Influenza in Vietnam: cross immunity from prior infection and its effect on population-level estimates of infection

A thesis submitted in accordance with the requirements and regulations of the

University of Liverpool for

A Degree of Doctor of Philosophy

By

Stacy Todd

MbChB (University of Aberdeen) MRCP (London) MRes (University of Liverpool)

DTM&H (Liverpool School of Tropical Medicine)

21 September 2016

Liverpool School of Tropical Medicine
UK

&

Institute of Infection and Global Health
University of Liverpool
UK

&

Wellcome Trust Major Overseas Programme – Vietnam
Oxford University Clinical Research Unit
Hospital for Tropical Diseases, Ho Chi Minh City
Vietnam

DECLARATION

Except for the assistance outlined in the acknowledgements, the work described is my own and has not been submitted for a degree or other qualification to this or any other university.

FUNDING

ABSTRACT

Introduction: Repeat infections with influenza A occur because of the continual evolution of the virus. A specific humoral response occurs after each infection and this develops into an immunological profile for each individual based on their prior exposure history and the cross-reaction between antigenically similar viruses. Understanding how this profile changes following acute infection is important for interpretation of seroepidemiological studies.

Methods: This observational study was designed to investigate non-severe influenza and influenza like illness in a tropical, urban primary care setting. A prospective, observational study of patients with ILI in Ho Chi Minh City, Vietnam has been running since August 2013. Influenza A & B PCR and antibody testing to a panel of 11 human and 5 avian strains is performed using a novel protein microarray technique. A subset of subjects are followed up clinically and serologically for 7 months. This doctoral research is designed to address questions around the effect of cross-immunity from previous influenza infection on susceptibility to currently circulating strains, understand the short term antibody dynamics of both current and historical strains post infection and determine optimal serological determinants of recent infection.

Results: 953 ILI patients were recruited between August 2013 & May 2015. 274 and 136 subjects had influenza A and B respectively. Three peaks of influenza activity were detected, H3N2 peak between April and June 2014, influenza B peak in July to December 2014 and mixed H3N2 and H1N1 peak March to May 2015. Lower baseline titre was associated with higher risk of influenza but response to H3N2 2005 and 2009 were most predictive of current susceptibility even when original antigenic sin was taken into account. Levels of protection offered by a fixed titre changed depending on force of infection.

186 ILI patients were recruited to the longitudinal study. The largest response was within subtype which peaked at around 30 days. A boosting of historic response was also seen. The response in historic strains waned quicker than response to recent strains. A smaller but significant between subtype increase in titre was also detected. Following an acute rise after infection within six months most individuals had returned to a standard rate of decline of 1 log2 titre unit each one to two years.

The optimal serological marker of recent infection was investigated using 470 sample. For the most recent H3N2 2011 strain the optimal threshold was greater than log2 5.5 for all age groups at all time points. Sensitivity was greater than 90% for all thresholds but specificity was poor. Specificity was improved by using a multi-strain approach as measured by diversity index. The measured

sensitivity and specificity will lead to a significant over estimation of influenza sero-prevalence unless test accuracy is adjusted for.

Conclusion: Influenza in southern Vietnam has complex transmission dynamics including periods of intense influenza activity. Titre rise is seen within and across subtypes which would lead to repeat boosting of titre levels across many years. Further work should be performed to establish if this titre rise is accompanied by a change in neutralisation activity. Seroepidemiology of influenza is challenging because of cross-reaction. Methods exist to improve the estimation from these methods and should be employed for non-pandemic influenza serosurveillance.

ACKNOWLEDGEMENTS

Supervision:

I would like to express my extreme gratitude to all my supervisors whose support and encouragement has been invaluable. Dr Jonathan Read (University of Liverpool) for his unrelenting enthusiasm and introducing me to the world of R; Dr Maciej Boni (OUCRU) for allowing me to join his group and run with my project ideas; and Prof David Lalloo (LSTM) for being a constant source of practicality and reminding me about the bigger clinical picture.

Finally, I would like to thank my family and especially my husband Hon-Wing Shek for his constant support through this endeavour and in particular during my time in Vietnam. Without this, it would have been a very different PhD.

ABBREVIATIONS AND NOTATIONS

ADCC	Antibody Dependent Cell Cytotoxicity
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
AUC	Area under the Curve
CI	Confidence Intervals or Credible Intervals
CONSISE	Consortium for the Standardization of Influenza Seroepidemiology
CRF	Case Record Form
CtL	Closest to the Top Left
ECDC	European Centre for Disease Control
edf	Estimated Degrees of Freedom
ELISA	Enzyme Linked Immunosorbent Assay
FOI	Force of Infection
GAM	Generalised Additive Model
GMT	Geometric Mean Titre
HA	Haemagglutinin
HCMC	Ho Chi Minh City
HI	Haemagglutination Inhibition Assay
HPAI	Highly Pathogenic Avian Influenza
ICH GCP	International Committee for Harmonisation of Good Clinical Practice
ILI	Influenza like Illness
IQR	Inter Quartile Range
NA	Neuraminidase
NT	Neutralisation Titre Assay
PCR	Polymerase Chain Reaction
PMA	Protein Microarray
PMH	Past Medical History
PMH	Past Medical History
RIVM	Dutch Institute of Public Health
RNA	Ribonucleic Acid
ROC	Receiver Operator Curve
RT-PCR	Reverse transcription polymerase chain reaction
SARI	Severe Acute Respiratory Infection
SE Asia	South East Asia
Tukey HSD	Tukey Honest Significant Difference Test
VN	Vietnam
WHO	World Health Organisation

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION AND LITERATURE REVIEW

1.1 INFLUENZA BACKGROUND

Influenza viruses circulate globally and seasonal epidemics are thought to be associated with 3-5 million severe clinical infections and 250,000 to 500,000 deaths each year (World Health Organisation, 2014c, Global Burden of Disease 2013 Mortality and Causes of Death Collaborators, 2015, Global Burden of Disease Study 2013 Collaborators, 2015). Morbidity and mortality are highest in the extremes of age (Cohen et al., 2010, Nair et al., 2011) but all ages are affected with repeat infections throughout an individual's lifetime. In temperate countries predicable winter epidemics occur with annual incidence concentrated into a two to three month peak of activity with limited or no activity recorded between these annual peaks (Finkelman et al., 2007, Nelson and Holmes, 2007, Rambaut et al., 2008) . For many decades influenza was thought not to be a disease of importance in tropical countries due to a lack of observable peak in activity and poor availability of routine surveillance. This has changed considerably in the last 10 years with influenza now recognised to be at least as big a health problem in tropical countries as it is in temperate countries (Yang et al., 2011, Fischer II et al., 2014, Ng and Gordon, 2015). In contrast to the predictable peaks in temperate countries, influenza transmission in tropical areas is less predictable with asynchronous peaks of activity, co-circulation of strains and apparent year round transmission (Moura, 2010, Cheng et al., 2013, Le et al., 2013, Nelson et al., 2014) with the drivers for this difference not yet being fully understood (Tamerius et al., 2011, Paynter, 2015).

Global movement of influenza A viruses can be tracked using phylogenetic analysis. Results to date have been far from conclusive although East and South-East (SE) Asia are likely to play a major role in influenza evolution and persistence (so called 'SE Asian source-sink' model) (Figure 1) (Cox and Subbarao, 2000, Nelson and Holmes, 2007, Rambaut et al., 2008, Russell et al., 2008, Bahl et al., 2011, Kenah et al., 2011). Influenza evolution occurs through a series of 'shifts' and 'drifts', with gradual antigenic changes occurring each year and larger jumps occurring every three to five years (Smith et al., 2004, Bedford et al., 2014). In addition to these changes the sporadic emergence of novel subtypes occur with a subsequent rapid spread globally. These pandemics are known to have occurred in 1918, 1957, 1968, 1977 and 2009. Prior to 1977 these pandemic strains replaced the previously dominant circulating subtype; however, since 1977 H1N1 and H3N2 have both continued to circulate in the human population (Webster, 2013). Influenza B types also circulate globally although will not be covered in depth in this review. The interaction between these two influenza

types are under researched and likely to be an important determinant in our overall understanding of influenza (Caini et al., 2015).
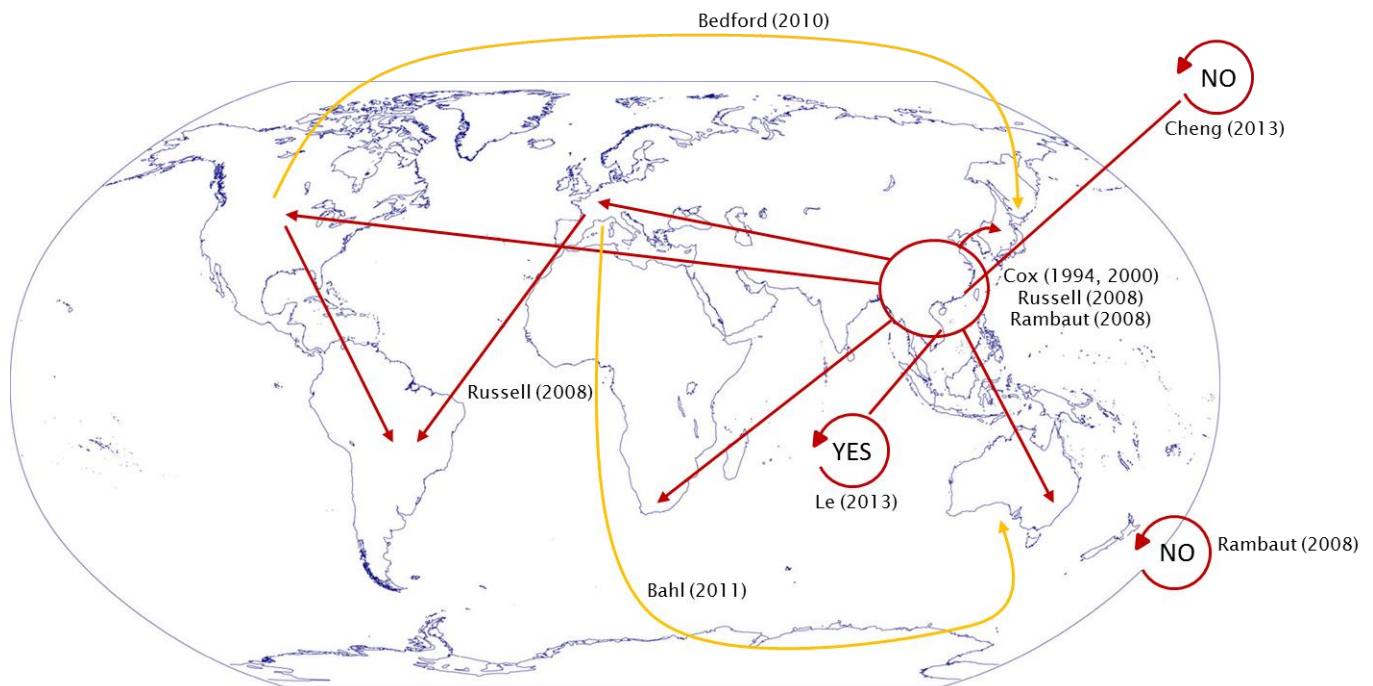
Figure 1 Transmission Routes of Influenza Viruses determined by phylogenetic analysis

Global influenza transmission dynamics are a complex system thought to be the result of interplay between the virus, immunity in the human population and environmental factors (Figure 2). In order for individuals to have repeat infections during the course of their life influenza viruses must continually evolve to evade the human immune system (Ferguson et al., 2003). Infants are generally protected by maternal antibodies in the first six months of life and in temperate regions most children have had their primary infection by six years of age (Bodewes et al., 2011). At the other end of the age spectrum older individuals have a general decline in immune function, a phenomenon known as immunosenescence (Sasaki et al., 2011, Pera et al., 2015), which could partly explain higher rates of morbidity and mortality in this age group. It is worth noting that advancing age is not always detrimental with older individuals being protected in the first wave of the 2009 pandemic due to preexisting antibodies (Van Kerkhove et al., 2013b). That antibodies to strains encountered many years prior to testing exist has been known almost as long as our ability to test for antibodies to influenza (Davenport et al., 1953, Hennessy et al., 1955, Hilleman et al., 1958). The 'original antigenic sin' hypothesis stated that the first infection an individual encountered left the most significant impact on the humoral immune system and could result in a deminished response to future infection. This now appears to be a simplification and work to understand this in the context of both individual susceptibility and population immunity continues to be explored (Kim et al., 2009,

Kucharski and Gog, 2012, Lessler et al., 2012, Miller et al., 2013, Fonville et al., 2014).
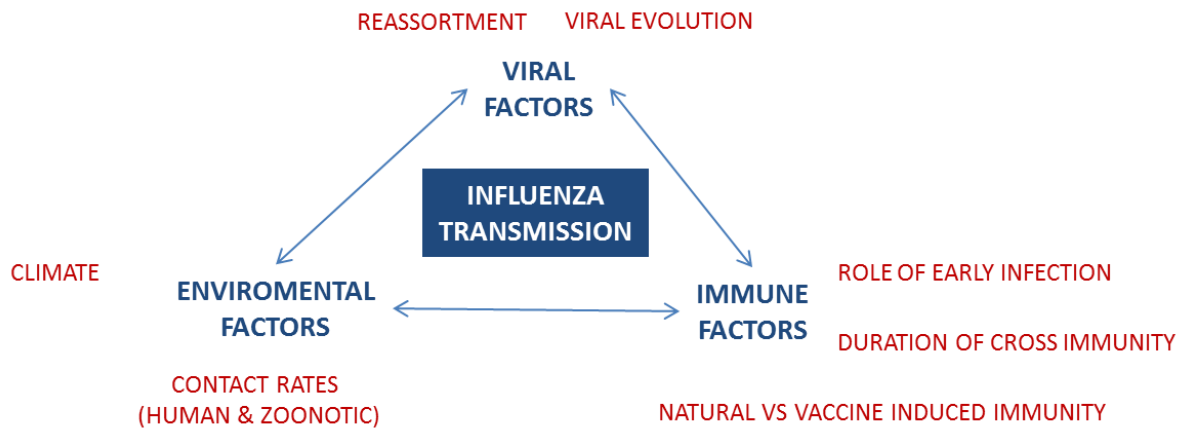
Figure 2 Ongoing influenza transmission and factors thought to play a role in global circulation.

Influenza is a vaccine preventable disease and the WHO recommends all countries consider vaccination of young children, pregnant women, those with chronic medical conditions and the elderly (World Health Organisation, 2012). Due to the continual evolution of the influenza virus the composition of the influenza vaccine must be updated regularly. WHO makes recommendations approximately eight months prior to the winter season for Northern and Southern temperate regions (World Health Organisation, 2015c) to allow time for vaccine manufacture, distribution and administration before the onset of the annual epidemic. These recommendations are based on international surveillance systems looking at the rate of evolution of the virus and analysis of the antigenic characteristics of most recently circulating viruses (World Health Organisation, 2007). This is a complex process which does not always identify the ideal vaccine composition, as seen during the 2014-2015 northern winter season with a mismatch between H3N2 vaccine component and the circulating H3N2 strain, which resulted in reduced vaccine efficacy (<25%) and an increase in severe cases (Broberg et al., 2015, Molbak et al., 2015, Pebody et al., 2015). Current global vaccination planning is designed around the timing of temperate-zone winters and previous advice for tropical countries had recommended following the nearest temperate season nevertheless this has proven to be inadequate (Mello et al., 2009, Luna and Gattás, 2010). Increasingly novel strategies are being considered to provide vaccine for regions with year round transmission (Lambach et al., 2015). Work is ongoing to develop a vaccine which can induce broadly protective antibodies which would not require annual updates (Wei et al., 2010, Jegaskanda et al., 2014, Lu et al., 2014, Lee and Wilson, 2015); however, this is many years from large scale human trials or routine use in public health. We therefore must maximise our use of existing vaccine technologies through better understanding of global influenza dynamics and as a result improve our ability to anticipate the best vaccine composition through the consideration of novel vaccine administration strategies.

Despite being recognised for over 70 years, there are a number of unanswered questions regarding the role of cross immunity in susceptibility to influenza. These include the duration of protection offered by previous influenza infection, the role of infections contracted early in life and the impact that cross immunity has on the transmission dynamics of influenza. The emergence of the H1N1 2009 pandemic strain led to a re-examination of the role of serological studies in outbreak description and how, by means of additional techniques such as mathematical modelling, results of these studies can be used in policy decisions (Van Kerkhove and Ferguson, 2012, Van Kerkhove et al., 2013a, Van Kerkhove et al., 2013b). A lack of data from tropical settings, and the need for better understanding of the effect of recent infection and cross immunity on serological profiles, was recognised. This thesis is designed, within the parameters of the study data, to address some of these deficiencies.

1.2 SEARCH STRATEGY AND SELECTION CRITERIA

References for this review were identified through searches of PubMed for articles published from January 1930 to February 2016, by use of the terms "influenza", "transmission", "serology" (and related terms through wild card searches "sero*"), "humoral immunity" and "Vietnam". Articles resulting from these searches were narrowed down by review of abstract. Older papers without abstracts were reviewed in full. Articles published in English were included. Appropriate articles focusing on the role of humoral immunity in the evolution of influenza were included. Relevant references cited in included articles were reviewed and included as appropriate.

1.3 INFLUENZA VIROLOGY

Influenza viruses are enveloped, single-stranded RNA viruses belonging to the *orthomyxoviridae* family (Webster, 2013). They circulate in mammalian and avian hosts, transcribing and replicating their genome within the nucleus of infected cells meaning that host to host transmission is required for ongoing survival. Influenza can be divided into three types (A, B and C), all of which can infect humans, although A and B are the predominant circulating types. Influenza is typically spherical in shape measuring ~100nm in diameter. It is made up of 3 components (Figure 3) i. Lipid bilayer envelope with the 3 transmembrane proteins haemagglutinin (HA), neuraminidase (NA) and ion channel (M2); ii) matrix layer with abundant sub-membrane protein M; iii) virus core with helical ribonucleoproteins (RNPs) which house the negative stranded genomics viral RNAs and nuclear proteins (NP) (Horimoto and Kawaoka, 2005, Webster, 2013). Haemagglutinin (HA) and neuraminidase (NA) surface glycoproteins are the key surface proteins responsible for the antigenic variation which allows continual transmission in humans and are also the primary targets of the human humoral immune response (Ferguson et al., 2003, Knossow and Skehel, 2006).
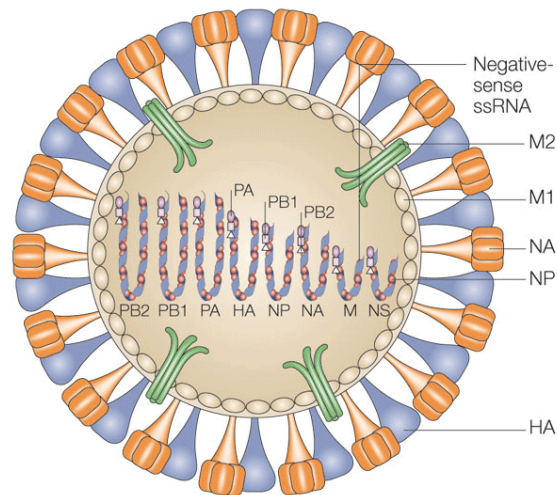
Copyright © 2005 Nature Publishing Group
Nature Reviews | Microbiology

Figure 3 Schematic of influenza viral structure. From (Horimoto and Kawaoka, 2005) Two surface glycoproteins, haemagglutinin (HA) and neuraminidase (NA), and the M2 ion-channel protein are embedded in the viral envelope, which is derived from the host plasma membrane. The ribonucleoprotein complex comprises a viral RNA segment associated with the nucleoprotein (NP) and three polymerase proteins (PA, PB1 and PB2). The matrix (M1) protein is associated with both ribonucleoprotein and the viral envelope. A small amount of non-structural protein 2 is also present, but its location within the virion is unknown.

Briefly, the viral life cycle is as follows (Skehel and Wiley, 2000, Lakadamyali et al., 2003, Rust et al., 2004, Webster, 2013): i) HA binds virus to cell-surface sialylated glycoproteins from which the bound virus is taken into cells by endocytosis mainly via clathrin-coated pits (Rust et al., 2004). NA aids cell entry by causing virus elution. ii) Following viral endocytosis an acidic environment is generated allowing the activated HA to mediate membrane fusion (Skehel and Wiley, 2000). The fusion process requires proximity between viral and cell membranes, for this to occur HA must undergo major structural changes. The cleavage of HA0 to HA1 and HA2 brings the surfaces together while preserving the globular head structure of the HA protein (Figure 4). iii) Following fusion processes M1- ribonucleoprotein (RNP) complexes are separated releasing RNP. The acidic environment opens the M2 ion channel allowing the free RNPs to be transported into the host nucleus for transcription and replication of viral RNA (Ye et al., 1999). iv) After transcription and replication the viral components are brought and assembled at the budding site, a step which involves the transmembrane envelope proteins being brought to that site (Palese et al., 1974, Nayak et al., 2009).

Current anti-influenza viral drugs act on the processes of uncoating and releasing of RNP by blocking the M2 proton channel (amantadine and rimantadine) or by preventing the spread of virus particles budding by inhibiting the enzymatic activity of NA (oseltamivir and zanamivir).
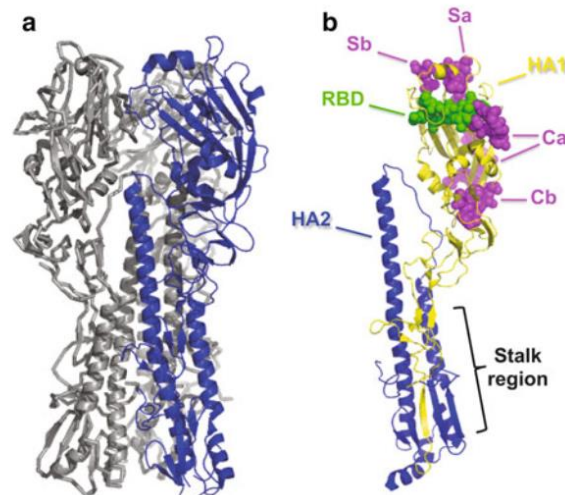
Figure 4 Haemagglutinin Structure of Influenza A/PR8/34 (H1N1). Figure from (Chiu et al., 2015). A. Depiction of hemagglutinin as a trimer (with one monomer colored blue) and B. the monomer. Each monomer is composed of two subunits: HA1 (yellow) and HA2 (blue). Within the HA1 subunit, well-defined neutralizing epitopes within the globular head region (Sa, Sb, Ca, and Cb) are shown in magenta. The receptor-binding domain (RBD) is shown in green. The HA stalk domain is formed by the N- and C-terminal domains of HA1 plus the ectodomain part of HA2.

Haemaggultinin is the most abundant of the envelope proteins accounting for more than 80% of the viral proteins on the virus's surface (Wilson and Cox, 1990), it forms trimetric spikes with receptor binding sites and epitope for neutralising antibody (Figure 4). Neuraminidase is the second most abundant envelope protein (~17%) and tends to group in clusters surrounded by HA for reasons that are currently unclear. Differences in the HA and NA protein sequence are what determine the subtype of influenza A and give rise to the subtype nomenclature (e.g. H1N1, H3N2). 18 different HA groups have been identified and 11 NA groups (Schrauwen and Fouchier, 2014). Originally this subtype classification was performed using hyper-immune sera and viral solution (World Health Organisation, 1980) but is now performed using gene sequencing and phylogenetic analysis (Rohm et al., 1996).

Both HA and NA subtypes can be classified into one of two groups phylogenetically with a further categorisation into five and two clades respectively (Nobusawa et al., 1991). Group 1 HA viruses included H1, H2 and H5 whereas H3, H7 and H9 belong to group 2. The most diverse range of combinations is found in avian species, with human epidemics being restricted to H1N1, H2N2 and H3N2, with rare transmissions of avian subtypes such as H5N1 and H7N9 (Dugan et al., 2008, Schrauwen and Fouchier, 2014). This host restriction of subtypes is related to receptor specificity of HA and NA when binding host cells. Influenza binds to sialic acid containing either α2,3 or α2,6 linkages, with avian subtypes preferentially binding the former and human and other mammalian subtypes the later (Kobasa et al., 1999) . The significance of this is clear when you consider the

primary route of infection (and onward transmission) for each species; the human upper respiratory tract (nasal and trachea) cells express predominantly α2,6 linked sialic acid and the gastrointestinal tract of birds express predominantly α2,3 (Rogers and Paulson, 1983). When a species jump of a subtype has occurred (normally through reassortment with avian strains) such as in 1957 (H2N2) or 1968 (H3N2) there has been an accompanying shift in receptor specificity from avian α2,3 to human α2,6 preference. This occurs rapidly in the HA receptors, more slowly in the NA (Kobasa et al., 1999). This receptor specificity is the reason that avian subtypes, which make repeated incursions into the human population such as H5N1 and the recent H7N9, outbreak have yet to result in sustained human to human transmission. Nonetheless, it is also likely to be a contributing factor to high rates of severe disease and mortality as the lower respiratory tract in humans (bronchiolar and alveolar tissues) expresses a roughly equal proportion of α2,3 and α2,6-linked cells meaning that any infection with α2,3 tropism is more likely to result in lower respiratory disease, which is generally more severe than upper respiratory tract symptoms (Peiris et al., 2007).

1.4   IMMUNITY TO INFLUENZA

As the first line of defence, the innate immune system produces a range of responses which aims to stop viral entry into respiratory epithelial cells. This includes the production of mucus and collectins on the epithelial mucosa surface of the upper and lower respiratory tract (Webster, 2013).   If this physical barrier is breached and epithelial cells are infected then the innate immune system is key to detecting the presence of viral RNA within the host cell. Pattern recognition receptors such as toll like receptors recognise and bind to viral RNA and produces pro-inflammtory cytokines and type 1 interferon (Pang and Iwasaki, 2011). These pathways have direct viral neutralising activity but are also important for stimulation of the adaptive immune system including dendritic cells stimulation and recruitment of B cells to regional lymph nodes early in infection.

Once viral products have entered peripheral dendritic cells then they migrate via the lymphatic system to the local draining lymph node. Viral proteins are degraded within the dendritic cells and immune-peptides (epitopes) are then presented by MHC class I or class II antigen presenting cells to CD4 and CD8 T helper cells (GeurtsvanKessel et al., 2008, Braciale et al., 2012). Stimulation of CD4 and CD8 T helper cells then result in activation of B and T cell immune cascades. Once stimulated, B cells can then undergo clonal expansion and the differentiate to plasmablasts which are relatively short lived or the stimulation of germinal centres which are full formed within a week and then last for weeks to month (Paus et al., 2006). In addition to the peripheral dendritic cells, there is increasing evidence that a specialised area within the germinal centres of lymphatic tissues (follicular dendritic cells; FDC) also presents antigen to B cells. Although not mobile, these FDC foci do not

appear to be permanent with increase in the size and expression of these centres after stimulation with inflammatory mediators. There is mounting evidence that FDC also stimulate B cells through direct antigen presentation and through activation of secondary lymphoid tissues within the respiratory tract (bronchus-associated lymphoid tissue). These germinal centres can still be detected up to 5 months post infection and it is postulated that ongoing antigen presentation by FDC is the driver for this, helping to invoke long lasting immune responses. Longer lasting plasma cells are generated in the lungs and bone marrow during this time and persist for years. It is this memory response which is relied upon in the event of re-exposure to influenza. The rapid response of memory B cells has been demonstrated with responses often detected within days of infection with a homotypic virus (Wrammert et al., 2008). Memory B cells demonstrate high affinity for antigen and rapidly differentiate to antibody secreting cells after exposure to these antigens and are thought to be a significant part of the early plasmablast wave seen in the days after infection. As well as being a specific immune response on exposure to homotypic virus, there is increasing evidence of broadly protective responses in both seasonal and pandemic situations (Ekiert et al., 2009, Li et al., 2012).

Once antibodies have been produced they have a number of targets. HA-specific antibodies bind to the globular head of the HA protein stopping viral attachment and cell entry. Immunogenic sites have been identified in the region surrounding the receptor binding sites (Whittle et al., 2011, Koel et al., 2013). Anti-HA antibodies have also been implicated in antibody dependent cell cytotoxicity (ADCC) where cells which have viral envelope proteins present on the cell surface (e.g. during process of cell fusion) can be bound by antibody and natural killer cells induce phagocytosis via Fc receptor expressing cells (Jegaskanda et al., 2013). Anti-HA antibodies have also recently been detected against the conserved stalk region which is formed through cleavage of part of the HA1 and all of the HA2 polypeptides (Figure 4). Antibodies targeted at this region are able to prevent viral fusion with the human cell membrane (Lu et al., 2014). Antibodies to other viral proteins have also been identified but tend to get less attention compared to HA antibodies. NA specific antibodies stop the cleavage of the newly formed virus particle from the cell surface preventing the release and spread of the virus (Marcelin et al., 2012). NA antibodies also facilitate ADCC and may have a role in limiting disease severity. Antibodies to NP and M2 proteins of the influenza virus appear to have a role in limiting replication efficiency rather than stopping altogether (Song et al., 2011). However both of these antigens are important targets for the cellular immune system

1.5 SEROEPIDEMIOLOGY BACKGROUND

Serological surveys are a long standing method of investigating infectious disease epidemiology with two major goals i) determination of disease incidence and ii) assessment of susceptibility to disease (Van Kerkhove et al., 2013a). However, many of the existing statistical methods rely on assumptions of life-long immunity following infection which does not hold in influenza in non-pandemic situations (Hens et al., 2012). Following the emergence of H1N1 2009 pandemic strain, age-specific seroprevalence studies were widely performed demonstrating both differences in attack rates in different age groups and different levels of pre-existing immunity (Van Kerkhove et al., 2013b). In the post-pandemic period there has been review of the methods used and discussions around how this can be improved for future pandemic preparedness, as well as in relation to seasonal influenza (Cauchemez et al., 2012, Wu et al., 2014). Understanding of the humoral immune response is vital to achieving these aims both in the context of recent infection and lifetime of infections.

Measuring the type and strength of an individual's immune response forms the basis of the population-level measurements that result from seroepidemiological studies. Neutralisation Titres (NT) and Haemagglutination Inhibition (HI) assays are WHO reference standards with recognised correlates of protection (HI ≥ 40 gives 50% protection against infection) or diagnosis (4-fold increase in paired samples) (World Health Organisation, 2011). However, they have a number of limitations, including lack of reproducibility between laboratories, because of inter–observer differences in reading of HI and quality and type of animal red blood cells used in the process (Laurie et al., 2015a). In order to test for multiple subtypes and strains a large volume of serum is required along with the storage and usage of multiple viruses. Enzyme Linked Immunosorbent Assay (ELISA) has been investigated as a virus free approach however it has had problems with cross reactivity which limited its usefulness (Murphy et al., 1981). With these limitations of existing techniques in mind a novel protein microarray was developed by the Dutch Institute of Public Health (RIVM) to measure antibodies to the HA1 subunit of the influenza virus. This method benefits from high throughput testing to multiple strains using only minimal volume of serum. Published work to date has focused on response to the pandemic H1N1 2009 (Koopmans et al., 2011, Baas et al., 2013, Huijskens et al., 2013, de Bruin et al., 2014, te Beest et al., 2014a, te Beest et al., 2014b).

Seroepidemiology employs a number of study designs to assess immunity at either the individual or population level. These include large cross sectional surveys at the end of a winter season in temperate countries (Fragaszy et al., 2015) or through longitudinal serological sampling, either in a specific observational cohort (Chen et al., 2010, Horby et al., 2012) or convenience sampling of repeat blood donors (McVernon et al., 2010, Wu et al., 2010, Sauerbrei et al., 2014). These methods

have often been developed and employed in the face of an emerging pandemic where there is limited prior immunity and therefore most of the population have undetectable antibodies making seroprevalence curves easier to estimate (Van Kerkhove et al., 2013b). During the most recent pandemic in 2009, there was suggestion that the standard threshold of 1:40 was too high and was underestimating the true community attack rate (Cauchemez et al., 2012, Wu et al., 2014). However, most influenza infections do not occur during a pandemic period and most seroepidemiology studies are performed to assess attack rates of subtypes which have caused multiple infections during a lifetime.

1.6   INFLUENZA IN VIETNAM

Vietnam is a low middle income country in SE Asia. With a population of over 93 million, the majority of whom are working age, and increasingly living in urban environments, it has undergone massive cultural and economic change in the last 20 years (CIA World Factbook, 2015). Health expenditure accounts for 6% of GDP and public provision of healthcare is structured around strict institutional hierarchies (World Health Organisation & MOH Vietnam, 2012). In addition to public healthcare provision there is a private market which mostly consists of small clinics which many urban Vietnamese use as the first point of healthcare contact. A National Influenza Surveillance System (NISS) has been running since 2006 and is coordinated by the National Institute of Hygiene and Epidemiology (NIHE) within the Ministry of Health. This operates using sentinel sites across the country, within the public healthcare system, four central hospitals, two provincial hospitals, seven district hospitals, and two polyclinics. Symptomatic surveillance for ILI and SARI is conducted (Nguyen et al., 2009, Nguyen et al., 2013) along with virus subtyping and sequencing as part of the WHO Global Influenza Surveillance Network (Le et al., 2013).

Year round transmission of influenza has been detected in Vietnam through both national surveillance and other epidemiological studies. Simultaneous circulation of multiple influenza strains and types has been documented (Nguyen et al., 2007, Li et al., 2008, Nguyen et al., 2009, Horby et al., 2012, Le et al., 2013) along with patients having more than one influenza infection in one season (Horby et al., 2012). Previous serological work in Vietnam has included small scale serological surveys looking for evidence of avian influenza exposure (Liem and Lim, 2005, Uyeki et al., 2012, Boni et al., 2013, Todd et al., 2014) and a larger prospective cohort (Cauchemez et al., 2012, Horby et al., 2012, Fonville et al., 2014).

15-20% of patients presenting to hospitals with influenza-like illness (ILI) have virologically confirmed influenza (Nguyen et al., 2009, Yoshida et al., 2010, Nguyen et al., 2013) and influenza is thought to

be associated with up to 14% of community acquired pneumonia presenting to hospital (Takahashi et al., 2013). However, complementary data on community ILI is limited. As national surveillance programmes and previous research have focused largely on hospital sites, incidence and burden estimates from these studies are likely to underestimate true clinical burden and attack rates in the community.

Seasonal influenza vaccine is not currently recommended in Vietnam (United Nations in Vietnam, 2009) and there is minimal, although increasing, use in the private sector (Palache et al., 2014). Vaccine production capacity has been developed over the past decade in Vietnam, due to the great risk posed by highly pathogenic avian influenza H5N1 (HPAI) (Monto et al., 2011). To date work on Vietnamese influenza vaccination strategies has focused on regional distribution of poultry vaccines based on prevalence and localized outbreaks of HPAI. Research and production capacity has been expanded to produce human pandemic influenza vaccines in Vietnam, although regular production is yet to begin (Hoa et al., 2011, PATH, 2012). The optimal vaccination strategy for seasonal influenza in Vietnam is still to be established (Vuong et al., 2012a).

1.7 THESIS AIMS AND OBJECTIVES

The work presented in this thesis is part of a wider body of research which aims to understand the effect of cross-immunity from previous infections on susceptibility now and how this impacts on transmission of influenza at the population level within Vietnam. This thesis presents the results of an observational study of individuals presenting with influenza-like-illness to primary care clinics in Ho Chi Minh City, Vietnam.

As discussed, there are challenges involved with the use of serology for epidemiology in conditions where you have repeat infections. In order to address these we need to separate the two major aims of seroepidemiology, i) identifying recent infection and ii) making an assessment of population immunity. The second of these aims can again be separated into a knowledge of i) what levels of serological titre are associated with increased risk of infection and ii) after recent infection how does antibody decline to reach susceptibility again. All of this is to be in the context of recent and historical strains. The primary aims of the doctoral research presented in this thesis is therefore to address questions around the effect of cross-immunity from previous influenza infection on susceptibility to currently circulating strains, understand the short term antibody dynamics of both current and historical strains post infection and determine optimal serological determinants of recent infection.

2 COMMON METHODS

2.1 BACKGROUND

As detailed in Chapter 1, this doctoral research is designed to address questions around the effect of cross-immunity from previous influenza infection on susceptibility to currently circulating strains, understand the short term antibody dynamics of both current and historical strains post infection and determine optimal serological determinants of recent infection. This chapter will detail the common methods used throughout this work, principally the prospective observational studies and the laboratory techniques for virological and serological testing. Analytical and statistical methods will be described in each of the subsequent data chapters.

The work presented in this thesis is part of a wider body of work designed to address questions around influenza A transmission in central and southern Vietnam and the role of serology (Figure 5). A combination of research methodologies are being utilised including i) spatio-temporal analysis of influenza serology at a population-level; ii) a clinical observational study of influenza-like-illness; iii) mathematical modelling of influenza transmission dynamics. The clinical observational study which is the predominant source of work presented in this thesis was designed, implemented and analysed by myself with the exception of laboratory analysis detailed in the 'Acknowledgements'. The seroepidemiology study was designed by my supervisor Dr Maciej Boni in conjunction with Prof Jeremy Farrar and Prof Marion Koopmans following the emergence H1N1 2009 pandemic strain. This dataset was used in the development of the study design and sample size calculation of the clinical observational study. I conducted the primary analysis of the seroepidemiology dataset for this purpose.

	OUCRU-HCMC	Shared Methods	Other
Population-level Estimates: Influenza Incidence and Transmission	Seroepidemiology of Influenza in central and southern Vietnam (02FL)	PMA	Ha Nam cohort (northern Vietnam, run by OUCRU-Hanoi)
Primary Care: Syndromic Surveillance	GP ILI reporting using mobile telephones (03FL)	ECDC ILI definition, Molecular testing protocol	
	Clinical, demographic and serological features of individuals with mild ILI (10FL)	ECDC ILI definition, Molecular testing protocol, PMA	
Secondary Care: Severe Acute Respiratory Illness	Management of H1N1 2009 in early pandemic phase	WHO SARI/ILI definition, Molecular testing protocol	NIHE National Influenza Surveillance System
	Management of avian influenza cases	WHO SARI/ILI definition, Molecular testing protocol, PMA	

Figure 5 Influenza Clinical Research at OUCRU-Vietnam. Methodologies shared across studies conducted through Ho Chi Minh City are detailed. This thesis comprises work produced through the 10FL study highlighted in red.

2.2   CLINICAL OBSERVATIONAL STUDY OF INFLUENZA-LIKE ILLNESS IN HO CHI MINH CITY, VIETNAM (10FL)

2.2.1   Summary of Study Design

This observational study (10FL) enrolled subjects aged between 10 and 70, attending outpatient clinics and community medical practitioners in Ho Chi Minh City (HCMC) with influenza like symptoms to provided epidemiological, virological and immunological data.

Patients with ILI symptoms (as defined in section 2.2.3.2) were invited to participate in the study. Interested patients went through an informed consent process prior to any study interventions being performed. Once informed consent was obtained, nose and throat swabbing (respiratory sampling) was performed to test for influenza A & B PCR (by subtype) and 5ml venous blood was collected to test for 16 historical influenza antigens by protein microarray (Table 2.4). A baseline questionnaire collecting information on current illness, chronic medical problems and socio-demographic factors including household structure, occupation and exposure to animals was collected for each recruited subject. Information on participant's social contacts from the previous day was collected; this anonymised contact information was restricted to the number of contacts

made in particular age categories. Copies of the Case Record Forms (CRF) used in the study can be found in Supplementary Appendices.

A subset of patients were invited to join a longitudinal sub-study. Subjects attended for repeat serum sampling at 30, 90, 150 and 210 days (+/- 5 days) after recruitment. Active respiratory symptom surveillance was conducted during the follow-up period with telephone follow-up at 60, 120 and 180 days (+/- 5 days). Subjects were asked to attend clinic within five days of onset of new ILI symptoms where repeat nasal and throat swabbing was performed if they met ECDC ILI case definition.

Clinical characteristics of the ILI episode was reported by subjects to study staff, any subsequent medical or hospital attendance were reported by subject at next visit. No inpatient data was collect on subsequent hospitalisation. Contact data relied on participant's recollection of the previous day. No additional objective information was available from the subject's usual treating clinician.

2.2.2   Collaborating sites

Primary healthcare services within Vietnam are still developing and most provision is delivered by hospital outpatient departments and private clinics (World Health Organisation & MOH Vietnam, 2012). Study sites were identified through an existing General Practitioner ILI Surveillance Network and existing hospital collaborations within HCMC (Lam et al, Manuscript in Review; (Oxford University Clinical Research Unit, 2015)). The two hospital outpatient sites (Hospital for Tropical Diseases and Cho Ray Hospital) recruited patients attending for review of acute ILI symptoms that did not require admission. Four private clinics recruited outpatients attending with acute ILI. All collaborating sites and staff were trained in study protocol and procedures as well as receiving ICH GCP training.

2.2.3   Study Participants

2.2.3.1   Overall Description of Study Participants

Participants with influenza like illness (ILI) between the ages of 10 and 70 inclusive were recruited from collaborating sites in HCMC. Study participants were identified in three ways:

i.   Individuals who attend clinic for medical advice who fulfil the ILI case definition

ii.   Relatives/Guardians of individuals who attend clinic for medical advice who also fulfil the ILI case definition

iii.   Contacts of study participants who subsequently develop symptoms which fulfil the ILI case definition

Participants in groups i) and ii) were identified by treating clinicians at the recruiting clinics and referred to our study team. Participants in group iii) contacted the study team through a telephone number provided to original study participants. There was no obligation for participants to recruit household members and written information regarding this process was provided at the time of enrolment.

A subset of patients were recruited to the longitudinal sub-study. All subjects testing positive for influenza A were eligible to join the sub-study. Negative control subjects were matched by age (+/- 5 years) and gender to influenza A positive subjects included in the follow-up study. Those who were agreeable and able to commit to the follow-up visits were invited by the study team to return for follow up.

2.2.3.2 Inclusion Criteria

- Age 10-70 years of age inclusive [1]
- European CDC Influenza Like Illness (European Centre for Disease Prevention and Control (ECDC), 2015):
    - Sudden onset of symptoms
    - At least one of: Fever or feverishness (chills); Malaise; Headache; Muscle pain
    - At least one of: Cough; Sore throat; Shortness of breath
- Illness onset within the previous 72 hours
- Willing to participate in the study
- Written informed consent

[1] Recruitment was originally restricted to ages 10–50. The original upper age bracket was selected because of concerns regarding interpretation of the protein micro-array in older age groups. Subsequent analysis of the complementary seroepidemiology population data set which was using the same protein microarray revealed that this was unlikely to be an issue. Increasing our age band allowed an increase in overall study numbers. The upper band was increased in March 2014. The lower age band was selected to ensure that individuals had at least one previous influenza infection (Bodewes et al., 2011).

ECDC definition of ILI was used as the inclusion criteria for this study as this was already in use at several study sites which were part of an existing GP ILI Surveillance Network (Figure 5). This surveillance project reported cases of ILI and overall attendances via mobile phone networks. The major difference in the ECDC definition compared with WHO or CDC definition is the lack of a defined temperature threshold, instead relying on self-reported fever. This improves the sensitivity of the clinical criteria but with an expectant drop in specificity. There is an ongoing international

discussion on which is the optimal definition but an important consideration is the need for consistency in case identification and reporting (Jiang et al., 2015, Priest and Kelly, 2015). In order to have consistency between the studies being conducted within our research group it was agreed to continue using ECDC definition.

2.2.3.3 Exclusion Criteria

- Use of long term immunosuppressive drugs as reported by the patient.

2.2.4 Study Procedures

STUDY VISITS	DESCRIPTION	STUDY PROCEDURES	ESTIMATED NUMBER OF PARTICIPANTS	APPROX TIME
Visit 1	Patients with ILI Symptoms	1. Informed Consent 2. Enrolment Questionnaire 3. 1x Nasal Swab & 1x Throat Swab for Influenza A & B PCR 4. 1x 5ml Venous Blood Collection	1500-1600	15 min
Visit 2 (Day 30) Visit 3 (Day 90) Visit 4 (Day 150) Visit 5 (Day 210)	All subjects recruited to follow up sub study	1. 1x 5ml Venous Blood Collection 2. Follow-up Questionnaire	250	15 min
Respiratory Surveillance Telephone Call (Day 60, 120 & 180)	All subjects recruited to follow up sub study	1. Surveillance Questionnaire	250	5 min
Unscheduled Visit (between Day 30 – 210)	Subjects recruited to sub study with new ILI symptoms	1. 1x Nasal Swab & 1x Throat Swab for Influenza A & B PCR 2. Follow-up Questionnaire	20	15 min

Table 2.1 Summary of Study Procedures

Study documentation (Informed Consent Form, Clinical Record Form) is included in Appendix A.

2.2.5 Statistics and Analysis

Details of statistical analysis methods used in this thesis are detailed in appropriate chapter methods sections.

2.2.5.1 Sample Size Calculation – Primary Study

The sample size required to meet the primary objective (risk of current infection with influenza A given previous infection history) was calculated using multinomial likelihood. From preliminary work on the 02FL seroepidemiology data (see section 2.3), H3N2 strains (Table 2.4) appear to cluster in to three groups according to year of emergence a) 2003 & 2005, b) 2007 and c) 2009 & 2011 using

visual inspection of titre response, Pearson's correlation and basic Principle Components Analysis of titre response. From these clusters it was decided to define five categories based on magnitude of titre response across these groupings:

- i)   No evidence of recent infection (a, b & c all low)
- ii)   2003 & 2005 only (a only)
- iii)   2007 only (b only)
- iv)   2009 & 2011 only (c only)
- v)   (2003 & 2005) & (2007) OR (2009 & 2011) (2 from a, b or c)

It is presumed that group (i) should be the most susceptible to influenza infection in 2013-2015, while group (v) should be the least susceptible. The null hypothesis is that each of these groups is equally susceptible, and this corresponds to assigning each group a relative probability of 20% of being infected (relative to the other groups). Four alternative hypotheses ($H_{1:1}$ to $H_{1:4}$) in which the relative susceptibilities are highest in group (i) and lowest in group (v). In other words, for each of these groups, a prior probability for the likelihood of acquiring influenza A was defined as an alternative hypothesis (Table 2.2). For each of the alternative hypotheses and a given sample size (tested between 100 and 450) the Chi Square Test Statistic was calculated by sampling from a multinomial distribution with the five probabilities listed in Table 2.2, and a likelihood ratio test was used to determine if the null hypothesis could be rejected in favour of the alternative hypothesis $H_{1:j}$. This simulation was repeated 1000 times. A cut off χ^2 statistic for significance was defined as 13.28 (99% significance, df=4), the proportion of simulated runs which reached significance was recorded for each sample size.

		Titre Response Category				
		i)	ii)	iii)	iv)	v)
EQUALLY LIKELY TO ACQUIRE THIS CURRENT INFECTION REGARDLESS OF PREVIOUS INFECTION PATTERNS	H_0	0.2	0.2	0.2	0.2	0.2
MORE LIKELY TO ACQUIRE THIS CURRENT INFECTION GIVEN PREVIOUS INFECTION PATTERNS	$H_{1:1}$	0.3	0.25	0.2	0.15	0.1
	$H_{1:2}$	0.25	0.2	0.2	0.2	0.15
	$H_{1:3}$	0.25	0.25	0.2	0.2	0.1
	$H_{1:4}$	0.25	0.25	0.2	0.15	0.15

Table 2.2 Prior Probability for Likelihood of Acquiring Influenza A

A sample size of 350 is sufficient to give 80% power to reject the null hypothesis with probabilities $H_{1:1}$, $H_{1:3}$ and $H_{1:4}$ (Figure 6). Hypothesis $H_{1:2}$ is very close to the null hypothesis and would need a very large sample size to be rejected. In order to recruit 350 participants with confirmed influenza A, we estimate approximately 1500-1600 participants with ILI need to be recruited. This is based on previous results from Vietnam where 15-20% of ILI cases test positive for influenza (Horby et al., 2012). Sample size calculations were performed using R Statistical Software v2.15.2 (R Core Team, 2015).
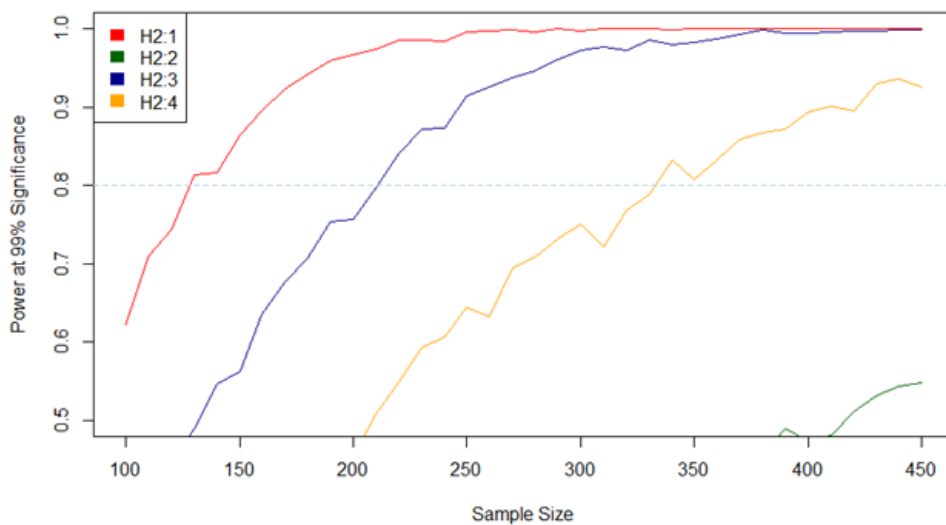
Figure 6 Sample Size Calculation for Observational Study

2.2.5.2   Sample Size Calculation – Longitudinal Sub Study

The longitudinal sub-study was designed to measure the rate of waning for antibody responses boosted by recent infection; this will be particularly relevant for the most recent H1 or H3 strains present on the microarray. Sample size estimates were performed using simulated data sets and maximum likelihood methods based on preliminary work on the 02FL seroepidemiology data (see section 2.3) and in-house comparisons between protein microarray titres and HI tests. We assumed a 100-day 'titre half-life', which is in line with VN-MOP data obtained during the 2009 pandemic (unpublished). A minimum sample size of 50 will allow the statistical identification of the titre half-life with a 95% confidence width range of between 42 – 61 days. This is robust to variation in assumed titre half-life up to 150 days. A sample size of 75 gives a titre half-life estimate with a 95% CI width between 39 and 42 days. A sample size of 75 was selected for the sub study to allow for loss to follow-up and improved precision.

The longitudinal study was expanded after six months recruitment (March 2014).   Analysis of 02FL study (see section 2.3) which utilised the same protein micro-array for antibody measurement demonstrated marked heterogeneity in response (unpublished). This included broad responses where titres are similar across the majority of strains and other narrower responses with high titre to one or two strains only. It was not known if this represented a difference in time since last infection, a difference in individual host response or other factors.   The extension to the sub study was designed to help address some of this uncertainly. An influenza A & B negative control group was added which was age and sex matched to the influenza A positive arm of the extended sub study (birth year +/- 5 years). The sample size was been increased from 75 to 125 to maintain statistical power to detect titre half-life in the event that the different antibody response groups have different waning rates.

2.2.6   Ethics and Funding

Both studies were approved by the Scientific and Ethical Committee of the Hospital for Tropical Disease, Ho Chi Minh City, Vietnam and Liverpool School of Tropical Medicine Research Ethics Committee, UK. Letters of agreement supporting the involvement of the community medical clinics were obtained from the Ho Chi Minh City Department of Health.

This study was funded by the Wellcome Trust Clinical PhD Fellowship awarded to Stacy Todd and Liverpool School of Tropical Medicine (Award 097465/B/11/Z).

2.3   SEROEPIDEMIOLOGY OF INFLUENZA A IN SOUTHERN AND CENTRAL VIETNAM (02FL STUDY)

As discussed in the chapter background, this thesis is part of a series of integrated studies aiming to address questions around the transmission of influenza A in central and southern Vietnam (Figure 5). Although no results from the population seroepidemiology study are presented in this thesis provisional results were used in the development of the study methods for the clinical observational study. An overview of the study methods of the population study are therefore given here.

2.3.1   Summary of Study Design

Since 2009, age-stratified sera from 10 hospitals in central and southern Vietnam were prospectively collected for the purpose of measuring influenza antibody titres. 200 samples were collected every two months from each site, simplified demographic information are available for each sample (age, sex, date of collection, hospital department).

2.3.2   Collaborating Sites

Data from four sites was used for the analysis contained in this thesis, Hospital for Tropical Diseases (Ho Chi Minh City), and Khanh Hoa, Dak Lak and Hue Provincial Hospitals. These sites covered a range of geographic regions from urban (Ho Chi Minh City), coastal semi-urban (Khanh Hoa), Central Highlands (Dak Lak) and Central Coast (Hue) with associated differences in socioeconomic status, living conditions, population density, access to healthcare, and potential exposure to avian influenza viruses.

2.3.3   Study Participants

Residual samples from routine biochemistry and haematology laboratories were included. The serum samples were intended to represent the general population in each hospital's catchment region. For each collection period each site provided a minimum of 50 samples in each age bands i) 0-19, ii) 20-39, iii) 40+. Samples from wards dedicated to HIV care were excluded from this study. Samples from dedicated respiratory wards as well as general wards were included in this study. Seasonal influenza vaccination in Vietnam was thought to be uncommon at the commencement of the study (<0.5% (Palache et al., 2014)) based on confidential sales data from Sanofi-Pasteur and GlaxoSmithKline who provide the human influenza vaccine available at private clinics in Vietnam. All samples were anonymized and unlinked to original hospital ID.

2.3.4   Ethics and Funding

The research protocol was approved by the Oxford Tropical Research Ethics Committee at the University of Oxford, and the Scientific and Ethical Committee of the Hospital for Tropical Diseases in HCMC.

This work was supported by the Wellcome Trust (098511/Z/12/Z, 089276/B/09/7, 084368/Z/07/Z), the British Medical Association (HC Roscoe 2011) and the Dutch Ministry of Economic Affairs, Agriculture, and Innovation, Castellum Project.

2.4   VIROLOGICAL TESTING FOR INFLUENZA A & B

Virological testing was performed by two laboratory technicians, Nguyen Thanh Hung and Nguyen Ha Thao Vy.

Molecular diagnostic techniques are commonly used for clinical detection and identification of infecting influenza viruses. Reverse-transcription polymerase chain reaction (RT-PCR) was utilised in this study for i) identification of influenza A or B infection and ii) subtyping of identified influenza A viruses. Respiratory samples were collected at the time of clinical symptoms in viral transport

medium (VTM), initially stored at -20°C within 24 hours of collection and then and stored at -80°C after initial PCR testing. RT-PCR was performed in accordance with international protocols (World Health Organisation, 2014b). Positive and negative controls were provided in house at OUCRU, Vietnam.

In brief, Viral RNA extraction was performed using commercial kit (MagNA Pure 96 DNA and Viral NA Small Volume Kit). Commercially available PCR Master Mix (LightCycler 480 RNA Probe Master, Roche, USA) and primers were used (Table 2.3). All laboratory staff had appropriate technical and ICH-GLP training.

Timing	Type/Subtype	Primer
Initial RT-PCR Testing	FluA_F (CDC)	GACCRATCCTGTCACCTCTGAC
	FluA_R (CDC)	AGGGCATTYTGGACAAAKCGTCTA
	FluA_probe (CDC)	TGCAGTCCTCGCTCACTGGGCACG
	FluB_F (CDC)	TCC TCA ACT CAC TCT TCG AGC G
	FluB_R (CDC)	CGG TGC TCT TGA CCA AAT GG
	FluB_Probe (CDC)	CCA ATT CGA GCA GCT GAA ACT GCG GTG
Secondary Subtype Testing	H3_F (CDC)	AAGCATTCCYAATGACAAACC
	H3_R (CDC)	ATTGCRCCRAATATGCCTCTAGT
	H3_Probe (CDC)	FAM-CAGGATCACATATGGGSCCTGTCCCAG
	SwPanH1_F	GTTACCCAGGAGATTTCATCGA
	SwPanH1_R	CATGCTGCCGTTACACCTTTG
	SwPanH1_Probe	AAGTTCATGGCCCAATCATGACTCGA

Table 2.3 Influenza RT-PCR Primers

2.5 SEROLOGICAL TESTING FOR INFLUENZA A

2.5.1 Background

Serology for influenza has been a cornerstone of clinical diagnosis and epidemiology for over 50 years. Neutralisation Titres (NT) and Haemagglutination Inhibition (HI) assays are WHO reference standards with recognised correlates of protection (HI ≥ 40 gives 50% protection against infection) or diagnosis (4-fold increase in paired samples) (World Health Organisation, 2011). However these longstanding methods have a number of limitations has led to the development of a novel protein micro array method to allow the measurement of titre response to multiple strains using only minimal volume of serum (Koopmans et al., 2011, Baas et al., 2013, Huijskens et al., 2013). This technique was developed by the Dutch Institute of Public Health (RIVM) who custom produced the slides used in this and the 02FL study.

2.5.2   Protein Microarray Slide Printing and Preparation

Microarray slides were printed for this project by Erwin de Bruin at the Dutch Institute for Public

Health and the Environment (RIVM).

Commercially available HA1 proteins of 16 influenza strains (Table 2.4) were produced in human embryonic kidney cells (HEK293) and purified by HIS-tag purification as per manufacturer instruction (Immune Technology, New York, NY, USA). Recombinant proteins were spotted in 2 drops of 333 pL in protein array buffer (Whatman, Maidstone, Kent, UK) using a non-contact spotter (PerkinElmer, Waltham, MA, USA) on a 64 pad nitrocellulose pad (Oncyte acid, Grac biolabs, Bends, OR, USA). Antigens were spotted in duplicate on each array. Slides were custom prepared for this study at RIVM laboratories, Netherlands and shipped in batches to Vietnam. They were stored at ambient temperature in a humidity and light controlled environment prior to processing.

Antigen	Influenza Virus Strain	Manufacturer
H1N1 1918	A/South Carolina/1/1918	Immune Technology Corp.
H1N1 1977	A/USSR/92/1977	Immune Technology Corp.
H1N1 1999	A/New Caledonia/20/1999	Immune Technology Corp.
H1N1 2007	A/Brisbane/59/2007	Immune Technology Corp.
H1N1 2009	A/California/6/2009	Immune Technology Corp.
H3N2 1968	A/Aichi/2/1968	Sino Biological
H3N2 2003	A/Wyoming/3/2003	Immune Technology Corp.
H3N2 2005	A/Wisconsin/67/2005	Sino Biological
H3N2 2007	A/Brisbane/10/2007	Immune Technology Corp.
H3N2 2009	A/Victoria/210/2009	Immune Technology Corp.
H3N2 2011	A/Victoria/361/2011	Immune Technology Corp.
H9N2 1999	A/Guinea Fowl/Hong Kong/WF10/1999	Immune Technology Corp.
H7N7 2003	A/Chicken/Netherlands/1/2003	Immune Technology Corp.
H5N1 2004	A/Vietnam/1194/2004	Immune Technology Corp.
H5N1 2007	A/Cambodia/R0405050/2007	Sino Biological
H5N1 2010	A/Hubei/1/2010	Immune Technology Corp.

Table 2.4 HA1 Antigens and Manufacturers on Protein Microarray

2.5.3   Protein Microarray Laboratory Methods

Serum samples from the 02FL and 10FL studies were processed in the laboratory by Tran Thi Nhu

Thao and Nguyen Ha Thao Vy.

Fifteen serum samples and one positive control serum can be tested on one slide. Samples are heat inactivated at 56°C processed in BSL-2 laboratory conditions. Fourfold dilutions of sera (1:20, 1:80, 1:360, 1:1280), four-fold dilutions of positive control (1:40, 1:160, 1:640, 1:2560), blocking agent (Blotto blocking buffer in TBS, 5% (w/v) non-fat powdered milk (Pierce 37530), and Surfact Ampt (Pierce, 28320)) were pipetted onto the slides and washed off according to the protocol (Koopmans et al., 2011). Briefly, slides are incubated in Blotto solution (as above) then washed six times with wash buffer (Protein Array Wash Buffer, Maine Manufacturing, USA, 10485330). Prepared serum

dilutions are then added to slides and incubated for one hour at 37℃. After this time they are washed then immunofluorescent conjugate added (Alexa Fluor 647 – AffiniPure Goat Anti-Human IgG, Fcᵞ Fragment Specific (Jackson Immunoresearch, 109-605-008). These are incubated for one hour at 37℃ and then washed thoroughly. After washing, slides are hand dried then stored in an ambient temperature humidity and light controlled environment. Slides are shipped in batches to RIVM laboratories for scanning using ScanArray GX Plus microarray scanner (PerkinElmer), scanned images are then sent electronically to Vietnam for further processing.

2.5.4   Protein Microarray Luminescence and Titre Calculation

Slides were read in with ScanArray software by Tran Minh Quan. Titre calculations and assay standardization to the positive control were designed and calibrated by Nguyen Thi Duy Nhat.

Spot fluorescence intensity was measured using ScanArray Express software (version 4.0, PerkinElmer) and optimal positioning for spot luminescence reading was performed manually (Figure 7). For each sample tested, eight luminescence values are available and these are mapped to a single protein microarray titre by a curve fitting process adapted from log-logistic dose-response curve used commonly in pharmacokinetics.
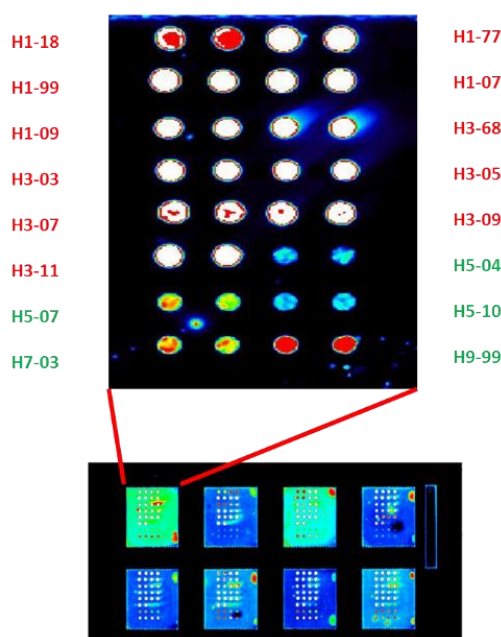
Figure 7 ScanArray Image of Scanned Slide

The median luminescence score for each of the four dilutions were regressed by a four-parameter log-logistic curve. The limits of detection of the scanning equipment are 3000 and 65536 and the titre value was defined as the x-value of the inflection point at the median luminescence (34268) between the limits of detection. Least squares fitting was used to identify the best-fit the logistic

curve, and the inflection point was taken from the best-fit curve (Figure 8). If the luminescence score at dilution 1:20 was less that the inflection point (34268) the sample was assigned a titre of less than 20. If the score at 1:1280 was greater than the inflection point the sample was assigned a value greater than 1280. For data processing these were assigned values of 10 and 1810 respectively,

$$Lum_i = c + \frac{d - c}{1 + \exp\left(b * \log\left(\frac{dilution_i}{titer}\right)\right)}$$
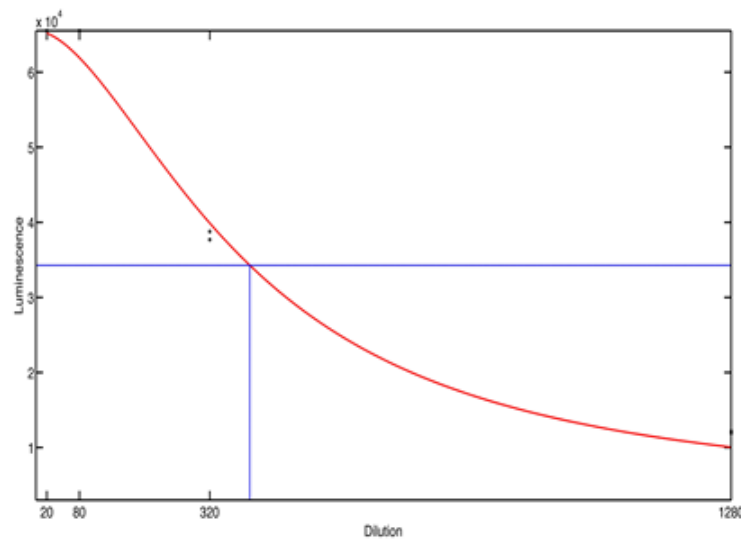
Figure 8 Luminescence and Sample Dilution fitted to four-parameter log-logistic curve; Lum_i: luminescence at dilution i, c=minimum luminescence (3000), d=maximum luminescence (65535), titre=dilution at inflection point, b=slope at inflection point

Standardisation of the assay is performed using the ISH-1 positive control included on each slide. This is designed to correct for inter-laboratory, inter-technician and across slide variation. The H1N1 2009 strain is used as the reference antigen for correction as it has the smallest standard deviation on repeated testing. Slides were produced in "batches" and "series" and differences in antigen spot quality were noted between these. Standardisation therefore occurs within series (i.e. across batches) and then across series. For within series titre correction, all titres on a single slide are normalized based on the mean H1N1 2009 titres within the series that the slide belongs to. The within-series normalization factor for all samples (*s*) on slide *i* in series *k* is calculated as:

$$CF_{ik} = \frac{GMT_{ISH1,H1\_09}}{T_{ISH1,i,H1\_09}}$$

in which $GMT_{ISH1,H1\_09}$ is the geometric mean of ISH-1 titres for H1N1 2009 antigen in series *k*. This correction factor is the same for all antigens *a* and all samples *s* on the same slide. This correction

factor also acts as a quality control criterion as slides are rejected when the H1N1 2009 titre of the positive control differs by more than 1.2 dilutions from the series GMT of H1N1 2009. To correct for the inter-series variation, the mean titre of antigen a in series k is inflated/deflated over its corresponding mean from series 1:

$$CF'_{ak} = \frac{GMT_{ISH1,a,1}}{GMT_{ISH1,a,k}}$$

Individual titre of antigen a for all samples s on slide i in series k is then corrected as:

$$Titres_{saik} = CF_{ik} * CF'_{ak} * Fitted\_Titre_{saik}$$

3   INFLUENZA IN HO CHI MINH CITY 2013-2015

ABSTRACT

Background: Year round transmission of influenza has been detected in Vietnam through both national surveillance and other epidemiological studies. Understanding the demographic and clinical features of influenza-like-illness (ILI) presenting to primary care in urban Vietnam is vital to understand these transmission dynamics. This impact of humoral immune response to current and historic strains on susceptibility will be explored.

Methods: A prospective, observational study of patients with ILI in Ho Chi Minh City, Vietnam has been running since August 2013. Influenza A & B PCR and antibody testing to a panel of 11 human and 5 avian strains is performed using a novel protein microarray technique. A subset of subjects are followed up clinically and serologically for seven months.

Results: 953 ILI patients were recruited between August 2013 & May 2015. 274 and 136 subjects had influenza A and B respectively. Three peaks of influenza activity were detected, H3N2 peak between April and June 2014, influenza B peak in July to December 2014 and mixed H3N2 and H1N1 peak March to May 2015.   Study subjects were younger than the general Vietnamese population but there was no difference between influenza and non-influenza ILI. Lower baseline titre was associated with higher risk of influenza but response to H3N2 2005 and 2009 were most predictive of current susceptibility even when original antigenic sin was taken into account. Levels of protection offered by a fixed titre changed depending on force of infection.

Conclusion: Influenza in southern Vietnam has complex transmission dynamics including periods of intense influenza activity. Serological techniques can be used to give insight into fundamental questions regarding the role of population level susceptibility.

3.1   BACKGROUND

Influenza viruses circulate globally and seasonal epidemics are thought to be associated with 3-5 million severe clinical infections and 250,000 to 500,000 deaths each year (World Health Organisation, 2014c, Global Burden of Disease 2013 Mortality and Causes of Death Collaborators, 2015, Global Burden of Disease Study 2013 Collaborators, 2015). Morbidity and mortality are highest in the extremes of age (Cohen et al., 2010, Nair et al., 2011) but all ages are affected with repeat infections throughout an individual's lifetime. In temperate countries predicable winter epidemics occur however transmission dynamics in tropical settings are much more complex. For many

decades influenza was thought not to be a disease of importance in tropical countries (Viboud et al., 2006) likely due to a lack of observable peak in activity or poor availability of routine surveillance. This has changed considerably in the last 10 years with influenza now recognised to be at least as big a health problem in tropical countries as it is in temperate countries (Yang et al., 2011, Fischer II et al., 2014, Ng and Gordon, 2015). Research efforts are ongoing to increase knowledge and understanding of how this feeds into the global drivers of influenza transmission. Vietnam is a potentially globally important site for influenza dynamics both from its possible contribution to global virus evolution through the 'Southeast Asian Source-Sink' hypothesis (Rambaut et al., 2008, Russell et al., 2008) as well as being considered risk for the emergence of novel pandemic strains such as highly pathogenic avian influenza H5N1.

Year round transmission of influenza has been detected in Vietnam through both national surveillance and other epidemiological studies. Simultaneous circulation of multiple influenza strains and types has been documented (Nguyen et al., 2007, Li et al., 2008, Nguyen et al., 2009, Horby et al., 2012) along with patients having more than one influenza infection in one season (Horby et al., 2012). There are differences in transmission dynamics within Vietnam with the northern subtropical regions having a more predictable seasonality than the tropical central and southern regions (Thai et al., 2015). 15-20% of patients presenting to hospitals with influenza-like illness (ILI) have virologically confirmed influenza (Nguyen et al., 2009, Yoshida et al., 2010, Nguyen et al., 2013) and influenza is thought to be associated with up to 14% of community acquired pneumonia presenting to hospital (Takahashi et al., 2013). However complementary data on community ILI is limited. As national surveillance programmes (Nguyen et al., 2009, Nguyen et al., 2013) and previous research have focused largely on hospital sites, incidence and burden estimates from these studies are likely to underestimate true clinical burden and attack rates in the community. Seasonal influenza vaccine is not currently recommended in Vietnam (United Nations in Vietnam, 2009) and there is minimal, although increasing, use in the private sector (Palache et al., 2014). The optimal vaccination strategy for seasonal influenza in Vietnam is still to be established (Vuong et al., 2012b).

In addition to human strains of influenza, Vietnam is considered to be an important area for human infection with non-human influenza (avian and swine). This cross species transmission is a potential source of new pandemic strains of influenza. Since 2003 there have been 844 confirmed human cases of avian influenza A H5N1 globally (World Health Organisation, 2015b). Vietnam has the third highest attack rate (after Indonesia and Egypt) with 127 confirmed cases. Case fatality rate in confirmed cases is 53.2% globally (449/844) and 50.2% within Vietnam (64/127). However, the reported frequency and severity is thought to be heavily biased by the under-detection of mild and asymptomatic cases. Serological surveys have been performed to try and better estimate the total

case numbers but have been plagued by a number of issues. Existing serological techniques for H5N1 are affected by low sensitivity where even virologically confirmed cases have minimal serological response (as defined by WHO). Several high profile publications in 2012 resulted in controversy regarding the extrapolation of results and whether existing surveillance is 'missing millions' of H5N1 cases (Osterholm and Kelley, 2012, Palese and Wang, 2012a, Palese and Wang, 2012b, Van Kerkhove et al., 2012, Wang et al., 2012). Despite this controversy, serological surveys have remained a key part of the response to the recent emergence of H7N9 in China (Liu et al., 2014, Wang et al., 2014). Recent studies have suggested that seropositivity to avian strains in humans may be related to cross reaction to previous infection or vaccination with human strains (Boni et al., 2013, Molesti et al., 2014, Oshansky et al., 2014, Todd et al., 2014) and that these cross reactions can provide neutralizing ability against those avian strains (Henry Dunand et al., 2015).

Repeat infection with influenza is possible due to the continual evolution of the influenza virus which means that specific immunity generated to one strain is no longer protective when the circulating virus is sufficiently different (Ferguson et al., 2003). Since its emergence in 1968 H3N2 has demonstrated continual change in the haemagglutinin surface protein (HA) termed antigenic drift. In recent years it has been demonstrated that although this accumulation of mutations is a continual process, antigenic evolution is not in parallel to genetic evolution (Bedford et al., 2014). Antigenically similar viruses group into clusters and see a step change to antigenically dissimilar strains every few years (Smith et al., 2004). These cluster changes are the result of amino acid substitutions immediately adjacent to the HA receptor binding site (Koel et al., 2013). All of these changes mean that in order to understand pattern of humoral response we need to take into account the specific response generated by exposure to a virus during infection and the cross reaction that could occur when testing for antibodies to strains which, although the individual has not been infected by, are antigenically similar to one they have previously been infected by. The 'original antigenic sin' hypothesis was generated from the observation that antibody responses to infections early in life 'dominate' the immune response to subsequent infections, and that this 'lesser' response to currently circulating strains resulted in a negative impact for the individuals immune response (Francis et al., 1953). However, more recent work has suggested that response from older strains generated by memory B cells may reduce the viral load, with the resultant new (specific) response being lower because of this (Kim et al., 2009). Key to understanding this, is increasing the number of studies which look at multiple influenza strains and use these to describe not only the pattern of antibody response by age but also to understand the impact that this has on susceptibility to infection and how this response changes after acute infection.

HI testing has traditionally been the gold standard in measuring humoral immunity to influenza because of its correlation with neutralisation assays. The seroprotection threshold of HI titres is 1:40 corresponding to a 50% reduction of the risk of contracting influenza in a susceptible adult population and was originally defined for use in vaccine licensing studies (Hobson et al., 1972). It remains the assumed level of protection in studies looking at population immunity (both vaccine and natural infection induced) despite several studies demonstrating that seroprotection is likely to be on a continuous scale rather than a fixed threshold (Nauta et al., 2009, Coudeville et al., 2010, Ohmit et al., 2011) as well as containing mechanisms beyond purely the inhibition of receptor binding. Another challenge is that levels of seroprotection are derived where there is a match between the circulating virus and the tested strain. Where there is a mismatch, the influence of cross-reaction and response to previous infections are likely to be an important factor. The development of protein microarray (PMA) for multiple influenza strains is aimed to improve some of the deficiencies associated with HI testing and in particular between laboratory variability and to test in a high throughput manner. Although PMA titres are higher than HI titres, there is a good correlation in the rise following infection (Koopmans et al., 2011, Huijskens et al., 2013). The ability of PMA to measure functional virus neutralization rather than virus binding capacity is still to be fully understood.

3.2 METHODS

The objectives of this portion of the study was to describe the dynamics of ILI presenting to primary care services within an urban setting in Vietnam over a 21 month period and the proportion of influenza A and B as a cause of ILI. To determine the demographic and clinical characteristics of individuals presenting with ILI and to determine whether there was any differences in these characteristics between influenza and non-influenza ILI. To describe the antibody response to current and historical human strains of influenza A in individuals presenting with ILI, look for differences in response by age and determine the risk of current infection with influenza for a given antibody response. To describe the antibody response to avian strains of influenza A and look for differences in response by age and poultry exposure. A subset of patients were recruited for longitudinal follow-up to determine clinical outcome from acute ILI and to determine rates of repeat ILI in the short term following original ILI illness. A summary of study procedures relevant to the results of this chapter are included below, a full description of study procedures and laboratory testing is given in Chapter 2. Longitudinal serological response after ILI will be reported in Chapter 4.

3.2.1   Summary of Design and Conduct of the Observational Study

This observational study was conducted between August 2013 and May 2015 at outpatient clinics and community medical practitioners in Ho Chi Minh City (HCMC), Vietnam. These clinics were considered representative of primary care within Vietnam (World Health Organisation & MOH Vietnam, 2012). Individuals were invited to join the study if they were between 10 and 70 years of age with symptoms for less than 72 hours and fitting the ECDC ILI definition (European Centre for Disease Prevention and Control (ECDC), 2015) One anterior nasal swab and one throat swab were collected at recruitment and transported in a single tube of viral transport medium to central laboratory before being stored at -20°C within 24 hours. A 5ml serum sample was also collected at baseline and stored at -20°C within 24 hours.

A subset of patients were invited to join a longitudinal sub-study. All subjects testing positive for influenza A were eligible to join the sub-study. Negative control subjects were matched by age (+/- 5 years) and gender to influenza A positive subjects included in the follow-up study. Subjects attended for repeat serum sampling at 30, 90, 150 and 210 days (+/- 5 days) after recruitment. Active respiratory symptom surveillance was conducted during the follow-up period with telephone follow-up at 60, 120 and 180 days. Subjects were asked to attend clinic within five days of onset of new ILI symptoms where repeat nasal and throat swabbing was performed if they met ECDC ILI case definition.

Clinical characteristics of the ILI episode was reported by subjects to study staff, any subsequent medical or hospital attendance were reported by subject at next visit. No inpatient data was collect on subsequent hospitalisation. Contact data relied on participant's recollection of the previous day. No additional objective information was available from the subject's usual treating clinician.

Both studies were approved by the Scientific and Ethical Committee of the Hospital for Tropical Disease, Ho Chi Minh City, Vietnam and Liverpool School of Tropical Medicine Research Ethics Committee, UK. Letters of agreement supporting the involvement of the community medical clinics were obtained from the Ho Chi Minh City Department of Health.

3.2.2   Sample Analysis

Respiratory samples were batched tested monthly for influenza A & B using standard polymerase chain reaction (PCR) techniques (World Health Organisation, 2011). Samples were stored at -20°C within 24 hours of collection and then stored at -80°C after initial PCR testing. Samples positive for influenza A were then tested for H3N2 and H1N1 subtypes. All influenza A positive subjects and subset of all other subjects (50%) had their baseline serum sample tested against a panel of 16

influenza A strains (11 human, 5 avian strains) using a protein microarray technique previously described (Koopmans et al., 2011). Negative and influenza B controls were age and sex matched to the positive subjects included in serological testing. In brief, serum samples were tested in 4 fold seral dilutions from 1:20 to 1:1280 on 64 pad nitrocellulose slides. Each pad was spotted (in duplicate) with recombinant HA1 proteins of the 16 strains (Table 3.1). Inter-assay variability was monitored by the inclusion of ISH-1 control. Microarray slides were scanned using ScanArray GX Plus microarray scanner (PerkinElmer) and spot fluorescence intensity was measured using ScanArray Express software (version 4.0, PerkinElmer). Titres were calculated from the inflection point of the titration curve. Full methods described in Chapter 2.

HUMAN INFLUENZA ANTIGENS		AVIAN INFLUENZA ANTIGENS	
H1N1	A/South Carolina/1/1918	H9N2	A/Guinea Fowl/Hong Kong/WF10/1999
	A/USSR/92/1977	H7N7	A/Chicken/Netherlands/1/2003
	A/New Caledonia/20/1999	H5N1	A/Vietnam/1194/2004
	A/Brisbane/59/2007		A/Cambodia/R0405050/2007
	A/California/6/2009		A/Hubei/1/2010
H3N2	A/Aichi/2/1968		
	A/Wyoming/3/2003		
	A/Wisconsin/67/2005		
	A/Brisbane/10/2007		
	A/Victoria/210/2009		
	A/Victoria/361/2011		

Table 3.1 HA1 antigens included on microarray

3.2.3   Statistical Analyses

The primary outcome was PCR-confirmed influenza A in nasal/throat swabs. This included both single influenza A infections and co-infections with influenza B.

Continuous variables which were normally distributed were compared with t-test or ANOVA as appropriate. Tukey's HSD was used for posthoc testing following ANOVA where appropriate. Continuous variables which were non-normally distributed were compared using Wilcox and Kruskal Wallis rank sum depending on the number of groups (2 vs more than 2 respectively). Categorical variables were compared using the Fisher exact test, Mann-Whitey test or Chi Square test as appropriate.

Analysis of overall prevalence of influenza as a cause of ILI was performed on the total study population. As well as overall prevalence, weekly prevalence was calculated. Weekly influenza transmission intensity was categorised by percentage of ILI subjects testing influenza positive using WHO thresholds (World Health Organisation, 2015a) as zero (0%), low (1-10%), medium (11-20%), high (21-30%), very high (>30%). A peak of influenza activity was defined as four or more consecutive weeks where influenza transmission intensity was high or very high. Weeks where no recruitment

was performed because of clinic closures were excluded from analysis, weeks where clinics were open but no patients were recruited were included in analysis. Repeat ILI episodes during respiratory follow-up were not included in the calculation of overall or weekly proportions. Incidence of secondary ILI was calculated on the follow-up cohort as a rate of events per 1000 days of follow-up and rate ratio comparisons. Time to repeat ILI was investigated using Cox proportional hazards.

Demographic, clinical and contact characteristics were compared between predefined groups based on influenza infection status. Initial analysis was planned to be performed between three groups; influenza A, influenza B and non-influenza ILI. Non-influenza ILI includes those who have influenza A or B but test negative by PCR (likely to be a very small number due to sensitivity of molecular testing), those infected with a non-influenza respiratory virus and those who have a non-respiratory virus cause of their symptoms. As it is currently not possible to distinguish between these groups they will be considered as one group referred to as non-influenza ILI or influenza negative. Analysis was planned to be performed between influenza A subtypes (H1N1 and H3N2) if numbers were sufficient.   Where appropriate, analysis was stratified by age. Age was preferentially used as a continuous variable, otherwise age was categorised as per recommendations from the Consortium for the Standardisation of Influenza Seroepidemiology (5-9, 10-19, 20-44, 45-65, 65+) (Van Kerkhove et al., 2013a).

Logistic regression was used to investigate the effect of household age structure (total number of household contacts, total number of household contacts in each CONSISE age class), recent household ILI on the risk of any influenza (influenza A or influenza B) compared to non-influenza ILI.

Age, gender and household size data was compared to national Vietnamese Census Data (General Statistics Office of Vietnam, 2009) according to influenza infections status and all study subjects combined. Age and gender comparisons were made to both Vietnamese National and Ho Chi Minh City census data. Household size comparisons were made to Vietnamese National Urban Average census data. For each of the infection status groups the expected proportion was the census point estimate, 95% binomial confidence intervals of the expected proportion was calculated using the number of subjects in that category. Chi square goodness of fit using the census proportions as expected probabilities detected if study distributions were significantly different from the general Vietnamese population. Testing whether the national or HCMC census data provided a better fit to study data was performed by performing Student's T-test of the mean Pearson Residual from the chi square test.

Statistical analysis of baseline serology was performed on all study subjects and then compared between influenza infection status groups (influenza A H1N1, influenza A H3N2, influenza B and non-influenza ILI). Microarray strains were categorised into i) Most Recent Circulating Strains (H1N1 2009, H3N2 2011), ii) Historical Human Strains (all other human strains), and iii) Avian Strains (Table 3.1). Titre responses will be categorised as homosubtypic where the infecting subtype and the microarray strain match (i.e. infected with H1N1, homosubtypic titre response is H1N1 strains of the microarray). Heterosubtypic titres are where the infecting subtype and the microarray strain do not match (i.e. infected with H1N1, heterosubtypic titre response is H3N2 strains on the microarray). Analysis was performed on log2 transformed titres unless explicitly stated.

Linear regression and non-parametric general additive models were used to investigate the relationship between baseline titre and age at recruitment, gender, pregnancy, chronic respiratory disease, smoking and previous influenza vaccination. In addition to calendar age of a subject, age at the time of strain emergence would be investigated to look for the effect of original antigenic sin and antigenic seniority.

Univariate logistic regression was performed to investigate level of protection offered by microarray titre at baseline. The outcome of this analysis was susceptibility to PCR positive influenza. Analysis was performed for H1N1 and H3N2 separately. An indicator variable of 1 was given to individuals infected with that subtype, i.e. in H3N2 analysis individuals infected with H3N2 are coded as one and individuals not infected (all of H1N1 PCR positive, influenza B positive and influenza A & B negative) are coded as zero. A multivariable logistic regression was performed to investigate the effect of age, gender, respiratory disease and smoking on seroprotection. Study time (week of study recruitment) was used as a proxy for the changing force of infection. GAM were used to investigate non-linear effects of baseline titre and study time. The goal was to look at how the level of protection offered by a fixed titre changed as the force of infection changed.

Historic human strains were investigated using multivariable logistic regression to investigate the effect of combination of strains titres on the predicted risk of developing influenza. All homosubtypic strain combinations were tested along with age. In order to investigate the original antigenic sin hypothesis, a variable of "Earliest Strain" was created for the strain which emerged closest to an individual's first 10 years of life. Model comparison was performed using a comparisons of AIC, adjusted R^2 and percentage deviance of the model fit. Further model testing investigating the effect of study time of multi-strain protection to be performed on the best reduced model and the full model.

All statistical analysis were performed using R Statistical Software v3.2.1 (R Core Team, 2015). Packages used for analysis were ggplot2, gamm4, mgcv, Epi, survival and MASS.

3.3 RESULTS

953 subjects with ILI were recruited between 8 August 2013 and 31 May 2015. The majority of patients were recruited from hospital outpatient settings (655/953, 68.7%). Two subjects were relatives who did not attend clinic for a healthcare assessment, all other subjects attended clinics to be reviewed by a medical practitioner. Only 20 subjects did not consent to be contacted to join the follow-up study. 186 subjects had at least one follow-up visit, 64.5% of these were influenza A PCR positive.
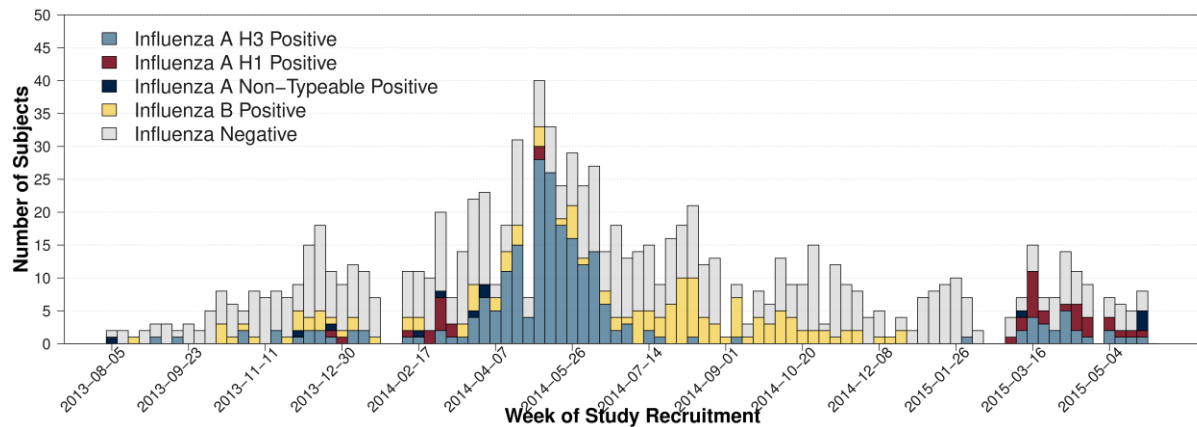
Figure 9 Weekly Study Recruitment Timeline by Influenza PCR Result. Study clinics were closed to recruitment during national public holidays corresponding to study weeks 26, 27, 81, 82 & 91.

3.3.1 ILI Dynamics

42.8% of ILI cases had PCR confirmed influenza (n= 410/953) (Figure 9). Influenza A was detected in 274 subjects (28.7%) and influenza B in 136 subjects (14.3%). H3N2 was the commonest circulating influenza A subtype (81.4%, n=223/274). H1N1 was detected in 39 subjects (14.2%) and no influenza A subtype was detectable in 4.4% of influenza A cases. No recruitment was performed in the five weeks where clinics were closed because of public holidays (study weeks 26, 27, 81, 82 and 91). Influenza was detected in the majority of weeks where study recruitment was performed (83.5%, n=76/90). Influenza A was detected in 50 study weeks (55.6%, n=50/90), with H3N2 present more frequently than H1N1 (46 vs 18 weeks, 51.1% vs 20%). Influenza B was detected in 51 study weeks (56.7%, n=51/90). Three periods of sustained high or very high transmission were identified (Figure 10). An H3N2 peak occurred between 31 March 2014 and 30 June 2014 (study weeks 35 to 48) which was immediately followed by an influenza B peak between 07 July 2014 and 22 December 2014

(study weeks 49 to 73). Finally a mixed H1N1 and H3N2 peak in activity was detected between 02 March 2015 and 31 May 2015 (study weeks 83 to 95).
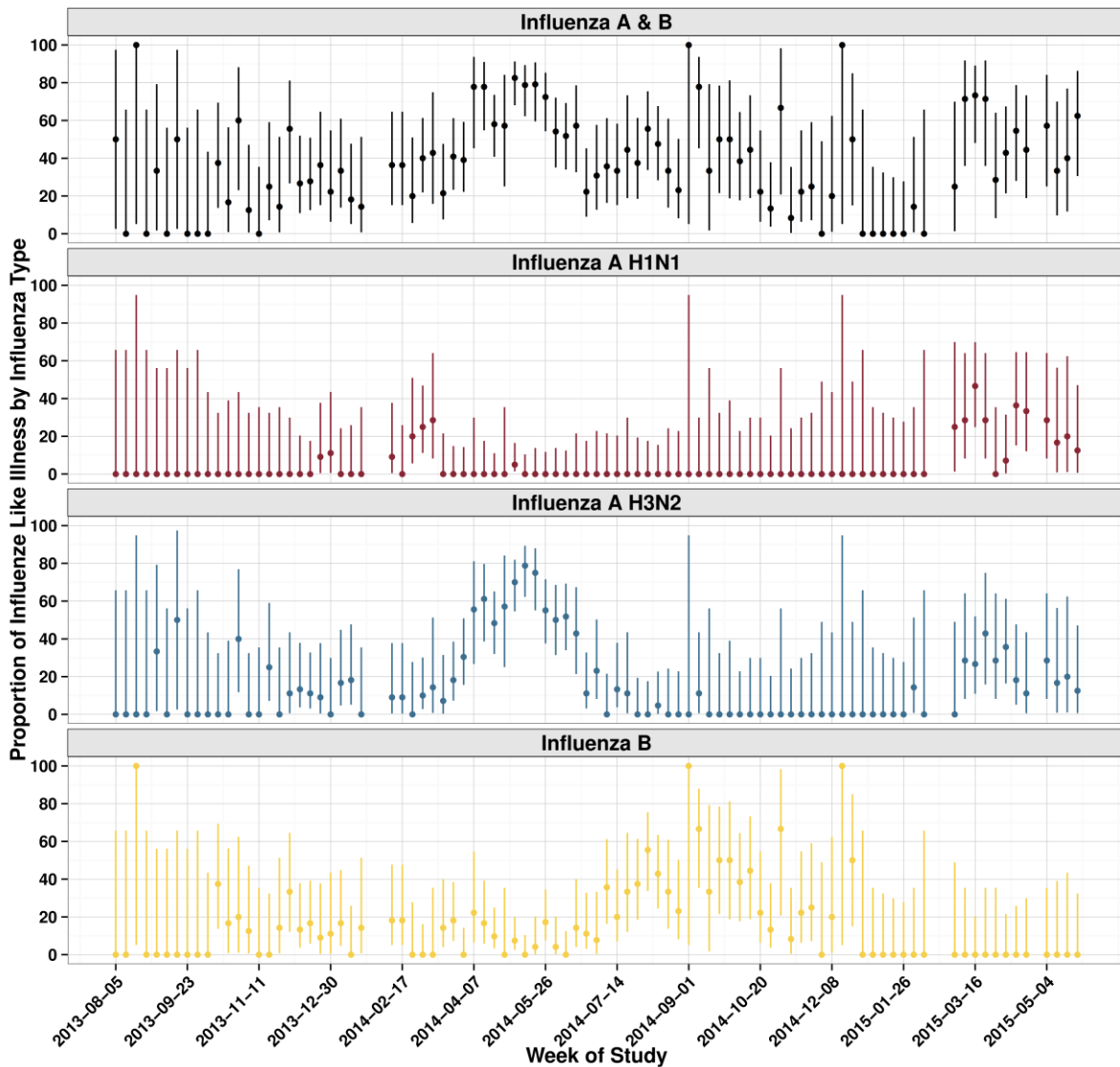
Figure 10 Proportion of ILI caused by PCR confirmed Influenza with Binomial 95% Confidence Intervals.

3.3.2   Repeat ILI during Follow Up

Follow up data was available 186 subjects, 64% of which were influenza A positive (n=120/186). Median follow up period was 187 days (IQR 66-216 days) and 14 repeat episodes of ILI were reported during the follow-up period. 6.6% of subjects from influenza A positive group had a further episode of ILI during the follow-up period (n=8/120) compared with 9% of subjects who were negative for influenza A & B at original presentation (n=6/66). Nose and throat swabs for molecular testing were available for 10 subjects during these repeat episodes. None of these tested positive for influenza A or B so analysis on risk of PCR confirmed ILI could not be conducted. The influenza A follow-up cohort had more days of follow-up than the influenza negative group (15496 days vs

10597 days). This gave a rate of repeat ILI of 0.516 episodes per 1000 person days following influenza A and 0.567 episodes/1000 person days following influenza B; rate ratio 1.098 (95% CI 0.381 – 3.164). Subject age or original week of recruitment had no significant effect on rates of repeat ILI. There did not appear to be clustering of repeat ILI cases to suggest an outbreak of another respiratory infection by visual inspection of follow up timeline (Figure 11). Time to repeat ILI was investigated using Cox Proportional Hazards. No significant difference was detected between those originally infected with influenza or not. Given the small number of repeat episodes of ILI during the follow-up period these results should be interpreted with caution.
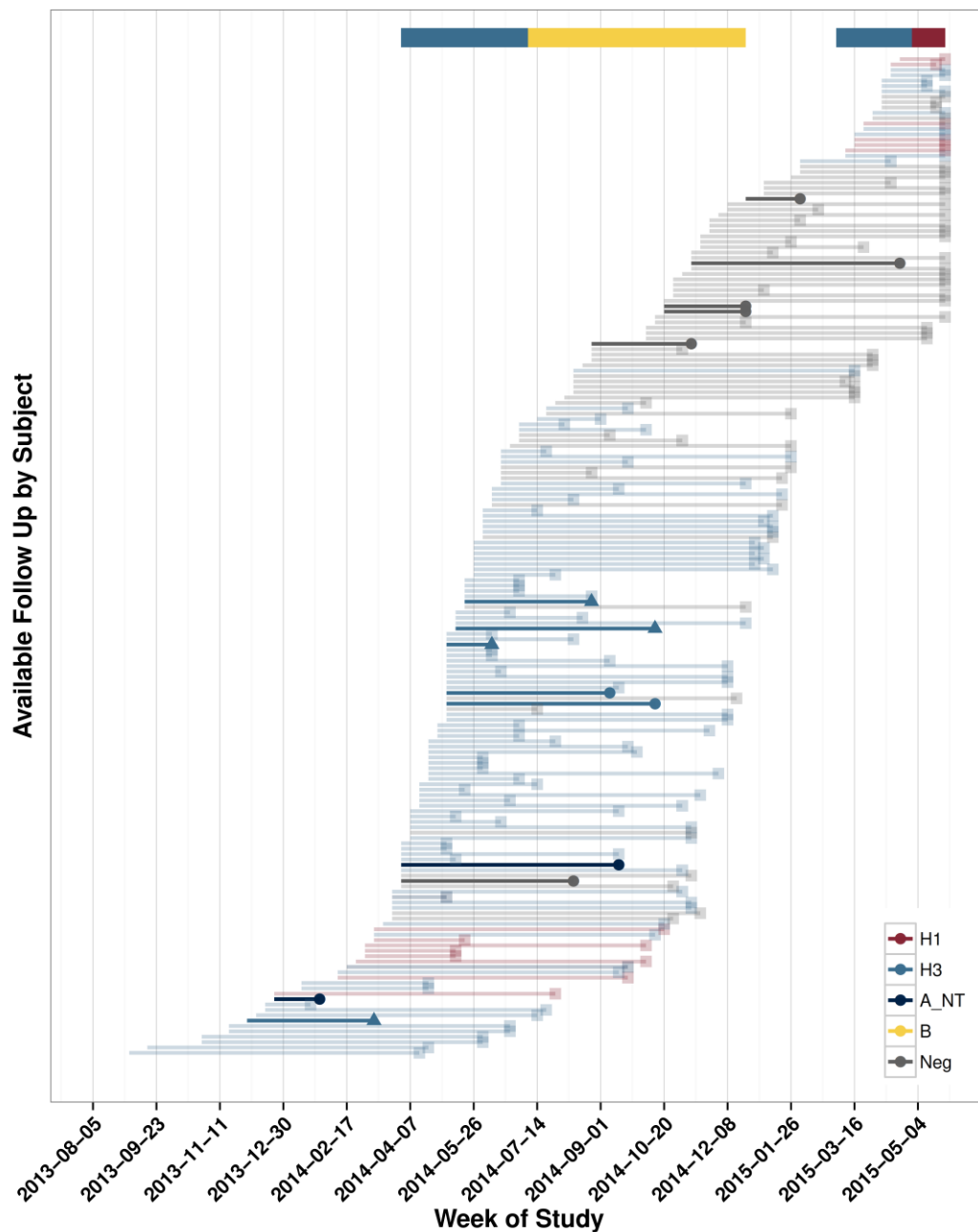
Figure 11 Follow-up durations and Repeat ILI Episodes. Darker lines are individuals who had repeat ILI. End symbol; square= final follow-up no ILI, circle=repeat ILI at this time, PCR available, triangle=repeat ILI at this time, no PCR. Top bars are periods of high influenza transmission by subtype.

3.3.3   Age Distribution

Age of recruited study participants was left-skewed with a median age of 25.4 (IQR 18.9-33.7). Three study participants were outside the protocol defined age limits, two aged 9 and one aged 74. These individuals are included in the analysis. No difference in the age distribution by gender or influenza infection status was seen on visual inspection. Median age of study participants was not statistically significantly different by gender (Mann Whitney U W=116000, p-value 0.4072) or influenza infection status (Kruskal Wallis $\chi2(2)=0.687$, p-value=0.5431). The gender specific age distribution of study participants was compared to Vietnam Census estimates (National and HCMC) using Chi Square Goodness of Fit (Figure 12, HCMC data only).
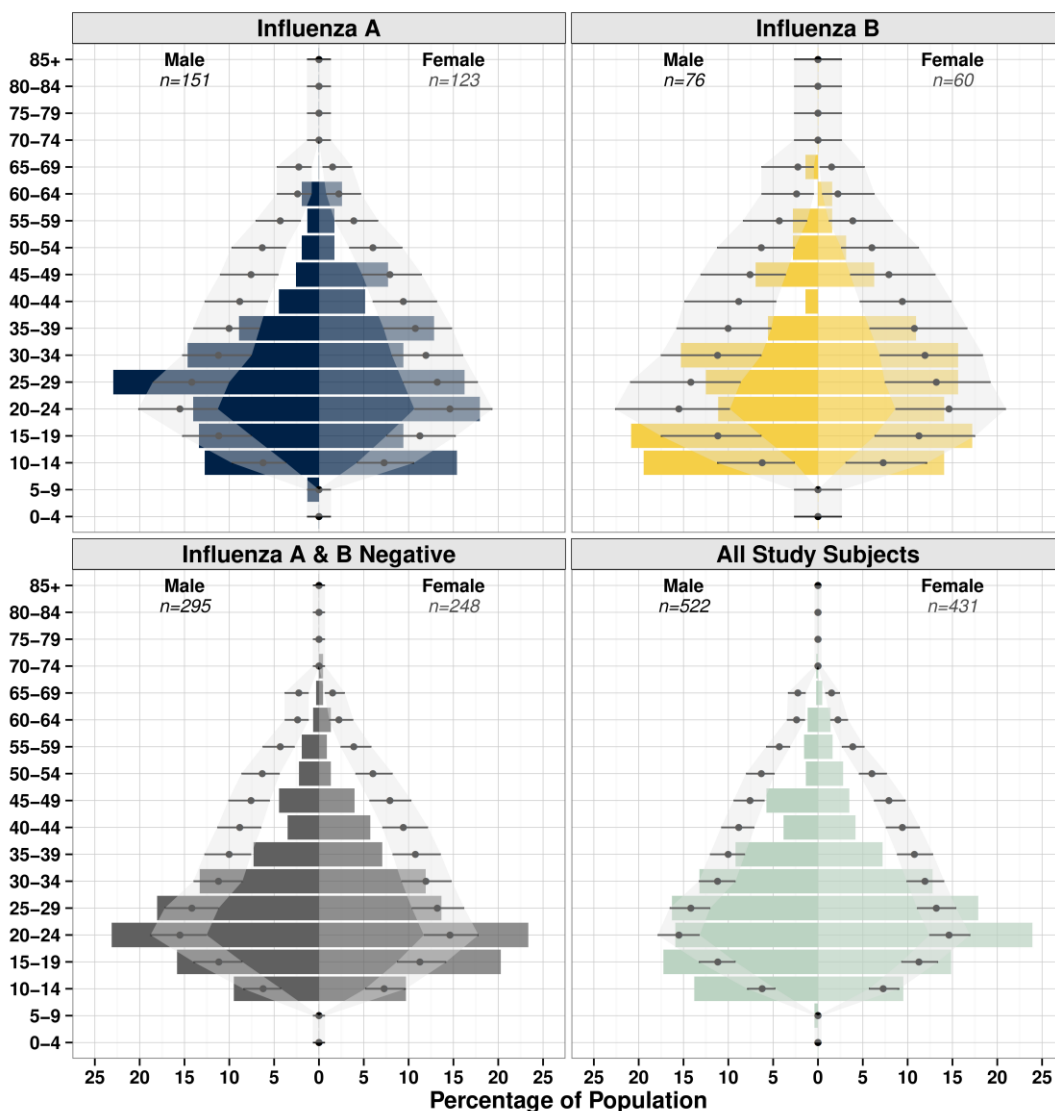
Figure 12 Age and Sex Distribution by Infecting Influenza Type. Point and bars represent expected percentage for HCMC with binomial 95% CI. Expected values were limited to age groups included in the study protocol (10 to 70 inclusive).

With the exception of subjects infected with influenza B, the gender specific age distributions of study participants were statistically significantly different to the general population (Table 3.2). Pearson residuals from the Chi Square model show that younger individuals tend to be over represented in our study population (reflected by positive residuals) whereas individuals over the age of 35 tend to be under represented (reflected by negative residuals) compared to expected rates from the population (Figure 13). There is no difference using national data or HCMC specific gender specific age distribution data for comparison (t test, all p values >0.4).

		Influenza A		Influenza B		Influenza A & B Negative		All Study Subjects	
		χ^2	p value	χ^2	p value	χ^2	p value	χ^2	p value
Vietnam National	**M**	41.363	0.0005	18.684	0.0745	80.882	0.0005		0.0005
Age Distribution	**F**	23.105	0.0165	16.195	0.1239	68.213	0.0005	126.18	0.0005
Ho Chi Minh City	**M**	42.969	0.0005	38.087	0.0010	66.584	0.0005	129.95	0.0005
Age Distribution	**F**	22.915	0.0210	16.211	0.1414	57.097	0.0005		0.0005

Table 3.2 Age and Gender Distribution by Infecting Influenza Type. Pearson's Chi Square Goodness of Fit with expected proportions in population. Symptotic p values calculated using Monte Carlo testing.
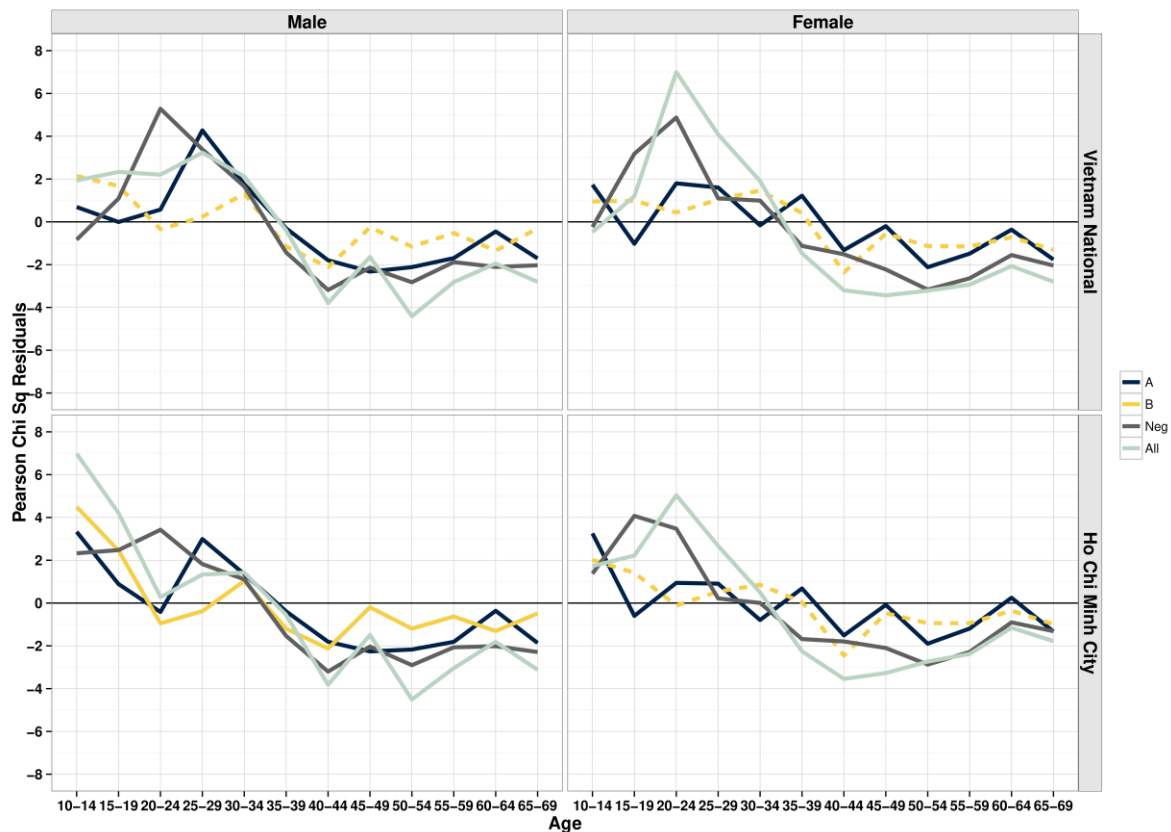
Figure 13 Age Group deviation from Population Census Data by Infecting Influenza Type. Pearson Chi Square Residuals. Dashed line chi square p value >0.05.

3.3.4 Clinical Characteristics

3.3.4.1 Past Medical History

At baseline there was no significant difference in relevant past medical history (PMH) of individuals presenting with PCR confirmed influenza or non-influenza ILI (Table 3.3). Chronic respiratory disease was uncommon in recruited subjects (2.4%, n=23/953) but was reported more frequently in individuals aged 45 years or over (6.3% vs 1.8%, 6/95 vs 17/858; χ^2 5.106 p values 0.0238). Current smoking was reported in 14.6% of the study population all of whom were male (n=139/952). 36% of males over the age of 20 are current smokers (n=129/354). None of the current smokers reported chronic respiratory disease. 2.8% (n=27/953) of the study population reported an indication for influenza vaccination as defined by WHO (Age >65; PMH of COPD, congenital heart disease, heart failure, diabetes or asthma; healthcare worker). 2.4% of the study population had received vaccination at some point in their life (n=23/953), but only one of these had a reported indication. All subjects who reported vaccination had received it after the emergence of the 2009 influenza pandemic (median 2012, IQR 2011-2013).

		Influenza A n (%)/ med (IQR)	Influenza B n (%)/ med (IQR)	Influenza Negative n (%)/ med (IQR)	χ^2	p value
	Age	26.65 (19.12-	25.4 (17.58-	24.8 (19.35-		
Age Category	5 to 9	2 (0.7)	0 (0)	0 (0)		
	10 to 19	70 (25.5)	49 (36)	148 (27.3)		
	20 to 44	174 (63.5)	69 (50.7)	346 (63.7)		
	45 to 64	28 (10.2)	17 (12.5)	46 (8.5)		
	65+	0 (0)	1 (0.7)	3 (0.6)		
Gender F		123 (44.9)	60 (44.1)	248 (45.7)		
M		151 (55.1)	76 (55.9)	295 (54.3)	0.1235	0.9401
Pregnant Yes		5 (4.1)	2 (3.3)	4 (1.6)		
No		116 (95.1)	58 (96.7)	241 (97.2)		
DK		1 (0.8)	0 (0)	3 (1.2)	2.9582	0.5648
Current Smoker Yes		36 (13.2)	19 (14)	84 (15.5)		
No		237 (86.8)	117 (86)	458 (84.3)		
Refused		0 (0)	0 (0)	1 (0.2)	1.5844	0.8116
COPD Yes		0 (0)	0 (0)	2 (0.4)		
No		274 (100)	136 (100)	541 (99.6)	1.5133	0.4692
Asthma Yes		4 (1.5)	1 (0.7)	9 (1.7)		
No		270 (98.5)	135 (99.3)	534 (98.3)	0.6392	0.7264
Any Resp Disease Yes		8 (2.9)	2 (1.5)	13 (2.4)		
No		266 (97.1)	134 (98.5)	530 (97.6)	0.8124	0.6662
Vaccine Indication Yes		5 (1.8)	3 (2.2)	19 (3.5)		
No		268 (98.2)	133 (97.8)	524 (96.5)	2.0619	0.3567
Ever Received Vaccine		3 (1.1)	4 (2.9)	16 (2.9)		
No		262 (95.6)	127 (93.4)	520 (95.8)		
DK		9 (3.3)	5 (3.7)	7 (1.3)	7.69	0.1036

Table 3.3 Past Medical History at Baseline

3.3.4.2   Clinical Symptoms at Presentation

Most subjects presented on the second or third day of symptoms. Those with influenza B presented later than those with influenza A or who were influenza negative (ANOVA F(2,950)=15.9, p <0.001) however the difference in time to presentation was 0.32 of a day (95%CI 0.17-0.47) which is unlikely to be clinically significant. All subjects had mild disease, with less than 1% of subjects reporting that they could not carry out their normal daily activities (n=5/948), although this was more common in those who were infected with influenza A (1.5%, 4/274; χ^2 6.52, p value=0.038), the numbers of subjects in this group mean this result should be interpreted with caution. Rhinorrhoea and cough were reported more frequently in those infected with influenza (Table 3.4). Antibiotic use prior to enrolment in the study was common in all groups (51.7%, n=492/952) but was highest in those who tested positive for influenza B (66.2%, n=90/136; χ^2 15.1, p value=0.0005).

	Influenza A n (%)/ med (IQR)	Influenza B n (%)/ med (IQR)	Influenza Negative n (%)/ med (IQR)	χ^2	p value
Days Symptom Onset	2 (2-3)	3 (2-3)	2 (2-3)		
Normal Tasks Yes	270 (98.5)	136 (100)	542 (99.8)		
No	4 (1.5)	0 (0)	1 (0.2)	6.5151	0.0385
Fever Yes	256 (93.4)	118 (86.8)	482 (88.8)		
No	18 (6.6)	18 (13.2)	61 (11.2)	5.9555	0.0509
Temp If Known	38.5 (38-39)	38.1 (38-38.8)	38 (37.8-38.85)		
Headache Yes	254 (92.7)	117 (86)	488 (90)		
No	20 (7.3)	19 (14)	54 (10)	4.6429	0.0981
Rhinorrhoea Yes	236 (86.1)	110 (80.9)	354 (65.2)		
No	38 (13.9)	26 (19.1)	189 (34.8)	45.433	<0.000
Cough Yes	259 (94.5)	129 (94.9)	463 (85.3)		
No	15 (5.5)	7 (5.1)	80 (14.7)	21.456	<0.000
Sore Throat Yes	243 (88.7)	120 (88.2)	475 (87.5)		
No	31 (11.3)	16 (11.8)	68 (12.5)	0.2646	0.8761
Myalgia Yes	246 (89.8)	123 (90.4)	490 (90.2)		
No	28 (10.2)	13 (9.6)	53 (9.8)	0.0596	0.9706
GI Symptoms Yes	18 (6.6)	13 (9.6)	36 (6.6)		
No	256 (93.4)	123 (90.4)	507 (93.4)	1.5526	0.4601
Malaise Yes	267 (97.4)	129 (94.9)	520 (95.8)		
No	7 (2.6)	7 (5.1)	23 (4.2)	2.0588	0.3572
Paracetamol Yes	227 (82.8)	119 (87.5)	452 (83.4)		
No	47 (17.2)	17 (12.5)	90 (16.6)	1.6219	0.4444
Antiviral Yes	1 (0.4)	1 (0.7)	2 (0.4)		
No	273 (99.6)	135 (99.3)	540 (99.6)	0.3767	0.8283
Antibacterial Yes	144 (52.6)	90 (66.2)	258 (47.6)		
No	130 (47.4)	46 (33.8)	284 (52.4)	15.139	0.0005
Vitamin Yes	102 (37.2)	67 (49.3)	185 (34.1)		
No	172 (62.8)	69 (50.7)	358 (65.9)	10.756	0.0046

Table 3.4 Clinical Symptoms at Presentation

3.3.4.3 Recovery following Acute Influenza like Illness

Follow-up information was available for 186 subjects (19.4%); 120 influenza A positive & 66 influenza A & B negative. Three individuals reported that their symptoms got worse after their original presentation (two influenza A (1.6%), one influenza negative (1.5%)). One individual consulted a community pharmacist but no other healthcare intervention was required. No study subjects required hospital admission.

3.3.5 Demographic Characteristics

The commonest reported occupation was school pupil or student (27.8%, n=265/952) followed by manual workers (19.7%, n=188/952) and shop assistants/traders (14.9%, n=142/952). Highest rates of influenza A were seen in at office workers who did not have contact with the wider public as part of their job and home carers (Table 3.5). Contact with pigs was uncommon but contact with poultry was reported more frequently with 5.9% and 11.4% of subjects reporting weekly contact with live or dead poultry respectively.

	Influenza A n (%)/ med (IQR)	Influenza B n (%)/ med (IQR)	Influenza Neg n (%)/ med (IQR)	χ^2	p value
Occupation: At Home	39 (14.3)	13 (9.6)	54 (9.9)		
Student	77 (28.2)	42 (30.9)	146 (26.9)		
Teacher	2 (0.7)	4 (2.9)	7 (1.3)		
Office:No Public Contact	23 (8.4)	3 (2.2)	33 (6.1)		
Office:Public Contact	26 (9.5)	10 (7.4)	44 (8.1)		
Healthcare	1 (0.4)	0 (0)	2 (0.4)		
Driver	8 (2.9)	7 (5.1)	17 (3.1)		
Trader	35 (12.8)	16 (11.8)	91 (16.8)		
Manual Work	49 (17.9)	38 (27.9)	101 (18.6)		
Other	13 (4.8)	3 (2.2)	48 (8.8)	33.42	0.0148
Household Contacts	3 (2-4)	3 (2-4)	3 (2-4)		
Live Poultry Contact:	260 (95.9)	130 (96.3)	477 (90.5)		
rarely	0 (0)	0 (0)	1 (0.2)		
monthly	1 (0.4)	0 (0)	9 (1.7)		
weekly	10 (3.7)	5 (3.7)	40 (7.6)	12.119	0.0594
Dead Poultry	218 (80.7)	115 (85.8)	434 (82.5)		
rarely	3 (1.1)	1 (0.7)	8 (1.5)		
monthly	14 (5.2)	5 (3.7)	26 (4.9)		
weekly	35 (13)	13 (9.7)	58 (11)	2.2724	0.893
Live Pig Contact: never	267 (98.9)	132 (98.5)	510 (97.1)		
rarely	0 (0)	0 (0)	1 (0.2)		
monthly	1 (0.4)	1 (0.7)	2 (0.4)		
weekly	2 (0.7)	1 (0.7)	12 (2.3)	4.5621	0.6011
Dead Pig Contact: never	267 (98.9)	132 (98.5)	510 (97.1)		
rarely	0 (0)	0 (0)	1 (0.2)		
monthly	1 (0.4)	1 (0.7)	2 (0.4)		
weekly	2 (0.7)	1 (0.7)	12 (2.3)	4.5621	0.6011

Table 3.5 Demographic Characteristics at Baseline

Median number of household members (including the study subject) was four (IQR 3-5), however this was heavily right skewed with the largest household having 17 members. The median household size was significantly higher than the national urban average of 3.66 (Wilcoxon Rank Sum V= 283830, p-value=<0.001). The distribution of household size was also statistically significantly different to the national urban average with fewer one person households and more households with five or more occupants (Figure 14, Table 3.6; Chi Square Goodness of Fit, all p values <0.005). Distribution of household size was similar for all infection groups and median number of household members was not statistically significant between groups (Kruskal Wallis χ2 (2)=1.8958, p-value=0.3875). There was no relationship between the age of the recruited subject and their number of household members as assessed by linear regression (F(1, 950) = 0.7552, p 0.385).
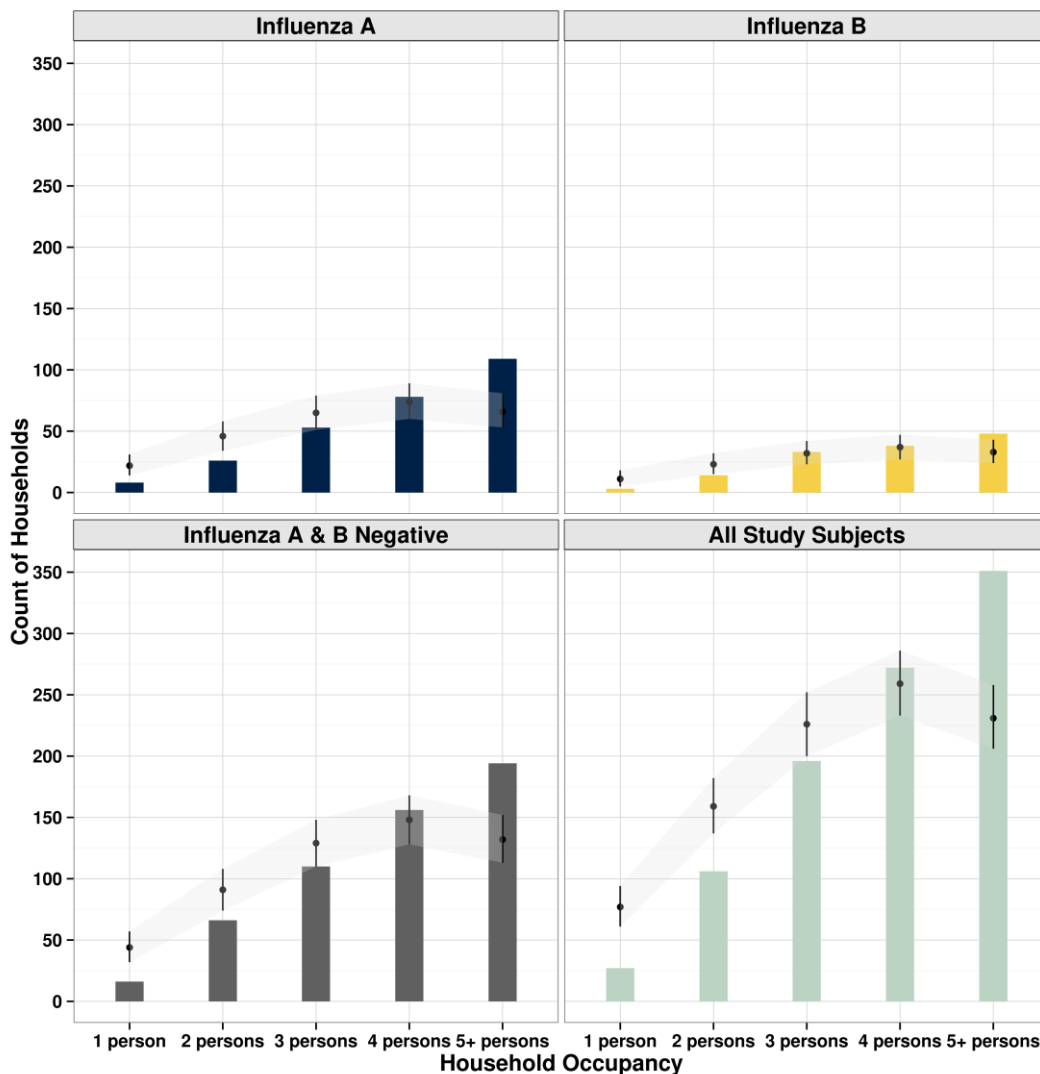
Figure 14 Counts of Household Size by Infecting Type and in All Study Subjects. Point and bars are the expected counts with 95% binomial CI from the national household distribution.

	Influenza A (%)	Influenza B (%)	Influenza Negative (%)	National Urban Average (%)
1 person	2.92	2.21	2.95	8.1
2 persons	9.49	10.29	12.18	16.7
3 persons	19.34	24.26	20.3	23.7
4 persons	28.47	27.94	28.78	27.2
5+ persons	39.78	35.29	35.79	24.3
χ^2	46.9	15.98	56.99	
p value	0.0004	0.004	0.0005	

Table 3.6 Distribution of Household Size in Study Population and National Urban Average for Vietnam (2009). Chi Square Goodness of Fit.

Detailed information on household members was given by 894 subjects (93.8%) where they lived with at least one other person. 8.5% of subjects reported that at least one member of their household had ILI symptoms in the preceding week (n=76/894). There was no difference in the presence of household ILI between influenza positive or negative groups (Kruskal Wallis χ^2 (2)=0.1019, p-value=0.95). Overall 3.5% of household contacts were reported to have ILI symptoms (n=100/2894). The proportion of households with dependent members (under 10 and over 65, including study subjects) was lower than the national average (Figure 15). In univariate and multivariable logistic regression there was no significant effect of household age structure or recent household ILI on risk of flu vs non-influenza ILI or on risk of repeat ILI during the follow up period.
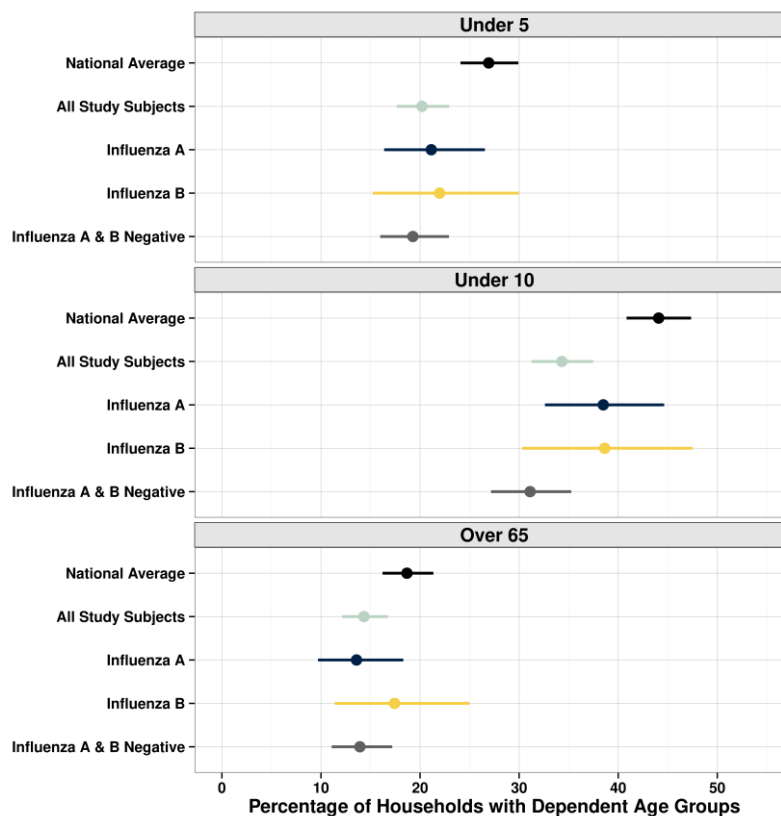
Figure 15 Percentage of Households with Dependent Age Groups

3.3.6   Contact Patterns

Patients were asked to report contact patterns for the previous day at baseline, planned and symptomatic follow-up visits. Although 91.3% of subjects gave this information at baseline (n=870/953) and 80% during symptomatic follow-up visits (n=8/10), only 5.7% of subjects agreed to share this information during routine follow-up visits (n=28/489). Linear regression demonstrated no difference in either total numbers of face to face conversations or total number of physical contacts when individuals currently had ILI (p value >0.05). Due to the low number of follow-up contact patterns, no further analysis was performed stratifying by current ILI. All available contact pattern data was used in subsequent analysis but not stratified by current symptoms.

Mean number of reported face to face conversations and physical contacts were 4.73 (95% CI 4.28-5.17) and 3.49 (95% CI 3.24 - 3.73) respectively. However this was highly variable with several individuals reporting upwards of 50 contacts per day (Figure 16). There were significant differences in the number of reported contacts in different age groups (Kruskal Wallis χ^2(3)=13.497, p-value 0.003 and χ^2 (3)=21.894, p-value <0.001 respectively). The mean number of contacts was highest in the 65 and over age group, however this was from only four reports of contact details. Individuals under the age of 45 had a wider distribution of number of contacts and reported more contacts outside the home (Table 3.7).
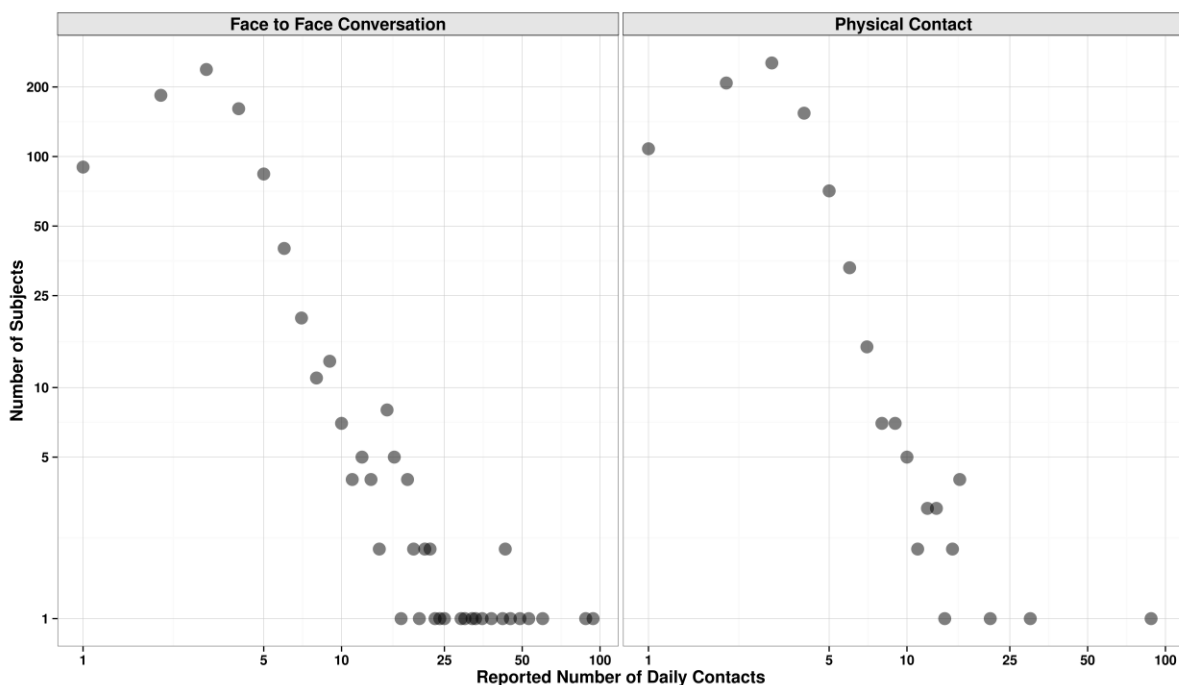
Figure 16 Log-Log Distribution of Number of Daily Contacts Reported by Participants

Age Category of Subject	n	Face to Face Conversation			Physical Contact		
		Total	Home	Work	Total	Home	Work
5 to 19	263	3.236 (3.041-3.43)	3.236 (3.041-3.43)	2.065 (1.243-	3.862 (3.529-	3.173 (2.975-	0.573 (0.296-0.85)
20 to 44	556	3.121 (2.769-	3.121 (2.769-	0.998 0.549-	3.409 (3.043-	2.927 (2.565-	0.146 (0.066-
45 to 64	83	2.916 (2.574-	2.916 (2.574-	0.88 (-0.55-	2.805 (2.493-	2.683 (2.39-2.976)	0 (NaN-NaN)
65+	4	3.25 (0.532-	3.25 (0.532-	0 (NaN-NaN)	5 (2.516-	4 (1.516-	0 (NaN-NaN)

Table 3.7 Mean Number of Reported Contacts by Age Group with 95% CI of mean. 'Work' includes school/university attendance.

3.3.7 Baseline Serology

Baseline serology was available for 65.8% of recruited subjects (n=627/953) as per the planned testing protocol (Table 3.8). All subjects with a PCR detectable subtype for influenza A had serology available. Approximately 50% of influenza B and influenza negative samples were tested, age and sex distribution of these subjects was similar to influenza A positive participants as per the planned matching process. The distribution of study week of recruitment was different between infecting subtypes consistent with the peaks of activity seen in Figure 9. Samples where an influenza A subtype could not be identified were excluded from the subsequent analysis unless explicitly stated.

	Influenza A H1N1 n (%)/ med (IQR)	Influenza A H3N2 n (%)/ med (IQR)	Influenza A NT n (%)/ med (IQR)	Influenza B n (%)/ med (IQR)	Influenza Negative n (%)/ med (IQR)
Number	39 (100)	223 (100)	9 (75)	62 (45.6)	294 (54.1)
Age	28.4 (21.5-37.45)	26.2 (18.25-34.1)	31.3 (26.6-40.4)	24.2 (17.8-34.1)	24.6 (19.9-33)
Female	23 (59)	94 (42.2)	5 (55.6)	30 (48.4)	124 (42.2)
Study Week	85 (31.5-89)	41 (38-44)	31 (21-35)	50 (34.5-54)	41 (28-63)

Table 3.8 Baseline Serology Samples by Influenza Subtype. Number of serum samples is presented as percentage of total study numbers.

3.3.7.1 Serology to Most Recent Circulating Strains

The log2 transformed titres to both H1N1 2009 and H3N2 2011 are normally distributed but have censored values. In contrast to HI titres the microarray titre is on a continuous scale. Left censoring at titre 10 is seen more frequently in H1N1 2009, whereas right censored at titre 1810 is more common in H3N2 2011 (Figure 17). A value of 10 represents no detectable serological response on the microarray and likely means this individual has not been exposed to H1N1 2009 or has any detectable cross reaction to this. The highest measurable dilution on the microarray is 1280, any sample which has a reaction greater than this is censored at 1810 so these samples have a value greater than 1280 but the true value is not known. The mean titre value of H3N2 2011 is significantly greater than H1N1 2009 (log 2 titre 3.57 vs 5.16 (GMT 119 vs 354), 95% CI of difference 1.37-1.77. p

value <0.001). This remains true on exclusion of censored values (log 2 titre 4.14 vs 4.6 (GMT 177 vs 244), 95% CI 0.29-0.63. p value <0.001).
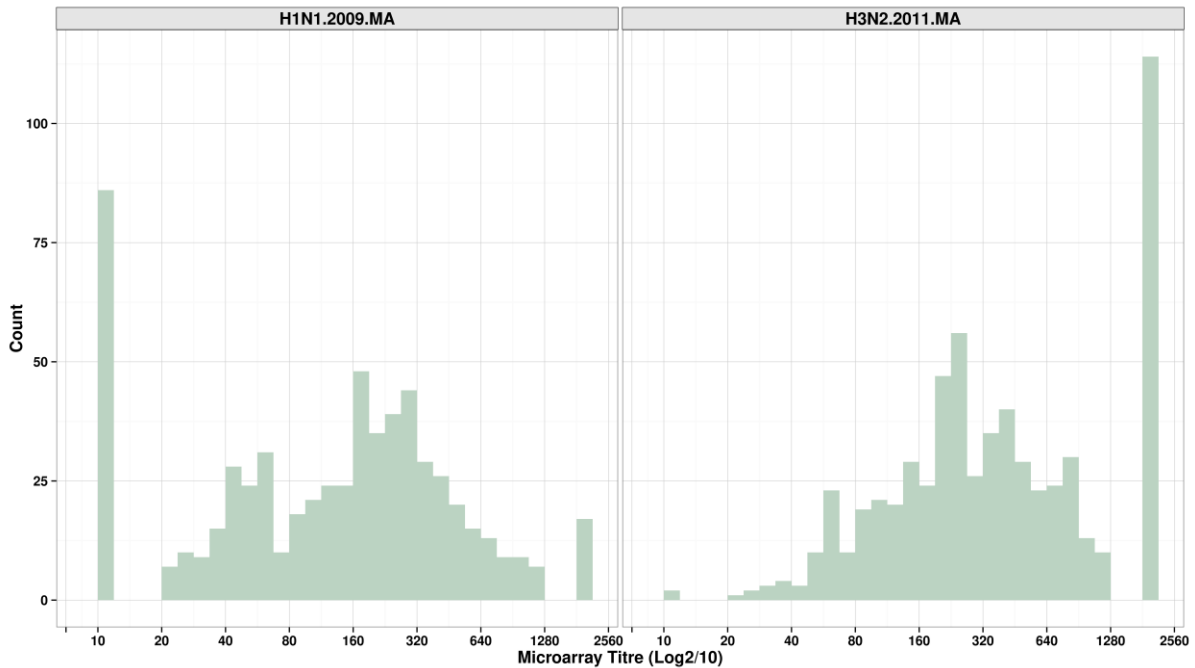
Figure 17 Histogram of Baseline Serology, all samples. Values are presented on log2 scale.

The baseline titre to recent strains was significantly associated with subject's age at recruitment. In simple linear regressions, mean baseline titre reduced as age increased (H1N1 2009 p value<0.001, H3N2 2011 p value 0.018) (Figure 18). The model fit for both H1N1 and H3N2 was improved by using a non-linear spline which suggests a more rapid decline in titre followed by a plateau around age 25 (ANOVA model comparison, both p <0.001). Despite this improvement in model fit, the percentage deviance explained remains low, even with non-linear fit (H1N1 7.1 % (Adjusted R^2 0.0606); H3N2 3.9% (Adjusted R^2 0.0326)). Although both models have reasonable predictions of the mean response with narrow confidence intervals, it fails to take into account the range of individual response. There was no effect of age on censored titre values (logistic regression, p value >0.05). In multivariable analysis, there was no effect of gender, pregnancy, chronic respiratory disease or smoking on baseline titre. Having previously received influenza vaccine significantly increased baseline H1N1 2009 titre (1.61 (0.16-3.06), p 0.03) but not H3N2 (0.75 (-0.42-1.9) p 0.212) in multivariable model with age.
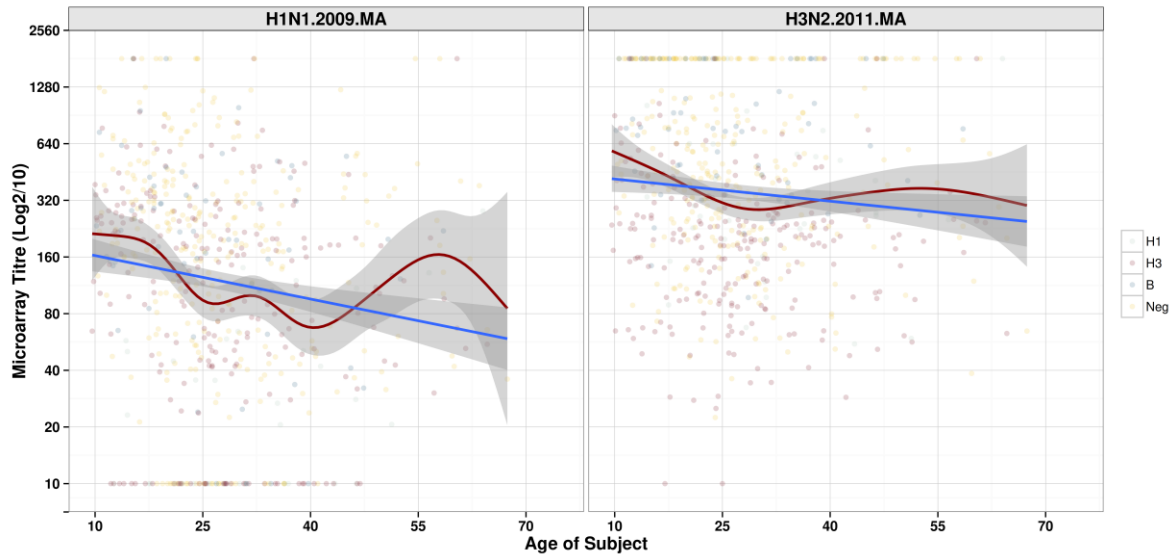
Figure 18 Titre to Recent Strains by Age of Subject. Blue line linear regression; Red line general additive model with spline for age. Points represents titre measurements from study, colour reflects infecting subtype.

Two way ANOVA was performed to examine baseline titre response by infecting subtype and age category. Main effects for both of these factors were significant (Table 3.9, Table 3.10). There was no significant interaction between the factors (H1N1 2009 $F_{(6,604)}=0.70$, p 0.652; H3N2 2011 $F_{(6,604)}=0.538$, p 0.78).

Baseline titre response was significantly different when stratified by infecting subtype (Table 3.9). A subtype specific effect is seen where the lowest titre response to that strain is seen in subjects infected with that subtype (Tukey HSD post-hoc testing). Those infected with H1N1 had a mean microarray titre to H1N1 2009 of 26 (95% CI 18-36) suggesting this is the first time these individuals have been infected with the H1N1 pandemic strain. Most individuals infected with H3N2 already had detectable titre to H3N2 2011 (mean 183 (95% CI 162-206)) but this was still significantly lower than those who were not infected with H3N2. Of note, the highest response to H1N1 2009 and H3N2 2011 was seen in individuals infected with influenza B. There remained a significant difference between groups when the homosubtypic infection group is removed for both subtypes (p values <0.005).

	Influenza A H1N1	Influenza A H3N2	Influenza B	Influenza Negative	ANOVA
H1N1 2009 Mean (95%CI)	26 (18-36)	96 (80-114)	169 (123-232)	160 (137-186)	$F_{(3,604)}=28.0$, p <0.001
H3N2 2011 Mean (95%CI)	359 (260-497)	183 (162-206)	731 (599-891)	501 (443-566)	$F_{(3,604)}=60.0$, p <0.001

Table 3.9 Mean Titre Response to Most Recent Strains by Infecting Subtype. ANOVA performed on log2 titres, transformed titres presented for clarity.

Mean titre response was also significantly different when stratified by age category when assessed by ANOVA (Table 3.10). On post-hoc testing (Tukey HSD) significant differences were seen between 10 to 19 age group and 20 to 45 age group for both H1N1 and H3N2 (p values <0.001) but not between other age groups.

	5 to 9	10 to 19	20 to 45	45 to 65	ANOVA
H1N1 2009 Mean (95%CI)	88 (2-4139)	208 (171-252)	94 (82-108)	120 (88-163)	$F_{(4,604)}=10.6$, p <0.001
H3N2 2011 Mean (95%CI)	203 (28-1443)	495 (418-588)	304 (274-338)	392 (289-532)	$F_{(4,604)}=8.4$, p <0.001

Table 3.10 Mean Titre Response to Most Recent Strains by Age Category. All subjects included in analysis. ANOVA performed on log2 titres, transformed titres presented for clarity. Category 65 over removed as CI incalculable due to small number.

Logistic regression was performed to investigate the level of protection offered by microarray titre at baseline. The outcome of this analysis was susceptibility to PCR positive influenza. Analysis was performed for H1N1 and H3N2 separately. An indicator variable of one is given to individuals infected with that subtype, i.e. in H3N2 analysis individuals infected with H3N2 are coded as one and individuals not infected (all of H1N1 PCR positive, influenza B positive and influenza A & B negative) are coded as zero.

Univariate logistic regression for the effect of baseline H3N2 2011 on susceptibility to PCR confirmed H3N2 was statistically significant. A one unit increase in log2 titre resulted in a halving of susceptibility (OR 0.48, 95% CI 0.42-0.56, Wald Z Statistic -10.05, p value <0.001). This gave a 50% threshold of protection of log2 titre 4.13 (95% CI 3.85-4.41; non-logged titre 175 (95% CI 143-213)) (Figure 19). Multivariable logistic regression was performed to investigate the effect of age, gender, respiratory disease and smoking on seroprotection. No additional significant factor was found.

Logistic regression for the effect of baseline H1N1 2009 titre of protection was also statistically significant. A one unit rise in log2 titre also resulted in an approximate halving of susceptibility (OR 0.54, 95% CI 0.44-0.65, Wald Z Statistic -6.31, p value <0.001). However, there were far fewer H1N1 PCR confirmed infections and of the individuals who were not infected with H1N1 who had low titres to H1N1 2009 suggesting they had never been infected with H1N1 in the post pandemic period. This gave an unusual appearance of the logistic regression curve and a nonsensical 50% threshold of protection of -1.81 (95% CI -3.0 - -0.53; non-logged titres 2.84 (95% CI 1.25-6.45)) (Figure 19).
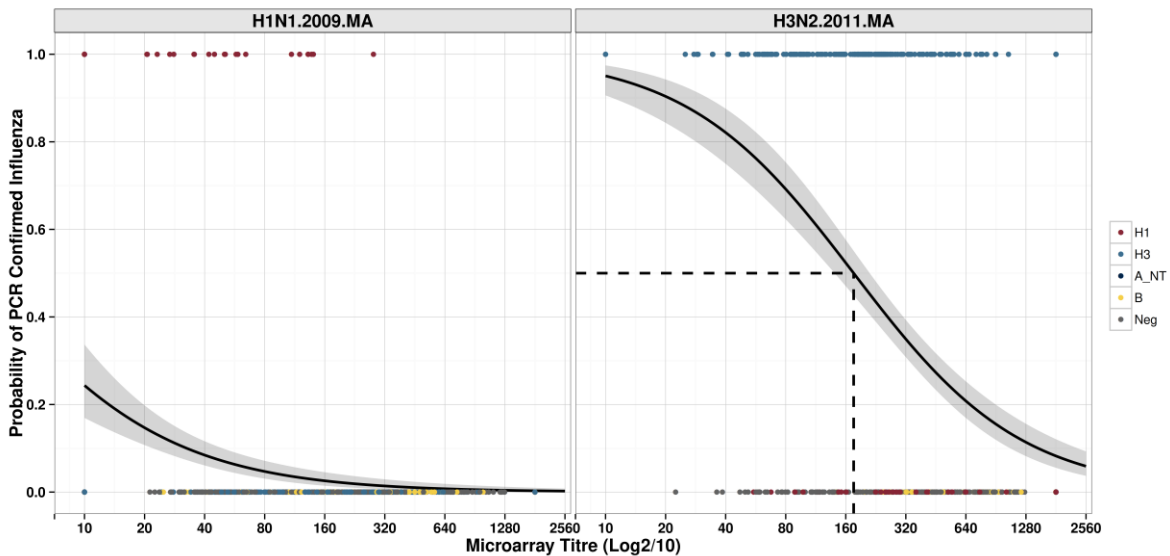
Figure 19 Logistic Regression for Protection Against Influenza A Subtype by Titre Response to Recent Circulating Strain. LR performed on log2 titres. Points represents titre measurements from study, colour reflects infecting subtype.

The risk of having PCR confirmed influenza A also depends on whether influenza is present in the population at that time. As shown in Figure 9 the risk of exposure fluctuated through the study and so incorporating this into a model of susceptibility is important. Initially a subtype specific indicator variable was used to indicate if there was substantial levels of influenza detected. If study week ILI percentage was greater than 10% (Figure 10) then this indicator was one. The presence or absence indicator was not significant for either H1N1 or H3N2 when included in logistic regression (p values >0.3).

As the underlying risk of exposure is not known, study time (week of study recruitment) was used as a proxy for changing force of infection. General additive models (GAM) were used to investigate non-linear effects of titre and time. Inclusion of study week in a model of H3N2 susceptibility led to a much improved model fit compared to one that only included titre to H3N2 2011 (R^2 0.469, AIC 487 vs R^2 0.21, AIC 669). A similar improvement is seen when fitting study time for H1N1 susceptibility (R^2 0.479, AIC 142 vs R^2 0.09, AIC 243). Fitting splines with an interaction term did not improve model fit compared to two separate splines for either H3N2 or H1N1.

Inclusion of an age term gave a marginal improvement in model fit for H3N2 (R^2 0.491, AIC 476) and a significant spline term for age (p value 0.006). This spline demonstrated increased odds of having PCR confirmed H3N2 if aged below 20 for the same titre to H3N2 2011 and same week of recruitment. Inclusion of age in model for H1N1 did not change model fit and spline term for age was non-significant. Using age category as a factor rather than a continuous age variable was non-significant for both H3N2 and H1N1.
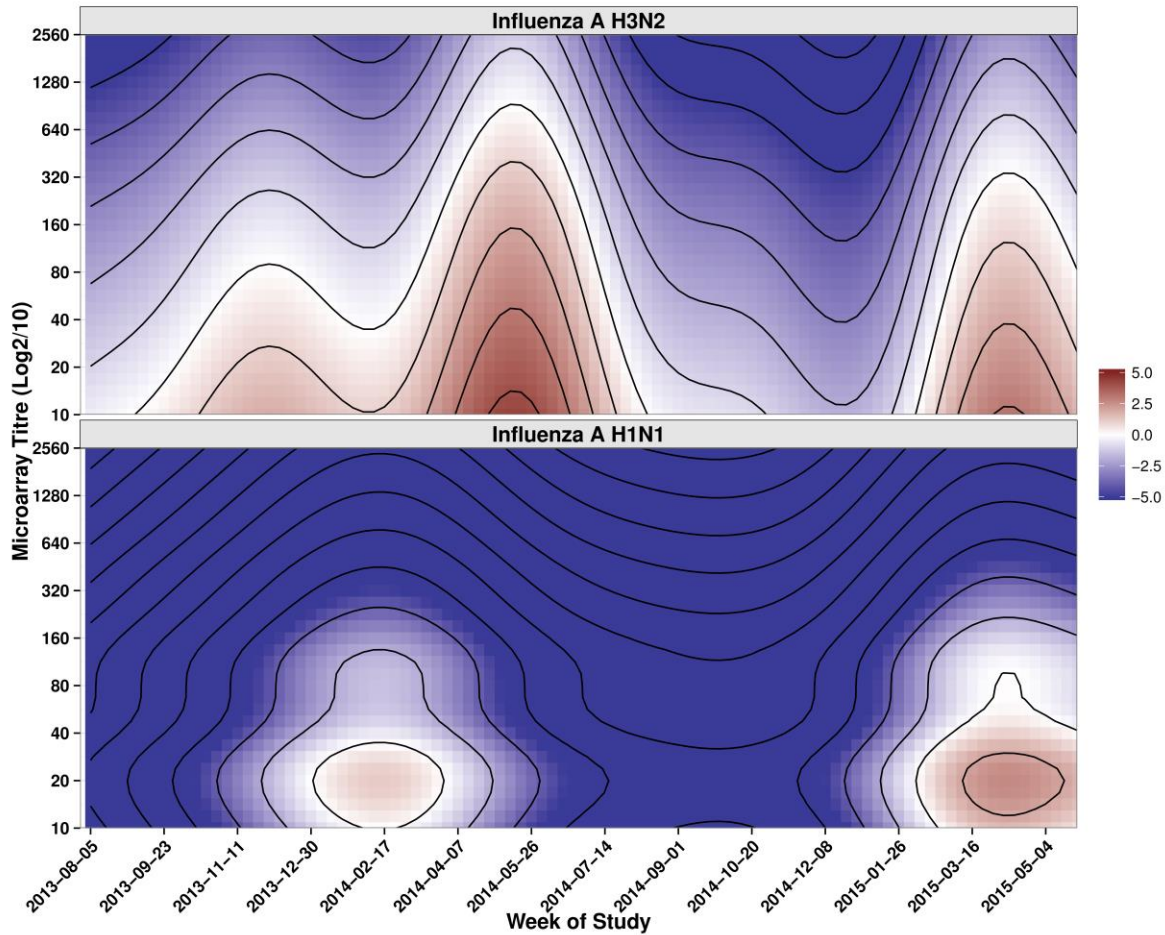
Figure 20 Log odds of having PCR confirmed influenza by titre at week of study for 30 year old. Log odds below -5 censored for clarity. Minimum log odd -15.

Using the optimally fitted models, the log odds of having PCR confirmed influenza was assessed across the full range of titre values and study weeks (age was fixed at 30) (Figure 20). The log odds of PCR confirmed influenza for a specific titre clearly changes in relation to how much influenza is circulating (Figure 9, Figure 10). As would be expected, when a subtype is not circulating the log odds of having PCR confirmed influenza in reduced (represented by blue in heat-map). However, when influenza is circulating the log odds of being infected for a particular titre varies depending on the relative force of infection. For example, in the week beginning 16 December 2013, an individual with H3N2 2011 log2 titre of 4 (160) had a log odds of PCR positive H3N2 -0.92 (95%CI -1.79 - -0.05; OR 0.39, 0.17-0.95). For an individual in week beginning 19 May 2014 with log titre of 4, their log odds of being PCR positive for H3N2 is 1.70 (95%CI 1.2-2.2; OR 5.49; 3.32-9.04). If the level of protection offered by a specific titre was independent of force of infection then it would be expected that the log odds would be unchanged for a higher force of infection only that more people below the level of protection would be infected. Within this study the probability of being infected when you had higher titres to the most recent strains was greater when there was more influenza circulating (Figure 21).
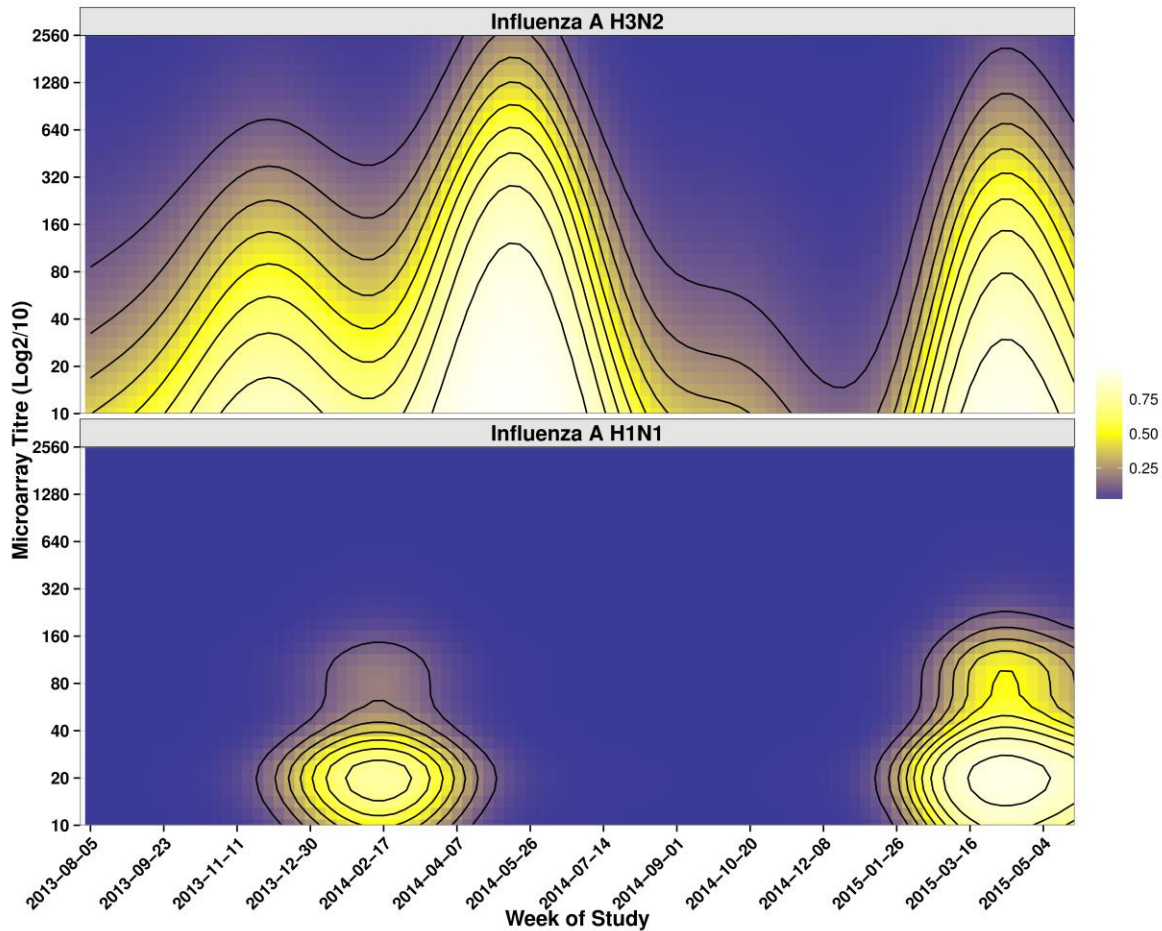
Figure 21 Probability of PCR confirmed influenza A H3N2 by titre to H3N2 2011 and study week for 30 year old.

3.3.7.2   Serology to Historical Human Strains

The log2 transformed titres to historic human strains are normally distributed with censored values (Figure 22). Significant differences in mean titre between different strains are seen as assessed by one way ANOVA $F_{(8, 5553)}$=239.7, p <0.001). Post-hoc testing on difference in mean titre (Tukey HSD) showed significant differences for all pairwise comparisons except H1N1 1918 & 1977, H1N1 1999 & 2007, H3N2 1968 & 2007, H3N2 2007 & 2009, and H3N2 1968 & H1N1 1999. Mean responses to historical H3N2 strains were higher than to historical H1N1 even when time since emergence is considered.

The relationship between baseline titre to historic strains and subjects' age is more complicated than that seen in the age related response to recent strains (Figure 23). Significant linear relationships were seen for all strains with age except H1N1 1999 and H3N2 2007 and 2009. When a non-linear spline of age was fitted, significant relationships were seen for all strains except H3N2 2009. The use of a non-parametric spline improved model fit for all strains as assessed by AIC and ANOVA except H3N2 2009 where optimal model was a linear relationship. The direction of the relationship between age and titre response is related to when the strain emerged, with older strains having a positive

relationship between titre and increasing age. The percentage of deviance explained is much better in older strains as demonstrated by higher R^2 with the exception of H1N1 1918.
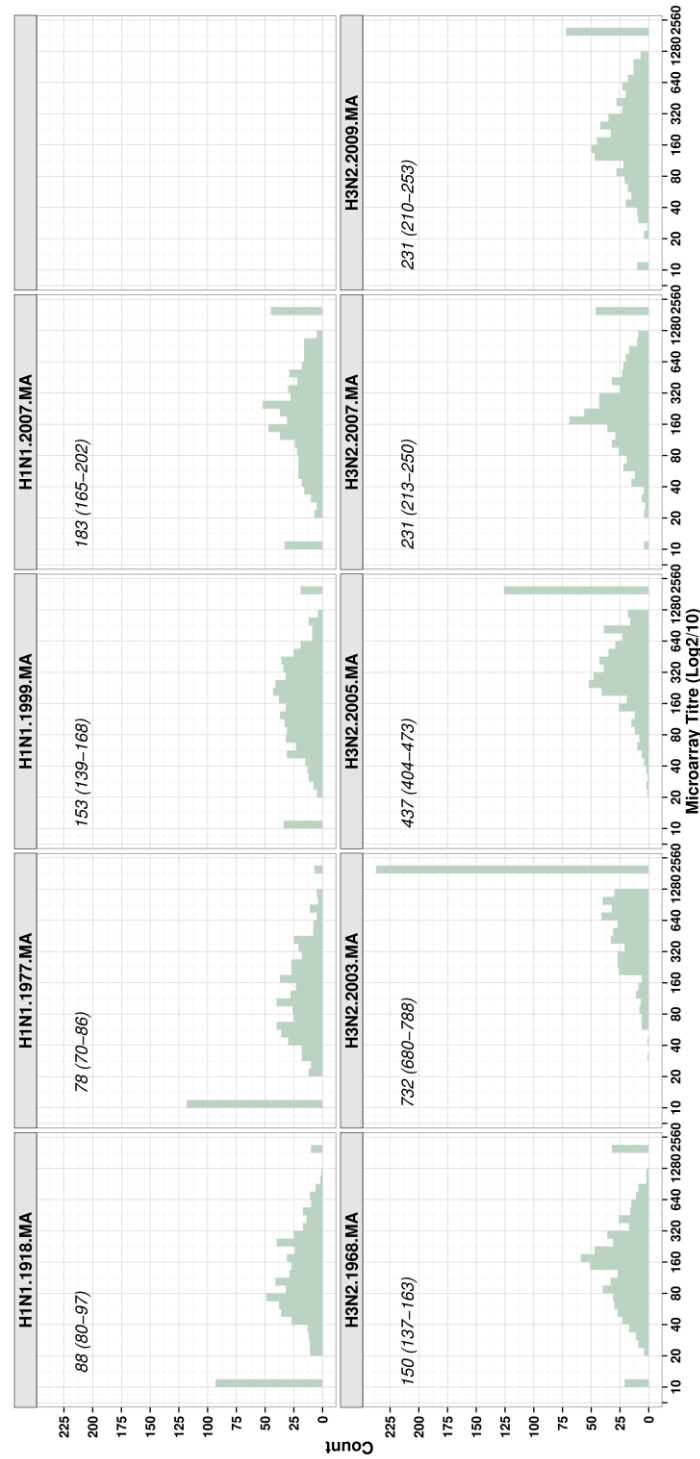
Figure 22 Histogram of Historic Human Strains, all strains. Geometric mean titre & 95% CI of mean displayed in top left.
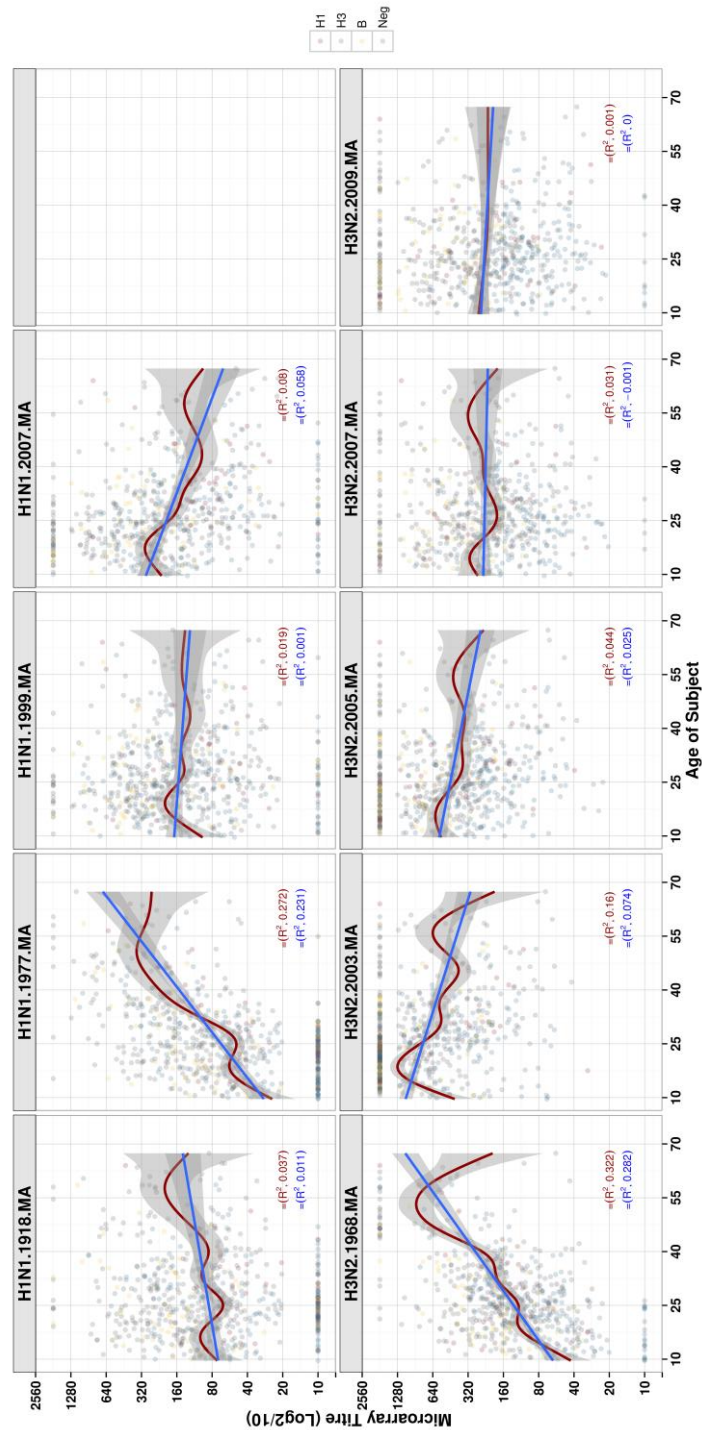
Figure 23 Titre to Historical Human Strains by Age of Subject. Blue line linear regression; Red line general additive model with spline for age, R² for each model included. Points represents titre measurements from study, colour reflects infecting subtype.

As would be expected, changing the age parameter from age at study recruitment to age at strain emergence does not change the overall quality of the model fits for individual historic strain models. GAM models were fitted looking for overall prediction of historical strain response according to age with individual intercepts for each strain (Figure 24). Age and all strains except H1N1 1977 had a statistically significant intercept and again the flexible fitted spline was better than linear age term

as assessed by AIC and ANOVA (spline model GCV 2.67, adjusted R^2 0.273, AIC 21247, p values <0.001). This model was improved by changing age at presentation to age at the time of strain emergence (spline model GCV 2.49, adjusted R^2 0.322, AIC 20861, p values <0.001). This shows that even accounting for strain differences, an individual's highest titre response appears to be to strains that were circulating when the subject was under 20 years of age.

Significant differences in mean response to historical strains were seen between the different infection groups (Table 3.11). Again a subtype specific response is seen with the lowest responses in historical strains of the subtype individuals were infected with. There remained significant differences between infecting subtypes when the homosubtypic response was removed from the ANOVA for all historical strains (p value <0.05) with the exception of H3N2 1968 ($F_{(3,392)}$=1.46, p value 0.233).

	Influenza A H1N1	Influenza A H3N2	Influenza B	Influenza Negative	ANOVA
H1N1 1918 Mean (95%CI)	38 (27-54)	65 (56-76)	136 (99-185)	113 (98-131)	$F_{(3,614)}$=17.6, p <0.001
H1N1 1977 Mean (95%CI)	52 (35-76)	63 (53-75)	102 (74-140)	92 (79-107)	$F_{(3,614)}$=5.78, p <0.001
H1N1 1999 Mean (95%CI)	113 (86-149)	118 (102-137)	220 (166-292)	179 (156-205)	$F_{(3,614)}$=8.61, p <0.001
H1N1 2007 Mean (95%CI)	115 (84-157)	142 (121-166)	262 (193-356)	218 (187-253)	$F_{(3,614)}$=8.64, p <0.001
H3N2 1968 Mean (95%CI)	153 (111-210)	97 (84-112)	218 (172-276)	192 (170-216)	$F_{(3,614)}$=20.9, p <0.001
H3N2 2003 Mean (95%CI)	606 (442-833)	513 (452-582)	990 (819-1197)	923 (836-1018)	$F_{(3,614)}$=21.4, p <0.001
H3N2 2005 Mean (95%CI)	400 (289-554)	258 (230-290)	774 (640-936)	584 (523-653)	$F_{(3,614)}$=42.02, p <0.001
H3N2 2007 Mean (95%CI)	236 (174-320)	131 (117-146)	396 (325-482)	317 (282-356)	$F_{(3,614)}$=46.5, p <0.001
H3N2 2009 Mean (95%CI)	226 (159-321)	107 (96-121)	461 (367-578)	357 (314-405)	$F_{(3,614)}$=71.0, p <0.001

Table 3.11 Mean Titre Response to Historical Human Strains by Infecting Subtype. ANOVA performed on log2 titres, transformed titres presented for clarity.
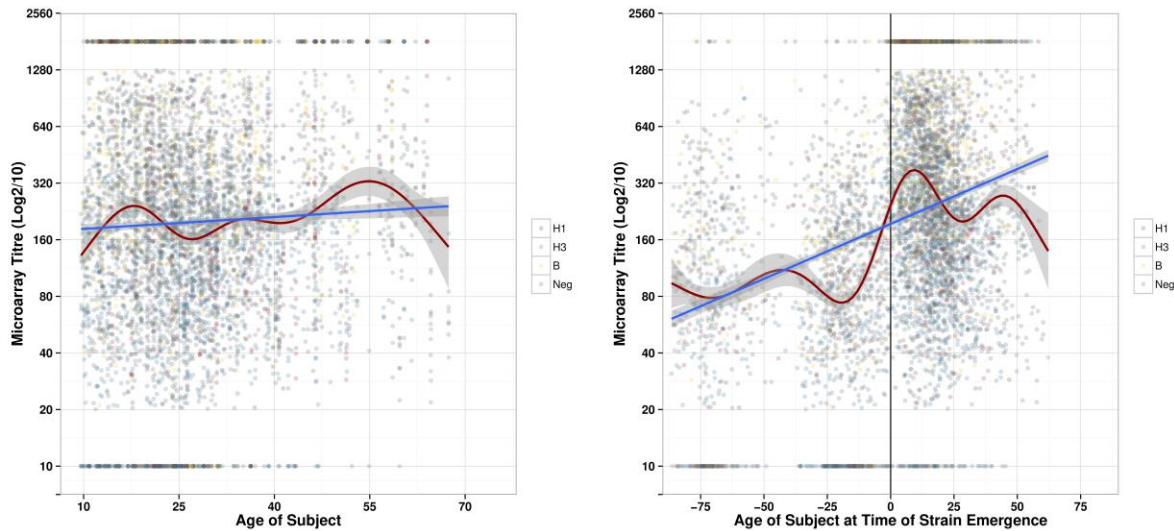
Figure 24 Titre to Historical Human Strains by Age at Emergence for strain. Adjusted for individuals strain effect. Blue line linear regression; Red line general additive model with spline for age. Points represent titre measurements from study, colour reflects infecting subtype.

Logistic regression was used to investigate the influence of historic titres at baseline on susceptibility. Outcome of this analysis was susceptibility to PCR positive influenza. Analysis was performed only for H3N2. In order to investigate the original antigenic sin hypothesis, a variable of "Earliest Strain" was created for the strain which emerged closest to an individual's first 10 years of life. 40% of study subjects had H3N2 1968 as their earliest H3 strain (n=385), 33.4% had H3N2 2003 (n=319), 6.7% H3N2 2005 (n=64), 5.3% H3N2 2007 (n=51) and 14% H3N2 2009 (n=134). In univariate analysis, all historic strains and the earliest strain were significant predictors of having PCR confirmed H3N2 (Table 3.12).

	Univariate		Multivariable	
	OR (95% CI)	p value	OR (95% CI)	p value
H3N2 1968	0.65 (0.570-0.727)	<0.001	1.056 (0.886-1.26)	0.544
H3N2 2003	0.636 (0.558-0.724)	<0.001	0.951 (0.76-1.182)	0.651
H3N2 2005	0.528 (0.457-0.604)	<0.001	1.53 (1.039- 2.262)	0.032
H3N2 2007	0.473 (0.402-0.55)	<0.001	0.91 (0.624-1.311)	0.602
H3N2 2009	0.428 (0.364-0.499)	<0.001	0.310 (0.207-0.45)	<0.001
H3N2 2011	0.489 (0.433-0.560)	<0.001	1.106 (0.718-1.715)	0.64
Earliest H3N2	0.663 (0.595-0.735)	<0.001	0.907 (0.786-1.046)	0.179

Table 3.12 Odds Ratio for PCR confirmed H3N2 for each 1 unit rise of baseline log2 titre. Earliest H3N2 is the strain which emerged closest to the first decade of life.

When a multivariable analysis was fit, only H3N2 2005 and 2009 remained significant with a higher baseline titre to H3N2 2005 increasing the odds of having PCR confirmed influenza. The maximal model of all strains, earliest strain and age was fitted using splines to look for any non-linear relationships between strains. As in the linear multivariable analysis only 2005 and 2009 strains were significant (R^2 0.306, deviance explained 28.9%, AIC 624) (Figure 20).

In order to investigate the optimal combination of strains, multivariable analysis was carried out for each unique combination of historic strains. For H3N2 this meant 31 models were fitted. Model fit was ranked by AIC (Table 3.13). Using historic strains gave a better model fit than fitting only a spline to the most recent H3N2 2011 (AIC 669.7, R^2=0.21, deviance explained=18.1%). Fitting with only H3N2 2009 gave a better explanation of susceptibility than only including the 2011 strain in the model. These two strains are highly correlated (Pearson's Correlation 0.93 (95%CI 0.92-0.94)) and this is consistent between all age groups. Titre response to H3N2 2011 is significantly higher than H3N2 2009 (5.15 vs 4.53 (95%CI of difference 0.42-0.80), p value <0.001) but with similar variance (2.54 vs 2.82 (F test ratio 0.90 (0.77-1.06) p value 0.197).

Strains included in Model	AIC/δAIC	aRsq	Dev Exp
H3N2 2005, H3N2 2009	627.196	0.265	0.236
H3N2 2003, H3N2 2005, H3N2 2009	0.944	0.266	0.237
H3N2 1968, H3N2 2005, H3N2 2009	1.633	0.266	0.238
H3N2 2005, H3N2 2007, H3N2 2009	1.866	0.265	0.236
H3N2 1968, H3N2 2003, H3N2 2005, H3N2 2009	2.694	0.266	0.239
H3N2 2003, H3N2 2005, H3N2 2007, H3N2 2009	2.716	0.266	0.238
H3N2 1968, H3N2 2005, H3N2 2007, H3N2 2009	3.483	0.265	0.239
H3N2 1968, H3N2 2003, H3N2 2005, H3N2 2007, H3N2 2009	4.508	0.266	0.24
H3N2 2009	5.078	0.254	0.228
H3N2 2007, H3N2 2009	6.376	0.253	0.228

Table 3.13 Logistic Regression Model Fit for PCR confirmed H3N2. Spline for Historic strains only. Top ten models presented. Delta AIC=AIC_i – AIC_{min}.

All combinations of historical H3N2 strains were then then refit in combination with H3N2 2011(Table 3.14). There was no improvement in the explanatory power of models which included the more recent strain and the penalty of the additional term is seen in the slightly higher AIC. The exclusion of H3N2 2009 in the models results in a considerable drop in fit quality. The best fitting model which excluded H3N2 2009 (H3N2 2011 & 1968) had an adjusted R^2 of 0.218 and an AIC of 667.

The inclusion of an age did not result in an improvement in the best fitting models with no change in AIC and an approximately 1% improvement in deviation explained compared to models with the same strain but without an age term. The age term was non-significant in the optimally fitted models however was significant in strain combinations with poorer explanatory power ($R^2 < 0.2$). These strains did not contain H3N2 2009 in the strain combinations.

Strain included in Model	AIC/δAIC	aRsq	Dev Exp
H3N2 2011, H3N2 2005, H3N2 2009	626.506	0.271	0.243
H3N2 2011, H3N2 2003, H3N2 2005, H3N2 2009	0.865	0.272	0.244
H3N2 2011, H3N2 2005, H3N2 2007, H3N2 2009	1.461	0.271	0.244
H3N2 2011, H3N2 1968, H3N2 2005, H3N2 2009	1.982	0.27	0.243
H3N2 2011, H3N2 2003, H3N2 2005, H3N2 2007, H3N2 2009	2.245	0.273	0.245
H3N2 2011, H3N2 1968, H3N2 2003, H3N2 2005, H3N2 2009	2.779	0.271	0.244
H3N2 2011, H3N2 2009	2.926	0.265	0.236
H3N2 2011, H3N2 1968, H3N2 2005, H3N2 2007, H3N2 2009	3.414	0.27	0.244
H3N2 2011, H3N2 1968, H3N2 2003, H3N2 2005, H3N2 2007, H3N2 2009	4.236	0.271	0.245
H3N2 2011, H3N2 1968, H3N2 2009	4.827	0.264	0.236

Table 3.14 Logistic Regression Model Fit for PCR confirmed H3N2. Spline for Most Recent and Historic strains. Top ten models presented. Delta AIC=$AIC_i - AIC_{min}$.

From this point onwards the full model (all H3N2 strains, titre to earliest H3N2 and age) and the best reduced model (H3N2 2005 and 2009 only) were used. Direct comparison between these two models shows the full model is significantly better with reduced residual deviance (575 vs 618, ANOVA, 0.002). Despite this, in the full model the only significant terms were for 2005 and 2009 strains. Comparison of the changing log odds of being PCR positive for H3N2 shows the highest odds

are for individuals with high titre to 2005 strain and low titres to 2009 (Figure 25). These odds appear to be even higher than for individuals who have low titre for both.

With the addition of study time, both models improve significantly as in the analysis for recent strains only. For the best reduced model R^2 increases to 0.514 (47.2% deviance explained, AIC 451) and for the full model R^2 0.528 (49.4% deviance explained, AIC 454). As in the analysis for recent strains, the log odds changes in relation to how much influenza is circulating (Figure 26). This again suggests that a higher force of infection leads to individuals with higher titres becoming infected.
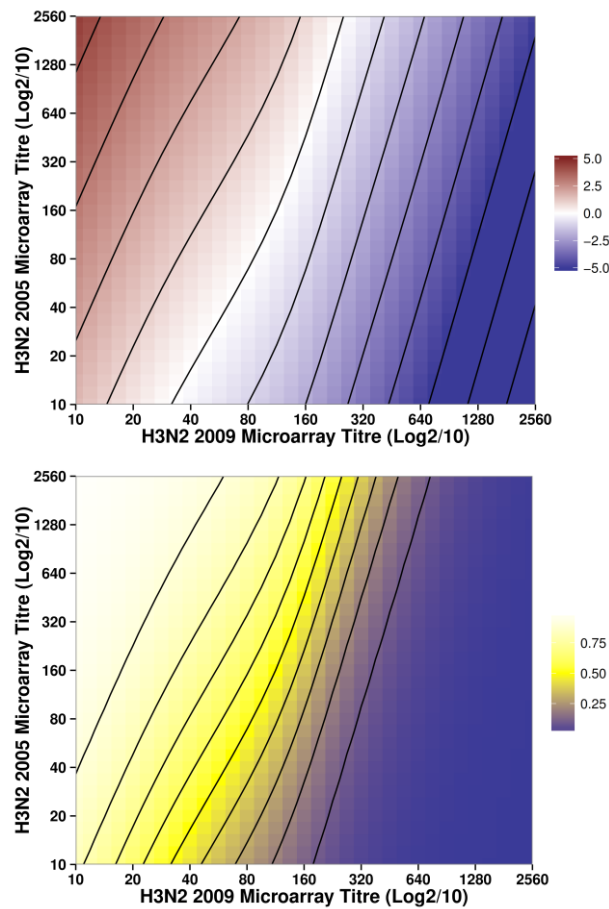
Figure 25 Predicted Log Odds (upper panel) and Probability (lower panel) of H3N2 PCR confirmed influenza by baseline titre to H3N2 2005 and 2009. Log odds censored at -5 for clarity.
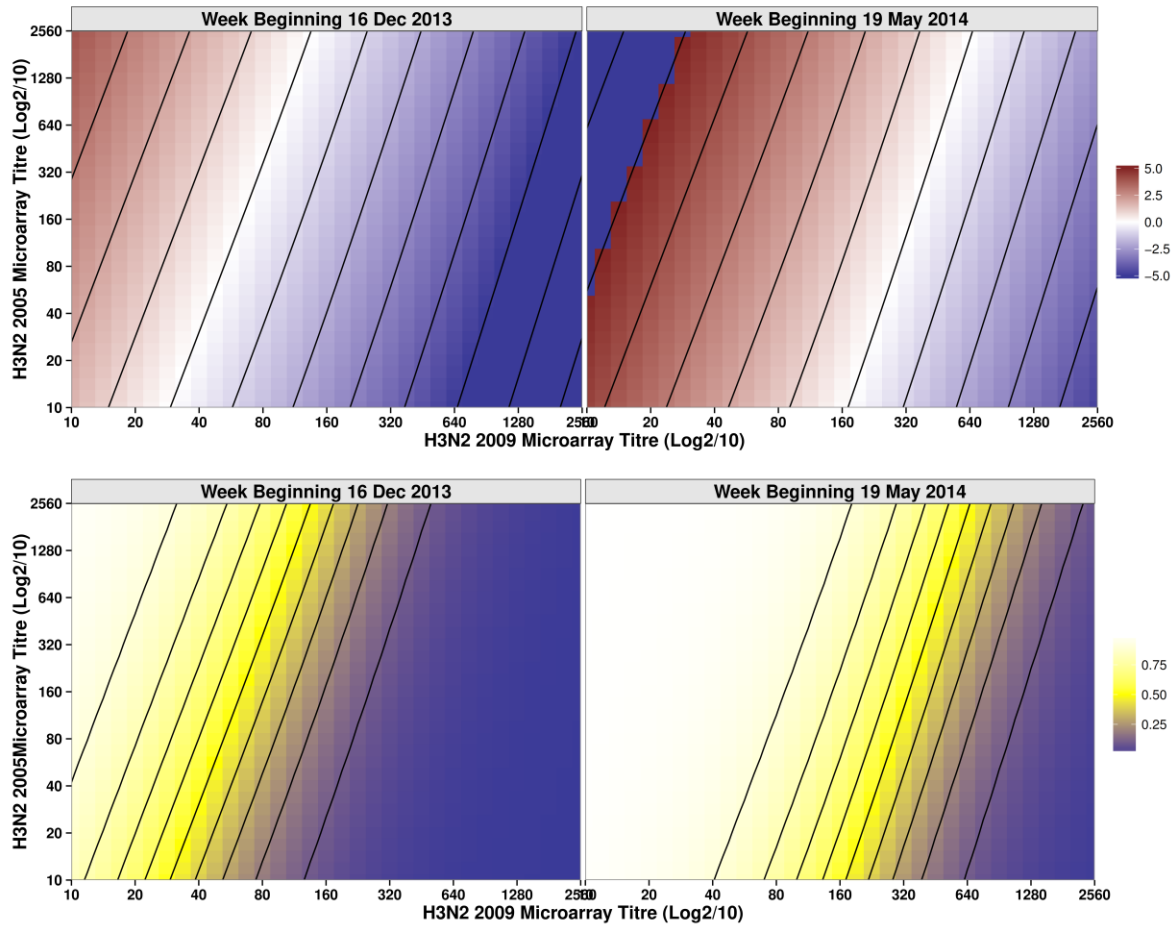
Figure 26 Predicted Log Odds and Probability of H3N2 PCR confirmed influenza by baseline titre to H3N2 2005 and 2009. Model included study week of recruitment, weeks 20 and 42 presented here. Log odds censored at -5 for clarity.

3.3.7.3 Serology to Avian Strains

In contrast to recent and historic human strains, titres to avian strains are heavily skewed to the left. The proportion of samples with measurable titres (log2 titre >0) varied according to strain (Table 3.15). A small number of individuals had very high titres to avian influenza detected. Significant differences were seen in mean (measurable) titre as assessed by one way ANOVA $F(4, 783)=4.482$, p 0.001). Post-hoc testing on difference in mean titre (Tukey HSD) showed only significant difference between H7N7 2003 & H5N1 2007, and H9N2 1999 & H5N1 2007.
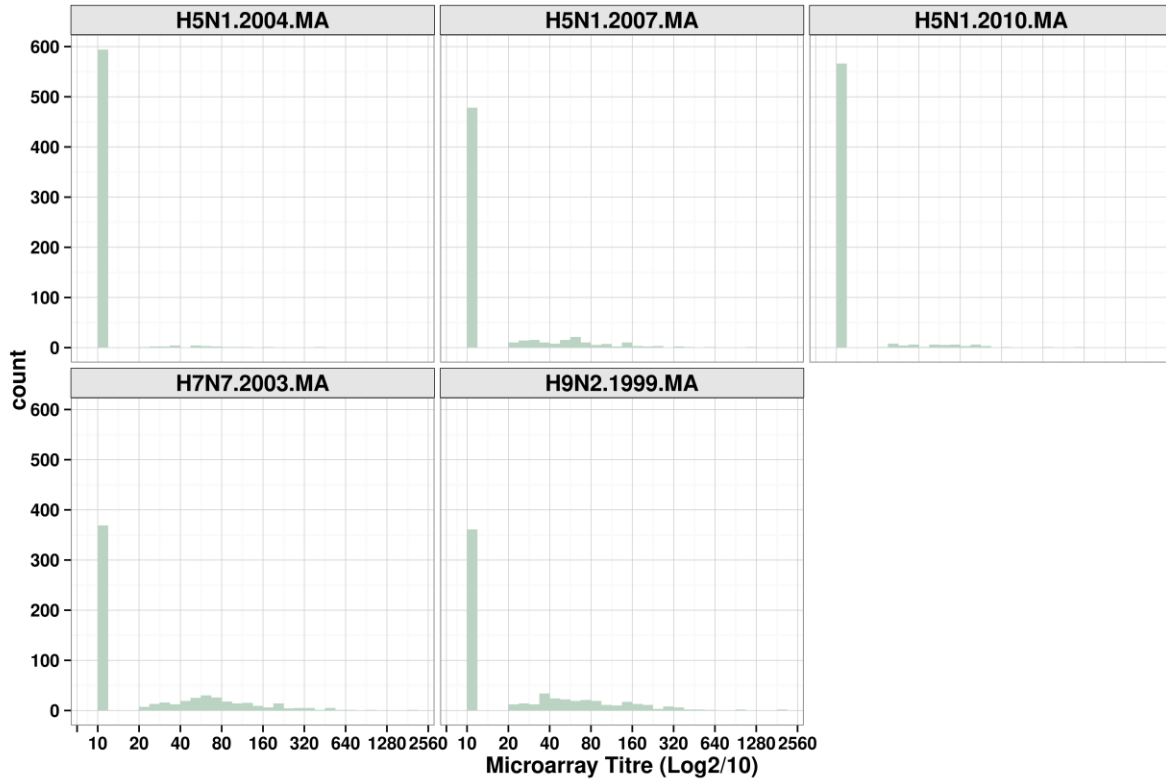
Figure 27 Histogram of Titre Response to Avian Strains. All Study Subjects.

Strain	Number (%)	Mean (95% CI)	Max Titre
H5N1 2004	23 (3.7)	55 (41-74)	418
H5N1 2007	140 (22.7)	59 (52-67)	1223
H5N1 2010	52 (8.4)	58 (48-70)	571
H7N7 2003	248 (40.2)	78 (71-86)	1810
H9N2 1999	265 (42.3)	75 (68-83)	1810

Table 3.15 Baseline Titre Response to Avian Strains with measurable titre. Calculations performed on log 2 titre.

There is a significant relationship between the titre response to avian strains and the age of the study participant (Figure 28). All fitted models were significant (p values <0.05), models with a non-linear spline of age were superior as measured by AIC and ANOVA with the exception of H7N7 2003. The optimal model for this strain is a linear term, even when allowed to fit flexibly the spline model selected for this. All strains had a positive relationship with increasing titre as age increased.
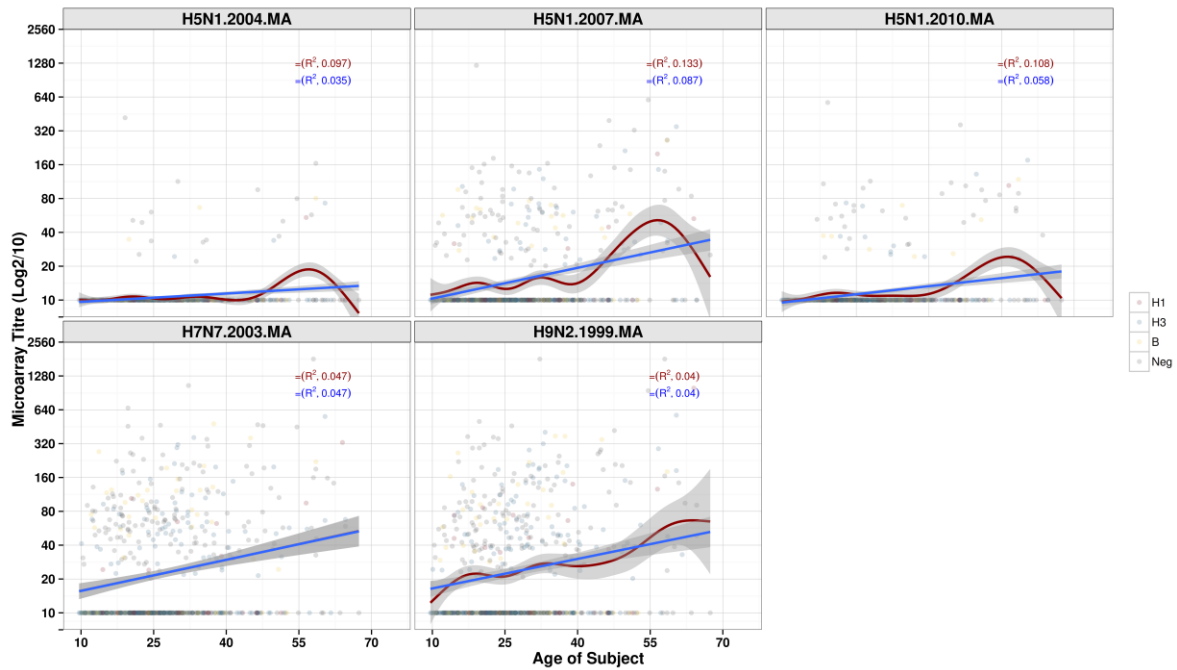
Figure 28 Titre to Avian Strains by Age of Subject. Blue line linear regression; Red line general additive model with spline for age, R^2 for each model included. Points represents titre measurements from study, colour reflects infecting subtype.

Changing the age parameter from age at study recruitment to age at strain emergence does not change the overall quality of the model fits or the explanatory power for individual avian strains. An overall prediction of avian strain response to age compared to age at time of emergence was performed. Age at emergence was a marginally better fit (adjusted R^2 0.187 vs 0.17, p values <0.001) but the difference was not as marked as in historical human strains. In contrast to human strains which demonstrate the highest response to infections early in life, higher responses to avian strains emerge after 35 years of age (Figure 29).
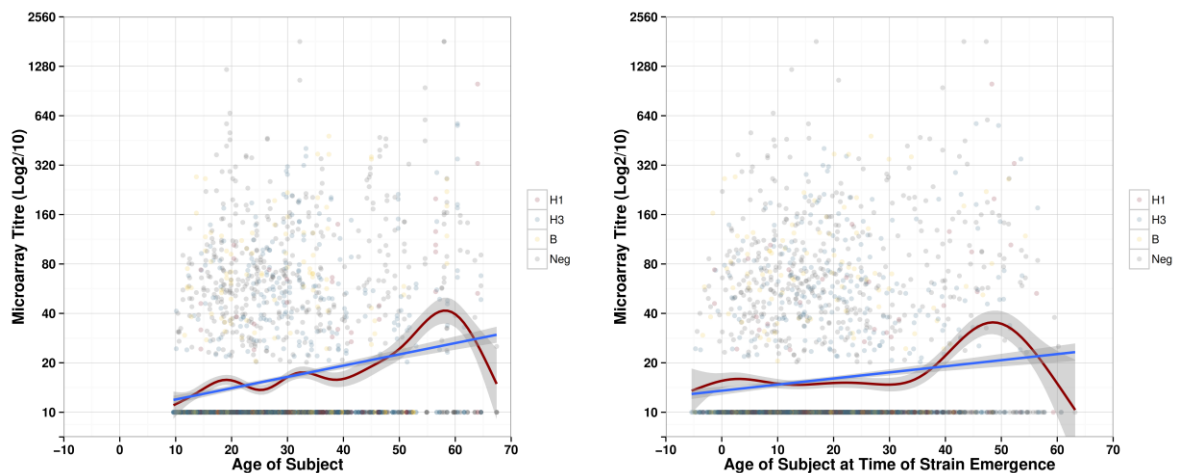
Figure 29 Titre to Avian Strains by Age and Age at Emergence for Strain. Adjusted for individuals strain effect. Blue line linear regression; Red line general additive model with spline for age. Points represent titre measurements from study, colour reflects infecting subtype

The effect of animal contact patterns on titre to avian strains was investigated. In univariate analysis, the only significant effect was an increase in H7N7 2007 associated with weekly contact with live poultry (Table 3.16). In multivariable analysis, age was a highly significant term (all p values <0.001) and the only other significant finding was an association between H5N1 2010 titre and live poultry exposure. However, this showed a significant reduction in titre for both individuals with weekly exposure and those who are never exposed (-0.64, p 0.04 and -0.59 p 0.04 respectively). Given the multiple testing that was performed during this analysis, minimally significant p values and the contradictory nature it is likely this was a chance finding.

	H5N1 2004		H5N1 2007		H5N1 2010		H7N7 2007		H9N2 1999	
	Est	p value	Est	p value	Est	p value	Est	p value	Est	p value
Live Poultry										
never	-0.085	0.602	-0.176	0.645	-0.369	0.132	0.156	0.762	-0.09	0.863
weekly	0.028	0.879	0.347	0.417	-0.224	0.415	*1.305*	*0.024*	0.89	0.129
Dead Poultry										
never	-0.026	0.782	-0.257	0.25	-0.019	0.894	-0.085	0.778	-0.197	0.512
rarely	0.188	0.41	-0.142	0.792	0.081	0.814	-0.576	0.428	-0.309	0.674
weekly	0.022	0.839	-0.289	0.261	0.015	0.925	-0.051	0.883	0.03	0.93
Live Pig										
never	0.093	0.798	0.042	0.96	0.208	0.701	1.174	0.306	0.594	0.533
weekly	0.233	0.556	1.15	0.215	0.962	0.105	*2.456*	*0.05*	1.519	0.161
Dead Pig										
never	-0.083	0.466	-0.293	0.275	0.031	0.857	0.15	0.679	-0.237	0.519
rarely	-0.169	0.594	-0.872	0.241	-0.179	0.707	-1.048	0.297	-1.439	0.158
weekly	-0.035	0.776	-0.263	0.368	0.114	0.544	0.195	0.621	-0.044	0.913

Table 3.16 Univariate Linear Regression on animal contact rates and baseline titre to avian strains.

GAM models were fit to look at the linear effect of human strains on titre response to avian strains (Table 3.17). All models included a non-linear age term which remained significant (Figure 30). Even accounting for age, historical human titre responses were significant predictors of avian strain response. Response to most recent strains were not significant predictors of response. Historical strain response were more predictive of H7N7 and H9N2 response than H5N1 strains with higher R^2. Adding animal contact into the models with human strains and age did not improve model fit and did not add significant terms.

Mean response to avian strains are significantly lower than the response to human strains (Table 3.18). Although a significant F statistic is seen for H5N1 2007, 2010 and H7N7 2003 between infecting subtype this should be interpreted with caution. Differences in titre values at this level are unlikely to be clinically significant, as this is at the limit of detection for the microarray.

	H5N1 2004		H5N1 2007		H5N1 2010		H7N7 2007		H9N2 1999	
	Est	p value	Est	p value	Est	p value	Est	p value	Est	p value
Recent Human Strains										
H1N1 2009	-0.025	0.154	-0.003	0.942	0.001	0.98	0.038	0.424	0.057	0.226
H3N2 2011	-0.034	0.405	-0.015	0.863	-0.048	0.431	-0.07	0.524	-0.08	0.467
Historic Human Strains										
H1N1 1918	0.044	0.062	*0.17*	*0.001*	0.062	0.07	*0.17*	*0.006*	*0.155*	*0.013*
H1N1 1977	0.04	0.051	0.081	0.057	0.044	0.143	*0.29*	*<0.001*	*0.355*	*<0.001*
H1N1 1999	-0.007	0.787	-0.079	0.121	-0.027	0.46	*-0.263*	*<0.001*	*-0.28*	*<0.001*
H1N1 2007	0.01	0.65	*0.09*	*0.049*	0.036	0.269	*0.118*	*0.045*	0.099	0.091
H3N2 1968	-0.01	0.689	0.098	0.066	-0.02	0.6	*0.179*	*0.007*	0.263	<0.001
H3N2 2003	0.012	0.625	0.071	0.163	0.018	0.615	0.09	0.161	0.094	0.145
H3N2 2005	0.054	0.161	0.051	0.52	-0.027	0.628	*-0.302*	*0.003*	*-0.277*	*0.007*
H3N2 2007	-0.01	0.78	-0.062	0.396	0.02	0.694	*0.192*	*0.038*	0.164	0.078
H3N2 2009	0.016	0.642	0.035	0.623	0.09	0.071	*0.241*	*0.008*	0.15	0.094
R² Dev	0.14	0.166	0.329	0.348	0.175	0.198	0.401	0.416	0.421	0.435

Table 3.17 Multivariable GAM Model for Titre to Avian Strains. All models contained non-linear age term which was significant (p values <0.001) Significant terms in italics.

	Influenza A H1N1	Influenza A H3N2	Influenza B	Influenza Negative	ANOVA
H5N1 2004 Mean (95%CI)	10 (10-11)	10 (10-10)	11 (10-12)	11 (10-12)	F(3,614)=2.3, p 0.079
H5N1 2007 Mean (95%CI)	12 (10-15)	13 (12-14)	17 (13-20)	17 (15-19)	F(3,614)=5.68, p <0.001
H5N1 2010 Mean (95%CI)	11 (9-12)	11 (10-11)	12 (11-14)	12 (11-13)	F(3,614)=2.9, p 0.035
H7N7 2003 Mean (95%CI)	15 (11-20)	18 (16-21)	29 (21-40)	27 (24-31)	F(3,614)=8.2, p <0.001
H9N2 1999 Mean (95%CI)	19 (14-27)	21 (18-24)	29 (21-39)	26 (22-29)	F(3,614)=1.97, p 0.098

Table 3.18 Mean Titre Response to Avian Strains. ANOVA performed on log2 titres, transformed titres presented for clarity.
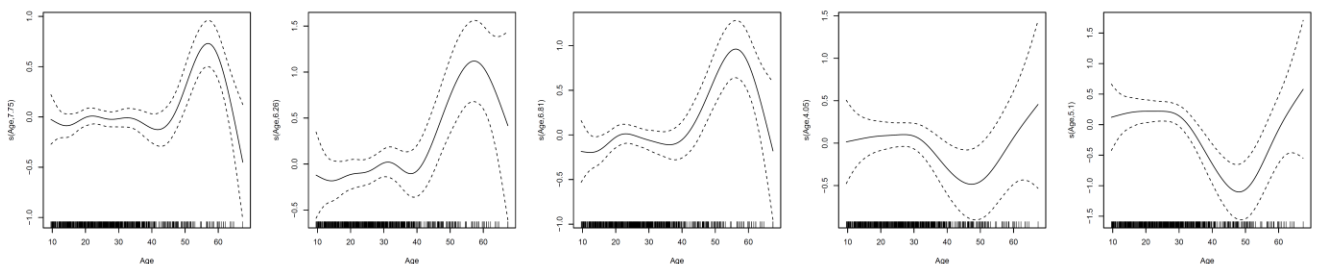
Figure 30 Multivariable GAM Model for Titre to Avian Strains. Strain specific non-linear age terms. All spline functions are significant (p value <0.001).

3.4 DISCUSSION

Influenza surveillance in Vietnam has previously centred on hospitalised patients. This observational study was designed to investigate non-severe influenza and influenza like illness in a tropical, urban primary care setting, the clinical and demographic features associated with it and the impact of pre-existing antibodies to influenza A strains.

Over the 90 weeks of the study, influenza was present the majority of the time. However, this was not a persistence of the same influenza strain but a mixture of high intensity peaks of single subtypes and co-circulation of types and subtypes at variable intensities. This was consistent with previous hospital based surveillance which demonstrated asynchronous peaks and co-circulation of different strains (Li et al., 2008, Nguyen et al., 2009). This study demonstrated higher rates of influenza positivity in ILI than are generally described and fewer periods where no influenza was detected (Kasper et al., 2010, Lutwama et al., 2012, Khamphaphongphane et al., 2013, Snacken et al., 2014). It should be noted that the frequency or duration of circulation of a particular strain does not necessarily reflect the number of cases presenting to clinics. Influenza A H3N2 was found to occur in a similar number of weeks to influenza B but almost 100 more cases were detected. Attack rates measured through healthcare settings are vulnerable to differences in healthcare seeking behaviour (Brooks-Pollock et al., 2011). Influenza B is generally considered to cause less severe symptoms than influenza A H3N2 (Glezen et al., 2013) which could mean that people with mild disease are less likely to present for clinical assessment. This could give an impression of smaller outbreak size despite considerable community transmission (Cowling et al., 2014, Caini et al., 2015). Focusing on hospitalised severe cases will therefore underestimate the circulation of flu in the community and improving surveillance to include non-severe cases in a primary care setting will help us better understand true community attack rates, although this approach has its own challenges and limitations (Ortiz et al., 2009).

One difficulty of utilising community surveillance is finding an appropriate symptom screening approach. Although fever, cough and rhinorrhoea were reported more commonly in those who had influenza, more than 60% of patients with non-influenza ILI reported such symptoms, suggesting that they have poor discriminatory power as a screening test for influenza. Other studies in SE Asia have suggested that a revision of ILI definitions developed in temperate regions may be necessary for their use in tropical settings with cough an important feature in discriminating from other causes of febrile illness (Jiang et al., 2015). Our study was conducted in clinics which also participated in an existing GP ILI study which utilised mobile phone reporting of ILI to assess ILI dynamics. This study has demonstrated that ILI symptoms are much more consistent through the year compared to

temperate countries and that peaks in ILI activity do not always correlate with peaks in influenza activity (Lam et al, Manuscript under Review, (Oxford University Clinical Research Unit, 2015)). All of this suggests that the current ILI definition as a syndromic surveillance for influenza in tropical countries is unlikely to be as clear cut as it is in temperate countries. This study focused on respiratory symptoms which therefore means that it is difficult to comment whether inclusion of other symptoms would lead to better syndromic definitions for surveillance. Refinement of existing definitions may be achieved by combining the results of this study with those of other fever studies conducted in similar populations in conjunction with the Hospital for Tropical Disease

Intense periods of influenza A transmission was seen in the second quarter of both 2014 and 2015 with noticeable increases after the lunar new year public holidays. This coincides with the start of the wet season in southern Vietnam which has previously been linked to influenza seasonality in Vietnam (Thai et al., 2015). Internationally, there was a significant shift in influenza A H3N2 antigenically between 2014 and 2015. This manifested in a mismatch between H3N2 vaccine component and the circulating strain which resulted in reduced vaccine efficacy (<25%) and an increase in severe cases in winter 2014-2015 in northern temperate countries (Broberg et al., 2015, Molbak et al., 2015, Pebody et al., 2015). Both periods of H3N2 activity in this study occurred at the end of the northern temperate influenza season. At present we do not know which strain was circulating in Vietnam between 2014 and 2015 and if this could be the result of introduction of the novel strain from northern temperate regions at the end of their winter season. Genome sequencing of viruses collected during this study is planned to investigate this further.

There was no detectable difference between ages of individuals presenting with influenza or non-influenza ILI. However our study population had more individuals under the age of 35 than would be expected for the general Vietnamese population. Young children are thought to be responsible for much of the community transmission of influenza (Fox et al., 1982, Longini et al., 1982) and although this study did not recruit very young children, the under 35's are likely to be siblings or parents of this group. Having younger children in the household did not change the likelihood of having influenza compared to non-influenza ILI. This could mean that having younger children in the household increases your risk of getting any viral respiratory infection rather than increasing the risk of influenza only. Households also tended to be larger than the Vietnamese average with a multi-generational structure. This may change the age distribution of ILI compared to temperate countries where multi-generational homes are less common.

Although these age and household structures could impact on transmission dynamics it may also represent a difference in care seeking behaviour. Individuals who live in smaller households are

likely to be more affluent and may not access healthcare through the clinics included in this study, either because they self-care, have difficulties attending clinics because of work patterns or they attend more expensive polyclinics which suit their lifestyle. More individuals of working age (20-64) were recruited from private clinics than from the government run outpatient clinics and were more likely to be in a professional work rather than manual labour. Recruitment at the hospital clinics was limited to between 8am and 4pm whereas private clinics were able to recruit patients in the evening. All clinics had limited recruitment over the weekend period. These restrictions are likely to have impacted on the patient group which was recruited. The considerable heterogeneity in subject demographics are important to consider when interpreting healthcare surveillance figures. In systems where patients have a relative choice regarding healthcare providers (rather than single providers like the NHS in the UK) it is important to consider alternative sources of surveillance to give a better overall picture of attack rates (Ortiz et al., 2009, Ong et al., 2010). The expanding middle class of Vietnam means that this issue is likely to become more important and that surveillance of private facilities should be incorporated into future public health surveillance planning

Another impact of the changing economic environment in Vietnam is the use of influenza vaccination. There is no public provision for influenza vaccination in Vietnam and prior to 2009 the use in the general public was very rare with less than 1% of the population having ever received it (Palache et al., 2014). In this study, 5% of individuals who presented to private clinics reported receiving vaccination compared to only 1.2% those who attended the hospital outpatient clinic. However, only a single individual who reported having received vaccination had a recommended indication (this was recruitment from a private clinic). As influenza vaccine is only available from private clinics its current use is limited by an individual's ability to pay. Although preventative care through primary care is the strategic goal of the Vietnamese MoH (World Health Organisation & MOH Vietnam, 2012), it remains unclear what the optimal vaccination strategy is for Vietnam (Lambach et al., 2015). Preliminary work looking the matches between circulating strains and vaccine composition by the Vietnamese MoH have looked at northern hemisphere recommendations (Vuong et al., 2012c). Where study participants reported the month of vaccination it varied between June and November which is the period covered by annual southern hemisphere WHO guidance. This also corresponds to the wet season in HCMC where local belief is that there are higher rates of influenza. Clarification on what the optimal strategy for vaccination is vital for private providers of vaccination in order to give people the best protection for upcoming influenza seasons. It is of note that having received influenza vaccination at any time did not have any protective effect against influenza in this study.

The rates of chronic respiratory disease in this study was consistent with national estimates (Global Burden of Disease Study 2013 Collaborators, 2015). Smoking is a leading cause of morbidity and mortality in Vietnam. The rates reported in this study are far below national estimates (47.4% of men, 1.4% of women) (World Health Organisation, 2010). It is unusual that none of the smokers reported chronic respiratory disease. The main source of recruitment for this study was Hospital for Tropical Disease and private clinics. Within Ho Chi Minh City, a separate hospital exists for chronic lung disease and TB. Patients with these conditions may already attend this hospital for follow-up and so may attend if they develop acute symptoms. Our study also concentrated on non-severe influenza, by not including hospitalised patients we may have excluded more of those with chronic health problems. This is an important distinction when examining burden of disease and planning interventions to reduce severe morbidity and mortality. As the main focus of our study was to investigate transmission dynamics, it was reasonable to focus on non-severe cases as they are responsible for the majority of transmission.

Antibiotic use at enrolment was high in all groups although higher in influenza B. All individuals presented early in their illness and had accessed antibiotics prior to this attendance at clinic. Antibiotics can easily be purchased in pharmacies across Vietnam (Nga et al., 2014). Most ILI will have a viral aetiology and given the timings of the study, it is unlikely that individuals have developed a secondary bacterial infection requiring antibiotics. Therefore this represents a significant problem of unnecessary antibiotic prescribing. Antibiotic resistance in Vietnam is an issue, as it is globally (Kim et al., 2012). Education of medical professionals, pharmacists and the public about the appropriate use of antibiotics is urgently required (Wertheim et al., 2013).

As would be expected, lower titre response at baseline is associated with a higher risk of PCR confirmed influenza of the same subtype for all human strains (homosubtypic response, recent and historic strains). The homosubtypic response to the most recently circulating strains showed differences between H3N2 and H1N1. For individuals infected with H1N1, most had a low or undetectable microarray titre at the time of infection. Although the mean titre was significantly lower than those not infected with H1N1, across all study subjects a considerable proportion had PMA titres less than 80 which were levels observed prior to the emergence of 'swine flu' and the global pandemic of H1N1 in 2009 (de Bruin et al., 2014, te Beest et al., 2014a). This group could represent individuals who have not been infected with H1N1 in the post pandemic period or individuals who had been infected previously but their antibody response has waned rapidly to pre-pandemic levels. The relative lack of antigenic change in the circulating H1N1 since 2009 suggests that there are still enough susceptible individuals to infect without selecting for antigenic shift. As waned immunity would still be expected to give a brisk immune response when re-challenged with a

very similar virus it seems more likely that this is explained by individuals having their primary infection with H1N1 2009 up to 5 years after it emerged.

In contrast, most individuals infected with H3N2 had detectable titre to H3N2 2011, likely due to cross reaction to previously encountered H3N2 strains. Younger individuals had the highest response to H3N2 2011 although this did not change their probability of being infected. This suggests that their higher response is due to H3N2 2011 being antigenically similar to their primary H3N2 infection and higher rates of cross reaction are seen. The same picture is seen across historic strains with the highest response being to strains seen early in life. The peak in the first decade of life is consistent with other studies (Lessler et al., 2012, Miller et al., 2013) but it should be noted that the age criteria to enter the study and the time span of the strains on our assay may have contributed to this finding.

Although all H3N2 strains can predict odds of being infected when a single strain is included in model, when multivariate analysis is performed with a combination of strains, more recent strains seem to be having the greatest contribution to current susceptibility. Strain response to H3N2 2005 and 2009 consistently come out as the most significant predictors of current susceptibility. All H3N2 strains included on the PMA belong to different antigenic clusters (Barr et al., 2014, Bedford et al., 2014) and the globally predominant circulating strain was antigenically similar to H3N2 A/Victoria/361/2011 until emergence of an antigenically distinct strain in 2014-2015 (Broberg et al., 2015, World Health Organisation, 2015c). H3N2 2005 and 2009 therefore represented the two previous antigenic clusters at the beginning of this study but this antigenic positioning could have changed through the recruitment period. That the highest risk of infection is in individuals with high response to H3N2 2005 and low response to H3N2 2009 could be interpreted in terms of time since last infection and that this is more important in susceptibility than the absolute titre (although the two are inextricably linked). If your last infection was more than one antigenic cluster away from the current circulating strain you are at highest risk of infection. As studies have suggested that adult individuals are infected once or twice in a decade this would fit with this interpretation (Kucharski et al., 2015). This also suggests that in terms of the microarray titre, high titres due to infections contracted early in life are less predictive of susceptibility now within a continually circulating subtype. These assumptions are not likely to hold for the emergence of a new pandemic strain such as H1N1 2009 where the historical strain response to H1N1 1918 was shown to be beneficial in the first waves of the pandemic (Van Kerkhove et al., 2013b).

How best to include changing exposure in models of susceptibility is a problem when considering epidemic disease rather than endemic particularly where seasonality is not predicable. The use of

week of recruitment as a proxy demonstrated interesting findings in this study. It suggested that levels of protection offered by a fixed titre changed depending on force of infection. An obvious explanation for this could be an antigenic change in the circulating virus meaning that the protection offered by existing antibodies is no longer sufficient. In this situation it demonstrates the importance of considering changes throughout an influenza 'season' and that assessments of population susceptibility rely on having a good match between the circulating and testing strain. If this is not due to an antigenic change in the virus then it raises other interesting possibilities. At lower prevalence's influenza may preferentially infect those with a lower titre because it is easier with a lower infecting dose and the lowest barrier to infection. Where influenza is circulating at higher prevalence, susceptible individuals are more likely to come into contact with infected individuals as there are more of these in the population leading to multiple challenges with influenza viruses. This could lead to a situation where an individual with a higher titre is protected against a single challenge but subsequent challenges in a short period of time may overcome the existing protection. Animal models have demonstrated interference between subtypes if exposed sequentially over a short period but repeat challenge with the same subtype was not explored (Cao et al., 2015, Laurie et al., 2015b). Another possibility is that during peaks of activity more individuals have higher levels of viral shedding, which again could overcome the immune barrier to infection. Climatic factors during wet or cold seasons may lead to changes in human behaviour which result in more frequent contacts between individuals which could give an ideal situation to allow the virus to overcome high barriers to infection. Evaluation of vaccine programs including influenza have suggested that vaccine efficacy changes across different transmission intensities (Gomes et al., 2014). This would likely be the same phenomenon described in this study.

That individuals infected with influenza B had the highest baseline titres to both H1N1 and H3N2 titres poses interesting questions. Influenza B was detected throughout the study period but a peak of activity was seen to follow a large peak of influenza A H3N2. Interference between influenza A and B have been noted in natural infection states previously (Camacho et al., 2011) and more recently modelling and animal experiments have suggested that infection with influenza A may change the susceptibility to influenza B in the short term (Cao et al., 2015, Laurie et al., 2015b). If being infected with influenza A does increase the chance of being infected with influenza B this could be a contributing factor to the different influenza epidemiology in tropical regions. If influenza A is present throughout the year then this could increase the amount of influenza B activity and explain why influenza B is seen more frequently in tropical countries (Caini et al., 2015) and may have a higher mortality than in temperate countries (Feng et al., 2012). The observation from this study needs to be explored in more detail and in particular look at the likely time since last influenza

A infection for those infected with influenza B, using the techniques described in Chapter 5. Further prospective studies are required to explore the interference of multiple influenza types and subtypes in relation to influenza transmission dynamics.

The serological response to avian strains in this study was consistent with work suggesting that much of the detectable response is related to cross reaction to previous infection with human strains (Boni et al., 2013, Molesti et al., 2014, Oshansky et al., 2014, Todd et al., 2014, Freidl et al., 2016). This is particularly noticeable in the response to H7N7 2003 and H9N2 1999 where over 40% of subjects have a detectable PMA titre. A small number of individuals have a high titre response which is similar to levels seen in response to human strains an individual could have been exposed to. Whether these high responses can be attributed to cross reaction requires further investigation including work from the longitudinal study described in Chapters 4 and 5 and how responses to avian strains change following recent infection with human strains. Further work is planned to look at titre response in individuals who were known to be infected with H5N1 as trying to determine appropriate cut offs for infection with avian strains is crucial to understand the true global attack rates (Wang et al., 2012).

There are a number of limitations to this study. In an ideal world this study would have been designed as prospective study recruiting individuals prior to their influenza illness to get a true picture of their influenza serology prior to infection. Given the numbers needed to recruit to reach the sample size of 350 influenza A positive individuals this was not feasible. A pragmatic approach was used to limit recruited to the first three days of illness so as to minimise the humoral immunity response which occurred following infection (Sealy et al., 2003). As historic B cell response can activate prior to three days there may already be a rise in response to historic strains at baseline. Any response in this way is likely to make the baseline titre differences between influenza and non-influenza less detectable, as a significant difference was found despite this these results are still interpretable. By recruiting individuals who attended clinic we are still missing those with mild or sub-clinical disease who may have a different antibody profile at baseline. Household studies with prospective follow-up of secondary cases can help to identify these individuals and would be beneficial to understand population level transmission. At the other end of the clinical spectrum, individuals with severe disease have been excluded and understanding whether differences in response to historic and current strains at the time of infection and also following infection could explain differences in clinical presentation should be explored. This could also address the question of whether original antigenic sin is indeed a sin.

Further work is required to relate the results from the PMA with HI testing as well as functional neutralisation assays. The more sensitive PMA assay is likely to be subject to more issues relating to cross reaction and being able to understand these in relation to susceptibility and recent infection will be explored in more detail in Chapters 4 and 5. Many of the questions raised in this chapter require an understanding of which virus was circulating during the recruitment period and where this fits antigenically with the most recent strains on the assay. Vietnam's national surveillance programme monitors for predominant circulating strain and these results will be available in the near future. In addition to this there are plans to sequence a subset of the viruses from this study. This study continued to recruit until the end of November 2015. Much of the analysis has focused on H3N2 because this was the predominant subtype identified. The mixed H3N2 and H1N1 wave detected in May 2015 carried on for two months. These additional subjects infected with H1N1 will allow further exploration of the effect of historic strains on susceptibility to a recently emerged subtype.

4 INFLUENZA ANTIBODY DYNAMICS FOLLOWING ACUTE RESPIRATORY INFECTION

ABSTRACT

Background: Repeat infections with influenza A occur because of the continual evolution of the virus. A specific humoral response occurs after each infection and this develops into an immunological profile for each individual based on their prior exposure history and the cross-reaction between antigenically similar viruses. Understanding how this profile changes following acute infection is important for interpretation of seroepidemiological studies.

Methods: A prospective, observational study of patients with ILI in Ho Chi Minh City, Vietnam has been running since August 2013. Influenza A & B PCR and antibody testing to a panel of 11 human and 5 avian strains is performed using a novel protein microarray technique. A subset of subjects are followed up clinically and serologically for seven months.

Results: 186 ILI patients were recruited between August 2013 & May 2015. 102 and 14 subjects had influenza A H3N2 and H1N1 respectively. The largest response was within subtype which peaked at around 30 days. A boosting of historic response was also seen. The response in historic strains waned quicker than response to recent strains. A smaller but significant between subtype increase in titre was also detected. Following an acute rise after infection within six months most individuals had returned to a standard rate of decline of 1 log2 titre unit each one to two years.

Conclusion: Titre rise is seen within and across subtypes which would lead to repeat boosting of titre levels across many years. Further work should be performed to establish if this titre rise is accompanied by a change in neutralisation activity.

4.1 BACKGROUND

Repeat infections with influenza are possible due to the continual evolution of the influenza virus which means that specific immunity generated to one strain is no longer protective when the circulating virus is sufficiently different (Ferguson et al., 2003). Response to the strains an individual has been exposed to during their life remain detectable many years after this exposure with some of the highest responses to strains exposed early in life (Hennessy et al., 1955, Lessler et al., 2012). This develops into an immunological profile for each individual based on their prior exposure history and the cross-reaction between antigenically similar viruses (Smith et al., 2004). Following infection with

a strain there is the development of a specific humoral immune response which then wanes over time. It is assumed that once antibodies have waned sufficiently that an individual is again susceptible to further infection. How the waning of specific immunity and the simultaneous evolution of the virus are linked in terms of susceptibility is not well understood. With improvements in computational methods and more multi-strain studies being conducted it is hoped that these questions can be addressed.

Short term antibody dynamics to the infecting strains have been investigated for many decades. A rapid rise with a peak around one month has been demonstrated previously using HI (Horsfall and Rickard, 1941, Morris et al., 1966, Sonoguchi et al., 1986). Where paired serology is available for sero-diagnosis a minimum of one month is preferred (World Health Organisation, 2014a). The pattern and speed of antibody waning is less clearly defined and in particular the response in historic strains (Horsfall and Rickard, 1941, Foy et al., 1980, Lerman et al., 1980, Ochiai et al., 1986, Sonoguchi et al., 1986). Understanding this rate of waning is vital for design and interpretation of seroepidemiology studies (Vinh and Boni, 2015). Much of the recent work around antibody dynamics has been around the antibody dynamics following vaccination (Petrie et al., 2015) or the dynamics following first infection with a pandemic strain. These will not necessarily be the same as the antibody response in individuals who have never been vaccinated and therefore is an important distinction when modelling response in countries such as Vietnam with very low rates of vaccination. For drifted strains such as H3N2, understanding the historic antibody response following infection is crucial to interpreting historical and current attack rates (Fonville et al., 2014, Kucharski et al., 2015). Finally, it is important to recognise any differences in measured antibody dynamics in novel testing techniques such as the protein microarray utilised in this study from traditional HI methods (Huijskens et al., 2013). One of the major benefits of these techniques is to allow high throughput testing of large population representative serum banks. To be able to do this we need to understand differences between these and existing technologies to ensure inferences from these studies are reasonable. The work presented in this chapter is designed to address some of these queries.

4.2 METHODS

A full description of the study procedures and laboratory testing is given in Chapter 2.

4.2.1 Summary of Design and Conduct of the Observational Study

This observational study was conducted between August 2013 and May 2015 at outpatient clinics and community medical practitioners in Ho Chi Minh City (HCMC), Vietnam. Individuals were invited to join the study if they were between 10 and 70 years of age with symptoms for less than 72 hours

and fitted the ECDC ILI definition (European Centre for Disease Prevention and Control (ECDC), 2015). One anterior nasal swab and one throat swab and one 5ml serum sample was collected at baseline and stored at -20°C within 24 hours. A subset of patients were invited to join a longitudinal sub-study. All subjects testing positive for influenza A were eligible to join the sub-study. Negative control subjects were matched by age (+/- 5 years) and gender to influenza A positive subjects. Subjects attended for repeat serum sampling at 30, 90, 150 and 210 days (+/- 5 days) after recruitment. Active respiratory symptom surveillance was conducted during the follow-up period with telephone follow-up at 60, 120 and 180 days.

Both studies were approved by the Scientific and Ethical Committee of the Hospital for Tropical Disease, Ho Chi Minh City, Vietnam and Liverpool School of Tropical Medicine Research Ethics Committee, UK. Letters of agreement supporting the involvement of the community medical clinics were obtained from the Ho Chi Minh City Department of Health.

4.2.2   Sample Analysis

Respiratory samples were batched tested monthly for influenza A & B using standard polymerase chain reaction (PCR) techniques (World Health Organisation, 2011). Samples positive for influenza A were then tested for H3N2 and H1N1 subtypes. All subjects recruited to the longitudinal study had their baseline and follow-up sera tested against a panel of 16 influenza A strains (11 human, 5 avian strains) using a protein microarray technique previously described (Koopmans et al., 2011). This included both influenza A PCR positive individuals and age matched negative controls (i.e. non-influenza ILI. Full methods are described in Chapter 2.

4.2.3   Outcome Variables and Statistical Analyses

The primary outcome was antibody titre as measured by protein microarray at one, three, five and seven months after acute respiratory infection.

Continuous variables which were normally distributed were compared with t-test or ANOVA as appropriate. Tukey's HSD was used for posthoc testing following ANOVA where appropriate. Continuous variables which were non-normally distributed were compared using Wilcox and Kruskal Wallis rank sum depending on the number of groups (2 vs more than 2 respectively). Categorical variables were compared using the Fisher exact test, Mann-Whitney test or Chi Square test as appropriate.

Antibody responses were compared between predefined groups based on PCR determined infecting virus subtype giving three groups; influenza A H3N2, influenza A H1N1 and non-influenza ILI. Microarray strains were categorised into i) Most Recent Circulating Strains (H1N1 2009, H3N2 2011),

ii) Historical Human Strains (all other human strains), and iii) Avian Strains (Table 3.1). Titre responses will be categorised as homosubtypic where the infecting subtype and the microarray strain match (i.e. infected with H1N1, homosubtypic titre response is to H1N1 strains of the microarray). Heterosubtypic titres are where the infecting subtype and the microarray strain do not match (i.e. infected with H1N1, heterosubtypic titre response is to H3N2 strains on the microarray). Analysis was performed on log2 transformed titres unless explicitly stated.

Where appropriate, analysis was stratified by age at recruitment. Age was preferentially used as a continuous variable, otherwise age was categorised as per recommendations from the Consortium for the Standardisation of Influenza Seroepidemiology (5-9, 10-19, 20-44, 45-65, 65+) (Van Kerkhove et al., 2013a).

Linear regression and non-parametric general additive models were used to investigate the titre change between visit one and visit two, and to look for significant predictors of titre change. This was performed for i) homosubtypic, ii) heterosubtypic and iii) negative responses separately. For the homosubtypic and heterosubtypic responses, individuals infected with H3N2 and H1N1 were grouped together and their response to recent strains were looked at together. A strain specific term for infecting subtype was included. Univariate and multivariable analysis was performed looking at the impact of infecting influenza strains, occurrence of new ILI during follow-up period, gender, week of study recruitment, days of symptoms at recruitment, paracetamol or antibiotic use prior to study recruitment (antiviral use not included due to very low number of subjects using this at baseline), any prior respiratory disease, current pregnancy, current smoker and whether they have ever received influenza vaccine.

Change in titre across the follow-up period were investigated using two methods, piecewise linear regression and non-parametric general additive models. Both of these methods were used to investigate the decline in antibody following post-infection peak and allow for monophasic and biphasic decline in titre.

A two-step process was used to investigate antibody response using piecewise regression. First the optimal breakpoint (or points) was determined using the R package segmented (Muggeo, 2008) which identifies the most likely breakpoint using maximum likelihood methods. 95% credible intervals of estimates were identified using the boot R package and 1000 replicates (Davison, 1997, Canty and Ripley, 2015). The mean breakpoint(s) were then used to identify the range of variation in the slopes surrounding the breakpoint. In order to fit the model, a prior estimate of break point was required, i) 30 and ii) 30 and 90 were used. Piecewise linear regression was also performed with a

random effects term allowing a different intercept for each subject. The segmented package does not allow the incorporation of individual random effects so the linear mixed methods package mcgv was used and user defined code for optimisation using the foo function within R. 1000 bootstrapped replicates were run to calculate mean and credible interval. In the sampling procedure sampling with replacement was allowed by participant ID but no resampling was performed on serum sample time points. This is consistent with the assumptions of random intercept but not slope for individuals and was coded by hand. User defined R code for resampling, bootstrapping and MLE of piecewise regression are included in the supplementary appendix. As in the initial regressions, models were fit to determine optimal breakpoint(s) followed by investigation of slope variability for the mean breakpoint(s). Quality of model fit was assessed using adjusted R^2 and comparison of relative model fit was performed using AIC. For the random effects models marginal and conditional pseudo-R^2 were calculated as per the methods in (Nakagawa and Schielzeth, 2013) using the *r.squaredGLMM* function from the MuMIn package (Barton, 2015).

Non-parametric generalised additive models and splines were used to investigate titre response. These methods allow a flexible approach to the shape of response accounting for individual variation when random effects are used and parameters such as age included. In addition to strain specific models, a model of all titre response with strain specific response was fitted. To explore possible mechanisms behind the age differences in the combined strain model, titre response was categorised depending on whether this was i) a titre response to a recent strain (H3N2 2011), ii) a titre response to historic strains an individual could have been exposed to (i.e. subject was alive at the time of circulation) ii) a titre response to historic strains an individual could not have been exposed to (i.e. subject was not alive at the time of circulation).

All statistical analysis were performed using R Statistical Software v3.2.1 (R Core Team, 2015). Packages used for analysis were ggplot2, gamm4, mgcv, segmented, boot.

4.3   RESULTS

4.3.1   Description of follow-up study participants

The longitudinal study recruited 186 subjects between 8 August 2013 and 31 May 2015. The majority of patients were recruited from hospital outpatient settings (n=124/186, 66.7%) in line with the overall pattern of pattern of study recruitment (χ^2 = 0.34, p=0.55). Seven month follow-up was available for 82 subjects (Table 4.1). 42 subjects (22.5%) withdrew from the study during the follow-up period. Drop out rates were higher in influenza positive groups (H1N1 3/14, 21.4%; H3N2 27/102,

26.4%; Negative 11/55, 16.6%) but this difference was not statistically significant ($\chi^2 = 0.34$, p=0.52). The commonest time point to drop out of the study was between visit two and three (n=33, 78%).

	Visit 1 (Day 1-3)	Visit 2 (Day 30)	Visit 3 (Day 90)	Visit 4 (Day 150)	Visit 5 (Day 210)
Influenza A H1N1	14 (100)	14 (100)	5 (35.7)	5 (35.7)	5 (35.7)
Influenza A H3N2	102 (100)	102 (100)	62 (60.8)	51 (50)	45 (44.1)
Influenza A NT	4 (100)	4 (100)	3 (75)	3 (75)	3 (75)
Influenza Negative	66 (100)	66 (100)	51 (77.3)	41 (62.1)	29 (43.9)

Table 4.1 Number of Follow-Up Clinic Visits before 31 May 2015. Percentage of follow-up subjects by infecting subtype in brackets.

Age and sex distribution suggested that the matching process between influenza positive and negative was adequate (Table 4.2). Study week of recruitment varied between infecting subtypes as would be expected given the differences in influenza circulation during the study period (Figure 31). Samples without an influenza A subtype were excluded from the subsequent analysis unless explicitly stated.

	Influenza A H1N1 n (%)/ med (IQR)	Influenza A H3N2 n (%)/ med (IQR)	Influenza A NT n (%)/ med (IQR)	Influenza Negative n (%)/ med (IQR)
Number	43 (100)	356 (98.3)	17 (100)	253 (100)
Age	30.1 (25.9-46.4)	26.3 (18.1-33.1)	40.4 (22.9-43.3)	25.8 (20.1-33.6)
Female	22 (51.2)	148 (41.6)	12 (70.6)	120 (47.4)
Study Week	31 (30-84)	40 (35-43)	29 (21-35)	62 (46-67)

Table 4.2 Serology Samples by Influenza Subtype. Number of serum samples is presented as percentage of total number of follow-up samples.
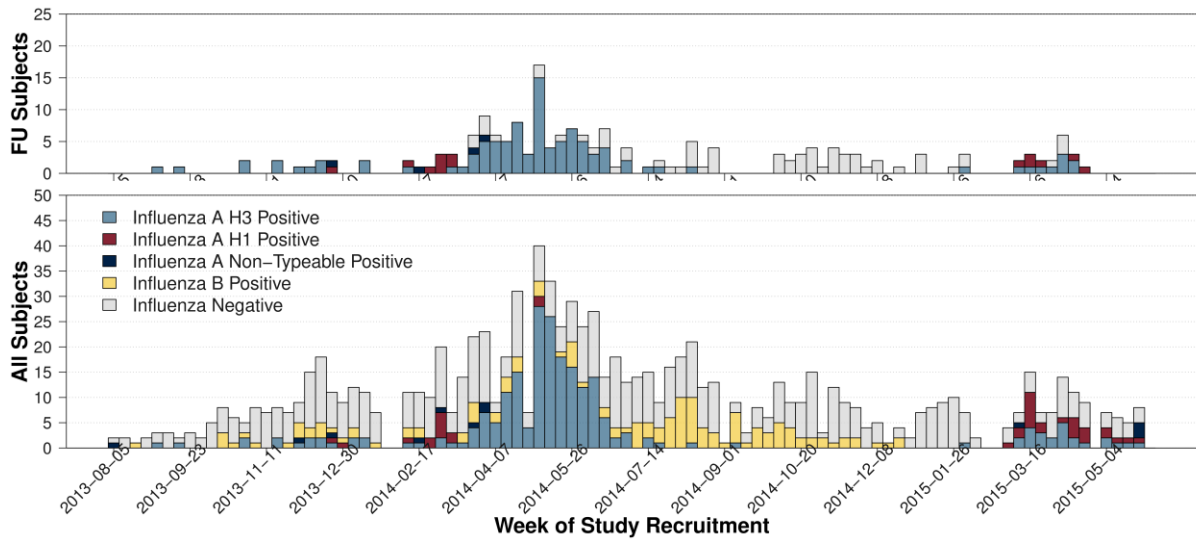
Figure 31 Study Timeline. Top panel is recruitment to follow-up study. Bottom panel main study recruitment.

Mean number of follow-up days at each visit was significantly different between infecting subtype as measured by two-way ANOVA ($F_{(12,655)}=2.428$, p 0.004). On post-hoc testing the only significant difference was between H1N1 positive and influenza negative groups at visit two.

	Visit 1 (Day 1-3)	Visit 2 (Day 30)	Visit 3 (Day 90)	Visit 4 (Day 150)	Visit 5 (Day 210)
Influenza A H1N1	2 (2-3)	33 (31-34)	95 (91-98)	157 (149-165)	218 (215-222)
Influenza A H3N2	2 (2-2)	35 (34-36)	95 (94-96)	158 (156-159)	217 (215-218)
Influenza Negative	2 (2-2)	37 (35-39)	95 (93-96)	155 (154-156)	216 (214-218)

Table 4.3 Follow-up Days at each Study Visit by Infecting Subtype. Mean and 95% CI presented.

4.3.2   PMA Titre Change between Visit One and Visit Two

4.3.2.1   *Change in Most Recent Circulating Strains*

The log titre change between serology at baseline (day 1-3 of illness) and approximately one month post infection showed significant differences depending on infecting subtype as measured by one way ANOVA (Table 4.4). Post-hoc testing shows significant differences in mean titre change between all infecting groups for both recent strains (Tukey HSD, all p values <0.05). The largest rise for each strain was where infecting and testing subtype match (e.g. testing strain H1N1 2009, infecting strain H1N1). However, significant positive responses are seen in the non-infecting subtype: these responses will be referred to as homosubtypic and heterosubtypic respectively. The mean response in subjects with non-influenza ILI was localised around zero. This is within the expected variation of the microarray assay. Outliers were seen in all infecting groups (Figure 32).

	Influenza A H1N1	Influenza A H3N2	Influenza Negative	ANOVA
H1N1 2009 Mean (95%CI)	3.67 (2.55-4.79)	0.82 (0.61-1.03)	-0.1 (-0.41-0.21)	$F_{(2,177)}=56.74$ p <0.001
H3N2 2011 Mean (95%CI)	0.86 (0.41-1.31)	2.97 (2.74-3.2)	0.01 (-0.27-0.3)	$F_{(2,177)}=140.1$ p <0.001

Table 4.4 Mean Log2 Titre Change between Visit 1 and 2 by Infecting Subtype.
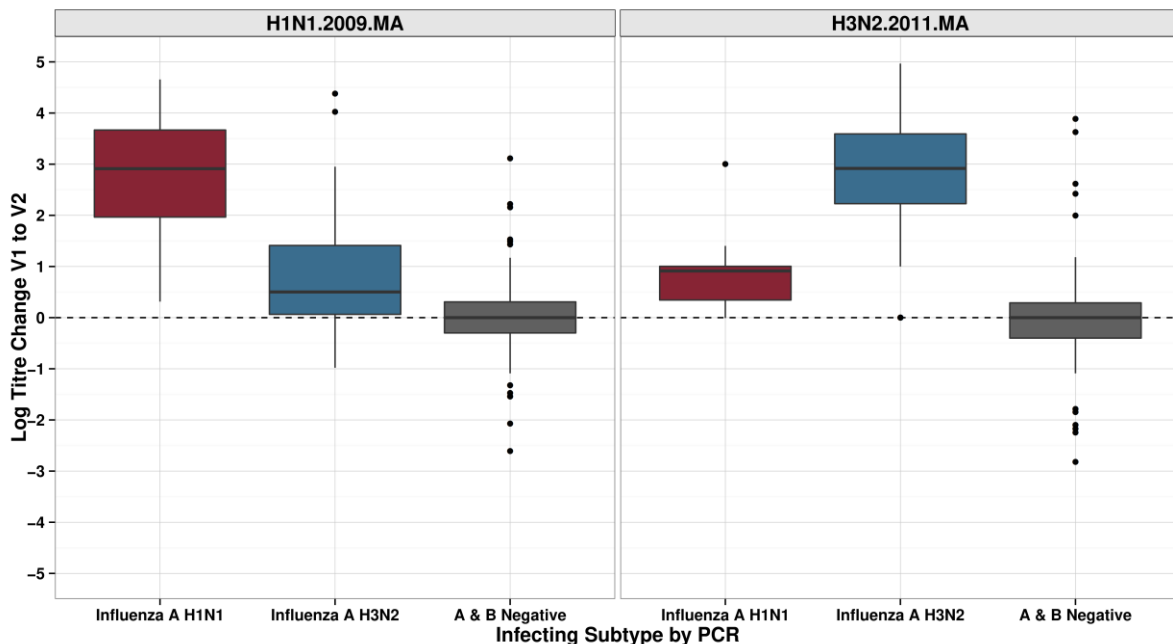
Figure 32 Log2 Titre Change between baseline and 30 days for Recent Strains by Infecting Subtype.

In univariate analysis of predictors of homosubtypic titre response between visit one and two, only titre at baseline of homosubtypic strain was a significant predictor of titre change (Table 4.5). In multivariable analysis, titre at baseline remained significant as was whether an individual was

infected with H3N2 or H1N1. In univariate analysis having H3N2 resulted in a lower titre change than being infected with H1N1 (non-significant). However, in multivariable analysis this became a significantly higher titre change than individuals infected with H1N1. The maximum titre measurable on the microarray is 1280 with a fixed value of 1810 for values which are higher than this dilution. This means that for individuals with a higher baseline titre there is a smaller possible increase than individuals who have a lower baseline titre. 85% (n=85/100) of individuals infected with H3N2 have a homosubtypic titre greater than 1280 at one month compared to only 21.4% (n=3/15) of individuals infected with H1N1.

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R^2	Estimate (95% CI)	p value	R^2
H3N2 vs H1N1	-0.701 (-1.42-0.018)	0.056	0.024	*1.325 (0.786-1.864)*	<0.001	0.684
Titre at Baseline	*-0.615 (-0.71--0.52)*	*<0.001*	*0.590*	-0.769 (-0.875--0.663)	<0.001	
Study Week	-0.011 (-0.024-0.001)	0.070	0.020	-0.002 (-0.011-0.006)	0.571	
Days since symptom onset	0.046 (-0.011-0.103)	0.111	0.014	0.025 (-0.009-0.058)	0.146	
Age	0.007 (-0.014-0.027)	0.505	-0.005	-0.011 (-0.024-0.001)	0.077	
Gender	-0.286 (-0.766-0.193)	0.239	0.004	-0.126 (-0.419-0.167)	0.395	
Influenza Vaccination	-1.383 (-3.192-0.425)	0.133	0.011	-0.284 (-1.336-0.767)	0.593	
Current Smoker	-0.217 (-1.029-0.594)	0.597	-0.006	0.043 (-0.455-0.542)	0.863	
Paracetamol at Baseline	0.039 (-0.579-0.658)	0.900	-0.009	-0.01 (-0.387-0.366)	0.957	
Antibacterial at Baseline	-0.338 (-0.814-0.138)	0.162	0.009	-0.144 (-0.454-0.167)	0.361	
Any Respiratory Disease	0.885 (-0.408-2.178)	0.178	0.007	0.311 (-0.483-1.105)	0.438	
Current Pregnancy	-0.513 (-2.459-1.432)	0.598	-0.015	-	-	

Table 4.5 Homosubtypic Titre Change between Visit 1 and Visit 2. Subjects infected with H1N1 and H3N2 included in linear regressions.

To investigate non-linear relationships, baseline titre, week of study recruitment, days since symptom onset and age were fitted as non-parametric splines along with an indicator for infecting subtype using GAMM package. The overall model fit was better than the multivariable linear model with an adjusted R^2 of 0.738 and AIC 240 (vs R^2 0.684 and AIC 263). Only titre at baseline had a significant smoothed spline term (p <0.001), age at baseline and infecting subtypes were both

significant linear terms (p <0.001 & 0.023 respectively). Due to a small number of H1N1 infections, fitting a separate spline for H1N1 and H3N2 infections results in an over fitted spline for H1N1 despite significance for this fitted term.

The GAM model demonstrates a plateauing of predicted visit two titres where baseline titre is greater than 80 (Figure 33 – left panel). When the predicted titre change is plotted against the maximum possible change between visit one and two (Figure 33 – right panel) it is likely that the presence of right censoring because of upper limit of detection on the microarray is responsible for this appearance.
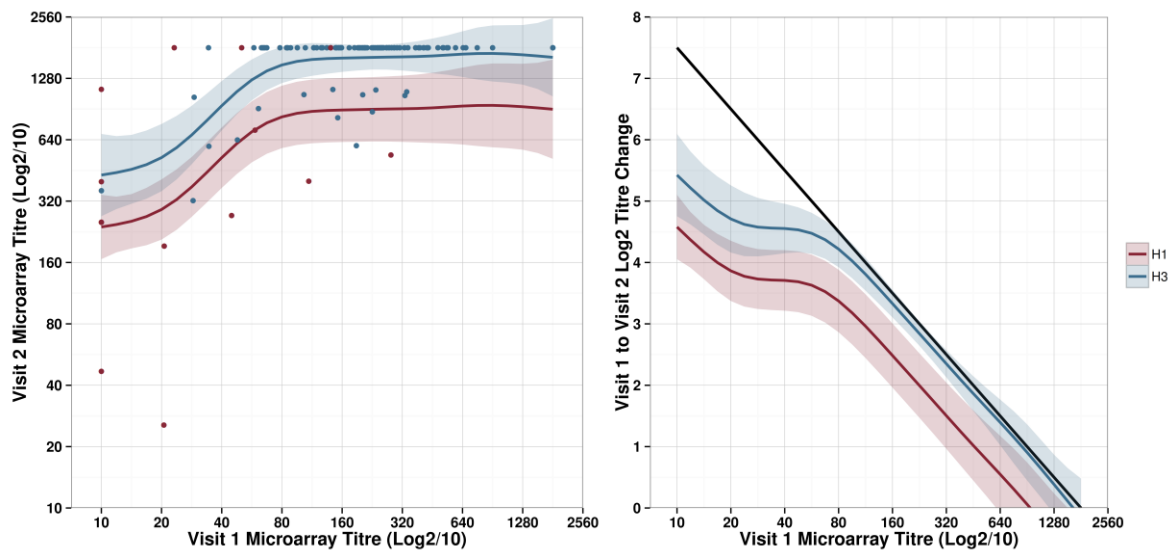
Figure 33 Predicted Visit 2 Homosubtypic Titre by Baseline Titre and Infecting Subtype. Left Panel Shows predicted titres with 95% CI of prediction. Right Panel shows titre change, black line is the maximum possible titre change for baseline titre. Points represent titre measurements from study, colour reflects infecting subtype.

In univariate analysis of heterosubtypic response, titre at baseline, days since symptom onset and having received antibiotics at baseline were predictive of a response (Table 4.6). In multivariable analysis, titre at baseline and antibiotic use at baseline remain significant. As in homosubtypic response, a GAM was fitted using non-parametric splines; titre at baseline was the only significant term. However, this was optimally fit with a linear term rather than a smoothed spline when given the option of fitting either model. In contrast to the homosubtypic response, four individuals had titre values at the limit of detection for the assay so the titre behaviour seen in Figure 33 is not replicated in Figure 34. A higher predicted response in heterosubtypic strain is seen where baseline heterosubtypic titre is low (Figure 34).

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R^2	Estimate (95% CI)	p value	R^2
H3N2 vs H1N1	-0.042 (-0.616-0.532)	0.886	-0.009	-0.421 (-0.988-0.147)	0.144	0.279
Titre at Baseline	*-0.282 (-0.37--0.193)*	*<0.001*	*0.257*	*-0.255 (-0.356--0.154)*	*<0.001*	
Study Week	0.001 (-0.009-0.01)	0.918	-0.009	0.002 (-0.008-0.011)	0.755	
Days since symptom onset	*0.067 (0.024-0.11)*	*0.003*	*0.070*	0.036 (-0.005-0.078)	0.084	
Age	0.015 (-0.001-0.031)	0.065	0.022	0.001 (-0.014-0.016)	0.866	
Gender	0.188 (-0.189-0.565)	0.326	0.000	0.197 (-0.152-0.545)	0.265	
Influenza Vaccination	-0.552 (-1.984-0.879)	0.446	-0.004	-0.291 (-1.539-0.958)	0.645	
Current Smoker	0.011 (-0.627-0.65)	0.972	-0.009	-0.165 (-0.756-0.426)	0.580	
Paracetamol at Baseline	0.176 (-0.309-0.661)	0.474	-0.004	0.361 (-0.085-0.807)	0.111	
Antibacterial at Baseline	*-0.443 (-0.81--0.075)*	*0.019*	*0.040*	*-0.398 (-0.772--0.025)*	*0.037*	
Any Respiratory Disease	0.147 (-0.877-1.171)	0.777	-0.008	-0.143 (-1.088-0.801)	0.764	
Current Pregnancy	0.5 (-0.746-1.745)	0.424	-0.007	-	-	

Table 4.6 Heterosubtypic Titre Change between Visit 1 and Visit 2. Subjects infected with H1N1 and H3N2 included in linear regressions.
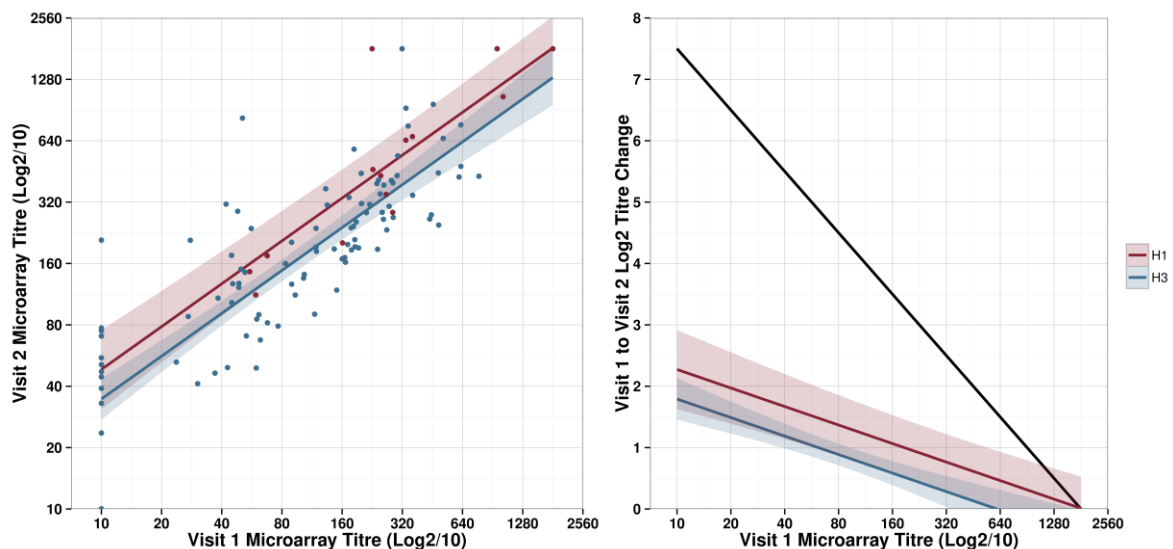
Figure 34 Predicted Visit 2 Heterosubtypic Titre by Baseline Titre and Infecting Subtype. Left Panel Shows predicted titres with 95% CI of prediction. Right Panel shows titre change, black line is the maximum possible titre change for baseline titre. Points represent titre measurements from study, colour reflects infecting subtype.

4.3.2.2 Change in Historic Human Strains

As with the response to more recent strains, the log titre change between baseline and one month post-infection of historic strains between baseline and one month post infection is significantly different between infecting subtypes (one way ANOVA, Table 4.7). Again the largest change was in the heterosubtypic response but a consistent positive rise was seen in homosubtypic historic strains (Figure 35). In post-hoc testing (Tukey HSD) the homosubtypic response was always statistically different from influenza negative individuals. The heterosubtypic response was not statistically significantly different from negative individuals for the more recent H1N1 strains (1999) or for most of the H3N2 strains (1968, 2005, and 2009). There are a relatively small number of individuals infected with H1N1 which may account for the lack of statistical significance although it is worth noting that the trend of the change is consistently positive. There is no significant difference in titre change of H3N2 2003 between individuals infected with H3N2 or H1N1. This is likely to be because H3N2 2003 had the highest baseline titre of all strains (see Chapter 3) and so has the lowest possible rise as described above.

	Influenza A H1N1	Influenza A H3N2	Influenza Negative	ANOVA
H1N1 1918 Mean (95%CI)	2.97 (1.96-3.99)	1.16 (0.94-1.37)	-0.06 (-0.3-0.17)	$F_{(2,177)}=52.56$ $p <0.001$
H1N1 1977 Mean (95%CI)	1.93 (1.31-2.55)	0.89 (0.69-1.08)	0.05 (-0.13-0.23)	$F_{(2,177)}=32.22$ $p <0.001$
H1N1 1999 Mean (95%CI)	1.31 (0.73-1.9)	0.34 (0.21-0.48)	0.1 (-0.15-0.35)	$F_{(2,177)}=12.16$ $p <0.001$
H1N1 2007 Mean (95%CI)	1.73 (1.15-2.31)	0.46 (0.27-0.65)	0.02 (-0.28-0.32)	$F_{(2,177)}=15.1$ $p <0.001$
H3N2 1968 Mean (95%CI)	0.75 (0.41-1.09)	2.84 (2.54-3.14)	0.02 (-0.28-0.33)	$F_{(2,177)}=89.52$ $p <0.001$
H3N2 2003 Mean (95%CI)	1.02 (0.4-1.63)	1.69 (1.42-1.96)	0.13 (-0.05-0.31)	$F_{(2,177)}=36.4$ $p <0.001$
H3N2 2005 Mean (95%CI)	0.77 (0.35-1.19)	2.54 (2.3-2.78)	0.1 (-0.16-0.36)	$F_{(2,177)}=98.3$ $p <0.001$
H3N2 2007 Mean (95%CI)	0.8 (0.41-1.19)	3.17 (2.96-3.38)	0.01 (-0.26-0.28)	$F_{(2,177)}=181.4$ $p <0.001$
H3N2 2009 Mean (95%CI)	0.89 (0.29-1.5)	3.69 (3.45-3.93)	0.04 (-0.33-0.4)	$F_{(2,177)}=159.1$ $p <0.001$

Table 4.7 Mean Log2 Titre Change of Historic Strains between Visit 1 and 2 by Infecting Subtype
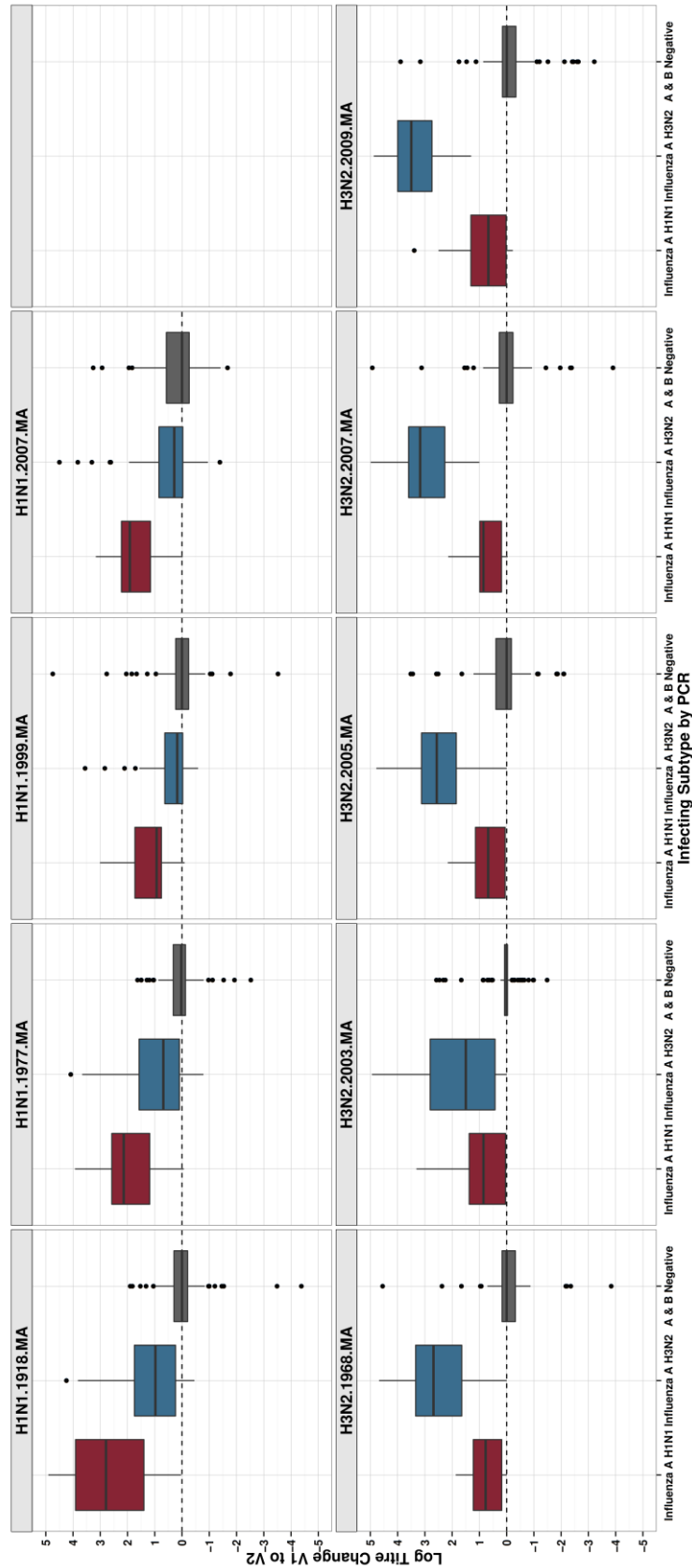
Figure 35 Log2 Titre Change between baseline and 30 days for Historic Strains by Infecting Subtype

For historic strains regression models were fit to look for predictors of titre change. For each strain the homosubtypic and heterosubtypic response were investigated separately. Because of the nature

of the historic strain data, it was necessary to consider individuals infected with H3N2 and H1N1 separately for this particular analysis, unlike with the response to current strains, ice homosubtypic response to H1N1 1918 uses data only from subjects infected with H1N1, heterosubtypic response to H1N1 1918 uses data only from subjects infected with H3N2. Results tables of univariate and multivariable analyses are in the supplementary appendix.

In univariate analysis of homosubtypic response, titre at baseline was a significant predictor of titre change for all historic human H3N2 strains (p values <0.001) but not H1N1 strains (p > 0.05). Age was a significant predictor for titre change for H3N2 1968, 2003 and 2005 although with differing effects. Higher age at baseline was associated with a lower titre change in 1968 but higher changes in 2003 and 2005 although all these effects were small with a 0.2 log2 titre change for each 10 years of age. Days since symptom onset was a significant predictor for H3N2 2005 but the p value was 0.048 and this is likely to be a chance finding. No other variables were significant in univariate analysis.

In multivariable analysis, titre at baseline remained a significant predictor of titre change for all historic H3N2 strains (p values <0.001). Age was a significant predictor of more recent H3N2 strains (H3N2 2009, p value 0.02) even when baseline titre was accounted for with older age groups having a lower response but again with a small effect of less than 0.2 log2 titre change for each 10 years of age. No significant terms were found in multivariable model of H1N1 homosubtypic response although there are only 14 individuals in this group.

As with the response to the most recent strains, higher baseline titre in historic human strains results in a lower log titre change. The number of right censored samples on visit two samples are likely to contribute to this (Table 4.8).

	Influenza A H1N1 (n=14)	Influenza A H3N2 (n=102)	Influenza Negative (n=66)
H1N1 1918 n (%)	1 (7.1)	3 (3)	1 (1.5)
H1N1 1977 n (%)	0 (0)	2 (2)	0 (0)
H1N1 1999 n (%)	1 (7.1)	2 (2)	0 (0)
H1N1 2007 n (%)	4 (28.6)	3 (3)	4 (6.1)
H1N1 2009 n (%)	3 (21.4)	1 (1)	0 (0)
H3N2 1968 n (%)	1 (7.1)	32 (32)	5 (7.6)
H3N2 2003 n (%)	8 (57.1)	89 (89)	31 (47)
H3N2 2005 n (%)	2 (14.3)	87 (87)	18 (27.3)
H3N2 2007 n (%)	0 (0)	63 (63)	9 (13.6)
H3N2 2009 n (%)	2 (14.3)	75 (76.5)	14 (21.2)
H3N2 2011 n (%)	3 (21.4)	85 (85)	17 (25.8)

Table 4.8 Number of Samples (%) with right censored values at visit 2 (>1280).

In univariate analysis of heterosubtypic response in historic human strains, titre at baseline was the only significant predictor of log titre change between visit one and two (H3N2 strains p values <0.05, H1N1 strains p values <0.001). This was true of all strains except heterotypic response to H3N2 1968 (i.e. response in individuals infected with H1N1). In multivariable regression, baseline titre remains the only significant predictor of heterosubtypic titre change in individuals infected with H3N2 (response in H1N1 strains with exception of H1N1 1977 p value <0.01). No significant terms were found in the model of heterosubtypic response for individuals infected with H1N1. As in the homosubtypic response, the small numbers in this group make it difficult to interpret this response.

4.3.3   Homosubtypic Antibody Dynamics

Following acute influenza infection there is a rise in the homosubtypic and heterosubtypic strain responses as detailed in section 4.3.2. In looking at the short term dynamics after infection, the raw titres rapidly rise followed by a decline (Figure 36). It is not clear if this decline is monophasic with a fixed rate of decline from the peak response or if there is a more complex decline. We employ several methods to investigate this: in the first instance piecewise linear regression will be used to fit multiple linear regressions through the data accounting for rise and fall of antibody followed by generalised additive models using flexible non-parametric spline fits. Because of the small numbers of H1N1 infections with longitudinal data to seven months, all analyses in this section will only include subjects infected with H3N2 and negative individuals.
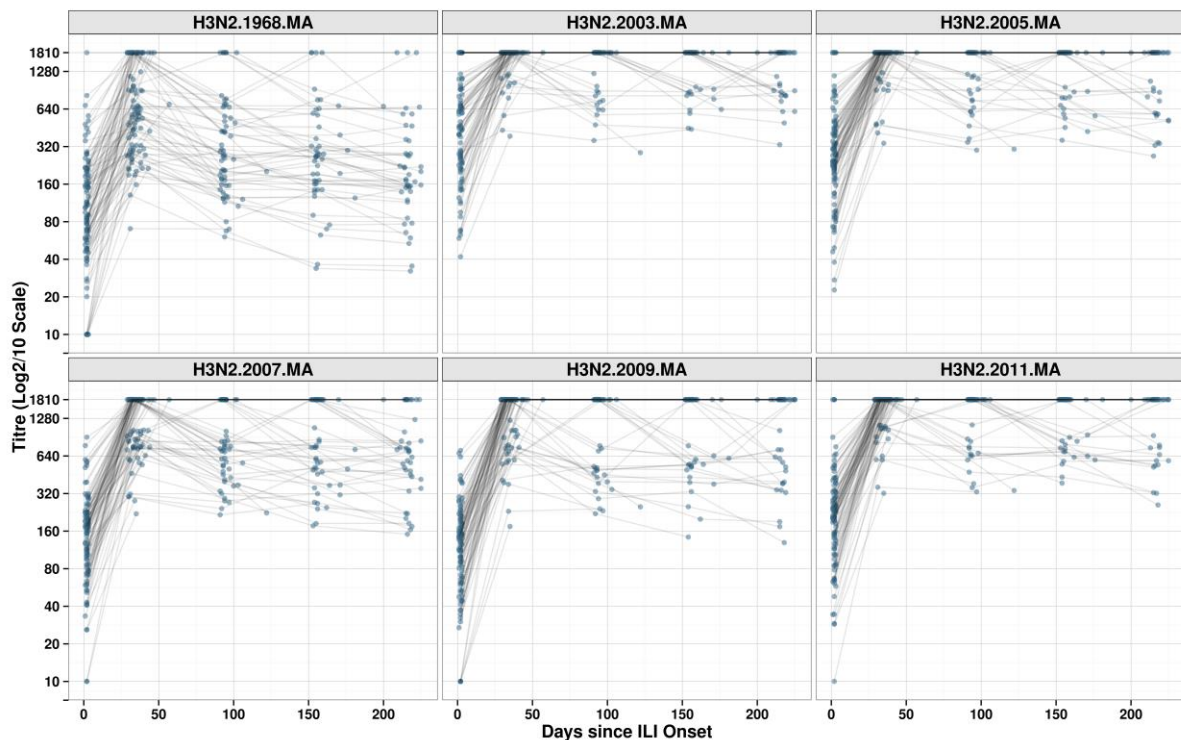
Figure 36 H3N2 Homosubtypic Strain Response. Individual trajectories in grey.

4.3.3.1 Piecewise Linear Regression

A two-step process was used to investigate antibody response using piecewise regression. First the optimal breakpoint (or points) was determined using bootstrapping methods and the segmented package. The mean breakpoint(s) were then used to identify the range of variation in the slopes surrounding the breakpoint. In order to fit the model, a prior estimate of break point was required, i) 30 and ii) 30 and 90 were used, which were estimated from visual inspection of Figure 36.

Mean estimates of the first breakpoint (which represents the peak of titre response post infection) were localised around 30 days for all homosubtypic strains with the exception of H3N2 2003 for both models (one and two breakpoints, Table 4.9) . The upper limit of credible interval was always less than 40 days where a single breakpoint was estimated and only higher for H3N2 1968 and 2005 where two breakpoints were estimated. The optimal peak for H3N2 2003 was earlier and it is again possible that this is due to the higher baseline titres to this strain and that this represents the time to reach the maximum possible titre measurable on the assay rather than a time to reach peak response. This could also have affected strains which have a large number of censored titres in follow-up samples and may explain the wider credible intervals for more recent strains (Table 4.10).

	Single Breakpoint				Two Breakpoints				
	Break (95%CI)	AIC	a R^2	Conv. (n)	Break 1 (95%CI)	Break 2 (95%CI)	AIC	a R^2	Conv. (n)
H3N2 1968	32.16 (10.8-35.5)	1235.91	0.37	1000	36.01 (28.8-86.51)	96.21 (32.2-216.1)	1229.77	0.38	603
H3N2 2003	10.41 (4.69-35.86)	948.27	0.38	1000	8.47 (2.63-37.7)	33.63 (30.5-51.21)	950.12	0.38	734
H3N2 2005	32.09 (9.29-36.31)	946.48	0.59	1000	32.49 (8.15-56.32)	83.17 (30.7-214.5)	947.91	0.59	669
H3N2 2007	32.37 (31.5-36.47)	1036.85	0.61	1000	32.32 (5.96-35.15)	103.92 (30.9-218.3)	1035.75	0.61	752
H3N2 2009	29.14 (20.4-32.71)	1054.53	0.68	1000	29.04 (8.15-33)	122.06 (30.4-218.9)	1055.13	0.68	764
H3N2 2011	31.6 (15.73-34.4)	933.95	0.67	1000	31.64 (6.09-36.82)	127.2 (31.6-220.4)	935.29	0.67	703

Table 4.9 Homotypic Response to H3N2 Infection. Piecewise linear regression. Bootstrapped 95% Credible Interval from 1000 replicates. AIC and adjusted R^2 for the model using best predicted break point. Comparison of AIC within strain only, i.e. across rows only. Conv. is the number of models which successfully converged during Maximum Likelihood Estimation using the segmented package.

There was more variability in the estimation of the second breakpoint although mean estimates and credible intervals may suggest a pattern of second breakpoint between 90 and 120 days after infection (Table 4.9). It should be noted that having two breakpoints did not significantly improve model fit or percentage of deviance explained compared to a single breakpoint as measured by AIC and adjusted R^2 for the optimal breakpoint (within strain comparison only). The model with two

breakpoints was also less likely to converge during maximum likelihood optimization (60-76% of bootstrapped datasets converged). This demonstrates that for the available data it is difficult to make interpretations about a second breakpoint. This could be because it exists but we have too few data points (either too few subjects or too few sampled time points) or because there is not an optimal second time point.

	Visit 1 n (%)	Visit 2 n (%)	Visit 3 n (%)	Visit 4 n (%)	Visit 5 n (%)
H3N2 1968	1 (1)	32 (32)	7 (11.5)	4 (8.2)	3 (6.8)
H3N2 2003	25 (24.5)	89 (89)	47 (77)	36 (73.5)	29 (65.9)
H3N2 2005	6 (5.9)	87 (87)	43 (70.5)	35 (71.4)	28 (63.6)
H3N2 2007	0 (0)	63 (63)	25 (41)	20 (40.8)	13 (29.5)
H3N2 2009	0 (0)	75 (76.5)	41 (67.2)	30 (61.2)	27 (61.4)
H3N2 2011	4 (3.9)	85 (85)	46 (75.4)	33 (67.3)	31 (70.5)

Table 4.10 Number of samples (%) with right censored values (>1280) by visit number. H3N2 PCR positive subjects only.

The next step was to fit slopes using the optimal breakpoints. This was again performed with 1000 bootstrapped datasets for both models of one or two break points. As would be expected for strains with similar breakpoints, the rate of titre increase was similar for all strains except H3N2 2003 where slopes for a single breakpoint were estimated (Table 4.11). The rate of decline was less consistent between strains with the mean rate of titre decline varying between 102 and 589 days/one log unit decline. This again is likely to be affected by the number of censored values and it is worth noting that strains with fewer censored values (H3N2 1968 and 2007) have faster rates of decline and narrower credible intervals.

	Break Point	Slope 1: Days to Rise 1 Log Unit	Slope 2: Days to Fall 1 Log Unit	Fit (n)
H3N2 1968	33	11.4 (10.1-13.1)	102.4 (81.6-131.3)	1000
H3N2 2003	8	4.6 (3.9-5.5)	589 (369-1561)	1000
H3N2 2005	32	11.7 (10.4-13)	399.2 (252.5-836.5)	1000
H3N2 2007	32	9.5 (8.8-10.4)	201.1 (148.6-318.8)	1000
H3N2 2009	29	7.4 (6.9-8.1)	359.2 (215.3-1309.5)	1000
H3N2 2011	32	10.1 (9.2-11.1)	462.3 (277.3-1066.3)	1000

Table 4.11 Homotypic Response to H3N2 Infection. Using single fixed break point slope estimates from piecewise linear regression. 1000 bootstrapped replicates performed, mean and 95% credible interval of slope presented.

Where two breakpoints were fitted, the slope estimates for a two phase decline were less consistent (Table 4.12). The rates of initial titre rise were consistent with estimates for a single breakpoint. The credible intervals included rates of decline which extended beyond 10 years to drop one unit. Again it is likely that titre censoring is playing a part here as the most interpretable strain is H3N2 1968 with a fast initial decline followed by a slower rate of decline which is in keeping with prior work. The relatively short follow-up period would also make it difficult to measure a biphasic decline using this method.

	Break Point 1	Break Point 2	Slope 1: Days to Rise 1 Log Unit	Slope 2: Days to Fall 1 Log Unit	Slope 3: Days to Fall 1 Log Unit	Fit (n)
H3N2 1968	36	96	11.2 (9.8-12.8)	44.7 (33.8-66.1)	222.2 (99.3-1731.5)	1000
H3N2 2003	8	34	4.3 (6.6-104)	81.8 (13159.4-26.3)	557.5 (330.1-1293.8)	1000
H3N2 2005	32	83	11.5 (10.4-12.9)	216.3 (104.9-1742.1)	596 (1485->3000)	1000
H3N2 2007	32	104	9.3 (8.5-10.1)	110.2 (73.6-217.9)	468.8 (629.8->3000)	1000
H3N2 2009	29	122	7.3 (6.7-8)	188.7 (102.3-719.6)	109746.5 (>3000->3000)	1000
H3N2 2011	32	127	9.9 (9.1-10.9)	275.9 (159.4-1087.3)	3000.9 (1408.6->3000)	1000

Table 4.12 Homotypic Response to H3N2 Infection. Using two fixed break points slope estimates from piecewise linear regression. 1000 bootstrapped replicates performed, mean and 95% credible interval of slope presented. Rates in italics include a positive slope 2 or 3 (i.e. an increase in titre) in confidence intervals.

As can be seen in Figure 36 there is considerable variation between individuals in the magnitude of their response. To try and account for this, piecewise linear regression was refit with a random effects term allowing a different intercept for each subject. There were not enough data points to allow different breakpoints and slopes for each individual but this also fits in with the assumption that the overall trend of response is similar between all individuals although the magnitude of response may differ. 1000 bootstrapped replicates were run to calculate mean estimate and 95% credible intervals. In the sampling procedure, sampling with replacement was allowed by participant ID but no resampling was performed on serum sample time points. As in the initial regressions, models were fit to determine optimal breakpoint(s) followed by investigation of slope variability for the mean breakpoint(s).

Compared with the initial regression, the introduction of random effects gives a similar range of time to peak response but with narrower 95% CI (Table 4.13). Of note, the time of peak response in H3N2 2003 is later at 22 days (rather than 10 days without random effects). The inclusion of random effects is likely to be accounting for the higher baseline and censored values to give a better overall

idea of trend. Where two breakpoints were fit, the second breakpoint was more consistent between strains in both mean and credible interval estimates. Using random effects all models were able to converge in contrast to the initial regressions not accounting for individual differences. Using the methods described by (Nakagawa and Schielzeth, 2013) it is possible to determine the adjusted R^2 of random effects model. The first measurement describes the proportion of variance explained by the fixed factors only (marginal R^2). The second is the conditional R^2, which describes the proportion of variance explained by both the fixed and random factors. As would be expected, the marginal R^2 for the random effects models (Table 4.13) are similar to the R^2 for the models without random effects (Table 4.9). The inclusion of the random effects shows a considerable improvement in the R^2 and demonstrates the importance of individual variability and baseline titre on estimation of titre dynamics. The inclusion of random effects leads to a bigger proportion of variance explained in older strains compared to more recent strains, likely due to the larger variability in baseline titre (see Chapter 3). Only H3N2 1968 showed an improvement in fit as measured by AIC between one and two breakpoints even with random effects.

| | Single Breakpoint | | | | | Two Breakpoints | | | | | |
	Break (95%CI)	AIC	aR² M	aR² C	Fit (n)	Break 1 (95%CI)	Break 2 (95%CI)	AIC	a R² M	a R² C	Fit (n)
H3N2 1968	30.3 (17.1-32.6)	1103.97	0.38	0.76	1000	30.3 (18.1-32.5)	92.3 (57.6-118.1)	1096.86	0.39	0.78	1000
H3N2 2003	22.6 (5.5-36)	934.75	0.36	0.61	1000	23 (5.3-36)	117.4 (44-176)	946.05	0.35	0.61	1000
H3N2 2005	28.3 (8.8-35.8)	912.35	0.57	0.76	1000	28.4 (8.5-35.8)	103.9 (91-205.5)	923.84	0.57	0.76	1000
H3N2 2007	31.5 (31-33)	949.08	0.60	0.82	1000	31.6 (31-33)	99.4 (59.6-126.4)	956.24	0.60	0.82	1000
H3N2 2009	30.5 (22.4-34.7)	1003.36	0.67	0.83	1000	30.6 (29-33.7)	137.5 (91-210)	1013.01	0.67	0.83	1000
H3N2 2011	32.5 (17.1-36)	896.82	0.65	0.82	1000	32.6 (17.9-36)	126.8 91-206.9)	907.53	0.65	0.82	1000

Table 4.13 Homotypic Response to H3N2 Infection. Piecewise linear regression with random effects for individual intercept. Bootstrapped 95% Credible Interval from 1000 replicates. Adjusted R^2 calculated using methods described in (Nakagawa and Schielzeth 2013), M = marginal R^2, C = conditional R^2. Comparison of AIC within strain only, i.e. across rows only

	Break Point	Slope 1: Days to Rise 1 Log Unit	Slope 2: Days to Fall 1 Log Unit	Fit (n)
H3N2 1968	30	10.4 (9.4-11.5)	109.8 (94.6-127.5)	1000
H3N2 2003	23	12.4 (10.5-14.4)	432.6 (299.4-643.4)	1000
H3N2 2005	28	10.3 (9.4-11.3)	398.8 (259.3-655.8)	1000
H3N2 2007	32	9.6 (8.9-10.2)	184.3 (140.6-244.6)	1000
H3N2 2009	31	8 (7.5-8.5)	304.1 (202.1-477.1)	1000
H3N2 2011	33	10.3 (9.6-11.1)	388.7 (257.5-644.9)	1000

Table 4.14 Homotypic Response to H3N2 Infection. Using single fixed break point slope estimates from piecewise linear regression with random intercept for individual response. 1000 bootstrapped replicates performed, mean and 95% credible interval of slope presented.

In the models with a single breakpoint, the credible intervals of the rate of decline were narrower where random effects were included (Table 4.14). The fastest rate of decline was in H3N2 1968 and 2007. Although it is feasible that titre rises relating to the boosting of previously exposed strain last a shorter length of time than a new response to the most recent circulating strains, the presence of censored titres currently means these findings cannot be explored further. As in the model without random effects, the two break point model includes rises within the credible intervals where there is a significant proportion of censored values (Table 4.15).

	Break Point 1	Break Point 2	Slope 1: Days to Rise 1 Log Unit	Slope 2: Days to Fall 1 Log Unit	Slope 3: Days to Fall 1 Log Unit	Fit (n)
H3N2 1968	30	92	9.6 (8.7-10.7)	53.8 (44.1-68.1)	226.7 (161.4-331.7)	1000
H3N2 2003	23	117	12.2 (10.5-14.2)	328.5 (174.2-708.8)	-1051.8 (->3000->3000)	1000
H3N2 2005	28	104	10.1 (9.3-11.1)	287.6 (166.5-490.7)	626.9 (228.9-2356.7)	1000
H3N2 2007	32	99	9.3 (8.7-9.9)	108.7 (78.7-154.7)	355.8 (198.5-692.8)	1000
H3N2 2009	31	138	7.9 (7.4-8.3)	198 (128.4-308.3)	84.2 (-9783-7880.3)	1000
H3N2 2011	33	127	10.2 (9.5-11)	254.1 (159-443.1)	1146.6 (->3000->3000)	1000

Table 4.15 Homotypic Response to H3N2 Infection. Using two fixed break points slope estimates from piecewise linear regression with random intercept for individual response. 1000 bootstrapped replicates performed, mean and 95% credible interval of slope presented. Rates in italics include a positive slope 2 of slope presented

4.3.3.2 Non-linear model of response

An alternative to piecewise linear regression is to use a non-parametric approach to titre response using generalised additive models and splines. These methods allow a flexible approach to the shape of response accounting for individual variation when random effects are used. For homosubtypic responses in individuals infected with H3N2, the fitted splines show a rapid rise and predominantly biphasic decline which is most pronounced in strains with the fewest censored values in follow-up (Figure 37). All strain specific splines were highly significant (p values <0.001). When the non-linear splines of antibody response were compared to piecewise linear regression with one or two breakpoints, there was little difference in the adjusted R^2 on the fixed component (marginal R^2) of the regression models (Table 4.16). The adjusted R^2 was highest in the most recent strains for both modelling approaches. AIC was lower for the simpler piecewise regressions for all strains but a significant improvement in AIC between one or two breaks was only seen where there were few censored values (H3N2 1968 and 2007).

	GAM		Piecewise One Break		Piecewise Two Break	
	AIC	R^2	AIC	R^2	AIC	R^2
H3N2 1968	1092.067	0.378	1104.099	0.38	1097.25	0.391
H3N2 2003	934.191	0.371	934.794	0.355	946.052	0.354
H3N2 2005	913.534	0.583	912.343	0.567	923.829	0.566
H3N2 2007	952.844	0.604	949.132	0.599	956.017	0.6
H3N2 2009	1006.155	0.675	1003.329	0.672	1012.91	0.672
H3N2 2011	899.258	0.667	896.679	0.654	907.254	0.654

Table 4.16 Comparison of Fixed Effects for Homosubtypic Antibody Response in Individuals Infected with Influenza A H3N2. Comparison of AIC within strain only, i.e. across rows only.

Figure 37 Antibody response to Homosubtypic Strains in Individuals infected with Influenza A H3N2. Red line is GAM fit for titre against time since infection. Grey band is 95% CI of mean estimated response. Vertical dashed lines are the strain specific breakpoints from piecewise linear regression estimates.

In both the GAM and piecewise regressions, a random intercept for each individual was allowed which makes the assumption that the baseline titre will vary the magnitude but not shape of response. One explanation for the rising R^2 in more recent strains could be that this is a new immune response to infection with current strains and that most people have similar kinetics regardless of age and therefore prior exposure to this strain. As seen in Chapter 3 most infected individuals had similar baseline titres of recent strains so these kinetics are reasonably consistent between individuals. For older strains (1968 and 2003), the adjusted R^2 was lower for post infection response because factors other than just baseline titre dictate the shape of the antibody dynamics. Although the simpler piecewise regression models were better fitting than the GAM model of titre response, the flexibility allowed by the GAM modelling approach will allow easier exploration of these additional features such as age, time of infection and baseline titre.

Age was incorporated in the model of homosubtypic antibody dynamics in two ways. The first was to allow a different spline of titre response for each age category. The other was to fit a model including separate splines of time since infection and age. Fitting splines for different age categories did not significantly improve model fit (Table 4.17) although all splines were highly significant where there were adequate numbers of data points (Figure 38). In contrast, the inclusion of a single age

spline showed a significant improvement compared to GAM with only time since infection or separate splines. The adjusted R^2 of the fixed components are most improved in the historic strains although the AIC of the simpler piecewise regressions are still better than the more complicated GAM models.

	Separate Splines for each Age Category		Single Age Spline	
	AIC	R^2	AIC	R^2
H3N2 1968	1109.59	0.397	1051.427	0.57
H3N2 2003	963.501	0.366	930.406	0.431
H3N2 2005	948.559	0.58	909.735	0.608
H3N2 2007	1001.357	0.592	954.205	0.619
H3N2 2009	1051.534	0.667	1008.569	0.681
H3N2 2011	938.482	0.663	895.812	0.686

Table 4.17 Comparison of Fixed Effects for Homosubtypic Antibody Response by Age in Individuals Infected with Influenza A H3N2. Comparison of AIC within strain only, i.e. across rows only.
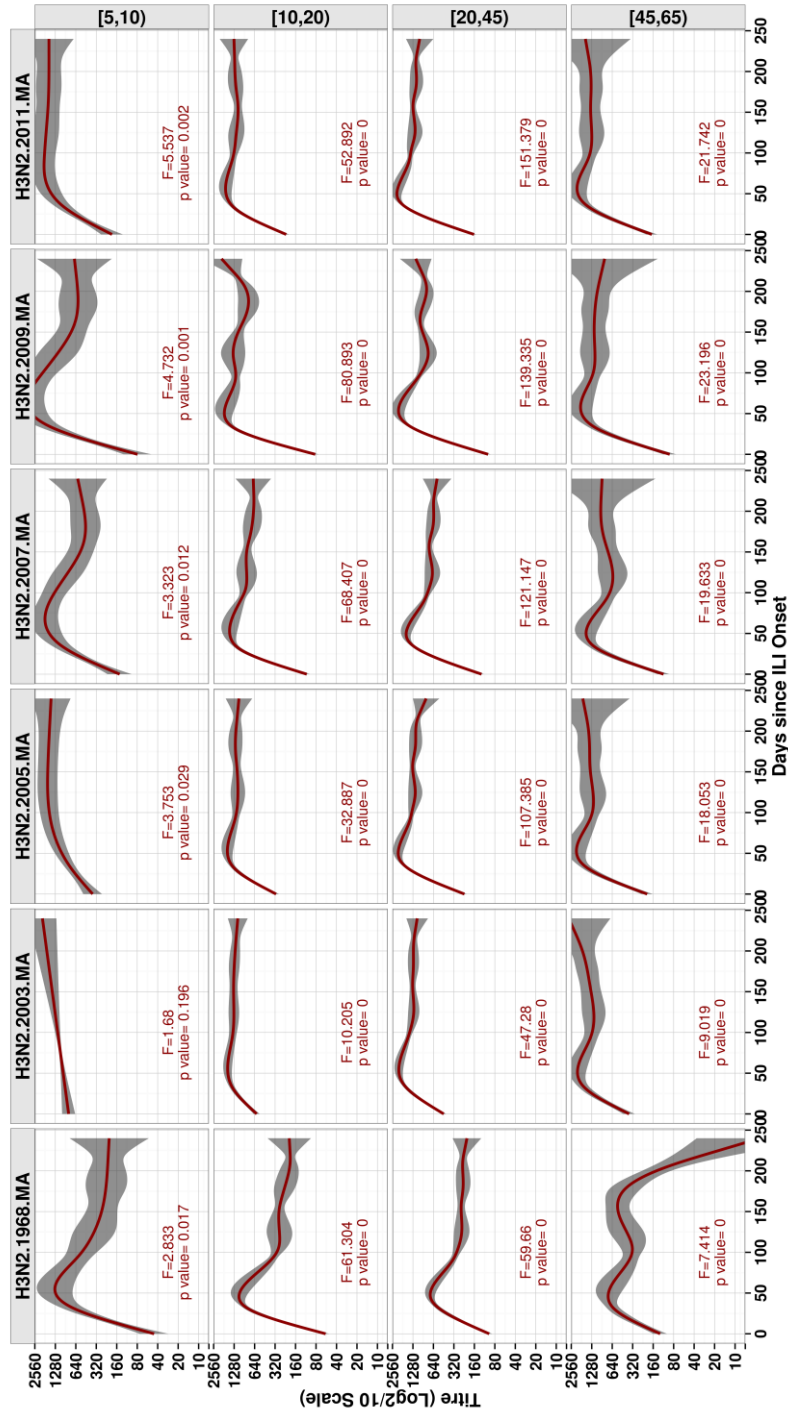
Figure 38 Antibody response to Homosubtypic Strains by Age Category in Individuals infected with Influenza A H3N2. Points represent measured titre. Red line is GAM fit for titre against time since infection with random effects for individual response. Grey band is 95% CI of mean estimated response.

If the GAM with splines of age and days since infection are used to predict the antibody response for different ages (15, 25, 40 and 60) then a pattern of response can be seen (Figure 39). The predicted response to H3N2 1968 shows that a higher response is expected in older individuals however in

more recent strains this changes, with highest response seen in younger individuals with less differentiation between other age groups.
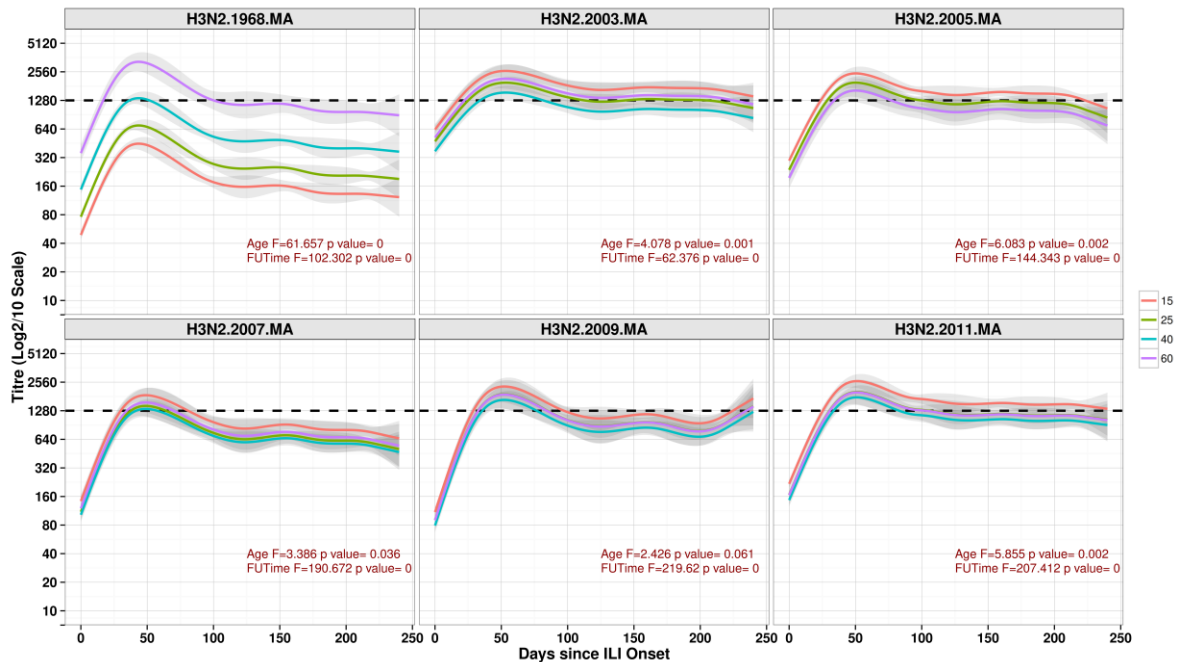
Figure 39 Predicted Antibody Response to Homosubtypic Strains for Selected Ages in Individuals infected with Influenza A H3N2. Age is included as a spline of response. Grey band is 95%CI of Mean Response. Horizontal dashed line is current limit of detection for Microarray.

To this point all strain specific models have been fit separately. An alternative approach is to analyse all titre responses together with strain specific splines while maintaining random effects for individual subjects. Initially this model was fit without age and all strains remained highly significant. The combined model led to an improvement in fit for some strains but poorer for more recent strains (adjusted R^2 for combined model 0.474 vs Table 4.16). Using this approach the model is unable to converge when a spline of age or age at emergence was included.

To explore possible mechanisms behind the age differences, titre response was categorised depending on whether this was i) a titre response to a recent strain (H3N2 2011), ii) a titre response to historic strains an individual could have been exposed to (i.e. subject was alive at the time of circulation) iii) a titre response to historic strains an individual could not have been exposed to (i.e. subject was not alive at the time of circulation). The overall model fit was not significantly improved compared to using splines for individuals strains (adjusted R^2 0.459) and the addition of age did not improve this (adjusted R^2 0.459). Splines for the exposure period were highly significant in both of these models but age spline was not (Figure 40). Compared to both responses to recent strains and to historic strains individuals could have encountered, the titre response in the historic strains individuals could not have encountered appear to give a higher peak and faster decline. However, it

is worth noting that this response is entirely based on titre response to H3N2 1968 strains (Table 4.18), it would be important to explore this in more detail with a broader range of strain years.

Figure 40 Homosubtypic Titre Response by Strain Exposure in Individuals infected with Influenza A H3N2. Age is included as a spline of response in bottom row. Grey band is 95% CI of Mean Response. Horizontal dashed line is current limit of detection for Microarray.

	Historic Exposed	Historic Not Exposed	Recent
H3N2 1968	33	323	0
H3N2 2003	356	0	0
H3N2 2005	356	0	0
H3N2 2007	356	0	0
H3N2 2009	354	0	0
H3N2 2011	0	0	356

Table 4.18 Number of Titres by Strain for Each Strain Exposure Category

4.3.4   Heterosubtypic Antibody Dynamics

In order to investigate the apparent heterosubtypic rise following acute infection, a combination of piecewise linear regression and GAM with splines were used to investigate the dynamics in the same way as the homosubtypic response was investigated. This analysis will look at the response to H1N1 strains in individuals infected with H3N2 only. Figure 41 shows the individual subject trajectories throughout the follow-up period, although the increase between baseline and first follow-up can be

seen clearly, there appears to be a pattern of decline that is less consistent than in the homosubtypic response for the same individuals plotted in Figure 36.
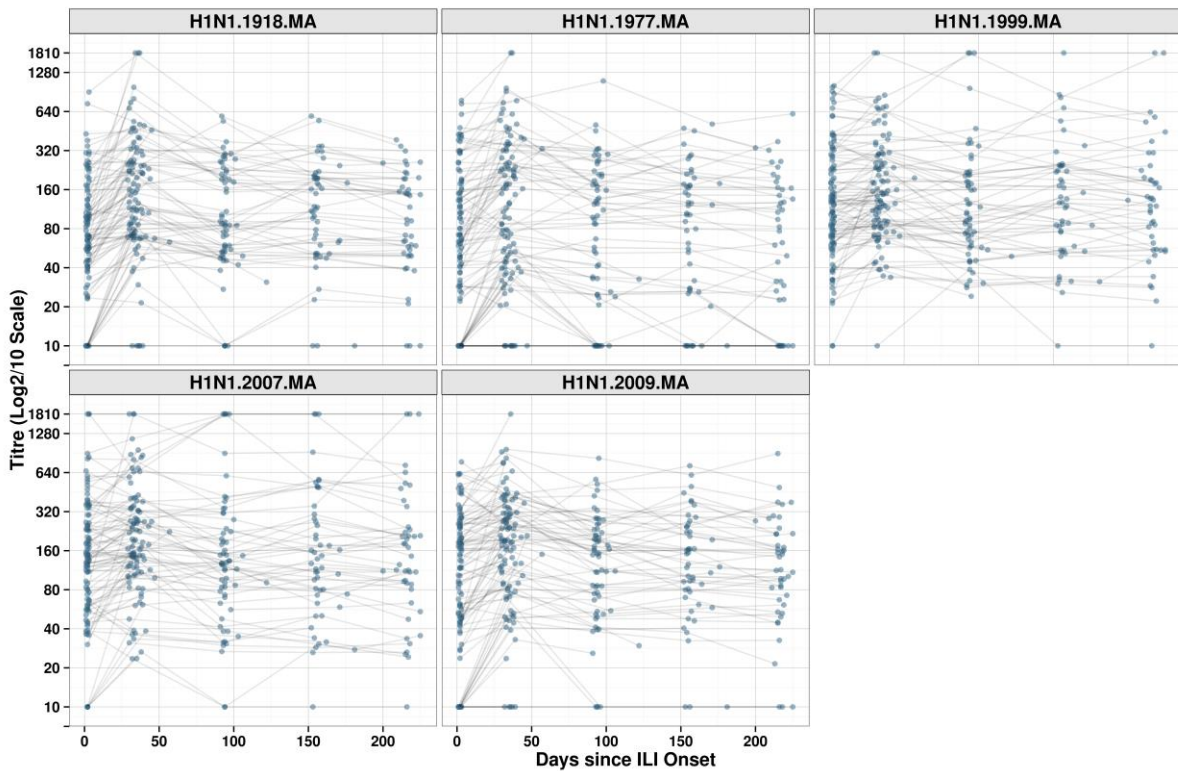
Figure 41 Heterosubtypic Strain Response in individuals infected with H3N2. Individual trajectories in grey.

4.3.4.1   Piecewise Linear Regression

For the heterosubtypic response all models included individual random effects as this method clearly demonstrated better fit in the homosubtypic response. As before a two-step process was used with the determination of optimal breakpoints followed by estimation of the slope variation around these breakpoints. 1000 bootstrapped replicates were run to calculate mean and credible interval. In the sampling procedure, sampling with replacement was allowed by participant ID but no resampling was performed on serum sample time points.

As with the homosubtypic response, the first breakpoint localises around 30 days although there is more variation than in the homosubtypic response (Table 4.19). The 95% CI are widest for H1N1 1999 and 2007 which did not demonstrate a significant titre change between Visit one and two in section 4.3.2.2. Where a second breakpoint was fit the mean break was at around day 100 as with the homosubtypic response.

	Single Breakpoint					Two Breakpoints					
	Break (95%CI)	AIC	aR^2 M	aR^2 C	Fit (n)	Break 1 (95%CI)	Break 2 (95%CI)	AIC	aR^2 M	aR^2 C	Fit (n)
H1N1 1918	26.7 (6.4-33.9)	1015.4	0.07	0.84	1000	26.9 (7-33.5)	93.2 (70.1-96)	1013.50	0.07	0.85	1000
H1N1 1977	31.5 (30-35)	1022.3	0.03	0.90	1000	31.5 (30-35)	94.5 (57-121)	1024.13	0.03	0.91	1000
H1N1 1999	40.5 (6.4-129)	855.3	0.01	0.87	1000	41 (6-154)	104.4 (55.2-210)	860.73	0.01	0.88	1000
H1N1 2007	34.4 (4.9-122)	977.0	0.01	0.85	1000	35 (5-122)	97 (56.1-162.1)	982.25	0.01	0.85	1000
H1N1 2009	24.5 (4.3-36)	992.5	0.04	0.86	1000	25.4 (4.7-36.2)	97.3 (93-118.1)	996.09	0.04	0.87	1000

Table 4.19 Heterotypic Response to H3N2 Infection. Piecewise linear regression with random effects for individual intercept. Bootstrapped 95% Credible Interval from 1000 replicates. Adjusted R^2 calculated using methods described in (Nakagawa and Schielzeth 2013), M = marginal R^2, C = conditional R^2. Comparison of AIC within strain only, i.e. across rows only.

Slope variation for heterosubtypic response was consistent with the magnitude of change seen in section 4.3.2 with around a one log rise in the month following infection (Table 4.20). The rate of decline was more consistent for the strains where there was a significant rise (H1N1 1918, 1977 and 2009) with a one log unit decline over 260 days. This was slower than in many of the homosubtypic strains (Table 4.14) but had a narrower 95% CI. Most subjects did not have censored values for H1N1 strains and it is likely that this allows for better estimation and can explain the narrower credible intervals.

	Break Point	Slope 1: Days to Rise 1 Log Unit	Slope 2: Days to Fall 1 Log Unit	Fit (n)
H1N1 1918	27	23.5 (19.5-28.7)	260.1 (194-357.5)	1000
H1N1 1977	32	37.6 (30.1-47.7)	258.4 (196.7-345.5)	1000
H1N1 1999	41	132.2 (80.6-231.6)	17.1 (373.5- -2040.7)	1000
H1N1 2007	35	86.7 (54.2-155.1)	505.1 (301.5-988.5)	1000
H1N1 2009	25	30.4 (23.6-40)	257.7 (202.1-331.4)	1000

Table 4.20 Heterosubtypic Response to H3N2 Infection. Using single fixed break point slope estimates from piecewise linear regression with random intercept for individual response. 1000 bootstrapped replicates performed, mean and 95% credible interval of slope presented.

When a second breakpoint was fitted, the rate of decline between the two break points suggested that it would take between 100 and 200 days for the titre to drop one log unit. The rate of decline

after the second breakpoint included rates of decline which would be considered implausible (time to drop one log unit greater than 10 years, a rise in titre after the second breakpoint, Table 4.21).

	Break Point 1	Break Point 2	Slope 1: Days to Rise 1 Log Unit	Slope 2: Days to Fall 1 Log Unit	Slope 3: Days to Fall 1 Log Unit	Fit (n)
H1N1 1918	27	93	20.6 (17.2-24.7)	99.6 (75.6-137.7)	>3000 (>3000- ->3000)	1000
H1N1 1977	32	95	33.4 (26.7-41.7)	115 (80-172.7)	>3000 (>3000- ->3000)	1000
H1N1 1999	41	104	99.4 (70.1-144)	209.5 (113.6-453.2)	982.9 (>-3000- >3000)	1000
H1N1 2007	31	97	63.2 (42.7-97.2)	190.7 (117.1-393.8)	530.6 (>-3000 ->3000)	1000
H1N1 2009	26	97	27.8 (21.8-35.8)	122.2 (85.8-176.7)	754.9 (349.2-2636.2)	1000

Table 4.21 Heterosubtypic Response to H3N2 Infection. Using two fixed break points slope estimates from piecewise linear regression with random intercept for individual response. 1000 bootstrapped replicates performed, mean and 95% credible interval of slope presented. Rates in italics include a positive slope 2 of slope presented

4.3.4.2   Non-linear Model of Response

The use of non-parametric splines to investigate the heterosubtypic response demonstrated a mean titre rise of approximately one log unit which was short lived and appeared to return to near baseline values by 150 days (Figure 42). All strain specific splines were highly significant (p values <0.001). Despite the highly significant strain specific splines the percentage deviance explained (as measured by the adjusted R^2) by the splines remains very low when compared to the adjusted R^2 of the homosubtypic response (Table 4.22). In contrast to the homosubtypic response, the best AIC for each strain was always using a GAM approach rather than the piecewise linear model.

	GAM		Piecewise One Break		Piecewise Two Break	
	AIC	RSq	AIC	RSq	AIC	RSq
H1N1 1918	998.273	0.063	1015.456	0.067	1013.53	0.073
H1N1 1977	1008.38	0.023	1022.403	0.029	1024.042	0.032
H1N1 1999	840.838	0.001	855.48	0.006	860.718	0.009
H1N1 2007	964.274	0.003	977.057	0.011	982.25	0.015
H1N1 2009	976.922	0.02	992.504	0.035	996.051	0.039

Table 4.22 Comparison of Fixed Effects for Heterosubtypic Antibody Response in Individuals Infected with Influenza A H3N2. Comparison of AIC within strain only, i.e. across rows only.
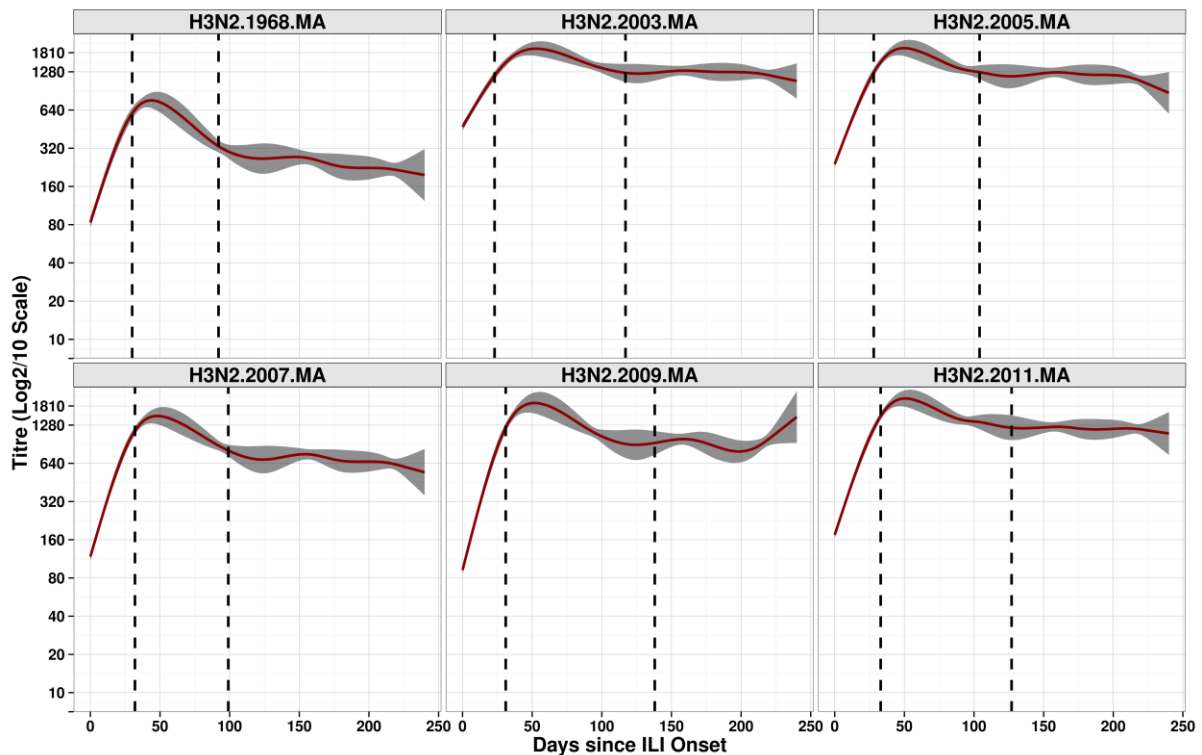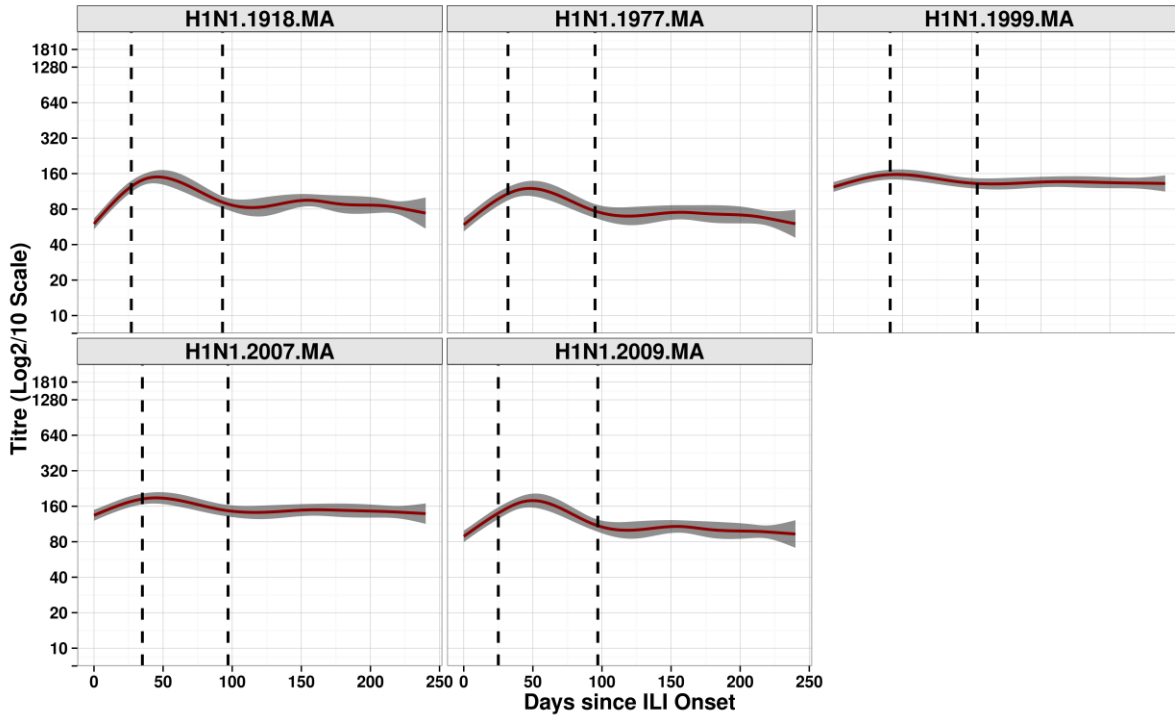
Figure 42 Antibody response to Heterosubtypic Strains in Individuals infected with Influenza A H3N2. Red line is GAM fit for titre against time since infection. Grey band is 95% CI of mean estimated response. Vertical dashed lines are the strain specific breakpoints from piecewise linear regression estimates.

Age was incorporated into the model of heterosubtypic response. As with the homosubtypic response, fitting splines for different age categories did not improve the model fit compared to time since ILI alone but using a spline of age did improve the adjusted R^2 of the fixed component (Table 4.23). The predicted mean antibody response for different ages (15, 25, 40 and 60) shows a faster rate of decline and return to baseline in younger individuals for H1N1 1918 and 2009 (Figure 43). When considered in terms of exposure the largest heterosubtypic response was for historic strains an individual could not have been exposed, suggesting this could be a either an assay effect or non-specific antibody production rather than a boosting of immunological memory (Figure 44).

	Separate Splines for each Age Category		Single Age Spline	
	AIC	aR^2	AIC	aR^2
H1N1 1918	1032.609	0.045	1003.852	0.079
H1N1 1977	1033.717	0.007	976.034	0.291
H1N1 1999	857.507	0.005	844.104	0.006
H1N1 2007	980.672	0.003	953.979	0.077
H1N1 2009	1002.173	0.006	970.205	0.068

Table 4.23 Comparison of Fixed Effects for Heterosubtypic Antibody Response by Age in Individuals Infected with Influenza A H3N2. Comparison of AIC within strain only, i.e. across rows only.
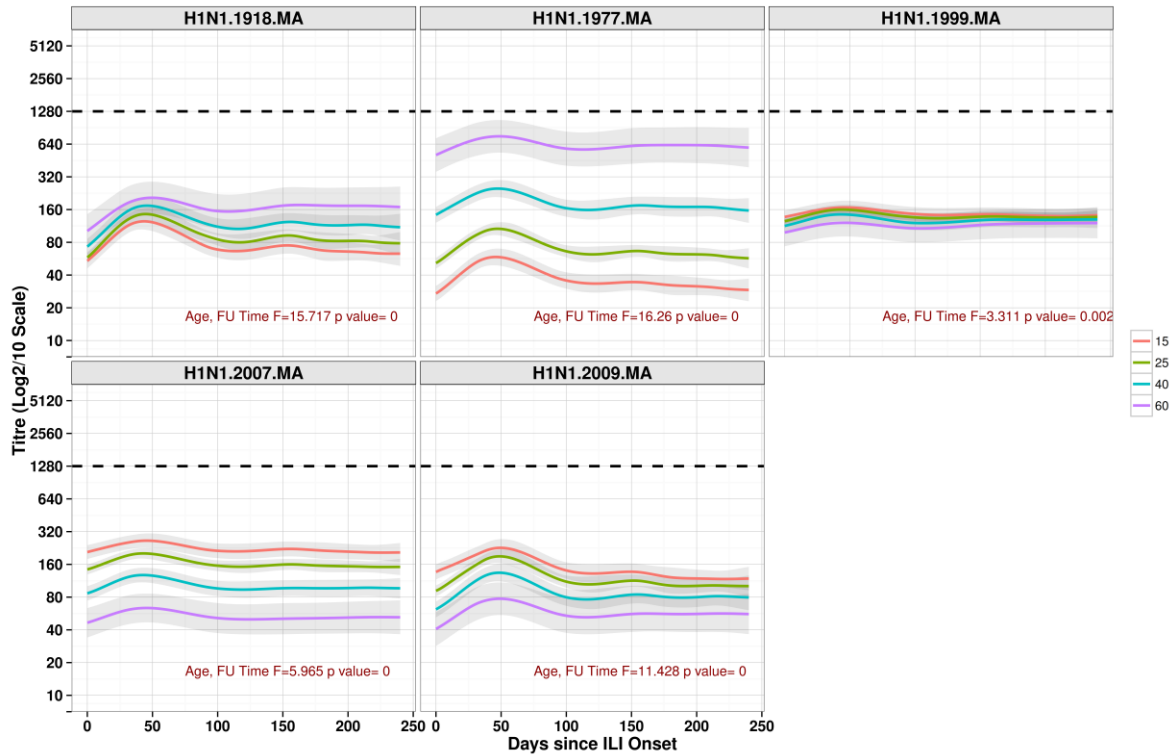
Figure 43 Predicted Antibody Response to Heterosubtypic Strains for Selected Ages in Individuals infected with Influenza A H3N2. Age is included as a spline of response. Grey band is 95% CI of Mean Response. Horizontal dashed line is current limit of detection for Microarray.
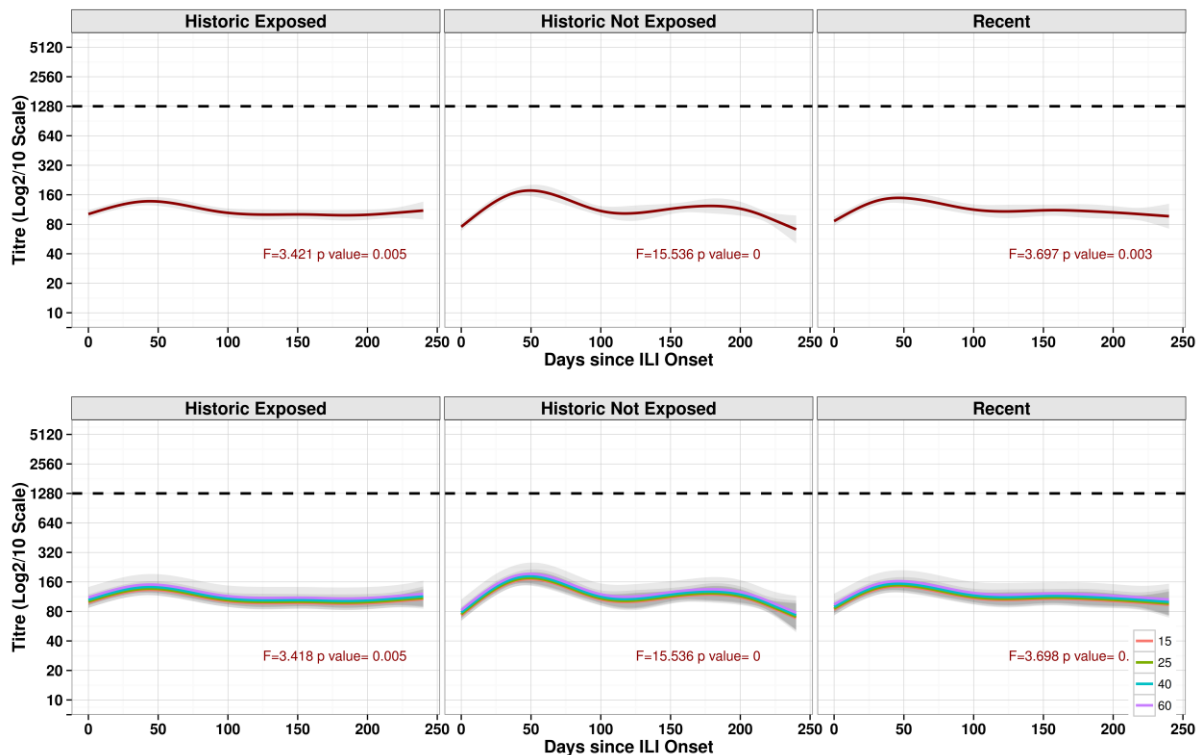
Figure 44 Heterosubtypic Titre Response by Strain Exposure in Individuals infected with Influenza A H3N2. Age is included as a spline of response in bottom row. Grey band is 95% CI of Mean Response. Horizontal dashed line is current limit of detection for Microarray.

4.3.5 Antibody Response after Non-Influenza ILI

In contrast to both the homo- and heterosubtypic responses in those infected with H3N2, there is not an obvious pattern of response seen in individuals who had non-influenza ILI (Figure 45).
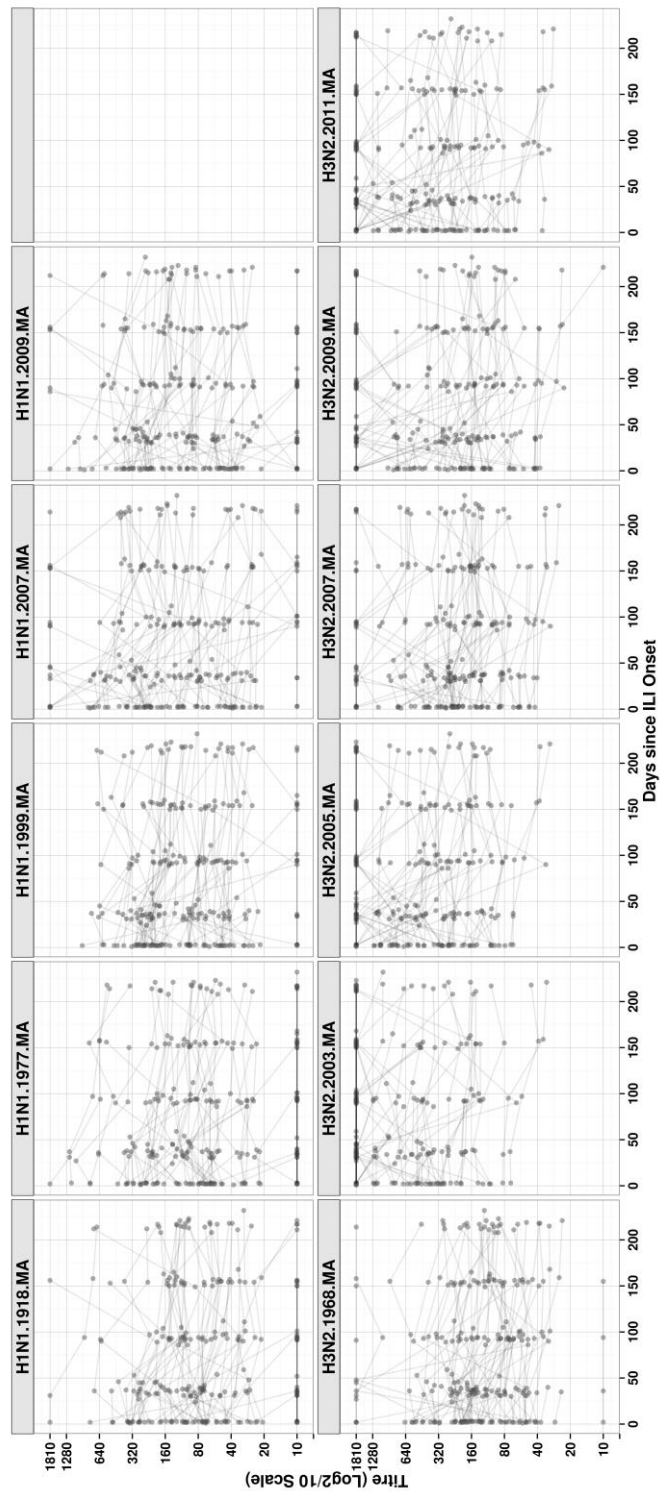
Figure 45 Strain Response in individuals with non-influenza ILI. Individual trajectories in grey.

4.3.5.1 Linear Model of Response

Piecewise linear regression was not performed as there was no suggestion from the raw data of a biphasic response. Instead, linear regression with individual random effects for intercept was used. 1000 bootstrapped replicates were run to calculate mean and credible interval of the slope. In the sampling procedure, sampling with replacement was allowed by participant ID but no resampling was performed on serum sample time points. Slope variation for those not infected with influenza was consistent with the decline in the homosubtypic response with a one log unit decline approximately every one to two years for most strains (Table 4.24).

Strain	H1N1 Slope 1: Days to Fall 1 Log Unit	Fit (n)	Strain	H3N2 Slope 1: Days to Fall 1 Log Unit	Fit (n)
H1N1 1918	375.2 (214.5-725.9)	1000	H3N2 1968	387.2 (243-678.1)	1000
H1N1 1977	369.7 (240.9-657)	1000	H3N2 2003	974.2 (355.3-2759.8)	1000
H1N1 1999	506.3 (275.5-938.9)	1000	H3N2 2005	691.5 (261-2660.2)	1000
H1N1 2007	374.6 (214.3-832)	1000	H3N2 2007	547.7 (279.8-1603.6)	1000
H1N1 2009	2353.3 (2596.1-3343.8)	1000	H3N2 2009	672.9 (3123.8-4288.2)	1000
			H3N2 2011	549.5 (253.5-2517.8)	1000

Table 4.24 Antibody Response following non-influenza ILI. Slope estimates from linear regression with random intercept for individual response. 1000 bootstrapped replicates performed, mean and 95% credible interval of slope presented.

4.3.5.2 Non-linear Model of Response

Using non parametric splines to investigate response in individuals presenting with non-influenza ILI demonstrated a marginal improvement in model fit compared with the linear model (Table 4.25). However the R^2 of the fixed effects was low for both GAM and linear models. H3N2 2003, 2009 and H1N1 1918 all preferentially fit linear models even when the option of fitting a spline was allowed (as determined by edf ~ 1). Where a spline was fit, a general decline was noted with a slight plateau at past 150 days which may be related to the number of samples available at this point (Figure 46). All strain specific splines were significant (p values <0.05). The inclusion of age did not significantly improve the model fit. Where models of exposure were fit, there was not a considerable difference in mean rates of decline (Figure 47).

	GAM			Linear Response	
	edf	AIC	RSq	AIC	RSq
H1N1 1918	1.646	770.852	0.006	777.74	0.015
H1N1 1977	1	683.056	0.004	689.604	0.013
H1N1 1999	2.169	722.82	0.007	729.86	0.013
H1N1 2007	1.416	848.654	0.01	855.289	0.014
H1N1 2009	1	853.114	0.01	859.663	0.005
H3N2 1968	1.593	740.34	0.009	747.082	0.019
H3N2 2003	1	705.131	0.002	711.68	0.005
H3N2 2005	2.561	767.91	0.001	776.426	0.008
H3N2 2007	2.179	743.555	0.001	751.513	0.01
H3N2 2009	1.726	876.644	0.001	883.634	0.006
H3N2 2011	2.052	804.968	0	812.838	0.008

Table 4.25 Comparison of Fixed Effects for Antibody Response in Individuals not Infected with Influenza. Comparison of AIC within strain only, i.e. across rows only.
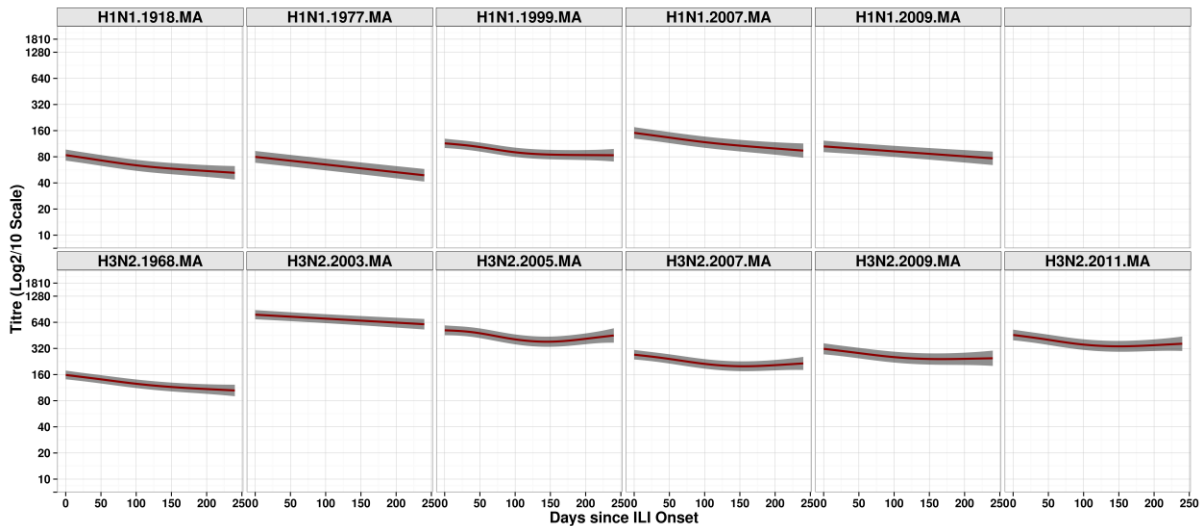
Figure 46 Antibody response in Individuals not infected with Influenza. Red line is GAM fit for titre against time since infection. Grey band is 95% CI of mean estimated response.

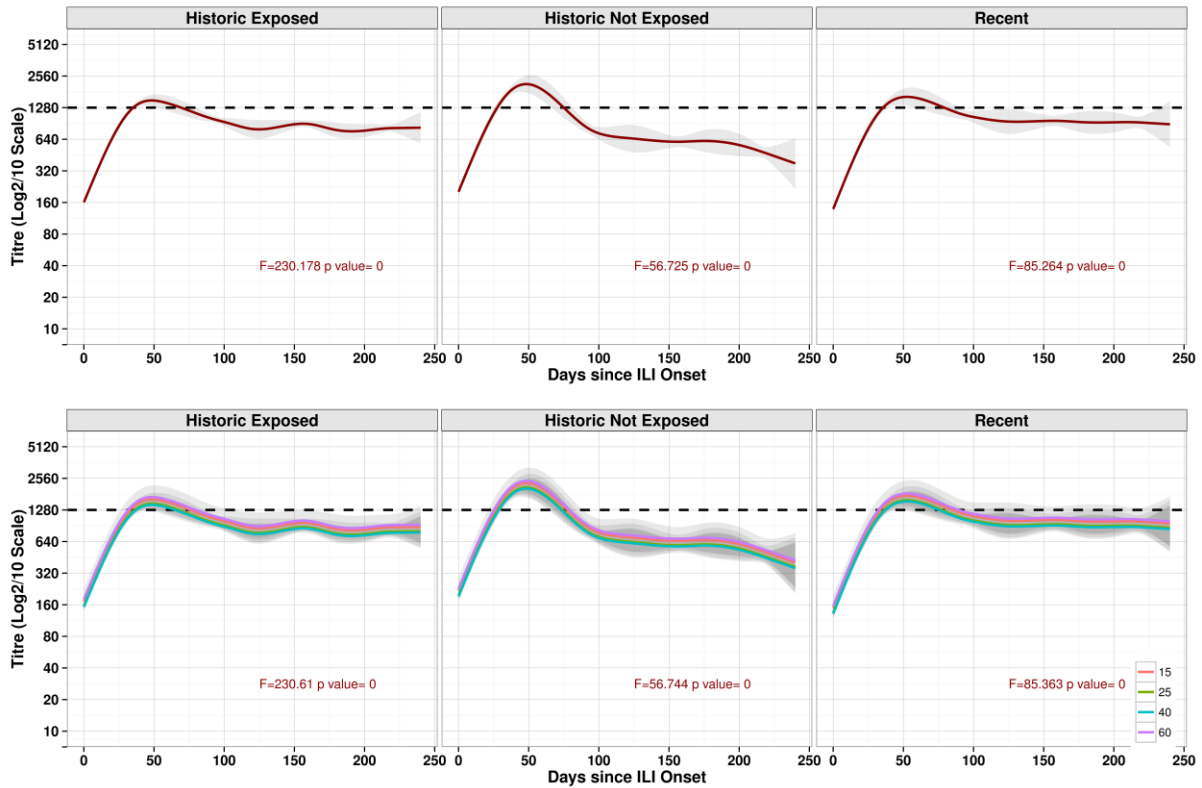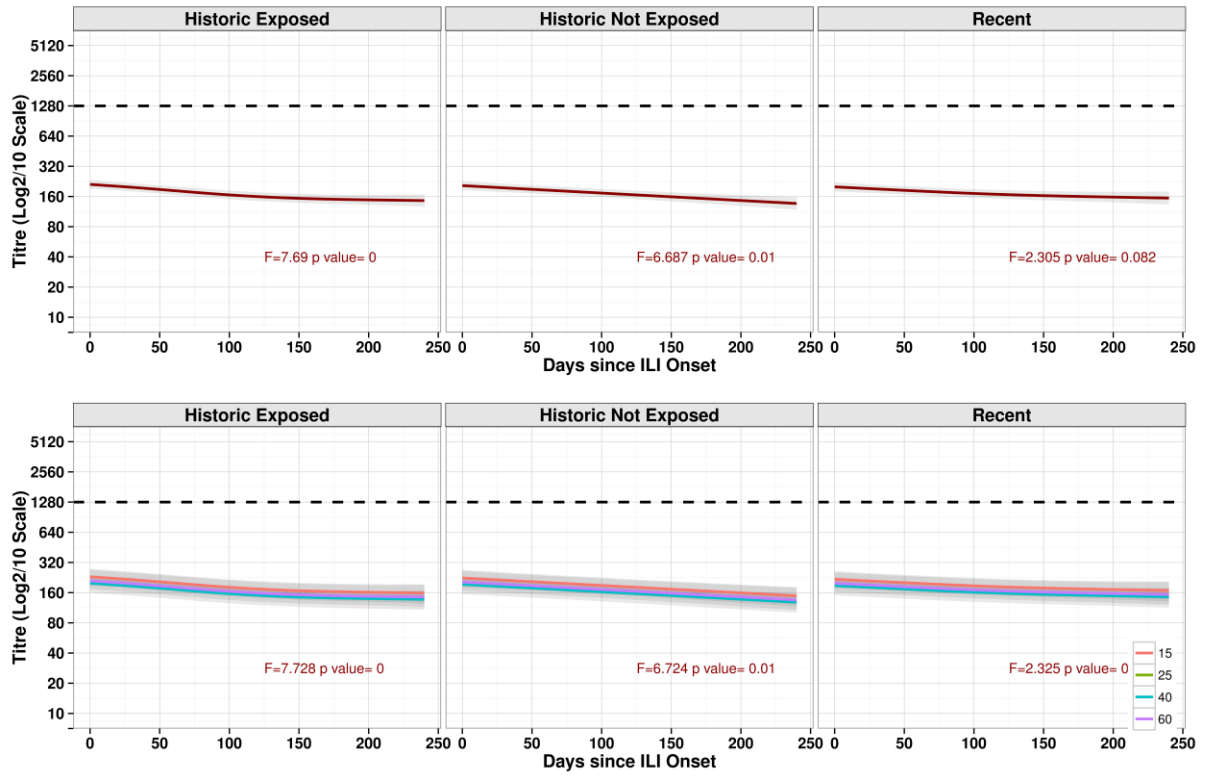*Figure 47 Antibody Response by Strain Exposure in Individuals not infected with Influenza. Age is included as a spline of response in bottom row. Grey band is 95% CI of Mean Response. Horizontal dashed line is current limit of detection for Microarray*

4.4 DISCUSSION

Influenza antibody dynamics have been studied for over 70 years. This has generally been restricted to early responses matched to the strain the individual was infected with. Despite being an integral part of influenza surveillance systems and pandemic planning through assessment of population immunity, our understanding of multi-strain and multi-subtype responses have been limited by the labour intensive nature of existing serological techniques. The development of the protein microarray used in this analysis and other high throughput techniques (Koopmans et al., 2011), along with a dramatic shift in the availability of computing power to analyse complex multivariate data sets not possible even a few decades ago, has given an opportunity to re-examine multi-strain and multi-subtype responses following infection.

As would be expected the most significant titre response post infection is in strains matched with the same subtype as the infecting virus (homosubtypic response). There is a higher mean change between visit one and two in more recent strains than in historic strains with most having a threefold increase in titre. Baseline titre is the strongest predictor of the magnitude of response but this result was heavily impacted by the current limit of detection of the microarray assay. For the most recent strains, those who had a very low baseline titre had a predicted titre change of four to five log2 titre increase (Figure 33). The current upper limit of detection of the microarray means that it is not possible to determine whether this magnitude of change is consistent across historic strains for a given baseline titre. Previous results have suggested that although the presence of homosubtypic response to historic strains increases with an individual's age (Horsfall and Rickard, 1941, Grilli et al., 1986, Fonville et al., 2014), the magnitude of rise following repeated exposures to live vaccine is lower than in primary exposure (Davenport and Hennessy, 1956, Yamane et al., 1981, Ochiai et al., 1986). As seen in Chapter 3, age was a significant predictor of baseline titre for most strains, and in multivariable analysis age was only an additional predictive value for H3N2 2009 but with a p-value of 0.02 this may represent chance finding given the number of tests performed.

Being infected with H3N2 resulted in a homosubtypic titre change 1.3 log higher than those infected with H1N1 in multivariable analysis. It is important to note that the initial univariate result which suggested a higher fold change in those infected with H1N1 was inaccurate as it had failed to take into account the higher baseline titres in H3N2 and the smaller potential rise because of the current limit of detection on the microarray. This phenomenon has been described previously and is an important consideration in vaccine studies where immunogenicity is an outcome (Nauta, 2011) and subjects have multiple infections or vaccinations during their lifetime. Methods to overcome this include the use of multivariable analysis with baseline titre as a specific variable as was performed in

this analysis, an alternative is to look at the covariance of change and this will be investigated further.

Time to peak antibody response of around 30 days was consistent with previous work (Horsfall and Rickard, 1941, Morris et al., 1966, Sonoguchi et al., 1986) and although memory response to historic strains are produced more quickly than response to new strains (Morris et al., 1966, Jao et al., 1970) this is only seen as a one to two day difference in reaching peak response in our study. This is likely to be heavily influenced by our study design which did not sample earlier than 30 days. Despite this limitation, the identification of time to peak response was possible using piecewise linear regression and was much improved when the magnitude of baseline titre was incorporated using random effects. The suggestion of a biphasic decline in antibody response has been demonstrated previously (Horsfall and Rickard, 1941, Foy et al., 1980, Lerman et al., 1980, Ochiai et al., 1986, Sonoguchi et al., 1986). Although the identification of a second breakpoint was less consistent than a model with a single breakpoint it was in keeping with previous results. The second breakpoint was identified more frequently in older strains. While this could be due to the presence of censored values, previous studies have demonstrated that the response to infecting strains persists longer than homosubtypic response to historic strains (Horsfall and Rickard, 1941, Grilli et al., 1986). As the follow-up period of this study was limited to seven months it is highly likely that there are too few sampling points after the second breakpoint for more recent strains for this to be identified. Previous work has also demonstrated that age appears to have an impact on rates of decline with the very youngest individuals (under 10 years of age) having a shortest duration of response but those in their second and third decade of life having the longest duration of response to contemporary strains (Hall et al., 1973, Boucher et al., 1979, Lerman et al., 1980). When exposure history is considered, historic strains that an individual cannot have been exposed to have the shortest duration of persistence after infection and the response is similar in magnitude and duration as the heterosubtypic response.

A consistent heterosubtypic response of approximately one log increase is detected for most strains. The lack of response in individuals with non-influenza ILI suggests this is an influenza specific effect rather than a non-specific activation of the respiratory immune system. Possible mechanisms of this include cross-reaction of infected strain specific antibodies (homosubtypic) in the acute phase ('sticky assay'), the release of heterosubtypic specific antibodies when infected acutely with the opposing strain or the release of specific antibodies against the stem component of the HA . The finding of heterosubtypic rise was more pronounced in subjects infected with H1N1 (i.e. a larger rise in H3N2 strains for individuals infected with H1N1). Previous results had suggested that heterosubtypic response was related to baseline antibody response (Morris et al., 1966, Pyhala,

1985) but there was not a clear relationship in these results. A higher heterosubtypic response was seen where baseline titre to the most recent heterosubtypic strain was lowest, this correlates with the rise in the homosubtypic titre suggesting that this could be cross reaction within the assay.

The duration of heterosubtypic response was shorter than the homosubtypic response with a more consistent second breakpoint at around 100 days for all strains. Although the rate of decline was rapid to this point, most strains had not returned to baseline at seven months. The results from those not infected with influenza suggest that a one log drop in antibody response can be expected every one to two years. The rate of antibody decline beyond the effect of recent infection was difficult to measure due to the limited duration of follow-up in the study but the credible intervals in the both the homosubtypic and heterosubtypic responses suggest that this could be a plausible rate of decline for all strains after the second breakpoint.

The results of this longitudinal study demonstrate repeat boosting of historic responses with each new infection. Although the most significant part of this response is within subtype, there appears to be a smaller but still significant between subtype boost which changes the overall trajectory of the decline. This boosting of historic responses has been described elsewhere (Lerman et al., 1980, Miller et al., 2013, Fonville et al., 2014) and would be an explanation for the persistence of antibody to strains encountered early in life if these were being boosted with each new infection throughout lifetime (Kucharski et al., 2015). One key question is whether these 'back boosted' titres to historic strains represent a functional change in immunity and an increase in strain specific immunity. The ability of PMA to measure neutralisation is yet to be established and correlation of these longitudinal results with HI is planned. As discussed in Chapter 3, when looking at multi-strain antibody response, H3N2 2005 and 2009 were most predictive of susceptibility to current strains. This suggests that back boosting of historical strains is less important for current susceptibility in drifted subtypes such as H3N2. However, in the face of novel pandemic strains, historical responses which have been boosted periodically over time are likely to provide protection (Van Kerkhove et al., 2013b). Some studies have shown that rises in specific HI titre over time are correlated with an increase in total IgG for group one or group two influenza subtype corresponding to the HI subtype (Miller et al., 2013). If this rise in total IgG is related to the production of antibodies to the conserved stem region of the haemagglutinin, then this may also be what is being measured in the short lived measurable titres to strains an individual has not been exposed to or heterosubtypic response (Lu et al., 2014). Further exploration of the antibody dynamics of avian strains on the microarray and in particular, within and across group responses (group 1 H1N1 & H5N1; group 2 H3N2 & H7N7) are planned.

Although all tests will have an upper limit of detection, it is important that this captures the majority of results we are interested in. The microarray used in this analysis was developed and validated predominantly using H1N1 rather than H3N2 strains (Koopmans et al., 2011, Baas et al., 2013, Huijskens et al., 2013). The upper limit of detection of 1280 is likely to be adequate for H1N1 with only 20% of subjects infected with H1N1 having values greater than this limit of detection one month after infection, the point likely to be the peak of response (Table 4.8). In contrast, up to 89% of H3N2 titres are above the limit of detection and this persists through the seven months of follow-up (Table 4.10). This demonstrates that the consideration of influenza A as a single entity is likely to be incorrect. This microarray was validated in the early post-pandemic phase following the emergence of H1N1 2009. Global public health emergencies such as this or Ebola in West Africa often provide the impetus for the development of novel diagnostics which are then used in the post pandemic phase. Serological diagnostics developed in a population with limited immunity are likely to perform differently in a population where they are measuring both historic responses and response to repeat infection. The determination of the maximum limit of detection of an array should relate to the questions that need to be answered. The use of a threshold approach to determine recent exposure (de Bruin et al., 2014) is likely to require fewer dilutions and a lower threshold than if investigating dynamics. Chapter 5 will discuss the issues around determining thresholds for recent exposure.

5 DETERMINANT OF ACUTE INFLUENZA INFECTION IN SEROEPIDEMIOLOGY STUDIES

ABSTRACT

Background: Accurate surveillance of influenza is essential to understand the global burden of disease. Serological surveillance is a longstanding method used to estimate population attack rates particularly in pandemic situations. In seasonal strains cross reaction from prior infection complicates this surveillance methods. Diagnostic accuracy of serological markers of recent infection is therefore crucial to accurately measure attack rates.

Methods: A prospective, observational study of patients with ILI in Ho Chi Minh City, Vietnam has been running since August 2013. Influenza A & B PCR and antibody testing to a panel of 11 human and 5 avian strains is performed using a novel protein microarray technique. A subset of subjects are followed up clinically and serologically for seven months, samples. Optimal threshold for sero-diagnosis was determined by ROC analysis using titre response to most recent strain and was compared to a multi-strain measure using a modified Simpson's diversity index.

Results: 470 samples from 186 ILI patients were available for analysis. For the most recent H3N2 2011 strain the optimal threshold was greater than log2 5.5 for all age groups at all time points. Sensitivity was greater than 90% for all thresholds but specificity was poor. Specificity was improved by using a multi-strain approach as measured by the diversity index. The sensitivity and specificity of serological measures of recent infection will lead to a significant over estimation of influenza seroprevalence unless test accuracy is adjusted for.

Conclusion: Seroepidemiology of influenza is challenging because of cross-reaction. Methods exist to improve the estimation from these methods and should be employed for non-pandemic influenza serosurveillance.

5.1 BACKGROUND

Influenza viruses circulate globally and seasonal epidemics are thought to be associated with three to five million severe clinical infections and 250,000 to 500,000 deaths each year (World Health Organisation, 2014c, Global Burden of Disease 2013 Mortality and Causes of Death Collaborators, 2015, Global Burden of Disease Study 2013 Collaborators, 2015). Morbidity and mortality are highest in the extremes of age (Cohen et al., 2010, Nair et al., 2011) but all ages are affected with repeat

infections throughout an individual's lifetime. In temperate countries predicable winter epidemics occur with annual incidence concentrated into a two to three month peak of activity with limited or no activity recorded between these annual peaks (Finkelman et al., 2007). In contrast to the predictable peaks in temperate countries, influenza transmission in tropical areas is less predictable with asynchronous peaks of activity, co-circulation of strains and apparent year round transmission (Moura, 2010) with the drivers for this difference not yet being fully understood (Tamerius et al., 2011, Paynter, 2015).

A variety of surveillance techniques are utilised to estimate the global impact of influenza including healthcare based syndromic and virological surveillance (World Health Organisation, 2014a). These systems have well recognised biases, in particular the under representation of clinically mild or asymptomatic cases (Van Kerkhove et al., 2013a). One approach used to try and quantify this is through the use of serological surveillance techniques, also known as seroepidemiology. There are a number of different ways that this method is employed, including large cross sectional surveys at the end of a winter season in temperate countries (Fragaszy et al., 2015) or through longitudinal serological sampling either in a specific observational cohort (Chen et al., 2010, Horby et al., 2012) or convenience sampling of repeat blood donors (McVernon et al., 2010, Wu et al., 2010, Sauerbrei et al., 2014). Haemagglutinin Inhibition assays (HI) are the traditional gold standard for seroepidemiology with long established diagnostic standards of a four-fold titre rise in paired samples or a titre of greater than 1:40 in cross sectional surveys. However, HI has challenges, including the time-intensive nature of performing the test and inter-laboratory variability.

These seroepidemiology methods have often been developed and employed in the face of an emerging pandemic where there is limited prior immunity and therefore most of the population have undetectable antibodies making seroprevalence curves easier to estimate (Van Kerkhove et al., 2013b). During the most recent pandemic in 2009, there was suggestion that the standard threshold of 1:40 was too high and was underestimating the true community attack rate (Cauchemez et al., 2012, Wu et al., 2014). However, most influenza infections do not occur during a pandemic period and most seroepidemiology studies are performed to assess attack rates of subtypes which have caused multiple infections during a lifetime. Cross reaction to the current circulating strain from previously encountered strains may result in detectable antibodies, reflecting prior infection rather than being a measure of recent infection (Smith et al., 2004, Kucharski et al., 2015). Most seroepidemiology studies only measure antibodies to current circulating strains due to the labour intensive nature of HI testing. High throughput technologies such as the protein microarray (PMA) have been designed to allow testing of multiple strains from a large number of samples but the appropriate markers of acute infection have yet to be determined. Microarray technology has been

employed previously to identify serological markers of response in infections which have considerable cross reaction (Sundaresh et al., 2007, Felgner et al., 2009, Crompton et al., 2010). The use of these methods in influenza have yet to be fully explored.

Serological surveys are a long standing method of investigating infectious diseases with two major goals i) determination of disease incidence and ii) assessment of susceptibility to disease (Van Kerkhove et al., 2013a). However, many of the existing statistical methods rely on assumptions of life-long immunity following infection which does not hold in influenza in non-pandemic situations (Hens et al., 2012). Following the emergence of H1N1 2009, age-specific seroprevalence studies were widely performed demonstrating both differences in attack rates in different age groups and different levels of pre-existing immunity (Van Kerkhove et al., 2013b). In the post-pandemic period there was a review of the methods used and discussions around how interpretation of influenza serology can be improved for future pandemic preparedness and understanding seasonal influenza (Cauchemez et al., 2012, Wu et al., 2014).

5.2 METHODS

A full description of the study procedures and laboratory testing is given in Chapter 2.

5.2.1 Design and Conduct of the Observational Study

This observational study was conducted between August 2013 and May 2015 at outpatient clinics and community medical practitioners in Ho Chi Minh City (HCMC), Vietnam. Individuals were invited to join the study if they were between 10 and 70 years of age with symptoms for less than 72 hours and if they fitted the ECDC ILI definition (European Centre for Disease Prevention and Control (ECDC), 2015) One anterior nasal swab and one throat swab were collected at recruitment and transported in a single tube of viral transport medium to central laboratory before being stored at -20°C within 24 hours. A 5ml serum sample was also collected at baseline and stored at -20°C within 24 hours.

A subset of patients were invited to join a longitudinal sub-study. All subjects testing positive for influenza A were eligible to join the sub-study. Negative control subjects were matched by age (+/- 5 years) and gender to influenza A positive subjects included in the follow-up study. Subjects attended for repeat serum sampling at 30, 90, 150 and 210 days (+/- 5 days) after recruitment. Active respiratory symptom surveillance was conducted during the follow-up period with telephone follow-up at 60, 120 and 180 days.

Both studies were approved by the Scientific and Ethical committees of Hospital for Tropical Disease, Ho Chi Minh City, Vietnam and Liverpool School of Tropical Medicine Research Ethics Committee, UK. Letters of agreement supporting the involvement of the community medical clinics were obtained from the Ho Chi Minh City Department of Health.

5.2.2   Sample Analysis

Respiratory samples were batched tested monthly for influenza A & B using standard polymerase chain reaction (PCR) techniques (World Health Organisation, 2011). All influenza A positive subjects and subset of all other subjects (50%) had their baseline serum sample tested against a panel of 16 influenza A strains (11 human, 5 avian strains) using a protein microarray technique previously described (Koopmans et al., 2011). Negative and influenza B controls were age and sex matched to the positive subjects included in longitudinal study. Baseline and follow-up sera were analysed for subjects recruited to the longitudinal study. Full methods described in Chapter 2.

5.2.3   Statistical Analyses

The primary outcome was diagnostic accuracy of serological markers of recent H3N2 infection. The gold standard for influenza diagnosis was PCR-confirmed influenza A H3N2 in nasal/throat samples. This included both single influenza A infections and co-infections with influenza B.

5.2.3.1   ROC Analysis of Threshold Values

To perform this analysis we compared H3N2 PCR positive individuals to H3N2 negative individuals in the longitudinal study who were either (i) PCR negative for both influenza A & B; ii) H1N1 PCR positive or iii) H3N2 PCR positive at recruitment but sampled after the time period of interest had elapsed. Three time periods of interest were explored for 'recent infection'; infection in the last i) 60 days, ii) 100 days or iii) 250 days. Analysis was performed on log2 transformed titres unless explicitly stated.

Where appropriate, analysis was stratified by age at recruitment. Age was preferentially used as a continuous variable, otherwise age was categorised as per recommendations from the Consortium for the Standardisation of Influenza Seroepidemiology (5-9, 10-19, 20-44, 45-65, 65+) (Van Kerkhove et al., 2013a).

Two methods of serological characterisation were investigated i) threshold cut-off using log2 titre to most recently circulating strain; ii) threshold cut-off using a modified Simpson's Diversity Index for within subtype (homosubtypic) microarray response. Analysis was performed for both threshold measures using area under the curve (AUC) of the receiver operating characteristic (ROC) curve or sensitivity and 1-specificity of gold standard (PCR positivity). Sensitivity and specificity was

calculated for different cut off thresholds varied by 0.25. 95% confidence intervals of the sensitivity and specificity was calculated using 2000 bootstrap replicates. Age stratified cut-offs using CONSISE recommended age categories was calculated. R package pROC was used for this analysis (Robin et al., 2011).

A Diversity Index is a method which gives a single value to summarise the overall magnitude and breadth of titre response across several strains. A modified Simpson's diversity index is used in this instance (Equation 2 from (Jost, 2006)). Within the diversity index equation there is a defined cut off value. This allows the distinction between "broad and low" and "broad and high" responses. This cut off is set to log titre 5.5 (non-transformed titre 452). This was selected based on results from Chapter 4 where most H3N2 strains had a visit two titre greater than this value in subjects infected with H3N2. The diversity index was calculated for both log transformed and non-transformed titres. R Code for the calculation of the Diversity Index is included in the Supplementary Appendix.

5.2.3.2   *Optimal Threshold Cut Off for Changing Prevalence*

Underlying influenza prevalence changes throughout the year meaning that positive and negative predictive values of any determinant will change as well. Two established measures used to determine optimal measures are the Youden's J Statistic and Closest-To-Top-Left. Both of these have incorporated weights for prevalence and cost of false negatives (Perkins and Schisterman, 2006).

Weights are added with

$$r = \frac{1 - prevalence}{cost * prevalence}$$

Youden's J Statistic (Youden, 1950) is the maximised distance to the identity (diagonal) line on ROC curve.

$$\max(sensitivities + r * specificities)$$

Closest to the top left is the point closest to the top left part of the ROC plot (i.e. point of perfect (100%) sensitivity/specificity).

$$\min(1 - sensitivities)^2 + r * (1 - specificities)^2)$$

The cost of false negative is set to 1. Optimal cut off thresholds are then calculated for a range of prevalence (0.5, 0.3, 0.1) for each age category. 95% credible intervals are calculated for the optimal

threshold and for the sensitivity and specificity using 2000 bootstrapped replicates. R package pROC was used for this analysis (Robin et al., 2011).

An alternative way to consider the question of optimal threshold is to look at how this will change across all possible values of prevalence (range 0-1). Because serology is imperfect we detect an apparent prevalence (or seroprevalence) (T+ = test positive) rather than the true prevalence (D+ = disease positive). The relationship between the two is demonstrated in:

$$P(T+) = P(T+|D+)P(D+) + P(T+|D-)[1 - P(D+)]$$

Which is the same as

$$P(T+) = Sensitivity * P(D+) + (1 - Specificity) * [1 - P(D+)] \qquad (1)$$

For each age group, sensitivity and specificity (95% CI) were calculated for threshold cut offs for infection in the last 60, 100 and 250 days. The median estimates of sensitivity and specificity were then used to calculate the apparent prevalence (i.e. seropositivity) for a given true prevalence using equation 1.

Up to this point, we have attempted to select a threshold value which means that P(T+) is closest to P(D+). Another way is the try and directly estimate P(D+). Equation 1 can be manipulated to give estimated true prevalence (P(d+)) (2):

$$P(d+) = \frac{P(T+) \quad + \quad P(T-|D-) - 1}{P(T+|D+) \quad + \quad P(T-|D-) - 1}$$

$$(2)$$

This is known as the Rogan-Gladen estimate (Rogan and Gladen, 1978) and can give a calculation of the true prevalence where the sensitivities and specificities are known. However, negative results can be produced where the apparent prevalence (P(T+)) measured is lower than the expected false positive rates. Sensitivity and specificity can be altered by factors such as laboratory differences, cross reaction (likely different at different ages in our cohort) meaning that they are unlikely to be fixed. One way to overcome this is to use a Bayesian approach which assumes an intrinsic distribution to sensitivity and specificity (Speybroeck et al., 2013). By incorporating this uncertainty a priori along with the observed seroprevalence (T+), we can obtain a probability distribution of the underlying prevalence (P(d+)).

For this analysis the estimated seroprevalence calculated using equation 1 were utilised (T+). The prior probabilities of the sensitivity and specificity were the 95% credible intervals calculated for a

threshold cut offs between log2 titre 5.5 to 7 for infection in last 60, 100 and 250 days. For each 'True Prevalence' we were interested in, a hypothetical 1000 samples were assumed. The seroprevalence for a given threshold was then calculated for the 1000 samples. OpenBUGS was used as the Bayesian modelling framework (Lunn et al., 2009) with the R2OpenBUGS package providing the interface for R programming (Sturtz et al., 2005). In OpenBUGS sampling 2 chains were started with a burn in of 1000 iterations with 6000 iteration used to calculated estimated true prevalence (P(d+)) with 95% credible interval. R Code is included in the Supplementary Appendix and is adapted from (Speybroeck et al., 2013).

All statistical analysis were performed using R Statistical Software v3.2.1(R Core Team, 2015). Packages used for analysis were ggplot2, pROC, R2OpenBUGS.

5.3   RESULTS

5.3.1   Data Description

As described in detail in Chapters 3 and 4, 953 subjects with ILI were recruited between 8 August 2013 and 31 May 2015 with 186 subjects in the longitudinal study. Serology samples were available for 686 subjects at baseline and seven month follow-up was available for 81 subjects (Table 5.1).

	Visit 1 (Day 1-3)	Visit 2 (Day 30)	Visit 3 (Day 90)	Visit 4 (Day 150)	Visit 5 (Day 210)
Influenza A H1N1	39	14 (100)	5 (35.7)	5 (35.7)	5 (35.7)
Influenza A H3N2	223	100 (100)	61 (61)	49 (49)	44 (44)
Influenza A NT	9	4 (100)	3 (75)	3 (75)	3 (75)
Influenza B	79	0	0	0	0
Influenza	336	66 (100)	51 (77.3)	41 (62.1)	29 (43.9)

Table 5.1 Number of Microarray Results Available before 31 May 2015. Percentage of follow-up subjects at each visit by infecting subtype in brackets.

The peak response post infection was approximately 30 days post infection. The titre to recent strains remained higher than baseline for the duration of the seven month follow-up.

5.3.2   Threshold Cut-off using Contemporary Strains

Using a single threshold cut off for all ages, the discriminatory power of titre thresholds below a titre of 640 (log2 titre of 6) was poor (Figure 48). There was a marked improvement as the test approached the limit of detection but the specificity for detecting infection in the last 100 days was still less than 70%. If the timing since infection was expanded to consider infection in the last 250 days, then the specificity improved with a compensatory drop in sensitivity.
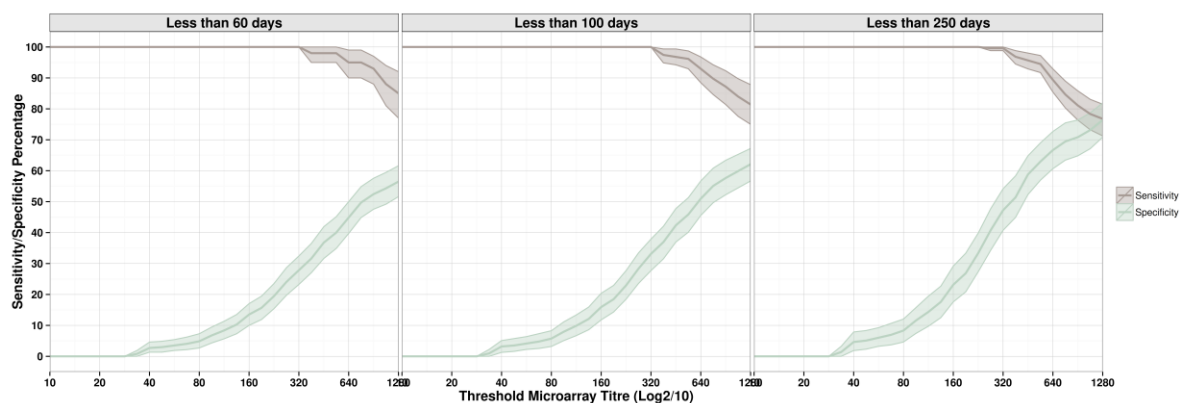
Figure 48 Sensitivity and Specificity for different threshold values for H3N2 2011. 95% CI calculated using 2000 bootstrapped replicates.

When the threshold analysis was performed for different age categories, it became clear that there are different sensitivities and specificities for different age groups (Figure 49). This is most marked in the 10 to 19 age group with high sensitivity but lower specificity even at the limits of detection of the assay.
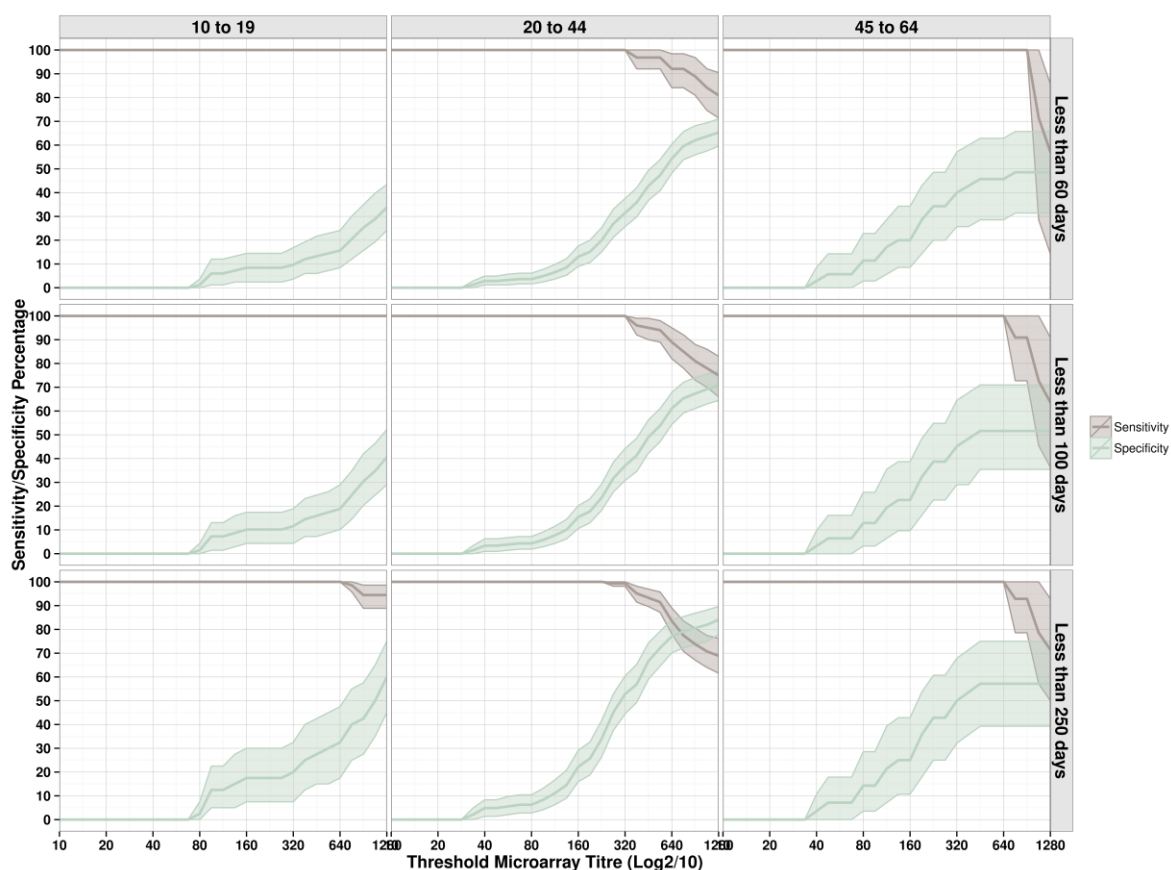
Figure 49 Sensitivity and Specificity for different threshold values for H3N2 2011 for different age categories. 95%CI calculated using 2000 bootstrapped replicates.

Estimates were made of the best threshold limit as measured by the Youden Index and Closest to Top Left (CtL) (Table 5.2). Comparing the two methods, the ideal thresholds identified by Youden

index have a higher sensitivity and lower specificity than the threshold identified by Closest to Top Left when all ages are combined and at all time points. This means that the Youden identified best thresholds will have fewer false negatives but more false positives compared to the Closest to the Top Left thresholds. When separate thresholds were identified for different age groups, Youden and CtL identified the same threshold if numbers of cases were smaller (10 to 19 and 45 to 64) and follow-up period was longer (250 days). The optimal threshold to identify recent infection was lower as individuals aged. A single threshold for all ages would result in more false positives for younger individuals but more false negatives in older individuals. This estimation was performed for a fixed prevalence of 50% which is much higher than the estimated annual attack rate for influenza.

Time since Infection	Age Group	Youden Index			Closest to Top Left		
		Threshold	Sensitivity	Specificity	Threshold	Sensitivity	Specificity
60 days	All Ages	6.51 (6.3-6.73)	0.94 (0.89-0.99)	0.53 (0.46-0.59)	7.18 (6.63-7.24)	0.88 (0.81-0.95)	0.56 (0.51-0.61)
	10 to 19	7.18 (7.16-7.18)	1 (1-1)	0.34 (0.24-0.45)	7.18 (7.16-7.18)	1 (1-1)	0.34 (0.24-0.45)
	20 to 44	6.45 (5.88-6.67)	0.92 (0.86-0.98)	0.62 (0.5-0.69)	6.63 (6.45-7.24)	0.89 (0.81-0.95)	0.64 (0.58-0.7)
	45 to 64	6.28 (5.75-6.44)	1 (1-1)	0.49 (0.31-0.66)	6.28 (5.75-6.44)	1 (1-1)	0.49 (0.31-0.66)
100 days	All Ages	6.33 (5.88-7.24)	0.92 (0.84-0.98)	0.56 (0.46-0.65)	7.18 (6.63-7.24)	0.84 (0.78-0.9)	0.61 (0.56-0.67)
	10 to 19	7.18 (7.16-7.18)	1 (1-1)	0.41 (0.29-0.52)	7.18 (7.16-7.18)	1 (1-1)	0.41 (0.29-0.52)
	20 to 44	5.88 (5.88-6.64)	0.93 (0.82-0.98)	0.62 (0.54-0.73)	6.63 (6-7.24)	0.83 (0.76-0.91)	0.69 (0.62-0.75)
	45 to 64	5.69 (5.52-6.03)	1 (1-1)	0.52 (0.35-0.68)	5.69 (5.41-6.03)	1 (1-1)	0.52 (0.32-0.71)
250 days	All Ages	5.87 (5.74-5.89)	0.94 (0.9-0.96)	0.67 (0.6-0.74)	6.64 (5.86-7.24)	0.82 (0.74-0.95)	0.73 (0.65-0.81)
	10 to 19	7.18 (7.11-7.18)	0.94 (0.89-0.99)	0.6 (0.45-0.75)	7.18 (7.16-7.18)	0.94 (0.89-0.99)	0.6 (0.45-0.75)
	20 to 44	5.87 (5.7-5.88)	0.9 (0.85-0.95)	0.77 (0.69-0.83)	5.87 (5.85-6.18)	0.9 (0.84-0.94)	0.78 (0.71-0.84)
	45 to 64	5.69 (5.47-6.07)	1 (1-1)	0.57 (0.39-0.75)	5.69 (5.47-6.07)	1 (1-1)	0.57 (0.39-0.75)

Table 5.2 Best Threshold Value for Recent Infection by Age Group by Youden Index or Closet to Top Left Estimates. Fixed prevalence of 50% used. 95% CI calculated using 2000 bootstrapped replicates.

When prevalence was reduced to 30% the optimal threshold for H3N2 2011 titre for all ages increased (Table 5.3). Only Closest to Top Left estimates are presented as the Youden index estimates failed to give credible intervals during fitting process. When prevalence was reduced to

10%, an optimal threshold for infection in previous 60 or 100 days is not identifiable. For infection in previous 250 days the optimal threshold was at the limit of detection on the microarray assay.

Time since Infection	Age Group	Prevalence 30%			Prevalence 10%		
		Threshold	Sensitivity	Specificity	Threshold	Sensitivity	Specificity
60 days	All Ages	7.24 (6.72-7.24)	0.85 (0.78-0.92)	0.56 (0.51-0.61)	Inf (Inf-Inf)	0 (0-0)	1 1-1
	10 to 19	Inf (7.16-Inf)	0 (0-1)	1 (0.35-1)	Inf (Inf-Inf)	0 (0-0)	1 (1-1)
	20 to 44	7.16 (6.62-7.24)	0.86 (0.76-0.94)	0.65 (0.59-0.71)	Inf (7.16-Inf)	0 (0-0.89)	1 (0.67-1)
	45 to 64	6.28 (5.75-Inf)	1 (0-1)	0.51 (0.37-1)	Inf (Inf-Inf)	0 (0-0)	1 (1-1)
100 days	All Ages	7.24 (6.65-7.24)	0.82 (0.76-0.88)	0.62 (0.57-0.67)	Inf (7.24-Inf)	0 (0-0.8)	1 (0.68-1)
	10 to 19	7.18 (7.16-Inf)	1 (0-1)	0.43 (0.35-1)	Inf (Inf-Inf)	0 (0-0)	1 (1-1)
	20 to 44	6.64 (6.43-7.24)	0.8 (0.71-0.88)	0.7 (0.64-0.76)	7.24 (6.63-Inf)	0.75 (0-0.85)	0.72 (0.68-1)
	45 to 64	5.69 (5.41-6.41)	1 (1-1)	0.52 (0.35-0.77)	Inf (5.69-Inf)	0 (0-1)	1 (0.68-1)
250 days	All Ages	7.24 (6.63-7.24)	0.77 (0.72-0.83)	0.76 (0.71-0.81)	7.24 (6.82-7.24)	0.77 (0.72-0.82)	0.76 (0.71-0.82)
	10 to 19	7.18 (7.16-7.18)	0.94 (0.89-0.99)	0.6 (0.45-0.75)	Inf (7.18-Inf)	0 (0-0.97)	1 (0.68-1)
	20 to 44	5.88 (5.85-7.24)	0.87 (0.7-0.93)	0.8 (0.73-0.87)	7.24 (6.18-7.24)	0.7 (0.62-0.82)	0.84 (0.78-0.9)
	45 to 64	5.69 (5.43-6.07)	1 (1-1)	0.57 (0.39-0.75)	Inf (5.59-Inf)	0 (0-1)	1 (0.68-1)

Table 5.3 Best Threshold Value for Recent Infection by Age Group by Closet to Top Left Estimates. Fixed prevalence of 30% and 10% used. 95% CI calculated using 2000 bootstrapped replicates.

From Table 5.2 & Table 5.3 the optimal threshold was never less than log titre 5.5 for all age groups. For each age group sensitivity and specificity was calculated for a threshold cut off of log2 titre 5.5 to 7 for infection in last 60, 100 and 250 days (Table 5.4). The median estimates of sensitivity and specificity were then used to calculate the apparent prevalence for a given true prevalence (Figure 50). As demonstrated in Figure 50, even when the threshold is at the limit of detection of the PMA (log2 titre =7) the number of false positives in the measured seroprevalence (T+) would lead to an overestimation of influenza attack rate where the true prevalence is less than 50%. If you extend the time period since infection to 250 days then the degree of overestimation is reduced. Younger individuals have a greater degree of over-estimation than individuals over the age of 20. As we estimate influenza to fluctuate between 0 and 30% prevalence in an inconsistent seasonal pattern,

the degree of overestimation at low prevalence and short outbreaks makes influenza surveillance challenging in tropical settings.

	Threshold	Less than 60 days		Less than 100 days		Less than 250	
		Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
All Ages	5.5	0.98 (0.95-1)	0.37 (0.32-0.42)	0.97 (0.94-0.99)	0.42 (0.37-0.48)	0.96 (0.93-0.98)	0.59 (0.53-0.65)
	6.0	0.95 (0.9-0.99)	0.45 (0.4-0.5)	0.93 (0.88-0.97)	0.51 (0.46-0.56)	0.89 (0.85-0.93)	0.67 (0.61-0.73)
	6.5	0.93 (0.88-0.97)	0.52 (0.47-0.58)	0.87 (0.81-0.92)	0.58 (0.52-0.63)	0.81 (0.76-0.86)	0.71 (0.65-0.77)
	7.0	0.85 (0.77-0.91)	0.56 (0.51-0.62)	0.81 (0.75-0.87)	0.62 (0.57-0.68)	0.77 (0.71-0.82)	0.76 (0.71-0.82)
10 to 19	5.5	1 (1-1)	0.13 (0.06-0.2)	1 (1-1)	0.16 (0.09-0.25)	1 (1-1)	0.28 (0.15-0.42)
	6.0	1 (1-1)	0.16 (0.08-0.24)	1 (1-1)	0.19 (0.1-0.29)	1 (1-1)	0.32 (0.2-0.48)
	6.5	1 (1-1)	0.25 (0.16-0.35)	1 (1-1)	0.3 (0.2-0.42)	0.94 (0.89-0.99)	0.42 (0.28-0.57)
	7.0	1 (1-1)	0.34 (0.23-0.45)	1 (1-1)	0.41 (0.29-0.52)	0.94 (0.89-0.99)	0.6 (0.45-0.75)
20 to 44	5.5	0.97 (0.92-1)	0.43 (0.37-0.49)	0.95 (0.9-0.99)	0.49 (0.42-0.56)	0.93 (0.9-0.97)	0.67 (0.59-0.74)
	6.0	0.92 (0.86-0.98)	0.54 (0.48-0.61)	0.89 (0.83-0.95)	0.61 (0.54-0.68)	0.84 (0.78-0.89)	0.77 (0.7-0.83)
	6.5	0.89 (0.81-0.97)	0.62 (0.56-...)	0.81 (0.73-0.89)	0.67 (0.61-0.74)	0.74 (0.68-0.8)	0.81 (0.74-0.87)
	7.0	0.81 (0.71-0.9)	0.65 (0.6-0.71)	0.75 (0.67-0.84)	0.71 (0.64-0.77)	0.69 (0.62-0.76)	0.84 (0.78-0.9)
45 to 64	5.5	1 (1-1)	0.46 (0.29-0.63)	1 (1-1)	0.52 (0.35-0.68)	1 (1-1)	0.57 (0.39-0.75)
	6.0	1 (1-1)	0.46 (0.29-0.63)	1 (1-1)	0.52 (0.35-0.68)	1 (1-1)	0.57 (0.39-0.75)
	6.5	1 (1-1)	0.49 (0.31-0.66)	0.91 (0.73-1)	0.52 (0.35-0.68)	0.93 (0.79-1)	0.57 (0.39-0.75)
	7.0	0.57 (0.29-0.86)	0.49 (0.31-0.66)	0.64 (0.36-0.91)	0.52 (0.35-0.68)	0.71 (0.5-0.93)	0.57 (0.39-0.75)

Table 5.4 Sensitivity and Specificity for Fixed Threshold Cut-Off for Recent H3N2 Infection using Titre to H3N2 2011. 95%CI from 2000 bootstrapped replicates.
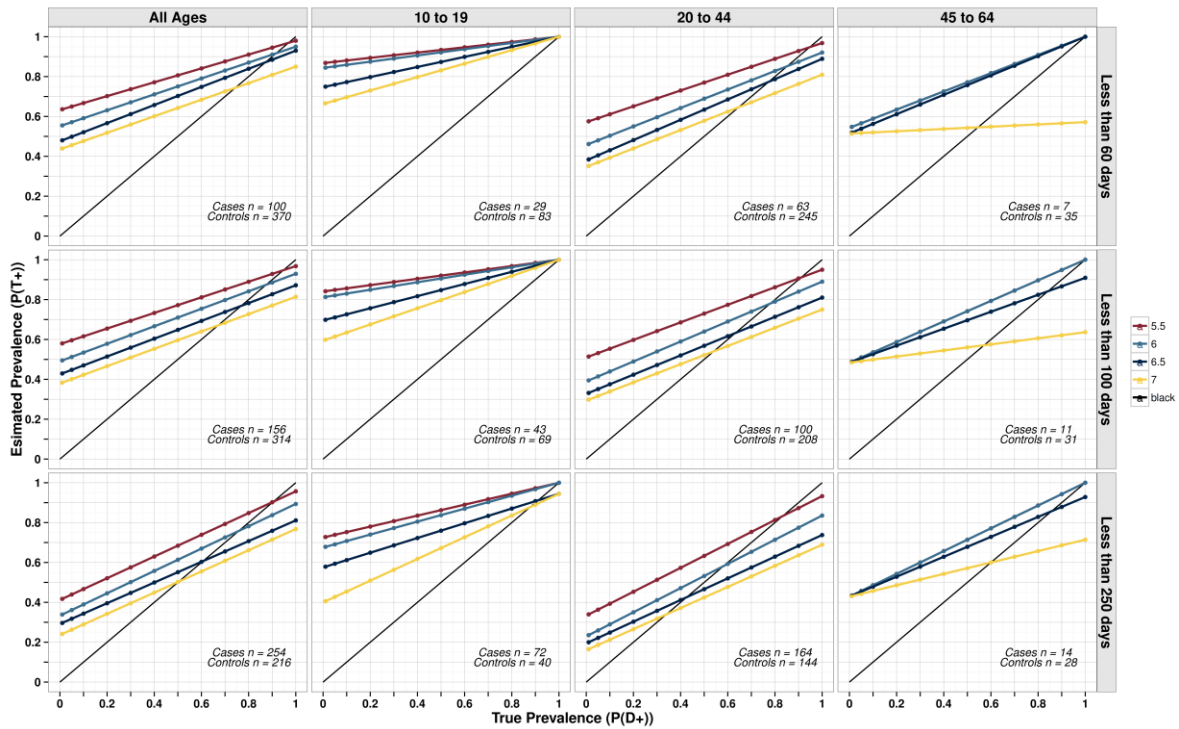
Figure 50 Estimated Seroprevalence for Underlying True Prevalence for given Sensitivity and Specificity using Threshold Cut-off of H3N2 2011 Titre. Number of cases and controls used in ROC curve estimation. Black line is the output a 'perfect test' would give.

Up to this point we have been trying to select a threshold value which means that P(T+) is closest to P(D+). Another way is to try and directly estimate P(D+) using the Rogan Gladen estimate of P(d+). For this analysis the estimated seroprevalence calculated for Figure 50 were used (T+). The prior probabilities of the sensitivity and specificity were the 95% CI presented in Table 5.4. For each 'True Prevalence,' a hypothetical 1000 samples were assumed to be collected. In OpenBUGS sampling 2 chains were started with a burn in of 1000 iterations with 6000 iterations used to calculated estimated true prevalence (P(d+)) with 95% CI (Figure 51).
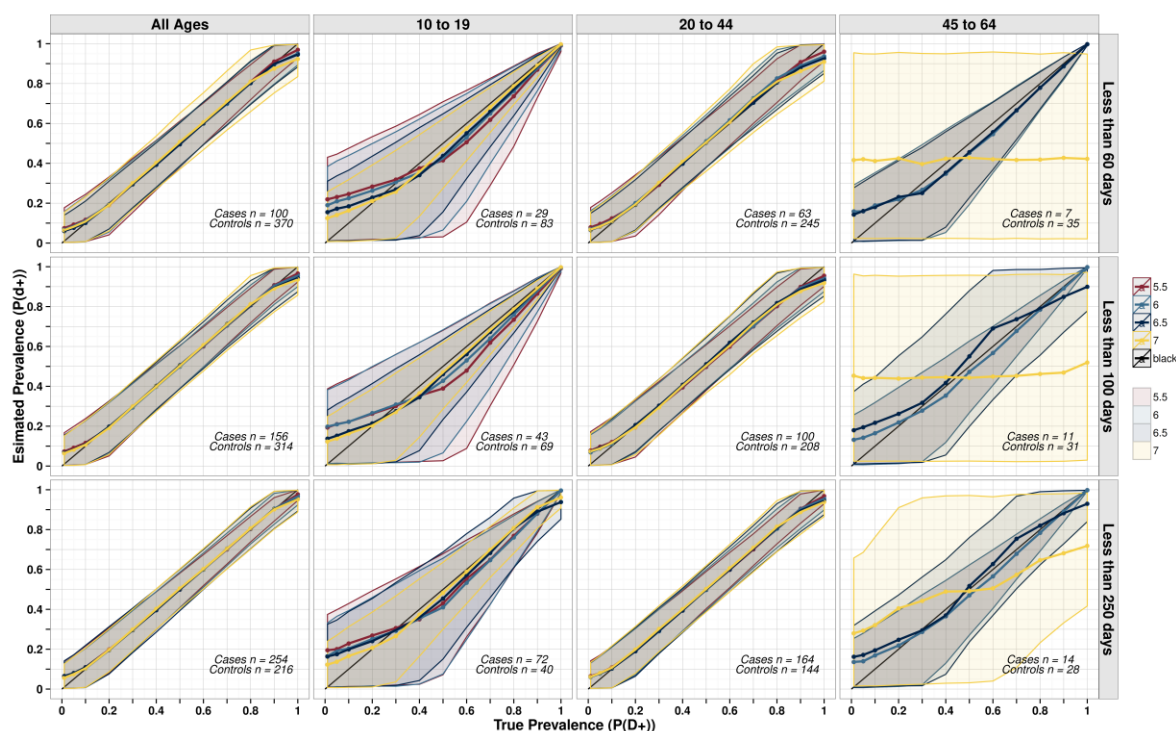
Figure 51 Bayesian Estimate of Prevalence for True Underlying Prevalence. Number of cases and controls used in ROC curve estimation of sensitivity and specificity presented. Black line is the output a 'perfect test' would give. Hypothetical 1000 samples taken at each true prevalence of interest (points). Ribbons are 95% CI for each threshold level. OpenBUGS burn in of 1000 iterations and 6000 iterations used.

The point estimates of prevalence (d+) are much improved (i.e. closer to D+) using this method compared to using uncorrected seroprevalence (T+).   Figure 51 shows that point estimates for all ages now run along the 'perfect test' line for true prevalence between 10 and 90%. At the extremes of prevalence, there continues to be an over or underestimate. When looking at ages separately, there are wider confidence intervals around the sensitivity and specificity in 10 to 19 and over 45's. It is possible that there is a greater variability in these groups (particularly in younger individuals), but these groups also had smaller sample sizes when performing ROC analysis which is likely to have contributed.

In addition, although the point estimates are much improved, the wide confidence intervals around these estimates may still lead to difficulty in interpretation. A median estimate of prevalence (P(d+)) 0.2 has a confidence interval of 0.1 to 0.3 meaning precise prevalence estimations are difficult.

5.3.3   Threshold Cut-off using Diversity Index

An alternative to using a single strain with a single threshold is to look at the overall response across several strains. As was discussed in Chapter 4, most strains had values greater than the current limit of detection on the microarray. When viewed as a heatmap, this shows a clear pattern of high response across all H3N2 strains for individuals recently infected (Figure 52). In contrast to using only a single strain, this "broad and high" response seems to be present for a relatively short period of time with the highest apparent frequency at visit two (~30 days post infection). One method to investigate this "broad and high" response is to use a diversity index which gives a single value to summarise the overall magnitude and breadth of titre response. A modified Simpson's diversity index is used in the instance calculated on titres to H3N2 strains only.
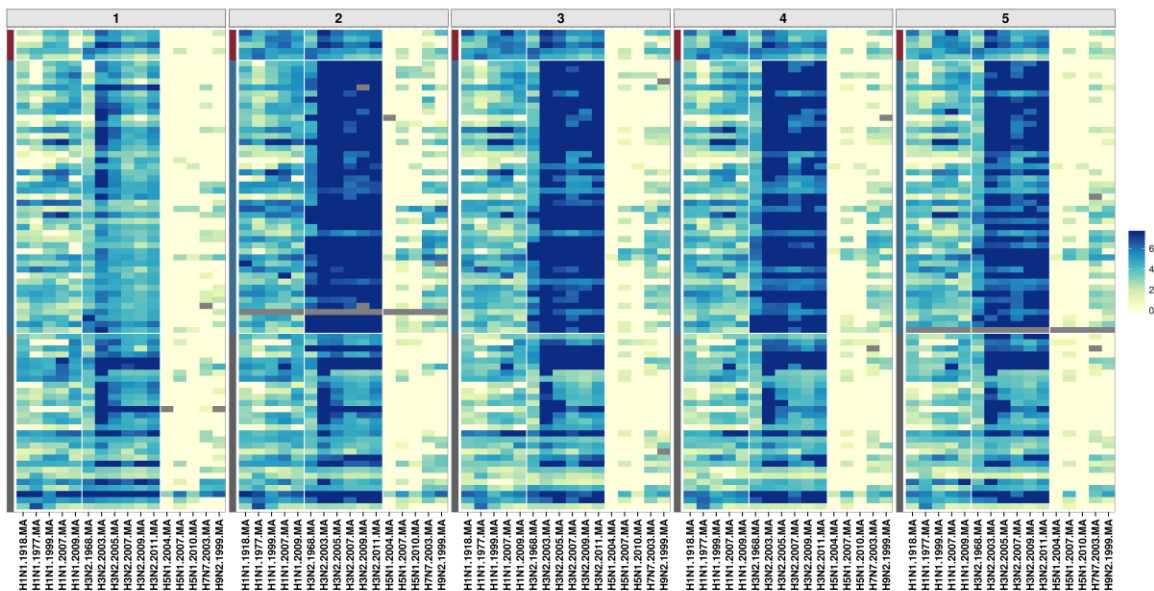
Figure 52 Heatmap of Log2 Titre by Visit Number. Each horizontal row represents the same subject, 79 subjects with follow-up to visit 5 included in figure. Coloured bar on left represents infecting subtype; Red=H1N1, Blue=H3N2, Grey= Negative. Within infecting subtype ordered by age.

Within the diversity index equation there is a defined cut off value. This allows the distinction between "broad and low" and "broad and high. This cut off is set to log titre 5.5 (non-transformed titre 452). This was selected based on results from Chapter 4 where most strains had a visit two titre greater than this value. When the logged titre values were used to calculate the diversity index, there was poor discrimination with median diversity index score of 5.8 and a narrow IQR of 5.5-5.9 across all visit numbers and infecting subtypes (Figure 53 – top panel). In contrast, if non transformed titres are used then a wider range of diversity scores occur with a median of 4.3 and wider IQR of 3.2-5.2 (Figure 53 – bottom panel). If the heatmap is replotted using non transformed titres, the pattern becomes very noticeable (Figure 54).
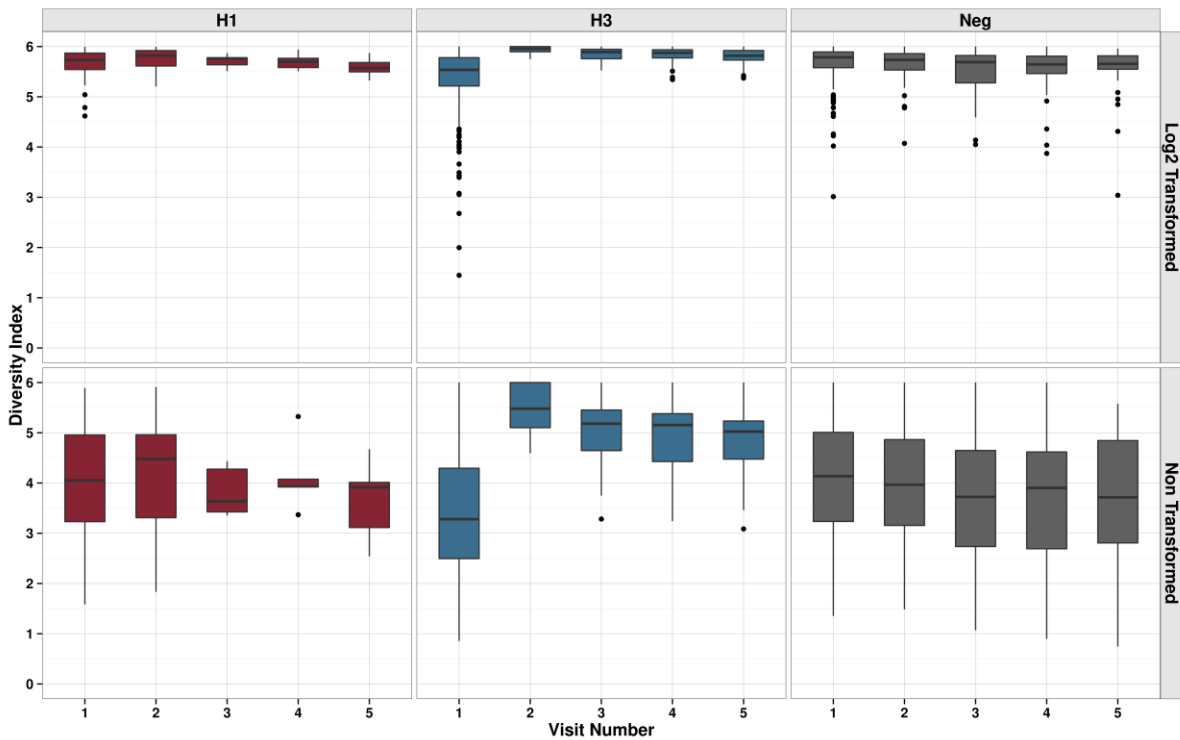
Figure 53 Boxplot of Diversity Index Scores by Infecting Subtype and Visit Number. Modified Simpson's Diversity Index shows diversity of H3 response only with cut-off of log2 5.5 or 452. Top panel with logged titres, bottom panel with non-transformed titres.
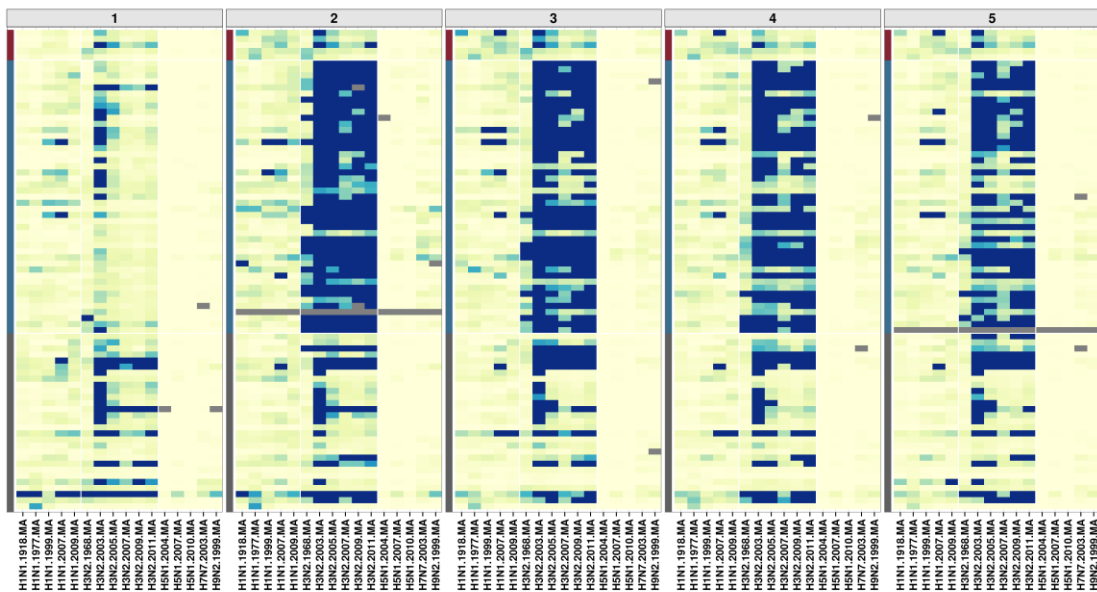
Figure 54 Heatmap of Non-Transformed Titre by Visit Number. Each horizontal row represents the same subject, 79 subjects with follow-up to visit 5 included in figure. Coloured bar on left represents infecting subtype; Red=H1N1, Blue=H3N2, Grey= Negative. Within infecting subtype ordered by age.

Compared to using a single strain threshold, using the diversity index gives an improved specificity (Figure 55) and is more consistent between age groups (Figure 56).
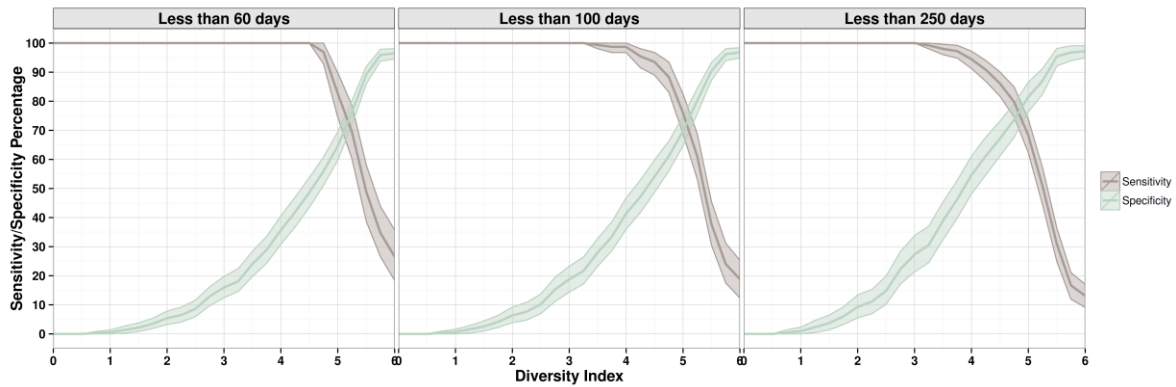
Figure 55 Sensitivity and Specificity for different threshold values for Modified Simpsons Diversity Index. 95% CI calculated using 2000 bootstrapped replicates.
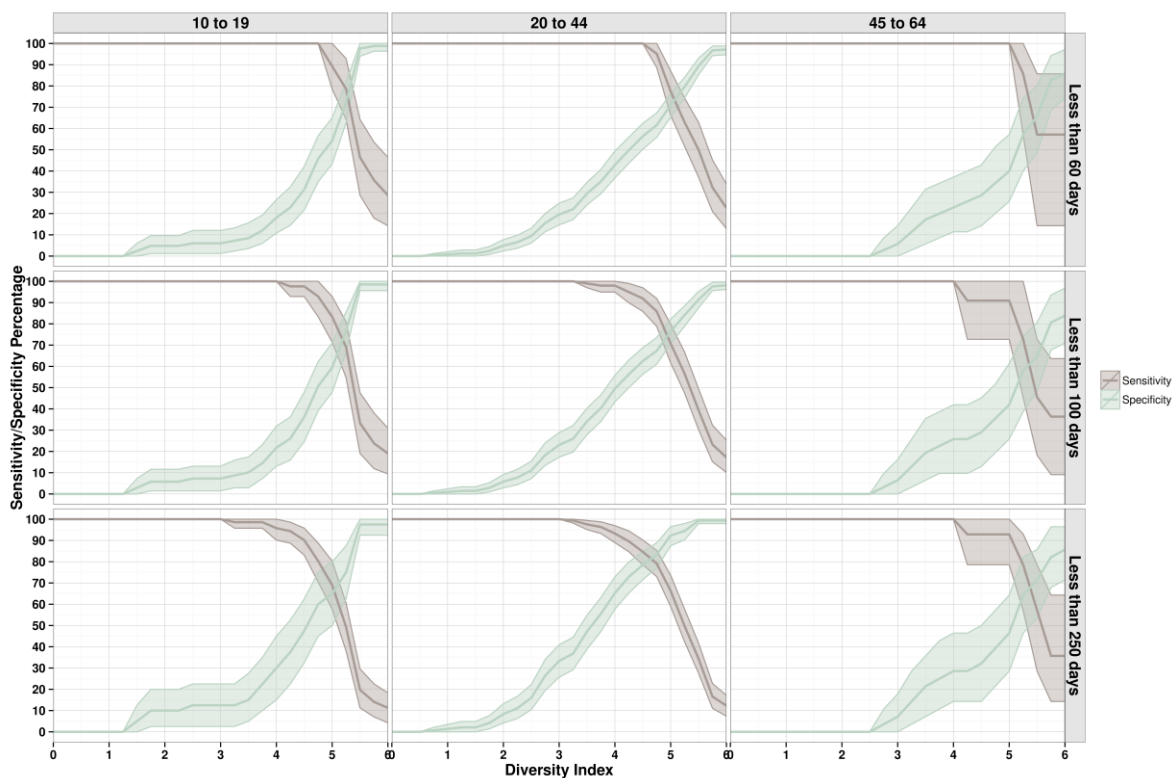
Figure 56 Sensitivity and Specificity for different threshold values for Modified Simpsons Diversity Index by Age Category. 95% CI calculated using 2000 bootstrapped replicates.

As before, for each age group sensitivity and specificity was calculated for a threshold cut off of Diversity Index 4.5 to 6 for infection in last 60, 100 and 250 days (Table 5.5). The median estimates of sensitivity and specificity were then used to calculate the apparent prevalence for a given true prevalence (Figure 57). Using the Diversity Index, the apparent prevalence (T+) is closer to the true prevalence than where a single strain threshold was used. There is a marked improvement in younger individuals when using a diversity index threshold greater than 5 with a switch to underestimation. The higher diversity index threshold is better for estimating recent infection (last 60 days); as the time since infection increases, this threshold will underestimate the number of cases and a lower diversity index is a better.

	Threshold	Less than 60 days		Less than 100 days		Less than 250	
		Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
All Ages	4.0	1 (1-1)	0.36 (0.31-0.4)	0.99 (0.97-1)	0.42 (0.36-0.47)	0.94 (0.92-0.97)	0.55 (0.48-0.61)
	4.5	1 (1-1)	0.48 (0.44-0.54)	0.94 (0.9-0.97)	0.54 (0.49-0.6)	0.86 (0.82-0.9)	0.67 (0.61-0.73)
	5.0	0.83 (0.74-0.9)	0.65 (0.59-0.69)	0.76 (0.69-0.82)	0.7 (0.65-0.75)	0.68 (0.62-0.74)	0.81 (0.76-0.87)
	5.5	0.49 (0.39-0.59)	0.89 (0.86-0.92)	0.38 (0.3-0.45)	0.9 (0.87-0.94)	0.31 (0.25-0.37)	0.95 (0.92-0.98)
	6.0	0.27 (0.18-0.36)	0.96 (0.95-0.98)	0.19 (0.13-0.25)	0.97 (0.95-0.99)	0.13 (0.09-0.17)	0.97 (0.95-0.99)
10 to 19	4.0	1 (1-1)	0.18 (0.1-0.27)	1 (1-1)	0.22 0.13-0.32)	0.96 (0.9-1)	0.3 (0.18-0.45)
	4.5	1 (1-1)	0.31 (0.22-0.41)	0.98 (0.93-1)	0.36 (0.25-0.48)	0.9 (0.83-0.96)	0.48 (0.32-0.62)
	5.0	0.89 (0.75-1)	0.54 (0.45-0.65)	0.83 (0.71-0.93)	0.59 (0.48-0.71)	0.69 (0.58-0.79)	0.65 (0.5-0.8)
	5.5	0.46 (0.29-0.64)	0.98 (0.94-1)	0.33 (0.19-0.48)	0.99 (0.96-1)	0.2 (0.11-0.3)	0.98 (0.92-1)
	6.0	0.29 (0.14-0.46)	0.99 (0.96-1)	0.19 (0.07-0.31)	0.99 (0.96-1)	0.11 (0.04-0.2)	0.98 (0.92-1)
20 to 44	4.0	1 (1-1)	0.43 (0.36-0.49)	0.98 (0.95-1)	0.5 (0.43-0.56)	0.93 (0.89-0.97)	0.65 (0.58-0.73)
	4.5	1 (1-1)	0.56 (0.5-0.63)	0.92 (0.86-0.97)	0.62 (0.56-0.69)	0.85 (0.79-0.9)	0.78 (0.72-0.85)
	5.0	0.77 (0.66-0.87)	0.71 (0.65-0.76)	0.71 (0.62-0.79)	0.76 (0.7-0.82)	0.66 (0.59-0.74)	0.92 (0.88-0.97)
	5.5	0.5 (0.37-0.63)	0.89 (0.86-0.93)	0.39 (0.3-0.48)	0.91 (0.87-0.95)	0.34 (0.27-0.42)	0.99 (0.98-1)
	6.0	0.23 (0.13-0.34)	0.97 (0.95-0.99)	0.17 (0.1-0.24)	0.98 (0.96-1)	0.12 (0.07-0.17)	0.99 (0.98-1)
45 to 64	4.0	1 (1-1)	0.23 (0.09-0.37)	1 (1-1)	0.26 (0.13-0.42)	1 (1-1)	0.29 (0.14-0.46)
	4.5	1 (1-1)	0.29 (0.14-0.43)	0.91 (0.73-1)	0.29 (0.13-0.45)	0.93 (0.79-1)	0.32 (0.14-0.5)
	5.0	1 (1-1)	0.4 (0.26-0.57)	0.91 (0.73-1)	0.42 (0.26-0.58)	0.93 0.79-1)	0.46 (0.29-0.64)
	5.5	0.57 (0.14-0.86)	0.66 (0.49-0.8)	0.45 (0.18-0.73)	0.65 (0.48-0.81)	0.57 (0.29-0.86)	0.71 (0.54-0.86)
	6.0	0.57 (0.14-0.86)	0.86 0.74-0.97)	0.36 (0.09-0.64)	0.84 (0.71-0.97)	0.36 (0.14-0.64)	0.86 (0.71-0.96)

Table 5.5 Sensitivity and Specificity for Fixed Threshold Cut-Off for Recent H3N2 Infection using Modified Simpson's Diversity Index. 95%CI from 2000 bootstrapped replicates.
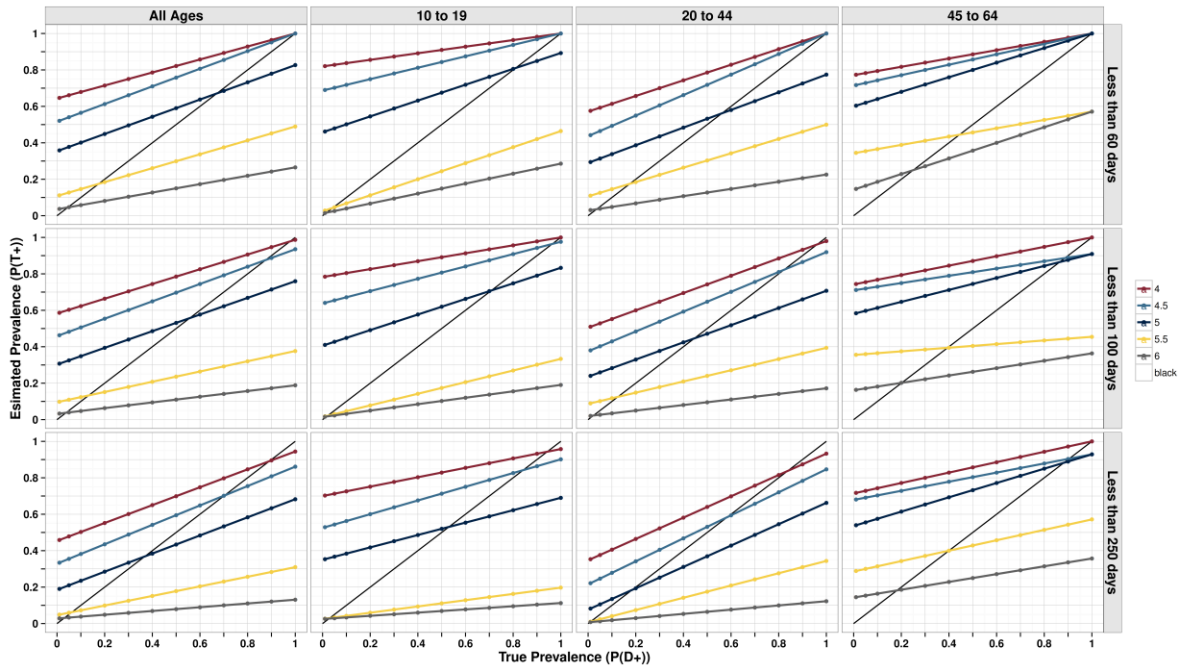
Figure 57 Estimated Seroprevalence for Underlying True Prevalence for given Sensitivity and Specificity using Modified Simpson's Diversity Index. Number of cases and controls used in ROC curve estimation. Black line is the output a 'perfect test' would give.
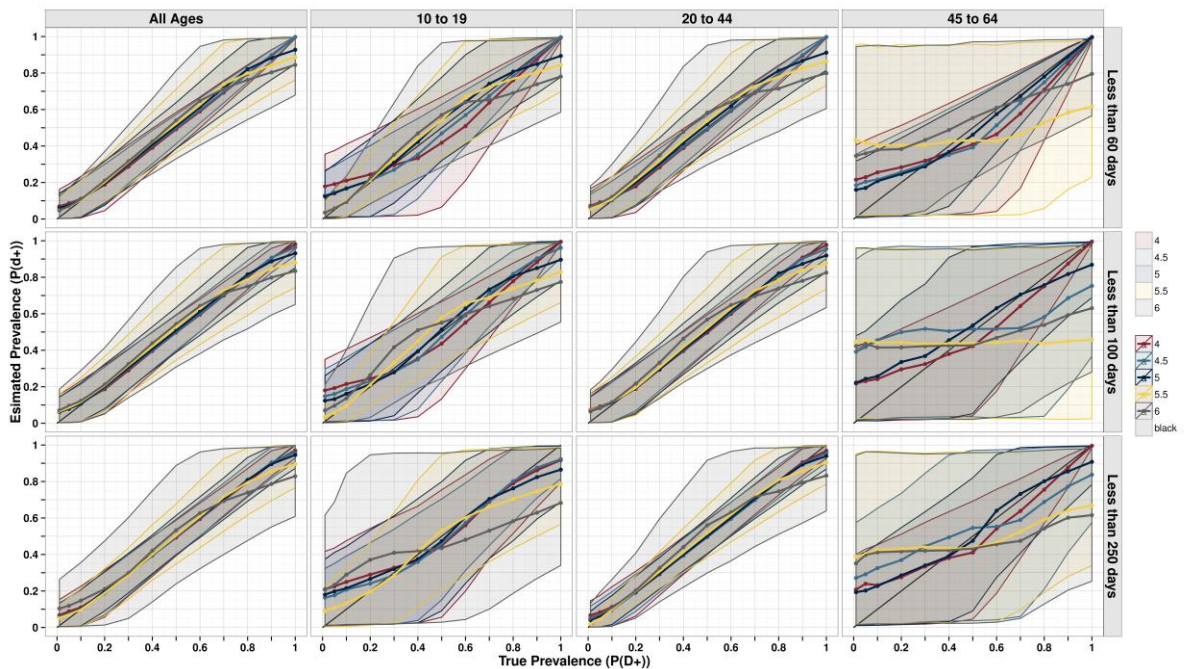
Figure 58 Bayesian Estimate of Prevalence for True Underlying Prevalence using Modified Simpson's Diversity Index. Number of cases and controls used in ROC curve estimation of sensitivity and specificity presented. Black line is the output a 'perfect test' would give. Hypothetical 1000 samples taken at each true prevalence of interest (points). Ribbons are 95% CI for each threshold level. OpenBUGS burn in of 1000 iterations and 6000 iterations used.

Using the estimated seroprevalence (T+) from Figure 57 and the 95% CI of the sensitivity and specificity presented in Table 5.5, the estimated true prevalence (P(d+)) with 95% CI were calculated using OpenBUGS (Figure 58). Again, the point estimates of prevalence are much improved using this

method compared to using only the estimated seroprevalence. Compared to using threshold to H3N2 2011 only, there is an improvement in point estimates for prevalence below 10% particularly for younger individuals. Again when looking at ages separately, there are wider confidence intervals around the sensitivity and specificity in 10 to 19 and over 45's. With high diversity index cut off (5.5 or 6) then the credible intervals are narrower at prevalence in the range that are likely to be observed in seasonal influenza (below 0.3).

5.4 DISCUSSION

One of the key purposes of seroepidemiology is the determination of disease incidence. Serology has been used for many years in influenza surveillance and has improved the understanding of disease attack rates through both seasonal epidemics and introductions of new viruses during pandemics. The increasing interest in understanding influenza in tropical settings with its different disease dynamics has brought new challenges.

The use of a single threshold for determination of acute infection is the current standard in influenza seroepidemiology. Traditionally an HI titre of greater than 40 is considered positive in cross sectional surveys. In surveys with paired samples, a fourfold rise between acute and convalescent samples are considered to be diagnostic (World Health Organisation, 2014a). The contradiction between these two definitions are evident, in paired samples the assumption of baseline titre value of 10 (i.e. no detectable titre) is not present. A baseline titre of greater than 10 in HI could be possible due to a cross-reaction with older responses. The duration of response is also important to consider in seroepidemiology. Once infected, titres can stay significantly elevated from many months and potentially years. If the titre threshold value for seroepidemiology is low, we will get a high sensitivity for infection at any time but poor specificity for infection in a recent period. In temperate countries this phenomenon can be obviated by doing a single cross sectional study at the end of the relatively short influenza season and, assuming there is a good match between infecting and testing strain, could give a reasonable season attack rate (Laurie et al., 2012, Fragaszy et al., 2015). In tropical settings where seasonality is less pronounced, this could lead to difficulties in timing and interpretation of cross sectional sero-studies. The development of a novel microarray allows a reassessment of these methods to optimise serological determinants of acute infection and how this can be applied in settings with less pronounced seasonal epidemics.

Defining an optimal titre cut-off to the most recent circulating strain on the microarray is an attempt to replicate the traditional 1:40 cut off using HI testing. The PMA is known to be more sensitive than HI and a PMA titre is higher than a corresponding HI test (Koopmans et al., 2011). Small studies in-

house and at the population level have suggested that an HI of 1:40 for H1N1 2009 corresponds to a PMA of approximately 1:100 (log2 3.3) (Boni, unpublished data; (te Beest et al., 2014a)). As measured by the Youden or Closest to the Top Left methods, the optimal threshold for recent infection on PMA is considerably higher than would be expected given the HI threshold of 1:40 (Table 5.2). The sensitivity of these thresholds are better but have a poorer specificity than those reported for HI cut off of 1:40 in the post-pandemic period with PCR confirmed H1N1 (Veguilla et al., 2011). Age specific differences in sensitivity and specificity were also identified in this study. In another study in the Netherlands, sensitivity and specificity of HI were much lower when estimated at the population level in the immediate pre and post emergence of H1N1 2009pdm but were improved using a PMA cut off of ~ 1:100 (te Beest et al., 2014a).

In contrast to the published estimates of sensitivity and specificity, this study looked at the threshold for H3N2, which has been circulating for over 40 years, rather than to a newly emerged subtype. Most individuals in this study have been exposed to H3N2 previously and have baseline values of H3N2 2011 greater than the 1:40 equivalent of 1:100 (see Chapter 3). This cross reactivity due to previous infection is overcome by increasing the threshold value and explains why sensitivity can be preserved or improved despite the presence of positive baseline titres. Specificity improves as the time since infection get longer when using the titre response to H3N2 2011. Where very few individuals have ever been exposed to a disease and therefore have undetectable or very low titres then specificity is improved. Where individuals have been exposed at some point in the past they will have detectable titres which then wane. Individuals were categorised as negative for H3N2 if they had non-influenza ILI or influenza A H1N1 at recruitment. We do not know when their last infection was prior to study recruitment which could have affected the specificity. Younger individuals have poorer specificity than older individuals even at the limit of detection. The presence of 'original antigenic sin' means that these individuals have higher titres to antigenically similar viruses which could explain the poorer specificity.

The use of multiple strains in the definition of a threshold cut off markedly improves specificity for all ages in this analysis. The use of a diversity index to look for a high titre across multiple strains takes into account the antibody dynamics which were observed in Chapter 4. The use of a weighting threshold gave a higher diversity score where all titres were above a certain value (log2 5.5 in this analysis) allowing a higher score for 'high and broad' microarray responses. This method has been explored to look for cross reaction following H1N1 2009 (Freidl et al., 2016) but this is the first time it has been used as a diagnostic marker in influenza. Further work is planned to investigate the optimal weighting value and the optimal number of strains to be included in the calculation of the index. Other methods that have been used to look at the inclusion of multiple strains for diagnosis include

a ratio of H1N1 2009 to 1918 which improved specificity compared to a single strain in the post-pandemic period (te Beest et al., 2014a). Microarrays have been used for diagnosis of non-influenza infectious disease and methods utilised have included k-means clustering and support vector machine learning which are standard methods for genomic microarray analysis (Sundaresh et al., 2006, Sundaresh et al., 2007, Felgner et al., 2009, Crompton et al., 2010). These approaches are now being applied to large scale population sero-surveys for malaria (King et al., 2015). It is planned to explore these more complicated methods to look for classification of recent subtype specific infection and susceptibility and apply this to the related population serum bank for central and southern Vietnam.

One challenge in serosurveillance is the continual evolution of the virus which can result in a mismatch between the infecting strain and testing strain. This can occur when there is antigenic drift within seasons but also in longer term studies over a number of seasons and batch testing. If using a titre threshold to a single strain, it is likely that the sensitivity and specificity of this will change as the virus evolves and these changes need to be taken into account when using newer serological markers. The advantage of the diversity index approach is that this pattern is likely to be present regardless of the match between the most recent strain on the assay and the infecting strain. It also leads to a more consistent sensitivity and specificity between different age groups. Age in the multiple strain approach is less important than in single strain as there is similarity in broad cross reactivity in the short term after infection leading to a more consistent sensitivity and specificity. This method is also affected less by presence of censored data than a single strain approach which may allow fewer dilutions to be performed.

The time since infection is an important consideration when performing surveillance for influenza as the assumption of a single infection during a lifetime, which underpins most other serosurveillance methods, does not hold true (Hens et al., 2012). The standard serological cut-off of 1:40 assumes infection at any time, which is not useful for estimating the force of infection of non-pandemic influenza using cross-sectional data (Lessler, 2014). There remain unanswered questions regarding the optimal designs of seroepidemiology for seasonal influenza (Van Kerkhove et al., 2013a). One particular question is what measure of incidence are we interested in and are we able to measure it. Age specific population attack rate is important to determine public health impact of influenza and plan interventions such as vaccination. The other use of seroepidemiology is to understand the influenza transmission dynamics at a population level including those with mild or asymptomatic disease. The second of these aims will require a finer time resolution than the first, i.e. a six month age specific attack rate is a useful figure but does not give information on transmission dynamics. This is of particular importance in tropical settings with less predictable seasonality. The use of

multiple strains appears to provide better discrimination between infections in the last 60 days. Further exploration of this is required as well as exploring if a probability of infection in the last 60 or 100 days can be determined.

Seroepidemiology in an environment without strong seasonal forcing is likely to be compromised by the lack of a perfect test. Few studies have adjusted for the sensitivity and specificity of influenza serology (Cox et al., 2011, Lee et al., 2011, Reed et al., 2012) and all of these used a fixed sensitivity and specificity value. The assumption of fixed sensitivity and specificity is likely not to hold, especially in the situation of a drifted seasonal virus which individuals have been exposed to multiple times. None of the studies where adjustment has been performed considered the changing underlying prevalence on positive and negative predictive values of the test and how this would impact the confidence intervals of the predictions. Our results demonstrate a considerable overestimation of prevalence if adjustments for sensitivity and specificity have not been made. This is particularly true in younger individuals even six months post infection. The use of a multi-strain approach with its improved specificity reduces this overestimation in all age groups. Bayesian techniques offer a method to improve these estimations but with the wide confidence intervals will require further refinement before they can be widely implemented.

The work presented in this thesis also raises important questions around how to identify the optimal measure of acute influenza infection using serological methods. To date a cut off of HI 1:40 has been used for all ages, in temperate and tropical settings, for both seasonal and pandemic strains, for all influenza A subtypes. The results presented in this thesis supports other published work which shows this simplified approach to seroepidemiology in influenza is in need of review and requires to be tailored to the specific questions that are being asked of the methodology (Van Kerkhove et al., 2013a). In the pandemic setting, detection of age specific attack rates is the priority and several papers suggest that a titre threshold of 1:20 is more appropriate to estimate the population attack rate (Cauchemez et al., 2012, Wu et al., 2014). The relative lack of cross reactivity in the initial waves after the emergence of a new strain means that a lack of specificity due to previous infection is not a problem in the way it is for a seasonal strains. For seasonal strains, cross reactivity and serological response to previous infection becomes an issue and here specificity of any marker of acute infection is likely to be at least as important as sensitivity, if not more so. Existing methods for defining the optimal threshold such as Youden's and Closest to the Top Left minimise the number of false negatives (Perkins and Schisterman, 2006). This is because these methods were developed for cancer screening tests where false negatives need to be avoided even at the expense of a large number of false positives. In a surveillance setting rather than screening, the goal is to get as close to 'true prevalence (D+)' as possible and have a diagnostic test which offers the best accuracy over the

range of likely prevalence (Banoo et al., 2010). For influenza this is likely to range between 0 and 30% prevalence for seasonal strains. In addition to this we need to use the information we possess on test accuracy by incorporating sensitivity and specificity into estimates. The Rogan-Gladen adjusted estimates of seroprevalence are used more commonly within veterinary epidemiology (Lewis and Torgerson, 2012, Brooks-Pollock et al., 2013). One reason for this difference between human and animal epidemiology could be that false positives can have a financial implication if culling is employed for disease control. With developments in computing and statistical software, the use of simple Bayesian models becomes much easier and within the grasp of those working in influenza epidemiology.

There are a number of limitations to this study. This work has been restricted to measure of acute infection with influenza A H3N2 due to the small number of subjects with PCR confirmed H1N1 within the longitudinal study. The mixed H3N2 and H1N1 wave detected in May 2015 carried on for two months and there were additional H1N1 positive subjects in the longitudinal portion of the study. Further analysis is planned to detect the optimal threshold for the most recent strain on the PMA as well as looking at the multi-strain response in a recently emerged strain (although no longer strictly a pandemic strain). No follow-up was conducted on subjects with influenza B as this study was designed to assess influenza A seroepidemiology. Subjects infected with influenza B had the highest mean titres to H3N2 at baseline and further work is planned to see if these would be classified as recent infection using the markers defined in this Chapter. The interaction between subtypes and influenza A and B still needs to be fully understood particularly in tropical settings where influenza B seems to be a more significant cause of disease than had previously been recognised (Caini et al., 2015). One of the benefits of seroepidemiology is that it is thought to give a more accurate depiction of population attack rate because it includes individuals with mild or asymptomatic disease who do not attend healthcare and are therefore missed by syndromic surveillance techniques. Although this study specifically recruited outpatients with mild disease these individuals still presented to healthcare setting for assessment. Antibody response in those with very mild or asymptomatic infection is likely to be lower than those who have symptoms severe enough to present to healthcare. It is important to consider this in any future marker of acute infection. Household studies may elicit further information on antibody response in those with asymptomatic shedding or very mild disease.

The way forward for this work will be to apply markers of acute infection to the related population serum bank which has been collected since 2009 in central and southern Vietnam. This would allow an assessment of the markers defined in a real life setting and should be compared to the results

from national surveillance system within Vietnam. In particular to look at whether the markers of acute influenza infection demonstrably give a measure of transmission dynamics in a tropical setting.

6 DISCUSSION

6.1 INTRODUCTION

As discussed in chapter 1, global influenza circulation is a complex interplay between viral evolution, population level immunity and environmental pressures which results in multiple infections during an individual's lifetime (Ferguson et al., 2003). Despite being studied for over 70 years this interplay is still not fully understood. Influenza evolution occurs through a series of 'shifts' and 'drifts', with gradual antigenic changes occurring each year and larger jumps occurring every three to five years (Smith et al., 2004, Bedford et al., 2014). In addition to these changes, the sporadic emergence of novel subtypes occur with a subsequent rapid spread globally. These pandemics are known to have occurred five times in the last 100 years and each time the new virus has become the predominant circulating virus in the human population (Webster, 2013). It is important to consider the differences between these forms of influenza, generally termed seasonal and pandemic respectively, particularly with regard to limiting the clinical impact of existing and novel strains.

Serological surveys are a long standing method of investigating infectious disease epidemiology with two major goals i) determination of disease incidence and ii) assessment of susceptibility to disease (Van Kerkhove et al., 2013a). However, many of the existing statistical methods rely on assumptions of life-long immunity following infection which does not hold true in influenza in non-pandemic situations (Hens et al., 2012).

This thesis presents the results of a prospective observational study of non-severe influenza like illness (ILI) in a tropical, urban primary care setting. This study was designed to address questions around the levels of protection against influenza A infection offered by pre-existing antibodies, understand the short term antibody dynamics of both recent and historical strains post infection and identify optimal serological determinants of recent infection.

These results will contribute to a wider body of work which aims to understand the impact on the humoral immune system of multiple influenza infections during an individual's lifetime and how this might impact on susceptibility and transmission of influenza at the population level. A pre-existing population level seroepidemiology study has been collecting residual serum samples from hospitals across central and southern Vietnam since 2009 and had a bank of over 40,000 samples at the end of 2015. These samples can be related to simplified demographic information but no information on recent infection or clinical status was available. This dataset provided rich spatio-temporal data on the changing antibody titre measurements at population level which has been used to investigate

the introduction and spread of H1N1 2009 in Vietnam (manuscript in preparation) and assessment of potential exposure to avian strains at the time of H7N9 emergence in southern China (Boni et al., 2013, Todd et al., 2014). The study presented in this thesis was developed to answer questions around individual level susceptibility to influenza in a non-pandemic period. This collected more detailed clinical and demographic information than the population study in the context of known current infection status and allowed short term follow-up of antibody response for individuals infected with influenza A and non-influenza ILI. Together these research studies offer a rare opportunity to improve our understanding of the interaction between individual and population measures of immunity and how this influenza transmission at the population level..

6.2   Key Findings in Context

Many of the observations in this thesis are consistent with and corroborate the existing influenza literature. Year round transmission is one of the hallmarks of influenza in tropical regions (Nguyen et al., 2007, Li et al., 2008, Nguyen et al., 2009, Moura, 2010, Tamerius et al., 2011, Horby et al., 2012) and this study supports this finding with influenza identified in over 80% of study weeks. However, it is important to note that this was not persistence of the same influenza type or all types simultaneously but a mixture of high intensity peaks of single subtypes and co-circulation of types and subtypes at variable intensities. Influenza is traditionally thought of as a single entity; however, this is likely to be a gross simplification and understanding the interactions between the different types and subtypes is vital. The presence of long lasting antibodies to strains encountered decades prior to testing has long been recognised (Davenport et al., 1953, Lessler et al., 2012, Miller et al., 2013). The results of this study have again demonstrated an age specific effect with the highest titres being to those strains an individual was exposed to in the first decade of life. The mechanism behind this 'original antigenic sin' or 'antigenic seniority' remains elusive. One hypothesis is the repeat boosting of strain responses after acute infection. The finding of short term boosting of historic responses after infection was identified in this and other recent multi-strain studies (Lerman et al., 1980, Miller et al., 2013, Fonville et al., 2014) and has been incorporated into models estimating the life course of infection with reasonable success (Kucharski et al., 2015). An alternative is the concept of 'antigenic trapping' where pre-existing antibodies reduce the viral (and therefore antigenic) load, meaning the new humoral response to the infecting strain is smaller (Hennessy et al., 1955). In this study, there appeared to be a higher fold change of titre to the most recent strain for individuals with very low titres at baseline. As these individuals were mostly younger, this could represent the antigenic trapping phenomenon but this result was highly confounded by the upper limit of detection on the microarray.

Some unexpected results were found during this analysis. That recent but not most recent strains were predictive of susceptibility to infection with H3N2 is consistent with the idea that an adult is infected once or twice each decade (Kucharski et al., 2015). It also suggests that the boosting of historic responses may not translate to an increase in neutralisation titre to drifted strains such as H3N2 but could provide protection against the emergence of novel strains such as H1N1 2009 (Van Kerkhove et al., 2013b). Another surprising finding was the extent to which protection offered by a fixed titre measurement changes with the force of infection. Although intuitively this finding makes sense there has been little work which explored it in detail (Gomes et al., 2014, Cao et al., 2015, Laurie et al., 2015b). Another intuitive finding, which has been overlooked, is the degree to which prevalence is overestimated if the imperfect nature of serology is not corrected for (Cox et al., 2011, Lee et al., 2011, Reed et al., 2012). This is the first time, to my knowledge, that an exploration of this nature has been made to quantify the degree of uncertainty around influenza seroprevalence and establish an optimal measure, based on the underlying likely prevalence.

6.3 IMPLICATIONS FOR PRACTICE

The two key messages from this thesis are related to the conduct and interpretation of seroepidemiology studies in influenza; i) surveillance methods developed in temperate settings cannot be directly moved to tropical settings without consideration; ii) surveillance and testing methodologies should be tailored specifically for pandemic or seasonal strains. This is a major shift from existing approaches which have traditionally used a universal approach for all settings. These findings have important implications for influenza public health policy and research.

6.3.1 Surveillance methods developed in temperate settings cannot be directly moved to tropical settings without consideration

Influenza is a global disease with differing behaviour in temperate and tropical settings (Moura, 2010). The year round persistence of all types of influenza rather than predictable winter peaks is the most significant difference. In considering the first goal of seroepidemiology, to determine disease incidence, year round presence of some type of influenza will lead to difficulty. Following infection, titre rise to most recent circulating strains persists out to seven months meaning that fine resolution of seroprevalence in cross-sectional studies is likely to prove difficult if relying on a response to most recent strains alone. The presence of heterosubtypic boosting within influenza A and the possibility of influenza B infection following influenza A are additional complicating factors. In temperate countries a single cross sectional study at the end of the relatively short influenza season could give a reasonable season attack rate estimate, assuming there is a good match

between infecting and testing strains. In tropical settings where seasonality is less pronounced, this could lead to difficulties in timing and interpretation of cross sectional sero-studies.

The other goal of seroepidemiology is to assess the susceptibility to disease in the population. The complex transmission dynamics in tropical countries highlights the question of how population immunity may be shaping these transmission dynamics. In temperate countries there is temporal clustering of infection where people are infected within a relatively short season. As waning of titre (and therefore immunity) appears to occur at a reasonably consistent rate between individuals, this may lead to a synchronicity in time to becoming susceptible again after infection. In tropical countries there is less clustering of infection in time and more low level transmission of different subtypes through the year, which could lead to greater variability in the population level waning of immunity. If the heterosubtypic boosting demonstrated in chapter 4 corresponds to a boost in neutralisation activity then this could also impact on the dynamics of population susceptibility. The other major finding with regard to susceptibility in this study is that levels of protection offered by a specific titre vary depending on the force of infection. If this observation is related to an antigenic change in the predominant H3N2 strain then it accentuates the need to match the serological assessment of susceptibility with the genetic and antigenic evolution of virus. If this finding is not due to mismatch between infecting and testing strains, then it poses interesting questions around changing population susceptibility through the year. Again this is likely to be a more significant problem in tropical countries where there is more variation in the incidence of influenza infection through the year rather than the condensed intense peaks of transmission in temperate regions. The longer periods of low transmission in tropical countries (rather than no transmission in temperate countries) could lead to variability in which portions of the population are susceptible at different times throughout the year.

Changes occurring across the tropics will pose new challenges in influenza surveillance. As seen in this study, vaccination is increasingly being used in the private sector in Vietnam, although in small numbers, and this will start to change the population immune landscape that has arisen due to a lifetime of natural infection. The challenge of incorporating multiple healthcare providers is likely to become more important in future public health surveillance planning

6.3.2   Surveillance and testing methodologies should be tailored specifically for pandemic and seasonal strains

Just as methods developed in temperate settings need to be introduced with due consideration to tropics, pandemic and seasonal influenza should also be treated differently. The H1N1 pandemic strain which emerged from Mexico in 2009 has replaced previously circulating H1N1 seasonal strain

and now can be considered to be a seasonal influenza type. However, the results presented in this thesis suggests that, even six years after its emergence, there still appear to be individuals encountering their first infection with this strain. This was a surprising finding and it is unclear if this is related to cross reaction of the assay with pre-existing H1N1 antibodies related to strains circulating between 1977 and 2009 (rather than pre 1957 H1N1 strains). The small number of H1N1 infections present in the study meant that it was not possible to perform the planned analysis presented in chapters 4 and 5 for H1N1. The study has continued to collect data and this should allow additional analysis of this finding.

Given the primary aim of seroepidemiology is to determine disease incidence, in the pandemic setting detection of age specific attack rates is a priority. Several papers have suggested that an HI titre threshold of 1:20 is more appropriate than the traditional 1:40 to estimate the population attack rate in pandemic periods (Cauchemez et al., 2012, Wu et al., 2014). The relative lack of cross reactivity in the initial waves after the emergence of a new strain means that a lack of specificity due to pre-existing titres, from previous infection, is not a problem in the way it is for seasonal strains and therefore a lower threshold with a higher sensitivity is optimal. For seasonal strains, cross reactivity and serological response to previous infection becomes an issue and here specificity of any marker of acute infection is likely to be at least as important as sensitivity. In order to determine disease incidence, the goal is to get as close to true prevalence as possible and have a diagnostic test which offers the best accuracy over the range of likely prevalence. This means that for seasonal strains, a higher threshold and likely one that takes into account cross reaction is required. The results presented in this thesis suggests a multi-strain diagnostic is likely to give a better estimate of disease incidence that titre to a single strain.

Serological surveys have also been performed to attempt to estimate true attack rates for non-human strains which are spilling over from animal and avian reservoirs to cause sporadic infections in humans (Wang et al., 2012, Liu et al., 2014). Results from this and other studies have suggested that cross reaction from repeat human infections could lead to cross reaction with avian strains on serological testing (Boni et al., 2013, Molesti et al., 2014, Oshansky et al., 2014, Todd et al., 2014). This has important implications both for assessment of prevalence but also for consideration of population susceptibility. The need for adjustment for imperfect test, raised in Chapter 5, applies to this type of serosurveillance and offers a new dimension to the debate on the 'missing millions' of H5N1 cases (Osterholm and Kelley, 2012, Palese and Wang, 2012a, Palese and Wang, 2012b, Van Kerkhove et al., 2012, Wang et al., 2012). With regard to population susceptibility, further work is required to assess whether this detectable cross reaction to avian strains corresponds to neutralisation activity. If initial results are replicated and this detectable cross-reactivity also reflects

cross-immunity (Henry Dunand et al., 2015) then this finding could be used to assess for the likelihood of non-human strains moving from intermittent spill over and small scale outbreaks to global pandemics.

Finally, where new technologies are developed in post-pandemic periods for pandemic strains their use and relevance should be assessed before translation to seasonal strains. As demonstrated in this study, the upper limit of detection of 1280 is likely to be adequate for H1N1 with only 20% of subjects infected with H1N1 having values greater than this limit of detection one month after infection. In contrast, up to 89% of H3N2 titres are above the limit of detection and this persists through the seven months of follow-up. Global public health emergencies often provide the impetus for the development of novel diagnostics which are then used in the post pandemic phase. Serological diagnostics developed in a population with limited immunity are likely to perform differently in a population where they are measuring both historic responses and response to repeat infection.

6.4 STRENGTHS AND LIMITATIONS

There are a number of strengths associated with the design of this study. That it was designed as part of a wider programme of research into influenza in Vietnam will be key to full exploration of these results and their wider implications. Conducting studies within a primary care setting in low and middle income countries has been traditionally an underserved area of research, this combined with the specific objective of recruiting those with mild disease reflects a major strength when aiming to understand serology in the context of disease transmission and population estimates of incidence. The use of the protein microarray technology in this study has allowed for the investigation of multi-strain serological response, which as demonstrated from this and other studies (Fonville et al., 2014) is becoming vital to the next steps in our understanding of the role of humoral immunity in influenza evolution. Sample quality is fundamental to the accurate interpretation of both serological and virological results. This study was designed with a view to ensuring sample consistency and quality including daily sample collections from study sites, and subsequent aliquoting and freezing within 24 hours. Standardisation across batches of PMA processing also reduces variability and increases confidence in study results. Further work is planned for results validation including HI testing of longitudinal samples, testing for other respiratory viruses and integration of results with population serology results, these will add to the existing strengths of this work and for future directions of influenza seroepidemiology and surveillance.

As with all research, there are important limitations which should be considered with regard to this study and its results. For pragmatic reasons this study recruited from healthcare clinics individuals who already had ILI symptoms. In studies investigating the humoral response and its role in susceptibility, the ideal design would be to sample individuals as close to the onset of symptoms as possible, preferably just prior to infection. Although cohort studies allow the sampling of individuals without ILI symptoms, 'baseline' samples can be collected up to a year prior to the onset of symptoms, depending on the study design (Horby et al., 2012, Lessler et al., 2012). Although this study recruited individuals with mild disease these subjects still attended for clinical assessment. Differences in healthcare seeking behaviour suggests that this study is likely not to be representative of the population as a whole. The majority of influenza infections are self-limiting but a significant proportion cause severe disease (Fischer II et al., 2014, Global Burden of Disease 2013 Mortality and Causes of Death Collaborators, 2015). This study specifically excluded these individuals and further work is required to understand if the findings in mild disease can be directly translated to those with severe complications. The selection of the strains present on the PMA was dictated by the commercial availability of HA1 proteins. The H3N2 strains do not cover all antigenic clusters since its emergence in 1968 and the analysis looking at original antigenic sin is likely to be impacted by this. As multi-strain seroepidemiology for influenza is expanded it is important to determine what the optimal number of strains and antigenic clusters should be included on a testing panel. This will likely depend on the question being asked.

As discussed there were a number of limitations associated with the study design, with these and other aspects of research bias in mind a number of steps were taken to attempt to minimise bias within the study design. By recruiting from a number of different clinics and different clinic types I attempted to recruit across demographic variations expected within HCMC. The inclusion of polyclinics expanded recruitment time into the evening and weekend which is likely to have resulted in greater number of working individuals to be recruited to the study. Allowing recruiting of accompanying family members and household contacts with ILI symptoms was done with the aim of reducing selection bias introduced by healthcare seeking behaviours. Unfortunately only a small number of individuals were recruited through this route and in future studies I would plan to change this to a more prospective recruitment pathway. To reduce loss to follow-up and attrition bias phone-calls were introduced at months two, four and six to encourage individuals to remain within the study, to remind them of study procedures and of the next follow-up date. An additional reminder phone-call was made shortly before the follow-up clinic visit to try and encourage retention. These additional contacts also helped in the reduction of recall bias for any new ILI symptoms which occurred during follow-up period.

In the analysis of the serology results comparisons were made between those who had ILI caused by influenza and those who had non-influenza ILI. The inclusion of non-influenza ILI group provided a comparator who had similar recent history of healthcare seeking behaviour and age and sex matching was aimed at reducing bias in the likely influenza strains that an individual could have been exposed to during their lifetime. The lack of an 'ILI negative' group (i.e. individuals who presented to healthcare for reasons other than ILI) was considered as a potential source of bias in the study design. The recruiting clinic sites reported a high rate of individuals presenting with non-ILI febrile illnesses which have a wide differential diagnoses. It is unclear if non-respiratory causes of febrile illnesses could affect longitudinal influenza serology and with a heterogeneous group of presentations it would be difficult to explore with explanatory power without a much larger sample size if using only the existing clinic sites. By recruiting only patients with ILI it was hoped that any comparison would be more consistent. In future studies designs the inclusion of age and sex matched non-infected control such as the use of elective surgery or trauma patients would be planned.

6.5 FUTURE DIRECTIONS

Building on the results presented in this thesis, further investigation into several areas are planned. The results in this thesis present data collected to the end of May 2015, study recruitment and follow-up continued until the end of November 2015. This included a period of sustained H1N1 transmission which will allow much of analysis performed only in H3N2 cases to be performed for this subtype. In order to further validate the microarray results, subjects included in the follow-up study will have HI assays performed using the same strains as on the microarray. Genomic sequencing is planned for the isolated influenza A viruses along with testing for other respiratory viruses in those who had non-influenza ILI. Finally, the results obtained from this study will be applied to the population dataset to advance assessments on the changing dynamics of disease incidence and susceptibility.

For the wider influenza field, the period since 2009 has led to considerable progress in our understanding of influenza humoral immunity and its contribution to transmission dynamics. This has included exciting new areas such as the identification of conserved regions of the haemagglutinin stalk region, giving hope of a universal vaccine (Lu et al., 2014) but also a renewed interest in old concepts such original antigenic sin and the role of early infection (Kucharski and Gog, 2012). The work presented in this thesis, along with wider literature published since the design and set up of this study, indicates several areas where future work should be concentrated. Understanding the interaction between individual and population level immunity is a vital piece of

the interplay between virus, human and environment. Human challenge studies of influenza are being used to investigate the impact of infection on the immune system and also the interference between different respiratory viruses (Habibi et al., 2015). This technique could be used to investigate the finding of different levels of protection offered by a fixed titre with different force of infections. A better understanding of the role of humoral immunity in disease severity is also important: whether the 'sin' component of original antigenic sin does lead to more serious disease outcomes is an important question. The ethics around both of these types of studies need careful consideration but are important enough to require the discussion. At the population level, trialling of novel vaccination schedules in tropical regions should be considered. In particular trialling of targeted vaccination schedules driven by an assessment of population immunity by assessing population profile of susceptibility, as in this study, or through the use of antigenic landscapes (Fonville et al., 2014). The most pressing area identified in this work is the need to improve our surveillance methods for influenza and further refine the testing and analysis techniques to take into account the difficulties associated with seroepidemiology. This study used a novel protein microarray and the results suggest that a high throughput multi-strain approach is the future of influenza seroepidemiology. Further refinement of the microarray technology to improve the specificity and sensitivity of the test across different age groups is likely to be beneficial. From a methodological perspective the results from this study on test accuracy, and how this will change global estimated disease surveillance, should be replicated in other settings.

6.6 CONCLUSION

Eradication of human influenza, in the manner of smallpox, is unlikely to be possible because of animal and avian reservoirs of disease. This means our focus in influenza control should be on mitigation of severe disease and attempts to minimise the risk of new pandemic strain emergence. Annual influenza vaccination remains the best method available for disease control, despite this there are still around half a million deaths per year. Although investigation of the humoral immune response is one of the oldest themes of influenza research, this thesis demonstrates we still have much to learn about the continually shifting landscape of influenza immunity.

7 REFERENCES

Baas, D. C., Koopmans, M. P., de Bruin, E., Hulscher, H. I. t., Buisman, A. M., Hendrikx, L. H., et al. (2013) 'Detection of influenza A virus homo- and heterosubtype-specific memory B-cells using a novel protein microarray-based analysis tool', *J Med Virol,* 85(5), pp. 899-909.

Bahl, J., Nelson, M. I., Chan, K. H., Chen, R., Vijaykrishna, D., Halpin, R. A., et al. (2011) 'Temporally structured metapopulation dynamics and persistence of influenza A H3N2 virus in humans', *Proceedings of the National Academy of Sciences,* 108(48), pp. 19359-19364.

Banoo, S., Bell, D., Bossuyt, P., Herring, A., Mabey, D., Poole, F., et al. (2010) 'Evaluation of diagnostic tests for infectious diseases: general principles', *Nat Rev Microbiol,* 8(12 Suppl), pp. S17-29.

Barr, I. G., Russell, C., Besselaar, T. G., Cox, N. J., Daniels, R. S., Donis, R., et al. (2014) 'WHO recommendations for the viruses used in the 2013–2014 Northern Hemisphere influenza vaccine: Epidemiology, antigenic and genetic characteristics of influenza A(H1N1)pdm09, A(H3N2) and B influenza viruses collected from October 2012 to January 2013', *Vaccine,* 32(37), pp. 4713-4725.

Barton, K. (2015) *MuMIn: R package*. Available at: https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf (Accessed: 26 Jan 2015).

Bedford, T., Suchard, M. A., Lemey, P., Dudas, G., Gregory, V., Hay, A. J., et al. (2014) 'Integrating influenza antigenic dynamics with molecular evolution', *eLife,* 3.

Bodewes, R., de Mutsert, G., van der Klis, F. R. M., Ventresca, M., Wilks, S., Smith, D. J., et al. (2011) 'Prevalence of Antibodies against Seasonal Influenza A and B Viruses in Children in Netherlands', *Clinical and Vaccine Immunology,* 18(3), pp. 469-476.

Boni, M. F., Chau, N. V. V., Dong, N., Todd, S., Nhat, N. T. D., de Bruin, E., et al. (2013) 'Population-Level Antibody Estimates to Novel Influenza A/H7N9', *Journal of Infectious Diseases,* 208(4), pp. 554-558.

Boucher, D. W., Contreras, G. and Furesz, J. (1979) 'Persistence of influenza serum antibodies in humans following immunization with a bivalent A/Victoria and A/New Jersey vaccine', *Can Med Assoc J,* 120(7), pp. 799-802, 831.

Braciale, T. J., Sun, J. and Kim, T. S. (2012) 'Regulating the adaptive immune response to respiratory virus infection', *Nat Rev Immunol,* 12(4), pp. 295-305.

Broberg, E., Snacken, R., Adlhoch, C., Beaute, J., Galinska, M., Pereyaslov, D., et al. (2015) 'Start of the 2014/15 influenza season in Europe: drifted influenza A(H3N2) viruses circulate as dominant subtype', *Euro Surveill,* 20(4).

Brooks-Pollock, E., Tilston, N., Edmunds, W. J. and Eames, K. T. (2011) 'Using an online survey of healthcare-seeking behaviour to estimate the magnitude and severity of the 2009 H1N1v influenza epidemic in England', *BMC infectious diseases,* 11, pp. 68.

Brooks-Pollock, E., Conlan, A. J. K., Mitchell, A. P., Blackwell, R., McKinley, T. J. and Wood, J. L. N. (2013) 'Age-dependent patterns of bovine tuberculosis in cattle', *Veterinary Research,* 44(1), pp. 97-97.

Caini, S., Huang, Q. S., Ciblak, M. A., Kusznierz, G., Owen, R., Wangchuk, S., et al. (2015) 'Epidemiological and virological characteristics of influenza B: results of the Global Influenza B Study', *Influenza and Other Respiratory Viruses,* 9, pp. 3-12.

Camacho, A., Ballesteros, S., Graham, A. L., Carrat, F., Ratmann, O. and Cazelles, B. (2011) 'Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study', *Proceedings of the Royal Society of London B: Biological Sciences,* 278(1725), pp. 3635-3643.

Canty, A. and Ripley, B. (2015) *boot: Bootstrap R (S-Plus) Functions.* Available at: https://cran.r-project.org/web/packages/boot/boot.pdf (Accessed: 26 Jan 2016).

Cao, P., Yan, A. W. C., Heffernan, J. M., Petrie, S., Moss, R. G., Carolan, L. A., et al. (2015) 'Innate Immunity and the Inter-exposure Interval Determine the Dynamics of Secondary Influenza

Virus Infection and Explain Observed Viral Hierarchies', *PLoS Comput Biol,* 11(8), pp. e1004334.

Cauchemez, S., Horby, P., Fox, A., Mai, L. Q., Thanh, L. T., Thai, P. Q., et al. (2012) 'Influenza Infection Rates, Measurement Errors and the Interpretation of Paired Serology', *PLoS Pathog,* 8(12), pp. e1003061.

Chen, M. I., Lee, V. J., Lim, W. Y., Barr, I. G., Lin, R. T., Koh, G. C., et al. (2010) '2009 influenza A(H1N1) seroconversion rates and risk factors among distinct adult cohorts in Singapore', *Jama,* 303(14), pp. 1383-91.

Cheng, X., Tan, Y., He, M., Lam, T. T.-Y., Lu, X., Viboud, C., et al. (2013) 'Epidemiological Dynamics and Phylogeography of Influenza Virus in Southern China', *Journal of Infectious Diseases,* 207(1), pp. 106-114.

Chiu, C., Ellebedy, A. H., Wrammert, J. and Ahmed, R. (2015) 'B cell responses to influenza infection and vaccination', *Curr Top Microbiol Immunol,* 386, pp. 381-98.

CIA World Factbook (2015) *Vietnam*. Available at: https://www.cia.gov/library/publications/the-world-factbook/geos/vm.html (Accessed: 09 November 2015).

Cohen, S. A., Klassen, A. C., Ahmed, S., Agree, E. M., Louis, T. A. and Naumova, E. N. (2010) 'Trends for influenza and pneumonia hospitalization in the older population: age, period, and cohort effects', *Epidemiology & Infection,* 138(08), pp. 1135-1145.

Coudeville, L., Bailleux, F., Riche, B., Megas, F., Andre, P. and Ecochard, R. (2010) 'Relationship between haemagglutination-inhibiting antibody titres and clinical protection against influenza: development and application of a bayesian random-effects model', *BMC Medical Research Methodology,* 10(1), pp. 1-11.

Cowling, B. J., Ip, D. K. M., Fang, V. J., Suntarattiwong, P., Olsen, S. J., Levy, J., et al. (2014) 'Modes of Transmission of Influenza B Virus in Households', *PLoS ONE,* 9(9), pp. e108850.

Cox, C. M., Goodin, K., Fisher, E., Dawood, F. S., Hamilton, J. J., Leparc, G. F., et al. (2011) 'Prevalence of 2009 Pandemic Influenza A (H1N1) Virus Antibodies, Tampa Bay Florida — November–December, 2009', *PLoS ONE,* 6(12), pp. e29301.

Cox, N. and Subbarao, K. (2000) 'Global epidemiology of influenza: past and present', *Annual review of medicine,* 51(1), pp. 407-421.

Crompton, P. D., Kayala, M. A., Traore, B., Kayentao, K., Ongoiba, A., Weiss, G. E., et al. (2010) 'A prospective analysis of the Ab response to Plasmodium falciparum before and after a malaria season by protein microarray', *Proceedings of the National Academy of Sciences,* 107(15), pp. 6958-6963.

Davenport, F. M., Hennessy, A. V., Francis, T. and Fabisch, W. t. T. A. o. P. (1953) 'Epidemiologic and immunologic significance of age distribution of antibody to antigenic variants of influenza virus', *J Exp Med,* 98(6), pp. 641-656.

Davenport, F. M. and Hennessy, A. V. (1956) 'A serologic recapitulation of past experiences with influenza A; antibody response to monovalent vaccine', *J Exp Med,* 104(1), pp. 85-97.

Davison, A. H., DV (1997) *Bootstrap Methods and Their Applications.* Cambridge: Cambridge University Press.

de Bruin, E., Loeber, J. G., Meijer, A., Castillo, G. M., Cepeda, M. L. G., Torres-Sepúlveda, M. R., et al. (2014) 'Evolution of an influenza pandemic in 13 countries from 5 continents monitored by protein microarray from neonatal screening bloodspots', *Journal of Clinical Virology,* 61(1), pp. 74-80.

Dugan, V. G., Chen, R., Spiro, D. J., Sengamalay, N., Zaborsky, J., Ghedin, E., et al. (2008) 'The Evolutionary Genetics and Emergence of Avian Influenza Viruses in Wild Birds', *PLoS Pathog,* 4(5), pp. e1000076.

Ekiert, D. C., Bhabha, G., Elsliger, M. A., Friesen, R. H., Jongeneelen, M., Throsby, M., et al. (2009) 'Antibody recognition of a highly conserved influenza virus epitope', *Science,* 324(5924), pp. 246-51.

European Centre for Disease Prevention and Control (ECDC) (2015) *Influenza Case Definitions*. http://ecdc.europa.eu/en/activities/surveillance/EISN/surveillance/Pages/influenza_case_definitions.aspx (Accessed: 16 November 2015).

Felgner, P. L., Kayala, M. A., Vigil, A., Burk, C., Nakajima-Sasaki, R., Pablo, J., et al. (2009) 'A Burkholderia pseudomallei protein microarray reveals serodiagnostic and cross-reactive antigens', *Proceedings of the National Academy of Sciences,* 106(32), pp. 13499-13504.

Feng, L., Shay, D. K., Jiang, Y., Zhou, H., Chen, X., Zheng, Y., et al. (2012) 'Influenza-associated mortality in temperate and subtropical Chinese cities, 2003–2008', *Bulletin of the World Health Organization,* 90(4), pp. 279-288B.

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

Finkelman, B. S., Viboud, C., Koelle, K., Ferrari, M. J., Bharti, N. and Grenfell, B. T. (2007) 'Global Patterns in Seasonal Activity of Influenza A/H3N2, A/H1N1, and B from 1997 to 2005: Viral Coexistence and Latitudinal Gradients', *PLoS ONE,* 2(12), pp. e1296.

Fischer II, W. A., Gong, M., Bhagwanjee, S. and Sevransky, J. (2014) 'Global burden of influenza as a cause of cardiopulmonary morbidity and mortality', *Glob Heart,* 9(3), pp. 325-36.

Fonville, J. M., Wilks, S. H., James, S. L., Fox, A., Ventresca, M., Aban, M., et al. (2014) 'Antibody landscapes after influenza virus infection or vaccination', *Science,* 346(6212), pp. 996-1000.

Fox, J. P., Hall, C. E., Cooney, M. K. and Foy, H. M. (1982) 'Influenza virus infections in Seattle families, 1975-1979.', *Am J Epidemiol,* 116(2), pp. 212-42.

Foy, H. M., Cooney, M. K., Taylor, J., Allan, I., Chuang, T. Y., Blumhagen, J., et al. (1980) 'Antibody response to influenza A/New Jersey and A/Victoria virus vaccines in 1976 and subsequent antibody levels after influenza A epidemics, 1977-1979', *J Infect Dis,* 142(2), pp. 139-44.

Fragaszy, E. B., Quinlivan, M., Breuer, J., Craig, R., Hutchings, S., Kidd, M., et al. (2015) 'Public Health Research', *Population-level susceptibility, severity and spread of pandemic influenza: design of, and initial results from, a pre-pandemic and hibernating pandemic phase study using cross-sectional data from the Health Survey for England (HSE)*. Southampton (UK): NIHR Journals Library Copyright (c) Queen's Printer and Controller of HMSO 2015. .

Francis, T., Jr., Davenport, F. M. and Hennessy, A. V. (1953) 'A serological recapitulation of human infection with different strains of influenza virus', *Trans Assoc Am Physicians,* 66, pp. 231-9.

Freidl, G. S., van den Ham, H.-J., Boni, M. F., de Bruin, E. and Koopmans, M. P. G. (2016) 'Changes in heterosubtypic antibody responses during the first year of the 2009 A(H1N1) influenza pandemic', *Scientific Reports,* 6, pp. 20385.

General Statistics Office of Vietnam (2009) *Vietnam Population and housing census 2009 - Age-Sex structure and marital status of the population in Vietnam*. Available at: https://www.gso.gov.vn/default_en.aspx?tabid=515&idmid=5&ItemID=11082 (Accessed: 14 December 2015).

GeurtsvanKessel, C. H., Willart, M. A., van Rijt, L. S., Muskens, F., Kool, M., Baas, C., et al. (2008) 'Clearance of influenza virus from the lung depends on migratory langerin+CD11b- but not plasmacytoid dendritic cells', *J Exp Med,* 205(7), pp. 1621-34.

Glezen, W. P., Schmier, J. K., Kuehn, C. M., Ryan, K. J. and Oxford, J. (2013) 'The Burden of Influenza B: A Structured Literature Review', *American Journal of Public Health,* 103(3), pp. e43-e51.

Global Burden of Disease 2013 Mortality and Causes of Death Collaborators (2015) 'Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013', *The Lancet,* 385(9963), pp. 117-171.

Global Burden of Disease Study 2013 Collaborators (2015) 'Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013', *The Lancet,* 386(9995), pp. 743-800.

Gomes, M. G. M., Lipsitch, M., Wargo, A. R., Kurath, G., Rebelo, C., Medley, G. F., et al. (2014) 'A Missing Dimension in Measures of Vaccination Impacts', *PLoS Pathog,* 10(3), pp. e1003849.

Grilli, E. A., Davies, J. R. and Smith, A. J. (1986) 'Infection with influenza A H1N1. 1. Production and persistence of antibody', *J Hyg (Lond),* 96(2), pp. 335-43.

Habibi, M. S., Jozwik, A., Makris, S., Dunning, J., Paras, A., DeVincenzo, J. P., et al. (2015) 'Impaired Antibody-mediated Protection and Defective IgA B-Cell Memory in Experimental Infection of Adults with Respiratory Syncytial Virus', *Am J Respir Crit Care Med,* 191(9), pp. 1040-9.

Hall, C. E., Cooney, M. K. and Fox, J. P. (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), pp. 365-80.

Hennessy, A. V., Davenport, F. M., Francis, T. and Fabisch, P. (1955) 'Studies of Antibodies to Strains of Influenza Virus in Persons of Different Ages in Sera Collected in a Postepidemic Period', *The Journal of Immunology,* 75(5), pp. 401-409.

Henry Dunand, C. J., Leon, P. E., Kaur, K., Tan, G. S., Zheng, N.-Y., Andrews, S., et al. (2015) 'Preexisting human antibodies neutralize recently emerged H7N9 influenza strains', *The Journal of Clinical Investigation,* 125(3), pp. 1255-1268.

Hens, N., Shkedy, Z., Aerts, M., Faes, C., Van Damme, P. and Beutels, P. (2012) *Modeling Infectious Disease Parameters Based on Serological and Social Contact Data. Statistics for Biology and Health* 1 edn. New York: Springer-Verlag New York.

Hilleman, M. R., Flatley, F. J., Anderson, S. A., Luecking, M. L. and Levinson, D. J. (1958) 'Distribution and Significance of Asian and Other Influenza Antibodies in the Human Population', *New England Journal of Medicine,* 258(20), pp. 969-974.

Hoa, L. K., Hiep, L. V. and Be, L. V. (2011) 'Development of pandemic influenza vaccine production capacity in Viet Nam', *Vaccine,* 29(s1), pp. A34-6.

Hobson, D., Curry, R. L., Beare, A. S. and Ward-Gardner, A. (1972) 'The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses', *The Journal of Hygiene,* 70(4), pp. 767-777.

Horby, P., Mai, L. Q., Fox, A., Thai, P. Q., Thi Thu Yen, N., Thanh, L. T., et al. (2012) 'The Epidemiology of Interpandemic and Pandemic Influenza in Vietnam, 2007-2010: The Ha Nam Household Cohort Study I', *Am J Epidemiol*.

Horimoto, T. and Kawaoka, Y. (2005) 'Influenza: lessons from past pandemics, warnings from current incidents', *Nat Rev Micro,* 3(8), pp. 591-600.

Horsfall, F. L. and Rickard, E. R. (1941) 'Neutralizing antibodies in human serum after influenza A: the lack of strain specificity in the immunological response', *J Exp Med,* 74(5), pp. 433-439.

Huijskens, E. G. W., Reimerink, J., Mulder, P. G. H., van Beek, J., Meijer, A., de Bruin, E., et al. (2013) 'Profiling of Humoral Response to Influenza A(H1N1)pdm09 Infection and Vaccination Measured by a Protein Microarray in Persons with and without History of Seasonal Vaccination', *PLoS One,* 8(1), pp. e54890.

Jao, R. L., Wheelock, E. F. and Jackson, G. G. (1970) 'Production of interferon in volunteers infected with Asian influenza', *J Infect Dis,* 121(4), pp. 419-26.

Jegaskanda, S., Job, E. R., Kramski, M., Laurie, K., Isitman, G., de Rose, R., et al. (2013) 'Cross-reactive influenza-specific antibody-dependent cellular cytotoxicity antibodies in the absence of neutralizing antibodies', *J Immunol,* 190(4), pp. 1837-48.

Jegaskanda, S., Reading, P. C. and Kent, S. J. (2014) 'Influenza-specific antibody-dependent cellular cytotoxicity: toward a universal influenza vaccine', *J Immunol,* 193(2), pp. 469-75.

Jiang, L., Lee, V. J., Lim, W. Y., Chen, M. I., Chen, Y., Tan, L., et al. (2015) 'Performance of case definitions for influenza surveillance', *Euro Surveill,* 20(22), pp. 21145.

Jost, L. (2006) 'Entropy and Diversity', *Oikos,* 113(2), pp. 363-375.

Kasper, M. R., Wierzba, T. F., Sovann, L., Blair, P. J. and Putnam, S. D. (2010) 'Evaluation of an influenza-like illness case definition in the diagnosis of influenza among patients with acute febrile illness in cambodia', *BMC Infectious Diseases,* 10(1), pp. 1-5.

Kenah, E., Chao, D. L., Matrajt, L., Halloran, M. E. and Longini, I. M., Jr. (2011) 'The global transmission and control of influenza', *PLoS ONE,* 6(5), pp. e19515.

Khamphaphongphane, B., Ketmayoon, P., Lewis, H. C., Phonekeo, D., Sisouk, T., Xayadeth, S., et al. (2013) 'Epidemiological and virological characteristics of seasonal and pandemic influenza in Lao PDR, 2008–2010', *Influenza and Other Respiratory Viruses,* 7(3), pp. 304-311.

Kim, J. H., Skountzou, I., Compans, R. and Jacob, J. (2009) 'Original Antigenic Sin Responses to Influenza Viruses', *The Journal of Immunology,* 183(5), pp. 3294-3301.

Kim, S. H., Song, J.-H., Chung, D. R., Thamlikitkul, V., Yang, Y., Wang, H., et al. (2012) 'Changing Trends in Antimicrobial Resistance and Serotypes of Streptococcus pneumoniae Isolates in Asian Countries: an Asian Network for Surveillance of Resistant Pathogens (ANSORP) Study', *Antimicrobial Agents and Chemotherapy,* 56(3), pp. 1418-1426.

King, C. L., Davies, D. H., Felgner, P., Baum, E., Jain, A., Randall, A., et al. (2015) 'Biosignatures of Exposure/Transmission and Immunity', *The American Journal of Tropical Medicine and Hygiene,* 93(3 Suppl), pp. 16-27.

Knossow, M. and Skehel, J. J. (2006) 'Variation and infectivity neutralization in influenza', *Immunology,* 119(1), pp. 1-7.

Kobasa, D., Kodihalli, S., Luo, M., Castrucci, M. R., Donatelli, I., Suzuki, Y., et al. (1999) 'Amino Acid Residues Contributing to the Substrate Specificity of the Influenza A Virus Neuraminidase', *Journal of Virology,* 73(8), pp. 6743-6751.

Koel, B. F., Burke, D. F., Bestebroer, T. M., van der Vliet, S., Zondag, G. C. M., Vervaet, G., et al. (2013) 'Substitutions Near the Receptor Binding Site Determine Major Antigenic Change During Influenza Virus Evolution', *Science,* 342(6161), pp. 976-979.

Koopmans, M., de Bruin, E., Godeke, G. J., Friesema, I., van Gageldonk, R., Schipper, M., et al. (2011) 'Profiling of humoral immune responses to influenza viruses by using protein microarray', *Clinical Microbiology and Infection*, pp. 797-807.

Kucharski, A. J. and Gog, J. R. (2012) 'Age profile of immunity to influenza: effect of original antigenic sin', *Theor Popul Biol,* 81(2), pp. 102-12.

Kucharski, A. J., Lessler, J., Read, J. M., Zhu, H., Jiang, C. Q., Guan, Y., et al. (2015) 'Estimating the Life Course of Influenza A(H3N2) Antibody Responses from Cross-Sectional Data', *PLoS Biol,* 13(3), pp. e1002082.

Lakadamyali, M., Rust, M. J., Babcock, H. P. and Zhuang, X. (2003) 'Visualizing infection of individual influenza viruses', *Proceedings of the National Academy of Sciences,* 100(16), pp. 9280-9285.

Lambach, P., Alvarez, A. M., Hirve, S., Ortiz, J. R., Hombach, J., Verweij, M., et al. (2015) 'Considerations of strategies to provide influenza vaccine year round', *Vaccine*.

Laurie, K. L., Huston, P., Riley, S., Katz, J. M., Willison, D. J., Tam, J. S., et al. (2012) 'Influenza serological studies to inform public health action: best practices to optimise timing, quality and reporting', *Influenza and Other Respiratory Viruses,* 7(2), pp. 211-24.

Laurie, K. L., Engelhardt, O. G., Wood, J., Heath, A., Katz, J. M., Peiris, M., et al. (2015a) 'International Laboratory Comparison of Influenza Microneutralization Assays for A(H1N1)pdm09, A(H3N2), and A(H5N1) Influenza Viruses by CONSISE', *Clinical and Vaccine Immunology,* 22(8), pp. 957-964.

Laurie, K. L., Guarnaccia, T. A., Carolan, L. A., Yan, A. W. C., Aban, M., Petrie, S., et al. (2015b) 'The time-interval between infections and viral hierarchies are determinants of viral interference following influenza virus infection in a ferret model', *Journal of Infectious Diseases,* 212(11), pp. 1701-1710.

Le, M. Q., Lam, H. M., Cuong, V. D., Lam, T. T.-Y., Halpin, R. A., Wentworth, D. E., et al. (2013) 'Migration and Persistence of Human Influenza A Viruses, Vietnam, 2001–2008', *Emerging Infectious Diseases,* 19(11), pp. 1756-1765.

Lee, P. S. and Wilson, I. A. (2015) 'Structural characterization of viral epitopes recognized by broadly cross-reactive antibodies', *Curr Top Microbiol Immunol,* 386, pp. 323-41.

Lee, V. J., Chen, M. I., Yap, J., Ong, J., Lim, W.-Y., Lin, R. T. P., et al. (2011) 'Comparability of Different Methods for Estimating Influenza Infection Rates Over a Single Epidemic Wave', *American Journal of Epidemiology,* 174(4), pp. 468-478.

Lerman, S. J., Wright, P. F. and Patil, K. D. (1980) 'Antibody decline in children following A/New Jersey/76 influenza virus immunization', *J Pediatr,* 96(2), pp. 271-4.

Lessler, J., Riley, S., Read, J. M., Wang, S., Zhu, H., Smith, G. J., et al. (2012) 'Evidence for antigenic seniority in influenza A (H3N2) antibody responses in southern China', *PLoS Pathog,* 8(7), pp. e1002802.

Lessler, J. (2014) 'Charting the life-course epidemiology of influenza', *Science,* 346(6212), pp. 919-920.

Lewis, F. I. and Torgerson, P. R. (2012) 'A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic', *Emerging Themes in Epidemiology,* 9(1), pp. 1-8.

Li, D., Saito, R., Le, M. T., Nguyen, H. L., Suzuki, Y., Shobugawa, Y., et al. (2008) 'Genetic analysis of influenza A/H3N2 and A/H1N1 viruses circulating in Vietnam from 2001 to 2006', *J Clin Microbiol,* 46(2), pp. 399-405.

Li, G.-M., Chiu, C., Wrammert, J., McCausland, M., Andrews, S. F., Zheng, N.-Y., et al. (2012) 'Pandemic H1N1 influenza vaccine induces a recall response in humans that favors broadly cross-reactive memory B cells', *Proceedings of the National Academy of Sciences,* 109(23), pp. 9047-9052.

Liem, N. T. and Lim, W. (2005) 'Lack of H5N1 avian influenza transmission to hospital employees, Hanoi, 2004', *Emerg Infect Dis,* 11(2), pp. 210-5.

Liu, B., Havers, F., Chen, E., Yuan, Z., Yuan, H., Ou, J., et al. (2014) 'Risk Factors for Influenza A(H7N9) Disease—China, 2013', *Clinical Infectious Diseases,* 59(6), pp. 787-794.

Longini, I. M., Jr., Koopman, J. S., Monto, A. S. and Fox, J. P. (1982) 'Estimating household and community transmission parameters for influenza', *Am J Epidemiol,* 115(5), pp. 736-51.

Lu, Y., Welsh, J. P. and Swartz, J. R. (2014) 'Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines', *Proceedings of the National Academy of Sciences,* 111(1), pp. 125-130.

Luna, E. J. A. and Gattás, V. L. (2010) 'Effectiveness of the Brazilian influenza vaccination policy, a systematic review', *Revista do Instituto de Medicina Tropical de São Paulo,* 52, pp. 175-181.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009) 'The BUGS project: Evolution, critique and future directions', *Stat Med,* 28(25), pp. 3049-67.

Lutwama, J. J., Bakamutumaho, B., Kayiwa, J. T., Chiiza, R., Namagambo, B., Katz, M. A., et al. (2012) 'Clinic- and Hospital-Based Sentinel Influenza Surveillance, Uganda 2007–2010', *Journal of Infectious Diseases,* 206(suppl 1), pp. S87-S93.

Marcelin, G., Sandbulte, M. R. and Webby, R. J. (2012) 'Contribution of antibody production against neuraminidase to the protection afforded by influenza vaccines', *Rev Med Virol,* 22(4), pp. 267-79.

McVernon, J., Laurie, K., Nolan, T., Owen, R., Irving, D., Capper, H., et al. (2010) 'Seroprevalence of 2009 pandemic influenza A(H1N1) virus in Australian blood donors, October - December 2009', *Euro Surveill,* 15(40).

Mello, W. A. d., Paiva, T. M. d., Ishida, M. A., Benega, M. A., Santos, M. C. d., Viboud, C., et al. (2009) 'The Dilemma of Influenza Vaccine Recommendations when Applied to the Tropics: The Brazilian Case Examined Under Alternative Scenarios', *PLoS One,* 4(4), pp. e5095.

Miller, M. S., Gardner, T. J., Krammer, F., Aguado, L. C., Tortorella, D., Basler, C. F., et al. (2013) 'Neutralizing Antibodies Against Previously Encountered Influenza Virus Strains Increase over Time: A Longitudinal Analysis', *Science Translational Medicine,* 5(198), pp. 198ra107.

Molbak, K., Espenhain, L., Nielsen, J., Tersago, K., Bossuyt, N., Denissov, G., et al. (2015) 'Excess mortality among the elderly in European countries, December 2014 to February 2015', *Euro Surveill,* 20(11).

Molesti, E., Ferrara, F., Lapini, G., Montomoli, E. and Temperton, N. (2014) 'Discordant Correlation between Serological Assays Observed When Measuring Heterosubtypic Responses against Avian Influenza H5 and H7 Viruses in Unexposed Individuals', *BioMed Research International,* 2014, pp. 231365.

Monto, A. S., Black, S., Plotkin, S. A. and Orenstein, W. A. (2011) 'Response to the 2009 pandemic: Effect on influenza control in wealthy and poor countries', *Vaccine,* 29(38), pp. 6427-6431.

Morris, J. A., Kasel, J. A., Saglam, M., Knight, V. and Loda, F. A. (1966) 'Immunity to influenza to antibody levels', *N Engl J Med,* 274(10), pp. 527-35.

Moura, F. E. (2010) 'Influenza in the tropics', *Curr Opin Infect Dis,* 23(5), pp. 415-20.

Muggeo, V. M. (2008) 'Segmented: an R package to fit regression models with broken-line relationships.', *R News,* 8(1), pp. 20-25.

Murphy, B. R., Phelan, M. A., Nelson, D. L., Yarchoan, R., Tierney, E. L., Alling, D. W., et al. (1981) 'Hemagglutinin-specific enzyme-linked immunosorbent assay for antibodies to influenza A and B viruses', *J Clin Microbiol,* 13(3), pp. 554-60.

Nair, H., Brooks, W. A., Katz, M., Roca, A., Berkley, J. A., Madhi, S. A., et al. (2011) 'Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis', *The Lancet,* 378(9807), pp. 1917-1930.

Nakagawa, S. and Schielzeth, H. (2013) 'A general and simple method for obtaining R2 from generalized linear mixed-effects models', *Methods in Ecology and Evolution,* 4(2), pp. 133-142.

Nauta, J. (2011) 'Chapter 5: Adjusting for Imbalance in Pre-Vaccination State', *Statistics in Clinical Vaccine Trials*. New York: Springer, pp. 57-66.

Nauta, J. J. P., Beyer, W. E. P. and Osterhaus, A. D. M. E. (2009) 'On the relationship between mean antibody level, seroprotection and clinical protection from influenza', *Biologicals,* 37(4), pp. 216-221.

Nayak, D. P., Balogun, R. A., Yamada, H., Zhou, Z. H. and Barman, S. (2009) 'Influenza virus morphogenesis and budding', *Virus Res,* 143(2), pp. 147-61.

Nelson, M. I. and Holmes, E. C. (2007) 'The evolution of epidemic influenza', *Nat Rev Genet,* 8(3), pp. 196-205.

Nelson, M. I., Njouom, R., Viboud, C., Niang, M. N. D., Kadjo, H., Ampofo, W., et al. (2014) 'Multiyear Persistence of 2 Pandemic A/H1N1 Influenza Virus Lineages in West Africa', *Journal of Infectious Diseases,* 210(1), pp. 121-125.

Ng, S. and Gordon, A. (2015) 'Influenza Burden and Transmission in the Tropics', *Curr Epidemiol Rep,* 2(2), pp. 89-100.

Nga, D. T. T., Chuc, N. T. K., Hoa, N. P., Hoa, N. Q., Nguyen, N. T. T., Loan, H. T., et al. (2014) 'Antibiotic sales in rural and urban pharmacies in northern Vietnam: an observational study', *BMC Pharmacology and Toxicology,* 15(1), pp. 1-10.

Nguyen, H. L., Saito, R., Ngiem, H. K., Nishikawa, M., Shobugawa, Y., Nguyen, D. C., et al. (2007) 'Epidemiology of influenza in Hanoi, Vietnam, from 2001 to 2003', *J Infect,* 55(1), pp. 58-63.

Nguyen, H. T., Dharan, N. J., Le, M. T., Nguyen, N. B., Nguyen, C. T., Hoang, D. V., et al. (2009) 'National influenza surveillance in Vietnam, 2006-2007', *Vaccine,* 28(2), pp. 398-402.

Nguyen, Y. T., Graitcer, S. B., Nguyen, T. H., Tran, D. N., Pham, T. D., Le, M. T. Q., et al. (2013) 'National surveillance for influenza and influenza-like illness in Vietnam, 2006–2010', *Vaccine,* 31(40), pp. 4368-4374.

Nobusawa, E., Aoyama, T., Kato, H., Suzuki, Y., Tateno, Y. and Nakajima, K. (1991) 'Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza A viruses', *Virology,* 182(2), pp. 475-85.

Ochiai, H., Shibata, M., Kamimura, K. and Niwayama, S. (1986) 'Evaluation of the efficacy of split-product trivalent A(H1N1), A(H3N2), and B influenza vaccines: reactogenicity, immunogenicity and persistence of antibodies following two doses of vaccines', *Microbiol Immunol,* 30(11), pp. 1141-9.

Ohmit, S. E., Petrie, J. G., Cross, R. T., Johnson, E. and Monto, A. S. (2011) 'Influenza Hemagglutination-Inhibition Antibody Titer as a Correlate of Vaccine-Induced Protection', *Journal of Infectious Diseases,* 204(12), pp. 1879-1885.

Ong, J. B. S., Chen, M. I. C., Cook, A. R., Lee, H. C., Lee, V. J., Lin, R. T. P., et al. (2010) 'Real-Time Epidemic Monitoring and Forecasting of H1N1-2009 Using Influenza-Like Illness from General Practice and Family Doctor Clinics in Singapore', *PLoS ONE,* 5(4), pp. e10036.

Ortiz, J. R., Sotomayor, V., Uez, O. C., Oliva, O., Bettels, D. and McCarron, M. (2009) 'Strategy to enhance influenza surveillance worldwide', *Emerg Infect Dis [serial on the Internet],* Available from http://wwwnc.cdc.gov/eid/article/15/8/08-1422.htm.

Oshansky, C. M., Wong, S.-S., Jeevan, T., Smallwood, H. S., Webby, R. J., Shafir, S. C., et al. (2014) 'Seasonal Influenza Vaccination Is the Strongest Correlate of Cross-Reactive Antibody Responses in Migratory Bird Handlers', *mBio,* 5(6), pp. e02107-14.

Osterholm, M. T. and Kelley, N. S. (2012) 'H5N1 influenza virus seroepidemiological studies: The facts revisited', *Proceedings of the National Academy of Sciences,* 109(21), pp. E1332.

Oxford University Clinical Research Unit (2015) *Influenza-Like Illness in Ho Chi Minh City*. Available at: www.ili.vn/en (Accessed: 16 November 2015).

Palache, A., Oriol-Mathieu, V., Abelin, A. and Music, T. (2014) 'Seasonal influenza vaccine dose distribution in 157 countries (2004-2011)', *Vaccine,* 32(48), pp. 6369-76.

Palese, P., Tobita, K., Ueda, M. and Compans, R. W. (1974) 'Characterization of temperature sensitive influenza virus mutants defective in neuraminidase', *Virology,* 61(2), pp. 397-410.

Palese, P. and Wang, T. T. (2012a) 'Reply to Osterholm and Kelley: H5N1 in humans—the threat is not as bad as it may seem', *Proceedings of the National Academy of Sciences,* 109(21), pp. E1333.

Palese, P. and Wang, T. T. (2012b) 'H5N1 influenza viruses: Facts, not fear', *Proceedings of the National Academy of Sciences,* 109(7), pp. 2211-2213.

Pang, I. K. and Iwasaki, A. (2011) 'Inflammasomes as mediators of immunity against influenza virus', *Trends in Immunology,* 32(1), pp. 34-41.

PATH (2012) *Influenza vaccine production in Vietnam*. http://sites.path.org/vaccinedevelopment/influenza/vaccine-production-in-the-developing-world/ (Accessed: 27 March 2015).

Paus, D., Phan, T. G., Chan, T. D., Gardam, S., Basten, A. and Brink, R. (2006) 'Antigen recognition strength regulates the choice between extrafollicular plasma cell and germinal center B cell differentiation', *The Journal of Experimental Medicine,* 203(4), pp. 1081-1091.

Paynter, S. (2015) 'Humidity and respiratory virus transmission in tropical and temperate settings', *Epidemiol Infect,* 143(6), pp. 1110-8.

Pebody, R. G., Warburton, F., Ellis, J., Andrews, N., Thompson, C., von Wissmann, B., et al. (2015) 'Low effectiveness of seasonal influenza vaccine in preventing laboratory-confirmed influenza in primary care in the United Kingdom: 2014/15 mid-season results', *Euro Surveill,* 20(5), pp. 21025.

Peiris, J. S. M., de Jong, M. D. and Guan, Y. (2007) 'Avian Influenza Virus (H5N1): a Threat to Human Health', *Clinical Microbiology Reviews,* 20(2), pp. 243-267.

Pera, A., Campos, C., Lopez, N., Hassouneh, F., Alonso, C., Tarazona, R., et al. (2015) 'Immunosenescence: Implications for response to infection and vaccination in older people', *Maturitas,* 82(1), pp. 50-5.

Perkins, N. J. and Schisterman, E. F. (2006) 'The Inconsistency of "Optimal" Cutpoints Obtained using Two Criteria based on the Receiver Operating Characteristic Curve', *American Journal of Epidemiology,* 163(7), pp. 670-675.

Petrie, J. G., Ohmit, S. E., Johnson, E., Truscon, R. and Monto, A. S. (2015) 'Persistence of Antibodies to Influenza Hemagglutinin and Neuraminidase Following One or Two Years of Influenza Vaccination', *Journal of Infectious Diseases,* 212(12), pp. 1914-1922.

Priest, P. and Kelly, H. (2015) 'Influenza case definitions need to be fit for purpose', *Euro Surveill,* 20(25), pp. 29-30.

Pyhala, R. (1985) 'Antibody status to influenza A/Singapore/1/57(H2N2) in Finland during a period of outbreaks caused by H3N2 and H1N1 subtype viruses', *J Hyg (Lond),* 95(2), pp. 437-45.

R Core Team 2015. R: A Language and Environment for Statistical Computing. Vienna, Austria.

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

Reed, C., Katz, J. M., Hancock, K., Balish, A., Fry, A. M. and Group, H. N. S. W. (2012) 'Prevalence of seropositivity to pandemic influenza A/H1N1 virus in the United States following the 2009 pandemic', *PloS one,* 7(10), pp. e48187.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011) 'pROC: an open-source package for R and S+ to analyze and compare ROC curves', *BMC Bioinformatics,* 12(1), pp. 1-8.

Rogan, W. J. and Gladen, B. (1978) 'Estimating prevalence from the results of a screening test', *Am J Epidemiol,* 107(1), pp. 71-6.

Rogers, G. N. and Paulson, J. C. (1983) 'Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin', *Virology,* 127(2), pp. 361-73.

Rohm, C., Zhou, N., Suss, J., Mackenzie, J. and Webster, R. G. (1996) 'Characterization of a Novel Influenza Hemagglutinin, H15: Criteria for Determination of Influenza A Subtypes', *Virology,* 217(2), pp. 508-516.

Russell, C. A., Jones, T. C., Barr, I. G., Cox, N. J., Garten, R. J., Gregory, V., et al. (2008) 'The global circulation of seasonal influenza A (H3N2) viruses', *Science,* 320(5874), pp. 340-346.

Rust, M. J., Lakadamyali, M., Zhang, F. and Zhuang, X. (2004) 'Assembly of endocytic machinery around individual influenza viruses during viral entry', *Nat Struct Mol Biol,* 11(6), pp. 567-573.

Sasaki, S., Sullivan, M., Narvaez, C. F., Holmes, T. H., Furman, D., Zheng, N.-Y., et al. (2011) 'Limited efficacy of inactivated influenza vaccine in elderly individuals is associated with decreased production of vaccine-specific antibodies', *The Journal of Clinical Investigation,* 121(8), pp. 3109-3119.

Sauerbrei, A., Langenhan, T., Brandstadt, A., Schmidt-Ott, R., Krumbholz, A., Girschick, H., et al. (2014) 'Prevalence of antibodies against influenza A and B viruses in children in Germany, 2008 to 2010', *Euro Surveill,* 19(5).

Schrauwen, E. J. A. and Fouchier, R. A. M. (2014) 'Host adaptation and transmission of influenza A viruses in mammals', *Emerg Microbes Infect,* 3, pp. e9.

Sealy, R., Surman, S., Hurwitz, J. L. and Coleclough, C. (2003) 'Antibody response to influenza infection of mice: different patterns for glycoprotein and nucleocapsid antigens', *Immunology,* 108(4), pp. 431-439.

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

Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D. M. E., et al. (2004) 'Mapping the Antigenic and Genetic Evolution of Influenza Virus', *Science,* 305(5682), pp. 371-376.

Snacken, R., Broberg, E., BeautÉ, J., Lozano, J. E., Zucs, P. and Amato-Gauci, A. J. (2014) 'Influenza season 2012–2013 in Europe: moderate intensity, mixed (sub)types', *Epidemiology and Infection,* 142(9), pp. 1809-1812.

Song, J.-M., Wang, B.-Z., Park, K.-M., Van Rooijen, N., Quan, F.-S., Kim, M.-C., et al. (2011) 'Influenza Virus-Like Particles Containing M2 Induce Broadly Cross Protective Immunity', *PLoS ONE,* 6(1), pp. e14538.

Sonoguchi, T., Sakoh, M., Kunita, N., Satsuta, K., Noriki, H. and Fukumi, H. (1986) 'Reinfection with influenza A (H2N2, H3N2, and H1N1) viruses in soldiers and students in Japan', *J Infect Dis,* 153(1), pp. 33-40.

Speybroeck, N., Devleesschauwer, B., Joseph, L. and Berkvens, D. (2013) 'Misclassification errors in prevalence estimation: Bayesian handling with care', *Int J Public Health,* 58(5), pp. 791-5.

Sturtz, S., Ligges, U. and Gelman, A. (2005) 'R2WinBUGS: A Package for Running WinBUGS from R', *2005,* 12(3), pp. 16.

Sundaresh, S., Doolan, D. L., Hirst, S., Mu, Y., Unal, B., Davies, D. H., et al. (2006) 'Identification of humoral immune responses in protein microarrays using DNA microarray data analysis techniques', *Bioinformatics,* 22(14), pp. 1760-6.

Sundaresh, S., Randall, A., Unal, B., Petersen, J. M., Belisle, J. T., Gill Hartley, M., et al. (2007) 'From protein microarrays to diagnostic antigen discovery: a study of the pathogen Francisella tularensis', *Bioinformatics,* 23(13), pp. i508-i518.

Takahashi, K., Suzuki, M., Minh, L., Anh, N., Huong, L. T., Son, T. V., et al. (2013) 'The incidence and aetiology of hospitalised community-acquired pneumonia among Vietnamese adults: a prospective surveillance in Central Vietnam', *BMC Infectious Diseases,* 13(1), pp. 296.

Tamerius, J., Nelson, M. I., Zhou, S. Z., Viboud, C., Miller, M. A. and Alonso, W. J. (2011) 'Global influenza seasonality: reconciling patterns across temperate and tropical regions', *Environ Health Perspect,* 119(4), pp. 439-45.

te Beest, D., de Bruin, E., Imholz, S., Wallinga, J., Teunis, P., Koopmans, M., et al. (2014a) 'Discrimination of Influenza Infection (A/2009 H1N1) from Prior Exposure by Antibody Protein Microarray Analysis', *PLoS ONE,* 9(11), pp. e113021.

te Beest, D. E., Birrell, P. J., Wallinga, J., De Angelis, D. and van Boven, M. (2014b) 'Joint modelling of serological and hospitalization data reveals that high levels of pre-existing immunity and school holidays shaped the influenza A pandemic of 2009 in The Netherlands', *Journal of The Royal Society Interface,* 12(103).

Thai, P. Q., Choisy, M., Duong, T. N., Thiem, V. D., Yen, N. T., Hien, N. T., et al. (2015) 'Seasonality of absolute humidity explains seasonality of influenza-like illness in Vietnam', *Epidemics,* 13, pp. 65-73.

Todd, S., De Bruin, E., Nhat, N. T. D., Koopmans, M. and Boni, M. F. (2014) 'Reply to Pawar et al', *Journal of Infectious Diseases,* 210(1), pp. 161-163.

United Nations in Vietnam (2009) 'Review of Expanded Program of Immunisation', *http://www.un.org.vn/en/publications/cat_view/130-un-viet-nam-joint-publications/205-health-joint-un-publications.html* (Accessed 27 March 2016).

Uyeki, T. M., Nguyen, D. C., Rowe, T., Lu, X., Hu-Primmer, J., Huynh, L. P., et al. (2012) 'Seroprevalence of antibodies to avian influenza A (H5) and A (H9) viruses among market poultry workers, Hanoi, Vietnam, 2001', *PLoS One,* 7(8), pp. e43948.

Van Kerkhove, M. D. and Ferguson, N. M. (2012) 'Epidemic and intervention modelling--a scientific rationale for policy decisions? Lessons from the 2009 influenza pandemic', *Bull World Health Organ,* 90(4), pp. 306-10.

Van Kerkhove, M. D., Riley, S., Lipsitch, M., Guan, Y., Monto, A. S., Webster, R. G., et al. (2012) 'Comment on "Seroevidence for H5N1 Influenza Infections in Humans: Meta-Analysis"', *Science,* 336(6088), pp. 1506.

Van Kerkhove, M. D., Broberg, E., Engelhardt, O. G., Wood, J., Nicoll, A. and The, C. s. c. (2013a) 'The consortium for the standardization of influenza seroepidemiology (CONSISE): a global partnership to standardize influenza seroepidemiology and develop influenza investigation protocols to inform public health policy', *Influenza and Other Respiratory Viruses,* 7(3), pp. 231-4.

Van Kerkhove, M. D., Hirve, S., Koukounari, A., Mounts, A. W. and for the, H. N. p. s. w. g. (2013b) 'Estimating age-specific cumulative incidence for the 2009 influenza pandemic: a meta-analysis of A(H1N1)pdm09 serological studies from 19 countries', *Influenza and Other Respiratory Viruses,* 7(5), pp. 872-86.

Veguilla, V., Hancock, K., Schiffer, J., Gargiullo, P., Lu, X., Aranio, D., et al. (2011) 'Sensitivity and Specificity of Serologic Assays for Detection of Human Infection with 2009 Pandemic H1N1 Virus in U.S. Populations', *Journal of Clinical Microbiology,* 49(6), pp. 2210-2215.

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

Vinh, D. N. and Boni, M. F. (2015) 'Statistical identifiability and sample size calculations for serial seroepidemiology', *Epidemics,* 12, pp. 30-39.

Vuong, C. D., Hoang, P. M. V., Nguyen, H. L. K., Nguyen, H. T., Nguyen, T. C., Le, T. T., et al. (2012a) 'The genetic match between vaccine strains and circulating seasonal influenza A viruses in Vietnam, 2001–2009', *Influenza and Other Respiratory Viruses*, pp. no-no.

Vuong, C. D., Hoang, P. M. V., Nguyen, H. L. K., Nguyen, H. T., Nguyen, T. C., Le, T. T., et al. (2012b) 'The genetic match between vaccine strains and circulating seasonal influenza A viruses in Vietnam, 2001–2009', *Influenza and Other Respiratory Viruses,* 7(6), pp. 1151-7.

Vuong, C. D., Hoang, P. M. V., Nguyen, H. L. K., Nguyen, H. T., Nguyen, T. C., Le, T. T., et al. (2012c) 'The genetic match between vaccine strains and circulating seasonal influenza A viruses in Vietnam, 2001–2009', *Influenza and Other Respiratory Viruses,* 7(6), pp. 1151-6.

Wang, T. T., Parides, M. K. and Palese, P. (2012) 'Seroevidence for H5N1 influenza infections in humans: meta-analysis', *Science,* 335(6075), pp. 1463.

Wang, X., Fang, S., Lu, X., Xu, C., Cowling, B. J., Tang, X., et al. (2014) 'Seroprevalence to Avian Influenza A(H7N9) Virus Among Poultry Workers and the General Population in Southern China: A Longitudinal Study', *Clinical Infectious Diseases,* 59(6), pp. e76-e83.

Webster, R. G. M., Arnold S.; Braciale, Thomas J,; Lamb, Robert A; (2013) *Textbook of Influenza 2nd Edition.* 2 edn. Chichester, UK: Wiley Blackwell.

Wei, C.-J., Boyington, J. C., McTamney, P. M., Kong, W.-P., Pearce, M. B., Xu, L., et al. (2010) 'Induction of Broadly Neutralizing H1N1 Influenza Antibodies by Vaccination', *Science,* 329(5995), pp. 1060-1064.

Wertheim, H. F. L., Chandna, A., Vu, P. D., Pham, C. V., Nguyen, P. D. T., Lam, Y. M., et al. (2013) 'Providing Impetus, Tools, and Guidance to Strengthen National Capacity for Antimicrobial Stewardship in Viet Nam', *PLoS Med,* 10(5), pp. e1001429.

Whittle, J. R. R., Zhang, R., Khurana, S., King, L. R., Manischewitz, J., Golding, H., et al. (2011) 'Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin', *Proceedings of the National Academy of Sciences,* 108(34), pp. 14216-14221.

Wilson, I. A. and Cox, N. J. (1990) 'Structural basis of immune recognition of influenza virus hemagglutinin', *Annu Rev Immunol,* 8, pp. 737-71.

World Health Organisation (1980) 'A revision of the system of nomenclature for influenza viruses: a WHO Memorandum', *Bulletin of the World Health Organization,* 58(4), pp. 585-591.

World Health Organisation (2007) *Process of influenza vaccine virus selection and development*. Available at: http://apps.who.int/gb/pip/pdf_files/Fluvaccvirusselection.pdf (Accessed: 05 November 2015).

World Health Organisation (2011) *WHO Global Influenza Surveillance Network. Manual for the laboratory diagnosis and virological surveillance of influenza.* Geneva, Switzerland. Available at: http://www.who.int/influenza/gisrs_laboratory/manual_diagnosis_surveillance_influenza/en/ (Accessed: 16 November 2015).

World Health Organisation (2012) 'Vaccines against influenza - WHO position paper', *Weekly Epidemiological Record,* 87(47), pp. 461-476.

World Health Organisation (2014a) *WHO Global Epidemiological Surveillance Standards for Influenza*. Available at: http://www.who.int/influenza/resources/documents/influenza_surveillance_manual/en/ (Accessed: 09 February 2016).

World Health Organisation (2014b) *WHO information for molecular diagnosis of influenza virus - update*. Geneva, Switzerland. Available at: http://www.who.int/influenza/gisrs_laboratory/molecular_diagnosis/en/ (Accessed: 16 November 2015).

World Health Organisation (2014c) 'Fact sheet 211: influenza (seasonal)', *http://www.who.int/mediacentre/factsheets/fs211/en/* (Accessed 07 November 2015).

World Health Organisation (2015a) *Influenza Update*. Available at: http://www.who.int/influenza/surveillance_monitoring/updates/latest_update_GIP_surveillance/en/ (Accessed: 08 December 2015).

World Health Organisation (2015b) *Cumulative number of confirmed human cases of avian influenza A(H5N1) reported to WHO 2003-2015 (as of 13 Nov 2015)*. Available at: http://www.who.int/influenza/human_animal_interface/H5N1_cumulative_table_archives/en/ (Accessed: 08 December 2015).

World Health Organisation (2015c) *WHO recommendations on the composition of influenza virus vaccines*: WHO. Available at: http://www.who.int/influenza/vaccines/virus/recommendations/en/ (Accessed: 08 November 2015).

World Health Organisation & MOH Vietnam (2012) *Health Service Delivery Profile: Vietnam 2012*. Available at: http://www.wpro.who.int/health_services/service_delivery_profile_vietnam.pdf (Accessed: 09 November 2015).

World Health Organisation, M. o. H. V. (2010) *Global Adult Tobacco Study (GATS): Vietnam 2010*. Available at: http://www.who.int/tobacco/surveillance/en_tfi_gats_vietnam_report.pdf (Accessed: 06 Jan 2016).

Wrammert, J., Smith, K., Miller, J., Langley, W. A., Kokko, K., Larsen, C., et al. (2008) 'Rapid cloning of high-affinity human monoclonal antibodies against influenza virus', *Nature,* 453(7195), pp. 667-671.

Wu, J. T., Ma, E. S. K., Lee, C. K., Chu, D. K. W., Ho, P.-L., Shen, A. L., et al. (2010) 'The Infection Attack Rate and Severity of 2009 Pandemic H1N1 Influenza in Hong Kong', *Clinical Infectious Diseases,* 51(10), pp. 1184-1191.

Wu, J. T., Leung, K., Perera, R. A. P. M., Chu, D. K. W., Lee, C. K., Hung, I. F. N., et al. (2014) 'Inferring Influenza Infection Attack Rate from Seroprevalence Data', *PLoS Pathog,* 10(4), pp. e1004054.

Yamane, N., Hiratsuka, M., Arikawa, J., Odagiri, T. and Ishida, N. (1981) 'Antibody responses after repeated influenza A virus immunizations among schoolchildren in Japan', *J Hyg (Lond),* 87(3), pp. 383-92.

Yang, L., Ma, S., Chen, P. Y., He, J. F., Chan, K. P., Chow, A., et al. (2011) 'Influenza associated mortality in the subtropics and tropics: Results from three Asian cities', *Vaccine,* 29(48), pp. 8909-8914.

Ye, Z., Liu, T., Offringa, D. P., McInnis, J. and Levandowski, R. A. (1999) 'Association of influenza virus matrix protein with ribonucleoproteins', *J Virol,* 73(9), pp. 7467-73.

Yoshida, L. M., Suzuki, M., Yamamoto, T., Nguyen, H. A., Nguyen, C. D., Nguyen, A. T., et al. (2010) 'Viral pathogens associated with acute respiratory infections in central vietnamese children', *Pediatr Infect Dis J,* 29(1), pp. 75-7.

Youden, W. J. (1950) 'Index for rating diagnostic tests', *Cancer,* 3(1), pp. 32-5.

8 SUPPLEMENTARY APPENDICES

8.1 PUBLICATIONS ARISING DIRECTLY FROM PHD

A

MF Boni, NVV Chau, N Dong, **S Todd**, NTD Nhat, E de Bruin, J van Beek, NT Hien, CP Simmons, J Farrar, M Koopmans 'Population-level antibody estimates to novel influenza A/H7N9' J Infect Dis. first published online May 17, 2013 doi:10.1093/infdis/jit224

S Todd, E De Bruin, NT Nhat, M Koopmans, MF Boni. 'Reply to Pawar et al' J Infect Dis. 2014 doi: 10.1093/infdis/jiu034

8.2   PUBLICATIONS ARISING INDIRECTLY FROM PHD

S Todd, PJ Diggle, PJ White, A Fearne, JM Read. 'The spatio-temporal association of non-prescription retail sales with cases during the 2009 influenza pandemic in Great Britain.' BMJ Open 2014; 4:e004869.doi:10.1136/bmjopen-2014-004869

8.3   STUDY DOCUMENTATION

Informed Consent Form   C

Patient Information Leaflet

Case Record Form

SUSCEPTIBILITY TO INFLUENZA A:
A STUDY IN HO CHI MINH CITY

You/Your child are(is) being asked to participate in a research study because you are suspected to have influenza based on your symptoms. Participation is your choice. Take as much time as you need to read the following information regarding this research study. You will be given a copy of this form to keep.

Why is this study being done?
Influenza is normally a mild, short lived illness. However in some groups it causes serious illness and death. The influenza virus continually changes and this means individuals may have multiple influenza infections during their life. The purpose of this study is to explore how previous influenza infections affect your risk of catching current influenza viruses. We are also interested in how important infections you catch early in life (before the age of 10) are.

What will happen if you/your child take(s) part in this study?
You/Your child will be asked to spend up to 15 minutes in the study today. If you agree to join the study the following investigations will be performed:

1. Investigations for all participants
   a. One nasal swab and one throat swab will be collected to give us information on whether influenza is causing your symptoms and if so the type of influenza you/your child has.
   b. 5ml venous blood sample will be collected by the study nurse.

We will collect some simple information from everyone who takes part in the study. This will include information on your previous health, on people you came into contact with yesterday and the people who you live with.

If anyone in your household develops similar symptoms in the next 7 days we would like to invite them to join the study. This is an optional part of the study and you are not under any obligation to inform your household members about this study.

We want to follow up a group of participants over the next few months. You may be contacted by a study staff to invite you to take part in this part of the study.

What will happen if you/your child agree to take part in the follow-up study?
Some people will be invited to have additional study visits over the next 7 months. If you agree to join the follow-up study the following investigations will be performed:

   a. 5ml venous blood sample to be taken at this clinic approximately 1 month, 3 months, 5 months and 7 months after today's visit.
   b. Monthly questionnaire on your recent symptoms. We will do this on the day you/your child attend for the blood test. For months where you don't have a clinic visit we will telephone you to ask these questions (approximately 2 months, 4 months and 6 months after today).
   c. If you/your child develop new symptoms similar to today we would like you to contact the study team

who will ask some simple questions to decide if influenza could be causing your symptoms. If we think it could be influenza we will invite you/your child to attend the clinic. At this visit you/ your child will have one nasal swab and one throat swab collected to give us information on whether influenza is causing the symptoms and if so the type of influenza you/your child has

Are there potential benefits to taking part in this study?

Participation in this study will have only limited direct benefit to you/your child. The swabs will not be processed right away and we will report the results back to your doctor and you/your child, but that they will not be available in time to change your treatment. For most people, specific anti-viral treatment for influenza is not needed and no treatment will be offered as part of the study.

You will receive 50,000VND compensation for participating in the study. If you agree to the follow-up study you will receive 220,000VND for each return study visit including any extra visits to the clinic if you develop new influenza like symptoms

Your/your child's participation will help the Ho Chi Minh City Health Department understand patterns of influenza in the city which may help to develop future advice on the use of vaccination and anti-virals.

What are the risks if you/your child take(s) part in this study?

A nasal swab may cause some slight discomfort, and rarely mild bleeding. A throat swab may cause a sensation of gagging and an urge to cough. Collection of blood samples may be painful and there is a very small risk of infection. The amount of blood we collect is very small and poses no risk to your health. The Liverpool School of Tropical Medicine is sponsoring this study and has insurance in place to cover the unlikely case of any harm coming to you as a result of participating in this study.

Will my health information be kept confidential?

All samples will be identified by only a number. No names or personal identifiers will be sent with the samples. The information collected from you/your child will be kept in a secure location. Study staff or the research ethics board may review the information collected for the study and will do so in the strictest confidence.

What will happen to the samples?

Swabs will be tested for influenza within the next few weeks. Blood samples will be stored in a freezer at the laboratory Hospital for Tropical Diseases. If you test positive for influenza then a small amount of your blood will be sent to The Netherlands for testing on immune reactions to previous influenza infections. This test cannot be performed in Vietnam at present. All results will be sent directly back to Vietnam for analysis. The results of these blood tests are only meaningful when looked at in a big group, we don't know what the results mean for an individual person. You will not receive the results of your individual blood test but we will send the results of the whole study to you/your child.

After this study is completed, samples may be stored in the freezer at HTD for up to 10 years. Further research on these samples may be performed in the future to improve our understanding of diseases in Viet Nam, including tests done by researchers outside Vietnam. These future studies will be approved by Hospital for Tropical Diseases. Some of this future research may include studies of your genetic code to try and understand why some people get

sick from this disease, while others do not. This will involve studying the DNA from your cells. Results will be made available to other researchers but you will not be identified and no one will know whose genetic information it is. You can choose if you agree to your samples being stored after this study is completed and if you also agree to further research on your samples by marking the end of this form.

What are your/your child's rights if you/your child take(s) part in this study?
Your/Your child's participation is voluntary. You/Your child may refuse to participate or may stop participating at any time without penalty. Choosing not to participate or stopping participation will not affect the quality of your/your child's health care. If you/your child decide to stop the study, just tell your doctor or a member of the study team.

What if you/your child have questions about this study?
If you/your child have(has) any questions about this research study or your/your child's medical care, you should contact Dr. Nguyen Thi Cam Huong at 09 8377 3915 or the study team at 09 3701 4962 Monday to Friday 0800-1600. If you/your child have any questions about your/your child's rights as a research subject or about research-related injuries, you/your child should contact the Clinical Research Unit of the Hospital for Tropical Diseases at 08 3924 1983.

SUSCEPTIBILITY TO INFLUENZA A:
A STUDY IN HO CHI MINH CITY

Participant's name: _____ Participant ID [__] [__] [__] [__] [__]

- I have read the participant information sheet for this study, and I have been told about the purpose, possible risks and benefits of taking part in this study.
- I have had a chance to discuss this information with study staff, and have got answers that I can understand to all my questions.
- I consent to study staff collecting and processing my information, including information about my / my child's health, and using this information for future medical research about influenza.
- I freely agree that I / my child will take part in this study.
- I understand that I / my child may withdraw from this study at any time, and that if I do leave the study it will not affect my future care. If I decide to leave the study, I agree that the information collected about me up to the point when I withdraw, may continue to be used.
- □ I AGREE **OR** □ I DO NOT AGREE to take part in the follow-up part of the study. I agree to attend the clinic for follow up visits 1, 3, 5 and 7 months after this visit and be contacted monthly about my symptoms.
- □ I AGREE **OR** □ I DO NOT AGREE that samples from me / my child may be stored and that further research on these samples may be undertaken in the future, including tests done by researchers outside Viet Nam.
- □ I AGREE **OR** □ I DO NOT AGREE that genetic tests on my blood sample can be performed and that this may be done outside of Viet Nam.

By signing my name here, I am saying that you gave me a copy of this form. I will keep my copy until my part in the study ends.

_____   _____   _____   _____
Signature of person giving   Print name   Relationship to   Date of signature
consent   participant

PERSON OBTAINING CONSENT:
I confirm the participant (and their parent/guardian if under 15) has understood the information in this form, has had a chance to ask questions and consider the answers to his/her questions and voluntarily agrees to be in the study.
If the subject is under 15, I confirm they have given verbal assent and agree to join the study □.

_____   _____   _____
Printed name of study staff   Signature of study staff   Date

WITNESS (if the person giving consent can not read, a witness should sign below):
I was present when this consent form was read accurately to the participant. I agree that they have given their consent to be in the study.

_____   _____   _____
Printed name of witness   Signature of witness   Date

SUSCEPTIBILITY TO INFLUENZA A:
A STUDY IN HO CHI MINH CITY

Thank you for taking part in our study. We would like to recruit other people who live in your house to our study if they develop similar symptoms to you in the next 7 days. This part of the study is voluntary and you can chose not to pass this information onto the people who live in your home.

Why is this part of the study being done?
Influenza viruses often transmit between people who live in the same house. We are interested in looking at how much the virus spread in households in Vietnam and whether everyone in a household has the same patterns of previous infection.

What symptoms should I look for?
We are interested if anyone develops symptoms of:

- Fever
- Cough
- Runny Nose
- Muscle Aches
- Head Aches
- Sore Throat

What will my household contact have to do as part of the study?
The person in your household will go through all the same study procedures as you did today. If they are interested in taking part, our study team will discuss it with them in detail before any procedures take place and they can make their own decision about whether to join the study.

Why do you need to know my study number?
If we have people from the same household or family it is useful for us to be able to link their samples together to compare results. Linking will be done through study numbers not your name and all results are confidential.

What is the next step?
If your household contact is interested in joining the study they can contact the study team through the dedicated study phone number 0937014962 Monday to Friday 0800-1600. The member of the study team will then invite them to attend the clinic discuss the study in more detail. We will also provide you with a business card with the study phone number which you can pass onto other members of your house.

Thank you again for helping us with this study. If you have any questions please contact us on the above number.

ENROLMENT	**ENROL**

Participant Number	Initials
10FL-[__|__]-[__|__|__]	[__|__|__|__|__|__]

STUDY PROCEDURES
ICF MUST BE SIGNED BEFORE ANY STUDY PROCEDURES ARE PERFORMED.

1. Date of ICF Signed: [__|__]/[__|__]/[__|__]

2. Recruitment Group: ○ i   ○ ii   ○ iii

3. Agreed to enter the follow-up sub-study: ○ YES ○ NO

DEMOGRAPHIC HISTORY

4. Sex   ○ MALE   ○ FEMALE

5. Date of birth: Day: [__|__]   Month: [__|__]   Year: [__|__|__|__]

CURRENT INFLUENZA LIKE ILLNESS

6. How many days ago did your symptoms start?   [__|__]

7. Which of these symptoms have you had since you have been ill?

a.	Fever	○ Yes	○ No
	Temperature (If known): [__|__].[__]		
b.	Headache	○ Yes	○ No
c.	Runny nose	○ Yes	○ No
d.	Cough	○ Yes	○ No
e.	Sore throat	○ Yes	○ No
f.	Muscle Aches	○ Yes	○ No
g.	Vomiting/Diarrhoea	○ Yes	○ No
h.	Shortness of Breath	○ Yes	○ No
i.	Malaise	○ Yes	○ No

8. Have you taken any of these treatments for this illness?

a.	Paracetamol	○ Yes	○ No
b.	Anti-viral	○ Yes	○ No
c.	Anti-bacterial	○ Yes	○ No
d.	Other	○ Yes	○ No

   If Yes, specify: [_____]

9. Have you been able to do your normal tasks during this illness? ○ Yes ○ No ○ Don't know

ENROLMENT	**ENROL**

Participant Number   Initials

10FL-[__|__]-[__|__|__]   [__|__|__|__|__|__]

4. PREVIOUS ILLNESS HISTORY

10. Have you ever suffered from any of these health problems?

     a.   COPD   ○ Yes   ○ No

     b.   Congenital Heart Disease   ○ Yes   ○ No

     c.   Heart Failure   ○ Yes   ○ No

     d.   Diabetes   ○ Yes   ○ No

     e.   Asthma   ○ Yes   ○ No

     f.   Other Respiratory Disease   ○ Yes   ○ No

         If yes, specify: [_____]

     g.   Other Health Problems   ○ Yes   ○ No

         If yes, specify: [_____]

11. Are you currently pregnant?   ○ Yes   ○ No   ○ Don't know   ○ N/A

12. Are you a smoker?   ○ Yes   ○ No   ○ Don't know   ○ N/A

13. Have you ever had influenza vaccination?   ○ Yes   ○ No   ○ Don't know

     If Yes, when? (mm/yyyy):   [__|__] / [__|__|__|__]

HOUSEHOLD AND CONTACT HISTORY

14. What is your usual/main occupation?

         ○ School/College Student

         ○ Stay at Home Parent/Carer

         ○ Professional or Office Work, **if yes**: Contact with Public?   ○ YES   ○ NO

         ○ Shop Assistant/Trader

         ○ Manual Work

         ○ Other, specify [_____]

15. How many people did you have a face to face conversation with yesterday? **Refused to answer** ☐

	0-4 YEARS	5-18 YEARS	19-44 YEARS	45-64 YEARS	65+ YEARS
HOME	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
WORK/SCHOOL	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
OTHER	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]

16. How many people did you have physical contact with yesterday? **Refused to answer** ☐

	0-4 YEARS	5-18 YEARS	19-44 YEARS	45-64 YEARS	65+ YEARS
HOME	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
WORK/SCHOOL	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
OTHER	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]

17. How many people (including you) have stayed in your house in the last 7 days: [__|__]

ENROLMENT		**ENROL**

Participant Number
10FL-[__|__]-[__|__|__]

Initials
[__|__|__|__|__|__]

18. For the people who stayed in your house, can you tell us their ages and if they have had any illness in the last 7 days: **Refused to answer** ☐

	AGE	RELATIONSHIP *	USUAL DAYTIME LOCATION	ILLNESS IN LAST 7 DAYS	ENROLLED IN 10FL
1	[__|__]	[__|__]	⭘ Home ⭘ School ⭘ Work ⭘ Travel	⭘ Yes ⭘ No ⭘ Don't Know	⭘ Yes ⭘ No ⭘ Don't Know If Yes, Study ID: [__|__]-[__|__|__]
2	[__|__]	[__|__]	⭘ Home ⭘ School ⭘ Work ⭘ Travel	⭘ Yes ⭘ No ⭘ Don't Know	⭘ Yes ⭘ No ⭘ Don't Know If Yes, Study ID: [__|__]-[__|__|__]
3	[__|__]	[__|__]	⭘ Home ⭘ School ⭘ Work ⭘ Travel	⭘ Yes ⭘ No ⭘ Don't Know	⭘ Yes ⭘ No ⭘ Don't Know If Yes, Study ID: [__|__]-[__|__|__]
4	[__|__]	[__|__]	⭘ Home ⭘ School ⭘ Work ⭘ Travel	⭘ Yes ⭘ No ⭘ Don't Know	⭘ Yes ⭘ No ⭘ Don't Know If Yes, Study ID: [__|__]-[__|__|__]
5	[__|__]	[__|__]	⭘ Home ⭘ School ⭘ Work ⭘ Travel	⭘ Yes ⭘ No ⭘ Don't Know	⭘ Yes ⭘ No ⭘ Don't Know If Yes, Study ID: [__|__]-[__|__|__]
6	[__|__]	[__|__]	⭘ Home ⭘ School ⭘ Work ⭘ Travel	⭘ Yes ⭘ No ⭘ Don't Know	⭘ Yes ⭘ No ⭘ Don't Know If Yes, Study ID: [__|__]-[__|__|__]
7	[__|__]	[__|__]	⭘ Home ⭘ School ⭘ Work ⭘ Travel	⭘ Yes ⭘ No ⭘ Don't Know	⭘ Yes ⭘ No ⭘ Don't Know If Yes, Study ID: [__|__]-[__|__|__]
8	[__|__]	[__|__]	⭘ Home ⭘ School ⭘ Work ⭘ Travel	⭘ Yes ⭘ No ⭘ Don't Know	⭘ Yes ⭘ No ⭘ Don't Know If Yes, Study ID: [__|__]-[__|__|__]

* *1.Parent; 2.Parent in Law; 3.Grandparent;   4.Grandparent in Law; 5.Uncle/Aunt; 6. Uncle/Aunt in law;*

   7. Nephew/Neice; 8.Nephew/Neice in law; 9.Cousin; 10.Friend; 11.Other 12. Son/Daughter 13. Son/Daughter in law 14. Wife/Husband

19. How often do you have skin contact with the follow animals: **Refused to answer** ☐

	4+ DAYS PER WEEK	1 – 3 DAYS PER WEEK	MORE THAN ONCE PER WEEK BUT LESS THAN ONCE PER MONTH	LESS THAN ONCE A MONTH	NEVER
LIVE POULTRY (CHICKENS, DUCKS, GEESE)	[ ]	[ ]	[ ]	[ ]	[ ]
DEAD POULTRY (CHICKENS, DUCKS, GEESE)	[ ]	[ ]	[ ]	[ ]	[ ]
LIVE PIGS	[ ]	[ ]	[ ]	[ ]	[ ]
DEAD PIGS	[ ]	[ ]	[ ]	[ ]	[ ]

FOLLOW UP	**FU**

Participant Number	Initials
10FL-[__|__]-[__|__|__]	[__|__|__|__|__|__]

STUDY PROCEDURES   1 / 2

ICF MUST BE SIGNED BEFORE ANY STUDY PROCEDURES ARE PERFORMED.

1. Date of ICF Signed: [__|__]/[__|__]/[__|__] *(dd/mm/yy)*

2. Visit number: [__|__]   Telephone Visit: ○ Yes   ○ No

3. Date of visit: [__|__]/[__|__]/[__|__] *(dd/mm/yy)*

4. Confirm participant agrees to continue in the follow up study and
   any additional questions have been addressed   ○ Yes   ○ No

5. Date of next follow up: [__|__]/[__|__]/[__|__] *(dd/mm/yy)*
   a. Study complete   ☐

PREVIOUS INFLUENZA LIKE ILLNESS (if this is visit 2)

6. After your last visit did your symptoms get worse?   ○ Yes ○ No ○ Don't Know

7. After your last visit did you seek any further medical advice?
   ○ No ○ Pharmacy   ○ Private Clinic ○Hospital Clinic   ○ Hospital Inpatient ○ Other

INFLUENZA LIKE ILLNESS SINCE LAST VISIT

8. Have you had any new respiratory symptoms since the last clinic visit?
   ○ Yes
   ○ No ○ Don't Know (go to question 14)

   a. If Yes, Did you attend for Symptomatic Follow Up Visit (SFU)
      ○ Yes (go to question 9 & 10)
      ○ No (go to question 11 – 13)

9. After your last visit did your symptoms get worse?   ○ Yes ○ No ○ Don't Know

10. After your last visit did you seek any further medical advice?
    ○ No ○ Pharmacy   ○ Private Clinic ○Hospital Clinic   ○ Hospital Inpatient ○ Other

11. Which of these symptoms did you have?

   a. Fever   ○ Yes   ○ No
      Temperature (If known): [__|__].[__]
   b. Headache   ○ Yes   ○ No
   c. Runny nose   ○ Yes   ○ No
   d. Cough   ○ Yes   ○ No
   e. Sore throat   ○ Yes   ○ No

FOLLOW UP **FU**

Participant Number   Initials

10FL-[__|__]-[__|__|__]   [__|__|__|__|__|__]

f. Muscle Aches	O Yes	O No	
g. Vomiting/Diarrhoea	O Yes	O No	
h. Shortness of Breath	O Yes	O No	
i. Malaise	O Yes	O No	

12. Did you take any of these treatments for this illness?

a. Paracetamol	O Yes	O No
b. Anti-viral	O Yes	O No
c. Anti-bacterial	O Yes	O No
d. Other	O Yes	O No

If Yes, specify: [_____]

13. Have you been able to do your normal tasks during this illness? O Yes   O No   O Don't know

CONTACT HISTORY

14. How many people did you have a face to face conversation with yesterday? **Refused to answer** ☐

	0-4 YEARS	5-18 YEARS	19-44 YEARS	45-64 YEARS	65+ YEARS
HOME	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
WORK/SCHOOL	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
OTHER	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]

15. How many people did you have physical contact with yesterday? **Refused to answer** ☐

	0-4 YEARS	5-18 YEARS	19-44 YEARS	45-64 YEARS	65+ YEARS
HOME	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
WORK/SCHOOL	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
OTHER	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]

16. How many people (including you) have stayed in your house in the last 7 days: [__|__]

FOLLOW UP	**TFU**

Participant Number
10FL-[__|__]-[__|__|__]

Initials
[__|__|__|__|__|__]

STUDY PROCEDURES

ICF MUST BE SIGNED BEFORE ANY STUDY PROCEDURES ARE PERFORMED.

1. Date of ICF Signed: [__|__]/[__|__]/[__|__] *(dd/mm/yy)*

2. Telephone contact number: [__|__]

3. Date of contact: [__|__]/[__|__]/[__|__] *(dd/mm/yy)*

4. Confirm participant agrees to continue in the follow up study and any additional questions have been addressed   ○ Yes   ○ No

5. Date of next follow up: [__|__]/[__|__]/[__|__] *(dd/mm/yy)*

INFLUENZA LIKE ILLNESS SINCE LAST VISIT

6. Have you had any new respiratory symptoms since the last clinic visit?
   ○ Yes   (go to question 7)
   ○ No ○ Don't Know   (end call)

7. How many days ago did your symptoms start?   [__|__] (if <5 invite to study clinic)

8. Did you seek any further medical advice?
   ○ No ○ Pharmacy ○ Private Clinic ○ Hospital Clinic ○ Hospital Inpatient ○ Other

9. Which of these symptoms did you have?

a.	Fever	○ Yes	○ No
	Temperature (If known): [__|__].[__]		
b.	Headache	○ Yes	○ No
c.	Runny nose	○ Yes	○ No
d.	Cough	○ Yes	○ No
e.	Sore throat	○ Yes	○ No
f.	Muscle Aches	○ Yes	○ No
g.	Vomiting/Diarrhoea	○ Yes	○ No
h.	Shortness of Breath	○ Yes	○ No
i.	Malaise	○ Yes	○ No

10. Did you take any of these treatments for this illness?

e.	Paracetamol	○ Yes	○ No

FOLLOW UP		**TFU**
Participant Number 10FL-[__|__]-[__|__|__]	Initials [__|__|__|__|__|__]	

f.	Anti-viral	⭕ Yes	⭕ No
g.	Anti-bacterial	⭕ Yes	⭕ No
h.	Other	⭕ Yes	⭕ No
	If Yes, specify: [_____]		

11. Have you been able to do your normal tasks during this illness? ⭕ Yes ⭕ No ⭕ Don't know

FOLLOW UP	**SFU**
Participant Number	Initials
10FL-[__|__]-[__|__|__]	[__|__|__|__|__|__]

STUDY PROCEDURES

ICF MUST BE SIGNED BEFORE ANY STUDY PROCEDURES ARE PERFORMED.

1. Date of ICF Signed: [__|__]/[__|__]/[__|__] *(dd/mm/yy)*

2. Symptomatic Visit: [__|__]

3. Date of visit: [__|__]/[__|__]/[__|__] *(dd/mm/yy)*

4. Confirm participant agrees to continue in the follow up study and any additional questions have been addressed   ○ Yes   ○ No

5. Date of next follow up: [__|__]/[__|__]/[__|__] *(dd/mm/yy)*

CURRENT INFLUENZA LIKE ILLNESS

6. How many days ago did your symptoms start?   [__|__]

7. Did you seek any further medical advice?
   ○ No ○ Pharmacy ○ Private Clinic ○ Hospital Clinic ○ Hospital Inpatient ○ Other

8. Which of these symptoms do you have?

   j. Fever   ○ Yes   ○ No

   Temperature (If known): [__|__].[__]

   k. Headache   ○ Yes   ○ No

   l. Runny nose   ○ Yes   ○ No

   m. Cough   ○ Yes   ○ No

   n. Sore throat   ○ Yes   ○ No

   o. Muscle Aches   ○ Yes   ○ No

   p. Vomiting/Diarrhoea   ○ Yes   ○ No

   q. Shortness of Breath   ○ Yes   ○ No

   r. Malaise   ○ Yes   ○ No

9. Did you take any of these treatments for this illness?

   i. Paracetamol   ○ Yes   ○ No

   j. Anti-viral   ○ Yes   ○ No

   k. Anti-bacterial   ○ Yes   ○ No

   l. Other   ○ Yes   ○ No

   If Yes, specify: [_____]

FOLLOW UP		**SFU**
Participant Number	Initials	
10FL-[__|__]-[__|__|__]	[__|__|__|__|__|__]	

10. Have you been able to do your normal tasks during this illness? ◯ Yes ◯ No ◯ Don't know

CONTACT HISTORY

11. How many people did you have a face to face conversation with yesterday? Refused to answer □

	0-4 YEARS	5-18 YEARS	19-44 YEARS	45-64 YEARS	65+ YEARS
HOME	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
WORK/SCHOOL	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
OTHER	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]

12. How many people did you have physical contact with yesterday? **Refused to answer** □

	0-4 YEARS	5-18 YEARS	19-44 YEARS	45-64 YEARS	65+ YEARS
HOME	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
WORK/SCHOOL	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]
OTHER	[__|__]	[__|__]	[__|__]	[__|__]	[__|__]

13. How many people (including you) have stayed in your house in the last 7 days: [__|__]

8.4 SUPPLEMENTARY TABLES

Heterosubtypic Titre Change between Visit 1 and Visit 2 of Historic Strains. Subjects infected with H1N1 and H3N2 included in multivariable linear regression.

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	-0.235 (-0.359--0.11)	<0.001	0.116	-0.226 (-0.359--0.093)	0.001	0.188
Study Week	0.008 (-0.005-0.021)	0.217	0.005	0.007 (-0.006-0.02)	0.310	
Days since symptom onset	0.034 (-0.016-0.083)	0.179	0.008	0.017 (-0.033-0.066)	0.505	
Age	0.001 (-0.018-0.02)	0.932	-0.010	0.002 (-0.018-0.021)	0.876	
Gender	0.283 (-0.147-0.714)	0.194	0.007	0.19 (-0.263-0.643)	0.407	
Influenza Vaccination	-0.631 (-2.161-0.899)	0.415	-0.003	-0.767 (-2.257-0.723)	0.309	
Current Smoker	0.051 (-0.636-0.738)	0.883	-0.010	-0.167 (-0.881-0.547)	0.644	
Paracetamol at Baseline	0.363 (-0.169-0.895)	0.179	0.008	0.473 (-0.085-1.031)	0.096	
Antibacterial at Baseline	-0.02 (-0.45-0.41)	0.926	-0.010	-0.129 (-0.601-0.344)	0.589	
Any Respiratory Disease	-0.502 (-2.034-1.029)	0.517	-0.006	-0.303 (-1.853-1.246)	0.698	
Current Pregnancy	0.494 (-1.123-2.111)	0.540	-0.015			

Table 1 Heterosubtypic Response H1N1 1918

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	-0.138 (-0.238--0.038)	0.008	0.061	-0.114 (-0.244-0.016)	0.084	0.123
Study Week	0 (-0.011-0.012)	0.973	-0.010	0.001 (-0.012-0.014)	0.883	
Days since symptom onset	0.03 (-0.015-0.075)	0.187	0.008	0.027 (-0.019-0.073)	0.241	
Age	-0.015 (-0.032-0.002)	0.082	0.021	-0.007 (-0.029-0.016)	0.544	
Gender	0.241 (-0.15-0.633)	0.224	0.005	0.195 (-0.234-0.624)	0.369	
Influenza Vaccination	-0.667 (-2.055-0.721)	0.342	-0.001	-0.502 (-1.912-0.908)	0.481	
Current Smoker	0.031 (-0.593-0.655)	0.922	-0.010	-0.127 (-0.802-0.548)	0.710	
Paracetamol at Baseline	0.234 (-0.252-0.72)	0.342	-0.001	0.402 (-0.125-0.93)	0.133	
Antibacterial at Baseline	-0.013 (-0.404-0.377)	0.946	-0.010	-0.181 (-0.626-0.263)	0.420	
Any Respiratory Disease	0.063 (-1.331-1.457)	0.929	-0.010	-0.179 (-1.642-1.284)	0.808	
Current Pregnancy	0.344 (-1.008-1.697)	0.610	-0.018			

Table 2 Heterosubtypic Response H1N1 1977

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	-0.19 (-0.274--0.105)	<0.001	0.160	-0.181 (-0.271--0.092)	<0.001	0.22
Study Week	0.005 (-0.003-0.013)	0.213	0.006	0.006 (-0.002-0.015)	0.117	
Days since symptom onset	0.011 (-0.019-0.042)	0.473	-0.005	0.006 (-0.023-0.035)	0.686	
Age	0.009 (-0.003-0.02)	0.138	0.012	0.004 (-0.007-0.016)	0.447	
Gender	0.122 (-0.143-0.387)	0.362	-0.002	0.108 (-0.165-0.38)	0.434	
Influenza Vaccination	-0.19 (-1.13-0.751)	0.690	-0.009	-0.278 (-1.173-0.618)	0.540	
Current Smoker	-0.095 (-0.516-0.325)	0.654	-0.008	-0.026 (-0.463-0.41)	0.905	
Paracetamol at Baseline	-0.029 (-0.358-0.301)	0.863	-0.010	-0.012 (-0.347-0.323)	0.945	
Antibacterial at Baseline	-0.103 (-0.366-0.16)	0.437	-0.004	-0.089 (-0.37-0.193)	0.533	
Any Respiratory Disease	-0.588 (-1.522-0.346)	0.214	0.006	-0.275 (-1.207-0.657)	0.559	
Current Pregnancy	-0.012 (-0.809-0.785)	0.976	-0.024			

Table 3 Heterosubtypic Response H1N1 1999

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	*-0.309 (-0.409--0.208)*	*<0.001*	*0.268*	*-0.274 (-0.388--0.159)*	*<0.001*	0.344
Study Week	0.001 (-0.01-0.012)	0.843	-0.010	0.002 (-0.009-0.012)	0.753	
Days since symptom onset	0.035 (-0.009-0.078)	0.120	0.015	0.022 (-0.017-0.061)	0.260	
Age	0.017 (0-0.033)	0.048	0.030	-0.003 (-0.019-0.014)	0.733	
Gender	0.191 (-0.19-0.572)	0.323	0.000	0.22 (-0.146-0.585)	0.235	
Influenza Vaccination	-0.038 (-1.393-1.317)	0.956	-0.010	-0.21 (-1.392-0.972)	0.725	
Current Smoker	-0.399 (-1-0.202)	0.191	0.007	-0.464 (-1.041-0.113)	0.113	
Paracetamol at Baseline	0.179 (-0.294-0.652)	0.454	-0.004	0.392 (-0.051-0.834)	0.082	
Antibacterial at Baseline	-0.348 (-0.722-0.025)	0.067	0.024	-0.344 (-0.714-0.027)	0.069	
Any Respiratory Disease	-0.9 (-2.243-0.443)	0.187	0.008	-0.635 (-1.868-0.599)	0.309	
Current Pregnancy	0.443 (-0.68-1.567)	0.430	-0.009			

Table 4 Heterosubtypic Response H1N1 2007

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	-0.156 (-0.379-0.068)	0.156	0.091	*-0.753 (-0.942--0.565)*	*<0.001*	0.513
Study Week	0 (-0.013-0.012)	0.949	-0.083	0.002 (-0.012-0.016)	0.788	
Days since symptom onset	0.042 (-0.093-0.177)	0.511	-0.043	0.026 (-0.026-0.079)	0.318	
Age	-0.018 (-0.042-0.006)	0.132	0.110	0.013 (-0.013-0.038)	0.321	
Gender	-0.253 (-0.962-0.457)	0.453	-0.032	0.223 (-0.261-0.707)	0.362	
Influenza Vaccination	-	-	-	0.44 (-1.172-2.051)	0.589	
Current Smoker	-	-	-	-0.381 (-1.15-0.388)	0.328	
Paracetamol at Baseline	0.808 (-0.493-2.109)	0.201	0.060	0.552 (-0.049-1.154)	0.071	
Antibacterial at Baseline	-0.216 (-0.93-0.499)	0.523	-0.046	-0.002 (-0.507-0.502)	0.993	
Any Respiratory Disease	0.056 (-0.971-1.084)	0.907	-0.082	-0.533 (-2.196-1.129)	0.526	
Current Pregnancy	-	-	-	-	-	-

Table 5 Heterosubtypic Response H3N2 1968

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	*-0.459 (-0.878--0.041)*	*0.034*	*0.266*	-0.496 (-1.631-0.639)	0.312	0.713
Study Week	-0.002 (-0.024-0.021)	0.876	-0.081	0.008 (-0.023-0.039)	0.546	
Days since symptom onset	-0.093 (-0.334-0.148)	0.417	-0.023	-0.153 (-0.543-0.236)	0.359	
Age	0.007 (-0.041-0.055)	0.751	-0.074	-0.047 (-0.189-0.094)	0.427	
Gender	-1.062 (-2.188-0.064)	0.062	0.199	-0.596 (-2.466-1.275)	0.450	
Influenza Vaccination	-	-	-	-	-	-
Current Smoker	-	-	-	-	-	-
Paracetamol at Baseline	1.094 (-1.326-3.515)	0.344	-0.002	-1.547 (-8.803-5.708)	0.607	
Antibacterial at Baseline	-0.32 (-1.614-0.974)	0.600	-0.058	0.282 (-1.791-2.354)	0.741	
Any Respiratory Disease	1.503 (-0.09-3.095)	0.062	0.199	1.856 (-0.946-4.658)	0.149	
Current Pregnancy	-	-	-	-	-	-

Table 6 Heterosubtypic Response H3N2 2003

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	*-0.313 (-0.519--0.107)*	*0.006*	*0.435*	-0.272 (-1.053-0.509)	0.411	0.527
Study Week	-0.007 (-0.022-0.008)	0.317	0.007	0.002 (-0.024-0.028)	0.833	0.527
Days since symptom onset	0.043 (-0.124-0.211)	0.585	-0.056	0.002 (-0.378-0.381)	0.991	0.527
Age	0.001 (-0.032-0.034)	0.949	-0.083	0.002 (-0.114-0.117)	0.969	0.527
Gender	-0.39 (-1.251-0.472)	0.344	-0.002	-0.065 (-1.854-1.725)	0.929	0.527
Influenza Vaccination	-	-	-	-	-	-
Current Smoker	-	-	-	-	-	-
Paracetamol at Baseline	0.828 (-0.812-2.468)	0.293	0.016	0.356 (-5.44-6.151)	0.881	0.527
Antibacterial at Baseline	-0.527 (-1.359-0.305)	0.192	0.065	-0.402 (-2.211-1.407)	0.592	0.527
Any Respiratory Disease	0.628 (-0.576-1.831)	0.278	0.022	-0.139 (-2.661-2.382)	0.893	0.527
Current Pregnancy	-	-	-	-	-	-

Table 7 Heterosubtypic Response H3N2 2005

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	*-0.259 (-0.513--0.005)*	*0.046*	*0.232*	-0.143 (-1.027-0.742)	0.696	0.441
Study Week	-0.007 (-0.021-0.007)	0.308	0.010	-0.002 (-0.029-0.025)	0.852	
Days since symptom onset	0.037 (-0.119-0.192)	0.615	-0.060	0.082 (-0.242-0.405)	0.545	
Age	-0.003 (-0.034-0.027)	0.821	-0.079	0.026 (-0.073-0.125)	0.523	
Gender	-0.428 (-1.213-0.356)	0.257	0.031	-0.409 (-2.086-1.267)	0.558	
Influenza Vaccination	-	-	-	-	-	-
Current Smoker	-	-	-	-	-	-
Paracetamol at Baseline	0.784 (-0.732-2.3)	0.282	0.020	0.963 (-4.61-6.535)	0.676	
Antibacterial at Baseline	-0.289 (-1.099-0.52)	0.451	-0.031	-0.292 (-2.081-1.497)	0.693	
Any Respiratory Disease	0.093 (-1.078-1.265)	0.865	-0.081	-0.911 (-3.4-1.577)	0.390	
Current Pregnancy	-	-	-	-	-	-

Table 8 Heterosubtypic Response H3N2 2007

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	*-0.345 (-0.656--0.034)*	*0.032*	*0.272*	-0.54 (-1.901-0.82)	0.354	0.385
Study Week	-0.008 (-0.03-0.014)	0.443	-0.029	0.003 (-0.041-0.046)	0.886	
Days since symptom onset	0.032 (-0.211-0.274)	0.781	-0.076	-0.015 (-0.57-0.539)	0.946	
Age	0.005 (-0.042-0.052)	0.828	-0.079	0.002 (-0.17-0.173)	0.979	
Gender	-0.433 (-1.687-0.821)	0.466	-0.035	-0.063 (-2.783-2.656)	0.955	
Influenza Vaccination	-	-	-	-	-	-
Current Smoker	-	-	-	-	-	-
Paracetamol at Baseline	0.963 (-1.427-3.353)	0.397	-0.018	-0.94 (-11.193-9.314)	0.823	
Antibacterial at Baseline	-0.348 (-1.612-0.916)	0.560	-0.052	0.235 (-2.927-3.397)	0.856	
Any Respiratory Disease	0.582 (-1.195-2.359)	0.489	-0.039	-0.548 (-4.579-3.484)	0.741	
Current Pregnancy	-	-	-	-	-	-

Table 9 Heterosubtypic Response H3N2 2009

Homosubtypic Titre Change between Visit 1 and Visit 2 of Historic Strains. Subjects infected with H1N1 and H3N2 included in multivariable linear regression.

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R^2	Estimate (95% CI)	p value	R^2
Titre at Baseline	-0.714 (-1.312--0.116)	0.023	0.308	-0.429 (-1.793-0.935)	0.455	0.591
Study Week	-0.01 (-0.047-0.027)	0.583	-0.055	0.017 (-0.042-0.077)	0.485	
Days since symptom onset	0.085 (-0.319-0.489)	0.655	-0.065	-0.179 (-0.819-0.461)	0.503	
Age	-0.048 (-0.121-0.025)	0.175	0.077	-0.067 (-0.241-0.106)	0.363	
Gender	-0.592 (-2.707-1.524)	0.554	-0.051	0.923 (-2.503-4.349)	0.519	
Influenza Vaccination	-	-	-	-	-	-
Current Smoker	-	-	-	-	-	-
Paracetamol at Baseline	2.9 (-0.803-6.603)	0.114	0.128	1.217 (-9.059-11.493)	0.773	
Antibacterial at Baseline	-1.014 (-3.066-1.037)	0.303	0.012	-1.492 (-5.179-2.195)	0.346	
Any Respiratory Disease	1.294 (-1.633-4.221)	0.354	-0.006	1.558 (-4.097-7.213)	0.510	
Current Pregnancy	-	-	-	-	-	-

Table 1 Homosubtypic Response H1N1 1918

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R^2	Estimate (95% CI)	p value	R^2
Titre at Baseline	-0.521 (-0.843--0.199)	0.004	0.467	-0.575 (-1.165-0.016)	0.054	0.754
Study Week	-0.011 (-0.033-0.011)	0.303	0.012	-0.002 (-0.029-0.026)	0.893	
Days since symptom onset	0.185 (-0.036-0.405)	0.093	0.152	0.011 (-0.294-0.317)	0.929	
Age	-0.032 (-0.076-0.012)	0.136	0.107	0.005 (-0.082-0.093)	0.879	
Gender	0.435 (-0.851-1.721)	0.475	-0.036	0.915 (-0.681-2.511)	0.201	
Influenza Vaccination	-	-	-	-	-	-
Current Smoker	-	-	-	-	-	-
Paracetamol at Baseline	1.398 (-0.971-3.767)	0.223	0.048	0.701 (-3.589-4.991)	0.692	
Antibacterial at Baseline	-0.295 (-1.597-1.007)	0.630	-0.062	-0.683 (-2.372-1.006)	0.346	
Any Respiratory Disease	0.049 (-1.81-1.909)	0.955	-0.083	-0.033 (-2.644-2.577)	0.975	
Current Pregnancy	-	-	-	-	-	-

Table 2 Homosubtypic Response H1N1 1977

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R^2	Estimate (95% CI)	p value	R^2
Titre at Baseline	-0.21 (-0.678-0.259)	0.349	-0.004	-0.156 (-1.279-0.967)	0.736	0.484
Study Week	-0.012 (-0.032-0.009)	0.231	0.043	-0.001 (-0.042-0.04)	0.967	
Days since symptom onset	0.043 (-0.191-0.277)	0.696	-0.069	-0.132 (-0.56-0.297)	0.465	
Age	-0.027 (-0.069-0.016)	0.196	0.063	-0.051 (-0.18-0.077)	0.351	
Gender	-0.093 (-1.334-1.148)	0.873	-0.081	0.7 (-1.478-2.878)	0.446	
Influenza Vaccination	-	-	-	-	-	-
Current Smoker	-	-	-	-	-	-
Paracetamol at Baseline	1.457 (-0.747-3.661)	0.175	0.076	0.379 (-7.626-8.383)	0.908	
Antibacterial at Baseline	-0.628 (-1.805-0.55)	0.268	0.026	-0.762 (-3.2-1.675)	0.458	
Any Respiratory Disease	0.629 (-1.083-2.34)	0.439	-0.028	0.727 (-2.926-4.38)	0.631	
Current Pregnancy	-	-	-	-	-	-

Table 3 Homosubtypic Response H1N1 1999

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	-0.098 (-0.539-0.343)	0.637	-0.063	0.045 (-1.256-1.346)	0.933	0.427
Study Week	-0.001 (-0.023-0.02)	0.900	-0.082	0.01 (-0.033-0.054)	0.575	
Days since symptom onset	-0.133 (-0.35-0.084)	0.206	0.057	-0.269 (-0.698-0.16)	0.168	
Age	-0.011 (-0.056-0.033)	0.592	-0.057	-0.041 (-0.184-0.103)	0.497	
Gender	-0.399 (-1.598-0.801)	0.483	-0.038	0.529 (-1.737-2.795)	0.575	
Influenza Vaccination	-	-	-	-	-	-
Current Smoker	-	-	-	-	-	-
Paracetamol at Baseline	0.649 (-1.669-2.968)	0.553	-0.051	0.951 (-8.277-10.178)	0.802	
Antibacterial at Baseline	-0.309 (-1.519-0.9)	0.588	-0.056	-0.616 (-3.153-1.922)	0.560	
Any Respiratory Disease	0.371 (-1.346-2.088)	0.646	-0.064	1.167 (-2.839-5.172)	0.488	
Current Pregnancy	-	-	-	-	-	-

Table 4 Homosubtypic Response H1N1 2007

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	*-0.68 (-0.826--0.534)*	*<0.001*	*0.460*	*-0.753 (-0.942--0.565)*	*<0.001*	*0.513*
Study Week	0.006 (-0.011-0.024)	0.490	-0.005	0.002 (-0.012-0.016)	0.788	
Days since symptom onset	0.014 (-0.055-0.084)	0.682	-0.008	0.026 (-0.026-0.079)	0.318	
Age	*-0.047 (-0.072--0.021)*	*<0.001*	*0.113*	0.013 (-0.013-0.038)	0.321	
Gender	0.466 (-0.13-1.062)	0.124	0.014	0.223 (-0.261-0.707)	0.362	
Influenza Vaccination	-0.252 (-2.386-1.882)	0.815	-0.010	0.44 (-1.172-2.051)	0.589	
Current Smoker	-0.033 (-0.988-0.922)	0.945	-0.010	-0.381 (-1.15-0.388)	0.328	
Paracetamol at Baseline	0.425 (-0.317-1.168)	0.258	0.003	0.552 (-0.049-1.154)	0.071	
Antibacterial at Baseline	0.576 (-0.01-1.163)	0.054	0.027	-0.002 (-0.507-0.502)	0.993	
Any Respiratory Disease	0.48 (-1.652-2.613)	0.656	-0.008	-0.533 (-2.196-1.129)	0.526	
Current Pregnancy	-0.883 (-3.065-1.299)	0.418	-0.008	-	-	-

Table 5 Homosubtypic Response H3N2 1968

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R²	Estimate (95% CI)	p value	R²
Titre at Baseline	*-0.896 (-0.947--0.844)*	*<0.001*	*0.923*	*-0.897 (-0.952--0.843)*	*<0.001*	*0.931*
Study Week	0.007 (-0.01-0.023)	0.419	-0.003	0.002 (-0.003-0.007)	0.442	
Days since symptom onset	0.05 (-0.012-0.113)	0.113	0.015	0.017 (-0.001-0.035)	0.062	
Age	*0.027 (0.003-0.051)*	*0.026*	*0.040*	-0.004 (-0.012-0.003)	0.236	
Gender	-0.007 (-0.557-0.543)	0.979	-0.010	-0.06 (-0.226-0.106)	0.475	
Influenza Vaccination	-0.292 (-2.237-1.652)	0.766	-0.009	-0.324 (-0.876-0.228)	0.247	
Current Smoker	0.017 (-0.853-0.888)	0.969	-0.010	0.041 (-0.225-0.306)	0.762	
Paracetamol at Baseline	0.303 (-0.376-0.981)	0.378	-0.002	0.104 (-0.103-0.31)	0.322	
Antibacterial at Baseline	0.255 (-0.288-0.797)	0.354	-0.001	-0.026 (-0.201-0.148)	0.767	
Any Respiratory Disease	-0.901 (-2.838-1.036)	0.358	-0.002	0.059 (-0.513-0.631)	0.838	
Current Pregnancy	-0.143 (-2.14-1.855)	0.886	-0.024	-	-	-

Table 6 Homosubtypic Response H3N2 2003

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R^2	Estimate (95% CI)	p value	R^2
Titre at Baseline	-0.844 (-0.911--0.776)	<0.001	0.860	-0.854 (-0.927--0.782)	<0.001	0.874
Study Week	-0.003 (-0.017-0.011)	0.654	-0.008	0.001 (-0.005-0.007)	0.642	
Days since symptom onset	0.055 (0-0.109)	0.048	0.029	0.024 (0.003-0.045)	0.027	
Age	0.023 (0.002-0.044)	0.033	0.036	-0.007 (-0.016-0.002)	0.137	
Gender	0.028 (-0.455-0.51)	0.910	-0.010	-0.11 (-0.307-0.086)	0.268	
Influenza Vaccination	-0.661 (-2.361-1.039)	0.442	-0.004	0.107 (-0.548-0.763)	0.746	
Current Smoker	0.163 (-0.599-0.925)	0.673	-0.008	0.048 (-0.265-0.361)	0.761	
Paracetamol at Baseline	-0.006 (-0.603-0.59)	0.983	-0.010	0.032 (-0.213-0.276)	0.797	
Antibacterial at Baseline	-0.017 (-0.495-0.461)	0.943	-0.010	-0.031 (-0.236-0.174)	0.764	
Any Respiratory Disease	-0.645 (-2.345-1.055)	0.453	-0.004	0.114 (-0.563-0.79)	0.739	
Current Pregnancy	-0.124 (-1.985-1.737)	0.894	-0.024	-	-	-

Table 7 Homosubtypic Response H3N2 2005

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R^2	Estimate (95% CI)	p value	R^2
Titre at Baseline	-0.71 (-0.828--0.591)	<0.001	0.586	-0.717 (-0.838--0.596)	<0.001	0.629
Study Week	-0.006 (-0.019-0.006)	0.333	-0.001	-0.001 (-0.01-0.008)	0.804	
Days since symptom onset	0.017 (-0.033-0.067)	0.503	-0.006	0.017 (-0.016-0.05)	0.309	
Age	-0.006 (-0.025-0.013)	0.546	-0.006	-0.013 (-0.026-0)	0.054	
Gender	0.081 (-0.354-0.515)	0.714	-0.009	-0.011 (-0.314-0.293)	0.945	
Influenza Vaccination	-1.005 (-2.53-0.52)	0.194	0.007	-0.216 (-1.231-0.799)	0.673	
Current Smoker	0.015 (-0.673-0.704)	0.964	-0.010	-0.065 (-0.549-0.419)	0.791	
Paracetamol at Baseline	0.275 (-0.261-0.811)	0.311	0.000	0.319 (-0.059-0.696)	0.097	
Antibacterial at Baseline	0.182 (-0.247-0.612)	0.402	-0.003	0.039 (-0.277-0.356)	0.806	
Any Respiratory Disease	0.184 (-1.354-1.721)	0.813	-0.010	0.206 (-0.84-1.253)	0.696	
Current Pregnancy	-0.739 (-2.414-0.937)	0.378	-0.005	-	-	

Table 8 Homosubtypic Response H3N2 2007

	Univariate			Multivariable		
	Estimate (95% CI)	p value	R^2	Estimate (95% CI)	p value	R^2
Titre at Baseline	-0.794 (-0.901--0.687)	<0.001	0.690	-0.801 (-0.912--0.69)	<0.001	0.724
Study Week	-0.01 (-0.025-0.004)	0.153	0.011	0.003 (-0.006-0.012)	0.521	
Days since symptom onset	0.027 (-0.03-0.083)	0.351	-0.001	0.012 (-0.02-0.044)	0.457	
Age	-0.008 (-0.03-0.014)	0.481	-0.005	-0.016 (-0.029--0.003)	0.020	
Gender	0.071 (-0.422-0.564)	0.776	-0.010	-0.024 (-0.324-0.275)	0.873	
Influenza Vaccination	-1.578 (-3.279-0.124)	0.069	0.024	-0.642 (-1.628-0.344)	0.199	
Current Smoker	-0.046 (-0.821-0.729)	0.906	-0.010	0.105 (-0.372-0.582)	0.662	
Paracetamol at Baseline	0.071 (-0.536-0.679)	0.816	-0.010	0.116 (-0.254-0.486)	0.536	
Antibacterial at Baseline	-0.024 (-0.514-0.466)	0.921	-0.010	-0.047 (-0.361-0.268)	0.770	
Any Respiratory Disease	0.598 (-1.129-2.325)	0.493	-0.005	0.266 (-0.752-1.284)	0.604	

Current Pregnancy	-0.461 (-2.355-1.433)	0.626	-0.018	-	-

Table 9 Homosubtypic Response H3N2 2009

8.5 R CODE

Thesis R Code

Stacy Todd

28 March 2016

Resampling Function used in Chapter 4 for Piecewise Linear Regression with Random Effects. Adapted from http://www.ats.ucla.edu/stat/r/dae/melogit.htm (http://www.ats.ucla.edu/stat/r/dae/melogit.htm)

```r
resample <- function(dat, cluster, replace) {

  # exit early for trivial data
  if(nrow(dat) == 1 || all(replace==FALSE))
      return(dat)

  # sample the clustering factor
  cls <- sample(unique(dat[[cluster[1]]]), replace=replace[1])

  # subset on the sampled clustering factors
  sub <- lapply(cls, function(b) subset(dat, dat[[cluster[1]]]==b))

  # sample lower levels of hierarchy (if any)
  if(length(cluster) > 1)
    sub <- lapply(sub, resample, cluster=cluster[-1], replace=replace[-1])

  n.sub <- seq(1, length(sub), 1)

  NewID <- unlist(mapply(function(x, y){rep(y, nrow(x))}, x=sub, y=n.sub))
  # join and return samples
  cbind(NewID, do.call(rbind, sub))

}
```

Piecewise linear regression with a random effects from Chapter 4. Adapted from segmented package.

```r
# One Break Point
#sample of ParNo but no replacement for VisitNo

H3only <- serology[which(serology$FluSubtype=="H3"),]



boot.break.RE <- function(data, formula){
 d <- resample(dat=data, cluster=c("ParNo", "VisitNo"), replace=c(T,F))

 foo <- function(bp){
    mod = lmer(formula, data = d, REML=FALSE)
    deviance(mod)
 }
search.range <- c(10, 210)
foo.opt <- optimize(foo, interval = search.range)
bp <- foo.opt$minimum
return(bp)}

H3only <- serology[which(serology$FluSubtype=="H3"),]

res.break1.RE <- list()
for(i in 1:6){
   st <- human.strains[5+i]
   formula1 <- paste(st, "~ FUTime+ I(pmax(FUTime-bp,0)) + (1|NewID)", collapse="")

   res.strain <- list()

   for(j in 1:1000){
       d <- resample(dat=H3only, cluster=c("ParNo", "VisitNo"), replace=c(T,F))

      foo <- function(bp){
         mod = lmer(as.formula(formula1), data = d, REML=FALSE)
         deviance(mod)
      }
     search.range <- c(3, 210)
     foo.opt <- tryCatch(optimize(foo, interval = search.range), error = function(e){N
A})
   if(is.na(foo.opt)){
     return(NA)
   } else {
     bp <- foo.opt$minimum
     res.strain[[j]] <-bp
     }}

   x <- do.call(rbind, res.strain)
   resEst <- round(mean(x, na.rm=T),1)
   resCILo <- round(quantile(x, prob=c(0.025)),1)
   resCIHi <- round(quantile(x, prob=c(0.975)),1)

   bestformula <- lmer(paste(st, "~ FUTime + I(pmax(FUTime-", resEst,",0)) + (1|ParNo)
", collapse=""),
                       data=H3only)
```

```
        res.break1.RE[[i]] <- c("resEst"=resEst, "resCILo"=resCILo, "resCIHi"=resCIHi,
               "AIC"=AIC(bestformula),
               "aRsqMarg"=r.squaredGLMM(bestformula)[1],
               "aRsqCond"=r.squaredGLMM(bestformula)[2],
               "n"=sum(!is.na(res.strain)))
}

  ttext <- as.data.frame(do.call("rbind", res.break1.RE))


  break1.RE.table <- data.frame("Strain" = human.strains[6:11],
     "Break1" = paste(ttext$resEst, " (", ttext$resCILo, "-", ttext$resCIHi, ")", sep
=""),
     "AIC" =ttext$AIC, "aRsqM"=ttext$aRsqMarg, "aRsqC"=ttext$aRsqCond,
     "n" = ttext$n)

  #wdGet()
  #  wdTable(break1.RE.table, caption="", row.names=FALSE)
     pander( break1.RE.table,split.table=160, emphasize.rownames=FALSE )
```

```r
#Two Break Points
#sample of ParNo but no replacement for VisitNo
H3only <- serology[which(serology$FluSubtype=="H3"),]

res.break2.RE <- list()
for(i in 4:6){
  st <- human.strains[5+i]

  res.strain <- list()

  for(j in 1:1000){
   d <- resample(dat=H3only, cluster=c("ParNo", "VisitNo"), replace=c(T,F))

   formula1 <- paste(st, "~ FUTime+ I(pmax(FUTime-bp1,0)) + (1|NewID)", collapse="")

    foo1 <- function(bp1){
        mod = lmer(as.formula(formula1), data = d, REML=FALSE)
        deviance(mod)
     }
    search.range <- c(3, 210)
    foo.opt1 <- tryCatch(optimize(foo1, interval = search.range), error = function(e)
{NA})
    if(is.na(foo.opt1)){
      bp1 <- NA
    } else {
      bp1 <- foo.opt1$minimum
    }

    formula2 <- paste(st, "~ FUTime+ I(pmax(FUTime-", bp1, ",0)) + I(pmax(FUTime-bp2,
0))+ (1|NewID)", collapse="")
       foo2 <- function(bp2){
        mod = lmer(as.formula(formula2), data = d, REML=FALSE)
        deviance(mod)
     }
    search.range <- c(bp1+5, 210)
    foo.opt2 <- tryCatch(optimize(foo2, interval = search.range), error = function(e)
{NA})
    if(is.na(foo.opt2)){
      bp2 <- NA
    } else {
      bp2 <- foo.opt2$minimum
    }
    res.strain[[j]] <- cbind(bp1, bp2)
    }

  x <- do.call(rbind, res.strain)
   resbp1Est <- round(mean(x[,1], na.rm=T),1)
   resbp1CILo <- round(quantile(x[,1], prob=c(0.025)),1)
   resbp1CIHi <- round(quantile(x[,1], prob=c(0.975)),1)
   resbp2Est <- round(mean(x[,2], na.rm=T),1)
   resbp2CILo <- round(quantile(x[,2], prob=c(0.025)),1)
   resbp2CIHi <- round(quantile(x[,2], prob=c(0.975)),1)
```

```
        bestformula <- lmer(paste(st, "~ FUTime + I(pmax(FUTime-", resbp1Est,",0)) +  I(pma
x(FUTime-", resbp2Est, ",0)) + (1|ParNo) ", collapse=""),
                            data=H3only)


    res.break2.RE[[i]] <- c("resbp1Est"=resbp1Est, "resbp1CILo"=resbp1CILo, "resbp1CIH
i"=resbp1CIHi,
                            "resbp2Est"=resbp2Est, "resbp2CILo"=resbp2CILo, "resbp2CIH
i"=resbp2CIHi,
            "AIC"=AIC(bestformula),
            "aRsqMarg"=r.squaredGLMM(bestformula)[1],
            "aRsqCond"=r.squaredGLMM(bestformula)[2],
            "nbp1"=sum(!is.na(x[,1])), "nbp2"=sum(!is.na(x[,2])))
}



  ttext <- as.data.frame(do.call("rbind", res.break2.RE))
  names(ttext) <- c("resbp1Est", "resbp1CILo", "resbp1CIHi", "resbp2Est", "resbp2CILo",
"resbp2CIHi",
                    "AIC", "aRsqM", "aRsqC", "nbp1", "nbp2")



  break2.RE.table <- data.frame("Strain" = human.strains[9:11],
    "Break1" = paste(ttext$resbp1Est, " (", ttext$resbp1CILo, "-", ttext$resbp1CIHi,
")", sep=""),
    "Break2" = paste(ttext$resbp2Est, " (", ttext$resbp2CILo, "-", ttext$resbp2CIHi,
")", sep=""),
    "AIC" =round(ttext$AIC, 2), "aRsqM"=round(ttext$aRsqM, 2), "aRsqC"=round(ttext$aRsq
C,2),
    "n1" = ttext$nbp1, "n2" = ttext$nbp2)

  wdGet()
 wdTable(break2.RE.table, caption="", row.names=FALSE)
    pander( break2.RE.table,split.table=160, emphasize.rownames=FALSE )
```

Modified Simpson's Diversity Index used in Chapter 5.

```r
# ecological diversity function from Jost (2006), equation (2)
# assumes minimum titer for anybody is a 10, which is 0 on a log-titer-scale
jost2006_GiniSimpson_diversity = function( v )
{
  q=2 # this is the order of the diversity measure
  sm = sum(v)
  if( sm==0.0 ) { return (0.0) }

  diversity_sum = 0.0

  for( i in 1:length(v) )
  {
    diversity_sum = diversity_sum + ((v[i]/sm)^q)
  }

  return( diversity_sum^(1/(1-q)) )

}

adapted_jost2006_GiniSimpson_diversity = function( vTiters, CutOff )
{
  dummy = CutOff
  m = max( vTiters )
 if( m > CutOff)
  {
   dummy = m
}
  v = c(vTiters,dummy)

  return( jost2006_GiniSimpson_diversity(v) - 1 )
}
```

Bayesian Sampling Framework Used in Chapter 5.

```r
  estTpos <-  c(lapply(res60d, function(x){x$res}),
               lapply(res100d, function(x){x$res}),
               lapply(res250d, function(x){x$res}))


 bayespriors <-c(lapply(res60d, function(x){x$summary}),
               lapply(res100d, function(x){x$summary}),
               lapply(res250d, function(x){x$summary}))


 thresh <- c(5.5, 6, 6.5, 7)



# This is the bayesian model to run in BUGS

  model2 <- function(){
    x ~ dbin(p, n)

    p <- pi*SE + (1-pi)*(1-SP)

    pi~ dunif(0,1)
    SE~ dunif(SensLO, SensHI)
    SP~ dunif(SpecLO, SpecHI)

  }

  bayes.res <- list()
  for(k in 1:12){
  # For each age and time period
  tmp.data <- estTpos[[k]]
  tmp.summary1 <- bayespriors[[k]]

  thresh.res <- list()
  for(i in 1:4){
    #For each threshold level
    tmp.thresh <- tmp.data[which(tmp.data==thresh[i]),]
    tmp.summary <- tmp.summary1[i,]

    prev.res <-  matrix(NA, nrow=12, ncol=4)

  for(j in 1:12){
    # For each prevalence
      x <- round(tmp.thresh[j, "Est"]*1000)
      n <- 1000
      SensLO <- as.numeric(tmp.summary["sensitivity.2.5."])
      SensHI <- as.numeric(tmp.summary["sensitivity.97.5."])
      SpecLO <- as.numeric(tmp.summary["specificity.2.5."])
      SpecHI <- as.numeric(tmp.summary["specificity.97.5."])

      model.file <- file.path(tempdir(),"model.txt")
      write.model(model2, model.file)
      data <- list("x", "n", "SensLO", "SensHI", "SpecLO", "SpecHI" )
      params <- c("pi")
      inits <- function() { list(p=0.5) }
      out <- bugs(data, inits, params, model.file,
```

```
                         n.chains=2,
                         n.burnin=1000,
                         n.iter=6000#,
                         #codaPkg=TRUE
                         )
        prev.res[j,] <- out$summary[1,c(1,3,7:8)]
    }
      thresh.res[[i]] <- data.frame("Thresh"= rep(thresh[i], 12),
                                   "Prev" = c(0.01, 0.05, seq(0.1, 1, 0.1)),
                                   "TposEst" = tmp.thresh[, "Est"],
                            prev.res)
    }
    bayes.res[[k]] <- do.call("rbind", thresh.res)
    }

    plot.data <- data.frame("time"= rep(c("Less than 60 days",
                         "Less than 100 days", "Less than 250 days"), each=192),
                         "Age.Cat"=rep(rep(c(age.class[3:5], "All Ages"),each=48),3),
                            do.call("rbind", bayes.res))
    plot.data$time <- factor(plot.data$time,
                                   levels=c("Less than 60 days", "Less than 100 days",
"Less than 250 days"))
    plot.data$Age.Cat <- factor(plot.data$Age.Cat,
                                   levels=c("All Ages", age.class[3:5]))
    names(plot.data)[6:9] <- c("dposEst", "dposCILo", "dposCIHi", "Rhat")
```