominantly associated with innate immune genes. *Journal of Infectious Diseases* 2007; **196**: 826–834.
70. **Ip WK, *et al.*** Mannose-binding lectin in severe acute respiratory syndrome coronavirus infection. *Journal of Infectious Diseases* 2005; **191**: 1697–1704.
71. **Chan VS, *et al.*** Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. *Nature Genetics* 2006; **38**: 38–46.
72. **Hamano E, *et al.*** Polymorphisms of interferon-inducible genes OAS-1 and MxA associated with SARS in the Vietnamese population. *Biochemical and Biophysical Research Communications* 2005; **329**: 1234–1239.
73. **He J, *et al.*** Association of SARS susceptibility with single nucleic acid polymorphisms of OAS1 and MxA genes: a case-control study. *BMC Infectious Diseases* 2006; **6**: 106.
74. **Neumann G, Shinya K, Kawaoka Y.** Molecular pathogenesis of H5N1 influenza virus infections. *Antiviral Therapy* 2007; **12**: 617–626.
75. **World Health Organization Global Influenza Programme Surveillance Network.** Evolution of H5N1 avian influenza viruses in Asia. *Emerging Infectious Diseases* 2005; **11**: 1515–1521.
76. **Hill AV.** Aspects of genetic susceptibility to human infectious diseases. *Annual Review of Genetics* 2006; **40**: 469–86.
77. **Frodsham AJ, Hill AV.** Genetics of infectious diseases. *Human Molecular Genetics* 2004; **13** (Spec. No. 2): R187–194.
78. **Burgner D, Jamieson SE, Blackwell JM.** Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? *Lancet Infectious Diseases* 2006; **6**: 653–663.

Chapter 11

Research paper 5

Title: The role of host genetics in susceptibility to influenza: a systematic review

Author(s): Peter Horby, Nhu Y Nguyen, Sarah J. Dunstan, Kenneth Baillie.

Journal/Publisher: PLoS One, Public Library of Science

Type of publication: Review article

Stage of publication: In press

Academic peer-reviewed: Yes

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Candidate's role: I conceived of the study, wrote the systematic review protocol and registered the protocol on the international prospective register of systematic reviews. I reviewed all the identified titles and abstracts, and read all of the selected full text articles as one of the two independent reviewers. I wrote the first and all subsequent drafts of the manuscript, submitted the manuscript for publication, and responded to all reviewers comments.

Candidate's signature:

Supervisor or senior author's signature to confirm Candidates role:

Title: The role of host genetics in susceptibility to influenza: a systematic review

Running title: Influenza host genetics

Peter Horby[1†], Nhu Y Nguyen[1], Sarah J. Dunstan[1], J. Kenneth Baillie[2].

Authors' Affiliations:

1. Oxford University Clinical Research Unit - Wellcome Trust Major Overseas Programme, Vietnam
2. The Roslin Institute, University of Edinburgh, UK

[†] **Corresponding author**

Word count abstract: 297

Abstract

Background

The World Health Organization has identified studies of the role of host genetics on susceptibility to severe influenza as a priority. A systematic review was conducted to summarize the current state of evidence on the role of host genetics in susceptibility to influenza (PROSPERO registration number: CRD42011001380).

Methods and findings

PubMed, Web of Science, the Cochrane Library, and OpenSIGLE were searched using a pre-defined strategy for all entries up to the date of the search. Two reviewers independently screened the title and abstract of 1371 unique articles, and 72 full text publications were selected for inclusion. Mouse models clearly demonstrate that host genetics plays a critical role in susceptibility to a range of human and avian influenza viruses. The *Mx* genes encoding interferon inducible proteins are the best studied but their relevance to susceptibility in humans is unknown. Although the *MxA* gene should be considered a candidate gene for further study in humans, over 100 other candidate genes have been proposed. There are however no data associating any of these candidate genes to susceptibility in humans, with the only published study in humans being under-powered. One genealogy study presents moderate evidence of a heritable component to the risk of influenza-associated death, and whilst the marked familial aggregation of H5N1 cases is suggestive of host genetic factors, this remains unproven.

Conclusion

The fundamental question *'Is susceptibility to severe influenza in humans heritable?'* remains unanswered. Not because of a lack of genotyping or analytic tools, nor because of insufficient severe influenza cases, but because of the absence of a coordinated effort to define and assemble cohorts of cases. The recent pandemic and the ongoing epizootic of H5N1 both represent rapidly closing windows of opportunity to increase understanding of the pathogenesis of severe influenza through multi-national host genetic studies.

Introduction

The on-going family clustering of highly pathogenic avian influenza A/H5N1 cases, as demonstrated by the deaths in 2011 of a mother and son in Cambodia, and of two siblings and their mother in Indonesia, has led to much speculation that host genetics play a critical role in susceptibility to H5N1 influenza [1-5]. Although H5N1 is an unusually virulent influenza virus, patterns of disease in other influenza epidemics also suggest a possible role for host genetics in susceptibility to severe influenza: around one-quarter to one-half of patients with severe pandemic influenza A/H1N1/09 were previously healthy, with no co-existing medical condition or other predisposing factors [6]. Whilst the viral genetic determinants of influenza severity have been intensively studied, host determinants are much less well studied.

A better understanding of the biological predispositions and pathways leading to severe influenza may lead to improved therapeutic options, and in 2009 the World Health Organization identified studies of the role of host genetic factors on susceptibility to severe influenza as a priority [7,8]. This systematic review was conducted with the objective of summarizing the current state of evidence that host genetic factors play a role in human susceptibility to influenza virus infection or disease.

METHODS

The systematic review was conducted and reported in accordance with the PRISMA guidelines and the protocol was registered on the international prospective register of systematic reviews (PROSPERO registration number: CRD42011001380. Available at: http://www.crd.york.ac.uk/prospero/). Briefly, we conducted a systematic review to summarize relevant published and unpublished evidence of host genetic factors influencing the risk of influenza infection or disease (illness following infection). This comprised a search of PubMed, Web of Science, the Cochrane Library, and OpenSIGLE (grey literature bibliographic database) using a pre-defined search strategy (supplemental material). Two reviewers independently screened all the titles and abstracts to identify publications that may be relevant. A third reviewer assessed the two independent lists of selected and rejected sources and made the final selection where there were discrepancies. The full text of all the sources in the final list was obtained and reviewed independently by two reviewers to decide if they met the inclusion/exclusion criteria. The reference list of all selected sources was reviewed to identify relevant articles that may have been missed by the search strategy. Individual researchers were contacted directly to obtain additional information where the source material could not be obtained or to enquire about on-going or unpublished research.

Certain categories of research were excluded from this review. A very large number of genes are up or down regulated during influenza infection and disease, and it was deemed outside the scope of this piece of work to review the extensive literature on the biological responses to natural or experimental influenza infection. These studies have recently been reviewed elsewhere [9]. Therefore we excluded studies of the molecular biology and pathogenesis of influenza except where the study directly compared the response to infection in genetically distinct animal strains with the objective of identifying host genetic determinants of response. We also excluded studies that solely examined the affects of gene knockouts, since a knockout mouse phenotype, although very useful for understanding pathogenesis, does not provide information on heritability of susceptibility under normal conditions, representing null alleles which rarely occur as such in the human population.

RESULTS

The search strategy was run on 26th June 2011 and identified 1371 unique articles published in English for which the title and abstract was reviewed. 58 met the criteria for full text review, of which 29 were considered relevant to the study and could be obtained (Figure 1). A further 43 relevant articles were identified through a review of the bibliographies of the 29 selected papers and through contact with lead authors. A total of 72 articles were therefore included in the review. The identified published evidence fell into the following categories: studies in animals of host genetics; studies or reports of familial aggregation or heritability; studies in humans of blood group; studies in humans of HLA type; and studies in humans of host genetics.

Animal studies

It has long been known that susceptibility to influenza varies between inbred mouse strains because most laboratory strains carry a mutation in the *Mx1* gene, which is a strong resistance locus for mouse-adapted influenza strains [10]. But more recently, it has been shown that genetic background also plays an important role for resistance or susceptibility, independent of the *Mx1* allele.

Myxovirus resistance gene

The resistance of certain inbred mouse strains to influenza A infection was first reported in 1962 [10] and was subsequently localised to the *Mx1* gene on chromosome 16 [11]. The *Mx1* and *Mx2* genes encode interferon inducible proteins, and Mx1 is able to inhibit influenza virus replication [12-21]. The role of Mx proteins in protection against influenza has recently been reviewed [22]. Susceptible mice have either deletions or a nonsense point mutation in the *Mx1* gene that results in non-functional Mx1 protein [18]. Mice expressing *Mx1* are also better protected from the high mortality caused by the lethal H5N1 (A/Vietnam/1203/04) and H1N1/1918 viruses and from the lung pathology mediated by these viruses [23,24]. Influenza viruses differ in their susceptibility to the action of Mx, with adaptive mutations permitting evasion of the Mx response or rapid viral replication outpacing the Mx response [22,24-28]. The H1N1/1918 and H5N1 (A/Vietnam/1203/04) viruses both demonstrate high replication efficiency and are highly pathogenic, and although both are sensitive to the antiviral activity of Mx, H5N1 (A/Vietnam/1203/04) is more

sensitive than H1N1/1918 [23,24,26]. Influenza virus strains of avian origin appear to have greater sensitivity to Mx than human influenza strains, indicating that adaptive mutations to escape Mx control may be required for successful cross-species transmission [26,28].

Mx gene homologues are found in many species and the homologue in humans is the MxA protein encoded by the *MxA* gene on chromosome 21 [29,30]. In humans MxA demonstrates antiviral activity [13,16,17,19-21] and whilst polymorphisms of the human *MxA* gene exist, their relevance to influenza susceptibility has not been examined.

Other susceptibility loci

Although *Mx* genes are the best studied, there are many other candidates genes for influenza susceptibility. Several groups have directly studied the influence of genetic background on the susceptibility of different mouse strains to influenza. All groups confirm that host genetic background plays a critical role in susceptibility to influenza and that highly susceptible mouse strains develop high viral loads, an elevated inflammatory response, and severe lung pathology following infection with a range of influenza viruses [27,31-33]. These studies were performed on inbred mouse strains that carried an *Mx1* mutant allele.

Toth *et al* have examined the genetic basis of differences between mouse strains (BALB/cByJ and C57BL/6J) in sleep patterns during influenza H3N2 A/Hong Kong/X31/68 (HK-X31) infection, identifying a quantitative trait loci (QTL) on chromosome 6 associated with influenza-induced slow-wave sleep patterns [34]. The group also showed large differences between mouse strains (BALB/cByJ and C57BL/6J) in the expression of genes in the lung following influenza HK-X31 infection [35]. In 2011 the same group showed significant strain differences in disease severity (as measured by survival and body temperature), viral titres and cytokine and chemokine concentrations in the lungs of four inbred strains of mice (BALB/cByJ, C57BL/6J, A/J, and DBA/2J) but did not demonstrate any statistically significant genetic loci associated with influenza HK-X31 severity using a QTL approach, although suggestive statistical associations were reported for regions on three

chromosomes (G-CSF chromosomes 5; CXCL10 chromosome 9, and IL-6 and CXCL1 on chromosome 18) [33].

Srivastava *et al* examined the susceptibility of seven inbred strains to influenza H1N1 A/Puerto Rico/8/34 (PR8) and identified one resistant (C57BL/6J) and one highly susceptible strain (DBA/2J) [31]. The response of these two strains to H7N7 A/Seal/Massachussetts/1/80 (SC35M) was also examined and DBA/2J mice were highly susceptible to SC35M virus infection compared to C57BL/6J mice. A cross between these two strains showed the resistant phenotype, although with a slightly higher weight loss than the parental resistant strain, suggesting that susceptibility in mice may be a polygenic trait. Further studies by this group examined differential gene expression following PR8 infection of susceptible versus resistant mouse strains [36]. Innate immune response genes were up regulated in both strains but to a greater extent in the susceptible strain, and overall a large number of genes were up or down regulated only in the susceptible strain (75, 538, and 993 on days 1, 2, and 3 after infection respectively). Blazejewska *et al* then looked at the effect of three mouse adapted H1N1 PR8 viruses (“low pathogenic” PR8M and PR8F, and “highly virulent” hvPR8) in two mouse strains that had previously been shown to be resistant (C57BL/6) and susceptible (DBA/2J) to PR8M [27,31]. They found that whilst PR8M showed differential pathogenicity in the two strains as previously observed, PR8F and hvPR8 replicated equally well in both strains and caused similar weight loss and mortality, demonstrating that pathogenicity is co-determined by both host and pathogen genetics. Additional studies of the relative sensitivity of the DBA/2J mouse strain compared to C57BL/6 have shown that the DBA/2J strain is susceptible to a wide range of human, avian and swine derived influenza viruses [37,38].

Boon *et al* explored the genetic determinants of susceptibility to an H5N1 virus containing 7 gene segments of A/Hong Kong/213/2003 H5N1 virus and the PB1 gene segment from A/Chicken/Hong Kong/Y0562/2002 H5N1 (termed HK213) infection using gene mapping of 66 strains of inbred mice (inbred between C57BL/6J and DBA/2J parent strains) and identified five genetic loci (quantitative trait loci 2, 7, 11, 15 and 17) associated with resistance to H5N1 HK213 disease [32]. This suggests that multiple genes determine H5N1 susceptibility in mice. A total of 121 genes located within these five loci were identified as candidates based on RNA expression

analysis, which was narrowed to 30 candidates based on differential expression between susceptible and resistant strains. In particular, there were 3, 14, 5, 2 and 6 candidate genes in QTL's 2, 7, 11, 15 and 17, respectively. The authors compared the outcome of HK213 infection in one mouse strain that expressed hemolytic complement and one that did not, finding that strains expressing hemolytic complement (*Hc*) gene, which is located on QTL 2, experienced increased survival rates at a 10-fold higher initial inoculum. However no association between *Hc* expression and susceptibility to influenza was observed in subsequent work by Trammel or Boon [33,39].

Boon *et al* further studied the susceptibility of 21 inbred mouse strains to H5N1 HK213 infection, demonstrating that although viral loads were much higher in susceptible strains, the relationship between viral load and cytokine concentrations was the same in resistant and susceptible strains [39]. The authors concluded that this indicates that mouse strain differences in susceptibility to H5N1 lies in a failure to control viral replication rather than the induction of an aberrant inflammatory response. Gene expression and pathway analysis in six strains showed that differential gene expression mostly consisted of up-regulation in susceptible strains of genes in proinflammatory pathways, indicating the immune response is quantitatively but not qualitatively different between strains. Resistant mouse strains (SMR, C57BL/6R, and BALB/cR) did not express a distinctive set of genes controlling replication or disease. 85 individual genes, again mostly associated with proinflammatory pathways, were identified whose expression was associated with susceptibility to severe disease. Three candidate genes identified in the 2009 study were also significant in the 2011 publication and are being further explored (*Grn, Ifi53,* and *Dhx58*). In summary the 2011 work by Boon *et al* suggest that genetic polymorphisms conferring susceptibility to severe H5N1 disease in mice lie in pathways that are involved in the early control of virus replication.

Summary of animal models

Mouse models clearly demonstrate a strong genetic effect on susceptibility to a range of influenza viruses. The *Mx* genes are the best studied but their relevance to susceptibility in humans is unknown and although the *MxA* gene should be considered

a candidate gene for further studies, there are many other candidates. Crossbred mouse strain studies have identified a large number of potential candidates.

Familial aggregation or heritability

Independent of genetic effects it is expected that influenza infection will aggregate in families since transmission of influenza is common within households. Family aggregation of severe influenza disease is however more likely to have a direct genetic component but such clustering might also be seen with indirect genetic effects (e.g. genetic predisposition to obesity) or non-genetic shared risk factors (e.g. air pollution).

From the perspective of genetic epidemiology, familial aggregation is said to occur when the frequency of a phenotype is more common amongst close relatives of people with the disease than in the general population [40]. Heritability is the proportion of the variation in the frequency of the phenotype that can be attributed to genetic variation. Familial aggregation can occur without heritability if the increased familial risk is due to shared non-genetic factors. On the other hand, genetics can still be important without any detectable heritability, since if there is no genetic variation in a population then heritability is zero, even though all cases may require a particular genetic background. However, significant heritability does suggest the presence of genetic factors that may be detectable by genotyping studies.

Genealogical studies

Two studies utilised large genealogical databases to look for evidence of heritability of susceptibility to death from influenza [41,42]. The study by Albright *et al* used a Utah database to look at 4855 deaths from influenza between 1904 and 2004 [41]. Gottfredsson *et al* concentrated on the 1918 influenza pandemic in Iceland and looked at 455 deaths over a six-week period [42]. Both studies found evidence of familial aggregation of influenza deaths but differed in their conclusions regarding heritability. Albright *et al* concluded that their results supported heritability since there was an increased relative risk of influenza death amongst relatives of people who died of influenza (relative risk 1.54; 95%CI 1.42–1.67; P-value <.0001), and this was greater than observed for relatives of spouses of individuals dying from influenza. Also, influenza deaths in relatives were frequently not associated closely in time (they

studied deaths over 100 years) and there was greater than expected relatedness amongst influenza deaths even after close relatives were excluded. Gottfredsson *et al* concluded that their results did not provide evidence of a heritable predisposition to death from 1918 influenza, as they did not identify a statistically significant difference in the relative risk of influenza death in relatives of people who died of influenza (relative risk in 1st degree relatives = 3.75; 95% CI 2.53-5.24) compared to relatives of their spouses (relative risk in 1st degree relatives = 2.95; 95% CI 2.01-4.49. *P*-value for comparison of relative risk in the two groups = 0.198). The apparently conflicting conclusions of these two studies was discussed by Dowell and Bresee, who highlighted the fact that the highest relative risk of influenza death in both studies was in the spouse of cases, so shared social and environmental conditions are important factors and the family aggregation of severe influenza (for whatever reason) offers opportunities to identify and target high risk individuals [43].

The study by Gottfredsson *et al* had ten-fold fewer subjects than the study by Albright *et al* and as such was considerably less well powered to detect differences in the risk of death in relatives of cases compared to relatives of spouses. Also Gottfredsson's study did not assess the relatedness of cases and was not able to examine deaths outside the six-week period studied, which would be less confounded by common exposures. As such the study by Albright *et al* provides moderate evidence of a heritable component to the risk of influenza death, whereas the Gottfredsson study is inconclusive.

Ethnicity

Racial differences in influenza attack rates have been described historically [44,45]. More recently, an increased risk of hospitalization or death with pandemic influenza H1N1 in indigenous and minority ethnic groups has been reported, particularly in the America's, Australasia and the Pacific [46-51]. Ethnic disparities are observed for many infectious diseases, much of which relates to inequalities in socioeconomic status and related differences in living conditions, access to health care, behaviours, and the prevalence of chronic diseases. No studies have been conducted to determine the genetic component of ethnic differences in rates of influenza hospitalization and death.

Familial aggregation of influenza H5N1

Influenza H5N1 is a rare human infection that displays clustering and familial aggregation of cases [3,5,52]. Around one third of all H5N1 cases occur in clusters and of the 54 H5N1 clusters summarised in January 2010, 50 were comprised only of blood relatives [5]. Pitzer *et al* have examined the familial aggregation of H5N1 cases and argued that although familial aggregation of H5N1 cases is observed, it is more consistent with non-genetic variation in household risk of exposure to H5N1 than host-genetic factors [53]. Horby *et al* have disputed the inferences drawn by Pitzer *et al* and argued that the totality of the epidemiological data is suggestive of a host genetic effect on susceptibility to H5N1 infection [3]. In addition to the familial aggregation of cases the evidence put forward by Horby *et al* includes: the low number of unrelated clusters, the occurrence of related cases that are separated in time and space (and therefore not compatible with common source exposure), and the poor correlation of exposure with risk [3].

Influenza associated encephalopathy (IAE)

Acute encephalitis is a rare but well recognized complication of influenza infection, that occurs mostly in children aged under 5 years and is reported more commonly in East Asia than elsewhere [54]. There is little data to assess if there is genetic susceptibility to IAE other than a report of a mother and daughter with H1N1/09 IAE, two siblings with H5N1 IAE, and an analysis of three IAE cases which reported a missense mutation in the TLR3 gene in one case [55-57]. Acute Necrotizing Encephalopathy (ANE) is a distinct clinical syndrome that is characterised by multiple necrotic brain lesions and is associated with influenza infection but also with other viral infections [58]. A subset of patients with recurrent or familial ANE (ANE1) have a missense mutation in the ran-binding protein 2 (RANBP2) gene on chromosome 2 (q12.3) [58-60]. The mechanism by which this mutation confers susceptibility to ANE is not yet established. ANE is a very distinct clinical syndrome that, whilst having a genetic basis, is unlikely to have any relation to more general susceptibility to influenza.

Summary of familial aggregation or heritability

Although the data are limited and historic, the two genealogy studies clearly demonstrate familial aggregation of the risk of influenza-associated death. The Utah

study presents moderate evidence of a heritable component to the risk of influenza-associated death. Whilst familial aggregation of H5N1 cases is generally accepted, there has been no formal estimation of the excess risk in relatives of cases compared to the general population. Such studies (e.g. familial relative risk studies) are theoretically feasible but challenging given the widespread distribution of H5N1 cases in time and space [61]. Estimating heritability of H5N1 is likely to be impossible since it is probably not feasible to disentangle genetic and non-genetic effects with such small numbers of cases.

Blood group

The 1960's and 70's saw a period of interest in the relationship between the ABO blood group and susceptibility to influenza infection. Studies involved observations of natural influenza infection [62-64], experimental infection [65], and serological studies [64,66-73]. The data are inconsistent, with authors reporting an increased risk of influenza in subjects with blood group O [62,65,66], groups O and B [63,68], B alone [67,74], A [73], A and B [64], AB [64,71], or no difference by blood group [69,70,72]. One group examined the ability to excrete soluble ABO blood group antigens in body fluids (secretor) and the risk of respiratory viral infections, and found a positive association between being a 'secretor' and influenza A infection [75].

Human leucocyte antigen (HLA)

Work in the 1970's by McMichael *et al* and extended by Doherty, Shaw and Biddison demonstrated that cell-mediated lysis of influenza infected cells is dependent on HLA specificities [76-81]. It is now well recognised that the HLA molecules plays a central role in antigen presentation to T-cells and indeed *HLA* is the classic example of genetic susceptibility to infectious diseases and of the influence of infectious diseases on human genomes [82]. Subsequent studies in mice and humans demonstrate that the HLA phenotype (H-2 in mice) influences the magnitude and specificity of the cytotoxic T lymphocyte (CTL) response to influenza infection [83-85]. Considerable work has also been undertaken to identify particular epitope-HLA molecule combinations that are associated with protective CTL responses in order to inform the design of vaccines targeting cell-mediated immunity [86,87]. However no genetic studies have been conducted to identify polymorphisms in HLA loci associated with susceptibility to influenza infection. Given the inherent diversity of HLA loci, the

complex interaction of *HLA* in determining responses to infection, and the linkage of *HLA* to other genes involved in innate immunity, such studies will be challenging [82,85].

Human genetic studies

Only one published human genetic study of susceptibility to influenza was identified. This study was a case control study that included 91 severe H1N1/09 cases and 98 exposed but asymptomatic, unrelated household controls [88]. The authors took a discovery rather than a candidate gene approach, using a commercial chip that incorporates around 50,000 SNPs in regions associated with cardiovascular, metabolic and inflammatory syndromes (HumanCVD Genotyping Beadchip). 28,368 SNPs were analyzed and four SNPs on three different chromosomes had *p*-value *of* <0.0001. These SNPs remained associated after controlling for the potential confounding factors of obesity, diabetes, arterial hypertension, age, gender, and smoking. Three of the SNPs were in genes: an immunoglobulin Fc receptor *(FCGR2A)*; a complement binding protein (*C1QBP);* and a protein that mediates the entry of replication protein A into the nucleus (*RPAIN*). Given the small size of the study, there is a reasonable probability that these are false positive findings, with the false discovery rate (the expected proportion of statistically significant findings that are false positives) for the four SNPs ranging from 22% to 56%.

Reviews

Five review articles were identified. The review by Trammel and Toth summarized animal and human data on genetic influences on influenza infection, with a particular focus on studies of differential gene expression [89]. This review highlighted the earlier work of Toth *et al* that identified 75 immune related genes (including 13 interferon related genes and 10 chemokine related genes) that were differentially expressed in C57BL/6J compared to BALB/cByJ mice in response to influenza H3N2 HK-X31 infection [35]. The review also identified increased expression of seven common genes in both H1N1/1918 and H3N2 HK-X31 infection of BALB/c mice, and 17 genes that showed increased expression in both human bronchial epithelial cell lines and mice infected with H3N2 (A/Udon/307/72 human bronchial epithelial cell, HK-X31 mice) [35,90-92]. The review by Zhang *et al* proposed a list of around 100 candidate genes that may be related to susceptibility to influenza infection based on

existing knowledge of the proteins involved in virus replication and the innate immune response [9]. An Editorial Commentary by Mubareka and Palese on the Utah genealogical study also discussed some potential candidate genes for host susceptibility to influenza, such as mannose-binding lectin, toll-like receptors, retinoic inducible gene I, 2'5'-oligoadenylate synthetase 1, and *MxA* [93]. Horisberger reviewed the data on the relationship between the *Mx1* gene and influenza as it stood in 1995 (see section on *Mx1*) [94]. Horby *et al* reviewed the epidemiological evidence for genetic susceptibility to H5N1 and concluded that the data are suggestive of a host genetic influence on susceptibility to H5N1 disease [3].

DISCUSSION

In mouse models the severity of influenza infection is clearly associated with both the pathogen and host genome. The observation that similar patterns of susceptibility or resistance of specific mouse strains are observed for a wide range of influenza viruses suggests that some of the host genetic determinants of susceptibility may be common across influenza subtypes. Susceptibility in mice is polygenic, and a number of candidate genes, including *MxA*, have been proposed. To date none of these candidate genes have been tested in studies of humans. Animal experiments will continue to be important for refining understanding of host-pathogen genetic interactions and for testing hypotheses about the pathogenesis of severe influenza.

In humans the best available evidence, relying on a single study of 4855 deaths, suggests a heritable component of susceptibility to death from seasonal and pandemic influenza. Given the numerous confounding factors, replication of this finding will require a similarly large study. Although heritability has not been quantified for H5N1, the marked familial aggregation and other epidemiological features suggest a stronger heritable predisposition. To date only one study of human host genetics and susceptibility to severe influenza has been published and no human genetic polymorphisms associated with susceptibility to seasonal, pandemic or avian influenza have been convincingly demonstrated.

Susceptibility to severe seasonal or pandemic influenza in humans is likely to be polygenic and is also likely to be co-determined by pathogen characteristics, prior infection history, co-morbidities, and environmental factors. In addition, the lack of evidence implicating any specific genes in humans suggests a hypothesis-free genome-wide approach should be taken. As such, very large studies will be required to identify genetic effects on susceptibility to severe influenza.

Pandemic H1N1 offers a rare opportunity to study genetic susceptibility to severe influenza in a context that, compared to seasonal influenza, is less confounded by infection history and pathogen diversity. However, large sample sizes will still be required to detect polygenic traits and case selection will need to consider confounding by cross-protective immunity and co-morbidity. Several groups have

compiled series of severe H1N1/09 cases but it seems very unlikely that any single group will have sufficient cases to conduct an adequately powered genome-wide association study [95]. To have a realistic prospect of identifying susceptibility loci for H1N1/09, groups will need to form a consortium, as has been successful for other diseases [96]. The chances of identifying susceptibility loci in H1N1/09 can be enhanced by adopting an 'extreme-trait' study design e.g. where cases are previously healthy young adults who develop very severe disease with high viral loads and no evidence of bacterial co-infection. Influenza encephalitis is another 'extreme-trait' where case cohorts should be assembled for comparison with other influenza disease cohorts. There may still be possibilities to study susceptibility to 1918 pandemic influenza through linkage studies within large genealogical cohorts, where pedigree and cause of death data stretch back to the early 1900's [41].

Susceptibility to H5N1 may be less complex than 'human influenza', since the phenotype appears to be more dichotomous than continuous, immunity probably plays a lesser role, co-morbidity seems less important, and familial aggregation is more marked. The importance of understanding the pathogenesis of highly pathogenic influenza and the possibility that a rare genetic variant with a moderate to large effect underlies H5N1 susceptibility makes efforts to assemble DNA from H5N1 cases worthwhile. Given the small number of H5N1 cases and the possibility of a rare variant with a moderate to large effect, genome-wide association studies may not be the optimal design and alternative approaches to identifying causal loci may be needed, such as sequencing candidate genes, the whole exome, or the even whole genome [97,98]. Purely epidemiological studies may contribute to understanding the genetic component of familial aggregation of H5N1 by quantifying heritability.

High viral replication efficiency, or from a host perspective a failure to control virus replication, is emerging as a key factor in severe influenza disease and is determined by both host and virus factors [27,39]. Thus studies of the determinants of influenza severity may benefit from a combined host-pathogen genetics approach, where the analysis of host genetic associations is conditioned upon the pathogen genotype in order to identify genotype-genotype interactions.

Conclusion

The fundamental question 'Is susceptibility to severe influenza in humans heritable?' remains unanswered. It is unanswered not because of a lack of genotyping or analytic tools, nor because of insufficient severe influenza cases, but because of the absence of a coordinated effort to define and assemble cohorts of cases. The recent pandemic and the ongoing epizootic of H5N1 both represent rapidly closing windows of opportunity to increase understanding of the pathogenesis of severe influenza through multi-national host genetic studies.

Financial Disclosure

This work was supported by the World Health Organization and the Wellcome Trust UK (grants 081613/Z/06/Z and 077078/Z/05/Z). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived the study and designed the experiments: PH, NYN, SD, KB. Performed the experiments: NYN, PH, SD, KB. Analyzed the data: PH, NYN. Wrote the first draft of paper: PH.

Competing Interests

All authors declare that no competing interests exist.

References

1. Sedyaningsih ER, Isfandari S, Setiawaty V, Rifati L, Harun S, et al. (2007) Epidemiology of cases of H5N1 virus infection in Indonesia, July 2005-June 2006. J Infect Dis 196: 522-527.
2. Kandun IN, Wibisono H, Sedyaningsih ER, Yusharmen, Hadisoedarsuno W, et al. (2006) Three Indonesian clusters of H5N1 virus infection in 2005. N Engl J Med 355: 2186-2194.
3. Horby P, Sudoyo H, Viprakasit V, Fox A, Thai PQ, et al. (2010) What is the evidence of a role for host genetics in susceptibility to influenza A/H5N1? Epidemiol Infect 138: 1550-1558.
4. Aditama TY, Samaan G, Kusriastuti R, Purba WH, Misriyah, et al. (2011) Risk Factors for Cluster Outbreaks of Avian Influenza A H5N1 Infection, Indonesia. Clin Infect Dis 53: 1237-1244.
5. (2010) Summary of human infection with highly pathogenic avian influenza A (H5N1) virus reported to WHO, January 2003-March 2009: cluster-associated cases. Wkly Epidemiol Rec 85: 13-20.
6. Bautista E, Chotpitayasunondh T, Gao Z, Harper SA, Shaw M, et al. (2010) Clinical aspects of pandemic 2009 influenza A (H1N1) virus infection. N Engl J Med 362: 1708-1719.
7. WHO (2009) WHO Public Health Research Agenda for Influenza. Geneva.
8. Konig R, Stertz S, Zhou Y, Inoue A, Hoffmann HH, et al. (2010) Human host factors required for influenza virus replication. Nature 463: 813-817.
9. Zhang L, Katz JM, Gwinn M, Dowling NF, Khoury MJ (2009) Systems-based candidate genes for human response to influenza infection. Infect Genet Evol 9: 1148-1157.
10. Lindenmann J (1962) Resistance of mice to mouse-adapted influenza A virus. Virology 16: 203-204.
11. Staeheli P, Pravtcheva D, Lundin LG, Acklin M, Ruddle F, et al. (1986) Interferon-regulated influenza virus resistance gene Mx is localized on mouse chromosome 16. J Virol 58: 967-969.
12. Arnheiter H, Haller O, Lindenmann J (1976) Pathology of influenza hepatitis in susceptible and genetically resistant mice. Exp Cell Biol 44: 95-107.
13. Haller O, Arnheiter H, Gresser I, Lindenmann J (1979) Genetically determined, interferon-dependent resistance to influenza virus in mice. J Exp Med 149: 601-612.
14. Haller O, Arnheiter H, Lindenmann J, Gresser I (1980) Host Gene Influences Sensitivity to Interferon Action Selectivity for Influenza-Virus. Nature 283: 660-662.
15. Arnheiter H, Haller O, Lindenmann J (1980) Host gene influence on interferon action in adult mouse hepatocytes: specificity for influenza virus. Virology 103: 11-20.
16. Arnheiter H, Haller O (1983) Mx gene control of interferon action: different kinetics of the antiviral state against influenza virus and vesicular stomatitis virus. J Virol 47: 626-630.
17. Staeheli P, Danielson P, Haller O, Sutcliffe JG (1986) Transcriptional activation of the mouse Mx gene by type I interferon. Mol Cell Biol 6: 4770-4774.
18. Staeheli P, Grob R, Meier E, Sutcliffe JG, Haller O (1988) Influenza Virus-Susceptible Mice Carry Mx-Genes with a Large Deletion or a Nonsense Mutation. Molecular and Cellular Biology 8: 4518-4523.
19. Arnheiter H, Skuntz S, Noteborn M, Chang S, Meier E (1990) Transgenic mice with intracellular immunity to influenza virus. Cell 62: 51-61.
20. Holzinger D, Jorns C, Stertz S, Boisson-Dupuis S, Thimme R, et al. (2007) Induction of MxA gene expression by influenza A virus requires type I or type III interferon signaling. J Virol 81: 7776-7785.
21. Koerner I, Kochs G, Kalinke U, Weiss S, Staeheli P (2007) Protective role of beta interferon in host defense against influenza A virus. J Virol 81: 2025-2030.

22. Haller O, Staeheli P, Kochs G (2009) Protective role of interferon-induced Mx GTPases against influenza viruses. Rev Sci Tech 28: 219-231.
23. Salomon R, Staeheli P, Kochs G, Yen HL, Franks J, et al. (2007) Mx1 gene protects mice against the highly lethal human H5N1 influenza virus. Cell Cycle 6: 2417-2421.
24. Tumpey TM, Szretter KJ, Van Hoeven N, Katz JM, Kochs G, et al. (2007) The Mx1 gene protects mice against the pandemic 1918 and highly lethal human H5N1 influenza viruses. J Virol 81: 10818-10821.
25. Grimm D, Staeheli P, Hufbauer M, Koerner I, Martinez-Sobrido L, et al. (2007) Replication fitness determines high virulence of influenza A virus in mice carrying functional Mx1 resistance gene. Proc Natl Acad Sci U S A 104: 6806-6811.
26. Dittmann J, Stertz S, Grimm D, Steel J, Garcia-Sastre A, et al. (2008) Influenza A virus strains differ in sensitivity to the antiviral action of Mx-GTPase. J Virol 82: 3624-3631.
27. Blazejewska P, Koscinski L, Viegas N, Anhlan D, Ludwig S, et al. (2011) Pathogenicity of different PR8 influenza A virus variants in mice is determined by both viral and host factors. Virology 412: 36-45.
28. Zimmermann P, Manz B, Haller O, Schwemmle M, Kochs G (2011) The viral nucleoprotein determines Mx sensitivity of influenza A viruses. J Virol 85: 8133-8140.
29. Reeves RH, O'Hara BF, Pavan WJ, Gearhart JD, Haller O (1988) Genetic mapping of the Mx influenza virus resistance gene within the region of mouse chromosome 16 that is homologous to human chromosome 21. J Virol 62: 4372-4375.
30. Malo D, Skamene E (1994) Genetic-Control of Host-Resistance to Infection. Trends in Genetics 10: 365-371.
31. Srivastava B, Blazejewska P, Hessmann M, Bruder D, Geffers R, et al. (2009) Host genetic background strongly influences the response to influenza a virus infections. PLoS One 4: e4857.
32. Boon AC, deBeauchamp J, Hollmann A, Luke J, Kotb M, et al. (2009) Host genetic variation affects resistance to infection with a highly pathogenic H5N1 influenza A virus in mice. J Virol 83: 10417-10426.
33. Trammell RA, Liberati TA, Toth LA (2011) Host genetic background and the innate inflammatory response of lung to influenza virus. Microbes Infect.
34. Toth LA, Williams RW (1999) A quantitative genetic analysis of slow-wave sleep in influenza-infected CXB recombinant inbred mice. Behav Genet 29: 339-348.
35. Ding M, Lu L, Toth LA (2008) Gene expression in lung and basal forebrain during influenza infection in mice. Genes Brain Behav 7: 173-183.
36. Alberts R, Srivastava B, Wu H, Viegas N, Geffers R, et al. (2010) Gene expression changes in the host response between resistant and susceptible inbred mouse strains after influenza A infection. Microbes Infect 12: 309-318.
37. Pica N, Iyer A, Ramos I, Bouvier NM, Fernandez-Sesma A, et al. (2011) The DBA.2 mouse is susceptible to disease following infection with a broad, but limited, range of influenza A and B viruses. J Virol.
38. Boon AC, deBeauchamp J, Krauss S, Rubrum A, Webb AD, et al. (2010) Cross-reactive neutralizing antibodies directed against pandemic H1N1 2009 virus are protective in a highly sensitive DBA/2 mouse influenza model. J Virol 84: 7662-7667.
39. Boon AC, Finkelstein D, Zheng M, Liao G, Allard J, et al. (2011) H5N1 Influenza Virus Pathogenesis in Genetically Diverse Mice Is Mediated at the Level of Viral Load. MBio 2.
40. Burton PR, Tobin MD, Hopper JL (2005) Key concepts in genetic epidemiology. Lancet 366: 941-951.
41. Albright FS, Orlando P, Pavia AT, Jackson GG, Cannon Albright LA (2008) Evidence for a heritable predisposition to death due to influenza. J Infect Dis 197: 18-24.
42. Gottfredsson M, Halldorsson BV, Jonsson S, Kristjansson M, Kristjansson K, et al. (2008) Lessons from the past: familial aggregation analysis of fatal pandemic influenza (Spanish flu) in Iceland in 1918. Proc Natl Acad Sci U S A 105: 1303-1308.

43. Dowell SF, Bresee JS (2008) Pandemic lessons from Iceland. Proc Natl Acad Sci U S A 105: 1109-1110.
44. Armstrong DB (1919) Influenza Observations in Framingham, Massachusetts. Am J Public Health (N Y) 9: 960-965.
45. (1950) Racial Susceptibility to Influenza. Lancet 258: 81-81.
46. La Ruche G, Tarantola A, Barboza P, Vaillant L, Gueguen J, et al. (2009) The 2009 pandemic H1N1 influenza and indigenous populations of the Americas and the Pacific. Euro Surveill 14.
47. Zarychanski R, Stuart TL, Kumar A, Doucette S, Elliott L, et al. (2010) Correlates of severe disease in patients with 2009 pandemic influenza (H1N1) virus infection. CMAJ 182: 257-264.
48. Thompson DL, Jungk J, Hancock E, Smelser C, Landen M, et al. (2011) Risk factors for 2009 pandemic influenza A (H1N1)-related hospitalization and death among racial/ethnic groups in New Mexico. Am J Public Health 101: 1776-1784.
49. (2009) Deaths related to 2009 pandemic influenza A (H1N1) among American Indian/Alaska Natives - 12 states, 2009. MMWR Morb Mortal Wkly Rep 58: 1341-1344.
50. Van Kerkhove MD, Vandemaele KA, Shinde V, Jaramillo-Gutierrez G, Koukounari A, et al. (2011) Risk factors for severe outcomes following 2009 influenza A (H1N1) infection: a global pooled analysis. PLoS Med 8: e1001053.
51. Bandaranayake D, Huang QS, Bissielo A, Wood T, Mackereth G, et al. (2010) Risk factors and immunity in a nationally representative population following the 2009 influenza A(H1N1) pandemic. PLoS ONE 5: e13211.
52. Olsen SJ, Ungchusak K, Sovann L, Uyeki TM, Dowell SF, et al. (2005) Family clustering of avian influenza A (H5N1). Emerg Infect Dis 11: 1799-1801.
53. Pitzer VE, Olsen SJ, Bergstrom CT, Dowell SF, Lipsitch M (2007) Little evidence for genetic susceptibility to influenza A (H5N1) from family clustering data. Emerg Infect Dis 13: 1074-1076.
54. Wang GF, Li W, Li K (2010) Acute encephalopathy and encephalitis caused by influenza virus infection. Curr Opin Neurol 23: 305-311.
55. Hidaka F, Matsuo S, Muta T, Takeshige K, Mizukami T, et al. (2006) A missense mutation of the Toll-like receptor 3 gene in a patient with influenza-associated encephalopathy. Clin Immunol 119: 188-194.
56. Gonzalez BE, Brust DG (2009) Novel influenza A (H1N1) presenting as an acute febrile encephalopathy in a mother and daughter. Clin Infect Dis 49: 1966-1967.
57. de Jong MD, Bach VC, Phan TQ, Vo MH, Tran TT, et al. (2005) Fatal avian influenza A (H5N1) in a child presenting with diarrhea followed by coma. N Engl J Med 352: 686-691.
58. Mizuguchi M, Abe J, Mikkaichi K, Noma S, Yoshida K, et al. (1995) Acute necrotising encephalopathy of childhood: a new syndrome presenting with multifocal, symmetric brain lesions. J Neurol Neurosurg Psychiatry 58: 555-561.
59. Neilson DE, Adams MD, Orr CM, Schelling DK, Eiben RM, et al. (2009) Infection-triggered familial or recurrent cases of acute necrotizing encephalopathy caused by mutations in a component of the nuclear pore, RANBP2. Am J Hum Genet 84: 44-51.
60. Gika AD, Rich P, Gupta S, Neilson DE, Clarke A Recurrent acute necrotizing encephalopathy following influenza A in a genetically predisposed family. Dev Med Child Neurol 52: 99-102.
61. Haralambous E, Weiss HA, Radalowicz A, Hibberd ML, Booy R, et al. (2003) Sibling familial risk ratio of meningococcal disease in UK Caucasians. Epidemiol Infect 130: 413-418.
62. McDonald JC, Zuckerman AJ (1962) ABO Blood Groups and Acute Respiratory Virus Disease. Br Med J 2: 89-90.
63. Frolov VK, Sokhin AA, Sotnik AY, Frolov AK, Lebedinsky AP, et al. (1975) Polymorphism of human blood groups and incidence of influenza A/Hong Kong (H3N2). Acta Virol 19: 406-412.

64. Lebiush M, Rannon L, Kark JD (1981) The relationship between epidemic influenza (A(H1N1) and ABO blood group. J Hyg (Lond) 87: 139-146.
65. Tyrrell DA, Sparrow P, Beare AS (1968) Relation between blood groups and resistance to infection with influenza and spome picornaviruses. Nature 220: 819-820.
66. Potter CW, Schild GC (1967) The incidence of HI antibody to Influenza virus A2/Singapore/ 1/57 in individuals of blood groups A and O. J Immunol 98: 1320-1325.
67. Mackenzie JS, Fimmel PJ (1978) The effect of ABO blood groups on the incidence of epidemic influenza and on the response to live attenuated and detergent split influenza virus vaccines. J Hyg (Lond) 80: 21-30.
68. Cuadrado RR, Davenport FM (1970) Antibodies of influenza viruses in military recruits from Argentina, Brazil and Colombia. Their relation to ABO blood group distribution. Bull World Health Organ 42: 873-884.
69. Evans AS, Shepard DA, Richards VA (1972) ABO blood groups and viral diseases. Yale J Biol Med 45: 81-92.
70. Potter CW (1969) HI antibody to various influenza viruses and adenoviruses in individuals of blood groups A and O. J Hyg (Lond) 67: 67-74.
71. Aho K, Pyhala R, Visakorpi R (1980) ABO associated genetic determinant in H1N1 influenza. Tissue Antigens 16: 310-313.
72. Watkin IJ, Tills D, Heath RB (1975) Studies of the genetic susceptibility of individuals to infection with influenza viruses. Humangenetik 30: 75-79.
73. Tyrrell DA, Peto M, King N (1967) Serological studies on infections by respiratory viruses of the inhabitants of Tristan da Cunha. J Hyg (Lond) 65: 327-341.
74. Mackenzie JS, Wetherall JD, Fimmel PJ, Hawkins BR, Dawkins RL (1977) Host factors and susceptibility to influenza A infection: the effect of ABO blood groups and HL-A antigens. Dev Biol Stand 39: 355-362.
75. Raza MW, Blackwell CC, Molyneaux P, James VS, Ogilvie MM, et al. (1991) Association between secretor status and respiratory viral illness. BMJ 303: 815-818.
76. McMichael AJ, Ting A, Zweerink HJ, Askonas BA (1977) HLA restriction of cell-mediated lysis of influenza virus-infected human cells. Nature 270: 524-526.
77. McMichael A (1978) HLA restriction of human cytotoxic T lymphocytes specific for influenza virus. Poor recognition of virus associated with HLA A2. J Exp Med 148: 1458-1467.
78. Shaw S, Biddison WE (1979) HLA-linked genetic control of the specicity of human cytotoxic T-cell responses to influenza virus. J Exp Med 149: 565-575.
79. Biddison WE, Shaw S (1979) Differences in HLA antigen recognition by human influenza virus-immune cytotoxic T cells. J Immunol 122: 1705-1709.
80. Shaw S, Shearer GM, Biddison WE (1980) Human cytotoxic T-cell responses to type A and type B influenza viruses can be restricted by different HLA antigens. Implications for HLA polymorphism and genetic regulation. J Exp Med 151: 235-245.
81. Doherty PC, Biddison WE, Bennink JR, Knowles BB (1978) Cytotoxic T-cell responses in mice infected with influenza and vaccinia viruses vary in magnitude with H-2 genotype. J Exp Med 148: 534-543.
82. Blackwell JM, Jamieson SE, Burgner D (2009) HLA and infectious diseases. Clin Microbiol Rev 22: 370-385, Table of Contents.
83. Boon AC, de Mutsert G, Graus YM, Fouchier RA, Sintnicolaas K, et al. (2002) The magnitude and specificity of influenza A virus-specific cytotoxic T-lymphocyte responses in humans is related to HLA-A and -B phenotype. J Virol 76: 582-590.
84. Belz GT, Stevenson PG, Doherty PC (2000) Contemporary analysis of MHC-related immunodominance hierarchies in the CD8+ T cell response to influenza A viruses. J Immunol 165: 2404-2409.
85. Day EB, Charlton KL, La Gruta NL, Doherty PC, Turner SJ (2011) Effect of MHC class I diversification on influenza epitope-specific CD8+ T cell precursor frequency and subsequent effector function. J Immunol 186: 6319-6328.

86. Wu C, Zanker D, Valkenburg S, Tan B, Kedzierska K, et al. (2011) Systematic identification of immunodominant CD8+ T-cell responses to influenza A virus in HLA-A2 individuals. Proc Natl Acad Sci U S A 108: 9178-9183.
87. Hertz T, Nolan D, James I, John M, Gaudieri S, et al. (2011) Mapping the landscape of host-pathogen coevolution: HLA class I binding and its relationship with evolutionary conservation in human and viral proteins. J Virol 85: 1310-1321.
88. Zuniga J, Buendia I, Zhao Y, Jimenez L, Torres D, et al. (2011) Genetic variants associated with severe pneumonia in A/H1N1 influenza infection. Eur Respir J.
89. Trammell RA, Toth LA (2008) Genetic susceptibility and resistance to influenza infection and disease in humans and mice. Expert Rev Mol Diagn 8: 515-529.
90. Kash JC, Basler CF, Garcia-Sastre A, Carter V, Billharz R, et al. (2004) Global host immune response: pathogenesis and transcriptional profiling of type A influenza viruses expressing the hemagglutinin and neuraminidase genes from the 1918 pandemic virus. J Virol 78: 9499-9511.
91. Kash JC, Tumpey TM, Proll SC, Carter V, Perwitasari O, et al. (2006) Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. Nature 443: 578-581.
92. Hayashi S, Jibiki I, Asai Y, Gon Y, Kobayashi T, et al. (2008) Analysis of gene expression in human bronchial epithelial cells upon influenza virus infection and regulation by p38 mitogen-activated protein kinase and c-Jun-N-terminal kinase. Respirology 13: 203-214.
93. Mubareka S, Palese P (2008) Human genes and influenza. J Infect Dis 197: 1-3.
94. Horisberger MA (1995) Interferons, Mx genes, and resistance to influenza virus. Am J Respir Crit Care Med 152: S67-71.
95. Calafell i Majo F, Gonzalez Candelas F (2011) [Genetic factors in severe cases of (H1N1) 2009 influenza]. Rev Esp Salud Publica 85: 33-36.
96. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678.
97. Cirulli ET, Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11: 415-425.
98. Ng SB, Nickerson DA, Bamshad MJ, Shendure J (2010) Massively parallel sequencing and rare disease. Hum Mol Genet 19: R119-124.

Table 1. Key studies of familial aggregation, heritability, or genetic susceptibility in humans

Author (Year)	Study / Investigation	Main Findings
Albright FS (2008) [41]	Study of 4855 deaths from influenza between 1904 and 2004 in a Utah genealogical database.	Evidence of heritability included: risk of influenza death greater in relatives of people who died of influenza than in relatives of the spouse of the person dying of influenza. Deaths in related people frequently did not occur close in time. Greater 'relatedness' amongst influenza deaths compared to age, gender and location matched controls.
Gottfredsson M (2008) [42]	Study of 455 deaths from 1918 influenza over a six-week period in Iceland.	Familial aggregation of deaths was observed but there was no detectable heritable component as the difference in the risk of death between relatives of people who died of influenza and relatives of their spouse was not statistically significant.
Mubareka S (2008) [93]	Commentary on the two genealogy studies	Heritability is unproven but the high risk in spouses identified in both studies indicates that people who share households with severe influenza cases are themselves at increased risk of severe influenza.
Pitzer VE (2007) [53]	Analysis of family clustering of H5N1 cases	A high proportion of household clusters would be expected to be limited to ‘blood relatives’ by chance alone.
Horby P (2010) [3]	Review of epidemiology of H5N1 cases	Epidemiological patterns that suggest host genetic susceptibility include familial aggregation of cases, related cases occurring separated by time and place, and low apparent risk in people who are highly exposed.
Olsen S (2005) [52]	Summary of family clustering of H5N1 cases	15 H5N1 clusters occurring between December 2003 and July 2005 were summarised.
WHO (2011) [5]	Summary of H5N1 clusters reported to WHO, January 2003-March 2009	Amongst a total of 480 Human H5N1 cases reported to WHO there were 54 clusters involving 138 cases (29% of cases). The remaining 342 cases were sporadic. In 50 clusters everyone was a blood relative. In the 4 remaining clusters, 2 clusters that included

		>3 people, 9/11 people were blood relatives; and in 2 clusters, each contained 2 unrelated people.
Zhang L (2009) [9]	Review of candidate genes for influenza disease and immunity.	Proposed a list of around 100 candidate genes based on published literature of their potential role in the pathogenesis of influenza.
Zuniga J (2011) [88]	Case-control genetic association study. 91 cases of A/H1N1/2009 associated pneumonia and 98 exposed but asymptomatic household contacts. Genotyped using a cardiovascular disease chip with around 50,000 SNPs.	Four SNPs were associated with severe pneumonia with a $p<0.0001$ after adjustment for gender and comorbidities (obesity, hypertension, and diabetes).

Table 2. Key studies of heritability or genetic susceptibility in mice

Author (Year)	Study / Investigation	Main Findings
Mx		
Lindenmann J (1962) [10]	Experimental inoculation of A2G mice with H1N1/NWS/1933 virus	A2G mice exhibit considerable resistance to intracerebral and intranasal to H1N1/NWS/1933 inoculation.
Staeheli P (1988) [18]	Molecular analysis of *Mx1* alleles using restriction fragment length polymorphism (RFLP) and southern blot analysis of classical inbred mouse strains.	The establishment of $Mx1^{+}$ and $Mx1^{-}$ mouse lines was due to a single nonsense mutation in the *Mx* gene, which was represented in present-day mice by the prototype strains A2G and CBA/J.
Horisberger MA (1995) [94]	Review article of *Mx* genes and influenza	
Salomon R (2007) [23]	Comparison of the effect in mice with and without a functional *Mx1* gene of inoculation with H5N1 A/Vietnam/1203/04 and reassortants with the non-lethal virus A/chicken/Vietnam/C58/04.	Compared to $Mx1^{-/-}$ mice, $Mx1^{+/+}$ mice were protected from A/Vietnam/1203/04, showing lower viral tires, less pathology, and no deaths.
Tumpey TM (2007) [24]	Comparison of the effect in mice with and without a functional *Mx1* gene of inoculation with H1N1/1918 and H5N1 A/Vietnam/1203/04.	Compared to $Mx1^{-/-}$ mice, $Mx1^{+/+}$ mice were protected from 1918 H1N1 and A/Vietnam/1203/04, showing lower viral tires, less weight loss, and fewer deaths.
Grimm D (2007) [25]	Characterization of influenza A H1N1 (PR8) that is unusually virulent in $Mx1^{+/+}$ mice.	Virulence of PR8 is due to high replication ability, not inherent resistance to Mx1.
Dittmann J (2008) [26]	In-vitro study of the inhibitory effect of mouse Mx1 protein and human MxA protein on different influenza strains in cell culture or minireplicon assay.	Influenza A viruses varied in their sensitivity to Mx proteins, with avian virus showing greater sensitivity than human viruses.
Haller O (2009) [22]	Review article of *Mx* genes and influenza	

Zimmermann P (2011) [28]	Study of the inhibitory effect of mouse Mx1 protein and human MxA protein on H1N1/09 (A/Hamburg/4/09) and highly pathogenic avian H5N1 (A/Thailand/1(KAN-1)/04)	H5N1 (A/Thailand/1(KAN-1)/04) was more sensitive to Mx proteins than H1N1/09 (A/Hamburg/4/09). This sensitivity was associated with the NP gene.
Other susceptibility loci		
Toth LA (1999) [34]	Study of strain associated variation in slow-wave-sleep patterns in response to influenza H3N2 (HK-X31) infection. Sleep measurement of 13 recombinant inbred strains, which were from a cross between C57BL/6ByJ and BALB/cByJ mice. Quantitative Trait Loci (QTL) linked to phenotype were identified using a genome wide linkage scan against 223 loci.	A 10- to 12-cM interval on chromosome 6 between *D6Mit74* and *D6Mit188* contains a quantitative trait loci (QTL) affecting the SWS response to influenza infection during the light phase.
Ding M (2008) [35]	Complementary DNA microarray analysis of lung and basal forebrain of influenza H3N2 (HK-X31) infected and uninfected BALB/cByJ and C57BL/6J mice	Gene expression in the lung in response to influenza infection was greater in BALB/cByJ. In lung, 361 different genes changed expression after influenza infection of BALB/cByJ mice as compared with 16 in C57BL/6J mice. Of 75 genes related to the immune response, 3 showed increased expression in the lungs of infected C57BL/6J mice, compared with 70 in infected BALB/cByJ mice.
Trammell RA (2008) [89]	Review article of human and animal data on host genetic susceptibility to influenza.	
Srivastava B (2009) [31]	Comparison of response to H1N1 (PR8) infection in seven inbred laboratory mouse	Different strains exhibited large differences in their response to PR8 infection. DBA/2J mice

	strains. Additional comparison of response to H7N7 (SC35M) infection in one of the susceptible strains (DBA/2J) and one of the more resistant strains (C57BL/6J).	were highly susceptible to both H1N1 (PR8) and H7N7 (SC35M) infection compared to C57BL/6J mice. DBA/2J mice showed higher viral loads, higher cytokine and chemokine expression, and greater lung pathology compared to C57BL/6J mice.
Boon AC (2009) [32]	Comparison of response of susceptible (DBA/2J) and resistant (C57BL/6J) mice, and 66 recombinant inbred mouse strains to H5N1 (HK213) infection using genome-wide linkage analysis and RNA expression analysis. HK213 was selected for its reduced lethality in C57BL/6J mice while retaining lethality in DBA/2J mice.	Following HK213 infection susceptible strains showed greater viral loads and pro-inflammatory cytokines than resistant strains. Gene mapping with recombinant inbred strains revealed five Quantitative Trait Loci located on Chromosomes 2, 7, 11, 15, and 17 associated with resistance to HK213 virus. 121 unique candidate genes were identified whose genetic polymorphisms or different expression levels may have affect H5N1 pathogenesis.
Alberts R (2010) [36]	Comparison of response to H1N1 (PR8) infection in susceptible (DBA/2J) versus resistant (C57BL/6J) mouse strains, analyzed by microarray gene expression analysis.	DBA/2J mice had a sstronger chemokine/cytokine response. Innate immune response genes were up regulated in both strains but to a greater extent in the susceptible strain, and overall a large number of genes were up or down regulated only in the susceptible strain.
Boon AC (2011) [39]	Comparison of viral loads and host responses in 21 inbred mouse strains infected with H5N1 (HK213). RNA expression and chemokine/cytokine analysis was undertaken	Susceptible strains exhibited higher viral loads and concentrations of proinflammatory mediators. There was increased expression of proinflammatory genes in susceptible strains

	in three susceptible strains (DBA/2S, 129/SvImS, and A/JS) and three resistant strains (SMR, C57BL/6R, and BALB/cR).	compared to resistant strains. Relationship between viral load and cytokine concentrations was the same in resistant and susceptible strains.
Trammell RA (2011) [33]	Evaluation of survival, viral load, and cytokine/chemokine responses induced in lung of four inbred mouse strains (BALB/cByJ, C57BL/6J, A/J, and DBA/2J) and QTL mapping 21 recombinant inbred strains following exposure to H3N2 (HK-X31).	DBA/2J mice demonstrated greater susceptibility to severe disease. There were variable response patterns of mouse strains after in vivo and in vitro exposure to HK-X31. No significant QTL were detected.
Blazejewska P (2011) [27]	Comparison between DBA/2J and C57BL/6J mice of infection with three mouse-adapted variants of the H1N1 PR8 strain: PR8M, PR8F and hvPR8.	The PR8F and the hvPR8 variants were lethal for both DBA/2J and C57BL/6J mouse strains; however, the PR8M variant is only lethal for DBA/2J mice. Infection of C57BL/6J mice with a re-assorted PR8 virus demonstrated that the HA gene is the primary determinant of virulence of the PR8F variant.

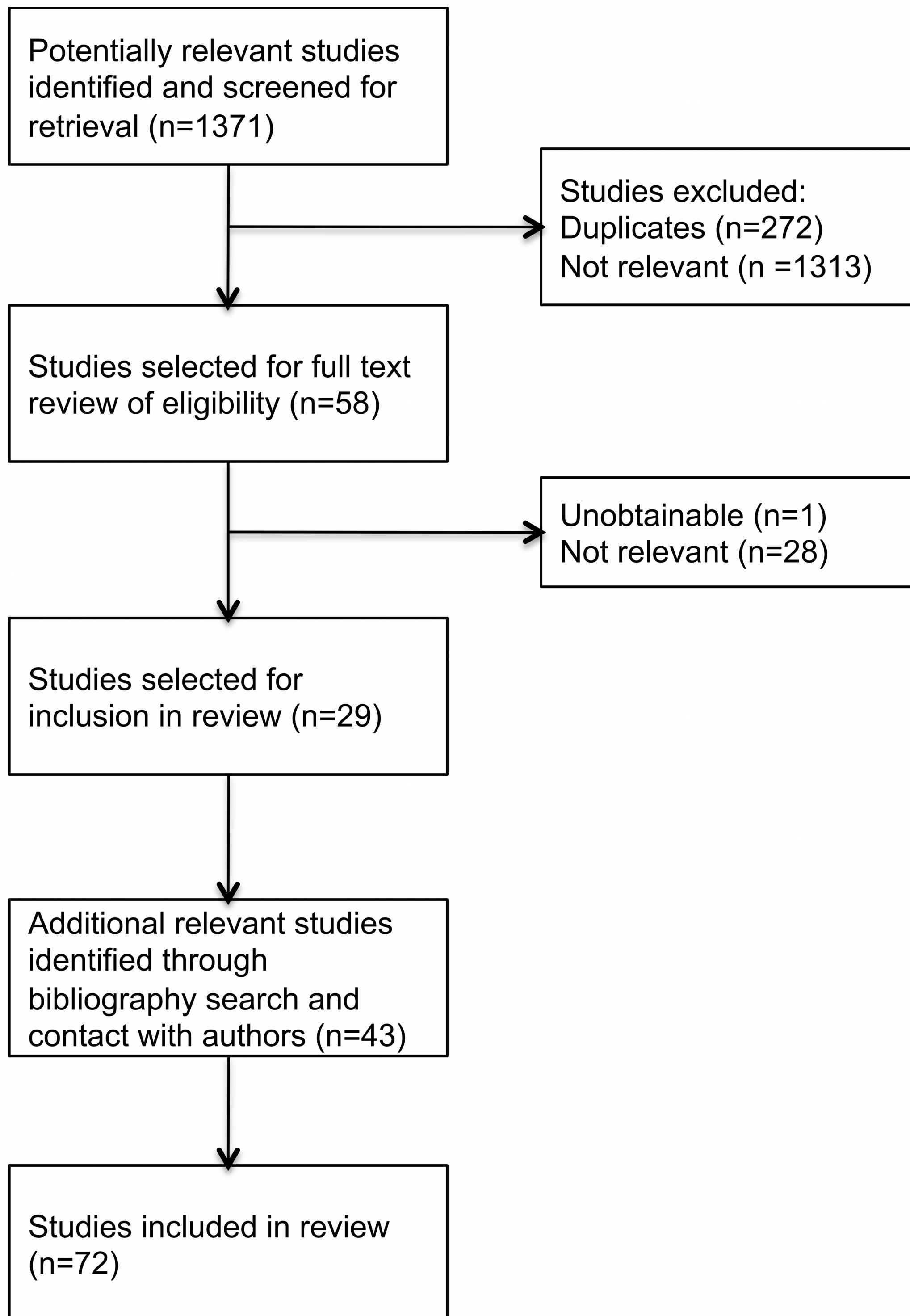


Figure 1. Identification and screening of articles for inclusion in systematic review.

Chapter 12

Research paper 6

Title: Susceptibility to highly pathogenic avian influenza A/H5N1 associated with interleukin 1 and TRPM8.

Author(s): Peter Horby, Chiea C. Khor, Annette Fox, Le Q. Mai, Surakameth Mahasirimongkol, Pham Q. Thai, Nuanjun Wichukchinda, Nguyen T. Yen, Sukanya Wattanapokayakit, Nusara Satproedprai, Luong K. Lan, Nguyen T. Hang, Vip Viprakasi, Sonia Davila, Jeremy Farrar, Tran T. Hien, Tawee Chotpitayasunondh, Sarah J. Dunstan1, Nguyen T. Hien, Martin Hibberd.

Journal/Publisher: PLoS Medicine, Public Library of Science

Type of publication: Research article

Stage of publication: Submitted

Academic peer-reviewed: Ongoing

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Candidate's role: I conceived and designed the study. All the data presented in this paper were collected through the H5N1 host genetics study of which I was the principal investigator. The genotyping and statistical analysis were performed at the Genome Institute of Singapore. I reviewed and commented on all the statistical analyses. The figures were prepared by Chiea C. Khor. I wrote the first and all subsequent drafts of the manuscript, submitted the manuscript for publication, and responded to all reviewers comments.

Candidate's signature:

Supervisor or senior author's signature to confirm Candidates role:

Title: Susceptibility to highly pathogenic avian influenza A/H5N1 associated with interleukin 1 and TRPM8.

Running title: H5N1 GWAS

Authors

Peter Horby[1†], Chiea C. Khor[2], Annette Fox[1], Le Q. Mai[3], Surakameth Mahasirimongkol[4], Pham Q. Thai[3], Nuanjun Wichukchinda[4], Nguyen T. Yen[3], Sukanya Wattanapokayakit[4], Nusara Satproedprai[4], Luong K. Lan[5], Nguyen T. Hang[1], Vip Viprakasit[6], Sonia Davila[2], Jeremy Farrar[1], Tran T. Hien[1], Tawee Chotpitayasunondh[4], Sarah J. Dunstan[1], Nguyen T. Hien[3], Martin Hibberd[2].

Authors' Affiliations:

1. Oxford University Clinical Research Unit - Wellcome Trust Major Overseas Programme, Vietnam
2. Genome Institute of Singapore, Singapore
3. National Institute for Hygiene and Epidemiology, Vietnam
4. Ministry of Public Health, Bangkok, Thailand
5. Hospital for Tropical Diseases, Ho Chi Minh City,Vietnam
6. Siriraj Hospital, Bangkok, Thailand.

[†] **Corresponding author**

Word count abstract: 179

Word count main text: 2237

Abstract

Background

Marked family clustering of highly pathogenic avian influenza A/H5N1 cases suggest that host factors may determine susceptibility to severe H5N1 disease.

Method and Findings

To identify single nucleotide polymorphisms associated with susceptibility to H5N1 we undertook a genome-wide, case-control discovery study in 51 H5N1 cases from Vietnam and Thailand. Five SNPs in two distinct genetic loci on Chromosome 2 were identified with odds ratios of 3-4 and $P \leq 5 \times 10^{-5}$ (three SNPs at Interleukin-1 gene cluster and two SNPs at *TRPM8*). One SNP at *TRPM8* (rs7560562) reached genome-wide significance at $P \leq 5 \times 10^{-8}$. When genetic variation at both loci was considered simultaneously we found strong evidence of an association with susceptibility to H5N1 (P-trend = 7.80×10^{-12}), with an odds ratio of 31 (95% ci = 9.84 – 98.18) for individuals carrying three copies of the risk alleles compared to wild-type individuals.

Conclusions

We have identified two potential candidates for genetic susceptibility to H5N1 that warrant further exploration and should encourage efforts to gather replication data sets from other countries with human H5N1 cases.

Introduction

Highly pathogenic avian influenza A/H5N1 continues to circulate widely in poultry in many countries and cause human cases with high mortality (~60%). Over 50 cases have been reported so far in 2011 and Cambodia has experienced its highest-ever annual number of cases (eight), with 100% mortality. Meanwhile the rapid evolution and genetic divergence of H5N1 viruses present a continuing risk of the emergence of strains with enhanced ability to infect humans [1]. A full understanding of the determinants of the high virulence of H5N1 is critical for developing predictive, preventive, and therapeutic tools against the public health threat of highly pathogenic influenza viruses.

Around one third of all H5N1 cases have occurred in clusters, and 50 of the 54 H5N1 clusters reported as of March 2009 were comprised entirely of blood relatives [2]. The deaths this year of a mother and son in Cambodia, and of two siblings and their mother in Indonesia demonstrate the continued family clustering of H5N1 cases. Several authors have commented that this familial clustering suggests an important host genetic effect [3-6]. In addition to the familial aggregation of cases, additional evidence that suggests a heritable component to susceptibility to H5N1 includes the small number of clusters of unrelated cases, the occurrence of related cases that are separated in time and space (and therefore not compatible with common source exposure), and the very low risk of infection in some highly exposed groups such as poultry cullers [5].

In mice, genetic background plays a substantial role in susceptibility to severe disease following infection with H5N1 and other influenza viruses, with susceptible mouse strains developing high viral loads, an elevated inflammatory response, and severe lung pathology [7-11]. Studies in cross-bred mouse strains, gene expression studies, and analysis of know biological pathways have resulted in a large number of candidate genes being proposed for susceptibility to severe influenza disease [9,12-14]. However, to date none of these candidate genes have been studied in humans.

As with non-infectious diseases, there is usually a complex interaction between the pathogen, environmental influences, and a range of innate and adaptive host factors [15,16]. Occasionally however, highly penetrant, single locus genotypes have been identified which clearly confer either enhanced susceptibility or protection against infection [17,18]. H5N1 may be a particularly fruitful 'outlier' in which to study host genetic determinants of influenza pathogenicity in humans since H5N1 displays marked familial clustering, an extremely virulent phenotype that does not seem to be associated with old age or pre-existing chronic illnesses, and is not complicated by pre-existing immunity [19]. Therefore the host determinants of H5N1 severity are likely to be less complex than for other types of influenza.

Although extremely challenging due to the small number of cases and the limited availability of DNA from many of these cases, the importance of H5N1 as an exemplar of severe human influenza warrants substantial efforts to determine host genetic factors associated with this infection. We report the first attempt to identify single nucleotide polymorphisms (SNPs) associated with H5N1 disease. A genome-wide, case-control, discovery approach was chosen since there are no prior genetic association studies in humans.

Methods

Ethical approval

The study was reviewed and approved by the ethical review boards of the University of Oxford Centre for Tropical Medicine; the National Institute for Hygiene and Epidemiology, Vietnam; and the Department of Medical Sciences of the Ministry of Public Health, Thailand. ClinicalTrials.gov Identifier: NCT01074736. All living subjects provided written informed consent. Written consent to attempt DNA extraction and genotyping from archived samples of people who had died from H5N1 was obtained from relatives of the deceased person.

Setting

The recruitment of cases and controls was conducted in Vietnam and Thailand between June 2008 and April 2010. Eligible H5N1 cases were identified from the national case lists maintained by the relevant national public health authority in each participating country. These were respectively, the National Institute for Hygiene and Epidemiology in Vietnam and the Ministry of Public Health in Thailand.

Participants

An individual was eligible for inclusion as a case of H5N1 infection if the individual had a clinically compatible illness (defined as respiratory symptoms and, an abnormal chest x-ray or encephalitis), and; influenza A/H5 RNA was identified in a clinical sample by reverse transcription-polymerase chain reaction, or influenza A/H5 was cultured from a clinical sample, or convalescent serum samples had high titres (1:80 or higher) of anti-H5 antibodies by microneutralization assay. The H5N1 case patients were all initially diagnosed between December 2003 and February 2009.

In Thailand ethnically matched unrelated controls were prospectively recruited from the same community as each H5N1 case patient. In Vietnam four different sources of controls were used. 1.)

Ethnically matched unrelated controls prospectively recruited from the same community as H5N1 case patients; 2.) Unrelated participants in a community cohort study of influenza transmission in an area where human H5N1 cases had previously occurred; 3.) Population-based, unmatched cord blood samples; 4.) Relatives of H5N1 case patients.

Community controls were used in order to reduce the chances of significant population stratification and to minimize other unmeasured biases. The cord blood samples represented a large, already genotyped data set that could be readily used as a reference panel to confirm Vietnamese control allele frequencies. The use of population-based, unmatched controls is valid when the studied disease is rare [20]. Relatives of H5N1 cases were recruited in an attempt to provide samples for a family-trio linkage analysis. Since insufficient family-trios were recruited for a family linkage study, non first-degree relatives were included in the analysis as controls.

Laboratory methods

DNA was extracted from samples (whole blood, mouthwash, and archived respiratory tract samples and serum/plasma samples) using standard protocols. Genotyping of the Vietnam and Thailand samples was performed using the Illumina Omni-Express chip according to manufacturer's instructions. Genotypes were assigned in batches for each country separately.

Statistical analysis

Statistical analysis was conducted using PLINK version 1.06. Both SNP-based and sample-based quality control checks were performed. For SNPs, the quality control criteria for inclusion in the analysis were a) Call rate > 95 percent and, b) Minor allele frequency > 5 percent and, c) No significant deviation from HWE as assessed by $P > 10^{-7}$ and, d) only autosomal SNPs. For samples, the quality control criteria for inclusion in the analysis were a) per-sample call rate > 95 percent, b) no first-degree relatives present within sample pairs, c) no significant outlying population ancestry.

Samples were then subjected to biological relationship verification using the principle of variability in allele sharing according to the degree of relationship, in order to identify and exclude 1st degree relatives. Those individuals who showed evidence of cryptic relatedness (possibly either due to duplicated or biologically related samples) were removed before principal component (PC) analysis was performed. PC analysis was undertaken to account for spurious associations resulting from ancestral differences of individual SNPs. Principal component analysis was conducted using multi-dimensional scaling.

Statistical tests of association were performed using the model-free allelic test to maximize statistical power. Meta-analysis was performed using the Cochran-Mantel-Haenszel stratified analysis, as previously described [21]. We also conducted an unguided 2 degrees of freedom genotype test (model free) to verify non-departure from the additive model for each SNP.

To assess the additive effect of the genotype of the two highest scoring SNPs, a two-locus association analysis was performed using a simple 1 d.f. score test (trend test per-copy of the risk allele at either genotype), with the odds ratios for individuals carrying ≥ 1 copy of the risk allele at either locus compared against individuals who are wild-type at both loci (reference odds ratio = 1.0). Only individuals with non-missing genotypes for both SNPs were included in this analysis.

Results

Forty-five Vietnamese H5N1 cases and 178 controls (11 unrelated, ethnically matched community controls for K'hor ethnic minority; 70 unrelated controls from a community cohort; 97 non first-degree relatives), plus an additional 2,018 Vietnamese cord blood controls passed the quality control filters. Six Thailand H5N1 cases and 33 unrelated community controls passed the quality control filters.

A starting number of 733,055 autosomal SNPs were present on the Illumina Omni-Express. In Vietnam, a total of 5,206 SNPs had call rates < 95 percent, and 84,795 SNPs had minor allele frequencies of less than 5 percent and were excluded from analysis. Another 7,882 SNPs showed significant deviation from Hardy-Weinberg equilibrium ($P < 10^{-7}$) and were also excluded from analysis. For Thailand, 6,435 SNPs had call rates < 95 percent, and 74,009 SNPs had minor allele frequencies of less than 5 percent. Another 36 SNPs showed significant deviation from Hardy-Weinberg equilibrium ($P < 10^{-7}$) and were also excluded from analysis. This left a total of 635,172 SNPs common in both Vietnam and Thailand sample collections for downstream statistical analysis. All samples genotyped showed per-sample call rates of > 95%. No duplicates or first-degree relatives were detected from the cases and controls used for analysis. Principal component analysis showed that cases and controls were well matched for both the Vietnam and Thailand sample collections (Figure S1).

Single-locus analysis modeled with allele-based tests of association showed no inflation of test statistics compared to the null distribution (Figure 1). Joint analysis of the Vietnam and Thailand data showed five SNPs in two distinct genetic loci on Chromosome 2 with $P \leq 5 \times 10^{-5}$ (three SNPs at the Interleukin-1 gene cluster, and two SNPs at transient receptor potential channel melastatin 8 [*TRPM8*]), with $P \leq 5 \times 10^{-8}$ at *TRPM8* rs7560562 (Table 1, Figure 2). An additional four SNPs in the *Interleukin-1 alpha* gene had $P \leq 6 \times 10^{-5}$ (Table 1). Review of the genotyping cluster plots in

both Vietnamese and Thailand collections showed distinct genotype clouds for SNPs, confirming their good quality (Figure S2). The allele frequencies of the Vietnamese controls were found to be consistent with the frequencies in a larger set of Vietnamese cord blood controls (Table S1).

The risk of H5N1 jointly conferred by the two lead SNPs at *TRPM8* (rs7560562) and *IL1A* / *IL1B* (rs4849124) was assessed. We observed evidence of association (*P*-trend = 7.80×10^{-12}) over and above that seen with single SNP analysis (Table S2). The risk of H5N1 infection was around 30-fold higher (OR 31.08; 95% ci = 9.84 – 98.18) in individuals carrying the maximal three risk alleles compared to individuals with no risk alleles (we did not observe any individuals carrying all four risk alleles at the two loci).

Discussion

Summary and interpretation

This study has identified five SNP's in two loci on chromosome 2 that are associated with influenza A/H5N1 disease with an odds ratio of 3-4 and a *P* value of ≤ 10^{-5}. Three of these SNPs are in the region encoding for the cytokines Interleukin-1 alpha (IL-1a) and Interleukin-1 beta (IL-1b), whilst two are in the region encoding for an ion channel, TRPM8. One SNP in the TRPM8 region reached genome wide significance [22]. The unavoidably small sample size requires that these findings are interpreted with caution since they may represent false positive rather than true positive associations [23], and even if real, the effect size (odds ratio) is likely to be overestimated [24,25]. Nevertheless, the epidemiology of human H5N1 cases is suggestive of a strong genetic effect and both the identified loci are plausible candidates and potential therapeutic targets. The greatly increased risk associated with a combination of risk alleles from both loci strengthens the confidence in their individual associations and suggests susceptibility is dependent on both genes.

Clinical relevance

Interleukin-1 cytokine gene cluster on chromosome 2

The cytokines IL-1a and IL-1b are key to the innate recognition of invading microorganisms and the early initiation of an inflammatory reaction [26]. High levels of pro-inflammatory cytokines, including IL-1, are found in natural and experimental H5N1 infection, and correlate with disease severity, suggesting that cytokine dysregulation plays a role in the pathogenesis of H5N1 disease [27-32]. Functional-genomics approaches have shown up-regulation of IL-1a/b genes in response to H5N1 infection [14,33], and studies in IL-1 receptor deficient mice show that IL-1 may be critical to the control of extra-pulmonary spread of H5N1 [34,35]. IL1-b has previously been proposed as a candidate gene for study of the host genetic control of response to influenza infection [13].

TRPM8

The *TRPM8* gene codes for a nonselective transmembrane ion channel that is involved in the detection by sensory nerve cells of cold temperatures and of chemicals, such as menthol, which produce a cooling sensation [36,37]. TRPM8 is also found in other tissues including human bronchial epithelial cells, and has recently been linked to respiratory diseases, with ongoing work suggesting that TRPM8 expressed in bronchial epithelial cells also responds to respiratory viruses [38-40]. Activation of the TRPM8 channel in human bronchial epithelial cells results in the expression of inflammatory cytokines, including both IL-1a and IL-1b [41]. TRP channel modulators are under development for a range of clinical conditions [38].

Conclusion

These data represent the first empirical evidence from humans of an association between genetic loci and susceptibility to H5N1. Although the identified SNPs may be false positive signals, the unusual threat posed by H5N1 warrants further exploration of these findings and should encourage efforts to gather replication datasets from other countries that have experienced H5N1 cases. Increases in sample size within the limits possible for a rare disease such as H5N1 may not however substantially increase the confidence in the results of any genome-wide analysis, and fine mapping or functional studies should therefore be pursued.

References

1. (2011) Continued evolution of highly pathogenic avian influenza A (H5N1): updated nomenclature. Influenza Other Respi Viruses.
2. (2010) Summary of human infection with highly pathogenic avian influenza A (H5N1) virus reported to WHO, January 2003-March 2009: cluster-associated cases. Wkly Epidemiol Rec 85: 13-20.
3. Sedyaningsih ER, Isfandari S, Setiawaty V, Rifati L, Harun S, et al. (2007) Epidemiology of cases of H5N1 virus infection in Indonesia, July 2005-June 2006. J Infect Dis 196: 522-527.
4. Kandun IN, Wibisono H, Sedyaningsih ER, Yusharmen, Hadisoedarsuno W, et al. (2006) Three Indonesian clusters of H5N1 virus infection in 2005. N Engl J Med 355: 2186-2194.
5. Horby P, Sudoyo H, Viprakasit V, Fox A, Thai PQ, et al. (2010) What is the evidence of a role for host genetics in susceptibility to influenza A/H5N1? Epidemiol Infect 138: 1550-1558.
6. Aditama TY, Samaan G, Kusriastuti R, Purba WH, Misriyah, et al. (2011) Risk Factors for Cluster Outbreaks of Avian Influenza A H5N1 Infection, Indonesia. Clin Infect Dis 53: 1237-1244.
7. Salomon R, Staeheli P, Kochs G, Yen HL, Franks J, et al. (2007) Mx1 gene protects mice against the highly lethal human H5N1 influenza virus. Cell Cycle 6: 2417-2421.
8. Srivastava B, Blazejewska P, Hessmann M, Bruder D, Geffers R, et al. (2009) Host genetic background strongly influences the response to influenza a virus infections. PLoS ONE 4: e4857.
9. Boon AC, deBeauchamp J, Hollmann A, Luke J, Kotb M, et al. (2009) Host genetic variation affects resistance to infection with a highly pathogenic H5N1 influenza A virus in mice. J Virol 83: 10417-10426.
10. Tumpey TM, Szretter KJ, Van Hoeven N, Katz JM, Kochs G, et al. (2007) The Mx1 gene protects mice against the pandemic 1918 and highly lethal human H5N1 influenza viruses. J Virol 81: 10818-10821.
11. Boon AC, Finkelstein D, Zheng M, Liao G, Allard J, et al. (2011) H5N1 Influenza Virus Pathogenesis in Genetically Diverse Mice Is Mediated at the Level of Viral Load. MBio 2.
12. Trammell RA, Toth LA (2008) Genetic susceptibility and resistance to influenza infection and disease in humans and mice. Expert Rev Mol Diagn 8: 515-529.
13. Zhang L, Katz JM, Gwinn M, Dowling NF, Khoury MJ (2009) Systems-based candidate genes for human response to influenza infection. Infect Genet Evol 9: 1148-1157.

14. Li C, Bankhead A, 3rd, Eisfeld AJ, Hatta Y, Jeng S, et al. (2011) Host Regulatory Network Response to Infection with Highly Pathogenic H5N1 Avian Influenza Virus. J Virol.
15. Hill AV (2006) Aspects of genetic susceptibility to human infectious diseases. Annu Rev Genet 40: 469-486.
16. Vannberg FO, Chapman SJ, Hill AV (2011) Human genetic susceptibility to intracellular pathogens. Immunol Rev 240: 105-116.
17. Newport MJ, Huxley CM, Huston S, Hawrylowicz CM, Oostra BA, et al. (1996) A mutation in the interferon-gamma-receptor gene and susceptibility to mycobacterial infection. N Engl J Med 335: 1941-1949.
18. Shioda T, Nakayama EE (2006) Human genetic polymorphisms affecting HIV-1 diseases. Int J Hematol 84: 12-17.
19. Adisasmito W, Chan PK, Lee N, Oner AF, Gasimov V, et al. (2010) Effectiveness of antiviral treatment in human influenza A(H5N1) infections: analysis of a Global Patient Registry. J Infect Dis 202: 1154-1160.
20. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678.
21. Graham RR, Kozyrev SV, Baechler EC, Reddy MV, Plenge RM, et al. (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat Genet 38: 550-555.
22. Hoggart CJ, Clark TG, De Iorio M, Whittaker JC, Balding DJ (2008) Genome-wide significance for dense SNP and resequencing data. Genet Epidemiol 32: 179-185.
23. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 96: 434-442.
24. Ioannidis JP (2008) Why most discovered true associations are inflated. Epidemiology 19: 640-648.
25. Wakefield J (2008) Reporting and interpretation in genome-wide association studies. Int J Epidemiol 37: 641-653.
26. Dinarello CA (2011) Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. Blood 117: 3720-3732.
27. Lipatov AS, Andreansky S, Webby RJ, Hulse DJ, Rehg JE, et al. (2005) Pathogenesis of Hong Kong H5N1 influenza virus NS gene reassortants in mice: the role of cytokines and B- and T-cell responses. J Gen Virol 86: 1121-1130.

28. de Jong MD, Simmons CP, Thanh TT, Hien VM, Smith GJ, et al. (2006) Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. Nat Med 12: 1203-1207.
29. Cheung CY, Poon LL, Lau AS, Luk W, Lau YL, et al. (2002) Induction of proinflammatory cytokines in human macrophages by influenza A (H5N1) viruses: a mechanism for the unusual severity of human disease? Lancet 360: 1831-1837.
30. Chan MC, Cheung CY, Chui WH, Tsao SW, Nicholls JM, et al. (2005) Proinflammatory cytokine responses induced by influenza A (H5N1) viruses in primary human alveolar and bronchial epithelial cells. Respir Res 6: 135.
31. Seo SH, Hoffmann E, Webster RG (2002) Lethal H5N1 influenza viruses escape host anti-viral cytokine responses. Nat Med 8: 950-954.
32. Perrone LA, Plowden JK, Garcia-Sastre A, Katz JM, Tumpey TM (2008) H5N1 and 1918 pandemic influenza virus infection results in early and excessive infiltration of macrophages and neutrophils in the lungs of mice. PLoS Pathog 4: e1000115.
33. Cilloniz C, Pantin-Jackwood MJ, Ni C, Goodman AG, Peng X, et al. (2010) Lethal dissemination of H5N1 influenza virus is associated with dysregulation of inflammation and lipoxin signaling in a mouse model of infection. J Virol 84: 7613-7624.
34. Szretter KJ, Gangappa S, Lu X, Smith C, Shieh WJ, et al. (2007) Role of host cytokine responses in the pathogenesis of avian H5N1 influenza viruses in mice. J Virol 81: 2736-2744.
35. Perrone LA, Szretter KJ, Katz JM, Mizgerd JP, Tumpey TM (2010) Mice lacking both TNF and IL-1 receptors exhibit reduced lung inflammation and delay in onset of death following infection with a highly virulent H5N1 virus. J Infect Dis 202: 1161-1170.
36. Clapham DE, Julius D, Montell C, Schultz G (2005) International Union of Pharmacology. XLIX. Nomenclature and structure-function relationships of transient receptor potential channels. Pharmacol Rev 57: 427-450.
37. Bautista DM, Siemens J, Glazer JM, Tsuruda PR, Basbaum AI, et al. (2007) The menthol receptor TRPM8 is the principal detector of environmental cold. Nature 448: 204-208.
38. Banner KH, Igney F, Poll C (2011) TRP channels: emerging targets for respiratory disease. Pharmacol Ther 130: 371-384.
39. Sabnis AS, Shadid M, Yost GS, Reilly CA (2008) Human lung epithelial cells express a functional cold-sensing TRPM8 variant. Am J Respir Cell Mol Biol 39: 466-474.

40. Abdullah H., Omar S., Heaney L, L. M, S.L.C (2011) The Effect of Respiratory Viruses Infection on Cough Receptors on Human Sensory Nerve and Human Primary Bronchial Epithelial Cells. XIII International Symposium on Respiratory Viral Infections. Rome.
41. Sabnis AS, Reilly CA, Veranth JM, Yost GS (2008) Increased transcription of cytokine genes in human lung epithelial cells through activation of a TRPM8 variant by cold temperatures. Am J Physiol Lung Cell Mol Physiol 295: L194-200.

Acknowledgments

We are grateful to the patients, their relatives and control subjects for agreeing to participate in this study. We also wish to thank the Ministry of Health of Vietnam and the Ministry of Public Health of Thailand for their continuing support of the Oxford University Clinical Research Unit. Cameron Simmons for access to the Vietnamese cord-blood controls. Salwaluk Panapipat for coordination in Thailand.

Financial Disclosure

This work was supported by the Wellcome Trust UK (grants 081613/Z/06/Z and 077078/Z/05/Z) and the South East Asia Infectious Disease Clinical Research Network (www.seaicrn.org). The SEAICRN is supported by the National Institute Allergy and Infectious Diseases (USA), the Wellcome Trust (UK), and the national authorities of Indonesia, Thailand, and Viet Nam. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived the study and designed the experiments: PH, JF, MH. Performed the experiments: PH, AF, LQM, SM, PQT, NW, NTTY, SW, NS, LKL NTH, VV, SD, JF TTH, TC, SJD, NTH, MH, CCK. Analyzed the data: CCK, MH, PH, SJD. Wrote the first draft of paper: PH, KCC, MH, SJD.

Competing Interests

All authors declare that no competing interests exist.

Table 1. Results of Single Nucleotide Polymorphism analysis

SNP	Alleles*	Function	Vietnam				Thailand				Meta-analysis			
			Freq affected	Freq unaffected	OR	*P* value	Freq affected	Freq unaffected	OR	*P* value	OR	95% CI	*P* value	Phet+
Interleukin 1, alpha														
rs4848300	G/A	Intergenic (3')	0.144	0.0435	3.71	6.22×10^{-4}	0.167	0.0152	13	0.012	4.15	(1.96 - 8.79)	6.53×10^{-5}	0.33
rs17561	A/C	Exonic (missense)	0.144	0.0435	3.71	6.22×10^{-4}	0.167	0.0152	13	0.012	4.15	(1.96 - 8.79)	6.53×10^{-5}	0.33
rs2856837	A/G	Intronic	0.148	0.0435	3.81	4.75×10^{-4}	0.167	0.0152	13	0.012	4.25	(2.01 - 9.02)	4.81×10^{-5}	0.34
rs1800587	A/G	Exonic (UTR-5)	0.144	0.0435	3.71	6.22×10^{-4}	0.167	0.0152	13	0.012	4.15	(1.96 - 8.79)	6.53×10^{-5}	0.33
rs4848304	G/A	Intergenic (5')	0.144	0.0435	3.71	6.22×10^{-4}	0.167	0.0152	13	0.012	4.15	(1.96 - 8.79)	6.53×10^{-5}	0.33
Interleukin 1, beta														
rs4849124	A/G	Intergenic (3')	0.156	0.0466	3.77	3.53×10^{-4}	0.167	0.0152	13	0.012	4.18	(2.01 - 8.67)	3.71×10^{-5}	0.33
rs2853550	A/G	Intergenic (nearGene-3')	0.156	0.0466	3.77	3.53×10^{-4}	0.167	0.0152	13	0.012	4.18	(2.01 - 8.67)	3.71×10^{-5}	0.33
TRPM8														
rs7560562	G/A	Intronic	0.422	0.202	2.89	1.97×10^{-5}	0.667	0.121	14.5	1.67×10^{-5}	3.44	(2.17 - 5.47)	4.75×10^{-8} †	0.03
rs6721761	A/C	Intronic	0.456	0.233	2.76	3.30×10^{-5}	0.667	0.152	11.2	9.77×10^{-5}	3.22	(2.04 - 5.07)	1.81×10^{-7}	0.053

* Reverse strand notation

\+ *P*-value for heterogeneity of effect

† $P = 1.03 \times 10^{-7}$ after continuity correction for significant test of heterogeneity of effect.

Table S1. Minor allele frequencies for SNPs rs7560562 and rs4849124 in Vietnamese H5N1 cases (N = 45), Vietnamese unrelated adult controls (N = 178), and Vietnamese cord blood controls (N = 2,018).

CHR	SNP	Gene	Base position	MAF H5N1 cases	MAF controls	MAF Vietnam cords
2	rs7560562	*TRPM8*	234576953	0.422	0.205	0.22
2	rs4849124	*IL-1*	113293372	0.156	0.045	0.0599

MAF: Minor allele frequency

Table S2. Two-locus association test for *TRPM8* rs7560562 and *IL1A/IL1B* rs4849124.

Number of risk alleles	Number of cord blood controls	Number of controls	Number of cases	Frequency of cord blood controls	Frequency of controls	Frequency of cases	Odds ratio	95% confidence interval of OR	Association test: Trend *P*
Thailand									
0	-	25	0	-	75.8	0	1.00	NA	
1	-	7	3	-	21.2	50	NA	NA	
2	-	1	2	-	3	33.3	NA	NA	
3	-	0	1	-	0	16.7	NA	NA	
Total		**33**	**6**		**100**	**100**			**0.000033**
Vietnam									
0	1092	95	13	54.1	59	28.9	1.00		
1	738	52	16	36.6	32.3	35.6	2.25	(1.08 - 4.70)	
2	174	14	12	8.62	8.7	26.7	6.26	(2.81 - 13.93)	
3	15	0	4	0.74	0	8.9	24.35	(7.11 - 83.39)	
Total	**2019**	**161**	**45**	**100**	**100**	**100**			**7.88×10^{-9} ***
Joint analysis									
0	1092	120	13	54.1	61.86	25.49	1		
1	738	59	19	36.6	30.41	37.25	2.22	(1.09 - 4.52)	
2	174	15	14	8.62	7.73	27.45	6.91	(3.20 - 14.93)	
3	15	0	5	0.74	0	9.80	31.08	(9.84 - 98.18)	
Total	**2019**	**194**	**51**		**100**	**100**			**7.80×10^{-12} ***

NA = Unable to accurately estimate the odds ratios (OR) due to cells with zero counts.
* Trend *P* for all controls including cord blood population controls

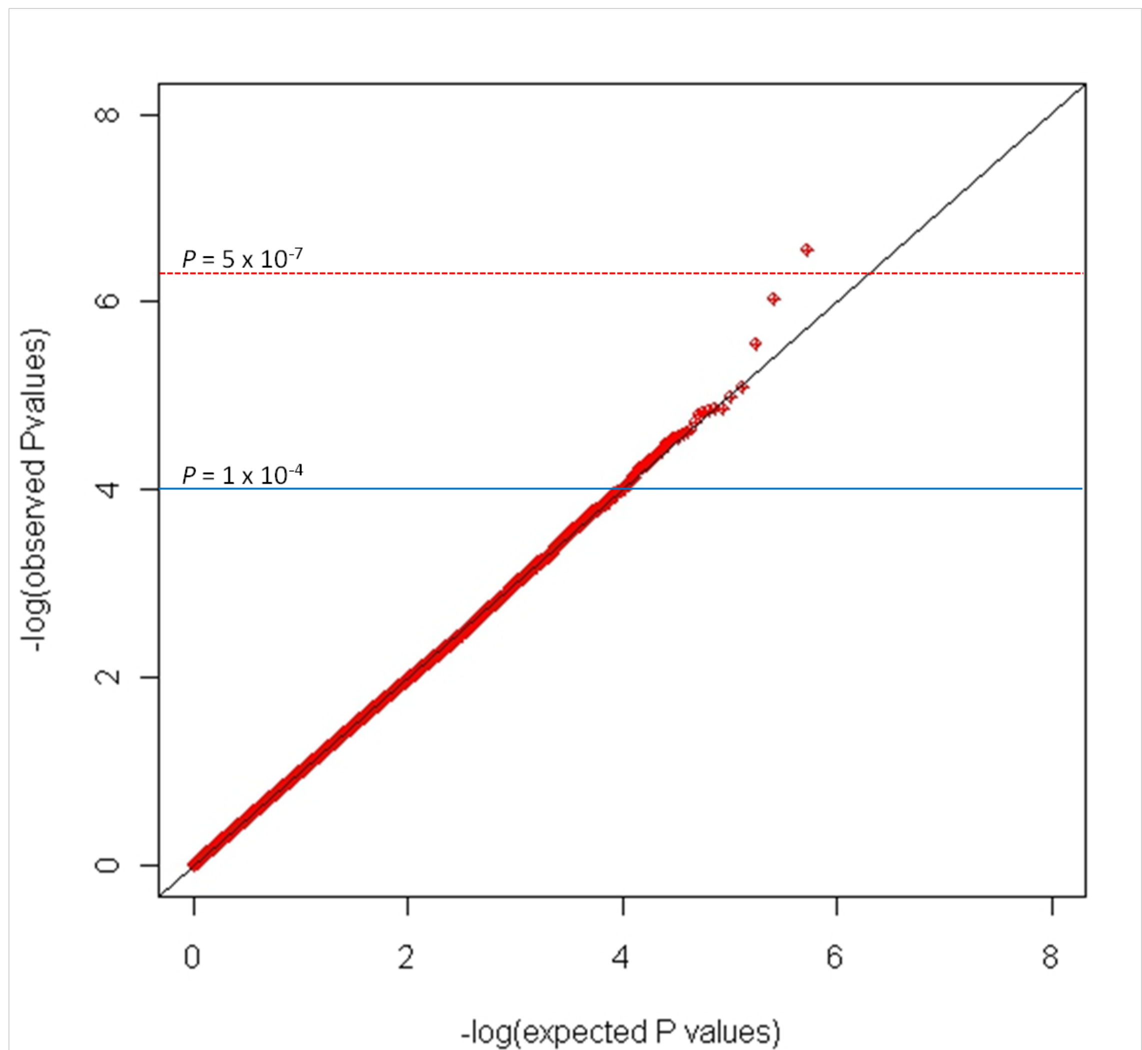


Figure 1. Quantile-quantile plot of the joint analysis of Vietnam and Thailand data.

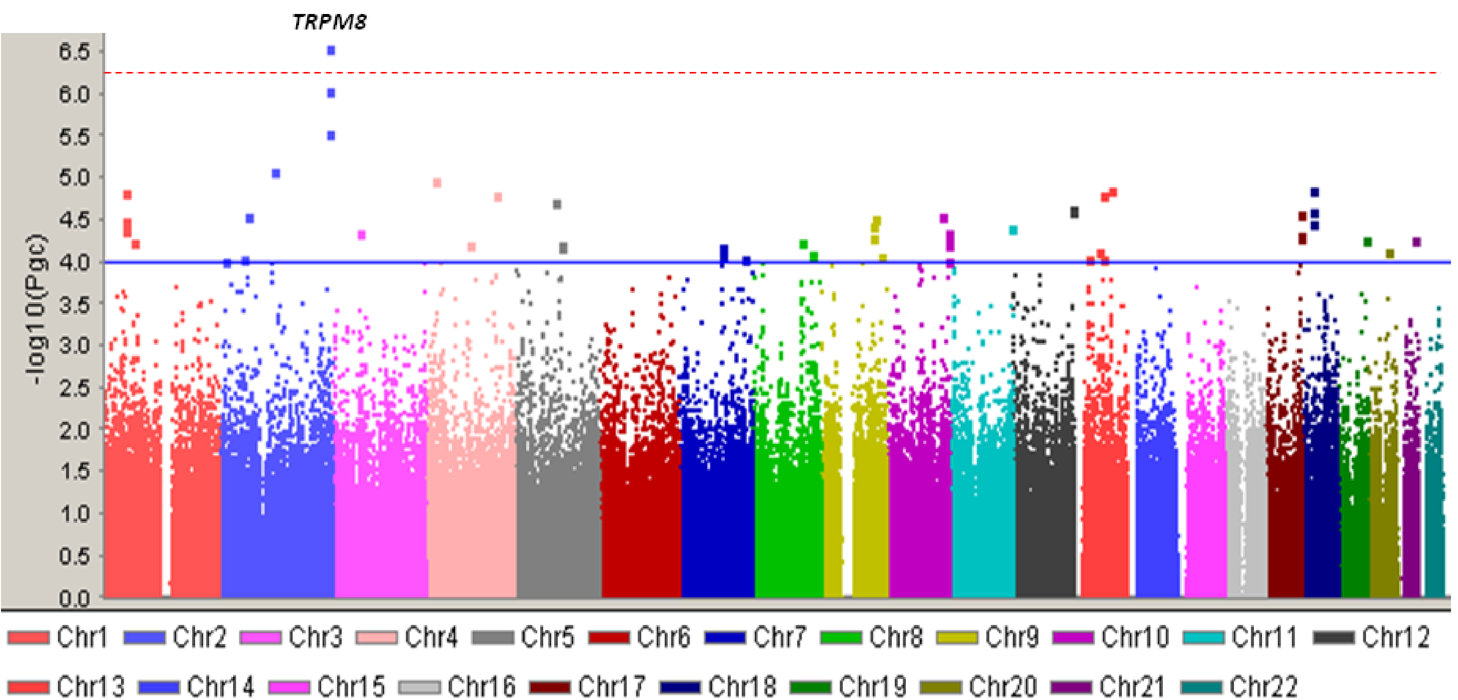


Figure 2. Manhattan plot of the joint analysis of Vietnam and Thailand data.

Manhattan plot using the Cochran Mantel-Haenszel stratified analysis. Vertical axis is -log10 of the P value for each SNP plotted against chromosome position on the horizontal axis. The lower horizontal line denotes P = 5 x 10-4. The upper horizontal line shows P = 5 x 10-7, a threshold indicating highly suggestive statistical association with disease.

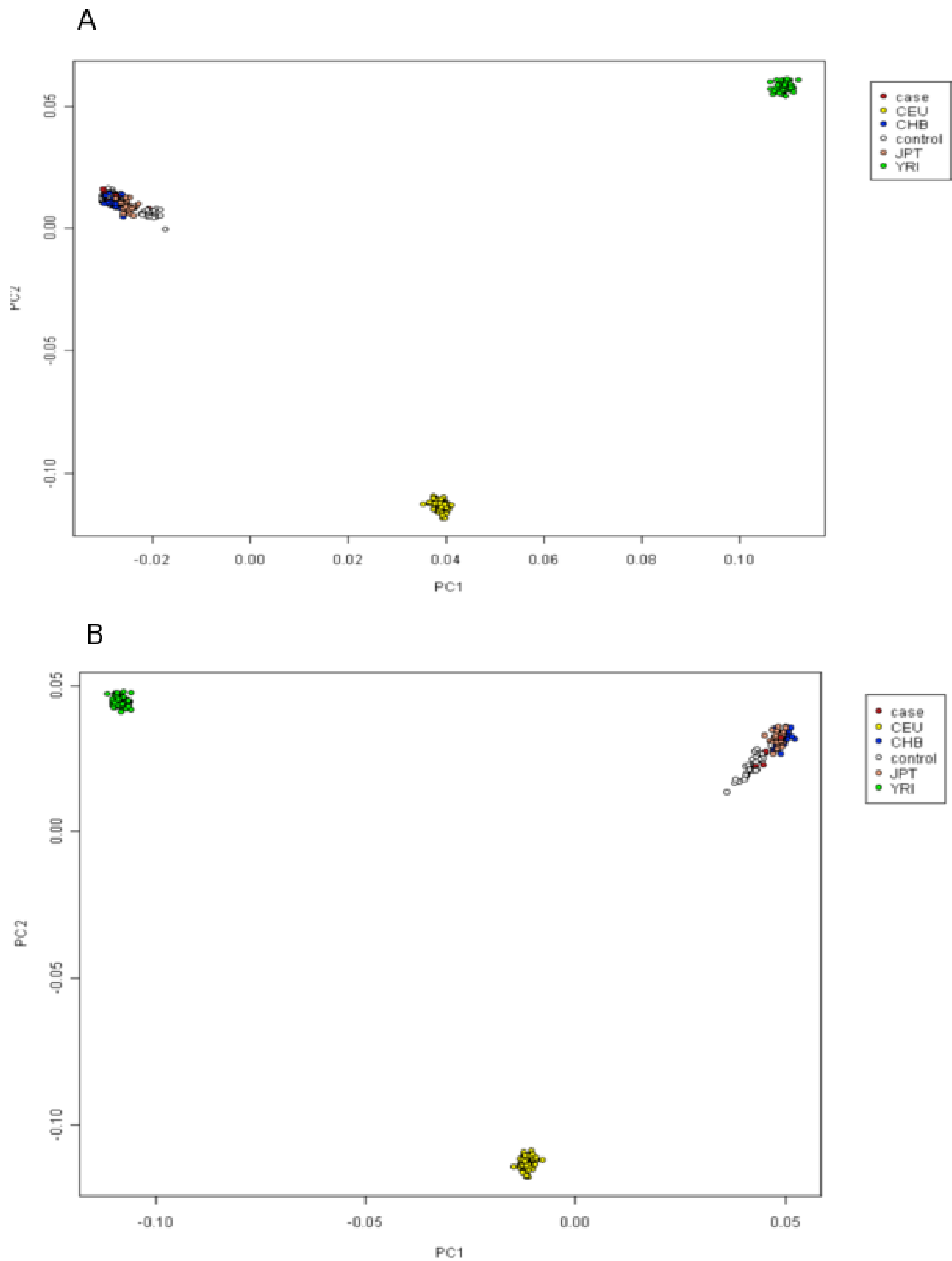


Figure S1. Analysis of genetic ancestry for H5N1 cases and controls against HapMap reference collections.

The top two principal components (PC) of genetic ancestry are plotted. Panel A: Vietnam. Panel B: Thailand.

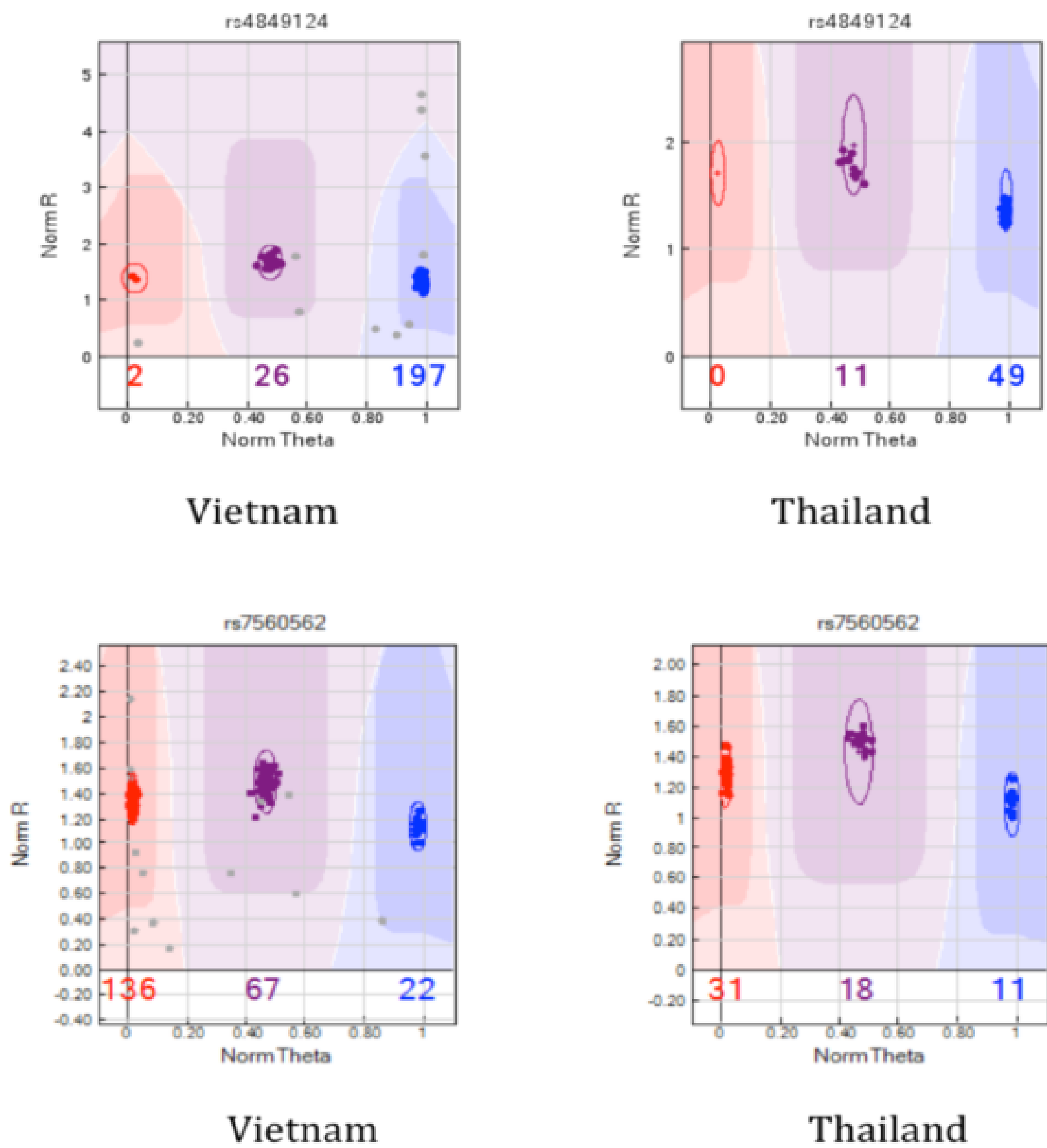


Figure S2. Illumina genotype cluster plots for SNPs rs4849124 and rs7560562 in Vietnam and Thailand H5N1 sample collections.

CHAPTER 13

DISCUSSION *(genetics study)*

13.1 Contribution to knowledge on host genetic susceptibility to H5N1

Although the epidemiology of H5N1 is highly suggestive of a strong familial risk of H5N1 (Aditama et al., 2011; Horby et al., 2010; WHO, 2010) a genome-wide association study for such a rare disease was always going to be a high-risk, high-return exercise. It is therefore both surprising and satisfying that the study has identified two possible candidate loci for susceptibility to H5N1 infection. Nevertheless, the small sample size, the large number of tests of association, and the low prior probability that any individual marker is associated with the outcome, means that the probability of false-positive results is high. However, a recent analysis of discovery and replication GWA studies has found that a threshold for genome-wide statistical significance of $P \leq 5 \times 10^{-8}$ may be too stringent, with 73% of borderline associations ($P > 5 \times 10^{-8}$ and $P \leq 5 \times 10^{-7}$) being successfully replicated (Panagiotou and Ioannidis, 2011). Given that the maximum number of H5N1 cases is unlikely to be sufficient to reach stringent criteria for replication (genome-wide significance of $P \leq 5 \times 10^{-8}$ excluding the discovery data set) and the potential importance of any true-positive association, the findings from the H5N1 GWAS deserve serious exploration.

The Interleukin-1 family is a group of 11 proteins produced by a variety of cell types, of which the two cytokines IL-1α (IL-1F1) and IL-1β (IL-1F2) were the first to be discovered (Barksby et al., 2007). The action of IL-1α and β is mediated through binding to type I IL-1 receptors (IL-1RI) and results in the expression of a wide range of proteins involved in the innate immune response and inflammation (Weber et al., 2010). As discussed in paper 6, IL-1α and IL-1β are highly plausible candidates for involvement in the pathogenesis of severe H5N1 disease since a hyper inflammatory response is characteristic of natural (de Jong et al., 2006; Deng et al., 2008; Peiris et al., 2004; To et al., 2001) and experimental HPAI H5N1 infection (Baskin et al., 2009; Cameron et al., 2008; Cheung et al., 2002; Cilloniz et al., 2010, 2009; Perrone et al., 2008; Sandbulte et al., 2008; Szretter et al., 2007). These loci are especially interesting since blocking IL-1 activity through the use of a recombinant IL1-receptor antagonist (anakinra) has already entered clinical practice for the treatment of chronic inflammatory disorders. However, Boon et al. (2011) have reported that H5N1 pathogenesis in genetically diverse mice is mediated by viral load rather than an aberrant immune response and Salomon et al. (2007) have reported that

mice deficient in key cytokine responses still succumb to H5N1 infection. Therefore the host responses that limit the severity of H5N1 infection may act through control of virus replication rather than immune responses.

The association with the *TRPM8* gene is a novel and unexpected finding, and validates the use of an hypothesis-free genome-wide approach rather than a candidate-gene approach. As discussed in paper 6, there are tantalising hints of biological plausibility, with TRPM8 proteins possessing sialic acid residues (Morenilla-Palao et al., 2009), being expressed on bronchial epithelial cells (Sabnis et al., 2008), responding to respiratory viruses (Abdullah et al., 2011), and activating inflammatory pathways (Sabnis et al., 2008). Forms of TRPM8 that are rich in sialic acid residues are also preferentially localised in lipid rafts within the cell membrane (Morenilla-Palao et al., 2009), and lipid rafts may act as a platform for influenza virus cell entry and budding (Chazal and Gerlier, 2003; Matrosovich et al., 2006; Scheiffele et al., 1997). TRPM8 molecules have received interest from the pharmaceutical sector as therapeutic targets for asthma and chronic obstructive pulmonary disease, and as such several TRP channel modulators are in development (Banner et al., 2011). Avenues for the development of therapeutic products therefore already exist if the association is substantiated.

However, many preliminary gene-disease associations cannot be replicated, the so called "winner's curse" (Kraft, 2008), so prudence is needed when stating the importance of these findings. Ideally the GWA study would have been better powered by including China, Indonesia, and Egypt from the outset, but it proved impossible to engage the serious interest of these countries until the results of the Vietnam and Thailand analysis became available. The rarity of H5N1 cases means that traditional GWA methods and statistical thresholds run a high risk of failing to detect susceptibility loci if they exist. This dilemma is the same for any rare disease and alternative approaches, such as whole-genome or whole-exome sequencing in highly selected patients, will be needed for the discovery of rare gene variants (Cirulli and Goldstein, 2010).

13.2 Further research directions

Substantial further work on host-genetic susceptibility to H5N1 is ongoing. In order to replicate the findings of the initial study, the case-control study has been extended to China, Cambodia, Indonesia, and Egypt. Sixteen case samples and 62 controls have been collected in China and are being genotyped in collaboration with the China Center for Disease Control, whilst DNA samples have been acquired from a mother and child cluster plus controls from Cambodia and are in storage at the Genome Institute of Singapore (GIS). The Indonesian Minister of Health has agreed to conduct the study in Indonesia, with the protocol currently being prepared for ethical review by the National Institute for Health Research and Development. The Ministry of Health of Egypt have also agreed

to participate and are currently preparing the protocol for ethical review. Sequencing of the full *IL-1*α, *IL-1*β, and *TRPM8* genes in the H5 cases already genotyped is ongoing at GIS. We are collaborating with Richard Webby at St. Jude Children's Research Hospital to conduct titration studies with H5N1/HK/213/03 in B6 mice with depletion using anti-IL-1α and anti-IL-1β antibodies and a TRMP8 antagonist compound. *IL-1*α, *IL-1*β, and *TRPM8* knock-out mice are also being bred for knock-out studies.

Chapter 14

Concluding remarks

In this thesis I have presented work that aims to provide insights into the epidemiology of influenza. The cohort study has added to the body of knowledge on the burden and epidemiology of inter-pandemic and pandemic influenza in the tropics by providing community level data that were previously absent, whilst the genome-wide association study offers a potential explanation for the observed clustering of H5N1 cases. The cohort data have value for planning influenza control programs in Southeast Asia but also for formulating and examining hypotheses about the distribution and nature of the forces driving the evolution of influenza. The GWA study has provided the first direct evidence of genetic loci associated with susceptibility to H5N1 and opens new avenues of research to test these findings and their relevance to the pathogenesis of H5N1 and other types of influenza. It is therefore hoped that this work constitutes research of both 'light' and 'fruit' (Bacon, 1620): research that illuminates fundamental processes but which is also useful for the practical task of controlling influenza.

The research has also stimulated ideas for a future research agenda. The publication of the results of the H5N1 GWA study along with results in press of an association between interferon-inducible transmembrane protein and severe influenza (personal communication Paul Kellam) will build momentum for further studies of host genetics and influenza. Whilst gene sequencing and functional studies are critical, there is also a necessity to present robust statistical evidence of an association between putative loci and disease risk, and to provide reliable estimates of the proportion of risk attributable to those loci. This will require the pooling of data since no individual research group has sufficient severe influenza cases to conduct an adequately powered GWA study. The pooling of data across influenza subtypes/types is also important since host and pathogen genomes co-evolve and interact to determine pathogenesis, and studies of host genetic factors influencing the risk of severe influenza disease should therefore be conditioned on pathogen genetics (Lambrechts, 2010). The challenge ahead is to unite centres working in isolation on small numbers of patients to crystallise a consortium with sufficient patients and expertise to conduct authoritative studies on the host determinants of severe influenza.

Whilst knowledge and understanding of influenza evolution and epidemiology has increased in recent years, it is probably fair to say that over the period that this research was conducted the control of influenza has not improved substantially. Public health interventions such as school-closures and hand-hygiene are short-term holding measures that are unlikely to have a significant or sustainable impact on the burden of influenza, whilst

influenza vaccination remains available to only a small minority of the global population at risk. A good starting point may be the rehabilitation or rebranding of influenza as an 'everyday' cause of disease and death. Pneumonia remains the biggest killer of children worldwide and it is likely that influenza causes 5-10% of these deaths and may contribute to a much larger proportion(Nair et al., 2011). A potential role in pneumonia-prevention programs of the immunisation of pregnant women and children against influenza deserves much greater attention. Whilst a broadly protective vaccine that does not require annual reformulation and delivery is an admirable aspiration, the epidemiology of influenza suggests that this is a formidable challenge. Currently available influenza vaccines, particularly live-attenuated vaccines, may however be practical tools for influenza control in developing countries. In addition, vaccine probe studies[1], as have been conducted for *Haemophilus influenzae* type b and pneumococcus, may be a particularly fruitful approach for gaining insights into the preventable fraction of influenza-associated disease in children in developing countries (Gessner et al., 2005).

A better understanding is also required of the critical components of influenza epidemiology that should be targeted for surveillance and control. Significant gaps remain in our understanding of the role of East and Southeast Asia in the global epidemiology of influenza, and although the number of HA gene sequences available from this region has increased greatly in recent years, sampling remains patchy. This leaves substantial room for phylogenetic and phylogeographic patterns to be biased by missing data and the selective choice of virus subsets for genotyping. The increasing awareness of the need to look beyond HA when considering viral evolutionary dynamics (Bhatt et al., 2011) means that full-genome sequencing of systematically collected and chronicled isolates is required. Population-based sero-surveillance is an under-developed area that also warrants attention. Sero-surveillance offers the tantalising prospect of being able to look 'upstream' to measure patterns and changes in the driver of virus evolution (population immunity), but is hampered by the lack of high-throughput and replicable antibody assays. The HI assay is sensitive to experimental conditions (e.g. red cell type and operator) and because of this experimental variability and the labour needed to process large numbers of samples in duplicate, is not well-utilised as a public health surveillance tool (Katz et al., 2011; Stephenson et al., 2007). What is more, increases in HI titres of only one dilution have traditionally been ignored as they may represent assay error, but it seems likely that such increases frequently represent a real response to infection. The application of a threshold of a four-fold rise to define infection may therefore provide an incomplete picture of the overall pattern and intensity of influenza infection, and the serological definition of infection for surveillance purposes should be reassessed. In addition, efforts are also needed to identify whether the antibody status of individuals can be aggregated to provide a meaningful measure of population immunity that has a quantifiable relationship to selection

[1] A vaccine trial designed also to 'probe' the vaccine-preventable burden of disease

pressures. These issues represent a long but potentially navigable pathway towards developing sero-surveillance as a valuable tool for monitoring and predicting global influenza epidemiology.

To conclude, the work presented in this thesis advances our knowledge of influenza epidemiology and also helps to define elements of a future research agenda, but it is undeniable that ample scope remains for deepening our understanding of influenza and translating this into practical tools for control that benefit everyone.

References

Abdullah, H., S. Cosby, L. Heaney, and L. McGarvey (2011). The effect of rhinovirus infection on cough receptors on human sensory nerve and human primary bronchial epithelial cells. *Thorax 66(S4)*, A57.

Aditama, T. Y., G. Samaan, R. Kusriastuti, W. H. Purba, Misriyah, H. Santoso, A. Bratasena, A. Maruf, E. Sariwati, V. Setiawaty, A. R. Cook, M. S. Clements, K. Lokuge, P. M. Kelly, and I. N. Kandun (2011). Risk factors for cluster outbreaks of avian influenza A H5N1 infection, Indonesia. *Clin Infect Dis 53*(12), 1237–1244.

Alexander, D. J. (2007). An overview of the epidemiology of avian influenza. *Vaccine 25*(30), 5637–5644.

Bacon, F. (1620). *The New Organon or True Directions Concerning the Interpretation of Nature.*

Badger, G. F., J. H. Dingle, A. E. Feller, R. G. Hodges, J. Jordan, W. S., and J. Rammelkamp, C. H. (1953a). A study of illness in a group of Cleveland families. II. Incidence of the common respiratory diseases. *Am J Hyg 58*(1), 31–40.

Badger, G. F., J. H. Dingle, A. E. Feller, R. G. Hodges, J. Jordan, W. S., and J. Rammelkamp, C. H. (1953b). A study of illness in a group of Cleveland families. IV. The spread of respiratory infections within the home. *Am J Hyg 58*(2), 174–8.

Bahl, J., M. I. Nelson, K. H. Chan, R. Chen, D. Vijaykrishna, R. A. Halpin, T. B. Stockwell, X. Lin, D. E. Wentworth, E. Ghedin, Y. Guan, J. S. M. Peiris, S. Riley, A. Rambaut, E. C. Holmes, and G. J. D. Smith (2011). Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans. *Proc Natl Acad Sci U S A 108*(48), 19359–19364.

Banner, K. H., F. Igney, and C. Poll (2011). TRP channels: emerging targets for respiratory disease. *Pharmacol Ther 130*(3), 371–84.

Barksby, H. E., S. R. Lea, P. M. Preshaw, and J. J. Taylor (2007). The expanding family of interleukin-1 cytokines and their role in destructive inflammatory disorders. *Clin Exp Immunol 149*(2), 217–225.

Baskin, C. R., H. Bielefeldt-Ohmann, T. M. Tumpey, P. J. Sabourin, J. P. Long, A. Garca-

Sastre, A.-E. Tolnay, R. Albrecht, J. A. Pyles, P. H. Olson, L. D. Aicher, E. R. Rosenzweig, K. Murali-Krishna, E. A. Clark, M. S. Kotur, J. L. Fornek, S. Proll, R. E. Palermo, C. L. Sabourin, and M. G. Katze (2009). Early and sustained innate immune response defines pathology and death in nonhuman primates infected by highly pathogenic influenza virus. *Proc Natl Acad Sci U S A 106*(9), 3455–3460.

Basler, C. F. and P. V. Aguilar (2008). Progress in identifying virulence determinants of the 1918 H1N1 and the Southeast Asian H5N1 influenza A viruses. *Antiviral Res 79*(3), 166–78.

Bedford, T., S. Cobey, and M. Pascual (2011). Strength and tempo of selection revealed in viral gene genealogies. *BMC Evol Biol 11*, 220.

Bhatt, S., E. C. Holmes, and O. G. Pybus (2011). The genomic rate of molecular adaptation of the human influenza A virus. *Mol Biol Evol 28*(9), 2443–2451.

Blackwell, J. M., S. E. Jamieson, and D. Burgner (2009). HLA and infectious diseases. *Clin Microbiol Rev 22*(2), 370–85, Table of Contents.

Boelle, P-Y.lle, P.-Y., S. Ansart, A. Cori, and A.-J. Valleron (2011). Transmission parameters of the A/H1N1 (2009) influenza virus pandemic: a review. *Influenza Other Respi Viruses 5*(5), 306–316.

Boni, M. F., B. H. Manh, P. Q. Thai, J. Farrar, T. T. Hien, N. T. Hien, N. Van Kinh, and P. Horby (2009). Modelling the progression of pandemic influenza A (H1N1) in Vietnam and the opportunities for reassortment with other influenza viruses. *BMC Med 7*, 43.

Boon, A. C. M., D. Finkelstein, M. Zheng, G. Liao, J. Allard, K. Klumpp, R. Webster, G. Peltz, and R. J. Webby (2011). H5N1 influenza virus pathogenesis in genetically diverse mice is mediated at the level of viral load. *MBio 2*(5), e00171–11.

Brankston, G., L. Gitterman, Z. Hirji, C. Lemieux, and M. Gardam (2007). Transmission of influenza A in human beings. *Lancet Infect Dis 7*(4), 257–265.

Brooks, W. A., D. Goswami, M. Rahman, K. Nahar, A. M. Fry, A. Balish, N. Iftekharuddin, T. Azim, X. Xu, A. Klimov, J. Bresee, C. Bridges, and S. Luby (2010, Mar). Influenza is a major contributor to childhood pneumonia in a tropical developing country. *Pediatr Infect Dis J 29*(3), 216–221.

Cameron, C. M., M. J. Cameron, J. F. Bermejo-Martin, L. Ran, L. Xu, P. V. Turner, R. Ran, A. Danesh, Y. Fang, P.-K. M. Chan, N. Mytle, T. J. Sullivan, T. L. Collins, M. G. Johnson, J. C. Medina, T. Rowe, and D. J. Kelvin (2008). Gene expression analysis of host innate immune responses during Lethal H5N1 infection in ferrets. *J Virol 82*(22), 11308–11317.

Cannell, J. J., M. Zasloff, C. F. Garland, R. Scragg, and E. Giovannucci (2008). On the epidemiology of influenza. *Virol J 5*, 29.

Cauchemez, S., F. Carrat, C. Viboud, A. J. Valleron, and P. Y. Boelle (2004). A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat Med 23*(22), 3469–87.

Cauchemez, S., C. A. Donnelly, C. Reed, A. C. Ghani, C. Fraser, C. K. Kent, L. Finelli, and N. M. Ferguson (2009). Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *N Engl J Med 361*(27), 2619–2627.

Cauchemez, S., N. M. Ferguson, C. Wachtel, A. Tegnell, G. Saour, B. Duncan, and A. Nicoll (2009). Closure of schools during an influenza pandemic. *Lancet Infect Dis 9*(8), 473–81.

CDC (2010). Estimates of deaths associated with seasonal influenza — United States, 1976-2007. *MMWR Morb Mortal Wkly Rep 59*(33), 1057–1062.

Chazal, N. and D. Gerlier (2003). Virus entry, assembly, budding, and membrane rafts. *Microbiol Mol Biol Rev 67*(2), 226–37, table of contents.

Cheung, C. Y., L. L. Poon, A. S. Lau, W. Luk, Y. L. Lau, K. F. Shortridge, S. Gordon, Y. Guan, and J. S. Peiris (2002). Induction of proinflammatory cytokines in human macrophages by influenza A (H5N1) viruses: a mechanism for the unusual severity of human disease? *Lancet 360*(9348), 1831–7.

Chew, F. T., S. Doraisingham, A. E. Ling, G. Kumarasinghe, and B. W. Lee (1998). Seasonal trends of viral respiratory tract infections in the tropics. *Epidemiol Infect 121*(1), 121–8.

Chiu, S. S., Y. L. Lau, K. H. Chan, W. H. S. Wong, and J. S. M. Peiris (2002). Influenza-related hospitalizations among children in Hong Kong. *N Engl J Med 347*(26), 2097–2103.

Chowell, G., C. Viboud, L. Simonsen, M. Miller, and W. J. Alonso (2010). The reproduction number of seasonal influenza epidemics in Brazil, 1996-2006. *Proc Biol Sci 277*(1689), 1857–66.

Chutinimitkul, S., S. Herfst, J. Steel, A. C. Lowen, J. Ye, D. van Riel, E. J. A. Schrauwen, T. M. Bestebroer, B. Koel, D. F. Burke, K. H. Sutherland-Cash, C. S. Whittleston, C. A. Russell, D. J. Wales, D. J. Smith, M. Jonges, A. Meijer, M. Koopmans, G. F. Rimmelzwaan, T. Kuiken, A. D. M. E. Osterhaus, A. Garca-Sastre, D. R. Perez, and R. A. M. Fouchier (2010). Virulence-associated substitution D222G in the hemagglutinin of 2009 pandemic influenza A(H1N1) virus affects receptor binding. *J Virol 84*(22), 11802–11813.

Cilloniz, C., M. J. Pantin-Jackwood, C. Ni, A. G. Goodman, X. Peng, S. C. Proll, V. S. Carter, E. R. Rosenzweig, K. J. Szretter, J. M. Katz, M. J. Korth, D. E. Swayne, T. M. Tumpey, and M. G. Katze (2010). Lethal dissemination of H5N1 influenza virus is associated with dysregulation of inflammation and lipoxin signaling in a mouse model

of infection. *J Virol 84*(15), 7613–24.

Cilloniz, C., K. Shinya, X. Peng, M. J. Korth, S. C. Proll, L. D. Aicher, V. S. Carter, J. H. Chang, D. Kobasa, F. Feldmann, J. E. Strong, H. Feldmann, Y. Kawaoka, and M. G. Katze (2009). Lethal influenza virus infection in macaques is associated with early dysregulation of inflammatory related genes. *PLoS Pathog 5*(10), e1000604.

Cirulli, E. T. and D. B. Goldstein (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet 11*(6), 415–425.

Clague, B., S. Chamany, C. Burapat, Y. Wannachaiwong, J. M. Simmerman, S. F. Dowell, and S. J. Olsen (2006). A household survey to assess the burden of influenza in rural Thailand. *Southeast Asian J Trop Med Public Health 37*(3), 488–93.

Cohen, C., L. Simonsen, J.-W. Kang, M. Miller, J. McAnerney, L. Blumberg, B. Schoub, S. A. Madhi, and C. Viboud (2010). Elevated influenza-related excess mortality in South African elderly individuals, 1998-2005. *Clin Infect Dis 51*(12), 1362–1369.

Corti, D., J. Voss, S. J. Gamblin, G. Codoni, A. Macagno, D. Jarrossay, S. G. Vachieri, D. Pinna, A. Minola, F. Vanzetta, C. Silacci, B. M. Fernandez-Rodriguez, G. Agatic, S. Bianchi, I. Giacchetto-Sasselli, L. Calder, F. Sallusto, P. Collins, L. F. Haire, N. Temperton, J. P. M. Langedijk, J. J. Skehel, and A. Lanzavecchia (2011). A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science 333*(6044), 850–856.

Cowling, B. J., K. H. Chan, V. J. Fang, L. L. H. Lau, H. C. So, R. O. P. Fung, E. S. K. Ma, A. S. K. Kwong, C.-W. Chan, W. W. S. Tsui, H.-Y. Ngai, D. W. S. Chu, P. W. Y. Lee, M.-C. Chiu, G. M. Leung, and J. S. M. Peiris (2010). Comparative epidemiology of pandemic and seasonal influenza A in households. *N Engl J Med 362*(23), 2175–2184.

Cowling, B. J., V. J. Fang, S. Riley, J. S. Malik Peiris, and G. M. Leung (2009). Estimation of the serial interval of influenza. *Epidemiology 20*(3), 344–347.

de Jong, M. D., C. P. Simmons, T. T. Thanh, V. M. Hien, G. J. Smith, T. N. Chau, D. M. Hoang, N. V. Chau, T. H. Khanh, V. C. Dong, P. T. Qui, B. V. Cam, Q. Ha do, Y. Guan, J. S. Peiris, N. T. Chinh, T. T. Hien, and J. Farrar (2006). Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nat Med 12*(10), 1203–7.

Deng, R., M. Lu, C. Korteweg, Z. Gao, M. A. McNutt, J. Ye, T. Zhang, and J. Gu (2008). Distinctly different expression of cytokines and chemokines in the lungs of two H5N1 avian influenza patients. *J Pathol 216*(3), 328–336.

Edmunds, W. J., C. J. O'Callaghan, and D. J. Nokes (1997). Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc Biol Sci 264*(1384), 949–957.

Ferdinands, J. M., A. M. Denison, N. F. Dowling, H. A. Jost, M. L. Gwinn, L. Liu, S. R. Zaki, and D. K. Shay (2011). A Pilot Study of Host Genetic Variants Associated with Influenza-associated Deaths among Children and Young Adults1. *Emerg Infect Dis 17*(12), 2294–2302.

Ferguson, N. M., D. A. Cummings, C. Fraser, J. C. Cajka, P. C. Cooley, and D. S. Burke (2006). Strategies for mitigating an influenza pandemic. *Nature 442*(7101), 448–52.

Ferguson, N. M., A. P. Galvani, and R. M. Bush (2003). Ecological and immunological determinants of influenza evolution. *Nature 422*(6930), 428–33.

Fouchier, R. A. M., P. M. Schneeberger, F. W. Rozendaal, J. M. Broekman, S. A. G. Kemink, V. Munster, T. Kuiken, G. F. Rimmelzwaan, M. Schutten, G. J. J. Van Doornum, G. Koch, A. Bosman, M. Koopmans, and A. D. M. E. Osterhaus (2004). Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc Natl Acad Sci U S A 101*(5), 1356–1361.

Fox, J. P. (1974). Family-based epidemiologic studies. The second Wade Hampton Frost Lecture. *Am J Epidemiol 99*(3), 165–179.

Fox, J. P., M. K. Cooney, C. E. Hall, and H. M. Foy (1982). Influenzavirus infections in Seattle families, 1975-1979. II. Pattern of infection in invaded households and relation of age and prior antibody to occurrence of infection and related illness. *Am J Epidemiol 116*(2), 228–42.

Fox, J. P., C. E. Hall, M. K. Cooney, and H. M. Foy (1982). Influenzavirus infections in Seattle families, 1975-1979. I. Study design, methods and the occurrence of infections by time and age. *Am J Epidemiol 116*(2), 212–27.

Fox, J. P., C. E. Hall, M. K. Cooney, R. E. Luce, and R. A. Kronmal (1972). The Seattle virus watch. II. Objectives, study population and its observation, data processing and summary of illnesses. *Am J Epidemiol 96*(4), 270–85.

Frank, A. L., L. H. Taber, W. P. Glezen, E. A. Geyer, S. McIlwain, and A. Paredes (1983). Influenza B virus infections in the community and the family. The epidemics of 1976-1977 and 1979-1980 in Houston, Texas. *Am J Epidemiol 118*(3), 313–25.

Frank, A. L., L. H. Taber, and J. M. Wells (1985). Comparison of infection rates and severity of illness for influenza A subtypes H1N1 and H3N2. *J Infect Dis 151*(1), 73–80.

Fukuyama, S. and Y. Kawaoka (2011). The pathogenesis of influenza virus infections: the contributions of virus and host factors. *Curr Opin Immunol 23*(4), 481–486.

Gamblin, S. J. and J. J. Skehel (2010). Influenza hemagglutinin and neuraminidase membrane glycoproteins. *J Biol Chem 285*(37), 28403–28409.

Gessner, B. D., A. Sutanto, M. Linehan, I. G. G. Djelantik, T. Fletcher, I. K. Gerudug, Ingerani, D. Mercer, V. Moniaga, L. H. Moulton, L. H. Moulton, K. Mulholland, C. Nel-

son, S. Soemohardjo, M. Steinhoff, A. Widjaya, P. Stoeckel, J. Maynard, and S. Arjoso (2005). Incidences of vaccine-preventable Haemophilus influenzae type b pneumonia and meningitis in Indonesian children: hamlet-randomised vaccine-probe trial. *Lancet 365*(9453), 43–52.

Glaser, L., G. Conenello, J. Paulson, and P. Palese (2007). Effective replication of human influenza viruses in mice lacking a major alpha2,6 sialyltransferase. *Virus Res 126*(1-2), 9–18.

Glass, K. and B. Barnes (2007). How much would closing schools reduce transmission during an influenza pandemic? *Epidemiology 18*(5), 623–8.

Glass, L. M. and R. J. Glass (2008). Social contact networks for the spread of pandemic influenza in children and teenagers. *BMC Public Health 8*, 61.

Glezen, W. P. and R. B. Couch (1978). Interpandemic influenza in the Houston area, 1974-76. *N Engl J Med 298*(11), 587–92.

Goeyvaerts, N., N. Hens, B. Ogunjimi, M. Aerts, Z. Shkedy, P. Van Damme, and P. Beutels (2010). Estimating infectious disease parameters from data on social contacts and serological status. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 59*, 255–277.

GSO (2010). The 2009 Population and Housing Census: Completed results. Technical report, Vietnam General Statistics Office.

Guan, Y., D. Vijaykrishna, J. Bahl, H. Zhu, J. Wang, and G. J. D. Smith (2010). The emergence of pandemic influenza viruses. *Protein Cell 1*(1), 9–13.

Hall, C. E., C. D. Brandt, T. E. Frothingham, I. Spigland, M. K. Cooney, and J. P. Fox (1971). The virus watch program: a continuing surveillance of viral infections in metropolitan New York families. IX. A comparison of infections with several respiratory pathogens in New York and New Orleans families. *Am J Epidemiol 94*(4), 367–85.

Hall, C. E., M. K. Cooney, and J. P. Fox (1973). The Seattle virus watch. IV. Comparative epidemiologic observations of infections with influenza A and B viruses, 1965-1969, in families with young children. *Am J Epidemiol 98*(5), 365–80.

Hanshaoworakul, W., J. M. Simmerman, U. Narueponjirakul, W. Sanasuttipun, V. Shinde, S. Kaewchana, D. Areechokechai, J. Levy, and K. Ungchusak (2009). Severe human influenza infections in Thailand: oseltamivir treatment and risk factors for fatal outcome. *PLoS One 4*(6), e6051.

Hoa, L. K., L. V. Hiep, and L. V. Be (2011). Development of pandemic influenza vaccine production capacity in Viet Nam. *Vaccine 29 Suppl 1*, A34–A36.

Hope-Simpson, R. E. (1984). Age and secular distributions of virus-proven influenza patients in successive epidemics 1961-1976 in Cirencester: epidemiological significance

discussed. *J Hyg (Lond) 92*(3), 303–36.

Horby, P., L. Mai, A. Fox, P. Q. Thai, N. Yen, L. Thanh, N. Hang, T. Duong, D. Thoang, J. Farrar, M. Wolbers, and N. Hien (2012). The epidemiology of interpandemic and pandemic influenza in Vietnam, 20072010: the Ha Nam household cohort study. *American Journal of Epidemiology In Press*.

Horby, P., N. Nhu, S. J. Dunstan, and J. Baillie. The role of host genetics in susceptibility to influenza: a systematic review. *PLoS One In Press*.

Horby, P., Q. T. Pham, N. Hens, T. T. Y. Nguyen, Q. M. Le, D. T. Dang, M. L. Nguyen, T. H. Nguyen, N. Alexander, W. J. Edmunds, N. D. Tran, A. Fox, and T. H. Nguyen (2011). Social contact patterns in Vietnam and implications for the control of infectious diseases. *PLoS One 6*(2), e16965.

Horby, P., H. Sudoyo, V. Viprakasit, A. Fox, P. Q. Thai, H. Yu, S. Davila, M. Hibberd, S. J. Dunstan, Y. Monteerarat, J. J. Farrar, S. Marzuki, and N. T. Hien (2010). What is the evidence of a role for host genetics in susceptibility to influenza A/H5N1? *Epidemiol Infect 138*(11), 1550–1558.

Hudson, K. L. (2011). Genomics, health care, and society. *N Engl J Med 365*(11), 1033–1041.

Hurwitz, E. S., M. Haber, A. Chang, T. Shope, S. Teo, M. Ginsberg, N. Waecker, and N. J. Cox (2000). Effectiveness of influenza vaccination of day care children in reducing influenza-related morbidity among household contacts. *JAMA 284*(13), 1677–82.

Ioannidis, J. P. A. (2007). Non-replication and inconsistency in the genome-wide association setting. *Hum Hered 64*(4), 203–213.

Jadhao, S. J., D. C. Nguyen, T. M. Uyeki, M. Shaw, T. Maines, T. Rowe, C. Smith, L. P. T. Huynh, H. K. Nghiem, D. H. T. Nguyen, H. K. L. Nguyen, H. H. T. Nguyen, L. T. Hoang, T. Nguyen, L. S. Phuong, A. Klimov, T. M. Tumpey, N. J. Cox, R. O. Donis, Y. Matsuoka, and J. M. Katz (2009). Genetic analysis of avian influenza A viruses isolated from domestic waterfowl in live-bird markets of Hanoi, Vietnam, preceding fatal H5N1 human infections in 2004. *Arch Virol 154*(8), 1249–1261.

Jamieson, D. J., R. N. Theiler, and S. A. Rasmussen (2006, Nov). Emerging infections and pregnancy. *Emerg Infect Dis 12*(11), 1638–1643.

Jordan, W. S., J., G. F. Badger, and J. H. Dingle (1958). A study of illness in a group of Cleveland families. XVI. The epidemiology of influenza, 1948-1953. *Am J Hyg 68*(2), 169–89.

Jordan, W. S., J., J. Denny, F. W., G. F. Badger, C. Curtiss, J. H. Dingle, R. Oseasohn, and D. A. Stevens (1958). A study of illness in a group of Cleveland families. XVII. The occurrence of Asian influenza. *Am J Hyg 68*(2), 190–212.

Katz, J. M., K. Hancock, and X. Xu (2011, Jun). Serologic assays for influenza surveillance, diagnosis and vaccine evaluation. *Expert Rev Anti Infect Ther 9*(6), 669–683.

Keeling, M. J. and L. Danon (2009). Mathematical modelling of infectious diseases. *Br Med Bull 92*, 33–42.

Kraft, P. (2008, Sep). Curses–winner's and otherwise–in genetic epidemiology. *Epidemiology 19*(5), 649–51; discussion 657–8.

Kuchipudi, S. V., R. Nelli, G. A. White, M. Bain, K. C. Chang, and S. Dunham (2009). Differences in influenza virus receptors in chickens and ducks: Implications for interspecies transmission. *J Mol Genet Med 3*(1), 143–151.

Kuiken, T., E. C. Holmes, J. McCauley, G. F. Rimmelzwaan, C. S. Williams, and B. T. Grenfell (2006). Host species barriers to influenza virus infections. *Science 312*(5772), 394–7.

Lambrechts, L. (2010). Dissecting the genetic architecture of host-pathogen specificity. *PLoS Pathog 6*(8), e1001019.

Lavenu, A., M. Leruez-Ville, M.-L. Chaix, P.-Y. Boelle, S. Rogez, F. Freymuth, A. Hay, C. Rouzioux, and F. Carrat (2006). Detailed analysis of the genetic evolution of influenza virus during the course of an epidemic. *Epidemiol Infect 134*(3), 514–520.

Lee, K. and D. Fidler (2007). Avian and pandemic influenza: progress and problems with global health governance. *Glob Public Health 2*(3), 215–234.

Lee, V. J., M. I. Chen, S. P. Chan, C. S. Wong, J. Cutter, K. T. Goh, and P. A. Tambyah (2007). Influenza pandemics in Singapore, a tropical, globally connected city. *Emerg Infect Dis 13*(7), 1052–1057.

Lee, V. J., J. Yap, J. B. Ong, K. P. Chan, R. T. Lin, S. P. Chan, K. T. Goh, Y. S. Leo, and M. I. Chen (2009). Influenza excess mortality from 1950-2000 in tropical Singapore. *PLoS One 4*(12), e8096.

Leo, Y.-S., D. C. Lye, and A. Chow (2009). Influenza in the tropics. *Lancet Infect Dis 9*(8), 457–458.

Li, F. C., B. C. Choi, T. Sly, and A. W. Pak (2008). Finding the real case-fatality rate of H5N1 avian influenza. *J Epidemiol Community Health 62*(6), 555–9.

Lidwell, O. M. and T. Sommerville (1951). Observations on the incidence and distribution of the common cold in a rural community during 1948 and 1949. *J Hyg (Lond) 49*(4), 365–81.

Lowen, A. and P. Palese (2009). Transmission of influenza virus in temperate zones is predominantly by aerosol, in the tropics by contact: a hypothesis. *PLoS Curr 1*, RRN1002.

Manolio, T. A., L. D. Brooks, and F. S. Collins (2008). A HapMap harvest of insights into the genetics of common disease. *J Clin Invest 118*(5), 1590–1605.

Matrosovich, M., T. Suzuki, Y. Hirabayashi, W. Garten, R. G. Webster, and H.-D. Klenk (2006). Gangliosides are not essential for influenza virus infection. *Glycoconj J 23*(1-2), 107–113.

McMichael, A. J., F. M. Gotch, G. R. Noble, and P. A. Beare (1983). Cytotoxic T-cell immunity to influenza. *N Engl J Med 309*(1), 13–17.

Melegaro, A., M. Jit, N. Gay, E. Zagheni, and W. J. Edmunds (2011). What types of contacts are important for the spread of infections?: using contact survey data to explore European mixing patterns. *Epidemics 3*(3-4), 143–151.

Middleton, D. and F. Gonzelez (2010). The extensive polymorphism of KIR genes. *Immunology 129*(1), 8–19.

Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman, and P. R. I. S. M. A. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med 6*(7), e1000097.

Monto, A. S. (1994). Studies of the community and family: acute respiratory illness and infection. *Epidemiol Rev 16*(2), 351–73.

Monto, A. S. (2008). Epidemiology of influenza. *Vaccine 26 Suppl 4*, D45–8.

Monto, A. S., F. M. Davenport, J. A. Napier, and J. Francis, T. (1970). Modification of an outbreak of influenza in Tecumseh, Michigan by vaccination of schoolchildren. *J Infect Dis 122*(1), 16–25.

Monto, A. S. and F. Kioumehr (1975). The Tecumseh Study of Respiratory Illness. IX. Occurence of influenza in the community, 1966–1971. *Am J Epidemiol 102*(6), 553–63.

Monto, A. S., J. S. Koopman, and J. Longini, I. M. (1985). Tecumseh study of illness. XIII. Influenza infection and disease, 1976-1981. *Am J Epidemiol 121*(6), 811–22.

Monto, A. S. and K. M. Sullivan (1993). Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiol Infect 110*(1), 145–60.

Morenilla-Palao, C., M. Pertusa, V. Meseguer, H. Cabedo, and F. Viana (2009). Lipid raft segregation modulates TRPM8 channel activity. *J Biol Chem 284*(14), 9215–9224.

Mossong, J., N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med 5*(3), e74.

Moura, F. E. A. (2010). Influenza in the tropics. *Curr Opin Infect Dis 23*(5), 415–420.

Nair, H., W. A. Brooks, M. Katz, A. Roca, J. A. Berkley, S. A. Madhi, J. M. Simmerman,

A. Gordon, M. Sato, S. Howie, A. Krishnan, M. Ope, K. A. Lindblade, P. Carosone-Link, M. Lucero, W. Ochieng, L. Kamimoto, E. Dueger, N. Bhat, S. Vong, E. Theodoratou, M. Chittaganpitch, O. Chimah, A. Balmaseda, P. Buchy, E. Harris, V. Evans, M. Katayose, B. Gaur, C. O'Callaghan-Gordo, D. Goswami, W. Arvelo, M. Venter, T. Briese, R. Tokarz, M.-A. Widdowson, A. W. Mounts, R. F. Breiman, D. R. Feikin, K. P. Klugman, S. J. Olsen, B. D. Gessner, P. F. Wright, I. Rudan, S. Broor, E. A. F. Simes, and H. Campbell (2011). Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis. *Lancet 378*(9807), 1917–1930.

Nelson, M. I., L. Simonsen, C. Viboud, M. A. Miller, and E. C. Holmes (2007). Phylogenetic analysis reveals the global migration of seasonal influenza A viruses. *PLoS Pathog 3*(9), 1220–8.

Nguyen, D. C., T. M. Uyeki, S. Jadhao, T. Maines, M. Shaw, Y. Matsuoka, C. Smith, T. Rowe, X. Lu, H. Hall, X. Xu, A. Balish, A. Klimov, T. M. Tumpey, D. E. Swayne, L. P. Huynh, H. K. Nghiem, H. H. Nguyen, L. T. Hoang, N. J. Cox, and J. M. Katz (2005). Isolation and characterization of avian influenza viruses, including highly pathogenic H5N1, from poultry in live bird markets in Hanoi, Vietnam, in 2001. *J Virol 79*(7), 4201–12.

Nguyen, D. H., H. H. Nguyen, T. V. Nguyen, T. M. Nguyen, T. M. Trinh, Q. T. Nguyen, T. Van Dinh, T. Shimbo, Y. Takahashi, Y. Kato, A. Kawana, S. Akita, and K. Kudo (2009). Human infection with highly pathogenic avian influenza virus (H5N1) in northern Vietnam, 2004-2005. *Emerg Infect Dis 15*(1), 19–23.

Nguyen, H. L. K., R. Saito, H. K. Ngiem, M. Nishikawa, Y. Shobugawa, D. C. Nguyen, L. T. Hoang, L. P. Huynh, and H. Suzuki (2007). Epidemiology of influenza in Hanoi, Vietnam, from 2001 to 2003. *J Infect 55*(1), 58–63.

Nicholls, J. M., R. W. Y. Chan, R. J. Russell, G. M. Air, and J. S. M. Peiris (2008). Evolving complexities of influenza virus and its receptors. *Trends Microbiol 16*(4), 149–157.

Nielsen, J., A. Mazick, S. Glismann, and K. Molbak (2011). Excess mortality related to seasonal influenza and extreme temperatures in Denmark, 1994-2010. *BMC Infect Dis 11*(1), 350.

Nunes, B., C. Viboud, A. Machado, C. Ringholz, H. Rebelo-de Andrade, P. Nogueira, and M. Miller (2011). Excess mortality associated with influenza epidemics in Portugal, 1980 to 2004. *PLoS One 6*(6), e20661.

Ortiz, J. R., J. A. Englund, and K. M. Neuzil (2011, Jun). Influenza vaccine for pregnant women in resource-constrained countries: a review of the evidence to inform policy decisions. *Vaccine 29*(27), 4439–4452.

Panagiotou, O. A. and J. P. A. f. t. G.-W. S. P. Ioannidis (2011). What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int. J. Epidemiol Epub ahead of print.*

Peiris, J. S., W. C. Yu, C. W. Leung, C. Y. Cheung, W. F. Ng, J. M. Nicholls, T. K. Ng, K. H. Chan, S. T. Lai, W. L. Lim, K. Y. Yuen, and Y. Guan (2004). Re-emergence of fatal human influenza A subtype H5N1 disease. *Lancet 363*(9409), 617–9.

Peiris, J. S. M., C. Y. Cheung, C. Y. H. Leung, and J. M. Nicholls (2009). Innate immune responses to influenza A H5N1: friend or foe? *Trends Immunol 30*(12), 574–584.

Perdue, M. L. and R. A. Bright (2011, Jul). United States of America Department of Health and Human Services support for advancing influenza vaccine manufacturing in the developing world. *Vaccine 29 Suppl 1*, A48–A50.

Perdue, M. L., M. Garca, D. Senne, and M. Fraire (1997). Virulence-associated sequence duplication at the hemagglutinin cleavage site of avian influenza viruses. *Virus Res 49*(2), 173–186.

Perrone, L. A., J. K. Plowden, A. Garcia-Sastre, J. M. Katz, and T. M. Tumpey (2008). H5N1 and 1918 pandemic influenza virus infection results in early and excessive infiltration of macrophages and neutrophils in the lungs of mice. *PLoS Pathog 4*(8), e1000115.

Philip, R. N., J. A. Bell, D. J. Davis, M. O. Beem, P. M. Beigelman, J. I. Engler, G. W. Mellin, J. H. Johnson, and A. M. Lerner (1961). Epidemiologic studies on influenza in familial and general population groups, 1951-1956. II. Characteristics of occurrence. *Am J Hyg 73*, 123–37.

Piel, F. B., A. P. Patil, R. E. Howes, O. A. Nyangiri, P. W. Gething, T. N. Williams, D. J. Weatherall, and S. I. Hay (2010). Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat Commun 1*, 104.

Pitzer, V. E., S. J. Olsen, C. T. Bergstrom, S. F. Dowell, and M. Lipsitch (2007). Little evidence for genetic susceptibility to influenza A (H5N1) from family clustering data. *Emerg Infect Dis 13*(7), 1074–6.

Powell, T. J., A. Fox, Y. Peng, L. T. Quynh Mai, V. T. K. Lien, N. L. K. Hang, L. Wang, L. Y.-H. Lee, C. P. Simmons, A. J. McMichael, J. J. Farrar, B. A. Askonas, T. N. Duong, P. Q. Thai, N. T. Thu Yen, S. L. Rowland-Jones, N. T. Hien, P. Horby, and T. Dong (2012). Identification of H5N1-Specific T-Cell Responses in a High-risk Cohort in Vietnam Indicates the Existence of Potential Asymptomatic Infections. *J Infect Dis 205*(1), 20–27.

Rambaut, A., O. G. Pybus, M. I. Nelson, C. Viboud, J. K. Taubenberger, and E. C. Holmes (2008). The genomic and epidemiological dynamics of human influenza A virus. *Nature 453*(7195), 615–9.

Rao, C., B. Osterberger, T. D. Anh, M. MacDonald, N. T. K. Chc, and P. S. Hill (2010). Compiling mortality statistics from civil registration systems in Viet Nam: the long road ahead. *Bull World Health Organ 88*(1), 58–65.

Reichert, T. A., N. Sugaya, D. S. Fedson, W. P. Glezen, L. Simonsen, and M. Tashiro (2001). The Japanese experience with vaccinating schoolchildren against influenza. *N Engl J Med 344*(12), 889–96.

Riley, S., K. O. Kwok, K. M. Wu, D. Y. Ning, B. J. Cowling, J. T. Wu, L.-M. Ho, T. Tsang, S.-V. Lo, D. K. W. Chu, E. S. K. Ma, and J. S. M. Peiris (2011). Epidemiological characteristics of 2009 (H1N1) pandemic influenza based on paired sera from a longitudinal community cohort study. *PLoS Med 8*(6), e1000442.

Russell, R. J., P. S. Kerry, D. J. Stevens, D. A. Steinhauer, S. R. Martin, S. J. Gamblin, and J. J. Skehel (2008). Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proc Natl Acad Sci U S A 105*(46), 17736–41.

Sabnis, A. S., C. A. Reilly, J. M. Veranth, and G. S. Yost (2008). Increased transcription of cytokine genes in human lung epithelial cells through activation of a TRPM8 variant by cold temperatures. *Am J Physiol Lung Cell Mol Physiol 295*(1), L194–200.

Sabnis, A. S., M. Shadid, G. S. Yost, and C. A. Reilly (2008). Human lung epithelial cells express a functional cold-sensing TRPM8 variant. *Am J Respir Cell Mol Biol 39*(4), 466–74.

Salomon, R., E. Hoffmann, and R. G. Webster (2007). Inhibition of the cytokine response does not protect against lethal H5N1 influenza infection. *Proc Natl Acad Sci U S A 104*(30), 12479–81.

Sandbulte, M. R., A. C. M. Boon, R. J. Webby, and J. M. Riberdy (2008). Analysis of cytokine secretion from human plasmacytoid dendritic cells infected with H5N1 or low-pathogenicity influenza viruses. *Virology 381*(1), 22–28.

Scheiffele, P., M. G. Roth, and K. Simons (1997). Interaction of influenza virus haemagglutinin with sphingolipid-cholesterol membrane domains via its transmembrane domain. *EMBO J 16*(18), 5501–5508.

Shaman, J., E. Goldstein, and M. Lipsitch (2011). Absolute humidity and pandemic versus epidemic influenza. *Am J Epidemiol 173*(2), 127–135.

Shek, L. P. and B. W. Lee (2003). Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatr Respir Rev 4*(2), 105–11.

Shelton, H., G. Ayora-Talavera, J. Ren, S. Loureiro, R. J. Pickles, W. S. Barclay, and I. M. Jones (2011). Receptor binding profiles of avian influenza virus hemagglutinin subtypes on human cells as a predictor of pandemic potential. *J Virol 85*(4), 1875–1880.

Sheng, Z.-M., D. S. Chertow, X. Ambroggio, S. McCall, R. M. Przygodzki, R. E. Cun-

ningham, O. A. Maximova, J. C. Kash, D. M. Morens, and J. K. Taubenberger (2011). Autopsy series of 68 cases dying before and during the 1918 influenza pandemic peak. *Proc Natl Acad Sci U S A 108*(39), 16416–16421.

Shinya, K., M. Ebina, S. Yamada, M. Ono, N. Kasai, and Y. Kawaoka (2006). Avian flu: influenza virus receptors in the human airway. *Nature 440*(7083), 435–6.

Shortridge, K. F. and C. H. Stuart-Harris (1982). An influenza epicentre? *Lancet 2*(8302), 812–813.

Simmerman, J. M., M. Chittaganpitch, J. Levy, S. Chantra, S. Maloney, T. Uyeki, P. Areerat, S. Thamthitiwat, S. J. Olsen, A. Fry, K. Ungchusak, H. C. Baggett, and S. Chunsuttiwat (2009). Incidence, seasonality and mortality associated with influenza pneumonia in Thailand: 2005-2008. *PLoS One 4*(11), e7776.

Simmerman, J. M. and T. M. Uyeki (2008). The burden of influenza in East and South-East Asia: a review of the English language literature. *Influenza Other Respi Viruses 2*(3), 81–92.

Skehel, J. J. and D. C. Wiley (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem 69*, 531–569.

Smith, D. J., A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. Osterhaus, and R. A. Fouchier (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science 305*(5682), 371–6.

Sokurenko, E. V., R. Gomulkiewicz, and D. E. Dykhuizen (2006). Source-sink dynamics of virulence evolution. *Nat Rev Microbiol 4*(7), 548–555.

Sorensen, T. I., G. G. Nielsen, P. K. Andersen, and T. W. Teasdale (1988). Genetic and environmental influences on premature death in adult adoptees. *N Engl J Med 318*(12), 727–732.

Stephenson, I., R. G. Das, J. M. Wood, and J. M. Katz (2007, May). Comparison of neutralising antibody assays for detection of antibody to influenza A/H3N2 viruses: an international collaborative study. *Vaccine 25*(20), 4056–4063.

Szretter, K. J., S. Gangappa, X. Lu, C. Smith, W. J. Shieh, S. R. Zaki, S. Sambhara, T. M. Tumpey, and J. M. Katz (2007). Role of host cytokine responses in the pathogenesis of avian H5N1 influenza viruses in mice. *J Virol 81*(6), 2736–44.

Tamerius, J., M. I. Nelson, S. Z. Zhou, C. Viboud, M. A. Miller, and W. J. Alonso (2011). Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environ Health Perspect 119*(4), 439–445.

Tellier, R. (2009). Aerosol transmission of influenza A virus: a review of new studies. *J R Soc Interface 6 Suppl 6*, S783–S790.

To, K. F., P. K. Chan, K. F. Chan, W. K. Lee, W. Y. Lam, K. F. Wong, N. L. Tang, D. N.

Tsang, R. Y. Sung, T. A. Buckley, J. S. Tam, and A. F. Cheng (2001). Pathology of fatal human infection associated with avian influenza A H5N1 virus. *J Med Virol 63*(3), 242–6.

Tran, T. H., T. L. Nguyen, T. D. Nguyen, T. S. Luong, P. M. Pham, V. C. Nguyen, T. S. Pham, C. D. Vo, T. Q. Le, T. T. Ngo, B. K. Dao, P. P. Le, T. T. Nguyen, T. L. Hoang, V. T. Cao, T. G. Le, D. T. Nguyen, H. N. Le, K. T. Nguyen, H. S. Le, V. T. Le, D. Christiane, T. T. Tran, J. Menno de, C. Schultsz, P. Cheng, W. Lim, P. Horby, and J. Farrar (2004). Avian influenza A (H5N1) in 10 patients in Vietnam. *N Engl J Med 350*(12), 1179–88.

Truscott, J., C. Fraser, S. Cauchemez, A. Meeyai, W. Hinsley, C. A. Donnelly, A. Ghani, and N. Ferguson (2012). Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza. *J R Soc Interface 9*(67), 304–312.

van Riel, D., M. A. den Bakker, L. M. E. Leijten, S. Chutinimitkul, V. J. Munster, E. de Wit, G. F. Rimmelzwaan, R. A. M. Fouchier, A. D. M. E. Osterhaus, and T. Kuiken (2010). Seasonal and pandemic human influenza viruses attach better to human upper respiratory tract epithelium than avian influenza viruses. *Am J Pathol 176*(4), 1614–1618.

van Riel, D., V. J. Munster, E. de Wit, G. F. Rimmelzwaan, R. A. Fouchier, A. D. Osterhaus, and T. Kuiken (2006). H5N1 Virus Attachment to Lower Respiratory Tract. *Science 312*(5772), 399.

Vannberg, F. O., S. J. Chapman, and A. V. S. Hill (2011). Human genetic susceptibility to intracellular pathogens. *Immunol Rev 240*(1), 105–116.

Viboud, C., W. J. Alonso, and L. Simonsen (2006). Influenza in tropical regions. *PLoS Med 3*(4), e89.

Viboud, C., P. Y. Boelle, S. Cauchemez, A. Lavenu, A. J. Valleron, A. Flahault, and F. Carrat (2004). Risk factors of influenza transmission in households. *Br J Gen Pract 54*(506), 684–9.

Viswanathan, K., A. Chandrasekaran, A. Srinivasan, R. Raman, V. Sasisekharan, and R. Sasisekharan (2010). Glycans as receptors for influenza pathogenesis. *Glycoconj J 27*(6), 561–570.

Wallinga, J., W. J. Edmunds, and M. Kretzschmar (1999). Perspective: human contact patterns and the spread of airborne infectious diseases. *Trends Microbiol 7*(9), 372–377.

Wallinga, J., P. Teunis, and M. Kretzschmar (2006). Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am J Epidemiol 164*(10), 936–44.

Watanabe, T., K. Shinya, S. Watanabe, M. Imai, M. Hatta, C. Li, B. F. Wolter,

G. Neumann, A. Hanson, M. Ozawa, S. Yamada, H. Imai, S. Sakabe, R. Takano, K. Iwatsuki-Horimoto, M. Kiso, M. Ito, S. Fukuyama, E. Kawakami, T. Gorai, H. A. Simmons, D. Schenkman, K. Brunner, S. V. Capuano, 3rd, J. T. Weinfurter, W. Nishio, Y. Maniwa, T. Igarashi, A. Makino, E. A. Travanty, J. Wang, A. Kilander, S. G. Dudman, M. Suresh, R. J. Mason, O. Hungnes, T. C. Friedrich, and Y. Kawaoka (2011). Avian-type receptor-binding ability can increase influenza virus pathogenicity in macaques. *J Virol 85*(24), 13195–13203.

Watanabe, Y., M. S. Ibrahim, H. F. Ellakany, N. Kawashita, R. Mizuike, H. Hiramatsu, N. Sriwilaijaroen, T. Takagi, Y. Suzuki, and K. Ikuta (2011). Acquisition of human-type receptor binding specificity by new H5N1 influenza virus sublineages during their emergence in birds in Egypt. *PLoS Pathog 7*(5), e1002068.

Weber, A., P. Wasiliew, and M. Kracht (2010). Interleukin-1 (IL-1) pathway. *Sci Signal 3*(105), cm1.

WHO (2005). H5N1 avian influenza: a chronology of key events.

WHO (2010). Summary of human infection with highly pathogenic avian influenza A (H5N1) virus reported to WHO, January 2003-March 2009: cluster-associated cases. *Wkly Epidemiol Rec 85*(3), 13–20.

Wiley, D. C. and J. J. Skehel (1987). The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu Rev Biochem 56*, 365–394.

Wong, C. M., L. Yang, K. P. Chan, G. M. Leung, K. H. Chan, Y. Guan, T. H. Lam, A. J. Hedley, and J. S. Peiris (2006). Influenza-associated hospitalization in a subtropical city. *PLoS Med 3*(4), e121.

Yang, L., S. Ma, P. Y. Chen, J. F. He, K. P. Chan, A. Chow, C. Q. Ou, A. P. Deng, A. J. Hedley, C. M. Wong, and J. S. M. Peiris (2011). Influenza associated mortality in the subtropics and tropics: results from three Asian cities. *Vaccine 29*(48), 8909–8914.

Zuniga, J., I. Buendia, Y. Zhao, L. Jimenez, G. S.nez, D. Torres, J. Romo, G. Ramirez, A. Cruz, G. Vargas-Alarcon, C.-C. Sheu, F. Chen, L. Su, A. M. Tager, A. Pardo, M. Selman, and D. C. Christiani (2011). Genetic variants associated with severe pneumonia in A/H1N1 influenza infection. *Eur Respir J Epub ahead of print.*

Appendix A

Supplementary research paper 1

Title: Identification of H5N1-Specific T-Cell Responses in a High-risk Cohort in Vietnam Indicates the Existence of Potential Asymptomatic Infections

Author(s): Powell TJ, Fox A, Peng Y, Quynh Mai le T, Lien VT, Hang NL, Wang L, Lee LY, Simmons CP, McMichael AJ, Farrar JJ, Askonas BA, Duong TN, Thai PQ, Thu Yen NT, Rowland-Jones SL, Hien NT, **Horby P**, Dong T.

Journal/Publisher: Journal of Infectious Diseases, Oxford University Press

Type of publication: Research article

Stage of publication: Published

Academic peer-reviewed: Yes.

Copyright: As part of my copyright agreement with Oxford University Press I have retained the right, after publication, to use all or part of the article and abstract, in the preparation of derivative works.

Candidate's role: The samples used in this work are derived from community cohort study of which I am the principal investigator. This work was conceived as a sub-study of the community cohort at the outset. I was involved in designing the experiments, interpreting the results, and drafting the manuscript.

Candidate's signature:

Supervisor or senior author's signature to confirm Candidates role:

MAJOR ARTICLE

Identification of H5N1-Specific T-Cell Responses in a High-risk Cohort in Vietnam Indicates the Existence of Potential Asymptomatic Infections

Timothy J. Powell,[1] Annette Fox,[2] Yanchun Peng,[1] Le Thi Quynh Mai,[3] Vu T. K. Lien,[3] Nguyen L. K. Hang,[3] LiLi Wang,[1] Laurel Y.-H. Lee,[1] Cameron P. Simmons,[4] Andrew J. McMichael,[1] Jeremy J. Farrar,[4] Brigitte A. Askonas,[1] Tran Nhu Duong,[3] Pham Quang Thai,[3] Nguyen Thi Thu Yen,[3] Sarah L. Rowland-Jones,[1] Nguyen Tran Hien,[3] Peter Horby,[2] and Tao Dong[1]

[1]MRC Human Immunology Unit, University of Oxford, John Radcliffe Hospital, United Kingdom; [2]Oxford University Clinical Research Unit and [3]National Institute of Hygiene and Epidemiology, Hanoi, and [4]Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam

(See the editorial commentary by Epstein, on pages 4–6.)

Background. Most reported human H5N1 viral infections have been severe and were detected after hospital admission. A case ascertainment bias may therefore exist, with mild cases or asymptomatic infections going undetected. We sought evidence of mild or asymptomatic H5N1 infection by examining H5N1-specific T-cell and antibody responses in a high-risk cohort in Vietnam.

Methods. Peripheral blood mononuclear cells were tested using interferon-γ enzyme-linked immunospot T assays measuring the response to peptides of influenza H5, H3, and H1 hemagglutinin (HA), N1 and N2 neuraminidase, and the internal proteins of H3N2. Horse erythrocyte hemagglutination inhibition assay was performed to detect antibodies against H5N1.

Results. Twenty-four of 747 individuals demonstrated H5-specific T-cell responses but little or no cross-reactivity with H3 or H1 HA peptides. H5N1 peptide-specific T-cell lines that did not cross-react with H1 or H3 influenza virus HA peptides were generated. Four individuals also had antibodies against H5N1.

Conclusions. This is the first report of ex vivo H5 HA-specific T-cell responses in a healthy but H5N1-exposed population. Our results indicate that the presence of H5N1-specific T cells could be an additional diagnostic tool for asymptomatic H5N1 infection.

Influenza H5N1 remains endemic in domestic poultry in large parts of Asia, and although the total number of human infections is relatively small, sporadic human cases with a high risk of death are still being reported [1, 2]. Since 2003, >500 human cases of highly pathogenic influenza A H5N1 have been reported, with 119 cases occurring in Vietnam [3]. At present, H5N1 influenza cannot be transmitted readily between humans, but the possibility remains of a recombination between H5N1 and other influenza viruses, resulting in a virulent and easily transmissible virus [4].

The reported frequency and severity of H5N1 infection in humans is almost certainly biased by the under-detection of mild or asymptomatic cases: leading to an underestimate of the number of cases and an overestimate of the case fatality rate. The extent of this bias is indicated by seroprevalence surveys that have reported anti-H5 antibody prevalence in exposed groups of between 0% and 12% [5–9]. The presence of virus-neutralizing antibody is important for protection against influenza, and antibodies that recognize specific hemagglutinin (HA) subtypes can give an indication of recent infection history [10, 11]. However, measurement of H5N1-specific neutralizing antibodies has been problematic because the

Received 8 April 2011; accepted 9 August 2011; electronically published 11 November 2011.

Presented in part: Options for the Control of Influenza VII, Hong Kong, China, 3–7 September 2010. Abstract 3783; British Society for Immunology Annual Congress, Liverpool, United Kingdom, 6–10 December 2010. Abstract 577.

Correspondence: Tao Dong, PhD, MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DS, United Kingdom (tao.dong@imm.ox.ac.uk).

The Journal of Infectious Diseases 2012;205:20–27

© The Author 2011. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

DOI: 10.1093/infdis/jir689

Downloaded from http://jid.oxfordjournals.org/ at Oxford University on December 28, 2011

traditional hemagglutination inhibition (HAI) assay has low sensitivity for the detection of H5N1 antibodies [12]. Alternative assays that have been used for the detection of H5N1 infection include horse erythrocyte HAI, microneutralization, and pseudoparticle assays [13]. These assays are all subject to false-positive reactions due to the presence of cross-subtype neutralizing antibodies [14]. Measuring interferon (IFN) γ secretion by peripheral blood mononuclear cells (PBMCs) after stimulation with pools of HA peptides from different influenza strains demonstrates the specificity and reactivity of T cells generated by strain-specific vaccination [15–17] or memory cells generated after natural infection [18–21].

In this study, we set out to investigate the rate of infection with H5N1 virus in a community in rural Vietnam that had previously experienced H5N1 cases in both poultry and humans by measuring the prevalence of specific T-cell responses against the HA and neuraminidase (NA) of H5N1 influenza virus. We also compared T-cell responses against the proteome of seasonal H3N2 and the HA and N1 of H5N1 and H1N1 influenza in the community cohort with those in a group of persons who had recovered from H5N1 infection and in healthy controls with no exposure to H5N1.

MATERIALS AND METHODS

Patient Cohorts

For the community cohort, a household-based cohort was established in a community in northern Vietnam that had previously experienced outbreaks of H5N1 in poultry and human H5N1 cases. Human H5N1 infections had occurred in the community in 2004, and additional human cases were identified in 2007 and 2008 in nearby villages. Poultry outbreaks of H5N1 had been detected in the province intermittently since 2004, including an outbreak 1 month before the beginning of the study. Households were randomly selected from a complete household register using a random number table. If a selected household declined to participate, the nearest neighbor was approached for participation. Members of the cohort provided blood for collection of PBMCs and plasma in December 2007.

For the recovered case patients, persons who were convalescent after H5N1 infection (n = 19) were recruited between July 2008 and March 2009, at intervals of 84–1449 days (median, 1300) after H5N1 onset (HHT Trang, A. Fox, T. Dong, LQ Mai, VTK Lien, T. Powell, TN Duong, NTT Yen, PQ Thai, NT Hien, P Horby, authors' unpublished data). For the healthy control group, PBMCs from volunteers (n = 271) from the United Kingdom and Vietnam were tested by enzyme-linked immunospot spot (ELISPOT) assays [19]. All participants provided written informed consent. The study was approved by the ethical review boards of the National Institute of Hygiene and Epidemiology (Hanoi, Vietnam), the University of Oxford (United Kingdom), and the London School of Hygiene & Tropical Medicine (United Kingdom).

Media, Peptides, and ELISPOT Screening

Dulbecco's phosphate-buffered saline (Sigma–Aldrich) and R10 (Roswell Park Memorial Institute 1640 medium plus 10% vol/vol fetal calf serum, glutamine, and penicillin streptomycin) medium were used as described elsewhere [19]. The sequences of the full influenza proteome from H3N2, *A/New York/388/2005* (HA and NA) and *A/New York/232/2004* (H3N2) (internal proteins), HA from H1N1 (*A/Hong Kong/1134/98* and *A/NewYork/228/2003*), and H5 and N1 from H5N1 (*A/Vietnam/CL26/2004*) were split into overlapping peptides of 18–21 residues overlapping by 10, using Peptgen (http://www.hiv.lanl.gov/content/sequence/PEPTGEN/peptgen.html) synthesized by Sigma-Aldrich. Peptides were divided into pools (Table 1). Overlapping peptides spanning the entire length of each HA protein were split up into 2 pools (eg, H5 into H5-1 and H5-2), and the amino acid range is shown for each pool. The exact sequences of the peptides are available on request.

Blood samples were taken from volunteers at local sites and transported to the National Institute of Hygiene and Epidemiology on the same day, where PBMCs were isolated and 2.5×10^5 were incubated with peptides at 2 μg/mL overnight, as described elsewhere [19]. Plates were read on an ELISPOT plate reader (CTL). Positive pools were defined using established criteria of 3 times average background and/or >10 spots per well [19]. Spot-forming units were the actual number of spots generated from a known number of cells.

Table 1. Peptide Pool Identities

Pool	Peptide Identity (No. of Peptides in Pool)[a]
HA1-1	H1 HA1 1–294 (37)
HA1-2	H1 HA1 285–565 (37)
HA3-1	H3 HA3 1–298 (37)
HA3-2	H3 HA3 289–556 (37)
HA5-1	H5 HA5 1–281 (38)
HA5-2	H5 HA5 272–565 (39)
NA1-1	N1 NA1 1–227 (30)
NA1-2	N1 NA1 218–442 (31)
NA2-1	N2 NA2 1–303 (40)
NA2-2	N2 NA2 294–467 (24)
PB1-1	H3N2 PB1 1–380 (50)
PB1-2	H3N2 PB1 370–757 (50)
PB2-1	H3N2 PB2 1–384 (52)
PB2-2	H3N3 PB2 375–759 (52)
M1/2	H3N2 M1 1–252 (34); M2 1–97 (13)
NP-1	H3N2 NP 1–261 (34)
NP-2	H3N2 NP 252–498 (34)
NS1/2	H3N2 NS1 1–230(29); NS2 1–121(17)
PA-1	H3N2 PA 1–370 (49)
PA-2	H3N2 PA 361–716 (46)

[a] Numbers are the range of the AMINO ACIDS contained within the pool.

Downloaded from http://jid.oxfordjournals.org/ at Oxford University on December 28, 2011

Generation of B-Cell Lines and Antigen-Specific T-Cell Lines

Epstein Barr virus (EBV)–transformed B-cell lines (BCLs) were generated by adding EBV supernatant, from a B958 cell line, to 1–2 × 10^6 PBMCs in a 96-well plate for 3–4 hours, followed by 2 µg/mL cyclosporine (Sandoz Pharmaceuticals) in R15 (RPMI medium with 15% vol/vol fetal calf serum). Antigen-specific short-term T-cell lines (STLs) were generated by pulsing PBMCs with peptide for 90 minutes, washing once, and then culturing in 96-well plates in H10 (10% vol/vol human AB serum; National Blood Service). Three days later, interleukin-2 (PeproTech EC) was added at a final concentration of 200 U/mL. STLs were maintained by restimulating with peptide-pulsed autologous BCLs every 10–15 days. For cultured ELISPOT assays, STLs were rested in H10 for 26–36 hours and then used in an ELISPOT assay. T cells (n = 40 000) were mixed with 10 or 2 µg/mL peptide pools and cultured for 18–20 hours, and then spots were developed using above protocol [22, 23]. T-cell lines that did not expand or were negative at ELISPOT assay were excluded from analysis.

Cloning of Cell Lines

Cell lines were stimulated with peptide-pulsed autologous BCLs for 3–4 hours followed by labeling with human IFN-γ capture kit (Miltenyi Biotec). High IFN-γ producers were sorted on a MoFlow cytometer (DakoCytomation). Clones were restimulated every 14–21 days using phytohemagglutin-treated irradiated allogeneic PBMCs, as described elsewhere [24].

Intracellular Cytokine Staining

Cells were stimulated with 10 µg/mL peptide for 1 hour followed by addition of BFA/Monesin (BD Biosciences). After an additional 12–16 hours, cells were washed, labeled with anti-CD4 Pacific blue (eBiosciences) anti-CD8 fluorescein isothiocyanate (BD Biosciences), permeabilized with FixPerm (BD) labeled with anti–tumor necrosis factor α Allophycocyanin (APC) (eBiosciences) and anti–IFN-γ Phycoerythrin PE (eBiosciences), then washed with Perm/Wash (BD) and fixed. Cells were analyzed on a CyAn flow cytometer (DakoCytomation).

HAI Assay and Antibody ELISPOT Assay

Plasma was tested in a standard HAI assay with antigens representing clade 1 and clade 2.3.4 H5N1 strains circulating in Vietnam and horse red blood cells, as described elsewhere [7, 12]. Donors were considered positive if they had an antibody titer of 1:40 or more [25].

RESULTS

Sufficient blood for ELISPOT assays was obtained from 747 participants in December 2007. Thirty-six participants had responses to H5 HA by ELISPOT assays. Twenty-four participants (3.2%) demonstrated specific responses to the H5 HA peptide pools but far lower (≥2-fold, but the majority of H5-specific responses were 5-fold) or no response to either H1 HA or H3

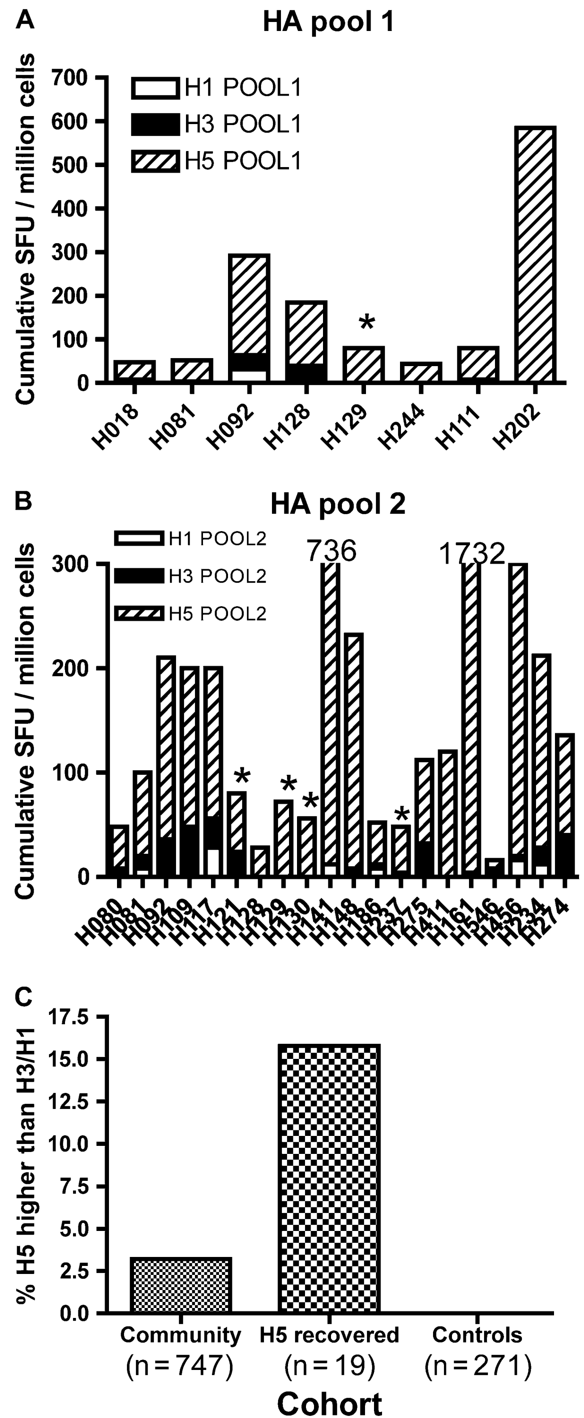


Figure 1. Enzyme-linked immunospot spot (ELISPOT) assay results in cohort volunteers with H5 peptide-specific responses. Peripheral blood mononuclear cells were stimulated with overlapping peptides from H1, H3, and H5 proteins to the first half (*A*) and second half (*B*) of the protein. Asterisks denote donors who were also positive for H5 antibody. *C*, Percentage of each cohort who had H5-specific T-cell responses by interferon-γ ELISPOT assays. Number of donors in each cohort is shown underneath each bar. Abbreviations: HA, hemagglutinin; SFU, spot-forming units.

Downloaded from http://jid.oxfordjournals.org/ at Oxford University on December 28, 2011

HA peptide pools. This included 6 donors who only made responses to H5 peptide pools and were negative for H3 and H1 pools (Figure 1*A*, *B*). There were 111 participants who had ELISPOT responses to H5 that are less than or equal to the H1 and/or H3 responses (Supplementary Table 1). However, none of 271 healthy controls at low risk of H5N1 exposure showed H5N1-specific responses, and 16% of person who had recovered from H5N1 infection made specific responses to H5 HA pools (Figure 1*C*). Of these recovered patients, tested during acute infection for H5 HAI, 5 of 14 were positive (HHT Trang, A. Fox, T. Dong, LQ Mai, VTK Lien, T. Powell, TN Duong, NTT Yen, PQ Thai, NT Hien, P Horby, authors' unpublished data). Thirty-seven (5%) of the community cohort participants had positive H5 antibody titers, and 4 also showed H5-specific T-cell responses (Figure 1*A*, *B*, denoted by asterisks). Within this group of individuals, 12 had T-cell responses (4 with H5-specific responses, a subset of these) and 25 had no T-cell responses.

For further analysis of the H5-specific responders, STLs and cultured ELISPOT experiments were performed. Tables 2 and 3 show the results of direct ex vivo and cultured ELISPOT from a number of different donors who were positive for H5 responses at initial screening, and PBMCs were then expanded as STL. Figure 2*A* shows 3 examples of HA STLs that show higher responses to H5N1 HA pools than to H1N1 or H3N2 HA pools. When similar experiments were done with other lines, 4 donors had H5-specific STL after expansion in vitro (Table 2). Ten donors had responses to internal pools that were either H3N2 peptide pool specific or cross-reactive with both H5N1 and H3N2 peptide pools (Table 3).

To determine the T-cell recognition of the single peptides containing potential epitopes from H5 HA pools in individuals with H5N1-specific T-cell responses, STLs were generated and tested by cultured ELISPOT assays using single peptides. The STL grown from donor 0081 for the HA5-1 pool was able to recognize only the peptide with the amino acid sequence 160–177 of $HA5_{160–77}$ FRNVVWLIKKNSTYPTIK and not the equivalent peptides from H1 or H3 (Figure 2*B*). Another line grown from this donor using the HA5-2 pool was able to respond only to the peptide $HA5_{439–56}$; TYNAELLVLMENERTLDF from the H5 strain of virus but not to the equivalent H1 and H3 peptides (Figure 2*C*). A similar response to $HA5_{439–56}$ was also found with a second donor (data not shown). A second line generated from donor 0081 using pool HA5-2 was specific to peptide $HA5_{344–64}$ KKRGLFGAIAGFIEGGWQGMV (Table 2). Notably, $HA5_{344–64}$ was CD4-restricted and $HA5_{439–56}$ was CD8-restricted, so the H5 HA specificity was not limited to only CD8 T cells.

To investigate whether this cohort had cross-reactive responses against peptides from the internal proteins of influenza, STLs were generated to internal peptide pools that had been found to be positive in initial ELISPOT screening. These lines were tested for reactivity against seasonal (H3) peptides and also H5 peptides. STLs from 5 donors were able to recognize internal peptide pools from both H3N2 and H5N1 (Figure 3 and Table 3). We found 1 CD4 clone that was able to respond to peptide $NS1_{163–85}$ GHTIEDVKNAIGVLIGGL, a peptide derived from an internal H3N2 NS1 protein that has not been documented elsewhere. We also identified several previously unknown individual peptides containing potential T-cell epitopes in H3N2 internal proteins that were cross-reactive with equivalent H5N1 peptides: $PA_{163–80}$

Downloaded from http://jid.oxfordjournals.org/ at Oxford University on December 28, 2011

Table 2. H5 Hemagglutinin (HA) Responders and Generation of HA5-Specific Short-Term T-Cell Lines (STLs)

	Ex Vivo Responses			STL	
Donor[a] and Stimulation Pool[b]	H5N1[c]	H3N2	H1N1	Single Peptide Identity[d]	CD4/CD8 Cell Type[e]
Donor 0081					
HA5-1	+	–	–	$HA5_{160–77}$	ND
HA5-2	+	–	–	$HA5_{439–56}$, $HA5_{344–64}$	CD8, CD4
Donor 0092					
HA5-1	++	–	–		CD8
HA5-2	++	–	–		CD8
Donor 0148					
HA5-2	++	–	–	$HA5_{344–64}$	CD4
Donor 0275					
HA5-2	+	–	–		CD4

Abbreviation: ND, no data.

[a] Donor number assigned during collection.

[b] Peptide pool used to stimulate the STL or peripheral blood mononuclear cells (PBMCs), as shown in Table 1.

[c] Ex vivo enzyme-linked immunospot response on fresh PBMCs; ++ indicates strong response (>100 spot-forming units [SFU]/10^6 PBMCs); +, medium response (40–99 SFU/10^6 PBMCs); –, no response (<40 SFU/10^6 cells).

[d] Single peptide defined by incubation of STLs with single peptides.

[e] Lymphocyte subset CD4/CD8 defined by tumor necrosis factor α or interferon-γ secretion in intracellular cytokine staining (ICS) after restimulation.

Table 3. Internal Peptide-Specific Short-Term T-Cell Lines (STLs)

Donor and Pool[a]	Ex Vivo[b,c]	Expanded STL Specificity		Single Peptide Identified[d]	CD4/CD8[e]
	H3N2	H5N1	H3N2		
Donor 0018					
NP-1	++	+++	+++	$NP_{199–216}$	CD8
Donor 0053					
NP-1	++	ND	+++		CD4
Donor 0080					
NP-1	++	++	++		CD8/CD4
NS1/2	++	++	++		CD4
Donor 0081					
PB2-2	++	ND	+++	$PB2_{427–44}$, $PB2_{637–55}$	ND
NS1/2	++	ND	++	$NS1_{163–85}$	CD4
Donor 0092					
PB1-1	++	ND	+		CD4
NP-1	++	+	++		ND
PA-1	++	ND	+		CD4
Donor 0109					
NP-1	+	ND	+		ND
Donor 0130					
NP-2	+	ND	++		ND
Donor 0141					
NP-1	++	+	++		ND
NP-2	++	+	++		CD4
Donor 0142					
M1/2	++	+	++		CD4
NP-1	++	+	++		CD8
Donor 0275					
M1/2	++	+	++		CD4
NP-2	++	+	++		CD4
Donor 0492					
NP-2	++	ND	++		CD8/CD4

Abbreviation: ND, no data.

[a] Peptide pool as defined in Table 1.

[b] Ex vivo enzyme-linked immunospot (ELISPOT) response on fresh peripheral blood mononuclear cells (PBMCs); ++ indicates strong response (>100 spot-forming units [SFU]/10^6 PBMCs); +, medium response (40–99 SFU/10^6 PBMCs); −, no response (<40 SFU/10^6 cells).

[c] Derived STL. Rested ELISPOT response: −. no response, +, ≤50 SFU/10^5 cells; ++, 50–100 SFU/10^5 cells; +++ > 100 SFU/10^5 cells.

[d] Single peptide identified by stimulation in ELISPOT.

[e] CD4/CD8 defined by tumor necrosis factor α or interferon-γ intracellular cytokine staining (ICS) after restimulation.

RIKTRLFTIRQEMASRGL, $PB2_{427–44}$ RLNTMHQLLRHFQK-DAKV, $PB2_{637–55}$ TVNVRGSGMRILVRGNSPV, $NP_{199–216}$ RGINDRNFWRGENGRRTR, and $PA_{406–21}$ KACELTDSIWIEL-DEI, noted in Table 3.

DISCUSSION

We used ELISPOT assays to analyze T-cell responses to influenza in members of a community exposed to H5N1 and found that 24 of 747 (3.2%) had specific responses to H5 HA peptides but little or no response to equivalent H3 or H1 HA peptides. H5-specific responses were further confirmed by cultured ELISPOT assays and by growing STLs and clones. Although H5 HA-specific CD4 T-cell responses can be generated in unexposed healthy individuals by in vitro expansion from PBMCs [26], this is the first study to detect H5 HA-specific T-cell responses directly ex vivo in a cohort at high risk of H5N1 exposure. In contrast, we did not detect any ex vivo H5 HA-specific T-cell responses in 271 unexposed healthy controls.

Almost 5% of participants (37 of 747; 4.9%) had horse erythrocyte HAI antibody titers ≥1:40, and 4 of them had both an H5N1-positive antibody titer and H5-specific T-cell responses. The poor correlation between the antibody and T-cell measurements may be a result of different kinetics of persistence after

Downloaded from http://jid.oxfordjournals.org/ at Oxford University on December 28, 2011

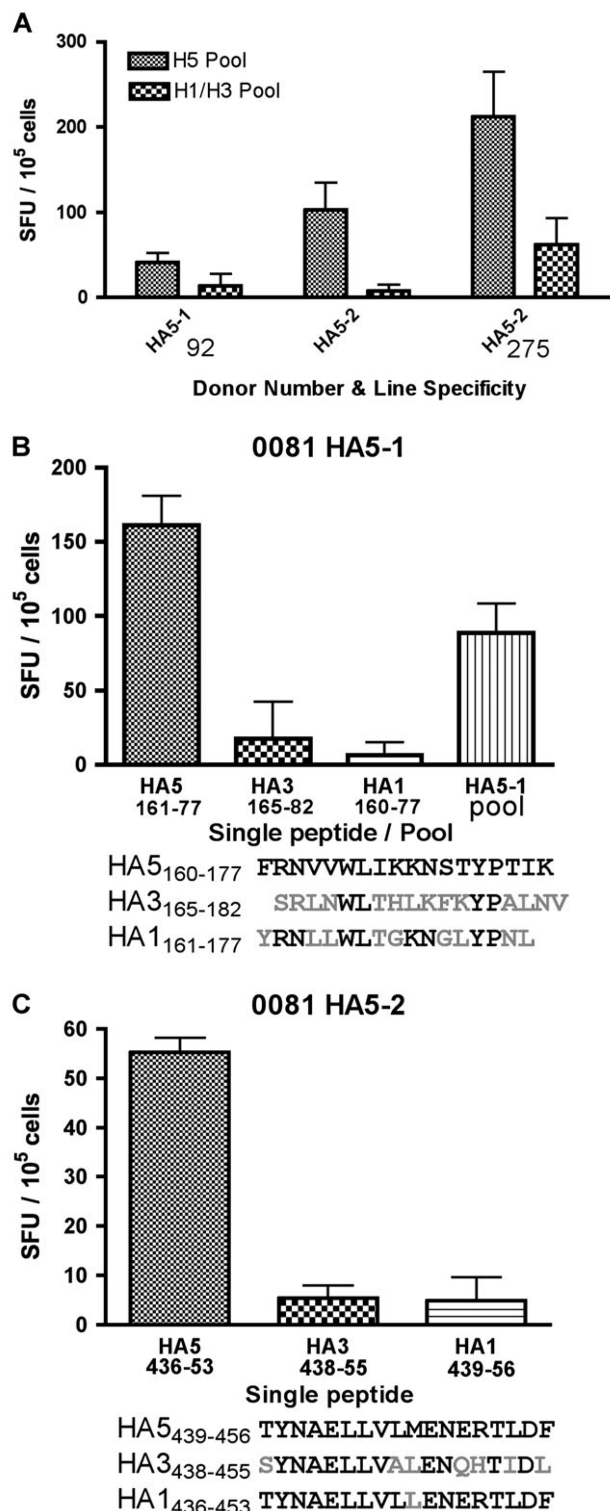


Figure 2. Hemagglutinin (HA)–specific short-term T-cell lines show specificity for H5 pools, but not H3 or H1 pools, and are specific to single H5 HA peptides. *A*, Peripheral blood mononuclear cells (PBMCs) were expanded in vitro using HA5-1 or HA5-2 peptide pools and interleukin (IL) 2. Ten days later, cells were rested overnight in IL-2–free media and then tested in an enzyme-linked immunospot spot (ELISPOT) assay using H5 pools, HA5-1 or HA5-2, depending on the specificity of the line, or a mixture of H3/H1 pools from the same corresponding region of the HA protein. *B*, *C*, STLs were restimulated with antigen pulsed autologous B cells. After 10-day stimulation and 30-hour rest, cells were tested in their response to specific peptides and the release of interferon-γ was measured by ELISPOT assay. *A*, *B*, *C*, Results shown are spot-forming units (SFU) per 10^5 cells of duplicate or triplicate wells (± standard deviations).

virus exposure. H5N1 antibodies have been shown to persist after severe infection [27] but decline after mild or asymptomatic infection [5]. H5-specific T-cell responses are seen only in a small proportion of confirmed cases, possibly because T-cell responses are short-lived, as seen with seasonal influenza [28], and it is not known how long specific H5 T-cell responses may persist. Both HAI and T-cell assays may not identify all infections, because there have been studies in which only 2 of 5 virologically confirmed H5N1 cases have antibodies detectable by HAI [12], and H5N1-specific T cells are not detected in all persons convalescing after confirmed H5N1 infection (HHT Trang, A. Fox, T. Dong, LQ Mai, VTK Lien, T. Powell, TN Duong, NTT Yen, PQ Thai, NT Hien, P Horby, authors' unpublished data). Therefore, measured rates of prevalence could be underestimated whichever assay (antibody or T cell) is used. Multiple time points and samples would answer questions of persistence [5, 27], and further studies are needed to explore this issue. Not many donors make both T-cell and antibody responses, and this may be due to underlying issues of immune repertoire between different donors or for the H5N1 cohort differences in clinical interventions [29]. Because the measured proportion of H5 HAI–positive results was greater in patients who had recovered from H5N1 infection than in the community cohort, HAI correlates to some extent with rate of infection or exposure. T-cell or antibody responses have been shown to persist for up to 6 months in vaccine studies [30, 31] but to decline after early time points. Most studies use only early time points [less than one month] after infection [15, 16]. Therefore, this study was undertaken during the influenza season to obtain samples during or shortly after influenza exposure [25].

The detection of subclinical H5N1 infection has several important implications. First, it can provide a more accurate assessment of the probability of animal-to-human transmission and of the severity of human H5N1 infection. Second, persons with subclinical or asymptomatic H5N1 infection will not be hospitalized and may be at risk of coinfection with another seasonal virus, which could reassort with H5N1 [32]. Third, identifying groups with different severities of H5N1 infection can contribute to our understanding of the pathogenesis of severe H5N1 influenza and factors that may influence susceptibility to severe disease [6].

Most donors have cross-reactive H5N1 T-cell responses to peptides from internal influenza genes that have been shown elsewhere to respond to infected target cells [19, 20]. There could be an important role for cross-reactive T-cell responses in protection against H5N1. Further characterization of these

Downloaded from http://jid.oxfordjournals.org/ at Oxford University on December 28, 2011

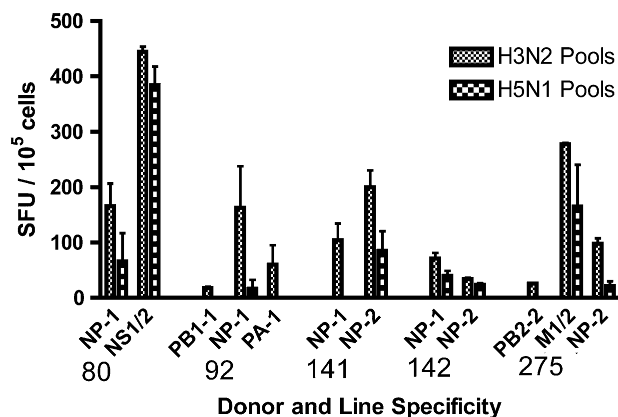


Figure 3. Short-term T-cell lines derived from H3N2 internal peptide pools show cross-reactive responses to both H3N2 and H5N1 peptide pools. Short-term lines were rested overnight in H10 media and then tested in a cultured enzyme-linked immunospot spot assay against peptide pools from internal genes of H3N2 and H5N1 peptide pools. Line specificity is shown on the axis, and duplicate values with background subtracted are shown ± standard deviations. Abbreviations: NP, nucleoprotein; NS, non-structural protein; PB1 and 2, polymerase basic protein 1 and 2; PA, polymerase acidic protein; M, matrix protein; SFU, spot-forming units.

recognized peptides will help in the design of universal influenza vaccines that target the less variable internal genes of the virus.

H5 HA-specific T cells are most likely to have been generated as a result of prior infection with, or exposure to, a low level of H5N1 virus. CD8 T cells are stimulated more readily by virus infection rather than inactivated vaccines [31, 33]. A low level of infection may have occurred because upper human respiratory tract lacks the α2,3-galactose sialic acid receptors that H5N1 viruses preferentially bind [34] so that the H5N1 virus is unable to replicate to a high titer.

In conclusion, we report evidence of possible subclinical H5N1 infection demonstrated with T-cell ELISPOT assays. These responses were not detectable by horse erythrocyte HAI. We consider that detection of H5N1-specific T-cell responses may be a useful adjunct to serology to identify the prevalence of infection with H5N1. Further research is needed with different cohorts in different geographical areas to determine whether this is universally applicable. Characterization and comparison of T-cell responses between asymptomatic responders and patients who have recovered from H5N1 infection may provide insights into the immune responses associated with severity of infection. Identification of cross-reactive epitopes in asymptomatic individuals may provide useful information for universal influenza vaccine design.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://www.oxfordjournals.org/our_journals/jid/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We acknowledge help and assistance from many coworkers in Vietnam and blood donors in Oxford and Vietnam. We thank Craig Waugh for cell sorting, Tim Rostron for HLA typing, and Alastair Waugh and Rebecca Horsfall for EBV production.

Financial support. This work was supported by the UK Medical Research Council (grant G0600520 to T. D.) and the Wellcome Trust UK (grants 081613/Z/06/Z and 077078/Z/05/Z).

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Author contributions. T.J.P., A.F., Y.P., L.T.Q.M.. V.T.K.L., N.L.K.H., and L.W. performed experiments; T.D., PH, A.F., N.T.H., S.L.R.J., N.T.T.Y., P.Q.T., T.N.D., B.A.A., C.P.S., J.J.F., A.J.M., and L.L.Y.L. designed the study; T.J.P., A.F., V.T.K.L., N.L.K.H., L.W., and T.D. analyzed data; T.J.P., T.D., A.F., and P.H. wrote the article.

References

1. Beigel JH, Farrar J, Han AM, et al. Avian influenza A (H5N1) infection in humans. N Engl J Med **2005**; 353:1374–85.
2. Gambotto A, Barratt-Boyes SM, de Jong MD, Neumann G, Kawaoka Y. Human infection with highly pathogenic H5N1 influenza virus. Lancet **2008**; 371:1464–75.
3. World Health Organization. Confirmed human cases of avian influenza A(H5N1). Avian Influenza Weekly Update 302. 14 October 2011. http://www.who.int/influenza/human_animal_interface/H5N1_cumulative_table_archives/en/index.html. Accessed 8 April 2011.
4. Salomon R, Webster RG. The influenza virus enigma. Cell **2009**; 136:402–10.
5. Buchy P, Vong S, Chu S, et al. Kinetics of neutralizing antibodies in patients naturally infected by H5N1 virus. PLoS One **2010**; 5:e10864.
6. Vong S, Ly S, Van Kerkhove MD, et al. Risk factors associated with subclinical human infection with avian influenza A (H5N1) virus–Cambodia, 2006. J Infect Dis **2009**; 199:1744–52.
7. Schultsz C, Nguyen VD, Hai le T, et al. Prevalence of antibodies against avian influenza A (H5N1) virus among cullers and poultry workers in Ho Chi Minh City, 2005. PLoS One **2009**; 4:e7948.
8. Katz JM, Lim W, Bridges CB, et al. Antibody response in individuals infected with avian influenza A (H5N1) viruses and detection of anti-H5 antibody among household and social contacts. J Infect Dis **1999**; 180:1763–70.
9. Ceyhan M, Yildirim I, Ferraris O, et al. Serosurveillance study on transmission of H5N1 virus during a 2006 avian influenza epidemic. Epidemiol Infect **2010**; 138:1274–80.
10. Barr IG, McCauley J, Cox N, et al. Epidemiological, antigenic and genetic characteristics of seasonal influenza A(H1N1), A(H3N2) and B influenza viruses: basis for the WHO recommendation on the composition of influenza vaccines for use in the 2009–2010 Northern Hemisphere season. Vaccine **2009**; 28:1156–67.
11. Nicholson KG, Wood JM, Zambon M. Influenza. Lancet **2003**; 362:1733–45.
12. Rowe T, Abernathy RA, Hu-Primmer J, et al. Detection of antibody to avian influenza A (H5N1) virus in human serum by using a combination of serologic assays. J Clin Microbiol **1999**; 37:937–43.
13. Temperton NJ, Hoschler K, Major D, et al. A sensitive retroviral pseudotype assay for influenza H5N1-neutralizing antibodies. Influenza Other Respi Viruses **2007**; 1:105–12.

Downloaded from http://jid.oxfordjournals.org/ at Oxford University on December 28, 2011

14. Garcia JM, Pepin S, Lagarde N, et al. Heterosubtype neutralizing responses to influenza A (H5N1) viruses are mediated by antibodies to virus haemagglutinin. PLoS One **2009**; 4:e7918.
15. McElhaney JE, Xie D, Hager WD, et al. T cell responses are better correlates of vaccine protection in the elderly. J Immunol **2006**; 176: 6333–9.
16. Hammitt LL, Bartlett JP, Li S, et al. Kinetics of viral shedding and immune responses in adults following administration of cold-adapted influenza vaccine. Vaccine **2009**; 27:7359–66.
17. Zinckgraf JW, Sposato M, Zielinski V, Powell D, Treanor JJ, von Hofe E. Identification of HLA class II H5N1 hemagglutinin epitopes following subvirion influenza A (H5N1) vaccination. Vaccine **2009**; 27: 5393–401.
18. Jameson J, Cruz J, Terajima M, Ennis FA. Human CD8+ and CD4+ T lymphocyte memory to influenza A viruses of swine and avian species. J Immunol **1999**; 162:7578–83.
19. Lee LY, Ha DL, Simmons C, et al. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. J Clin Invest **2008**; 118:3478–90.
20. Kreijtz JH, de Mutsert G, van Baalen CA, Fouchier RA, Osterhaus AD, Rimmelzwaan GF. Cross-recognition of avian H5N1 influenza virus by human cytotoxic T-lymphocyte populations directed to human influenza A virus. J Virol **2008**; 82:5161–6.
21. Babon JA, Cruz J, Orphin L, et al. Genome-wide screening of human T-cell epitopes in influenza A virus reveals a broad spectrum of $CD4^+$ T-cell responses to internal proteins, hemagglutinins, and neuraminidases. Hum Immunol **2009**; 70:711–21.
22. Vuola JM, Keating S, Webster DP, et al. Differential immunogenicity of various heterologous prime-boost vaccine regimens using DNA and viral vectors in healthy volunteers. J Immunol **2005**; 174:449–55.
23. Todryk SM, Pathan AA, Keating S, et al. The relationship between human effector and memory T cells measured by ex vivo and cultured ELISPOT following recent and distal priming. Immunology **2009**; 128:83–91.
24. Dong T, Stewart-Jones G, Chen N, et al. HIV-specific cytotoxic T cells from long-term survivors select a unique T cell receptor. J Exp Med **2004**; 200:1547–7.
25. Nguyen HT, Dharan NJ, Le MT, et al. National influenza surveillance in Vietnam, 2006–2007. Vaccine **2009**; 28:398–402.
26. Yang J, Gebe JA, Huston L, et al. H5N1 strain-specific hemagglutinin CD4+ T cell epitopes restricted by HLA DR4. Vaccine **2009**; 27:3862–9.
27. Kitphati R, Pooruk P, Lerdsamran H, et al. Kinetics and longevity of antibody response to influenza A H5N1 virus infection in humans. Clin Vaccine Immunol **2009**; 16:978–81.
28. McMichael AJ, Gotch FM, Dongworth DW, Clark A, Potter CW. Declining T-cell immunity to influenza, 1977–82. Lancet **1983**; 2:762–4.
29. de Jong MD, Simmons CP, Thanh TT, et al. Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. Nat Med **2006**; 12:1203–7.
30. Crowe BA, Bruhl P, Gerencer M, et al. Evaluation of the cellular immune responses induced by a non-adjuvanted inactivated whole virus A/H5N1/VN/1203 pandemic influenza vaccine in humans. Vaccine **2011**; 29:166–73.
31. Mbawuike IN, Piedra PA, Cate TR, Couch RB. Cytotoxic T lymphocyte responses of infants after natural infection or immunization with live cold-recombinant or inactivated influenza A virus vaccine. J Med Virol **1996**; 50:105–11.
32. Octaviani CP, Ozawa M, Yamada S, Goto H, Kawaoka Y. High level of genetic compatibility between swine-origin H1N1 and highly pathogenic avian H5N1 influenza viruses. J Virol **2010**; 84:10918–22.
33. Webster RG, Askonas BA. Cross-protection and cross-reactive cytotoxic T cells induced by influenza virus vaccines in mice. Eur J Immunol **1980**; 10:396–401.
34. Shinya K, Ebina M, Yamada S, Ono M, Kasai N, Kawaoka Y. Avian flu: influenza virus receptors in the human airway. Nature **2006**; 440:435–6.

Downloaded from http://jid.oxfordjournals.org/ at Oxford University on December 28, 2011

Appendix B

R code

```
# R-script for creating figure 2 in research paper 4.

# Proportion of cases occurring in household clusters by probability of
    infection for different household sizes.

# Clear workspace
#-------------------------------------------------------------
rm(list = ls())
library(Hmisc)
#-------------------------------------------------------------
# Set to output two graphs side by side
par(mfrow = c(2, 2))

#-------------------------------------------------------------
# SECTION 1 - DEFINE PARAMETERS

# tau is probability of infection (assumed equal for everyone)
tau.vector.length<-100
tau.vector<-seq(0, 1, length=tau.vector.length)

# Maximum nine persons per house in Ha Nam
n.vector<-(1:9)

# Total number of houses in Ha Nam study
no.house.category<-c(231,342,533,541,334,111,16,8,8)
# Proportion of households of size n defined by qn
qn<-c(no.house.category/sum(no.house.category))
qn
# [1] 0.108757062 0.161016949 0.250941620 0.254708098 0.157250471
# [6] 0.052259887 0.007532957 0.003766478 0.003766478

# average house size
sum(qn*(1:length(qn)))
# [1] 3.419021

#-------------------------------------------------------------
# SECTION 2 - DERIVE DATA FOR EACH HOUSEHOLD SIZE
```

```
# Loop through 'n.vector' in the same way.
proportion.in.clusters<-matrix(NA, nrow=length(tau.vector), ncol=length
    (n.vector))

for(i in 1:length(n.vector)){
   n        <-n.vector[i]

   for(j in 1:length(tau.vector)){
      # distribution of number of cases per house
      dist.cases.per.house<-dbinom(0:n, n, tau.vector[j])

      # average number of cases per house
      cases.per.house<-n*tau.vector[j]
      # (total number of cases is cases per house times m)

      # total number of cases in cluster houses is m times sum of (
          distribution times 2:n)
      # (clusters are houses with 2 or more cases)
      cluster.cases.per.house<-sum(dist.cases.per.house[3:I(n+1)]*I(2:n
          ))

      # proportion of cases in clusters
      proportion.in.clusters[j,i]<-cluster.cases.per.house/cases.per.
          house
}
}

# Set NA to equal zero - so proportion in clusters is zero when
    household size is 1
proportion.in.clusters[is.na(proportion.in.clusters)] <- 0

#--------------------------------------------------------------------
# SECTION 3 - SIMULATION TO DERIVE MEDIAN AND PREDICTION LIMITS

# size (number of houses) of village to simulate denoted by N
N<-2000

# Number of households of size n (unrounded)
Nqn<-c(N*qn)

# Number of simulations
 n.simulations<-10000

# Prepare matrix to accept data
```

```
proportion.in.clusters.simulation<-matrix(NA, nrow=length(tau.vector),
    ncol=3, dimnames=list(NULL, c('2.5 percentile', 'median', '97.5 
    percentile')))

# Number of households of size n rounded
Nqn.round<-round(N*qn)
# the rounding in the above may result in a total number of houses not
    equal to the specified number
# most common house size is 4; ensure the total is maintained by
    modifying this element
Nqn.round[4]<-N-sum(Nqn.round[c(1:3, 5:length(Nqn.round))])
# Check
sum(Nqn.round)
[1] 2000

# Prepare vector to accept data for simulation results
proportion.in.clusters.vector<-rep(NA, n.simulations)

for(i in 1:length(n.vector)){
   n        <-n.vector[i]

   for(j in 1:length(tau.vector)){
      # distribution of number of cases per house
      dist.cases.per.house<-dbinom(0:n, n, tau.vector[j])

      # simulate village
      # don't need to do this for every fixed house size
      # just do it for the first one
      if(i==1){
         if(tau.vector[j]==0){
            proportion.in.clusters.simulation[j,]<-c(0, 0, 0)
         }else{

            proportion.in.clusters.vector<-rep(NA, n.simulations)

            for(k in 1:n.simulations){
               house.sizes.long.simulation<-rep(1:9, Nqn.round)
               cases.by.house.simulation <-rbinom(
                  n    =length(house.sizes.long.simulation),
                  size=         house.sizes.long.simulation,
                  prob=tau.vector[j])
               # print(table(cases.by.house.simulation))
               # Expected total number of cases in households of size n
                   denoted by x
                total.cases.simulation<-sum(cases.by.house.simulation)
```

```
                cluster.case.simulation<-sum(cases.by.house.simulation[
                    cases.by.house.simulation>=2])
                # proportion of cases in clusters
                proportion.in.clusters.vector[k]<-cluster.case.
                    simulation/total.cases.simulation
            }
            # print(summary(proportion.in.clusters.vector))
            proportion.in.clusters.simulation[j,]<-
                quantile(proportion.in.clusters.vector, probs=c(0.025,
                    0.5, 0.975))
        }
    }
      }
}

#-----------------------------------------------------------------
# SECTION 4 - PLOT HOUSEHOLDS DATA AND SIMULATION

plot(tau.vector, proportion.in.clusters[,1], ylim=c(0,1), xlim=range(
    tau.vector), lty=2, type='l', xlab="Probability of infection if
    exposed",
ylab="Proportion of cases occuring in household clusters")
for(i in 2:length(n.vector)){
        lines(tau.vector, proportion.in.clusters[,i],lty=2)
        }
lines(tau.vector, proportion.in.clusters.simulation[,"2.5 percentile"],
     col=26, lwd=1, lty=3)
lines(tau.vector, proportion.in.clusters.simulation[,"median"], col=26,
     lwd=2, lty=1)
lines(tau.vector, proportion.in.clusters.simulation[,"97.5 percentile"
    ],col=26, lwd=1, lty=3)

# Add minor tick marks
library(Hmisc)
minor.tick(nx=5, ny=0, tick.ratio=0.5)

#-----------------------------------------------------------------
# SECTION 5 - MARK PROB INFECTION CORRESPONDING TO 22% CLUSTERING

lower.bound.tau.median<-min(tau.vector[proportion.in.clusters.
    simulation[,"median"]<=0.22])
upper.bound.tau.median<-min(tau.vector[proportion.in.clusters.
    simulation[,"median"]>=0.22])
lower.bound.proportion.median<-max(proportion.in.clusters.simulation[,"
    median"][proportion.in.clusters.simulation[,"median"]<=0.22])
```

```
upper.bound.proportion.median<-min(proportion.in.clusters.simulation[,"
    median"][proportion.in.clusters.simulation[,"median"]>=0.22])

slope      <-(upper.bound.tau.median-lower.bound.tau.median)/(upper.
    bound.proportion.median-lower.bound.proportion.median)

interp.tau<-lower.bound.tau.median + (slope*(0.22-lower.bound.
    proportion.median))

segments(x0=0,            x1=interp.tau, y0=0.22, y1=0.22, col=555, lwd
    =2)
segments(x0=interp.tau, x1=interp.tau, y0=0,     y1=0.22, col=555, lwd
    =2)

#-----------------------------------------------------------
# Re-plot to focus on low end of probability of infection

plot(tau.vector, proportion.in.clusters[,1], ylim=c(0,0.5), xlim=range
    (0,0.2), lty=2, type='l',xlab="Probability␣of␣infection␣if␣exposed"
    ,
ylab="Proportion␣of␣cases␣occuring␣in␣household␣clusters")
for(i in 2:length(n.vector)){
   lines(tau.vector, proportion.in.clusters[,i],lty=2)
}
lines(tau.vector, proportion.in.clusters.simulation[,"2.5␣percentile"],
      col=26, lwd=2, lty=3)
lines(tau.vector, proportion.in.clusters.simulation[,"median"],
              col=26, lwd=3, lty=1)
lines(tau.vector, proportion.in.clusters.simulation[,"97.5␣percentile"
    ], col=26, lwd=2, lty=3)

#-------------------------------------------------------------
# SECTION 5 - MARK PROB INFECTION CORRESPONDING TO 22% CLUSTERING

lower.bound.tau.2.5<-max(tau.vector[proportion.in.clusters.simulation[,
    "2.5␣percentile"]<=0.22])
upper.bound.tau.2.5<-min(tau.vector[proportion.in.clusters.simulation[,
    "2.5␣percentile"]>=0.22])
lower.bound.proportion.2.5<-max(proportion.in.clusters.simulation[,"2.5
    ␣percentile"][proportion.in.clusters.simulation[,"2.5␣percentile"
    ]<=0.22])
upper.bound.proportion.2.5<-min(proportion.in.clusters.simulation[,"2.5
    ␣percentile"][proportion.in.clusters.simulation[,"2.5␣percentile"
    ]>=0.22])
```

```
slope      <-(upper.bound.tau.2.5-lower.bound.tau.2.5)/(upper.bound.
    proportion.2.5-lower.bound.proportion.2.5)

interp.tau<-lower.bound.tau.2.5 + (slope*(0.22-lower.bound.proportion
    .2.5))

segments(x0=0,           x1=interp.tau, y0=0.22, y1=0.22, col=555, lwd
    =2)
segments(x0=interp.tau, x1=interp.tau, y0=0,    y1=0.22, col=555, lwd
    =2)

lower.bound.tau.median<-max(tau.vector[proportion.in.clusters.
    simulation[,"median"]<=0.22])
upper.bound.tau.median<-min(tau.vector[proportion.in.clusters.
    simulation[,"median"]>=0.22])
lower.bound.proportion.median<-max(proportion.in.clusters.simulation[,"
    median"][proportion.in.clusters.simulation[,"median"]<=0.22])
upper.bound.proportion.median<-min(proportion.in.clusters.simulation[,"
    median"][proportion.in.clusters.simulation[,"median"]>=0.22])

slope      <-(upper.bound.tau.median-lower.bound.tau.median)/(upper.
    bound.proportion.median-lower.bound.proportion.median)

interp.tau<-lower.bound.tau.median + (slope*(0.22-lower.bound.
    proportion.median))

segments(x0=0,           x1=interp.tau, y0=0.22, y1=0.22, col=555, lwd
    =2)
segments(x0=interp.tau, x1=interp.tau, y0=0,    y1=0.22, col=555, lwd
    =2)


lower.bound.tau.97.5<-max(tau.vector[proportion.in.clusters.simulation
    [,"97.5 percentile"]<=0.22])
upper.bound.tau.97.5<-min(tau.vector[proportion.in.clusters.simulation
    [,"97.5 percentile"]>=0.22])
lower.bound.proportion.97.5<-max(proportion.in.clusters.simulation[,"
    97.5 percentile"][proportion.in.clusters.simulation[,"97.5 
    percentile"]<=0.22])
upper.bound.proportion.97.5<-min(proportion.in.clusters.simulation[,"
    97.5 percentile"][proportion.in.clusters.simulation[,"97.5 
    percentile"]>=0.22])

slope.2      <-(upper.bound.tau.97.5-lower.bound.tau.97.5)/(upper.bound.
    proportion.97.5-lower.bound.proportion.97.5)
```

```
interp.tau.2<-lower.bound.tau.97.5 + (slope.2*(0.22-lower.bound.
    proportion.97.5))

segments(x0=interp.tau.2, x1=interp.tau.2, y0=0,     y1=0.22, col=555,
    lwd=2)

# Add minor tick marks
library(Hmisc)
minor.tick(nx=5, ny=5, tick.ratio=0.5)
#--------------------------------------------------------------
```