Evaluation of Data from an Urgent Out of Hours GP Service as a Potential Early Warning System for Influenza

Elaine D. Brabazon

A dissertation submitted to the University of Dublin,
in partial fulfillment of the requirements for the degree of
Master of Science in Health Informatics

2007

This thesis is dedicated to the memory of our baby son,

Tiarnán Brabazon Carton,

born 29th November 2006

"I will not forget you, I have carved you on the palm of my hand"

Acknowledgements

I would like to thank Dr. Lucy Hederman and Ms. Gaye Stephens for their support during this M.Sc. course.

Thanks are also due to all in the Department of Public Health (HSE-North East) but especially to Dr. Declan Bedford for helpful discussions and advice during this thesis and to Dr. Ciara Murray for her patience and rigorous proof-reading ability! Thanks are also extended to Ms. Denise McGranahan for her expertise in extracting data from the NE-Doc System. I am indebted to Dr. Mícheál Carton whose extensive knowledge of Microsoft Access was generously provided at crucial points during this project.

On a less formal note (!), love always to my best friend and husband, Mícheál, who is my rock and to my family who have always been there for me in good times and bad.

Declaration

I declare that the work described in this dissertation is, except where otherwise stated, entirely my own work, and has not been submitted as an exercise for a degree at this or any other university.

Signed: ___________________
Elaine D. Brabazon

Date: _____________________

Permission to lend and/or copy

I agree that the Trinity College Library may lend or copy this dissertation upon request.

Signed: ___________________
Elaine D. Brabazon

Date: _____________________

Abbreviations

Abbreviation	Explanation
CDC	Centre for Disease Control (USA)
CDSCNI	Communicable Disease Surveillance Centre, Northern Ireland
DD	Doctors Diagnosis Field
EISS	European Influenza Surveillance Scheme
GP	General Practitioner
HPA	Health Protection Agency (UK)
HPSC	Health Protection Surveillance Centre, formerly the NDSC (National Disease Surveillance Centre, Ireland)
HSE	Health Service Executive
ILI	Influenza like illness
NE-Doc	North East Doctor on Call Service
PRC	Patients Reported Condition Field
WHO	World Health Organisation

"In the highly interconnected and readily traversed 'global village' of our time, one nation's problem soon becomes every nation's problem"

From Smolinski et al., (2003): Microbial Threats to Health: Emergence, Detection, and Response

Summary

Introduction

With the ever increasing risk of a global influenza pandemic there is an onus on all nations to enhance influenza alert systems. These systems are essential to global public health as they aim to rapidly identify emerging outbreaks of infection and therefore allow early interventions which can impact on the spread of disease within the general population. Many countries are supplementing their routine influenza surveillance with new sources of information on influenza like illness (ILI) with, for example, healthcare call centre data. One such call centre service which covers the majority of the Irish population is the urgent out of hours GP system which is operated jointly by GPs and the HSE in the various regions. The aim of this thesis was to examine whether or not it would be possible to extract reliable and timely information on influenza related patient contacts with this service and to determine if this new data source has the potential to be utilised as a national influenza syndromic surveillance system, thus making a significant contribution to the overall surveillance of influenza in Ireland.

Methodology

Data on all calls, for three influenza seasons between 2003 and 2006, was requested from an urgent out of hours GP service covering the population of the North East and Midlands. Quantitative analysis of these data were conducted and Microsoft Access keyword queries were designed and validated to specifically extract influenza related call records from the free text patient reported condition field and doctors diagnosis field. These subsets of influenza related calls were compared graphically with current national influenza surveillance indicators, e.g. sentinel GP ILI consultations, using time series plots. Statistical analysis of the temporal relationship between surveillance datasets was assessed using the Spearmans Rank Correlation Coefficient test.

Findings

Over the three influenza seasons studied there were a total of 239,736 calls to the Doctor on Call service and over the same period the largest number of flu related calls extracted was 14,823 (6.2% of all calls to the service). Comparison of these flu related calls over time with national influenza indicators, in particular the sentinel GP ILI consultation rates and sentinel positive virological influenza swabs, demonstrated similar temporal patterns. Indeed in most instances, the peaks in influenza related calls occurred at least one week prior to the peaks in the national indicators. Statistical analysis between these datasets confirmed significant temporal correlation.

Conclusion

This thesis has demonstrated for the first time that data from an urgent out of hours GP service operating in a HSE region in Ireland can provide reliable and timely information on influenza activity in the community setting even in the absence of any clinical coding of the data. Furthermore, this data has the potential to act as an early warning of influenza activity in the community as it has been demonstrated that peaks in influenza related calls occur earlier than peaks in current national indicators. This work has described a scalable solution for the development of a system to collect syndromic influenza related data from all out of hours GP services in Ireland. Finally, the methodology described in this thesis, for the extraction of calls records from the urgent out of hours GP service dataset, affords an opportunity to realise benefits for other important areas of public health surveillance, beyond syndromic influenza surveillance.

TABLE OF CONTENTS

List of Figures & Tables

Chapter 1: Introduction

1.1 Introduction

We live in an era where the globe can be transversed in less than 24 hours and in which we are comfortable at the thoughts that advances in medical science can cure or treat most ailments. It is difficult, therefore, for us to imagine the dissemination of a highly transmissible infection which sweeps the globe within six months and which kills almost half of those it encounters. National borders are closed, public events are cancelled, schools and workplaces are deserted and the digging and filling of mass graves becomes a commonplace sight. This is not some futuristic tale – it is a true account of the devastation of the 1918 influenza pandemic and all leading public health authorities agree that the next influenza pandemic is just around the corner.

The requirement for accurate and timely monitoring of influenza activity has never been so important. Many countries are supplementing their routine influenza surveillance systems with extra influenza indicators, for example, healthcare call centre data, over the counter sales for particular drugs and absenteeism data from schools and workplaces. The expectation is that good pre-pandemic baseline data could provide essential epidemiological information for targeting at-risk groups for anti-retroviral therapy and vaccination. Analysis of the transmission of seasonal flu may provide an insight into the likely dissemination pattern of a pandemic flu strain perhaps allowing some form of intervention which may slow down the spread of disease within populations e.g. travel restrictions from epidemic areas. Finally collection of early pandemic data will only be possible if systems are designed, implemented and tested in a pre-pandemic phase. Clearly, in addition to directives from the World Health Organisation (WHO), every nation has an ethical responsibility to ensure that the best possible surveillance of influenza during this pre-pandemic period is maintained and enhanced.

1.2 Objectives

In light of the imminent public health crisis posed by influenza, this thesis aims to evaluate whether or not healthcare call centre data from an urgent out of hours "doctor on call" service in Ireland can be used as a early warning system for influenza surveillance.

This will be achieved by quantitative analysis of the data available from a typical "doctor on call" service covering two HSE regions in Ireland (north east and midlands) and by temporal comparison of influenza related records with routine surveillance indicators currently collected in Ireland. This study will be the first of its kind in Ireland to assess whether an urgent out of hours GP service can provide routine information on influenza activity in the general community and whether or not these data can provide more timely warning than current influenza surveillance methods. The results of this study will be presented to the Influenza Surveillance Subcommittee of the National Pandemic Influenza Expert Group.

1.3 Dissertation Structure

Chapter 2 presents the domain description for the epidemiology of the Influenza virus and provides an insight into the phenomenon of pandemic influenza. In addition, an overview of the discipline of infectious disease surveillance and the European and Irish legislation regarding infectious disease surveillance will be described.

Chapter 3 describes the state of the art in influenza surveillance in Ireland and other countries and will also review syndromic surveillance of influenza and influenza like illness with particular emphasis on healthcare call centre data. A potential source of healthcare call centre data in Ireland will also be described.

Chapter 4 outlines the methodology used in this study including literature review description, semi-structured interviewing, quantitative data analysis and design of a MS Access database for querying of free-text data from a "Doctor on Call" call centre service in Ireland.

Chapter 5 presents the results of the semi-structured interview and a description of the "Doctor on Call" service assessed in this study. Analysis of all calls to this service for the three influenza seasons (2003/2004, 2004/2005 and 2005/2006) will be presented. Finally, the sensitivity and specificity of queries designed to extract flu related records from these datasets will be assessed.

Chapter 6 details the temporal comparison of "Doctor on Call" flu related records with corresponding routine influenza surveillance indicators, for example, ILI GP consultations, influenza notifications and positive sentinel virological influenza data. The potential of the "Doctor on Call" data to provide an early warning for influenza or more timely information than current surveillance methods will be investigated.

Chapter 7 will contain a discussion of the results of the analysis of the "Doctor on Call" influenza related data and will outline the advantages and disadvantages of the data as a potential early warning system for influenza surveillance. The strengths and limitations of this study will be discussed and recommendations for incorporation of this new data source into routine influenza surveillance systems in Ireland will be described.

The Appendices will contain any information or data pertinent to this thesis but which was deemed too cumbersome to include in the general text. An example of the data included in this section includes the Dates and Week Numbers for all three Flu seasons and their corresponding nomenclature.

1.4. Conclusion

This thesis will describe for first time whether or not it is possible to use a "real-time", routinely collected, available data source on illness in the community (Doctor on Call service) for the surveillance of Influenza in Ireland. This urgent out of hours GP service which contains a huge repository of data has the potential to provide timely information on seasonal influenza outbreaks and may provide a valuable epidemiological baseline for comparison during the early stages of a pandemic.

Chapter 2: Domain Description – Influenza and Surveillance

2.1 Introduction

This chapter presents a domain description for the epidemiology of the influenza virus and also provides an overview of the discipline of infectious disease surveillance.

2.2 Impact of Infectious Disease

Throughout history, man has been plagued by a deadly foe that is not even visible to the naked eye – disease causing or pathogenic micro-organisms. These contagions have been responsible for many global pandemics and millions of deaths, including the typhoid fever epidemic in Athens circa 430BC (Papagrogorakis et al., 2006), the bubonic plague or "black death" in Europe during the 1300s killing an estimated 20 million (Riedel, 2005), the cholera pandemics of the 1800s (Kavic et al., 1999) and the virulent "Spanish flu" of 1918 which claimed at least 40 million lives globally (Reid & Taubenberg, 2003). Indeed even well known historical figures were not immune to the ravages of infectious disease: new evidence from the writings of Plutarch suggest that Alexander the Great died from West Nile virus encephalitis in 323BC (Marr & Calisher, 2003).

Today, man is no safer from epidemics – in recent years new emerging infectious diseases have swept across the globe. For example, the HIV/AIDS public health crisis was only first recognised in 1981 in the homosexual community in the US (Gottlieb, 1981). Transmission was subsequently shown to occur through contact with infected bodily fluids, i.e. through sexual intercourse, sharing of needles by intravenous drug users (IDU), treatment with infected blood products and vertical transmission mother to child. In 2005, HIV/AIDS was responsible for 2.8 million deaths worldwide (CDC 2006). Furthermore, the infection is no longer confined to the homosexual or IDU community, in many countries the heterosexual contraction rate has risen dramatically. Clearly the epidemiology of this 21st Century pandemic is progressing.

Another recent epidemic which brought infectious disease into sharp focus was the emergence of a novel coronavirus which caused Severe Acute Respiratory Syndrome (SARS). Originating in Southern China, the virus passed by human to human transmission into Hong Kong, VietNam and Singapore to name but a few of the Asian countries affected. Following this and with the help of air travel the virus was readily imported into Canada and Germany. SARS was ultimately responsible for 8,098 infections and 774 deaths in 26 countries in the 10-month period between November 2002 to July 2003 (WHO, 2006).

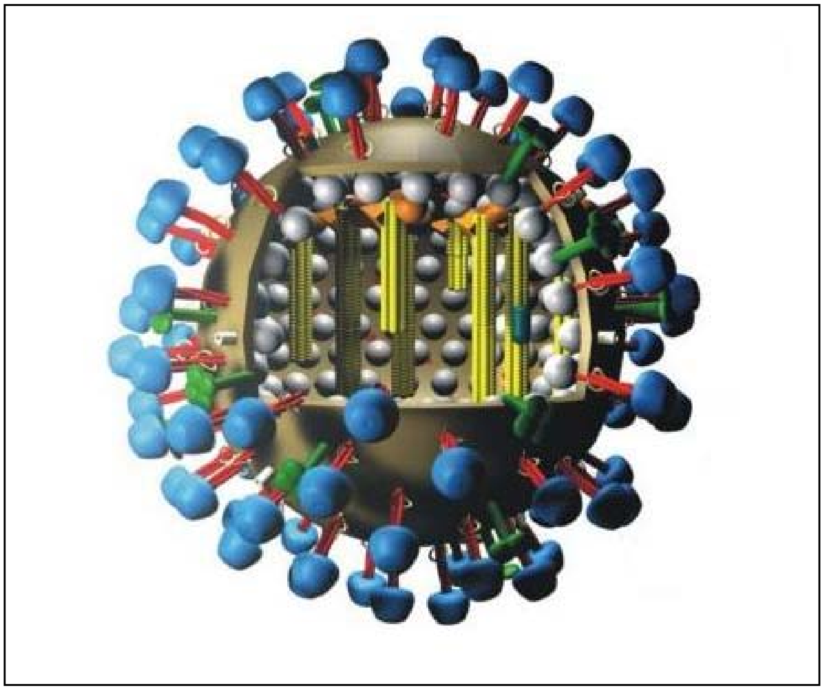
It is not always "new" emerging infectious diseases, however, that threaten public health. There is great concern that the next communicable disease to threaten mankind globally will be due to a common circulating virus: influenza.

2.3 The Influenza Virus

In the United States, seasonal influenza affects about 108 million Americans each year and is related to 20,000 to 40,000 deaths (Sullivan, 1996). Indeed, it is estimated for worldwide infection with influenza in any one year there are three to five million severe infections and 500,000 deaths (WHO, 2003). The virus circulates in the population through localised outbreaks, annual epidemics and less frequent but inevitable global pandemics. Common symptoms in humans include fever, nausea, sore throat, cough, muscle pain, weakness and fatigue. The virus is mainly transmitted in humans by coughing and sneezing. Infection in vulnerable populations, e.g. the elderly and children, often leads to pneumonia and death. Influenza generally reaches a peak in the winter months, probably due to the closer indoor contact among the general population.

The influenza virus, itself, is a negative stranded RNA virus of the family Orthomyxoviridiae and has the ability to infect all mammals and birds. The virus is approximately 80-100nm in diameter and has a spherical structure (see Fig. 2.1). There are three types of influenza virus, Influenza A, B and C. Types A and B are primarily responsible for all human infections. The virus has a segmented genome consisting of eight genes which encode 11 proteins, two of which are particularly well characterised surface glycoproteins – haemagluttinin (H) and neuraminidase (N). There are 16 haemagluttinin (labelled H1 – H16) and nine neuraminidase (labelled N1 – N9) protein variants and every strain of flu virus is named according to the protein variant it possesses, e.g. H5N1, H1N3 etc. Haemagluttinin is involved in the entry of the virus into the host cell and neuraminidase is directly involved in the release of progeny virus from

Fig. 2.1: Structure of the Influenza virus.



Taken from http://en.wikipedia.org/wiki/Influenza

Where 

represents the surface protein Haemogluttinin

represents the surface protein Neuraminidase an infected host cell. These proteins are, therefore, key components in the transmission and generation of the influenza virus.

The success of the influenza virus lies in its ability to undergo two distinct processes that allow the evolution of a new virus strain – antigenic shift and antigenic drift. Antigenic shift is the natural occurrence of mutations in particular genes producing a more virulent strain of the flu virus. Antigenic drift, however, is the recombination of genes between two distinct flu viruses giving rise to a more virulent hybrid influenza virus. These "new" viruses to which the human population has no natural immunity are typically responsible for the most severe of all influenza outbreaks – a global pandemic.

2.4 Pandemic Influenza

A pandemic is the worldwide dissemination of a highly transmissible infection. Influenza has been responsible for some of the worst pandemics in human history including the Asiatic Flu of 1889 and the Spanish flu of 1918. The latter pandemic was identified as H1N1 virus and was unusually deadly, striking down not only the vulnerable in the population but also healthy young adults with very severe symptoms:

> *"Totally different to anything ordinarily seen in the post mortem room, it involved a whole battery of impacts; the victims lungs were soaked, collapsed and pulped. Some sections would be so shattered that they were barely recognisable as lung tissue; others where the air sacs disintegrated, looked like bloody Gruyere. The victims would complain before they died that they were "all raw inside there"; they'd cough up bloody froth, or eight or ten ounces of pus in a day, and their breath rate would soar to 40, 50, even 60 breaths a minute. They died like victims of a gassing"*
> Comments by a noted physician on victims of 1918 "Spanish Flu" (Davies, Catching Cold)

It is estimated that the 1918 "Spanish" flu claimed more that 40 million lives. As this pandemic coincided with World War I, its impact in terms of deaths has never been fully appreciated by the general public. Two further but less ferocious human influenza pandemics have occurred in the 20th century – the Asian H2N2 Influenza of 1957 and the Hong Kong H3N2 Influenza of 1968. Together, these two pandemics were responsible for circa 100,000 deaths globally.

In the last decade, a new virus strain has emerged (H5N1) which was first identified in poultry flocks in the Guandong region of China in 1996. This avian flu virus was transmitted to 18 humans (who were living in close contact with infected birds) in 1997 resulting in six deaths. In January 2003, three people from one family were infected with this avian flu virus causing two fatalities. By the end of 2004 the virus had further spread in the poultry population in 10 Asian countries including China, Vietnam, and Indonesia. In 2005, this virus strain had spread to bird populations in Kazakhstan, Mongolia and Russia and finally was identified in a number of European bird populations including Romania and Croatia. In April 2006, a whooper swan was found to be infected with H5N1 in Cellardyke, UK. In 2007, a major outbreak of H5N1 was identified at a UK poultry farm in Holton, Suffolk. As this avian flu virus has spread through poultry and wild birds worldwide, there have been many human casualties too. As of 1st March 2007, a total of 277 people have been infected with this virus, of which, 167 or 60% have died (WHO, 2007a). A recombination or antigenic shift event between this avian flu virus, with such a high mortality rate, and an easily transmissible seasonal human flu virus, would produce the "new" feared global killer virus. This new pandemic virus is expected to have devastating consequences: the WHO has predicted (at a minimum) 134 – 233 million out-patient visits, 1.5 – 5.2 million hospital admissions and 2 – 7.4 million deaths worldwide (WHO, 2007b). In 2005, the WHO requested that all countries prepare national pandemic influenza plans based on their recommendations and begin the process of stockpiling anti-retroviral drugs. The WHO has designated 6 pandemic alert phases (see Table 2.1). Currently we are at Phase 3 – with evidence of human infection with a new influenza virus but no detectable human to human transmission. In Ireland a draft national pandemic plan was published for public consultation in January 2007 (HPSC, 2007a). In this report, a worst case pandemic

Table 2.1: Stages of an Influenza Pandemic.
Taken from "WHO global influenza prepardness plan, epidemic alert and response" WHO, 2005

PHASES	PUBLIC HEALTH GOALS
Interpandemic period	
Phase 1. No new influenza virus subtypes have been detected in humans. An influenza virus subtype that has caused human infection may be present in animals. If present in animals, risk of human infection or disease is considered to be low.	Strengthen influenza pandemic preparedness at the global, regional, national and the sub-national levels.
Phase 2. No new influenza virus subtypes have been detected in humans. However, a circulating animal influenza virus subtype poses a substantial risk of human disease.	Minimize the risk of transmission to humans; detect and report such transmission rapidly if it occurs.
Pandemic alert period	
Phase 3. Human infection(s) with a new subtype, but no human-to-human spread, or at most rare instances of spread to a close contact.	Ensure rapid characterization of the new virus subtype and early detection, notification and response to additional cases.
Phase 4. Small cluster(s) with limited human-to-human transmission but spread is highly localized, suggesting that the virus is not well adapted to humans.	Contain the new virus within limited foci or delay spread to gain time to implement preparedness measures, including vaccine development.
Phase 5. Larger cluster(s) but human-to-human spread still localized, suggesting that the virus is becoming increasingly better adapted to humans, but may not yet be fully transmissible (substantial pandemic risk).	Maximize efforts to contain or delay spread, to possibly avert a pandemic, and to gain time to implement pandemic response measures.
Pandemic period	
Phase 6. Pandemic: increased and sustained transmission in general population.	Minimize the impact of the pandemic.

scenario for Ireland based on clinical attack and mortality rates similar to the 1918 influenza pandemic estimated that there would be ~2 million influenza cases, ~78,000 influenza hospitalisations and ~52,000 influenza deaths. Clearly, in the event of a pandemic, with an overstretched healthcare system, worker absenteeism and public fear there will be severe widespread social disruption to add to the mix. Only by gathering information now on the epidemiology of influenza can we prepare for the next inevitable influenza pandemic in humans.

"The pandemic clock is ticking, we just don't know what time it is"

(Marcuse, 2005)

2.5 Infectious Disease Surveillance: Definition

Disease surveillance as defined by the World Health Organisation is the "systematic ongoing collection, collation and analysis of data and the timely dissemination of information to those who need to know so that action can be taken" (WHO, 2007c). Other similar definitions have also been advanced including "Surveillance, when applied to a disease, means the continued watchfulness over the distribution and trends of incidence through the systematic collection, consolidation and evaluation of morbidity and mortality reports and other relevant data. Intrinsic in the concept is the regular dissemination of the basic data and interpretations to all who have contributed and to all others who need to know" (Langmuir, 1963).

2.6 European Legislation covering Infectious Disease

In 1969, International Health Regulations required that the WHO be informed within 24 hours of the occurrence of cases of certain infectious diseases, including cholera, plague and smallpox. In 1992, Council Directive 92/117/EEC required EEC member states to ensure that an official state body gather epidemiology information on cases of particular zoonoses (infections that can be passed from animals to humans) including TB, brucellosis and salmonellosis. Following this, in 1998, a Decision of the European Parliament and Council (2199/98/EC) was adopted which recommended that a network be established for the surveillance and control of infectious diseases between members of the European Community. Several projects were established, including Euro TB, European Influenza Surveillance Scheme (EISS) and the Basic Surveillance Network which collected national data from contributing EU members. These projects were important in the promotion of co-operation between member states regarding infectious disease information and provided the basis for an early warning network within the EU. A further EU Decision 2000/96/EC recommended a list of 40 diseases for surveillance by member states and case definitions for reporting of these infectious disease at European level were proposed and adopted following a 2002 European Commission Decision (2002/253/EC). In 2004, legislation was passed to establish a new agency: the European Centre for Disease Control (ECDC). This centre became operational in May 2005 and is located at the Karolinska Institute in Solna, Stockholm. Its objective is "to reinforce and develop Europe's system of continent-wide disease surveillance, reinforce Europe's rapid alert systems against disease outbreaks, support the EU and its Member States in strengthening preparedness against epidemics and provide authoritative scientific advice on infectious diseases and the risks they pose".

2.7 Irish Legislation covering Infectious Disease

The Department of Health and Children in Ireland was first established in 1947 and a list of notifiable infectious diseases was first specified in the Health Regulations of 1948 (NDSC Report, 2001). This list and further regulations were updated in the 1981 Infectious Disease Regulations. Revisions to the regulations were made in 1985, 1988, 1996, 2000 and most recently in 2003 and the list of current notifiable diseases can be seen in Table 2.2. The infectious disease regulations legally oblige "a medical practitioner, as soon as he or she becomes aware or suspects that a person on whom he or she is in professional attendance is suffering from or is the carrier of an infectious disease, and a clinical director of a diagnostic laboratory as soon as an infectious disease is identified in that laboratory, shall .. forthwith transmit a written or electronic notification to a medical officer of health". Furthermore the regulations state that "a medical practitioner and a clinical director of a diagnostic laboratory shall notify to the medical officer of health any unusual clusters or changing patterns of any illness, and individual cases thereof, that may be of public health concern. The medical officer of health shall in turn notify the National Disease Surveillance Centre". In 1998 the National Disease Surveillance Centre, renamed in 2004 as the Health Protection Surveillance Centre, was established and took on the role of national infectious disease co-ordinator. The HPSC has a number of functions nationally including: maintaining national infectious disease databases, producing annual reports and recommendations and providing Irish epidemiological data to various EU projects e.g. Euro-TB, Euro CJD and Euro HIV. In addition, the HPSC is involved in national policy and guidance development for infectious disease. Within the Health Service Executive structure, the HPSC reports to the Assistant National Director of Population Health with a remit for Health Protection.

Table 2.2: List of Current Notifiable Diseases in Ireland.
Taken from HPSC Website: http://www.ndsc.ie/hpsc/

Acute anterior poliomyelitis
Acute infectious gastroenteritis
Ano-genital warts
Anthrax
Bacillus cereus food-borne infection/intoxication
Bacterial meningitis (not otherwise specified)
Botulism
Brucellosis
Campylobacter infection
Chancroid
Chlamydia trachomatis infection (genital)
Cholera
Clostridium perfringens (type A) food-borne disease
Creutzfeldt Jakob disease
nv Creutzfeldt Jakob disease
Cryptosporidiosis
Diphtheria
Echinococcosis
Enterococcal bacteraemia
Enterohaemorrhagic Escherichia coli
Escherichia coli infection (invasive)
Giardiasis
Gonorrhoea
Granuloma inguinale
Haemophilus influenzae disease (invasive)
Hepatitis A (acute) (Hepatitis A virus)
Hepatitis B (acute and chronic) (Hepatitis B virus)
Hepatitis C (Hepatitis C virus)
Herpes simplex (genital) (Herpes simplex virus)
Influenza (Influenza A and B virus)
Legionellosis (Legionella sp.)
Leptospirosis (Leptospira sp.)
Listeriosis (Listeria monocytogenes)
Lymphogranuloma venereum
Malaria (Plasmodium falciparum, vivax, ovale, malariae)
Measles (Measles virus)
Meningococcal disease (Neisseria meningitidis)
Mumps (Mumps virus)
Non-specific urethritis
Noroviral infection (Norovirus)
Paratyphoid (Salmonella paratyphi)
Pertussis (Bordetella pertussis)
Plague (Yersinia pestis)
Q Fever (Coxiella burnetii)
Rabies (Rabies virus)
Rubella (Rubella virus)
Salmonellosis (Salmonella enterica)
Severe Acute Respiratory Syndrome
Shigellosis (Shigella sp.)
Smallpox (Variola virus)
Staphylococcal food poisoning
Staphylococcus aureus bacteraemia (S. aureus (blood)
Streptococcus group A infection (invasive)
Streptococcus pneumoniae infection (invasive)
Syphilis (Treponema pallidum)
Tetanus (Clostridium tetani)
Toxoplasmosis (Toxoplasma gondii)
Trichinosis (Trichinella sp.)
Trichomoniasis (Trichomonas vaginalis)
Tuberculosis
Tularemia (Francisella tularensis)
Typhoid (Salmonella typhi)
Typhus (Rickettsia prowazekii)
Viral encephalitis
Viral meningitis
Viral haemorrhagic fevers
Yellow Fever (Yellow Fever virus)
Yersiniosis

2.8 Infectious Disease Surveillance: Purpose

Surveillance of infectious disease, which has its original roots in epidemiology, has evolved into a distinct public health discipline (Declich & Carter, 1994). Gathering of surveillance data and subsequent generation of surveillance information on infectious disease contributes to an understanding of the baseline epidemiology of disease in a population, allows the detection of epidemics and outbreaks and provides timely information for public health action. Secondary to this, surveillance data is important in stimulating research in the area, in the generation of epidemiological hypotheses, in the evaluation of control measures and the identification of "at-risk" groups or sub populations. Furthermore, gathering of complete epidemiological surveillance data can have an important impact on public health planning and policy making.

Surveillance is dependent on the fostering of good relationships with clinicians and local microbiology laboratories to encourage the reporting of infectious diseases and outbreaks. It is also important to be able to collate and analyse data regionally and nationally and provide feedback to those who contribute data for surveillance. International collaboration and co-operation in the sharing of surveillance data on infectious diseases provides an early warning mechanism and can allow individual countries assess the threat to its own population. In other words, infectious disease surveillance is primarily about detection, information, prevention and action.

2.9 Purpose of Influenza Surveillance

Surveillance of influenza can provide valuable information on the epidemiological trends and outbreaks of the disease prior to, during and after seasonal and pandemic influenza events. Identification of populations most at risk, the baseline influenza prevalence in the community, factors involved in influenza transmission and monitoring outcomes of intervention measures will form the basis of public health policy and protection measures. Forecasting of priority groups for vaccination and anti-retroviral therapy will be essential before and during a pandemic. In other words gathering of reliable influenza surveillance data in the pre-pandemic, inter-pandemic and pandemic phases can have a major impact on the detection, containment and dissemination of the infection.

2.10 Conclusion

This chapter has provided an overview of the Influenza virus and the impact of pandemic influenza. Furthermore, the European and Irish legislation covering infectious disease surveillance has been assessed. The following chapter will describe the current Irish system for influenza surveillance and assess the state of the art in syndromic surveillance.

Chapter 3: State of the Art - Influenza Surveillance Systems

3.1 Introduction

This chapter will outline the state of the art in influenza surveillance in Ireland and other countries and will also review syndromic surveillance of influenza and influenza like illness as an early warning system

3.2 The European Influenza Surveillance System

One of the main goals of Influenza surveillance is to provide an early warning for the occurrence of both seasonal flu and pandemic flu. The European Influenza Surveillance Scheme (EISS), established by a European Parliament and Council decision (2199/98/EC), currently collects national clinical and virological data from 30 European countries on influenza. The pooling of virological data on influenza allows epidemiological trends for Europe to be analysed and provides a means of alerting neighbouring countries of outbreaks within their vicinity (Fig 3.1). Furthermore, this virological data is also supplied to the WHO influenza surveillance network who are involved in predicting the circulating strains for the year ahead and recommending the composition of the associated influenza vaccine.

A comparison of the four main components of influenza surveillance systems across Europe was provided in a survey (Aguilera et al., 2001) carried out by EISS members in 2000 (Table 3.1). This survey demonstrated that 7/20 (35%) countries surveyed at the time collected hospital admission data as an indicator of influenza activity, 12/20 (60%) collected mortality data on influenza, 7/20 (35%) collected sickness absenteeism data as an indicator of influenza activity and almost all countries surveyed (18/20, 90%) collected clinical samples for virological identification from sentinel practices. This report and others has been used to provide a protocol for evaluation of the quality and timeliness of virological influenza surveillance data and it is hoped that eventually, through EISS, there will be a basis for the standardisation of influenza data collection across Europe.

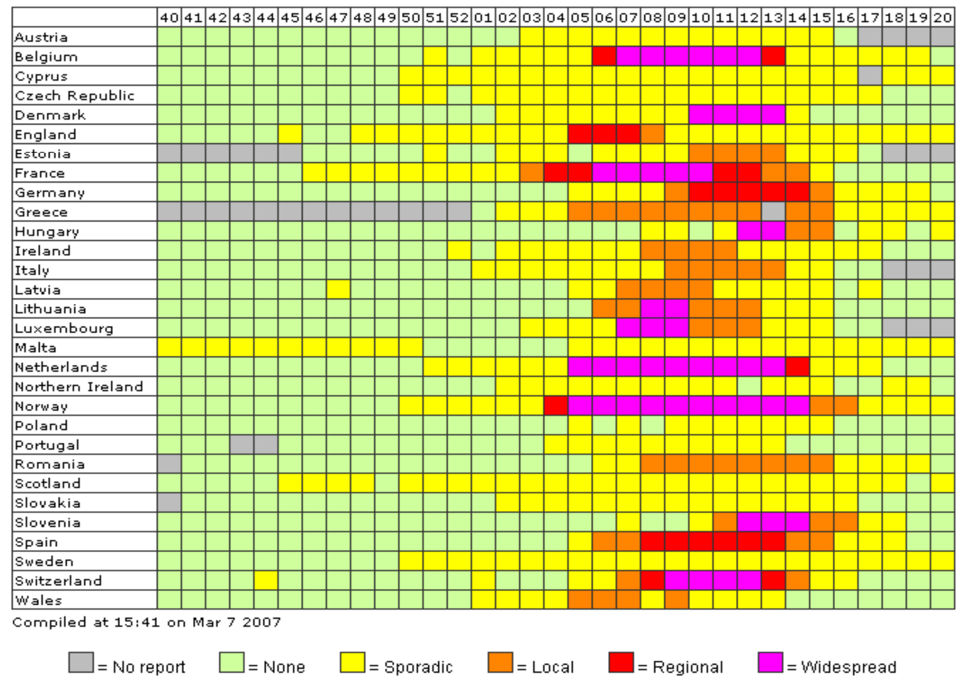
Table 3.1: Comparison of Flu Surveillance Systems in Europe.
Information taken from Aguilera et al., 2001.

COUNTRY	Collection of Hospital Data	Collection of Mortality Data	Collection of Absenteeism Data	Collection of Clinical Specimens	No. of practices collecting specimens	Average no. of swabs per week
BELGIUM	yes	yes	yes	yes	60	38
CZECH	no	yes	no	yes		45
DENMARK	no	yes	no	yes	150	10
ENGLAND	no	yes	no	yes	60	9
FRANCE	yes	no	yes	yes		122
GERMANY	no	yes	yes	yes	150	200
IRELAND	yes	no*	yes	yes	32	10
ITALY	no	no	no	yes	100	50
LATVIA	no	yes	yes	yes	40	87
NETHERLANDS	yes	yes	no	yes	33	20
NORWAY	no	no	no	yes	55	10
POLAND	yes	yes	no	yes	10	50-80
PORTUGAL	no	no	no	yes	40	30
SCOTLAND	yes	yes	no	yes	148	200
SLOVAKIA	no	yes	yes	yes		20
SLOVENIA	no	no	no	yes	44	3
SPAIN	no	no	no	yes	260	35
SWEDEN	no	yes	yes	no		
SWITZERLAND	yes	yes	no	yes	65	30
WALES	no	no	no	no		

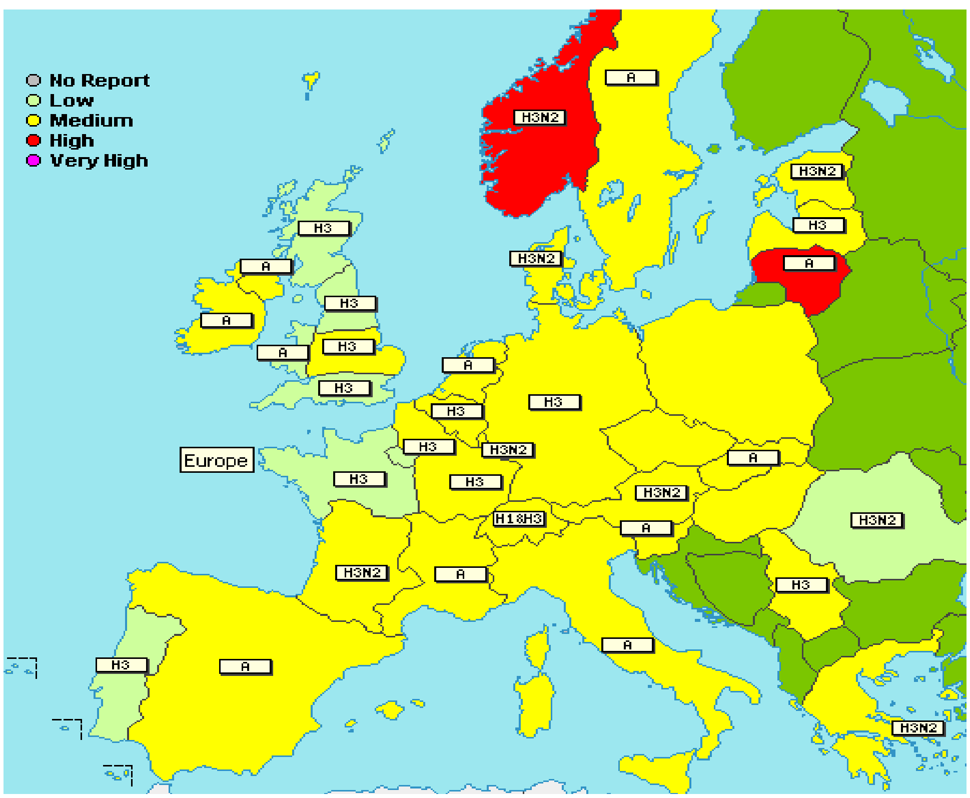
* Since new legislation was passed in 2004, the HPSC also gathers mortality data on influenza from the General Registers Office.

Fig 3.1: Typical information from EISS demonstrating the geographical mapping of influenza activity data and mapping of influenza virological data (EISS, 2007)

2005/2006 Flu Season



Week 8, 2007



3.3 The Irish Influenza Surveillance System

The Irish influenza surveillance system, managed nationally by the HPSC and which provides data to EISS, was first established in October 2000. The aim of the system is to monitor morbidity and mortality associated with influenza and influenza-like illness (ILI) in Ireland.

The system consists of a number of standard components as described below:

3.3.1 Influenza Notification System

Since 1st of January 2004, all medical officers are obliged to report any influenza cases through the normal statutory notification channels. Furthermore, under the legislation governing infectious diseases (Infectious Disease Regulations, 1981), the details of deaths attributed to a communicable disease, including influenza and pneumonia, are required to be reported weekly by the General Registers Office to the HPSC.

3.3.2 Sentinel and non sentinel virological surveillance

Sentinel GP's have been recruited through the Irish College of General Practitioners (ICGP) since 2000 to provide clinical data (GP id, patient date of birth, gender, diagnosis date, diagnosis) on all diagnosed influenza-like illness (ILI) seen in their practice on a weekly basis. In this context, ILI is defined as "the sudden onset of symptoms with a temperature of 38°C or more, with two or more of the following: headache, sore throat, dry cough and myalgia". For the 2006/2007 flu season 47 GPs are involved in this surveillance system. In addition, sentinel GPs are required to send a nasal and throat swab from at least one patient per week presenting with ILI to the National Virus Reference Laboratory (NVRL) for Influenza virological type confirmation. Virological data is sent by the NVRL to WHO. A recent evaluation of this sentinel GP component demonstrated that it complied with EISS requirements (McNamara, 2006). Additionally, positive non sentinel swabs obtained by the NVRL (usually from hospital paediatric wards and non sentinel GPs) are routinely reported to the HPSC during the Influenza season.

3.3.3 Sentinel surveillance: Hospital activity data

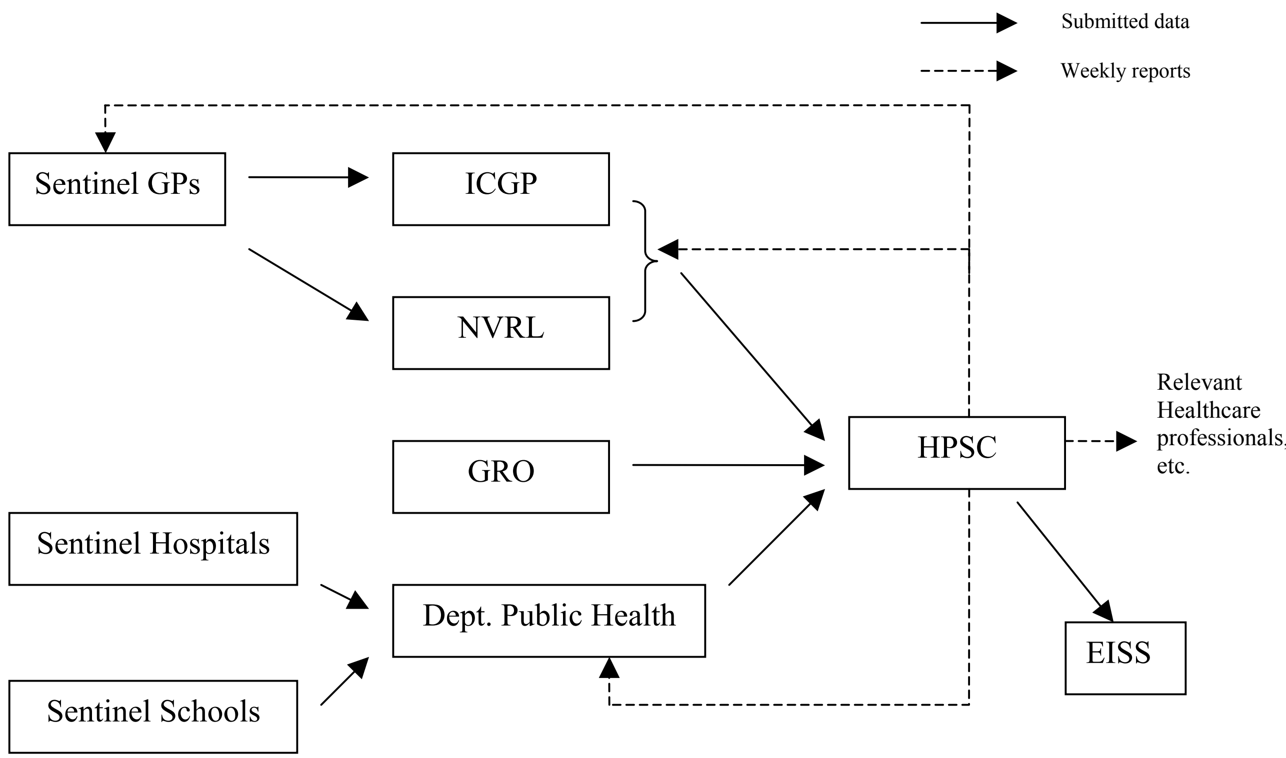
Each of the eight HSE-area Departments of Public Health are required to recruit a local hospital to participate in surveillance of influenza during the influenza season. Data on total admissions, respiratory admissions, A&E admissions are collected weekly in an electronic format from these sentinel hospitals. Respiratory admissions are defined as admission due to upper respiratory tract infection, lower respiratory tract infection, pneumonia, asthma, chronic bronchitis and exacerbations of chronic obstructive airways disease. For the 2004/2005 and 2005/2006 Influenza seasons, 7 hospitals provided data for influenza surveillance which was aggregated and reported as percentage respiratory admissions out of total admissions for each week of the season.

3.3.4 Sentinel surveillance: Absenteeism in schools

Each HSE-area Department of Public Health is also required to recruit a sentinel primary and secondary school, located near to the sentinel GP practice, to collect information on the average daily absenteeism levels.

Together this information provides the basis for an influenza activity index ranging from 1 to 5 where 1 represents “no report”, 2 represents “no influenza activity”, 3 represents “sporadic activity”, 4 represents “localised activity” and 5 represents ““regional or widespread influenza activity”. This index is comparable to EISS and WHO influenza activity indices and as such, data can be compared at a national level with that of other countries in order to assess the influenza threat to the population during the influenza season (the period when influenza is most likely to occur – from Week 40 to Week 20 of the following year). The HPSC enters the relevant data weekly on the EISS website and produces a weekly influenza report, with analysis of the national data, during the influenza season. A diagram outlining the flow of information in the Irish influenza surveillance system is shown in Fig. 3.2.

Fig. 3.2: Flow of information in the Irish influenza surveillance system.
Taken from McNamara, 2006



3.4 Evaluation of the Influenza Surveillance System

An evaluation of this influenza sentinel surveillance system in 2006 by the HPSC (McNamara, 2006) recognised the value and usefulness of the GP sentinel surveillance data, particularly the virological typing of circulating virus strains. Indeed, this component of the system is considered the major strength of the system. However, as there are only 46 sentinel GP practices out of approx. 2000 practices (estimate from ICGP) currently operating in Ireland it is likely that only a small fraction of all influenza cases ever have nose and throat swabs taken for virological identification and ever come under the radar of sentinel influenza surveillance. Furthermore, swabs taken for virological examination by the NVRL take 1-2 weeks to be processed before a result is available. Therefore, although the specificity of sentinel surveillance is quite high, its sensitivity to detect outbreaks or provide an "early alert" is low. According to the evaluation by McNamara, the HPSC recommend that an increased "population coverage" by increasing sentinel GPs numbers is required to improve the overall sensitivity of the system.

Reservations on the quality and timeliness of the sentinel school and hospital data were also expressed in the report. Some of the factors which affected the quality of this data included a lack of standardised definitions and a lack of standardised information technology infrastructure, particularly in the hospital setting which made it difficult to extract comparable "respiratory admissions". In fact, in one hospital it was impossible to extract admissions of any kind from the computerised system and only details of discharges could be provided. Furthermore, the timeliness of the school and hospital data was an issue. In some instances hospital admission data was only supplied on a 4-6 weekly basis and 5 out of 8 health boards described difficulties in obtaining regular school absenteeism data. Indeed, most health boards, felt that school absenteeism was a poor indicator of influenza and influenza like illness.

Under-reporting of statutorily notifiable diseases is another well recognised phenonomen which can mean that routine surveillance data is incomplete. Although all clinicians are legally obliged to notify infectious diseases (including influenza) there is a large repository of evidence both in other countries (Harvey, 1991; Davison et al., 2003; Barrett & Lau, 1996; Lerman et al., 1999; Karim & Karim, 1991, Strauss et al., 2003) and in Ireland (Collins, 1997; Brabazon et al., 2004; Brabazon et al., 2007) to suggest that not all cases are notified. Reasons for this underreporting include lack of time, lack of familiarity with the notifiable disease list, apathy regarding the notification process and concerns regarding patient confidentiality. It is reasonable to assume that there is also some level of under-reporting of influenza by Irish clinicians which in turn leads to an under-estimation of disease incidence in the population.

So, although there is information from routine and sentinel surveillance to provide an indication of influenza activity in the Irish population, whether the current system is sufficient to provide an "early warning" or "early alert" component is questionable. Clearly, enhancements to the system would strengthen its reliability.

3.5 Other Data Sources for Influenza Surveillance

With the threat of an imminent global influenza pandemic, there is a responsibility to ensure a reliable comprehensive system is in place for influenza surveillance and that all available data sources have been assessed for usefulness in detecting influenza in the community setting. In particular, data which are gathered electronically and which are maintained as part of routine medical care should be exploited to provide a timely and readily available source of information on influenza and ILI transmission.

Comparison of the Irish surveillance system with the current Northern Ireland and UK Influenza surveillance system is shown in Table 3.2. All surveillance systems collect information from statutory notifications and outbreaks, sentinel GP's, virological analysis of sentinel GP samples, death registrations and hospital admissions. But this is where the commonality ends. Each jurisdiction, in an attempt to provide an early warning of influenza activity, collects information from various different community based sources including over the counter drug sales, routine laboratory reporting, school absenteeism and boarding school illness indicators and healthcare call centre data. There are issues with most of these influenza indicators; for example, in the case of community pharmacy data, although there is a record of sales of specific medications there is no way of knowing when these medications are being used, whom they are being used by and if they are being used in the appropriate dosage and for the appropriate ailments. Furthermore, routine laboratory reporting suffers from under-reporting, school absenteeism rates may be dependent on factors other than illness and illness in a boarding school environment may not be representative of illness in the general community. Healthcare call centre data, however, which logs the direct interaction of a patient with a healthcare service and maintains a record of the patient's symptoms at a current point in time, does not suffer from under-reporting and is a direct measure of illness. The symptoms are considered severe enough by the patient to warrant interaction with a health service. This type of call centre system records information in an electronic format from all phone calls to a dedicated call number associated with a particular healthcare service, e.g. emergency room telephone triage, out of hours GP services, ambulance dispatch service. Analysis of trends in this type of pre-diagnosis data is a typical form of "syndromic" surveillance. These healthcare call centre services are believed to gather important unexploited real-time data on community wide illness.

Table 3.2: Comparison of current data sources for influenza surveillance in Ireland, Northern Ireland and the UK:

Data Source	Ireland	UK	CDSC, NI
Statutory Notifications & Outbreaks	Yes	Yes	Yes
Sentinel GP practices – Clinical Data	Yes	Yes	Yes
Viroloical identification of sentinel GP samples	Yes	Yes	Yes
Mortality Data	Yes	Yes	Yes
Hospital Admission data	Yes	Yes	Yes
Routine laboratory reports	No	Yes	Yes
Illness in children in boarding schools	No	Yes	No
Sentinel Community Pharmacies	No	No	Yes
School Absentism Data	Yes	No	No
Call Centre Data	No	Yes*	Yes#

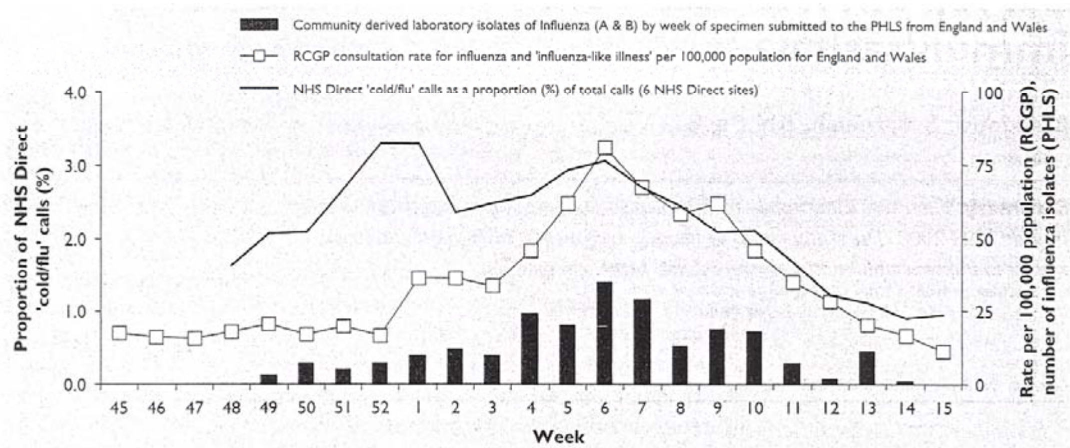
Taken from the following websites on reporting of influenza surveillance data:

- Ireland (Health Protection Surveillance Centre): http://www.ndsc.ie/hpsc/
- UK (Health Protection Agency): http://www.hpa.org.uk/infections/topics_az/influenza/seasonal/uk_data_sources.htm
- Northern Ireland (Communicable Disease Surveillance Centre): http://www.cdscni.org.uk/
- * NHS-Direct 24hr Nurse Advise Line
- # GP Out of Hours Service

3.6 Use of Call Centre Data in the UK for Influenza Surveillance

In the UK, a nurse-led 24hrs a day, 365 days a year national telephone helpline (NHS-Direct) was established in 1998 covering England and Wales. This system has been deemed as "the worlds largest provider of telephone-based healthcare" and has been investigated as a source of information on influenza activity (Cooper et al., 2002; Harcourt et al., 2001). This decision support system allows the nurse to identify or diagnose health-related problems and provide standardised advice to callers. A "cold/flu" call was defined as symptoms which lead the nurse to select a "cold/flu" clinical algorithm in the system. These studies demonstrated that the highest call rate for influenza like illness was in children less than 5 years old. Furthermore, peaks in the proportion of "cold/flu" calls coincided with an increase in influenza as determined by routine surveillance systems (Fig 3.3). These studies recommended the continued use of

Fig 3.3. Comparison of NHS-Direct call rate, GP consultation rates and number of Influenza isolates from sentinel GP practices for the influenza period 2000 to 2001 in England and Wales.
Note that NHS-Direct call rates follow a similar pattern as other indicators of influenza activity with the first peak in call rates even slightly ahead. Data taken from Cooper et al., 2002



NHS-Direct calls as a source of syndromic surveillance data for influenza. To this end a weekly report by the HPA using NHS-data (the NHS Direct Syndromic Surveillance Bulletin) was established including information on weekly call rates and respiratory calls. These respiratory calls are broken down as "cold/flu" calls, "cough" calls and "fever" calls and are compared to the previous four year averages for the same calls. This report has become an integral part of routine syndromic surveillance in the UK.

In Northern Ireland, during the influenza season, five out of hours GP centres also provide information on their total number of calls per week, broken down by age and sex (CDSCNI, 2006). Up to the 2005/2006 flu season, the data was not clinically coded, so total call numbers were analysed over time to identify peaks in call numbers which may be indicative of increased illness in the community. From this year, a new software system has been established in all out-of-hours centres in Northern Ireland which will allow the clinical coding of all incoming calls. It is therefore expected that the proportion of calls per week due to “flu-like illness” and “acute respiratory infection” can be extracted in an ongoing manner. This advancement will greatly enhance the current surveillance system in Northern Ireland.

3.7 Use of Call Centre Data in Other Countries for Influenza Surveillance

A report on use of call centre data for influenza syndromic surveillance in the United States describes the usefulness of an emergency room telephone triage facility that serves 10 hospitals and an after hours telephone triage for physicians in one state (Espino et al., 2003). The call centre triage data, which was collected over a one year period from September 2001 to August 2002, was syndrome mapped and all influenza related calls were compared with influenza activity data for the corresponding geographical region as reported by the CDC. Both call centre datasets were found to have similar temporal trends to the routine surveillance data for influenza. This study found that the influenza call data gathered from the emergency room was 1 to 5 weeks ahead of the routine surveillance data captured by the Centre for Disease Control. Furthermore, a gradual increase in the after hours telephone triage data was shown to occur several weeks before the peak in the routinely collected influenza activity data. This suggests that both of these datasets would provide a useful early warning of influenza activity in the community.

Two further studies in the US using nurse telephone triage data also demonstrated that the volume of calls for respiratory and ILI cases matched well with peaks in CDC’s routine influenza surveillance (Yih et al, 2006; Pinner et al., 2006). In New York, Mostashari et al., (2003) have demonstrated the usefulness of ambulance dispatch calls for monitoring outbreaks of influenza. Daily regression analysis allowed the identification of clusters of influenza cases which were shown to coincide with a period of peak influenza activity in each of six seasons examined. The authors conclude that this surveillance system “is sensitive to community wide respiratory outbreaks”.

In Victoria, Australia, a medical locum service that operates after normal working hours through a call centre has been investigated as a data source for influenza surveillance (Turner & Kelly, 2005). The details of the patient's call is forwarded to the doctor's pager and the patient is triaged as deemed necessary. The locum doctor provides a follow up on the case including presenting complaint, clinical findings and diagnosis and this information is entered onto a specific database. The patient's information and details of interaction with the service can be downloaded the following morning by the patient's own GP. Records containing the keywords "flu" or "influenza" were compared retrospectively and prospectively with routine ILI data from the sentinel surveillance system for influenza. Over a 6 year period from 1998 to 2003, it was demonstrated that there were similar patterns between the two datasets and that the medical locum service saw peaks in activity which corresponded with peak sentinel influenza activity. Indeed the medical locum service in each year saw a greater number of patients with ILI compared with the sentinel GPs. The authors demonstrate that this service has "a role in the recognition of emerging disease patterns". Clearly, these types of telephone triage or call centre health helplines can be an important addition in the early surveillance of influenza.

3.8 Irish Healthcare Call Centre Data

In Ireland, a common data source which has been, heretofore, overlooked as a source for influenza like illness in the community is the "Doctor on Call" call centre data. Since 1998, 11 out-of-hours co-operatives have been developed covering approximately 40% of the Irish population. The average population contact rate is 221 contacts/1000 persons/year (Bury et al., 2006). The aim of these co-operatives is to provide an urgent out of hours GP service to their patients and to provide an environment for continuity of care. This service, often in partnership with the local health board, provides access to a doctor through a call centre telephone number. The call is answered by trained personnel who record the call details in a computerised system. This information is passed to a doctor who returns the call promptly and will either provide advice over the phone, arrange for the patient to be seen at the local "Doc on Call" centre or provide a home visit where necessary. Details of the consultation with the doctor are also recorded and the information is passed to the patient's own GP before surgery opens the next morning ensuring that follow up can be carried out if required. This service, therefore, maintains a large data repository which contains records of self reported illness (syndromes) in the community and a clinician's subsequent diagnoses.

A survey on satisfaction levels with one of these GP co-operatives in Ireland (Health Matters, 2002) has provided some interesting data. Since its establishment in 2000, the co-op received over 100,000 calls over a two year period. Interestingly, 82% of people felt that their complaint couldn't wait until the next day and 66% were afraid that their condition would get worse. Of those who were surveyed, 82% were seen at a treatment centre by a doctor, 7% received a home visit by a doctor and 11% were dealt with over the phone by a doctor. In other words, these figures highlight the severity of patient's reported conditions to the call centre as, overall, 89% of callers surveyed were deemed to require immediate face to face consultation with a doctor.

This urgent out of hours service, therefore, collects valuable information on illness in the community setting and there is the potential to utilise this Irish data source as an early warning system for both seasonal and/or pandemic influenza. The advantage of this data is that it is readily available in an electronic format and can be easily extracted for further manipulation. Furthermore, the data is collected in real-time, i.e. as the call is being made, so the information in the system is current. Finally, this type of information has been shown to be a useful addition to influenza surveillance in other countries. Supplementing the Irish surveillance system, therefore, with healthcare call centre indicators for circulating disease in the community can only enhance the detection of the cause of an imminent public health crisis: influenza.

3.9 Conclusion

This chapter has described the state of the art in influenza surveillance. The use of healthcare call centre data in other countries for routine influenza surveillance strongly supports the premise that the out of hours GP service in Ireland has the potential to be used as a new data source for information on regional and national influenza activity.

Chapter 4: Methodology

4.1 Introduction

This chapter outlines the general methodology used in this thesis to analyse data from the north east urgent out of hours GP service (NE-Doc). Furthermore, the design and validation process of Microsoft Access queries for the extraction of records containing keywords (e.g. flu, influenza etc.) will be detailed.

4.2 Literature Review

Various sources of information were consulted in order to obtain the relevant background information for this project including searching of PubMed, IEEE and Google. Some of the basic search terms included: “Influenza”, “Flu”, “Pandemic”, “Surveillance”, “Telephone Triage”, “Call Centre Data”, “Doctor on Call”. Specific websites relating to infectious disease and public health were also reviewed including the HPSC (Ireland) website, the HPA (UK) website, the CDSCNI (Northern Ireland) website, the WHO (International) website and the CDC (US) website.

4.3 Background information on NE-Doc

A semi-structured interview was conducted with the IT Support Manager for North East Doctor on Call (NE-Doc) service (which covers the north east and midland regions) in order to gather background information on the system and the processes in place for capturing call data. The pre-determined questions for this interview are listed in Table 4.1.

Table 4.1: Pre-determined domain analysis interview questions for NE-Doc service

	Questions
1	When was the NE-Doc system established?
2	What is the name of the software that is used for data collection?
3	Is it an off the shelf, in house or customised system?
4	If off the shelf, what is the name of the supplier?
5	What is the architecture of the system?
6	Is the system being used by any other co-ops nationally or internationally?
7	Is there any clinical coding of the data in the system?
8	Are the data entry staff from clerical or nursing background?
9	How does the doctor’s diagnosis get added to the sytem?
10	Do they run any routine reports on the system?

4.4 Quantitative Data Analysis

Data on all calls for the 2003/2004, 2004/2005 and 2005/2006 influenza season (Table 4.2 and Appendix 1) were requested from the NE-Doc service. The data were extracted into .CSV files and were opened in Microsoft (MS) Excel. The fields in the dataset consisted of Call Number, Date of Call, Patient Age, Patient Sex, Patient's Reported Condition, Doctor's Diagnosis/Outcome. The patient's reported condition and doctor's diagnosis/outcome fields are free text fields.

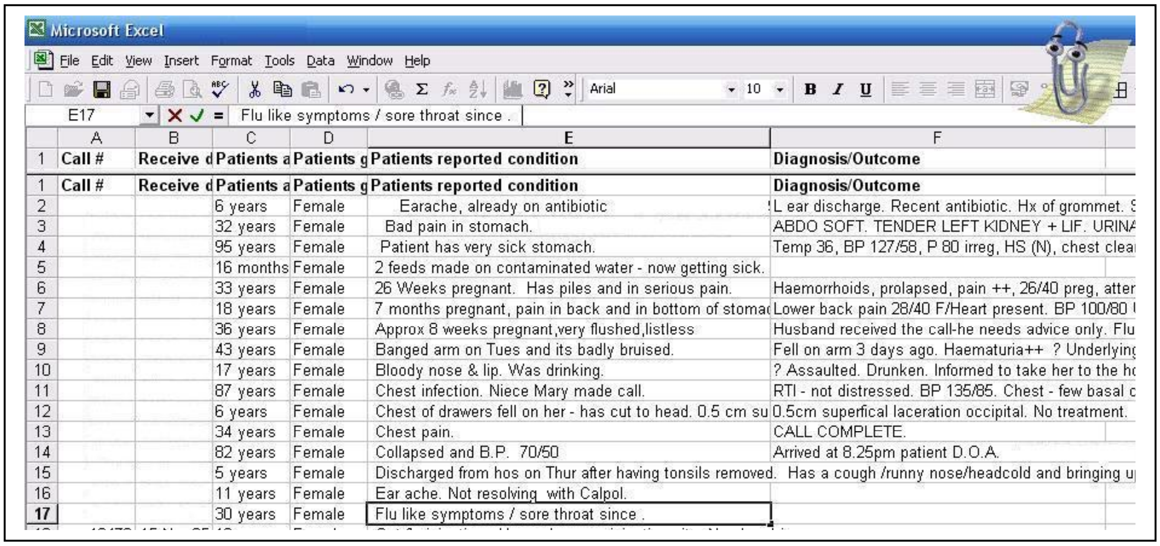
An example of the original data received from NE-Doc service is shown in Table 4.3. New fields that were manually added to the dataset included Flu Season Week Number, Day of Call and Age Group. All patient ages over 100 yrs were assumed to be typographical/entry errors and were categorised as age group "Unknown". The "Unknown" category was also used for records where the age of the patient was not recorded. Records which contained "test call" in the patients reported condition were deleted from the file – these referred to IT maintenance records. All subsequent statistical analysis was carried out in the statistics package, JMP (SAS Institute Inc). The data was also imported into Microsoft (MS) Access so that free text fields could be queried.

Table 4.2: Dates of Flu Seasons analysed in this study

Flu Season	Start Date (beginning Week 40)	End Date (ending Week 20)
2005/2006	3rd October 2005	21st May 2006
2004/2005	27th September 2004	22nd May 2005
2003/2004	29th September 2003	16th May 2004

In both hemispheres, influenza outbreaks are most common in the Winter and Spring months and surveillance of influenza is therefore concentrated to this influenza "season". In the Northern hemisphere, this season begins in Week 40 (end September, beginning October) and ends in Week 20 (generally mid-May). More detailed table outlining Week numbers during the Flu season can be found in Appendix 1.

Fig 4.1: Example of data received from NE-Doc



	A	B	C	D	E	F
1	Call #	Receive d	Patients a	Patients g	Patients reported condition	Diagnosis/Outcome
2			6 years	Female	Earache, already on antibiotic	L ear discharge. Recent antibiotic. Hx of grommet. S
3			32 years	Female	Bad pain in stomach.	ABDO SOFT. TENDER LEFT KIDNEY + LIF. URINA
4			95 years	Female	Patient has very sick stomach.	Temp 36, BP 127/58, P 80 irreg, HS (N), chest clea
5			16 months	Female	2 feeds made on contaminated water - now getting sick.	
6			33 years	Female	26 Weeks pregnant. Has piles and in serious pain.	Haemorrhoids, prolapsed, pain ++, 26/40 preg, atter
7			18 years	Female	7 months pregnant, pain in back and in bottom of stomac	Lower back pain 28/40 F/Heart present. BP 100/80
8			36 years	Female	Approx 8 weeks pregnant,very flushed,listless	Husband received the call-he needs advice only. Flu
9			43 years	Female	Banged arm on Tues and its badly bruised.	Fell on arm 3 days ago. Haematuria++ ? Underlying
10			17 years	Female	Bloody nose & lip. Was drinking.	? Assaulted. Drunken. Informed to take her to the ho
11			87 years	Female	Chest infection. Niece Mary made call.	RTI - not distressed. BP 135/85. Chest - few basal c
12			6 years	Female	Chest of drawers fell on her - has cut to head. 0.5 cm su	0.5cm superfical laceration occipital. No treatment.
13			34 years	Female	Chest pain.	CALL COMPLETE.
14			82 years	Female	Collapsed and B.P. 70/50	Arrived at 8.25pm patient D.O.A.
15			5 years	Female	Discharged from hos on Thur after having tonsils removed. Has a cough /runny nose/headcold and bringing u	
16			11 years	Female	Ear ache. Not resolving with Calpol.	
17			30 years	Female	Flu like symptoms / sore throat since .	

Note: Call Number (Call #) and Received date have not been shown for data protection reasons.

4.5 Calculation of Population Rates

The NE-Doc service does not maintain population numbers for their component practices, therefore in line with other studies (Bury et al., 2006), a nominal mean practice population of 1,500 persons per GP was used for population estimates in this study. This figure is deduced from a national profile of general practice in Ireland (ICGP, 1997) and the Department of Health's primary care strategy (Primary Care Strategy, 2001). The number of GPs involved in the cooperative at the end of each influenza season and the corresponding population estimate is shown in Table 4.3. Population contact rates were calculated as all contacts with patients per 1000 persons served by the co-operative.

Table 4.3: Number of GP's involved in Co-op during study period.

	No. of GPs	Population coverage estimate
2003/2004	199	298,500
2004/2005	224	336,000
2005/2006	224	336,000

4.6 Date Naming Convention

Through this thesis, the data for particular weeks will be named according to the Flu Season week and Flu Season year. For example, the 3rd October, 2005 to 10th October 2005 is Week 40 of the 2005/2006 Flu Season (Appendix 1).

4.7 Manual identification of a subset of data with flu-related keywords

The objective of this study is to automatically extract flu-related calls from a database of all calls to an urgent out of hours GP service covering two health board areas. This requires the construction of specific queries which can extract flu-related calls with high sensitivity and specificity (Section 4.9). In order to ensure that the queries are extracting the appropriate records from the all calls database or in other words to test, refine and validate these queries, a subset of data needs to be manually categorised for direct comparison with query results. Therefore, the patient's reported condition field from a subset of data from the 2005/2006 Flu season (Weeks 9 & 10 corresponding to a period of influenza activity as determined by ILI consultations reported to the HPSC) was manually inspected to identify records with key words relating to influenza and influenza like illness. This two-week period contained 5,732 records in total. Keywords relating to influenza that were chosen for inclusion in the analysis consisted of: Flu/Influenza, Cough, Temperature, Fever, Headache, Muscle Pain, Aches and Pains, Sore Throat and Sudden Onset. A combination of these terms represent the HPSC (Ireland) and CDC (USA) definitions of influenza like illness (see Chapter 5, Section 5.5 for further details). Aches and Pains was chosen instead of Myalgia in the HPSC ILI definition as it represents the common lay term for the medical phrase. A binary system was used to manually mark the records containing any of these keywords. The results of this analysis can be seen in Table 4.4.

Table 4.4: Analysis of Subset of NE-Doc Data

Binary Code	Count	Interpretation of Code (Keywords Present in PRC Field)	Meets Criteria for:
000000000	3722	No Key Words in Record	
000000010	157	Sore Throat	
000000100	16	Aches & Pains	
000000110	3	Aches & Pains, Sore Throat	
000010000	138	Headache	
000010010	18	Headache, Sore Throat	
000010110	1	Headache, Sore Throat, Aches & Pains	
000100000	14	Fever	
000100010	1	Fever, Sore Throat	CDC Definition
000110000	2	Fever, Headache	
001000000	753	Temperature	
001000010	99	Temperature, Sore Throat	CDC Definition
001000100	4	Temperature, Aches & Pains	
001000110	1	Temperature, Aches & Pains, Sore Throat	CDC & HPSC Definition
001010000	67	Temperature, Headache	
001010010	9	Temperature, Headache, Sore Throat	CDC & HPSC Definition
001010100	2	Temperature, Headache, Aches & Pains	HPSC Definition
001100000	2	Temperature, Fever	
001100010	1	Temperature, Fever, Sore Throat	CDC Definition
010000000	267	Cough	
010000010	35	Cough, Sore Throat	
010000110	1	Cough, Sore Throat, Aches & Pains	
010010000	9	Cough, Headache	
010010010	4	Cough, Headache, Sore Throat	
010100000	8	Cough, Fever	CDC Definition
010110000	1	Cough, Fever, Headache	CDC Definition
011000000	137	Cough, Temperature	CDC Definition
011000010	21	Cough, Temperature, Sore Throat	CDC & HPSC Definition
011010000	8	Cough, Temperature, Headache	CDC & HPSC Definition
011010010	3	Cough, Temperature, Headache, Sore Throat	CDC & HPSC Definition
011100000	1	Cough, Temperature, Fever	CDC Definition
100000000	157	Flu	
100000010	8	Flu, Sore Throat	
100000100	3	Flu, Aches & Pains	
100010000	5	Flu, Headache	
100010010	2	Flu, Headache, Sore Throat	
100100000	1	Flu, Fever	
101000000	24	Flu, Temperature	
101000010	4	Flu, Temperature, Sore Throat	CDC Definition
101100000	1	Flu, Temperature, Fever	
110000000	17	Flu, Cough	
110000010	1	Flu, Cough, Sore Throat	
111000000	4	Flu, Cough, Temperature	CDC Definition
Total	5732		

4.8. Design of Microsoft Access Queries for Extraction of Flu-related records

All NE-Doc records (5,732 in total) for the two week period of the 2005/2006 Flu season as described in Section 4.7 were imported into one table in a Microsoft Access database. This table was used to design Access queries which would extract records with particular keywords relating to influenza. The number of records extracted could be compared with the expected number of records counted manually (Table 4.5) in order to assess how well the query was performing. Modifications were made to the queries to try to increase the number of records being extracted in line with the manually identified records.

For example, initially a query was set up which extracted a list of all records from the subset with the term "Flu" from the patients reported condition field (Fig. 4.2) using the expression:

Like "*"+"Flu"+"*"

Fig. 4.2: Initial query design

Field:	Receive date	Week Year	Age Group	Patients reported condition
Table:	Week 9 and 10 Flu validation	Week 9 and 10 Flu validation	Week 9 and 10 Flu validation	Week 9 and 10 Flu validation
Sort:				
Show:	☑	☑	☑	☑
Criteria:				Like "*"+"Flu"+"*"
or:				

This query returns 262 records well above the 227 records identified with the keyword "Flu" during the manual search. On visual inspection of the "list query" this method was found to over estimate the number of true flu related patient reported conditions as it also returned words containing the "flu" term e.g. fluids, reflux, flushed, fluctuating, fluttering etc.

To modify the query to make it more specific for "Flu" or "Influenza" the expression

Like "*flu[!a-z]*" Or Like "*flu" Or Like "*influenza[!a-z]*" Or Like "*influenza"

was used. This expression only allows words that begin with Flu or Influenza to be included in the search. Using this query 228 records were returned making the query much more specific.

Fig. 4.3: Design of Query 1

Field:	Week No	Age Group	Patients reported condition
Table:	Week 9 and 10 Flu validation	Week 9 and 10 Flu validation	Week 9 and 10 Flu validation
Sort:			
Show:	☑	☑	☑
Criteria:			Like "*flu[!a-z]*" Or Like "*flu" Or Like "*influenza[!a-z]*" Or Like "*influenza"
or:			

In total, four queries were designed to optimise the extraction of flu related records using the method described, ie. by comparing with the manually identified records. Query 1 contained only the key search terms "flu" or "influenza" (using the expression outlined above, Fig. 4.3). Query 2 contained only the key search terms in the patients reported conditions field of "fever" or "high temperature" and "cough" or "sore throat" representing the CDC definition of ILI (Fig. 4.4). The third query was designed to extract, in the patients reported conditions field, the terms "fever" or "high temperature" and two or more of the following "headache", "sore throat", "cough", "aches and pains" representing the HPSC definition of ILI (Fig. 4.5). The final query combined all three definitions for influenza like illness (Fig 4.6). Further details on the choice of these ILI definitions is discussed in Chapter 5, Section 5.5. Query 1 was also applied to the doctors diagnosis field to extract diagnoses relating to Flu/Influenza.

Fig. 4.4: Design of Query 2

Field:	Receive date	Week Year	Patients reported condition
Table:	Week 9 and 10 Flu validation	Week 9 and 10 Flu validation	Week 9 and 10 Flu validation
Sort:			
Show:	☑	☑	☑
Criteria:			Like "*cough*" And (Like "*temp*" Or Like "*fever*")
or:			Like "*sore*" And Like "*throat*" And (Like "*temp*" Or Like "*fever*")

Fig. 4.5: Design of Query 3

Field:	Receive date	Week Year	Patients reported condition
Table:	Week 9 and 10 Flu	Week 9 and 10 Flu	Week 9 and 10 Flu validation
Sort:			
Show:	☑	☑	☑
Criteria:			Like "*Ache*" And Like "*Pains*" And Like "*head*" And Like "*ache*" And (Like "*temp*" Or Like "*fever*")
or:			Like "*Ache*" And Like "*Pains*" And Like "*sore*" And Like "*throat*" And (Like "*temp*" Or Like "*fever*")
			Like "*Ache*" And Like "*Pains*" And Like "*cough*" And (Like "*temp*" Or Like "*fever*")
			Like "*head*" And Like "*ache*" And Like "*sore*" And Like "*throat*" And (Like "*temp*" Or Like "*fever*")
			Like "*head*" And Like "*ache*" And Like "*cough*" And (Like "*temp*" Or Like "*fever*")
			Like "*sore*" And Like "*throat*" And Like "*cough*" And (Like "*temp*" Or Like "*fever*")

Fig. 4.6: Design of Query 4

Field:	Receive date	Week Year	Patients reported condition
Table:	Week 9 and 10 Flu validation	Week 9 and 10 Flu validation	Week 9 and 10 Flu validation
Sort:			
Show:	☑	☑	☑
Criteria:			Like "*Ache*" And Like "*Pains*" And Like "*head*" And Like "*ache*" And (Like "*temp*" Or Like "*fever*")
or:			Like "*Ache*" And Like "*Pains*" And Like "*sore*" And Like "*throat*" And (Like "*temp*" Or Like "*fever*")
			Like "*Ache*" And Like "*Pains*" And Like "*cough*" And (Like "*temp*" Or Like "*fever*")
			Like "*head*" And Like "*ache*" And Like "*sore*" And Like "*throat*" And (Like "*temp*" Or Like "*fever*")
			Like "*head*" And Like "*ache*" And Like "*cough*" And (Like "*temp*" Or Like "*fever*")
			Like "*sore*" And Like "*throat*" And Like "*cough*" And (Like "*temp*" Or Like "*fever*")
			Like "*cough*" And (Like "*temp*" Or Like "*fever*")
			Like "*sore*" And Like "*throat*" And (Like "*temp*" Or Like "*fever*")
			Like "*flu[!a-z]*" Or Like "*flu" Or Like "*influenza[!a-z]*" Or Like "*influenza"

4.9 Accuracy of the Queries

Queries were assessed to determine their sensitivity and specificity using the subset of data corresponding to Week 9 & 10 of the 2005/2006 Flu Season which were manually categorised as described. The relevant query was run in MS Access and the extracted data was visually inspected and categorised as either a “True Positive” record (in which the query extracted a record that truly related to influenza) or “False Positive” record (in which the query extracted a record that did not relate to influenza). The records that were not extracted were also visually inspected and categorised as either “True Negative” (in which the query did not extract the record as it truly does not relate to influenza) or “False Negative” (in which the query did not extract the record but should have as it does relate to influenza. The accuracy, sensitivity, specificity and positive predicted value for the query was calculated by knowing the numbers of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) values for each of the queries as described in Table 4.5.

Table 4.5: Definition and Calculation of Accuracy for Queries

Test Outcome	Definition	Calculation
Accuracy	An accuracy of 100% means that all positive and negatives are recognised correctly as such	= (TP + TN) /TP+FP+FN+TN
Sensitivity	Measures how often a test will be positive out of all true positive cases.	= TP/TP+FN
Specificity	Measures how often a test will be negative out of all true negative cases.	= TN/TN+FP
Positive Predictive Value	Measures the proportion of true positive test results out of all positive results	= TP/TP+FP

4.10 Statistical Analysis

In order to compare temporal trends in the extracted Doctor on Call data and national influenza indicators, the Spearman's rank correlation coefficient was used. In this test the two datasets under study must initially be ranked. The difference between these ranks (d) are calculated, squared and used in the following equation for calculation of the rank correlation coefficient (R^2):

$$(R^2) = 1 - \frac{6\sum d^2}{n^3 - n}$$

The R^2 value is designed to fall between +1 (representing a perfectly positively correlation) and –1 (representing perfect negative correlation). The statistical significance of the result can be calculated by obtaining a p-value (probability value) by comparison with the Spearman Rank significance table. Various websites provide access to free online calculators for the Spearman's correlation coefficient and associated p-value. The one used in this study can be found at:

http://www.fon.hum.uva.nl/Service/Statistics/RankCorrelation_coefficient.html

4.11 Source of Data on Routine Flu Surveillance Indicators

Data on routine flu surveillance indicators including GP ILI consultation rates by week, number of influenza positive virological specimens by week, percentage of respiratory admission/total admissions was obtained on request from the HPSC. Data on national influenza mortality was obtained from the 2003/2004, 2004/2005, 2005/2006 summary influenza reports published by the HPSC. Influenza has been a notifiable disease since 1st January 2004 and all data on national notifications for influenza was extracted from the

Computerised Infectious Disease Reporting System (CIDR). All of these data can be found in Appendix 2.

4.12 Conclusion

This chapter has outlined the main methodologies employed in this thesis. The next chapter will provide a description of the urgent out of hours GP service (NE-Doc) assessed in this study and will summarise three years of "all calls" data to this service.

Chapter 5: Analysis of Doc on Call Data

5.1 Introduction

This chapter will present an overview of all calls made to an urgent out of hours GP service (NE-Doc) for the 2003/2004, 2004/2005 and 2005/2006 influenza season. Furthermore, the results of the validation process for four queries designed to extract flu-related records from the NE-Doc service datasets will be analysed.

5.2 Interview Results

A semi structured interview was conducted with the IT Support Manager for the NE-Doc service using questions outlined in Chapter 4, Section 4.3. The purpose of the interview was to obtain an appreciation of the specific workflow within the NE-Doc service and gather information on the IT system that supports the service. Follow-up phonecalls and emails were used to clarify any outstanding issues. A description of the system based on this interview follows:

5.2.1 Description of the North East Doc on Call Service

The North East Doctor on Call (NE-Doc) service assessed in this study is an urgent out of hours GP service established in 2000 and managed jointly by GPs and the HSE-North East. This service covers two HSE areas – the North East and Midland regions. The service operates from 6pm to 8am Monday to Friday and 24 hours a day over weekends and on bank holidays. A dedicated telephone number connects the public to the service. When a patient rings the call centre, the call is answered by specifically trained clerical operatives who record the details of the case in a computerised database system. These details include patient name, patient address, date of birth, telephone number, symptoms of illness, patient's GP and patient's medical card details. The doctor on call in the centre reviews the details of the call on the computerised system and triages the case. The patient is contacted by an on-call doctor and depending on the severity of the case, the doctor will either provide medical advice over the telephone, request the patient to attend a designated treatment centre for further attention or visit the patient in their own home. The diagnosis/outcome details for the case are then either entered on the system by the triage doctor or entered at a later stage at the treatment centre. The following morning, the call details for each patient are automatically forwarded to the patient's GP by fax. The computerised system used to collect call details is an "off the shelf" system called Adastra developed by Adastra Software Ltd (UK). The system is located on a central file server and its architecture consists of a Windows "front-end" and SQL server database. There is no facility for clinical coding of symptoms in the version of software currently being used by the NE-Doc service. This Adastra software system is used in all GP co-ops in Ireland, by NHS-direct in the UK and by a specialist healthcare call management centre in Holland.

5.3 Analysis of All Calls from "Doc on Call" dataset

Call data for three flu seasons 2003/2004, 2004/2005 and 2005/2006 were extracted from the Adastra system of the NE-Doc service. The format of these data are described in Chapter 4 Section 4.4.

In total there were 71,703 calls for the 2003/2004 season, 83,269 calls for the 2004/2005 season and 84,764 calls for the 2005/2006 season (Table 5.1). The average number of calls per week for each season was 2173, 2449, 2568 respectively and the highest number of calls occurred during the Christmas period (Week 52/53) of each year. The lowest number of calls per week occurred early in the season: in Week 40 for the 2003/2004 and 2004/2005 flu season and in Week 45 for the 2005/2006 flu season. Population contact rates with the NE-Doc service for the three flu seasons ranged from 240.2 contacts per 1000 population during 2003/2004 flu season to 252.3 contacts per 1000 population during the 2005/2006 flu season. This compares well to the 2002 average population contact rate of 221 contacts/1000 population/year as described by

Bury et al., (2006) for all calls to the 11 out of hours "Doctor on Call" co-operatives in Ireland.

Table 5.1: Analysis of Total Calls for the 2003/3004, 2004/2005 and 2005/2006 Flu Seasons

	2003/2004	2005/2005	2005/2006
Total Number of Calls	71703	83269	84764
Average Weekly Call No	2173	2449	2568
Highest No of Calls/week during season	3871 (Week 52)	5709 (Week 53)	4454 (Week 52)
Lowest No of Calls/week during season	1634 (Week 40)	1865 (Week 40)	1916 (Week 45)
NE-Doc Population Coverage	298,500	336,000	336,000
Population Contact Rate	240.2	247.8	252.3

All call data for the three flu seasons were plotted as seven day totals (corresponding to flu season week numbers) over time (Fig. 5.1). The most prominent peak in the call numbers occurred during the Christmas holiday periods: weeks 52/03-04, 53/04-05, 52/05-06. Some other minor peaks can also be seen which correspond to periods including a Public Holiday (Table 5.2). These peaks are not surprising as the NE-Doc service operates 24hrs a day over bank holiday periods and therefore an increase volume of calls would be expected.

Fig. 5.1: Distribution of Calls by Week Number for the 2003/2004, 2004/2005 and 2005/2006 Flu Seasons

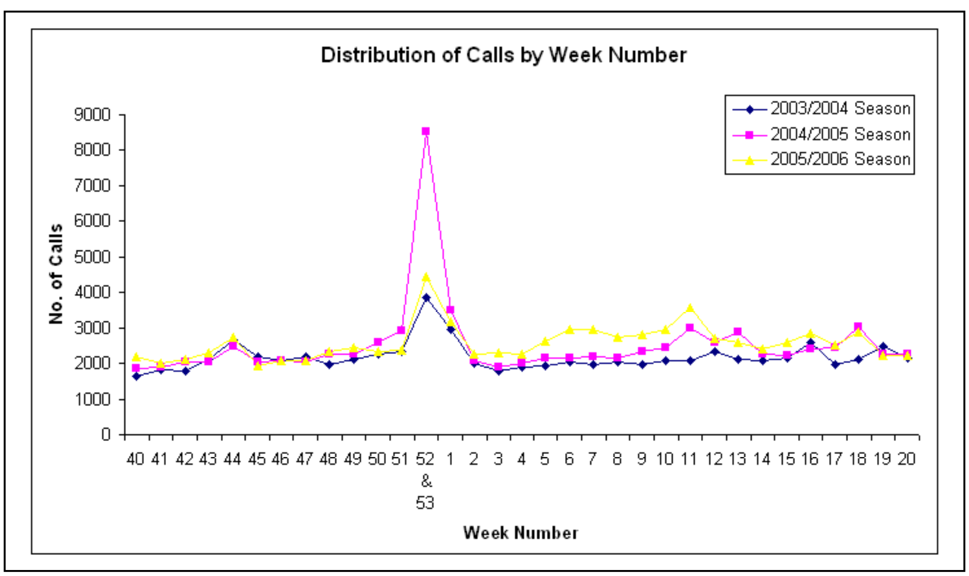


Table 5.2 Public Holidays during the Influenza Seasons

Public Holiday	2003/2004 Season		2004/2005 Season		2005/2006 Season	
	Dates	Corresponding Week No.	Dates	Corresponding Week No.	Dates	Corresponding Week No.
Halloween (Mon)	3-Nov-03	45	1-Nov-04	45	31-Oct-06	44
St. Patricks Day	17-Mar-04	12	17-Mar-05	11	17-Mar-06	11
Easter (Mon)	12-Apr-04	16	27-Mar-05	13	16-Apr-06	15
May Day (Mon)	3-May-04	19	2-May-06	18	1-May-06	18

The majority of calls for all three seasons occurred on Saturdays and Sundays (Fig. 5.2) corresponding with periods of 24 hour coverage. The most common patient age group for all three seasons was the 0 – 4 and 5 – 9 year groups followed by the 70+ years group (Fig. 5.3). Interestingly there was an increase in calls for the 0 – 4 yr age group during the 2004/2005 and 2005/2006 flu season. For each year there was a slightly larger number of calls regarding female patients than male patients (Fig. 5.4).

Fig. 5.2: Distribution of Calls by Day Received for the 2003/2004, 2004/2005 and 2005/2006 Flu Seasons

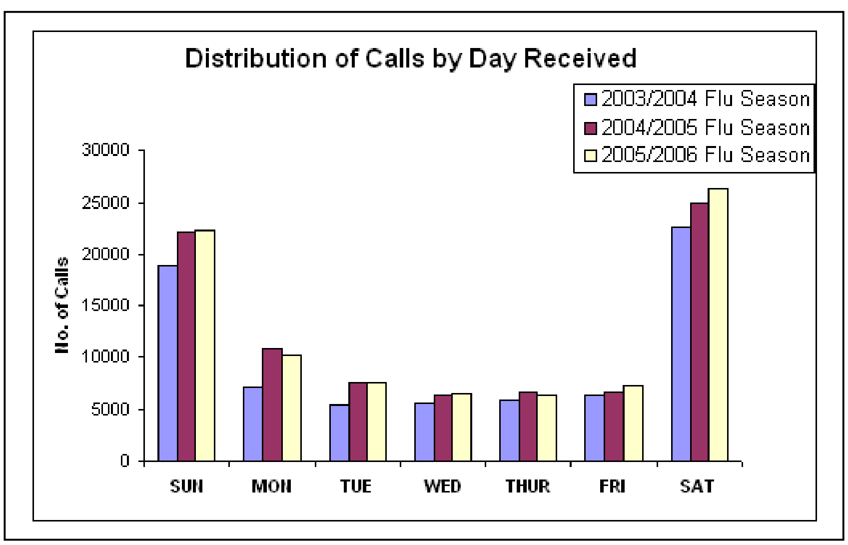


Fig. 5.3: Distribution of Calls by Patient Age Group

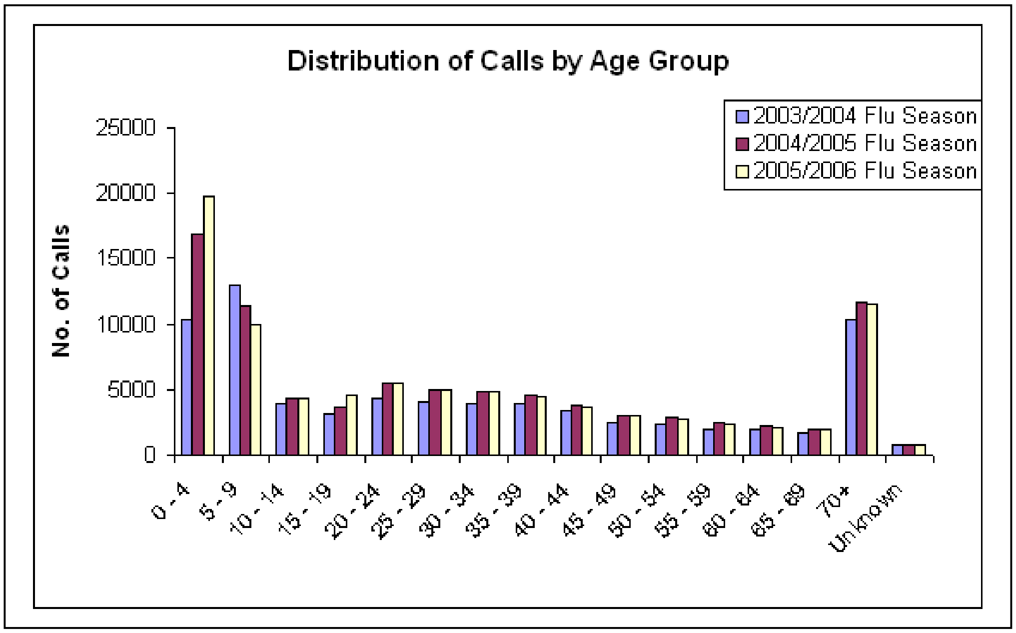
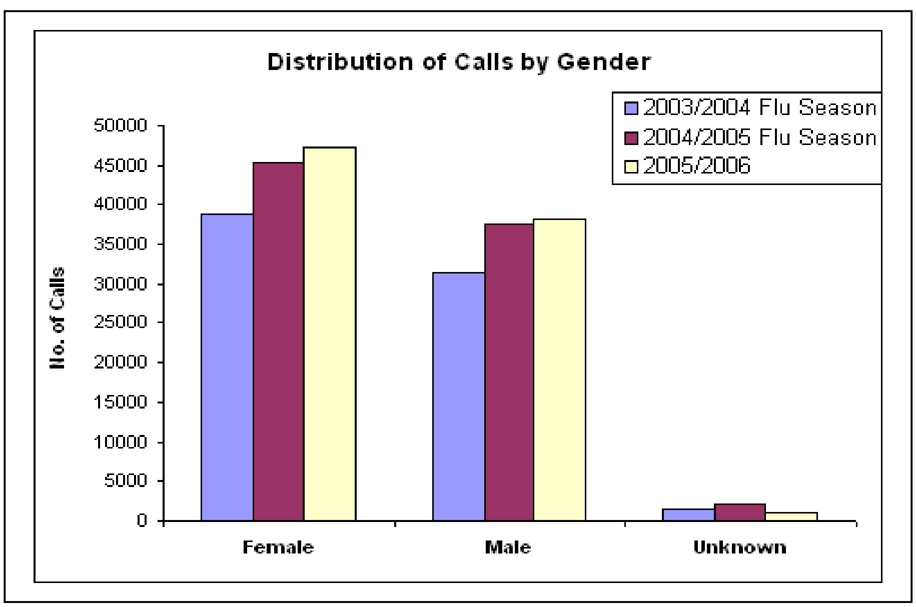


Fig. 5.4: Distribution of Calls by Patient Gender



5.4 Blank Cells in the Datasets

Datasets provided by the NE-Doc service for the three flu seasons were analysed to assess the amount of blank cells in each field. The reasons for this were two fold: firstly, this analysis would provide an insight into the recording habits of both the administrative personnel and medical personnel working on the system and secondly, this analysis would identify fields that could be used to extract the largest amount of flu-related records in subsequent queries.

There were no blank cells for four of the fields in the dataset – Call No, Date Received, Age of Patient, Gender of Patient. There were a small number of blank cells in the Patients Reported Condition Field (PRCF) for each of the three flu seasons (Table 5.3). However, the amount of blank cells in the PRCF has reduced over time from 0.75% blank fields out of all records for the 2003/2004 flu season to 0.2% blank cells out of all records for the 2005/2006 flu season. This may be due to a number of reasons including software upgrades which makes data entry easier and a policy to gather and record as much information as possible prior to passing the record to the on-call triage doctor. The amount of blank cells for the Doctors Diagnosis/Outcome Field (DD) is quite substantial with between 33.1% and 36.9% of all records with blank cells over the three flu seasons (Table 5.3). This may be due to time constraints in recording of data from patients that require consultations at designated treatment centres and at home rather than over the phone.

Table. 5.3: Blank Cells in the Datasets

DATA FIELD	2003/2004 BLANK CELLS (% OF TOTAL)	2004/2005 BLANK CELLS (% OF TOTAL)	2005/2006 BLANK CELLS (% OF TOTAL)
Total Number of Calls	71703	83269	84764
Call No.	0	0	0
Date Received	0	0	0
Age of Patient	0	0	0
Gender of Patient	0	0	0
Patients Reported condition (PRC)	178 (0.75%)	177 (0.2%)	142 (0.2%)
Doctors Diagnosis/outcome (DD)	23749 (33.1%)	29495 (35.4%)	31308 (36.9%)

5.5 Defining Search Terms for Extraction of Flu-related Calls

Case definitions for influenza like illness in sentinel surveillance programmes worldwide vary significantly. This is due in part to the many syndromes that can characterise a flu like illness. In their survey of ILI definitions across 22 European countries, Aguilera et al. (2003) found that the most commonly used symptoms were fever or high temperature, myalgia (muscle aches and pains), sudden onset and cough. Other terms often used in ILI definitions included sore throat and headache. In Ireland, ILI is defined by the HPSC as “the sudden onset of symptoms with a temperature of 38^{o}C or more, with two or more of the following: headache, sore throat, dry cough and myalgia”. In the United States, the CDC definition for ILI is much less restrictive being characterised as “fever >100^{o}F and cough or sore throat” (CDC, 2007).

In analysis of NHS Direct call data for influenza, Harcourt *et al*. (2001) defined a respiratory presenting complaint as any call with the following symptoms/syndromes: cold/flu, cough, fever, throat problems, headache. Since this first publication, a decision support system has been integrated into the data collection software to aid nurses code the incoming call into “cold/flu” calls, “fever” calls and “cough” calls. These three types of calls are used in the national syndromic surveillance bulletin as influenza or influenza like illness indicators. Turner & Kelly (2005), however, using call data from an Australian medical locum service felt that including all records relating to respiratory symptoms resulted in the inclusion of episodes that were non specific and which may affect data analysis. In their study, Turner & Kelly (2005) extracted records with the key words of “flu” and “influenza” only.

Thus, it can be difficult to decide on exact search extraction terms for influenza and influenza like illness from a free text field. So, to optimise the success of the search for ILI related records, in this study, four separate queries with specific search terms will be used. In the first query all records containing the terms “flu/influenza” will be extracted. The second query will extract all records with the term fever/high temperature and cough or sore throat and will represent the US Centres for Disease Control ILI definition. A third query was also designed to extract the terms “fever” or “high temperature” and two or more of the following headache, sore throat, cough, muscle pain (aches and pains). This third query represents the Irish case definition for influenza like illness as defined by the HPSC. A final query will be used which combines the three previous queries and provides records relating to Flu/Influenza keywords, the CDC ILI definition and the HPSC ILI definition.

5.6 Accuracy of Queries

In order to assess how well the designed queries are at extracting records with flu related keywords the accuracy, sensitivity, specificity and positive predictive value of each query was calculated. This was carried out by manual categorisation of a subset of records with specific key words and comparing the manual categorisation results with the corresponding query results for the same subset (Chapter 4, Section 4.8). This process allows the number of True Positive, False Positive, True Negative and False Negative values for each query to be calculated and concomitantly the accuracy, sensitivity, specificity and positive predictive value of each query.

5.6.1 Query 1

The accuracy of Query 1 for key search terms "Flu" and "Influenza" was assessed as described. The number of True Positives, False Positives, True Negatives and False Negatives for Query 1 was calculated and the results can be seen in Table 5.4.

Table 5.4: Results of validation for Query 1.

	True	False
Positive	225	3
Negative	5504	0

In total Query 1 correctly identified 228 out of 5732 records with the keywords "flu" or "influenza" however only 225 of these records were truly related to an influenza like illness. The three false positives that were identified contained the following text in their respective PRCF field and are clearly not ILI cases:

- "Gastric flu since last Tues."
- "Got flu injection yesterday, very sore, swollen & red round the area"
- "Patient has eaten 6 uniflu tablets."

There were 5504 True Negative records and 0 False Negative records identified. The accuracy, sensitivity, specificity and positive predictive value for Query 1 is shown in Table 5.5. Overall this query has a very high accuracy for identifying ILI records with the key words "flu" or "influenza".

Table 5.5: Calculation of Accuracy, Sensitivity, Specificity and Positive Predictive Value for Query 1

Test Outcome	Calculation	Query 1 Values	Results
Accuracy	= (TP + TN) /TP+FP+FN+TN	= 225+5504/225+3+0+5504 =5729/5732	= 99.9%
Sensitivity	= TP/TP+FN	= 225/225+0	=100.0%
Specificity	= TN/TN+FP	=5504/5504+3 =5504/5507	= 99.9%
Positive Predictive Value	= TP/TP+FP	= 225/225+3 = 225/228	= 98.6%

5.6.2. Query 2

The accuracy of Query 2 for key search terms "high temp or fever and cough or sore throat" was assessed as described previously and the results can be seen in Table 5.6 and 5.7.

Table 5.6: Results of validation for Query 2.

	True	False
Positive	294	0
Negative	5434	4

Table 5.7: Calculation of Accuracy, Sensitivity, Specificity and Positive Predictive Value for Query 2

Test Outcome	Calculation	Query 1 Values	Results
Accuracy	= (TP + TN) /TP+FP+FN+TN	=294 +5434/294+0+4+5434 =5728/5732	99.9%
Sensitivity	= TP/TP+FN	=294/294+4	98.6%
Specificity	= TN/TN+FP	=5434/5434+0	100%
Positive Predictive Value	= TP/TP+FP	=294/294+0	100%

In total Query 2 correctly identified 294 out of 298 records with the key search terms "high temp or fever and cough or sore throat". The four records which were not extracted by the query contained misspellings of the words "sore throat". Overall, the accuracy, sensitivity, specificity and positive predictive value for Query 2 is very high.

5.6.3 Query 3

The accuracy of Query 3 for key search terms "temperature or fever and two of the following: sore throat, cough, aches and pains, headache" was assessed as described previously and the results can be seen in Table 5.8 and Table 5.9.

Table 5.8: Results of validation for Query 3.

	True	False
Positive	44	3
Negative	5684	1

Table 5.9: Calculation of Accuracy, Sensitivity, Specificity and Positive Predictive Value for Query 3

Test Outcome	Calculation	Query 1 Values	Results
Accuracy	= (TP + TN) /TP+FP+FN+TN	=44+5684/5732	99.9%
Sensitivity	= TP/TP+FN	=44/44+1	97.7%
Specificity	= TN/TN+FP	=5684/5684+3	99.9%
Positive Predictive Value	= TP/TP+FP	=44/44+3	93.6%

In total Query 3 correctly identified 44 out of 45 records with the key search terms "high temp or fever and cough or sore throat" but also identified 3 False Positives. An example of one of these false positives includes the record containing the patients reported condition as "Child complaining of headaches, running a high temperature. Stomach pains". This record was extracted as the query was designed to extract keywords including "aches" and "pains". There was one false negative record which was not extracted by the query as it contained the term "Flu.Sore" and the access query did not recognise the "sore" keyword of the term "flu.sore" and therefore did not extract the record. Overall, the accuracy, sensitivity, specificity and positive predictive value for Query 3 is quite high but not as good as Query 1 and 2. Indeed, the positive predictive value is lower at 93.6% for Query 3 compared to 98.6% for Query 1 and 100% for Query 2.

5.6.4 Query 4

The accuracy of Query 4 (which combines the three flu definitions from Query 1, 2, and 3) was assessed as described previously and the results can be seen in Table 5.10 and 5.11.

Table 5.10: Results of validation for Query 4.

	True	False
Positive	512	6
Negative	5210	4

Table 5.11: Calculation of Accuracy, Sensitivity, Specificity and Positive Predictive Value for Query 4

Test Outcome	Calculation	Query 1 Values	Results
Accuracy	= (TP + TN) /TP+FP+FN+TN	= 512+5210/ 512+5210+4+6	99.8%
Sensitivity	= TP/TP+FN	= 512/512+4	99.2%
Specificity	= TN/TN+FP	= 5210/5210+6	99.9%
Positive Predictive Value	= TP/TP+FP	= 512/512+6	98.7%

In total, Query 4 correctly identified 512 records with the key search terms but also extracted six false positive records. These are the same false positive records as described for Query 1 and Query 3. Four false negative records were also identified that the query did not extract which corresponded to the same false negative records for Query 2. The remaining false negative record identified for Query 3 was extracted by Query 4 as it also contained the keyword Flu. Overall, the accuracy, sensitivity, specificity and positive predictive value for Query 4 are very high.

5.7 Comparison of Queries

Queries 1, 2 and 3 were designed for the extraction of common keywords which would recognise a flu related record according to various well established standardised definitions (i.e. CDC definition and HPSC definition). Each of these queries have high sensitivity, specificity and positive predictive values for the extraction of these keywords and each query extracts separate sets of flu related records with a small amount of overlap. Query 4 which combines the rules from Query 1, 2 and 3 is an attempt to extract as many flu related records as possible for each season. It is difficult to decide whether one particular query is “better” than another for the extraction of “real” flu cases in these circumstances but they do allow some comments to be made. For example, extraction of a dataset that only contain keywords of Flu or Influenza in the patients reported condition field represents those callers that have recognised their symptoms as flu like and have related their own “self diagnosis” to the Doctor on Call service. Clearly, callers to the service that only described their symptoms will not be included in this dataset and valuable information on real flu cases may be lost. Query 2 and 3 attempts to compensate for this fact. The CDC definition for influenza is quite broad requiring only that the patient report a temperature and either a cough or sore throat and may for this reason overestimate the number of “real” flu cases. On the other hand the HPSC definition of influenza is quite restrictive and this method may underestimate the number of “real” flu cases. Although these limitations are inherent in the extraction of data by using keyword queries, in the absence of clinical coding of the data in real time in is difficult to envisage any other easily available method for extraction of flu related records from the Doctor on Call service currently operating in Ireland. The majority of other papers relating to healthcare call centre data extract flu related records from systems that contain a decision support or clinical coding component (e.g. Espino et al., 2003; Cooper et al., 2002; Mostashari et al., 2003; Pinner et al., 2006).

5.8 Conclusion

This chapter has described all calls to an urgent out of hours Doctor on Call service in Ireland and provided a validated method for the extraction of flu-related records from this dataset.

Chapter 6: Evaluation of Doc on Call data

6.1 Introduction

This chapter will analyse patterns in both self reported influenza related records and GP diagnosed influenza records from an urgent out of hours GP service and will present the comparison of these data with National Flu Surveillance Indicators for three Flu Seasons spanning years from 2003 to 2006.

In order to assess trends in self reported illness from an urgent out of hours GP dataset, four separate queries (as designed in Chapter 5) will be used to extract flu related records from the patients reported condition field of the NE-Doc dataset. Summary data from each of these self reported illness influenza queries will be discussed. Graphical and statistical comparison of the patterns in self reported illness and national influenza activity indicators for each period will be presented. As a separate exercise, flu related records from the doctors diagnosis field of the NE-Doc dataset will also be analysed representing the equivalent of a GP ILI consultation. This analysis will provide a basis from which to assess if data from an urgent out of hours GP service can be used as a reliable source of information on influenza activity in the community. Furthermore, particular emphasis will be placed on investigating if peaks in flu related calls occur earlier than for current national influenza indicators.

6.2 Overview of Trends in National Influenza Activity Patterns (2003-2006)

National influenza activity peaked early in the 2003/2004 influenza season (HPSC, 2007b) with the highest number of ILI consultations from sentinel GPs reported in Week 46 (mid November). The majority of cases during this peak in activity were seen in 0 – 4 and 5-14 year age groups. In total over the season 624 ILI cases were reported from 35 sentinel general practices. Of the swabs sent from sentinel GPs to the NVRL for virological analysis, 149 were positive for influenza virus with influenza A/Fujian/411/2002(H3N2)-like being the most common strain. Two school outbreaks of Influenza A were reported during the 2003/2004 flu season, both occurring in September 2003. In most health board areas increases in sentinel hospital respiratory admissions and sentinel school absenteeism data coincided with the peak in ILI consultations observed early in the season. Two deaths related to influenza occurring in Week 47 and 48 were reported for the 2003/2004 season. Worldwide, influenza activity was also observed early in the 2003/2004 influenza season.

National influenza activity during the 2004/2005 season peaked during Week 1 (January) with 89 ILI sentinel GP consulations/100,000 population reported (HPSC 2007b). Unlike the 2003/2004 season, these consultations were mainly in the 15-64 year age group. In total over the season, 585 ILI cases were reported from 36 sentinel general practices. Of the swabs sent from sentinel GPs to the NVRL for virological analysis 142 were positive for influenza virus with Influenza A being the most predominant subtype. Three influenza outbreaks were reported during the 2004/2005 season in a school during week 48, a long stay care facility for the elderly in Week 3 and a school in Week 16. In most health board areas increases in sentinel hospital respiratory admissions and sentinel school absenteeism data coincided with the peak in ILI consultations. Two influenza related deaths were reported in Week 1 of the 2004/2005 season. In total 295 influenza cases were notified during the season with the highest number of notifications observed during Week 5 and Week 17. These peaks do not correspond with the ILI consultation rates for 2004/2005 and may be due to both delays in notification and notification of outbreak cases during the season. Across Europe and in the US and Canada, influenza activity also peaked during January and February 2005.

National influenza activity during the 2005/2006 flu season peaked in Week 10 with 82.5 ILI consultations/100000 population reported (HPSC, 2007b). Similar to the 2004/2005 season, the majority of cases were in the 15-64 year age group. In total over the season, 905 ILI cases were reported from 46 sentinel general practices. Of the swabs sent from sentinel GPs to the NVRL for virological analysis 132 were positive for influenza virus with Influenza A (H3) being the most predominant subtype. Four influenza outbreaks were reported during the 2005/2006 season in a nursing home during week 4 and in three schools during Week 9 and 10. In most health board areas increases in sentinel school absenteeism data coincided with the peak in ILI consultations. There was one influenza related death reported in Week 15. In total, 284 influenza cases were notified during the 2005/2006 season and peaked during Week 13, three weeks after the peak in ILI consultations for the season. Across Europe and in the US and Canada, influenza activity also peaked late in the influenza season.

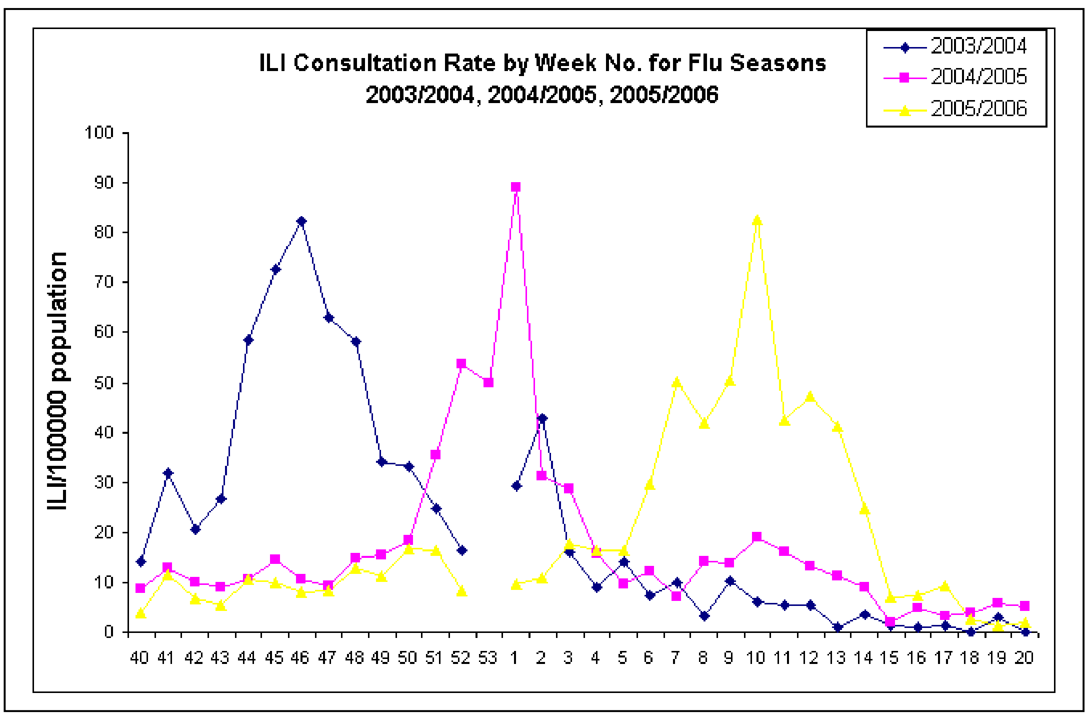
Interestingly, the major peak in national influenza activity for each the three Flu seasons as represented by ILI consultations/100,000 population do not overlap (Fig. 6.1). A further breakdown of figures for each influenza season can be found in Appendix 2.

Table 6.1: Summary data on National Flu Indicators for the 2003/2004, 2004/2005 and 2005/2006 Flu Seasons

	2003/2004	2004/2005	2005/2006
No. of General Practices participating	35	36	46
Total ILI Consultations	625	585	905
Peak ILI Rate/100,000	82.3	89	82.5
Week of Peak Clinical Activity	Week 46	Week 1	Week 10
No. of Postitive Sentinel Swabs	149	142	132
No. of Postitive Non Sentinel Swabs	112	52	36
Outbreaks of ILI	2	3	4
Notifications	53*	295	284
Influenza Related Deaths	2	2	1

Note: Influenza only became a notifiable disease on 1st January 2004

Fig. 6.1: National ILI Consultation Rates by Week Number for each of the three Flu Seasons



6.3 Overview of Extracted NE-Doc Flu Related Records (PRC Field)

The data for each of the three Flu seasons were imported into MS Access and flu related records were extracted using four specifically designed MS Access queries for searching of the "patients reported condition" free text field as there was a low percentage of blank cells in this field (see Table 5.3). Query 1 represents the extraction of all records with keywords of "flu" or "influenza", Query 2 represents the CDC definition of influenza (fever/high temperature and cough or sore throat), Query 3 represents the HPSC definition of influenza (fever/high temperature and two of the following: sore throat, cough, myalgia, headache) and Query 4 combines all three flu definitions i.e. combines Query 1, Query 2 and Query 3.

The number of flu-related records extracted from the NE-Doc dataset for each query is shown in Table 6.2. In total, for the three respective flu seasons 2376, 3004 and 2308 flu related records were extracted using Query 1 corresponding to 3.3%, 4% and 2.7% of all NE-Doc records. Similarly, 2128, 2464 and 2619 flu related records were extracted using Query 2 corresponding to 3%, 3% and 3.1% of all NE-Doc records for each respective flu season. There were 212, 263 and 334 flu related records extracted using Query 3 corresponding to 0.3%, 0.3% and 0.4% of all NE-Doc records for each respective flu season. Finally, for Query 4 which combines all three flu definitions there were 4467, 5435 and 4921 flu related records extracted corresponding to 6.2%, 6.5% and 5.8% of all NE-Doc records for each respective flu season. Quantitative analysis of the NE-Doc flu related records for each of the three flu seasons is shown in Tables 6.3, 6.4 and 6.5.

Table 6.2: Number of Flu Related Records extracted from NE-Doc dataset by Flu Season and Query.

	2003/2004 Records Extracted (% of Total Calls)	2004/2005 Records Extracted (% of Total Calls)	2005/2006 Records Extracted (% of Total Calls)	Total Flu Calls
Query 1	2376 (3.3%)	3004 (4.0%)	2308 (2.7%)	7688
Query 2	2128 (3.0%)	2464 (3.0%)	2619 (3.1%)	7211
Query 3	212 (0.3%)	263 (0.3%)	334 (0.4%)	809
Query 4	4467 (6.2%)	5435 (6.5%)	4921 (5.8%)	14823

Table 6.3: Analysis of NE Doc Flu Related Records by Query for 2003/2004 Season

	Query 1	Query 2	Query 3	Query 4
Records Extracted	2376	2128	212	4467
Male:Female Ratio for Patients	1:1.35	1.12:1	1:1.09	1:1.1
Most common age group of Patients	70+ (295)	5-9 (993)	5-9 (72)	5-9 (1263)
Week of highest no. records/total (%)	W44 (7.85%)	W44 (5.1%)	W42 (0.79%)	W44 (12.9%)
Week of ILI Consultations Peak	W46 (-2)*	W46 (-2)*	W46 (-4)*	W46 (-2)*
Flu contacts/1000 population/season	7.95	7.13	0.71	14.96

* Represents the number of weeks before (negative value) or after (positive value) the peak in national ILI consultations compared to the peak in % of NE-Doc flu related records/total number of records.

Table 6.4: Analysis of NE Doc Flu Related Records by Query for 2004/2005 Season

	Query 1	Query 2	Query 3	Query 4
Records Extracted	3004	2464	263	5435
Male:Female Ratio for Patients	1:1.23	1:1.01	1:1.18	1:1.13
Most common age group of Patients	70+ (292)	5-9 (956)	5-9 (99)	5-9 (1143)
Week of highest no. records/total (%)	W53 (11.1%)	W51 (5.79%)	W52 (0.64%)	W53 (15.2%)
Week of ILI Consultation Peak	W1(-1)*	W1 (-3)*	W1 (-2)*	W1 (-1)*
Flu contacts/1000 population/season	8.9	7.3	0.78	16.2

* Represents the number of weeks before (negative value) or after (positive value) the peak in national ILI consultations compared to the peak in % of NE-Doc flu related records/total number of records.

Table 6.5: Analysis of NE Doc Flu Related Records by Query for 2005/2006 Season

	Query 1	Query 2	Query 3	Query 4
Records Extracted	2308	2619	334	4921
Male:Female Ratio for Patients	1:1.27	1:1.05	1:1.02	1:1.15
Most common age group of Patients	70+ (238)	0-4 (1046)	5-9 (111)	0-4 (1275)
Week of highest no. records/total (%)	W6 (4.9%)	W10 (5.25%)	W6 (1.15%)	W6 (10.16%)
Week of ILI Consultation Peak*	W10 (-4)*	W10 (0)*	W10 (-4)*	W10 (-4)*
Flu contacts/1000 population/season	6.9	7.8	0.99	14.7

* Represents the number of weeks before (negative value) or after (positive value) the peak in national ILI consultations compared to the peak in % of NE-Doc flu related records/total number of records.

From analysis of Tables 6.3 – 6.5 it is clear that Query 1 and Query 2 pull out far more data than Query 3. This arises because the HPSC Flu definition is much more restrictive that the CDC definition or Flu/Influenza Keyword Query. The HPSC definition requires that "high temperature/fever" and two other flu related criteria must be met before a record can be extracted. The most common age group for Query 1 is the 70+ age group whereas for Query 2 and Query 3, the 0-4 and 5-9 year groups are the most common. Flu contacts/1000 population/season range from 6.9 to 8.9 for Query 1, 7.13 to 7.8 for Query 2 and 0.71 to 0.99 for Query 3. Query 4 represents the three combined definitions for flu and therefore the largest amount of data is extracted for this Query. The most common age group for this query is the 5-9 year group during the 2003/2004 and 2004/2005 season and the 0-4 year group during the 2005/2006 season. Flu contacts/1000 population/season ranged from 14.7 to 16.2 for Query 4 over the three Flu seasons.

Comparing the percentage of NE-Doc flu related calls out of the total number of calls per week for each query with the National ILI consultation rates per week for each season provides interesting information. In all queries for the three seasons bar one, the percentage of NE-Doc Flu calls peaked before the peak in ILI consultation rates. For example, during the 2003/2004 season, three of the four queries identified a peak in percentage of NE-Doc Flu calls during Week 44 which is two weeks earlier that the national ILI consultation rate peak reported for the 2003/2004 season (Week 46). In some cases the percentage of NE-Doc Flu calls peaked up to 4 weeks earlier than the ILI consultation rates (e.g. for Query 1, 3 and 4 during 2005/2006). The peak in the percentage of NE-Doc Flu calls never occurred after the peak in ILI consultation rates for any of the flu seasons studied.

6.4 Time Series Plots of NE-Doc Data with National Flu Surveillance Indicators

National data obtained from the HPSC on GP ILI Consultation Rates, GP Sentinel Influenza Positive Swabs, Non Sentinel Influenza Positive Swabs, Influenza Notifications, Influenza Mortalities and Respiratory Admission Data for the three Flu Seasons were compared graphical as time series plots with the NE-Doc Flu related data extracted from each of the Influenza Queries designed in Chapter 4, Section 4.8. The data was extracted based on keyword queries using the "patients reported condition" field as there was a low percentage of blank cells in this field (Chapter 5, Table 5.3). The NE-Doc data was plotted as the percentage of flu related records per week out of the total number of records per week in order to compensate for differences in weekly call rates to the NE-Doc service. The actual number of flu related records per week extracted by each query can be found in Appendix 3. The following sections present the graphical comparisons and analysis between the datasets for each of the three Flu Seasons using Query 1 only (representing "Flu/Influenza" keyword calls). The graphical comparisons and explanation for the remaining queries can be seen in Appendix 4, 5 and 6. This stratification was done for ease of reading throughout the remainder of this Chapter.

6.4.1 Comparison of 2003/2004 Flu/Influenza Calls with National Flu Indicators

Two prominent peaks, spanning Weeks 41-49 and Weeks 51-02, in the percentage of NE-Doc Flu/Influenza calls as extracted by Query 1 were identified during the early 2003/2004 Flu Season. There was a gradual tailing off of Flu/Influenza calls following these early peaks. The national ILI rate also peaked during Week 43-49 and there was a second smaller peak during Weeks 52-03 (Fig 6.2.a). The number of positive sentinel swabs contained one major peak spanning Weeks 43-49 which also coincides with the first peak in the percentage of Flu/Influenza calls (Fig. 6.2.b). The non-sentinel swabs displayed one large peak between Week 43-51 with a smaller peak during Week 01 (Fig. 6.2.c). In comparison of the percentage of Flu/Influenza calls with all three of these national indicators, the highest point of the percentage of Flu/Influenza calls occurred at least one week before the highest point of the national indicator. Comparison of the percentage of Flu/Influenza calls with Influenza Notifications during the 2003/2004 Season was not very informative as Influenza only became notifiable on 1st January 2004 and at this stage of the season, the increases in influenza activity for the season had passed (Fig. 6.2.d).

6.4.2 Comparison of 2004/2005 Flu/Influenza Calls with National Flu Indicators

One prominent sharp peak, spanning Weeks 50-03, in the percentage of NE-Doc Flu/Influenza calls as extracted by Query 1, was identified during the early 2004/2005

Flu Season. The national ILI rate also peaked sharply during Weeks 50-04 (Fig 6.3.a). The number of positive sentinel swabs contained one major peak spanning Weeks 51-03 which also coincides with the first peak in the percentage of Flu/Influenza calls (Fig. 6.3.b). The non-sentinel swabs displayed one large peak between Weeks 01-04 with some subsequent smaller peaks late in the season (Fig. 6.3.b). There was one major peak in influenza notifications occurring during Weeks 04-06 (Fig. 6.3.c), almost 6 weeks behind the peak in calls. The peak in the percentage of respiratory hospital admissions coincided with the peak in the percentage of Flu/Influenza calls occurring between Weeks 52-02 (Fig. 6.3.d). In comparison of the percentage of Flu/Influenza calls with the national ILI rate, non-sentinel swabs, influenza notifications and the percentage of respiratory admissions, the highest point of the percentage of Flu/Influenza calls occurred at least one week before the highest point of each of these national indicator. However, the percentage of Flu/Influenza calls were one week behind the week with the highest number of sentinel swabs.

6.4.3 Comparison of 2005/2006 Flu/Influenza Calls with National Flu Indicators

One prominent broad peak, spanning Weeks 04-14, in the percentage of NE-Doc Flu/Influenza calls as extracted by Query 1 were identified during the early 2004/2005 Flu Season. The national ILI rate also had a broad peak between Week 05-15 (Fig 6.4.a) which reached its highest point 4 weeks after the highest point in the percentage of Flu/Influenza calls. The number of positive sentinel swabs contained one major peak spanning Weeks 4-14 which also coincides with the peak in the percentage of Flu/Influenza calls (Fig. 6.4.b). The non-sentinel swabs displayed one large peak between Week 06-11 (Fig. 6.4.b) the highest point of which occurred 4 weeks after the highest point in the percentage of Flu/Influenza calls. There were two peaks in influenza notifications occurring between Weeks 08-15 (Fig. 6.4.c). There was one major peak in the percentage of Respiratory Hospital Admissions between Weeks 51-03 but this peak does not correspond with the prominent broad peak in the percentage of Flu/Influenza calls (Fig. 6.4.d). In comparison of the percentage of Flu/Influenza calls with all these national indicators, the highest point of the percentage of Flu/Influenza calls occurred at least three weeks before the highest point of the national indicator except for the percentage of Respiratory Admissions.

6.4.4 Use of data extracted from patients reported condition field as indicator for Flu

Flu-related records extracted from the patients reported condition field has the potential to be used as a new influenza indicator. The low rate of blank cells in this field make it ideal as a measure of self reported illness. This self reported illness indicator has been shown to mirror peaks in current national influenza indicators over three flu seasons. Further weight is added to this conclusion considering that peaks in influenza activity occurred during different time periods for each season. In most cases, peaks in these flu related calls were ahead of peaks in national indicators (up to 4 weeks ahead in some cases) meaning that this system could be used as an early alert or early warning of influenza activity in the community.

Fig. 6.2: Flu/Influenza Query Calls (Query 1) 2003/2004 Comparison with National Indicators.

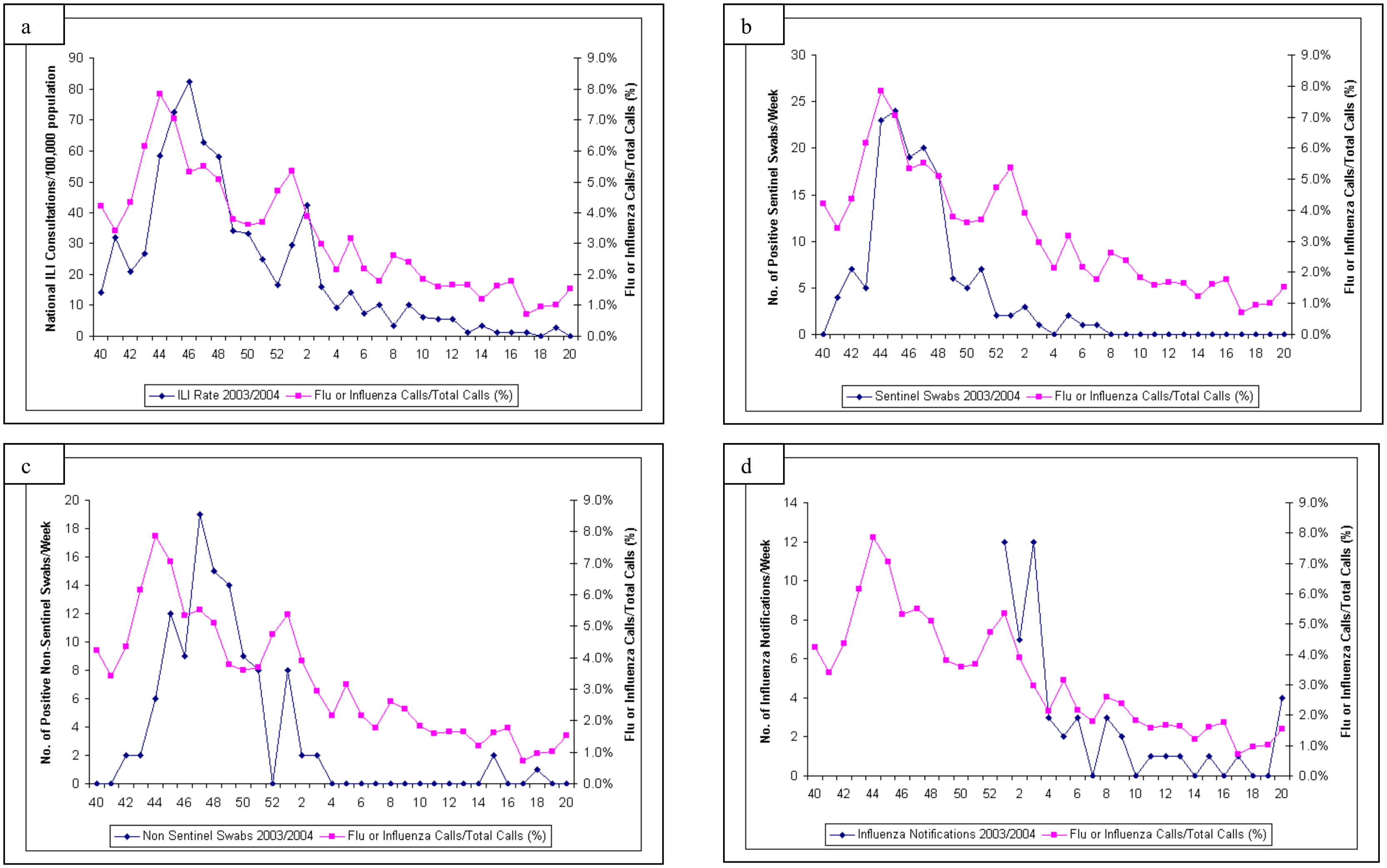


Fig. 6.3: Flu/Influenza Query Calls (Query 1) 2004/2005 Comparison with National Indicators.

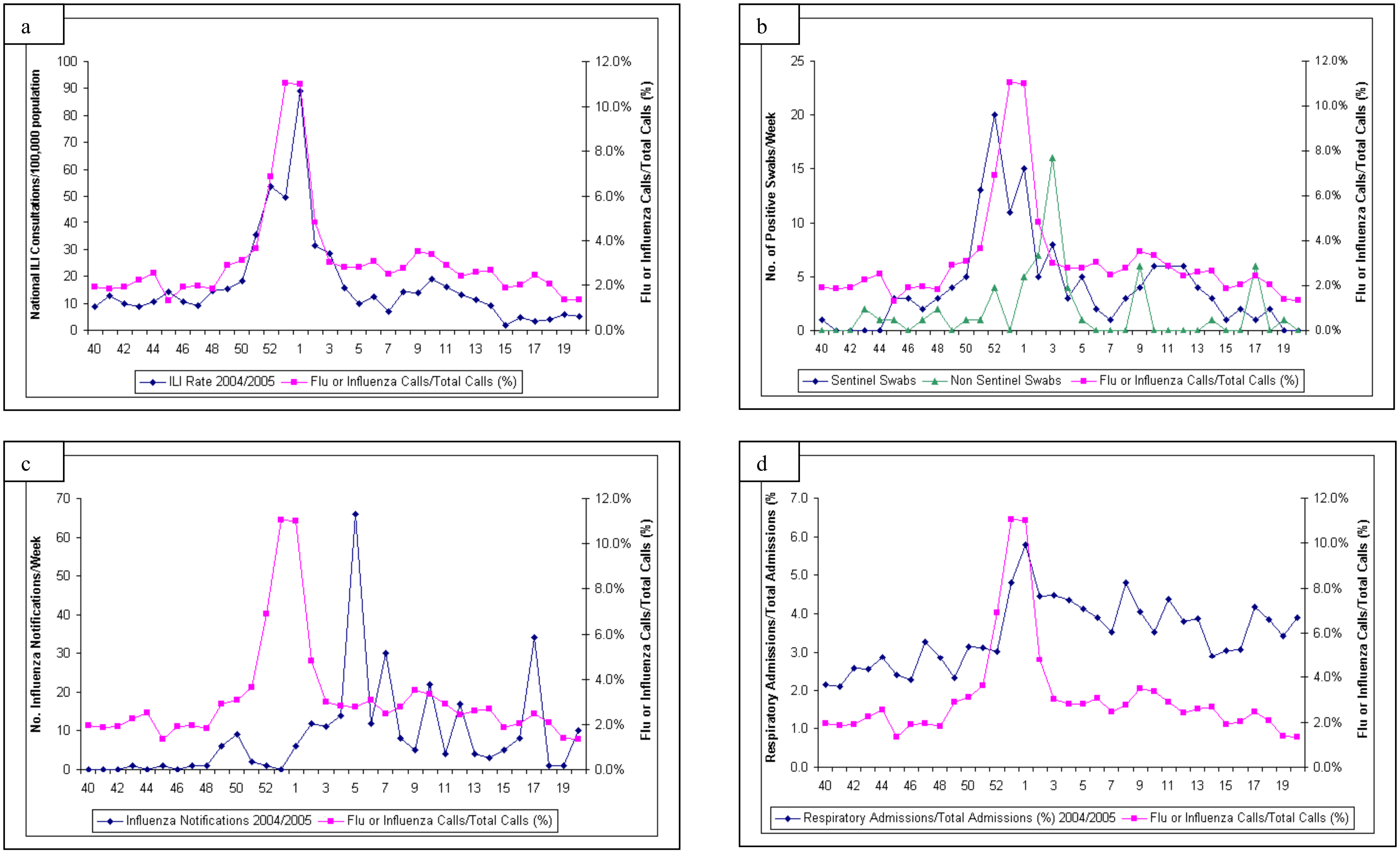
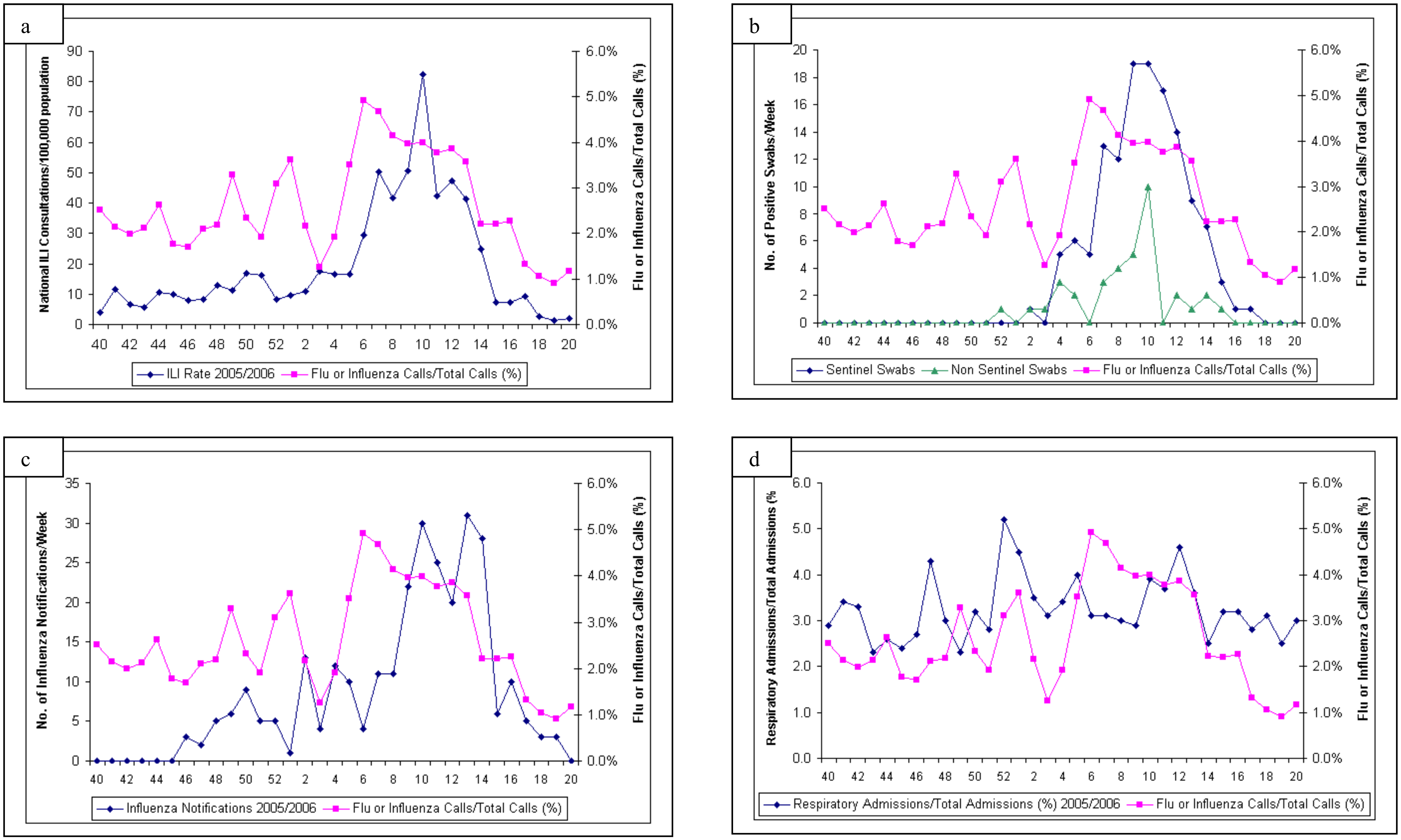


Fig. 6.4: Flu/Influenza Query Calls (Query 1) 2005/2006 Comparison with National Indicators.



6.5 Overview of Extracted NE-Doc Flu Related Records (DD Field)

National data obtained from the HPSC on GP ILI Consultation Rates for the three Flu Seasons were also compared with NE-Doc Flu related data extracted from the "doctors diagnosis/outcome" field. This keyword query was used in this instance as it was felt that it would be the most likely phrases used by doctors in diagnosing influenza. It must be kept in mind however, that for each year approximately one third of cells in this field were blank (Table 5.3). The NE-Doc data was plotted as the percentage of diagnosed Flu/Influenza related records per week out of the total number of records per week in order to compensate for differences in weekly call rates to the NE-Doc service.

Table 6.6: Analysis of NE Doc Flu/Influenza Diagnosis Related Records (DD field)

	2003/2004	2004/2005	2005/2006
Records Extracted	510	574	474
Male:Female Ratio for Patients	1:1.3	1:1.11	1:1.09
Most common age group of Patients	70+ (55)	25-29 (66)	0-4 (47)
Week of highest no. records/total (%)	W44 (1.91%)	W1 (2%)	W6 (1.25%)
Week of ILI Consultation Peak	W46 (-2)*	W1 (0)*	W10 (-4)*

* Represents the number of weeks before (negative value) or after (positive value) the peak in national ILI consultations compared to the peak in % of NE-Doc flu related records/total number of records.

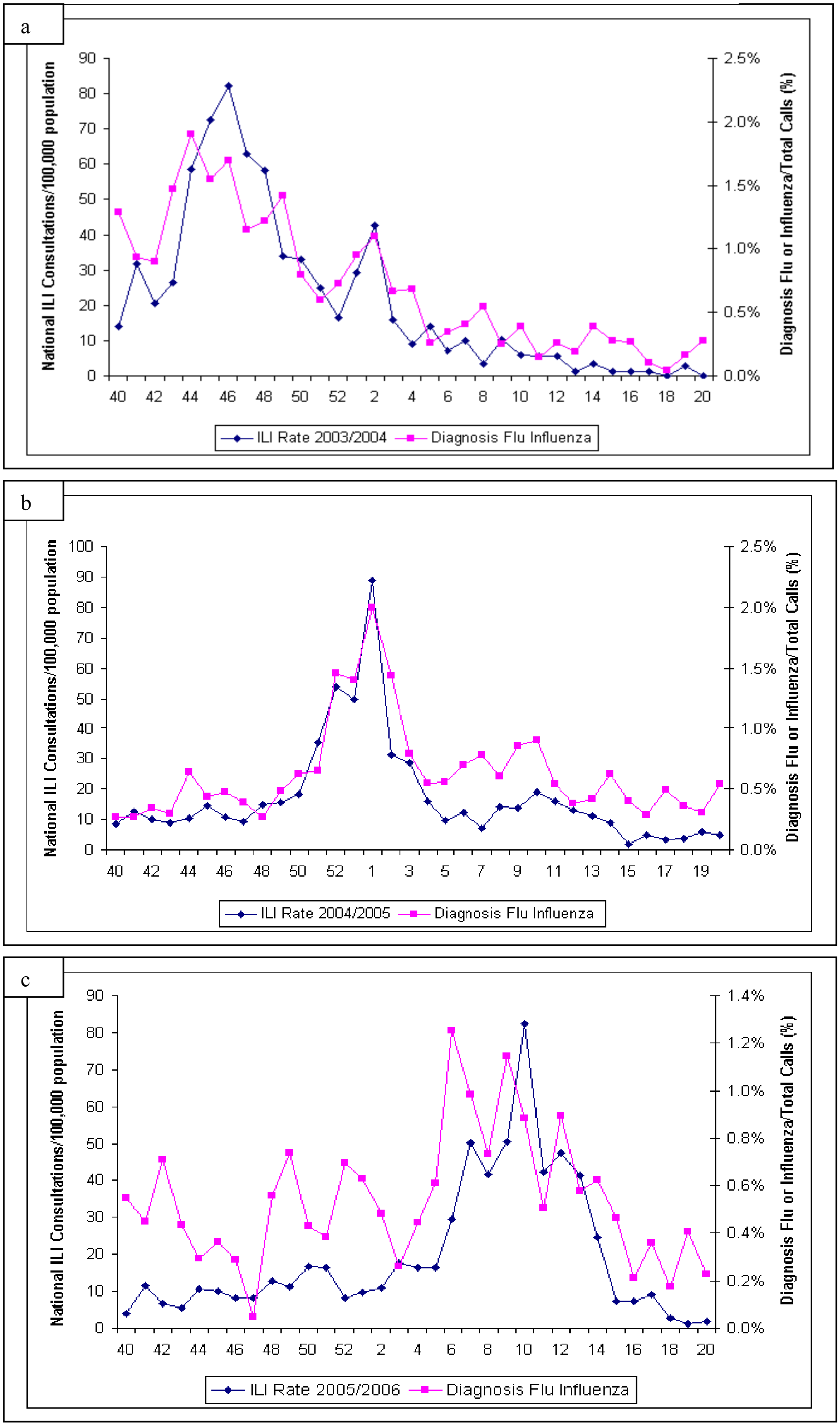
Table 6.6 presents the summary data for the extracted records containing the key words Flu or Influenza from the doctor's diagnosis field for the three flu seasons. Similar numbers of records were extracted for each season ranging from 474 to 574 records. Interestingly, the most common age group for these patients with a Flu/Influenza keyword in their diagnosis was different for each flu season. In the 2003/2004 and 2005/2005 flu season the peak in the percentage of flu/influenza diagnosis calls was ahead of the peak in national ILI consultations rates by 2 weeks and 4 weeks respectively. During 2004/2005, the peak in the percentage of flu/influenza diagnosis calls coincided with the peak in national ILI consultation rates.

Fig. 6.5 presents the time series graphical comparisons between the percentage of flu/influenza diagnosis calls and the national ILI consultation rates for each of the three Flu Seasons. During the 2003/2004 season two prominent peaks spanning Weeks 42-50 and Weeks 52-03 in the percentage of flu/influenza diagnosis calls as extracted by Query 1 were identified (Fig. 6.5.a). These peaks corresponded with the National ILI rate peaks spanning Weeks 43-49 and Weeks 52-03. During the 2004/2005 season one prominent peak spanning Weeks 51-04 in the percentage of flu/influenza diagnosis calls as extracted by Query 1 were identified (Fig. 6.5.b). This peak corresponded with the National ILI rate peak spanning Weeks 50-04. During the 2005/2006 flu season the picture is not so clear. There are a number of peaks present in the percentage of flu/influenza diagnosis calls (Fig. 6.5.c) but only one cluster of peaks spanning Weeks 05-15 coincide with the major broad peak in national ILI consultation rates (Weeks 06 – 14). This cluster of peaks, however, represents the highest percentage of flu/influenza diagnosis calls for the season.

6.5.1 Use of data extracted from doctors diagnosis field as indicator for Flu

Flu-related records extracted from the doctors diagnosis/outcome field has the potential to be used as a new influenza indicator. Although there are quite a high percentage of blank cells in this field which reduces the number of overall records extracted, those that are extracted have been medically diagnosed with a flu-related illness. This doctors diagnosis indicator has been shown to mirror quite well peaks in current national indicators over three flu seasons, albeit with a much lower number of extracted records.

Fig. 6.5: Doctor Diagnosis Flu/Influenza Query



6.6 Statistical analysis of Times Series

In order to determine how strong the temporal relationship is between the national Flu indicators and the NE-Doc flu related calls, Spearman's rank correlation coefficient for each time series pair was calculated. If there is a strong relationship between the two time series a trend in the ranked scatter data should be observed, in other words as one time series increases or decreases so does the other time series. The extent of this relatedness can be expressed statistically by calculating a rank correlation coefficient (Chapter 4, Section 4.10). If the correlation coefficient is equal to one, the two series are perfectly positively correlated. If the correlation coefficient is equal to minus one, the two series are perfectly negatively correlated. If the correlation coefficient is equal to zero, the two series are not related. In practice, if there is a relationship between the time series the correlation coefficient is rarely found to be exactly +/-1, however the closer to +/- 1 the greater the relationship between the series. The benefit of Spearman's rank correlation coefficient test is that unlike the usual Pearsons correlation coefficient, the test does not require that the underlying data be normally distributed. Finally, the Spearman's rank correlation coefficient has been used previously (Lewis et al., 2002; Doroshenko et al., 2005) to infer statistical correlation between syndromic surveillance datasets for influenza.

Table 6.7 shows the correlation coefficients for the graphical time series comparisons presented in Section 6.4 (i.e. for Query 1 Flu/Influenza Calls). All comparisons bar one with national indicators for each of the three flu seasons were statistically significant, in other words as one time series increased or decreased there was also a correlated increase or decrease in the corresponding national indicator. The only pair that was shown to have no statistical significance was the comparison of Flu/Influenza calls for 2004/2005 with 2004/2005 Flu Notifications. From visual inspection of the graphical comparison between these datasets (Fig. 6.3.c) it is clear that there is low similarity between these two time series. Indeed, for this particular year the notifications data did not coincide with the national ILI consultations peak either and delays in notifications and specific outbreaks have been deemed responsible for this. Overall, the NE-Doc Flu/Influenza call data for the three Flu seasons is positively correlated with the current national flu indicators suggesting at a minimum that this data is a good marker for influenza activity in the community.

Statistical testing of the three other queries with national indicators also demonstrated a high level of significance in most cases. However, as there is a time lag in some of the national indicators (between 1- 4 weeks) in comparison to the call data the statistical significance of the correlation coefficients could be further increased by shifting in time one of the datasets. For example, if the national ILI rate for 2005/2006 is shifted forward by 1 week and then compared with the Flu/Influenza (Query 1) calls, the Spearman's rank correlation coefficient increases from 0.679 to 0.773. These types of time lags are a common phenomenon in biological systems and therefore time series comparisons.

Table 6.7: Spearman's rank correlation coefficient and statistical probability for comparison of correlation between Query 1 (Flu/Influenza calls for patients reported condition field) and National Influenza indicators.

National Indicator	% Query 1 Calls/Total Calls 2003/2004	% Query 1 Calls/Total Calls 2004/2005	% Query 1 Calls/Total Calls 2005/2006
ILI Rate	0.901 ($p<0.001$)	0.709 ($p<0.001$)	0.679 ($p<0.001$)
Sentinel Swabs	0.854 ($p<0.001$)	0.764 ($p<0.001$)	0.672 ($p<0.001$)
Non-Sentinel Swabs	0.689 ($p<0.001$)	0.358 ($p<0.04$)	0.52 ($p<0.003$)
Notifications	NA	0.336 ($p<0.054$)*	0.522 ($p<0.003$)
%Respiratory Admissions	NA	0.546 ($p<0.001$)	0.349 ($p<0.048$)

*Not statistically significant

6.7 Conclusion

This chapter has demonstrated a statistically significant temporal relationship between influenza related urgent out of hours GP calls with current national influenza indicators over three influenza seasons. Furthermore, in the majority of cases, the peaks in healthcare call centre flu related data (both for self reported patient illness and doctors diagnosed illness) were ahead of the peaks in the national indicators, by up to 4 weeks in some cases, suggesting that this call data is an important source of early warning or early alert flu activity in the community.

Chapter 7: Discussion

7.1 Introduction

The purpose of this thesis was three fold. Firstly, to assess whether or not influenza related patient contact records from an urgent out of hours GP service dataset (covering an Irish population of approximately 336,000) could be easily extracted in the absence of any clinical coding. Secondly, to determine if the patterns in this extracted flu related call data closely matched the patterns seen in national flu activity data and finally, to test the hypothesis that peaks in these influenza related urgent out of hours calls may be more timely than peaks in current national influenza indicators and may therefore, better represent the true temporal epidemiology of flu in the community setting in Ireland. This chapter presents a discussion of the main findings, describes the main advantages and limitations of the study and provides recommendations for national collection of GP out of hours call data as a new influenza activity indicator.

7.2 Overview of Findings

A validated method for the extraction of flu related call data from a Doctor on Call service dataset using a commonly available database software package, Microsoft Access has been described in this study. Specific queries were designed based on standardised widely accepted definitions for influenza like illness with some minor modification which included using in one instance a commonly used lay term for a clinical phrase (i.e. "aches and pains" substituted for myalgia in the Irish ILI definition). This was necessary due to the nature of the Doctor on Call service which records the patient's symptoms in their own words. In the absence of any clinical coding or symptom to syndrome mapping of the data, these keyword queries are, essentially, the only way to identify flu related calls.

The three year period examined in this thesis represented time periods in which the peaks in national influenza activity varied over the three seasons, i.e. in 2003/2004 the peak in national ILI rates occurred early in the season, during 2004/2005 the peak occurred during the middle of the season and in 2005/2006, the peak occurred at the end of the flu season. Overall, patterns during each season in extracted flu related call data was shown to closely match patterns in national ILI rates and positive sentinel virological swabs. Interestingly, in most cases analysed, the week with the highest number of flu related calls for each season was at least 1 week ahead of the peaks in ILI rates and sentinel virological peaks. Other national indicators (e.g. non sentinel swabs, proportion of respiratory admissions) were less comparable with extracted call data due to inherent problems in collection of these data at national level including inadequate and under-reporting issues although they were presented for completeness sake in this thesis. A statistically significant temporal correlation between the extracted flu-related call data and national flu indicators was demonstrated using the Spearman's rank correlation coefficient. This statistical method has been used previously for comparison of influenza time series surveillance datasets in both the US (Lewis et al., 2002) and the UK (Duroshenko et al., 2005). Further statistical analysis could be done using a cross correlation coefficient function (for times series analysis) which would take into account any time lags between the datasets, however, this analysis is beyond the scope of this thesis. Overall, the correlation between the extracted flu related calls and national indicators suggests that this new source of data would at the very minimum be a valuable addition to the current surveillance system and has the potential to provide an early warning of heightened influenza activity in the community setting.

7.3 Evaluation of "Doc On Call" Dataset

In 2002, "11 GP co-operatives provided cover to 1.5m people (over 1/3 of the Irish population) and had almost 340,000 contacts with patients during out-of-hours periods" (Bury et al., 2006). Indeed, reliance on GP co-operatives for urgent on-call services has only increased in recent years. This represents a significant healthcare interaction, the details of which are routinely recorded electronically. This thesis set out to examine the suitability of this data source as a new surveillance indicator for influenza and assess the potential to utilise this data to provide a national influenza syndromic surveillance system for Ireland.

7.3.1 Advantages and Limitations of the Data

The development of the out of hours GP co-operatives and the acceptance of this new service by both the medical profession and the public has meant that the service has been widely used by a cross section of the Irish population. The dataset can, therefore, be considered representative of the general population having a good patient mix and geographic spread. The scale of the interaction of this service with the population and its electronic availability are not the only advantages that this dataset can provide. The data is entered onto the system in "real-time" during the patients call to the service which means that the information on the patient's reported condition and often the doctors diagnosis is available immediately for analysis. This adds a flexibility to the system which may have important implications during the early phases of a pandemic when daily monitoring of trends in influenza transmission may be required. Other than the IT requirement to extract the information once a day or once a week, depending on the surveillance protocol, there would be no additional costs, work load or change in work practices to the service envisaged as the data is being routinely collected electronically anyway. Trained call operatives would continue to answer and log calls and medically qualified personnel would continue to triage patients as before.

Another major benefit of the call data is that it should provide an early warning of influenza outbreaks. In the current influenza surveillance system, a national weekly report is produced every Thursday by the HPSC during the influenza season describing the data gathered on ILI consultations and virological samples for the previous epidemiological week. If the Doctor on Call data were to be included in this weekly report and provide information on the previous weeks influenza related calls to coincide with the reporting of other national indicators, the analysis from this thesis has demonstrated that the increase and peaks in the call data generally occurs prior to peaks in current national indicators and therefore this would be reflected during the collection of weekly data. In other words, by collecting the data from the Doctor on Call service in the same manner as data is currently collected and reported for other national indicators (i.e. one week behind), the flu-related call data would still provide an earlier indication of rises in influenza activity than other national indicators. Thus, both the medical profession and the public can be informed earlier of a threatened influenza outbreak allowing better allocation of limited resources and facilitate early interventions.

Another interesting aspect of the dataset which could be explored is that patient demographic data are linked to the call record. Although this information was not used in this study due to data protection issues, it may be possible in the future to spatially and temporally map postcodes for influenza related calls. This type of analysis has the potential to identify influenza outbreaks and their points of origin and could provide valuable insights on the geographical dissemination of a new influenza virus. Analysis of this type has already been employed with influenza military emergency and primary care surveillance data in the US (Lewis et al., 2002) using the geospatial mapping software, ArcView. A "moving picture" or tracking of cases through location and time was demonstrated to be possible and could clearly add a new dimension to the epidemiological analysis of influenza. Indeed this type of spatial analysis could be employed for any surveillance initiative for which this dataset is employed.

There are, of course, some caveats to using the data from the Doctor on Call service as an indicator of influenza activity. In the first instance, this service is not a 24 hour a day service – it operates only out of normal GP working hours (6pm to 8am). This means that some patients with ILI which present at the GP surgery during the day rather than contact the on-call service out of hours are still being lost to the surveillance process. Indeed, it is difficult to estimate what proportion of all GP consultations the out of hours service caters for. In the future, however, with the computerisation of all GP practices it may be possible not only to accurately estimate the proportion of all GP consultations the out of hours service represents but also to capture equivalent datasets for the normal working hours period. Finally, although the majority of GPs in Ireland do participate in these out of hours co-operatives, the service is not ubiquitious throughout Ireland and as the scheme is voluntary among GPs and the GP practice itself is an independent business, the service may never reach 100% coverage of the Irish population.

The other most obvious limitation of the data from the Doctor on Call service is the recording of the patient's reported condition and doctors subsequent diagnosis in a "free text" format only. Clearly, analysis of free text fields to identify trends and patterns in the data over time can be a difficult task. Other published studies for the analysis of flu related health care syndromic surveillance data all rely on the fact that the software systems incorporate some form of symptoms to syndrome mapping of the data collected (Espino et al., 2003; Yih et al., 2006) or utilises a decision support system designed to categorise all calls (Harcourt et al., 2001; Cooper et al., 2002; Doroshenko et al., 2005).

In the version of software currently used to collect data from the Doc on call service in Ireland there is no facility for the mapping of patient symptoms to syndromes or clinical coding of the data. There is the ability within the system, however, to incorporate a dropdown menu which could be selected by the data inputter or the doctor if they deemed a case was consistent with an ILI. This would mean that there could be a standard syntax used for identifying influenza related cases across Doctor on Call regions. In practice though this may be difficult to implement as every call would need to be assessed as flu-related or not. Additional time on each call and a change in work practice by the service would be necessary as well the provision of guidance to the service on the criteria for ILI related cases. Furthermore, there would then be an onus on the service to identify ILI cases by this method and this would mean, if the new protocol is not strictly adhered to, that true flu cases are under-reported. The benefit of the data mining technique described in this thesis is that personnel on the ground are not currently required to directly report flu related cases making the surveillance methodology initially more acceptable to the data providers and probably more complete for the data gatherers. As a future development, this aspect of the surveillance methodology could be reviewed with the Doctor on Call service and in conjunction with software updates may indeed become a viable alternative to the retrospective querying of free text fields for influenza related patient contacts.

The Doctor on Call service, therefore, appears to be an important and useful new data source on influenza activity in the community. The quality of the data is of a high standard as it compares well with currently used national indicators and similar syndromic systems for the surveillance of influenza have been implemented in other countries. Overall, this new data source has the potential to provide a very promising influenza early warning system but one pertinent question remains: on a practical level can the methodology employed in this thesis be extended to provide the basis for a national syndromic influenza surveillance system?

7.4 A Scalable Solution?

There are two main factors that must be taken into account in deciding whether or not this methodology can be "scaled up" to provide a national early warning flu syndromic surveillance system. In the first instance, will the designed queries be suitable for the extraction of flu-related calls across the various Doctor on Call services in the different HSE regions of Ireland? And secondly, can the turnaround time for the data be made short enough to allow advantage to be taken of the influenza early warning potential of the call data?

The methodology of this thesis is based on the design of an MS Access database allowing imported data to be directly queried. Clearly, the use of such a generic and widely available database program means that this methodology can be easily replicated. However, in order to utilise the queries designed in this thesis for a national surveillance system, a sample of data from each regional participating Doctor on Call service must be manually inspected or checked to identify any regional variations in the manner in which free text data and particularly flu related call data is entered on the system. In other words, the syntax in the free text fields may differ or vary from region to region and consequently the syntax of the queries may need to be adopted accordingly. Although this will not be a problem for simple keyword queries, like Query 1 designed in this thesis which only identifies records that contain the word “Flu” or “Influenza”, it will be more important when more complex queries are being used to search for flu related records. Overall, it is expected that there would not be wide variation between regions in the recording of commonly used terms like sore throat, cough and headache, etc., but it must be considered prior to collection of national syndromic data.

As discussed previously, it is envisaged that data from the Doctor on Call service could be available electronically on a Monday for the previous epidemiological flu week. This would allow two days for data analysis and inclusion of aggregate results in the Weekly Flu Surveillance Report produced by the HPSC. Therefore, it is entirely possible, with co-operation from relevant personnel during the flu season, to have aggregate numbers of flu related calls for the previous week available in quite a short period of time retaining the benefit of the early alert capacity of the data.

In summary, it is considered that this thesis provides a scalable solution which could be implemented as a national syndromic flu surveillance system.

7.5 Recommendations for National data collection and analysis

It is suggested that data collection from each participating Doctor on Call service be organised by each of the eight regional public health departments in Ireland with weekly reporting of aggregate results to the HPSC. This would operate in a similar manner to collection of data for school absenteeism figures. It would be essential to identify a dedicated IT professional in each Doctor on Call Service and a surveillance professional in each public health department to be involved in the timely collection, analysis and reporting of the data to the HPSC during the flu season. Furthermore the development of a general Standard Operating Procedure (SOP) and an initial national pilot project would be useful to assess how quickly the data could be turned around on a routine basis. The SOP would be important to identify the minimum dataset required from the Doctor on Call service, determine which query to use for the extraction of flu related records and which fields the query should be run on, i.e. the patients reported condition field and/or the doctors diagnosis field. As an initial national pilot, a minimum dataset containing a line listing of all calls with the fields, Date Called Received, Patients Reported Condition and/or Doctors Diagnosis, would be requested on a weekly basis from each participating Doctor on Call service. The designated person from the Doctor on Call service with responsibility for extraction of total calls for each week would liaise by email with a dedicated public health professional within each region, in most cases the Surveillance Scientist. The Surveillance Scientist would have the responsibility of stratifying the data by Week Number, importing the data into MS Access, querying the data to extract flu-related call records and emailing the aggregate results of the query to the HPSC for inclusion in the Weekly Influenza report. In practice, the extract for the previous week (Mon to Sun) would need to be emailed to local Public Health on the Monday of the current week. The data would need to be analysed by Tuesday of the current week and the aggregate results sent at the latest to the HPSC by lunchtime on Wednesday for inclusion in the Weekly Flu report which is issued every Thursday by the HPSC. Depending on the workload generated by this entire process, new queries could be incorporated at a later date.

In the future, it may be possible to reduce the amount of time to extract the relevant flu records even further by eliminating the need to export all call records from the Doctor on Call service. This could be achieved if, the Doctor on Call service agreed to directly query the Adastra database tables using, for example, the linked tables facility in Microsoft Access and the type of specific queries designed in this thesis. In this way, on a routine basis, only flu related records would be retrieved by the Doctor on Call service during their initial data extraction procedure. Apart from eliminating the time involved in exporting large amounts of data on all calls and subsequently importing such into an into a separate database for analysis, there are a number of other advantages to this approach to routine syndromic flu surveillance. The most important aspect is that queries would automatically only include relevant flu related call records, eliminating the extraction of non-flu related call records. This would be an added security in terms of data protection. In addition, the only downstream processing needed before results could be forwarded to the HPSC would be the counting of the number of extracted flu related records. Finally, file sizes would be significantly smaller and easier to handle. This method would require some detailed discussions with the Doctor on Call service but may in the long term prove the most efficient manner for extraction of flu related records.

In any event, the routine collection of this new flu indicator would not be a trivial task to implement nationally but would have long term benefits in relation to the timely monitoring of influenza activity in the community setting.

7.6 Conclusion

Results from this thesis has demonstrated that it is possible to extract flu-related patient contact records from an urgent out of hours GP service in Ireland and that the data has the potential to be utilised as an early alert system for seasonal and pandemic influenza. Clearly, a national flu syndromic surveillance system based on the Doctor on Call service would provide valuable epidemiological information at a relatively low resource and infrastructure cost. In the long term, improvements in routine influenza surveillance and outbreak alerts will impact on patient care and patient outcomes. Finally, this thesis not only describes a scalable solution for influenza surveillance but also demonstrates a methodology that could be applied to other important areas of Public Health surveillance including non-infectious disease related events. In other words, using the urgent out of hours GP service dataset as a marker for health in the community, keyword queries for the extraction of patient contact call records for other areas of public health interest could easily be developed in the future.

Bibliography

Abdool-Karim, S. S., & Abdool-Karim, Q. (1991). Under-reporting in Hepatitis B notifications. *South African Medical Journal* **79:** 242–244.

Adastra Software Ltd. Website (2007) Available at: http://www.adastra.com.

Aguilera J.F., Paget W.J., Manuguerra J.C. on behalf of EISS and EuroGROG. Survey of Influenza Surveillance Systems in Europe. EISS-EuroGROG Report, December 2001.

Barrett, P., & Lau, Y. K. (1996). Incompleteness of statutory notification of bacterial gastro-intestinal infection. *Public Health* **111:** 183–185.

Brabazon, E. D., O'Farrell, A., Murray, C., Carton, M., & Finnegan, P. (2007). Under-reporting of notifiable infectious disease hospitalisations in a health board region in Ireland: room for improvement? *Epidemiology & Infection,* 1-7 [Epub ahead of print].

Brabazon, E. D., O'Farrell, A., Murray, C., & Finnegan, P. (2004). Trends in viral meningitis hospitalisations and notifications in the North Eastern Health Board (1997 - 2001): a cause for concern? *Irish Medical Journal* **97:** 306-8.

Bury, G., Dowling, J., & Janes, D. (2006). General Practice Out of Hours Co-operatives - Population Contact Rates. *Irish Medical Journal* **99 (3):** 73-75.

CDC (2006) The Global HIV/AIDS pandemic, *MMWR Morb Mortal Wkly Rep* **55:** 841-4.

CDC (2007) USA ILI Definition. Available at: http://www.cdc.gov/flu/weekly/fluactivity.htm

CDSCNI. (2006). Communicable Disease Surveillance Centre Northern Ireland (CDSCNI) Report: Enhanced Surveillance of Influenza in Northern Ireland Summary Season 2005-2006.

Collins, G. (1997). Notification of infectious diseases (SHB). *Infoscan* **7**.

Cooper, D. L., Smith, G. E., Hollyoak, V. A., Joseph, C. A., Johnson, L., & Chaloner, R. (2002). Use of NHS Direct calls for surveillance of influenza - a second year's experience. *Communicable Disease and Public Health* **5:** 127-131.

Davies, P. (1999). Catching Cold. Published by Penguin Group, London. .

Davison, K. L., Crowcroft, N. S., Ramsay, M. E., Brown, D. W., & Andrews, N. J. (2003). Viral encephalitis in England, 1989–1998: what did we miss? *Emerging Infectious Diseases* **9:** 234–240.

Declich, S., & Carter, A. O. (1994). Public health surveillance: historical origins, methods and evaluation. *Bulletin of the World Health Organisation* **72:** 285-304.

Doroshenko, A., Cooper, D., Smith, G., Gerard, E., Chinemana, F., Verlander, N. & Nicoll, A. (2005) Evaluation of syndromic surveillance based on National Health Service Direct derived data--England and Wales. *MMWR Morb Mortal Wkly Rep.*, **54:** Suppl:117-22.

EISS (2007). European Influenza Surveillance Scheme Website. Available at: http://www.eiss.org/index.cgi

Espino, J. U., Hogan, W. R., & Wagner, M. M. (2003). Telephone Triage: A timely data source for Surveillance of Influenza-like Diseases. *AMIA Annu Symp Proc*, 215-219.

Gottlieb, M. S. (1981). Pneumocystis pneumonia--Los Angeles. *MMWR Morb Mortal Wkly Rep,* **30:** 250-252.

Harcourt, S. E., Smith, G. E., Hollyoak, V. A., Joseph, C. A., Chaloner, R., Rehman, Y., Warburton, F., Ejidokum, O. O., Watson, J. M., & Griffiths, R. K. (2001). Can calls to NHS-Direct be used for syndromic surveillance? *Communicable Disease and Public Health* **4:** 178-182.

Harvey, I. (1991). Infectious disease notification – a neglected legal requirement. *Health Trends* **23:** 73–74.

Health-Matters. (2002). Survey Highlights Satisfaction with NEDOC service. *Health Matters* **4**.

HPSC (2007a) Pandemic Influenza Prepardness for Ireland (Draft Report). Available at: http://www.ndsc.ie/hpsc/A-Z/EmergencyPlanning/AvianPandemicInfluenza/Guidance/

HPSC (2007b) Irish Influenza Surveillance Reports. Available at: http://www.ndsc.ie/hpsc/A-Z/Respiratory/Influenza/Publications/

Infectious-Disease-Regulations-Ireland. (1981). SI No. 390 of 1981 (amended by SI No. 268 of 1985, SI No. 288 of 1988, SI No. 384 of 1996, SI No. 151 of 2000, SI No 115 of 2003 and SI No 707 of 2003).

Kavic, S. M., Frehm, E. J., & Segal, A. S. (1999). Case studies in cholera: lessons in medical history and science. *Yale J Biol Med* **72:** 393-408.

Langmuir, A. D. (1963). The surveillance of communicable diseases of national importance. *New England Journal of Medicine* **268:** 182-192.
Lerman, Y., Chodik, G., Aloni, H., & Ashkenazi, S. (1999). How valid is the official data from the Health Department on reported morbidity in Israel? Hepatitis A as an example. *Harefuah* **136**.
Lewis, M.D., Pavlin, J.A., Mansfield, J.L., O'Brien, S., Boomsma, L.G., Elbert, Y. & Kelley, P.W. (2002) Disease Outbreak Detection System Using Syndromic Data in the Greater Washington DC Area. *Am J. Prev. Med*., **23(3):** 180-186.
Mackenzie, S. (2005). Official Chinese Bird Flu deaths could be "tip of the iceberg". Available at: http://www.newscientist.com/article.ns?id=dn8372
Marcuse, E. (2005). Quoted in US News and World Report. Available at: http://health.usnews.com/usnews/health/briefs/infectiousdiseases/hb050720a.htm
Marr, J. S., & Calisher, C. H. (2003). Alexander the Great and West Nile virus encephalitis. *Emerg Infect Dis* **9:** 1599-1603.
McNamara, A. (2006). Evaluation of the Irish Influenza Surveillance System (unpublished HPSC report). .
Meltzer, M. I., Cox, N. J., & Fukuda, K. (1999). The Economic Impact of Pandemic Influenza in the United States: Priorities for Intervention. *Emerging Infectious Diseases* **5**.
Mostashari, F., Fine, A., Das, D., Adams, J., & Layton, M. (2003). Use of ambulance dispatch data as an early warning system for community wide influenza like illness, New York City. *J Urban Health* **80:** 143-9.
NDSC. (2001). Review of Notifiable Diseases and the Process of Notification. Dublin: National Disease Surveillance Centre, 2001.
Papagrigorakis, M. J., Yapijakis, C., Synodinos, P. N., & Baziotopoulou-Valavani, E. (2006). DNA examination of ancient dental pulp incriminates typhoid fever as a probable cause of the Plague of Athens. *Int J Infect Dis* **10:** 206-14.
Pinner, R., Harmon, R., & Platt, R. (2006). Nurse Call Data for Detection of Influenza-Like Illness. *Advances in Disease Surveillance* **1:** 76.
Reid, A. H., & Taubenberger, J. K. (2003). The origin of the 1918 pandemic influenza virus: a continuing enigma. *Journal of General Virology* **84:** 2285 - 2292.
Rieldel, S. (2005). Plague: from natural disease to bioterrorism. *BUMC Proceedings* **18:** 116-124.
Smolinski, M.S., Hamburg, M.A. & Lederberg, J. (2003): Microbial Threats to Health: Emergence, Detection, and Response. *Editors*, Committee on Emerging Microbial Threats to Health in the 21st Century.
Strauss, R., Fulop, G., & Pfeifer, C. (2003). Hepatitis C in Austria1993–2000: reporting bias distort HCV epidemiology in Austria. *Eurosurveillance* **8:** 113–118.
Sullivan, K. M. (1996). Health impact of influenza in the United States. *Pharmacoeconomics* **9:** 26-33.
Turner, J. & Kelly, H. (2005) A medical locum service as a site for sentinel influenza surveillance. *Eurosurveillance,* **10(4):** 96-98.
WHO. (2003). Influenza Factsheet. Available at: http://www.who.int/mediacentre/factsheets/2003/fs211/en/
WHO. (2006). SARS. Available at: http://www.who.int/csr/sars/archive/en/
WHO. (2007a). Cumulative Number of Confirmed Human Cases of Avian Influenza A/(H5N1) Reported to WHO. Available at: http://www.who.int/csr/disease/avian_influenza/country/en/
WHO. (2007b). Pandemic Prepardness. Available at: http://www.who.int/csr/disease/influenza/pandemic/en/
WHO. (2007c). Surveillance. Available at: http://www.who.int/tobacco/surveillance/en/
Yih, K. W., Teates, K. S., Abrams, A., Kleinman, K., Pinner, R., Harmon, R. & Platt, R. (2006) Nurse Call Data for Detection of Influenza-Like Illness. *Advances in Disease Surveillance*, **1:** 76

Appendix 1: Flu Season Week Numbers and Corresponding Dates

Table A1.1: 2005/2006 Flu Season Week Numbers and Corresponding Dates

Week number	Week starting (Monday)	Week ending (Sunday)	Naming Convention
40	03/10/2005	09/10/2005	40/05-06
41	10/10/2005	16/10/2005	41/05-06
42	17/10/2005	23/10/2005	42/05-06
43	24/10/2005	30/10/2005	43/05-06
44	31/10/2005	06/11/2005	44/05-06
45	07/11/2005	13/11/2005	45/05-06
46	14/11/2005	20/11/2005	46/05-06
47	21/11/2005	27/11/2005	47/05-06
48	28/11/2005	04/12/2005	48/05-06
49	05/12/2005	11/12/2005	49/05-06
50	12/12/2005	18/12/2005	50/05-06
51	19/12/2005	25/12/2005	51/05-06
52	26/12/2005	01/01/2006	52/05-06
1	02/01/2006	08/01/2006	01/05-06
2	09/01/2006	15/01/2006	02/05-06
3	16/01/2006	22/01/2006	03/05-06
4	23/01/2006	29/01/2006	04/05-06
5	30/01/2006	05/02/2006	05/05-06
6	06/02/2006	12/02/2006	06/05-06
7	13/02/2006	19/02/2006	07/05-06
8	20/02/2006	26/02/2006	08/05-06
9	27/02/2006	05/03/2006	09/05-06
10	06/03/2006	12/03/2006	10/05-06
11	13/03/2006	19/03/2006	11/05-06
12	20/03/2006	26/03/2006	12/05-06
13	27/03/2006	02/04/2006	13/05-06
14	03/04/2006	09/04/2006	14/05-06
15	10/04/2006	16/04/2006	15/05-06
16	17/04/2006	23/04/2006	16/05-06
17	24/04/2006	30/04/2006	17/05-06
18	01/05/2006	07/05/2006	18/05-06
19	08/05/2006	14/05/2006	19/05-06
20	15/05/2006	21/05/2006	20/05-06

Table A1.2: 2004/2005 Flu Season Week Numbers and Corresponding Dates

Week number	Week starting (Monday)	Week ending (Sunday)	Naming Convention
40	27/09/2004	03/10/2004	40/04-05
41	04/10/2004	10/10/2004	41/04-05
42	11/10/2004	17/10/2004	42/04-05
43	18/10/2004	24/10/2004	43/04-05
44	25/10/2004	31/10/2004	44/04-05
45	01/11/2004	07/11/2004	45/04-05
46	08/11/2004	14/11/2004	46/04-05
47	15/11/2004	21/11/2004	47/04-05
48	22/11/2004	28/11/2004	48/04-05
49	29/11/2004	05/12/2004	49/04-05
50	06/12/2004	12/12/2004	50/04-05
51	13/12/2004	19/12/2004	51/04-05
52	20/12/2004	26/12/2004	52/04-05
53	27/11/2004	02/01/2005	53/04-05
1	03/01/2005	09/01/2005	01/04-05
2	10/01/2005	16/01/2005	02/04-05
3	17/01/2005	23/01/2005	03/04-05
4	24/01/2005	30/01/2005	04/04-05
5	31/01/2005	06/02/2005	05/04-05
6	07/02/2005	13/02/2005	06/04-05
7	14/02/2005	20/02/2005	07/04-05
8	21/02/2005	27/02/2005	08/04-05
9	28/02/2005	06/03/2005	09/04-05
10	07/03/2005	13/03/2005	10/04-05
11	14/03/2005	20/03/2005	11/04-05
12	21/03/2005	27/03/2005	12/04-05
13	28/03/2005	03/04/2005	13/04-05
14	04/04/2005	10/04/2005	14/04-05
15	11/04/2005	17/04/2005	15/04-05
16	18/04/2005	24/04/2005	16/04-05
17	25/04/2005	01/05/2005	17/04-05
18	02/05/2005	08/05/2005	18/04-05
19	09/05/2005	15/05/2005	19/04-05
20	16/05/2005	22/05/2005	20/04-05

Table A1.3: 2003/2004 Flu Season Week Numbers and Corresponding Dates

Week number	Week starting (Monday)	Week ending (Sunday)	Naming Convention
40	29/09/2003	05/10/2003	40/03-04
41	06/10/2003	12/10/2003	41/03-04
42	13/10/2003	19/10/2003	42/03-04
43	20/10/2003	26/10/2003	43/03-04
44	27/10/2003	02/11/2003	44/03-04
45	03/11/2003	09/11/2003	45/03-04
46	10/11/2003	16/11/2003	46/03-04
47	17/11/2003	23/11/2003	47/03-04
48	24/11/2003	30/11/2003	48/03-04
49	01/12/2003	07/12/2003	49/03-04
50	08/12/2003	14/12/2003	50/03-04
51	15/12/2003	21/12/2003	51/03-04
52	22/12/2003	28/12/2003	52/03-04
1	29/12/2003	04/01/2004	01/03-04
2	05/01/2004	11/01/2004	02/03-04
3	12/01/2004	18/01/2004	03/03-04
4	19/01/2004	25/01/2004	04/03-04
5	26/01/2004	01/02/2004	05/03-04
6	02/02/2004	08/02/2004	06/03-04
7	09/02/2004	15/02/2004	07/03-04
8	16/02/2004	22/02/2004	08/03-04
9	23/02/2004	29/02/2004	09/03-04
10	01/03/2004	07/03/2004	10/03-04
11	08/03/2004	14/03/2004	11/03-04
12	15/03/2004	21/03/2004	12/03-04
13	22/03/2004	28/03/2004	13/03-04
14	29/03/2004	04/04/2004	14/03-04
15	05/04/2004	11/04/2004	15/03-04
16	12/04/2004	18/04/2004	16/03-04
17	19/04/2004	25/04/2004	17/03-04
18	26/04/2004	02/05/2004	18/03-04
19	03/05/2004	09/05/2004	19/03-04
20	10/05/2004	16/05/2004	20/03-04

Appendix 2: Data on National Influenza Indicators

Table A2.1: Summary Data on National Influenza Indicators

Week No 2003/2004	ILI rate per 100,000 population#	Sentinel GP Positive Swabs#	Non-Sentinel Positive Swabs#	All Positive Swabs#	Influenza Notifications*	Influenza Related Deaths^
40	14.1	0	0	0	Not notifiable	0
41	31.8	4	0	4	Not notifiable	0
42	20.7	7	2	9	Not notifiable	0
43	26.6	5	2	7	Not notifiable	0
44	58.4	23	6	29	Not notifiable	0
45	72.7	24	12	36	Not notifiable	0
46	82.3	19	9	28	Not notifiable	0
47	62.9	20	19	39	Not notifiable	1
48	58.3	17	15	32	Not notifiable	1
49	34	6	14	20	Not notifiable	0
50	33.1	5	9	14	Not notifiable	0
51	24.9	7	8	15	Not notifiable	0
52	16.4	2	0	2	Not notifiable	0
1	29.4	2	8	10	12	0
2	42.7	3	2	5	7	0
3	16	1	2	3	12	0
4	9.1	0	0	0	3	0
5	14.1	2	0	2	2	0
6	7.3	1	0	1	3	0
7	10.1	1	0	1	0	0
8	3.3	0	0	0	3	0
9	10.2	0	0	0	2	0
10	6	0	0	0	0	0
11	5.5	0	0	0	1	0
12	5.5	0	0	0	1	0
13	1.1	0	0	0	1	0
14	3.5	0	0	0	0	0
15	1.2	0	2	2	1	0
16	1.1	0	0	0	0	0
17	1.2	0	0	0	1	0
18	0	0	1	1	0	0
19	2.9	0	0	0	0	0
20	0	0	0	0	4	0

Note: % Respiratory Admission/Total Admissions was not available for the 2003/2004 season
Obtained on request from the HPSC
* Extracted from the Computerised Infectious Disease Reporting System on 1st June 2007
^ Obtained from HPSC Summary Influenza Reports 2003/2004, 2004/2005, 2005/2006

Table A2.2: Summary Data on National Influenza Indicators

Week No 2004/2005	ILI rate per 100,000 population#	Sentinel GP Positive Swabs#	Non-Sentinel Positive Swabs#	All Positive Swabs#	Influenza Notifications *	Influenza Related Deaths^	% Respiratory Admissions$
40	8.6	1	0	1	0	0	2.2
41	12.8	0	0	0	0	0	2.1
42	9.9	0	0	0	0	0	2.6
43	8.9	0	2	2	1	0	2.6
44	10.5	0	1	1	0	0	2.9
45	14.4	3	1	4	1	0	2.4
46	10.7	3	0	3	0	0	2.3
47	9.2	2	1	3	1	0	3.3
48	14.7	3	2	5	1	0	2.8
49	15.5	4	0	4	6	0	2.3
50	18.4	5	1	6	9	0	3.2
51	35.4	13	1	14	2	0	3.1
52	53.8	20	4	24	1	0	3.0
53	49.8	11	0	11	0	0	4.8
1	89	15	5	20	6	2	5.8
2	31.3	5	7	12	12	0	4.5
3	28.5	8	16	24	11	0	4.5
4	15.8	3	4	7	14	0	4.4
5	9.7	5	1	6	66	0	4.1
6	12.3	2	0	2	12	0	3.9
7	7	1	0	1	30	0	3.5
8	14.2	3	0	3	8	0	4.8
9	13.7	4	6	10	5	0	4.0
10	18.9	6	0	6	22	0	3.5
11	16	6	0	6	4	0	4.4
12	13.1	6	0	6	17	0	3.8
13	11.2	4	0	4	4	0	3.9
14	9	3	1	4	3	0	2.9
15	1.9	1	0	1	5	0	3.0
16	4.9	2	0	2	8	0	3.1
17	3.3	1	6	7	34	0	4.2
18	3.9	2	0	2	1	0	3.9
19	5.8	0	1	1	1	0	3.4
20	5	0	0	0	10	0	3.9

Obtained on request from the HPSC
* Extracted from the Computerised Infectious Disease Reporting System on 1st June 2007
^ Obtained from HPSC Summary Influenza Reports 2003/2004, 2004/2005, 2005/2006
$ Obtained on request from the HPSC. Comprises data from seven sentinel Hospitals

Table A2.3: Summary Data on National Influenza Indicators

Week No 2004/2005	ILI rate per 100,000 population#	Sentinel GP Positive Swabs#	Non-Sentinel Positive Swabs#	All Positive Swabs#	Influenza Notifications *	Influenza Related Deaths^	% Respiratory Admissions$
40	4	0	0	0	0	0	2.9
41	11.5	0	0	0	0	0	3.4
42	6.7	0	0	0	0	0	3.3
43	5.6	0	0	0	0	0	2.3
44	10.5	0	0	0	0	0	2.6
45	9.9	0	0	0	0	0	2.4
46	8.1	0	0	0	3	0	2.7
47	8.3	0	0	0	2	0	4.3
48	12.9	0	0	0	5	0	3.0
49	11.3	0	0	0	6	0	2.3
50	16.8	0	0	0	9	0	3.2
51	16.3	0	0	0	5	0	2.8
52	8.2	0	1	1	5	0	5.2
1	9.7	0	0	0	1	0	4.5
2	10.8	1	1	2	13	0	3.5
3	17.7	0	1	1	4	0	3.1
4	16.5	5	3	8	12	0	3.4
5	16.5	6	2	8	10	0	4.0
6	29.6	5	0	5	4	0	3.1
7	50.2	13	3	16	11	0	3.1
8	41.7	12	4	16	11	0	3.0
9	50.5	19	5	24	22	0	2.9
10	82.5	19	10	29	30	0	3.9
11	42.3	17	0	17	25	0	3.7
12	47.4	14	2	16	20	0	4.6
13	41.2	9	1	10	31	0	3.6
14	24.7	7	2	9	28	0	2.5
15	7.2	3	1	4	6	1	3.2
16	7.3	1	0	1	10	0	3.2
17	9.2	1	0	1	5	0	2.8
18	2.7	0	0	0	3	0	3.1
19	1.3	0	0	0	3	0	2.5
20	1.9	0	0	0	0	0	3.0

Obtained on request from the HPSC
* Extracted from the Computerised Infectious Disease Reporting System on 1st June 2007
^ Obtained from HPSC Summary Influenza Reports 2003/2004, 2004/2005, 2005/2006
$ Obtained on request from the HPSC. Comprises data from seven sentinel Hospitals

Appendix 3: Data Extracted from NE-Doc Flu related records

Table. A3.1: Raw Data on Number of Calls Extracted by Query from Patient's Reported Condition (PRC) Field for each Flu Season

	No. of Calls Extracted by each query for each Flu Season														
	2003/2004					2004/2005					2005/2006				
Week No	Q1	Q2	Q3	Q4	Total Calls	Q1	Q2	Q3	Q4	Total Calls	Q1	Q2	Q3	Q4	Total Calls
40	69	53	6	122	1634	36	26	2	62	1865	55	45	4	102	2188
41	62	74	6	133	1818	35	49	4	84	1886	43	42	7	87	2007
42	77	82	14	158	1771	39	48	5	86	2043	42	46	6	89	2118
43	130	96	11	221	2113	46	49	4	94	2048	49	59	2	108	2303
44	210	136	13	345	2675	63	36	1	99	2489	72	69	8	140	2742
45	155	108	6	259	2201	27	40	3	66	2041	34	42	6	79	1916
46	110	85	5	191	2062	40	49	6	89	2095	35	46	0	81	2063
47	120	73	10	191	2175	40	48	5	88	2045	44	61	5	105	2089
48	100	61	8	159	1965	41	69	3	109	2250	51	59	6	111	2335
49	80	67	3	143	2111	66	90	10	153	2274	80	51	2	132	2437
50	81	75	2	153	2251	80	126	14	204	2588	54	57	6	112	2320
51	86	90	9	174	2329	106	169	18	274	2919	45	57	6	102	2352
52	183	137	19	316	3871	194	149	18	339	2814	138	126	11	259	4454
53	NA	NA	NA	NA	NA	631	246	33	866	5709	NA	NA	NA	NA	NA
1	158	86	7	242	2950	385	87	10	468	3504	115	73	3	188	3186
2	78	43	2	121	2001	100	60	11	161	2081	49	52	6	101	2272
3	53	36	6	91	1794	57	49	7	103	1897	29	65	10	91	2308
4	41	43	4	84	1910	56	43	4	98	2000	43	70	10	114	2247
5	61	50	10	110	1928	60	63	9	123	2148	92	118	24	208	2617
6	44	51	7	95	2031	66	42	5	108	2158	145	154	34	300	2953
7	35	37	5	72	1968	54	28	6	83	2177	138	140	28	283	2947
8	53	42	6	95	2029	60	34	5	92	2158	113	125	18	233	2731
9	47	47	4	93	1981	82	55	7	136	2334	111	140	24	250	2798
10	38	55	4	93	2071	82	79	8	161	2445	117	154	22	268	2934
11	33	47	4	81	2090	86	81	12	170	2979	134	152	27	285	3555
12	39	57	8	96	2343	63	74	8	138	2596	104	116	17	220	2691
13	35	49	2	84	2118	75	101	10	176	2880	93	83	7	175	2604
14	25	52	8	76	2070	60	51	3	111	2247	53	59	1	111	2395
15	35	57	4	93	2166	42	56	5	98	2231	57	72	3	129	2580
16	46	55	2	100	2592	49	65	5	113	2417	64	64	8	128	2829
17	14	26	0	40	1953	60	85	5	144	2448	33	57	4	90	2498
18	20	42	2	62	2116	62	93	8	155	3009	30	58	7	88	2861
19	25	57	4	82	2472	31	70	3	100	2249	20	44	6	62	2219
20	33	59	11	92	2144	30	54	6	84	2245	26	63	6	90	2215
Total	2376	2128	212	4467	71703	3004	2464	263	5435	83269	2308	2619	334	4921	84764

Table. A3.2: Raw Data on Number of Calls Extracted from Doctors Diagnosis (DD) field with keywords of Flu or Influenza (Query 1) for each Flu Season

	2003/2004		2004/2005		2005/2006	
Week No	DD Calls	Total Calls	DD Calls	Total Calls	DD Calls	Total Calls
40	21	1634	5	1865	12	2188
41	17	1818	5	1886	9	2007
42	16	1771	7	2043	15	2118
43	31	2113	6	2048	10	2303
44	51	2675	16	2489	8	2742
45	34	2201	9	2041	7	1916
46	35	2062	10	2095	6	2063
47	25	2175	8	2045	1	2089
48	24	1965	6	2250	13	2335
49	30	2111	11	2274	18	2437
50	18	2251	16	2588	10	2320
51	14	2329	19	2919	9	2352
52	28	3871	41	2814	31	4454
53	NA	NA	80	5709	NA	NA
1	28	2950	70	3504	20	3186
2	22	2001	30	2081	11	2272
3	12	1794	15	1897	6	2308
4	13	1910	11	2000	10	2247
5	5	1928	12	2148	16	2617
6	7	2031	15	2158	37	2953
7	8	1968	17	2177	29	2947
8	11	2029	13	2158	20	2731
9	5	1981	20	2334	32	2798
10	8	2071	22	2445	26	2934
11	3	2090	16	2979	18	3555
12	6	2343	10	2596	24	2691
13	4	2118	12	2880	15	2604
14	8	2070	14	2247	15	2395
15	6	2166	9	2231	12	2580
16	7	2592	7	2417	6	2829
17	2	1953	12	2448	9	2498
18	1	2116	11	3009	5	2861
19	4	2472	7	2249	9	2219
20	6	2144	12	2245	5	2215

Appendix 4:
Summary data for comparison of Query 2 with National Flu Indicators.

A4.1 Query 2: CDC Flu Definition Query 2003/2004

Again two prominent peaks, spanning Weeks 40-47 and Weeks 50-01, in the % of NE-Doc CDC Definition Calls, as extracted by Query 2, were identified during the early 2003/2004 Flu Season. These calls represent all calls with the keywords high temperature or fever and cough or sore throat. The national ILI rate also peaked during Week 43-49 and there was a second smaller peak during Weeks 52-03 (Fig A4.1.a). The number of positive sentinel swabs contained one major peak spanning Weeks 43-49 which also coincides with the first peak in the % of CDC Definition Calls (Fig. A4.1.b). The non-sentinel swabs displayed one large peak between Week 43-51 with a smaller peak during Week 01 (Fig. A4.1.c). In comparison of % CDC Definition Calls with all three of these national indicators, the highest point of the % CDC Definition Calls occurred at least one week before the highest point of the national indicator.

A4.2 Query 2: CDC Flu Definition Query 2004/2005

One prominent peak, spanning Weeks 48-02, in the % of NE-Doc CDC Definition Calls as extracted by Query 2, was identified during the 2004/2005 Flu Season. The national ILI rate peak spanned Weeks 50-04 (Fig A4.2.a). The number of positive sentinel swabs contained one major peak spanning Weeks 51-03 which also coincides with the first peak in the % of CDC Defintion Calls (Fig. A4.2.b). The non-sentinel swabs displayed one large peak between Weeks 01-04 with some subsequent smaller peaks late in the season (Fig. A4.2.b). There was one major peak in influenza notifications occurring during Weeks 04-06, almost 7 weeks behind the peak in calls (Fig. A4.2.c). The peak in % respiratory hospital admissions occurred 3 weeks after the peak in % CDC Definition Calls (Fig. A4.2.d). In comparison of % CDC Definition Calls with the national ILI rate, sentinel swabs, non-sentinel swabs, influenza notifications and % respiratory admissions, the highest point of the % CDC Definition Calls occurred at least one week before the highest point of each of these national indicators.

A4.3 Query 2: CDC Flu Definition Query 2005/2006

One prominent broad peak, spanning Weeks 03-14, in the % of NE-Doc CDC Definition Calls as extracted by Query 2 were identified during the early 2005/2006 Flu Season. The national ILI rate also had a broad peak between Week 05-15 (Fig A4.3.a) which reached its highest point the same week as the highest point in % CDC Definition calls. The number of positive sentinel swabs contained one major peak spanning Weeks 4-14 which also coincides with the peak in the % of CDC Definition Calls (Fig. A4.3.b). The non-sentinel swabs displayed one large peak between Week 06-11 (Fig. A4.3.b) the highest point of which occurred the same weeks as the highest point in the % of CDC Definition Calls. There were two peaks in influenza notifications occurring between Weeks 08-15 (Fig. A4.3.c). There was one major peak in the % Respiratory Hospital Admissions between Weeks 51-03 but this peak does not correspond with the prominent broad peak in % CDC Definition Calls (Fig. A4.3.d) and may therefore be due to respiratory infections other than influenza. In all cases, the highest point of the % CDC Definition Calls occurred the same week as the highest point in the national indicators

Fig. A4.1: CDC Definition Query Calls (Query 2) 2003/2004 Comparison with National Indicators

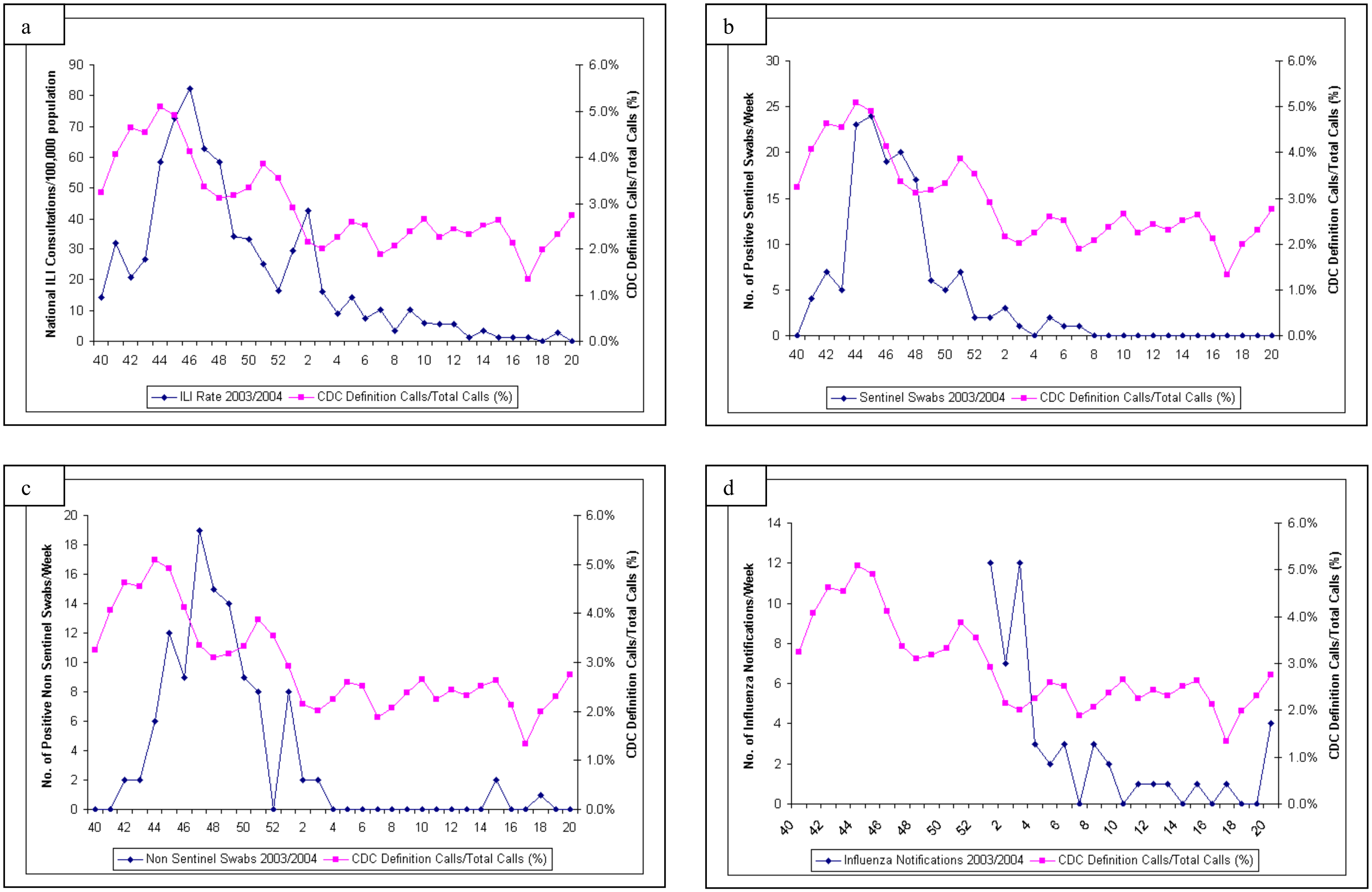


Fig. A4.2: CDC Definition Query Calls (Query 2) 2004/2005 Comparison with National Indicators

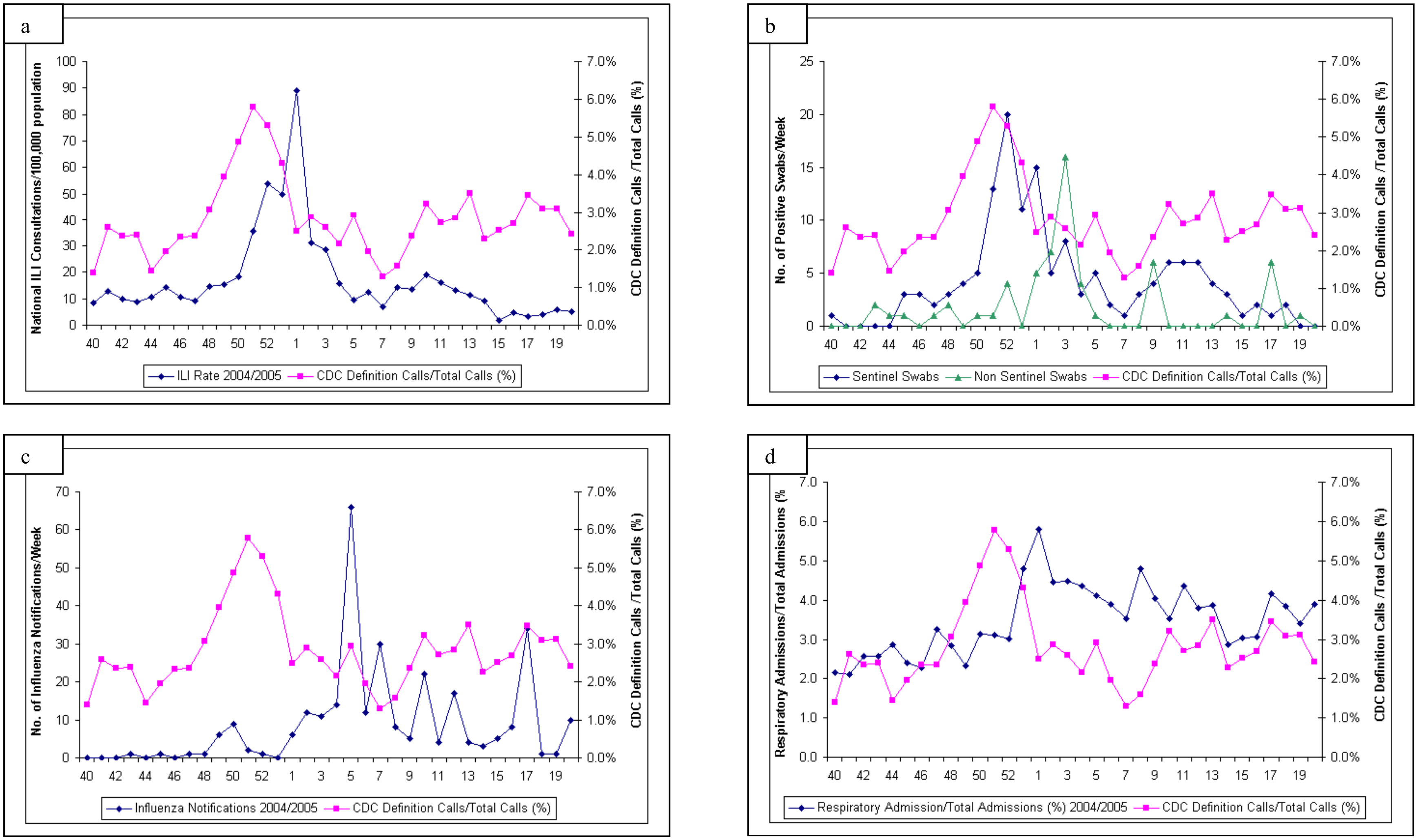
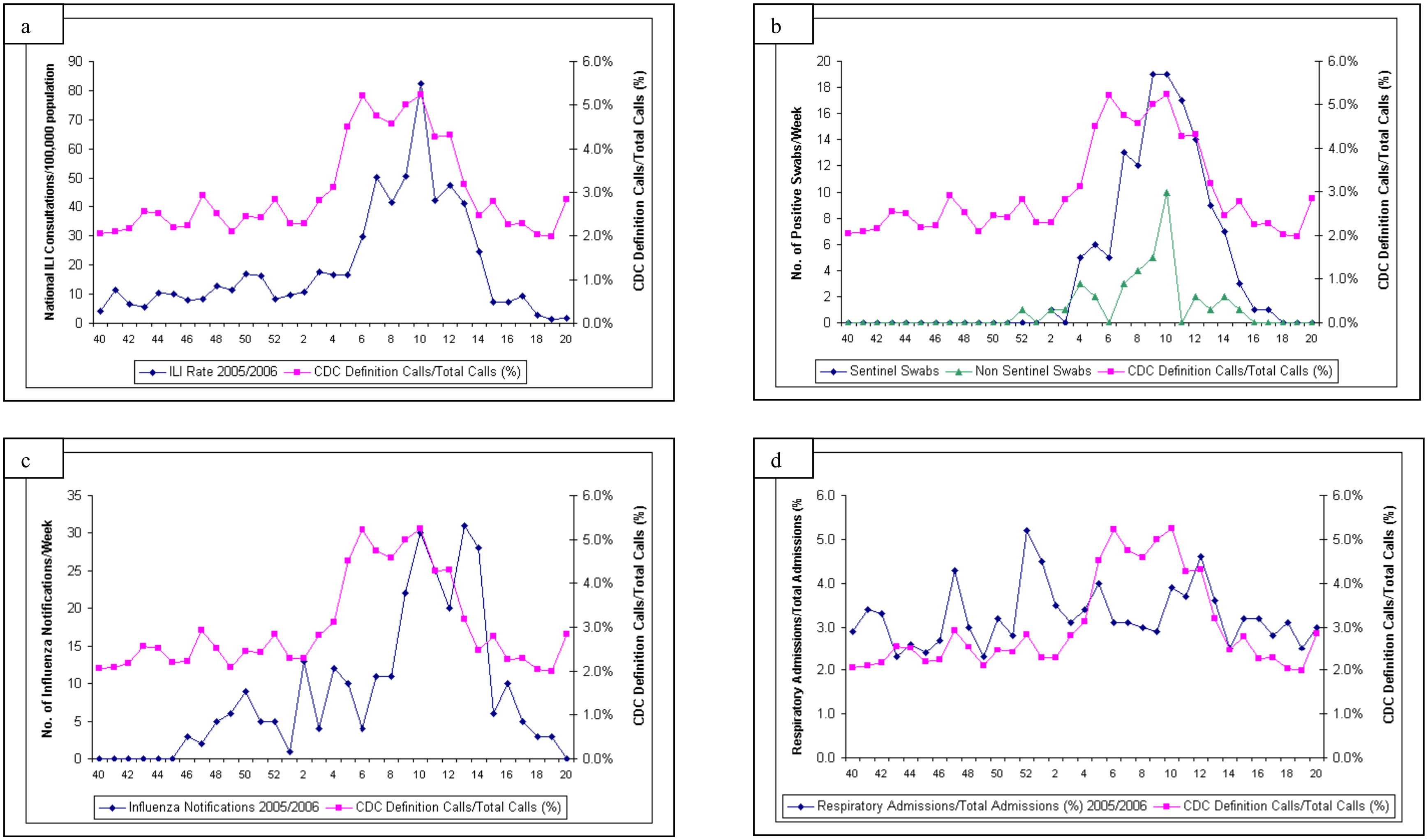


Fig. A4.3: CDC Definition Query Calls (Query 2) 2005/2006 Comparison with National Indicators



Appendix 5
Summary data for comparison of Query 3 with National Flu Indicators.

A5.1 Query 3: HPSC Flu Definition Query 2003/2004

One prominent peak, spanning Weeks 41-45, in the % of NE-Doc HPSC Definition Calls, as extracted by Query 3, was identified during the early 2003/2004 Flu Season. These calls represent all calls with the keywords high temperature or fever and two of the following: cough, sore throat, headache, aches and pains. The national ILI rate peaked during Week 43-49 and there was a second smaller peak during Weeks 52-03 (Fig A5.1.a). The number of positive sentinel swabs contained one major peak spanning Weeks 43-49 which also coincides with the first peak in the % of HPSC Definition Calls (Fig. A5.1.b). The non-sentinel swabs displayed one large peak between Week 43-51 with a smaller peak during Week 01 (Fig. A5.1.c). In comparison of % HPSC Definition Calls with all three of these national indicators, the highest point of the % HPSC Definition calls occurred at least three week before the highest point of the national indicator. However, as this query is quite restrictive and there is some variation in the data following the first major peak in HPSC Definition calls it is difficult to identify peaks in influenza activity as opposed to sporadic activity.

A5.2 Query 3: HPSC Flu Definition Query 2004/2005

Two prominent peaks, spanning Weeks 48-01 and Weeks 01-04, in the % of NE-Doc HPSC Definition Calls as extracted by Query 3, was identified during the 2004/2005 Flu Season. The national ILI rate peak spanned Weeks 50-04 (Fig A5.2.a) and the peak in calls occurred 2 weeks before the ILI rate peak. The number of positive sentinel swabs contained one major peak spanning Weeks 51-03 which also coincides with the first peak in the % of HPSC Definition Calls (Fig. A5.2.b). The non-sentinel swabs displayed one large peak between Weeks 01-04 with some subsequent smaller peaks late in the season (Fig. A5.2.b). There was one major peak in influenza notifications occurring during Weeks 04-06, almost 6 weeks behind the peak in calls (Fig. A5.2.c). The peak in % respiratory hospital admissions occurred 2 weeks after the peak in % HPSC Definition Calls (Fig. A5.2.d). In comparison of % HPSC Definition Calls with the national ILI rate, non-sentinel swabs, influenza notifications and % respiratory admissions, the highest point of the % HPSC Definition Calls occurred at least two week before the highest point of each of these national indicators. The peak in % HPSC Definition calls occurred the same week as the peak in Sentinel Swabs.

A5.3 Query 3: HPSC Flu Definition Query 2005/2006

One prominent broad peak, spanning Weeks 02-13, in the % of NE-Doc HPSC Definition Calls as extracted by Query 3 were identified during the early 2005/2006 Flu Season. The national ILI rate also had a broad peak between Week 05-15 (Fig A5.3.a) which reached its highest point the 4 weeks after the highest point in % HPSC Definition calls. The number of positive sentinel swabs contained one major peak spanning Weeks 4-14 which its highest point 3 weeks after the highest point in the % of HPSC Definition Calls (Fig. A5.3.b). The non-sentinel swabs displayed one large peak between Week 06-11 (Fig. A5.3.b) the highest point of which occurred 4 weeks after the highest point in the % of HPSC Definition Calls. There were two peaks in influenza notifications occurring between Weeks 08-15 (Fig. A5.3.c). There was one major peak in the % Respiratory Hospital Admissions between Weeks 51-03 but this peak does not correspond with the prominent broad peak in % HPSC Definition Calls (Fig. A5.3.d). In all cases, the highest point of the % HPSC Definition Calls occurred at least 3 weeks before the highest point in the national indicators (with the exception of the % Respiratory Admissions).

Fig. A5.1: HPSC Definition Query Calls (Query 3) 2003/2004 Comparison with National Indicators

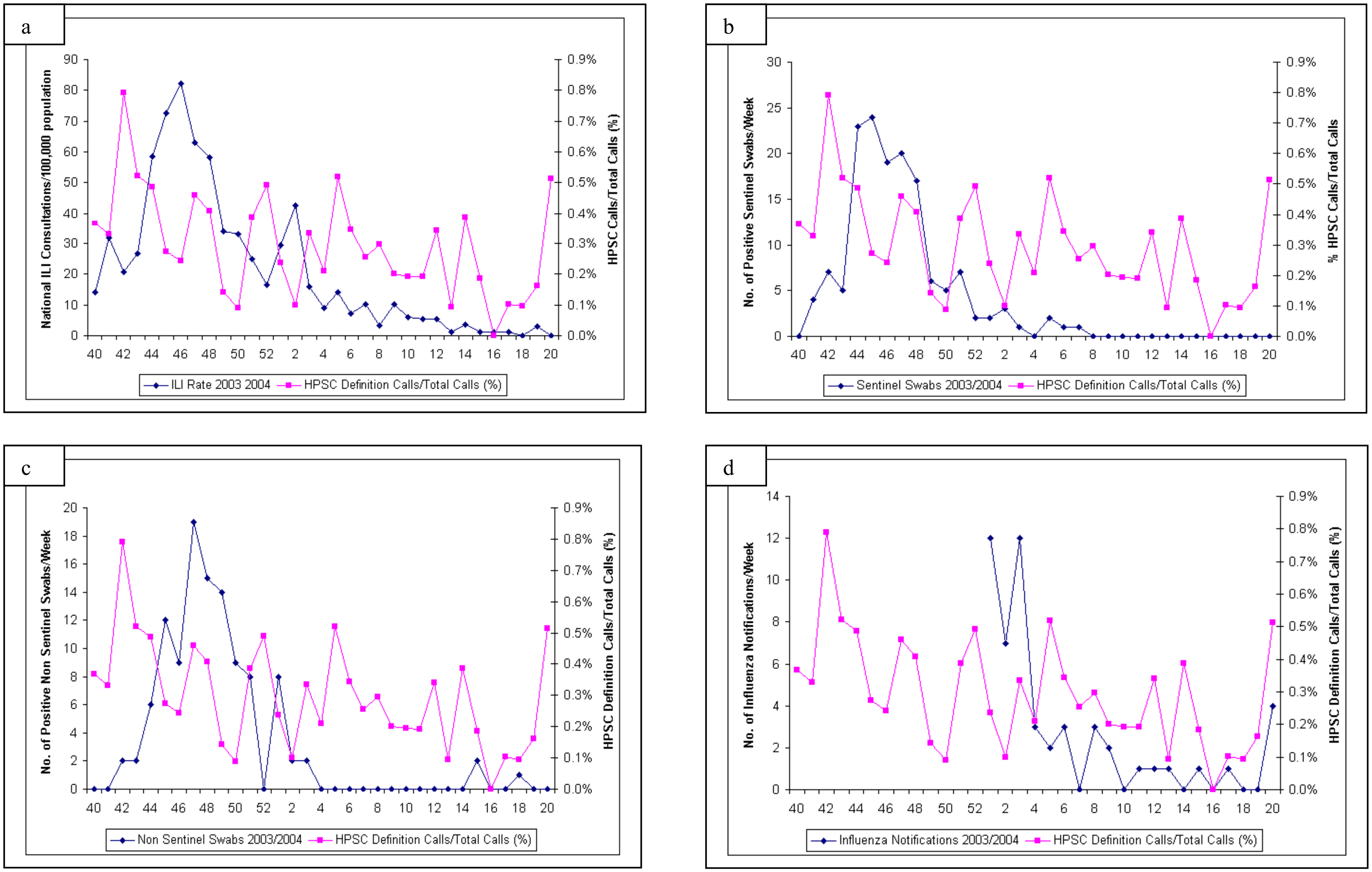


Fig. A5.2: HPSC Definition Query Calls (Query 3) 2004/2005 Comparison with National Indicators

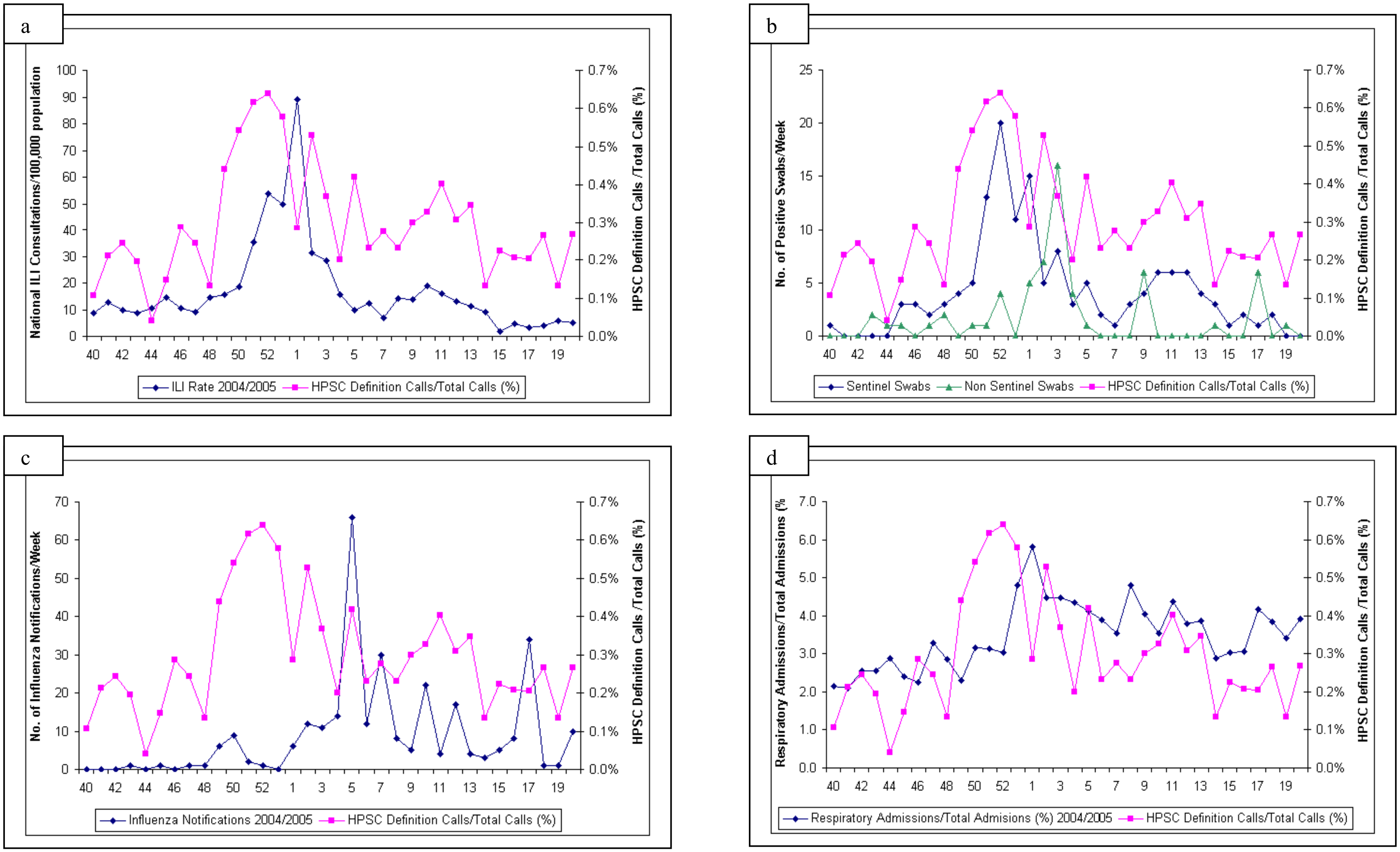
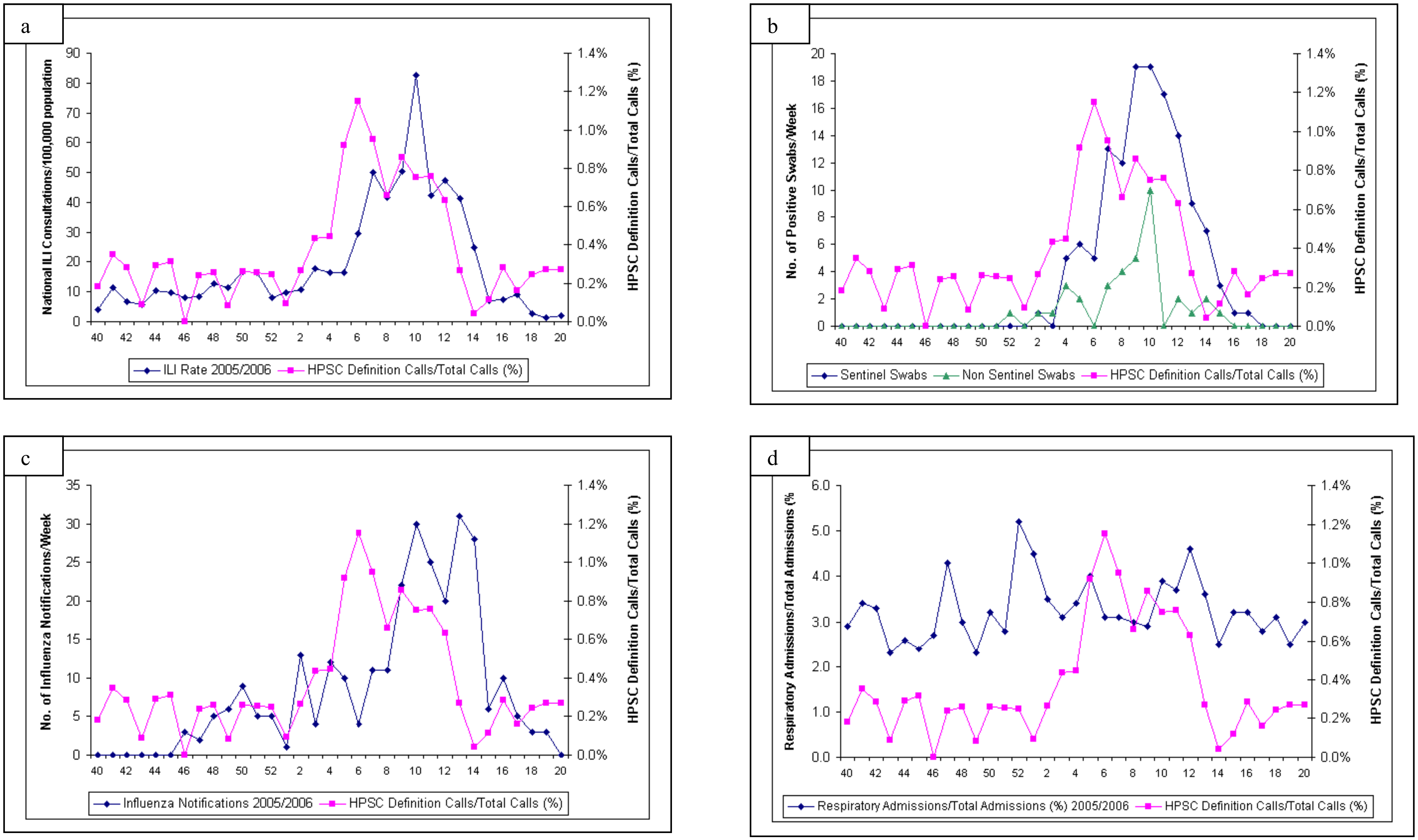


Fig. A5.3: HPSC Definition Query Calls (Query 3) 2005/2006 Comparison with National Indicators



Appendix 6
Summary data for comparison of Query 4 with National Flu Indicators.

A6.1 Query 4: Combined Flu Definition Query 2003/2004

Again two prominent peaks, spanning Weeks 41-49 and Weeks 50-02, in the % of NE-Doc Combined Definition Calls, as extracted by Query 4, were identified during the early 2003/2004 Flu Season. Calls gradually tailed off after these major peaks for the remainder of the season. These combined definition calls represent all flu related calls extracted using the Flu/Influenza keyword definition and the CDC Definition and the HPSC definition. The national ILI rate also peaked during Week 43-49 and there was a second smaller peak during Weeks 52-03 (Fig A6.1.a). The number of positive sentinel swabs contained one major peak spanning Weeks 43-49 which also coincides with the first peak in the % of Combined Definition Calls (Fig. A6.1.b). The non-sentinel swabs displayed one large peak between Week 43-51 with a smaller peak during Week 01 (Fig. A6.1.c). In comparison of % Combined Definition Calls with all three of these national indicators, the highest point of the % CDC Definition Calls occurred at least one week before the highest point of the national indicator.

A6.2 Query 4: Combined Flu Definition Query 2004/2005

One prominent peak, spanning Weeks 48-03, in the % of NE-Doc Combined Definition Calls as extracted by Query 4, was identified during the 2004/2005 Flu Season. The national ILI rate peak spanned Weeks 50-04 (Fig A6.2.a) and the peak in the ILI rate was one week behind the peak in calls. The number of positive sentinel swabs contained one major peak spanning Weeks 51-03 which also coincides with the first peak in the % of Combined Definition Calls (Fig. A6.2.b). The non-sentinel swabs displayed one large peak between Weeks 01-04 with some subsequent smaller peaks late in the season (Fig. A6.2.b). There was one major peak in influenza notifications occurring during Weeks 04-06, almost 5 weeks behind the peak in calls (Fig. A6.2.c). The peak in % respiratory hospital admissions occurred 1 week behind the peak in % Combined Definition Calls (Fig. A6.2.d). In comparison of % Combined Definition Calls with the national ILI rate, non-sentinel swabs, influenza notifications and % respiratory admissions, the highest point of the % Combined Definition Calls occurred at least one week before the highest point of each of these national indicators. Interestingly, the peak in % Combined Definition calls occurred the same week as the peak in Sentinel Swabs.

A6.3 Query 4: Combined Flu Definition Query 2005/2006

One prominent broad peak, spanning Weeks 04-14, in the % of NE-Doc Combined Definition Calls as extracted by Query 4 were identified during the early 2005/2006 Flu Season. The national ILI rate also had a broad peak between Week 05-15 (Fig A6.3.a) which reached its highest point the 4 weeks after the highest point in % Combined Definition calls. The number of positive sentinel swabs contained one major peak spanning Weeks 4-14 which its highest point 3 weeks after the highest point in the % of Combined Definition Calls (Fig. A6.3.b). The non-sentinel swabs displayed one large peak between Week 06-11 (Fig. A6.3.b) the highest point of which occurred 4 weeks after the highest point in the % of Combined Definition Calls. There were two peaks in influenza notifications occurring between Weeks 08-15 (Fig. A6.3.c). There was one major peak in the % Respiratory Hospital Admissions between Weeks 51-03 but this peak does not correspond with the prominent broad peak in % Combined Definition Calls (Fig. A6.3.d). In all cases, the highest point of the % HPSC Definition Calls occurred at least 3 weeks before the highest point in the national indicators (with the exception of the % Respiratory Admissions).

Fig. A6.1: Combined Definition Query Calls (Query 4) 2003/2004 Comparison with National Indicators

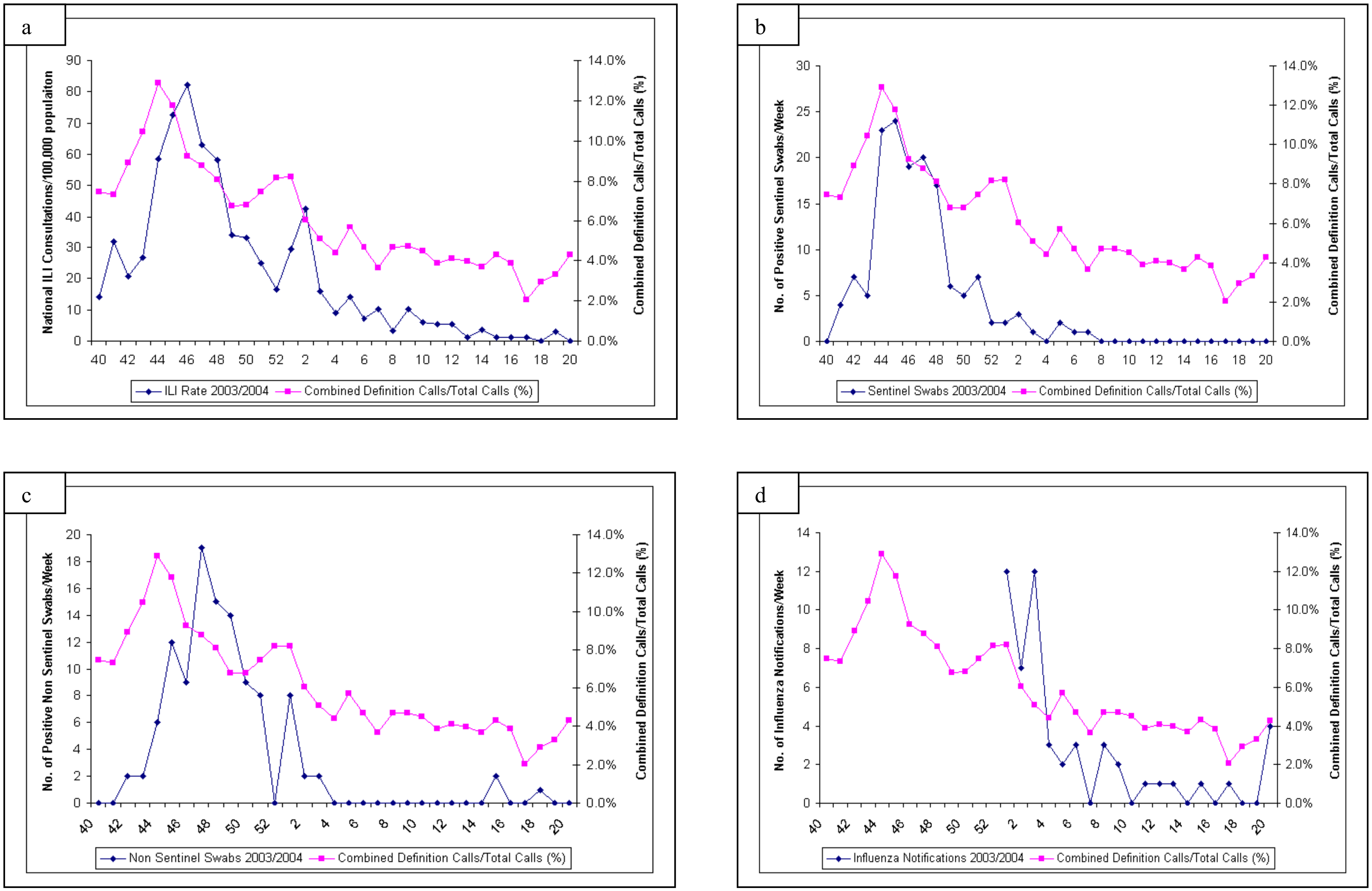


Fig. A6.2: Combined Definition Query Calls (Query 4) 2004/2005 Comparison with National Indicators

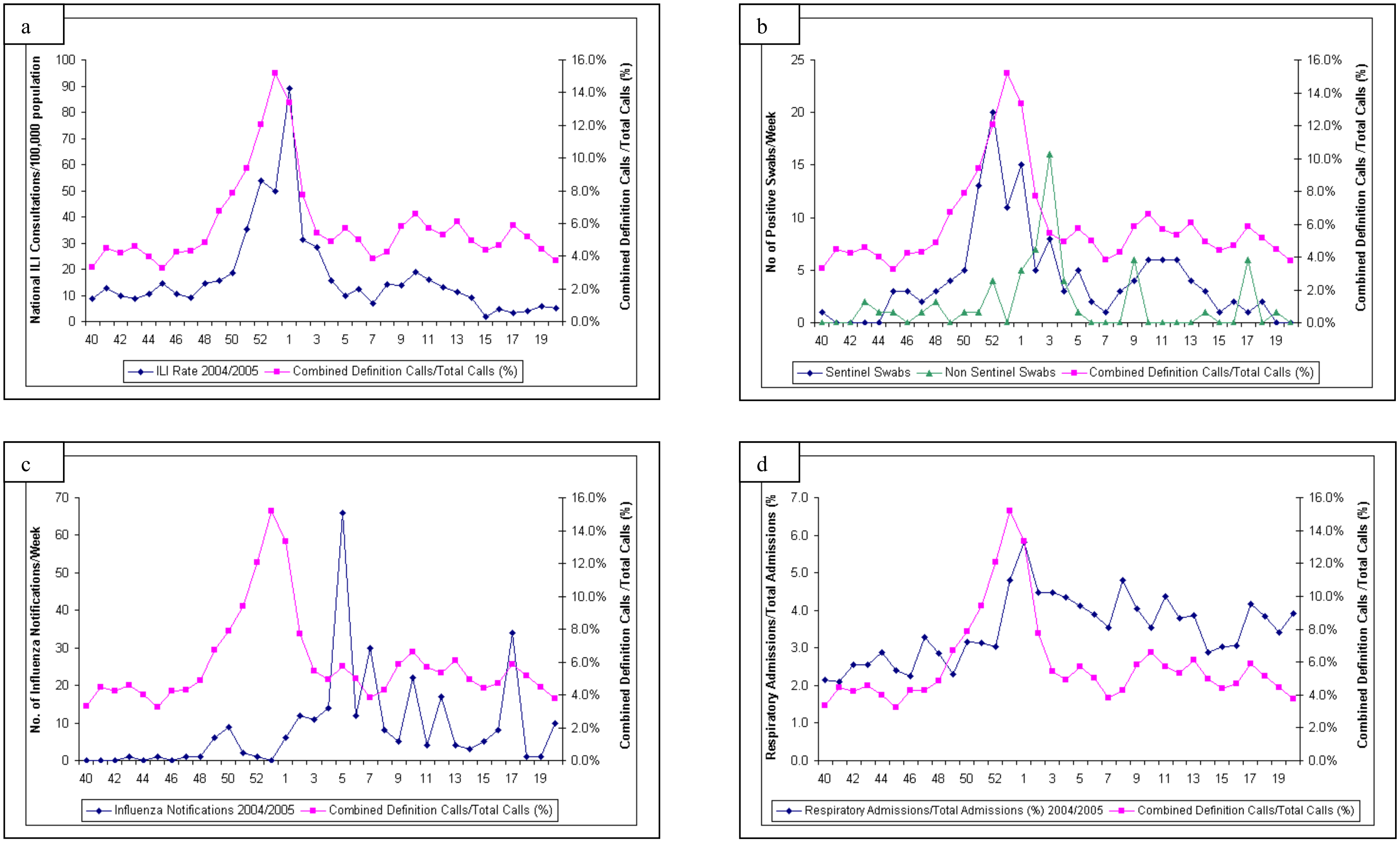


Fig. A6.3: Combined Definition Query Calls (Query 4) 2005/2006 Comparison with National Indicators