Creating a syndromic surveillance system

Submission for a Masters of Philosophy (Applied Epidemiology)

November 2017
MICA HARTLEY

Supervisors:
Dr. Kathryn Glass
Australian National University
Dr. Janice Biggs
Mr. Carlo Leonessa
Healthdirect Australia
This thesis is an account of research undertaken between February 2015 and November 2017 at Healthdirect Australia, to fulfil the requirements for a Masters of Philosophy (Applied Epidemiology) at the Australian National University.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Mica Hartley
November, 2017
There are many people who contributed to this thesis, and without their support I would not have been able to complete the task.

My biggest thanks must go to my supervisors, Kathryn Glass, Janice Biggs and Carlo Leonessa who were always available for questions. Since she sat next to me, Janice was always my first point of contact, and always considered my questions carefully - even though I asked at all kinds of inconvenient times. It was so valuable knowing she was always just around the corner. I also have to give huge thanks to Janice for continuing to supervise me after leaving the company! Katie was a bit further away, but her responses to all my questions were so quick she may as well have been just around the corner. Katie struck a great balance between keeping me on task and giving me enough flexibility to follow what I wanted to do, and I hugely appreciated her assistance in keeping me on task and making me manage my time wisely. Her lightning-fast comments on my chapters were also hugely appreciated. Finally, thanks to Carlo who was always able to provide the bigger picture and ask the big questions.

I was also part of a great team at Healthdirect Australia, and the Western Sydney Public Health units. Thanks go to Gail Hart, Michael Araco, Janelle Painter, Andrew Cole, Kay Law and Maureen Robinson at Healthdirect Australia. In the Western Sydney public health unit, Shopna Bag, Sophie Norton and Penelope Clark really helped me feel welcome and provided excellent training.

There are also a number of people outside of the work context who assisted me in this endeavour. Thanks to my MAE cohort, particularly to Meru Sheel and Katherine Todd in “Team Sydney”. Thanks to all my house mates over the two years - it was amazing being able to debrief at the end of a stressful day with such sympathetic ears. Thanks to my XPD team-mates and training partners, the Urinetown cast and crew, rogaining partners, the Rogue Ultimate Frisbee team, the Sydney Frisbee bushwalkers, the Sydney dancing crew and everyone else who helped me wind down with a huge range of de-stress activities. You guys all helped keep me sane. Special thanks to Rachel Poldy who went on a ton of adventures with me, but also helped proofread this thesis.

Finally, thanks to my partner, Mikey, who supported me every step of the way. Thanks for believing in me, for being flexible with your own career to support mine, for driving up to Sydney every other weekend, for moving with me to the Solomon Islands, for cooking me dinners, for listening, supporting and being in my camp always.

Here’s to the next adventure.
# Contents

## Acknowledgements

1 **Introduction**

1.0.1 Chapter 2: Surveillance ........................................... 2
1.0.2 Chapter 3: Data Analysis ......................................... 2
1.0.3 Chapter 4: Epidemiology Project ................................ 2
1.0.4 Chapter 5: Outbreak Investigation ............................ 2

1.1 Requirements ....................................................... 3
1.1.1 Field Projects .................................................. 3
1.1.2 Additional Requirements ....................................... 4
1.1.3 Teaching ......................................................... 5
1.1.4 Coursework ...................................................... 5

2 **Surveillance: Building a syndromic surveillance system using telephone and internet data**

2.1 Introduction ....................................................... 7
2.1.1 Objectives ...................................................... 9
2.2 Data Sources ...................................................... 10
2.3 Methods ........................................................... 12
2.3.1 Case Definition ............................................... 12
2.3.2 Indicators ....................................................... 15
2.3.3 Attributes of the surveillance system ......................... 16
2.4 Results ........................................................... 16
2.4.1 Indicators ....................................................... 16
2.4.2 Descriptive Epidemiology ..................................... 16
2.4.3 Consideration of Surveillance System Attributes .......... 20
2.5 Discussion ......................................................... 24
2.5.1 Indicators ....................................................... 25
2.5.2 Consideration of Surveillance System Attributes .......... 29
2.6 Implementation ..................................................... 29
2.7 Limitations ........................................................ 30
2.8 Public Health Implications ....................................... 30
2.9 Lessons Learnt ..................................................... 31

3 **Data Analysis: Investigating influenza-like illness using telephone and internet data** .................................................. 33

3.1 Introduction ....................................................... 33
3.2 Data Sources ...................................................... 35
3.3 Methods ........................................................... 36
3.3.1 Signal Identification ........................................... 36
3.3.2 Case Definition ............................................... 39
### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.3 Comparison of Data Sources</td>
<td>40</td>
</tr>
<tr>
<td>3.4 Results</td>
<td>42</td>
</tr>
<tr>
<td>3.4.1 Signal Analysis</td>
<td>42</td>
</tr>
<tr>
<td>3.4.2 Case Definition</td>
<td>42</td>
</tr>
<tr>
<td>3.4.3 Comparison of Data Sources</td>
<td>43</td>
</tr>
<tr>
<td>3.5 Discussion</td>
<td>46</td>
</tr>
<tr>
<td>3.5.1 Signal Analysis</td>
<td>46</td>
</tr>
<tr>
<td>3.5.2 Comparison of Data Sources</td>
<td>48</td>
</tr>
<tr>
<td>3.6 Limitations</td>
<td>51</td>
</tr>
<tr>
<td>3.7 Public Health Implications</td>
<td>52</td>
</tr>
<tr>
<td>3.8 Lessons Learnt</td>
<td>52</td>
</tr>
<tr>
<td>4 Epidemiology Chapter: Investigating the utility of conducting surveillance on a telephone helpline for six syndromes</td>
<td>55</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>55</td>
</tr>
<tr>
<td>4.2 My Role</td>
<td>57</td>
</tr>
<tr>
<td>4.3 Data Sources</td>
<td>57</td>
</tr>
<tr>
<td>4.4 Methods</td>
<td>58</td>
</tr>
<tr>
<td>4.4.1 Case Definitions</td>
<td>58</td>
</tr>
<tr>
<td>4.4.2 Spatial Investigation</td>
<td>60</td>
</tr>
<tr>
<td>4.5 Results</td>
<td>61</td>
</tr>
<tr>
<td>4.5.1 Gastroenteritis</td>
<td>61</td>
</tr>
<tr>
<td>4.5.2 Adverse Event Following Immunisation (AEFI)</td>
<td>65</td>
</tr>
<tr>
<td>4.5.3 Suicide Ideation and Acute Mental Health Events</td>
<td>66</td>
</tr>
<tr>
<td>4.5.4 Alcohol and Substance Abuse Events</td>
<td>66</td>
</tr>
<tr>
<td>4.6 Discussion</td>
<td>68</td>
</tr>
<tr>
<td>4.6.1 Gastroenteritis</td>
<td>68</td>
</tr>
<tr>
<td>4.6.2 Adverse Reactions to Vaccines</td>
<td>73</td>
</tr>
<tr>
<td>4.6.3 Suicide Ideation and Acute Mental Health Events</td>
<td>75</td>
</tr>
<tr>
<td>4.6.4 Alcohol and Substance Abuse Events</td>
<td>75</td>
</tr>
<tr>
<td>4.7 Public Health Implications</td>
<td>76</td>
</tr>
<tr>
<td>4.8 Lessons Learnt</td>
<td>77</td>
</tr>
<tr>
<td>5 Outbreak Investigation: A measles outbreak in western Sydney, Australia from March 2017 - May 2017</td>
<td>79</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>79</td>
</tr>
<tr>
<td>5.2 Methods</td>
<td>80</td>
</tr>
<tr>
<td>5.3 Results</td>
<td>81</td>
</tr>
<tr>
<td>5.4 Discussion</td>
<td>84</td>
</tr>
<tr>
<td>5.5 Public Health Implications</td>
<td>87</td>
</tr>
<tr>
<td>5.6 My Role</td>
<td>88</td>
</tr>
<tr>
<td>5.7 Lessons Learnt</td>
<td>89</td>
</tr>
<tr>
<td>6 Teaching</td>
<td>1</td>
</tr>
<tr>
<td>A Appendix: Deriving the Cumulative Sum Formuli</td>
<td>3</td>
</tr>
<tr>
<td>A.1 NSW Emergency DepartmentCuSum</td>
<td>3</td>
</tr>
<tr>
<td>A.2 Hutwagner and Cowling</td>
<td>4</td>
</tr>
</tbody>
</table>
## Contents

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>B Appendix: Writing for a non-scientific audience</td>
<td>5</td>
</tr>
<tr>
<td>C Appendix: Published Paper</td>
<td>17</td>
</tr>
<tr>
<td>D Appendix: Presentations</td>
<td>27</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

This thesis is a summary of the work completed as part of the Masters of Applied Epidemiology (MAE) program at the Australian National University. The thesis is made up of 5 chapters, and 4 appendices that together describe all the work required to complete the MAE program.

My field placement was at Healthdirect Australia, which is a Council of Australian Governments (COAG) organisation that delivers a range of telephone and digital health services across Australia. The main services used for disease surveillance are the healthdirect helpline and the healthdirect online symptom checker. Note that for the remainder of this thesis, Healthdirect Australia (HDA) is the company, and is printed in upper case whereas the healthdirect helpline and the healthdirect online symptom checker are branded services, with “healthdirect” written in lower case. Both these services provide health advice and information to consumers. Healthdirect Australia provides a range of smaller services, including My Aged Care, Pregnancy, Birth and Baby, Carer Gateway, NSW Quitline, the after hours GP helpline, Get Healthy, NSW Ambulance secondary triage service, NSW Palliative Care After Hours Helpline and the National Health Services Directory.

I was the first Masters of Applied Epidemiology student at Healthdirect Australia, so my role in the company was continually changing as my supervisor and I adapted to the requirements of the program. My main role in the company was to manage the syndromic surveillance project, which I report on in three of the chapters of this thesis. This included conducting research, developing a business plan, organising and chairing an expert advisory group, coordinating with public health departments, developing and managing the syndromic surveillance web page and presenting the work, both internally and externally. I also assisted in other research and evaluation projects at Healthdirect Australia, consulted with public health departments and universities, and contributed to the daily work of the company.

During the second year of my placement, I was seconded to the Western Sydney public health department to assist in an investigation of measles. This is a vibrant and busy public health department and the staff were very welcoming and taught me a lot during my time there. The majority of my work there was to assist with the measles investigation, but also to contribute to the routine activities of the team.

In the following 4 paragraphs, I introduce each of the projects completed. The first 3 chapters discuss the syndromic surveillance work completed at Healthdirect Australia,
and the last chapter details the measles investigation.

1.0.1 Chapter 2: Surveillance

This chapter discusses a syndromic surveillance system for influenza-like illness created using the healthdirect helpline and the healthdirect online symptom checker data. I discuss the data sources used for the surveillance system, describe the epidemiology of influenza-like illness contacts, then evaluate the proposed system using the US Centres for Disease Control and Prevention Guidelines. I conclude that the proposed surveillance system has excellent representativeness and timeliness, and submit a dashboard that could be used to publish the system. However, weaknesses of the system are flexibility and sensitivity, which are due to the limited details collected on each call or visit.

1.0.2 Chapter 3: Data Analysis

The data analysis chapter describes three projects undertaken to complement the creation of the surveillance system discussed in Chapter 2. The aims of these three projects were: to identify unusual activity in the data; to determine the correlation between Healthdirect Australia data sets with seven other data sets; and to establish a case definition that maximises correlation with other data sets. I find that the healthdirect helpline data correlates very well with other data sources, particularly with emergency department surveillance. I was also able to identify unusual behaviour during the 2009 pandemic year, and to identify the start and finish of an influenza season. I conclude that, due to the high degree of correlation between the Healthdirect Australia data and other data sources, this system could contribute to influenza surveillance in Australia by improving knowledge about the community level of disease.

1.0.3 Chapter 4: Epidemiology Project

In this chapter, I further explore the work of the previous 2 chapters by expanding to 6 further syndromes. Two of these syndromes - adverse reaction following immunisation and gastroenteritis - demonstrated high utility if they were to be added to the proposed influenza-like illness surveillance system. The adverse reaction following immunisation syndrome data shows the ability both to detect a spike in activity, and also to mirror the epidemiological characteristics known about the event. The healthdirect helpline data for gastroenteritis demonstrates similar behaviour to gastroenteritis visits to the emergency department. On the other hand, an investigation into geographical variations in gastroenteritis did not identify smaller regional outbreaks, nor did it show the geographical spread of illness. I conclude however that surveillance on both gastroenteritis and adverse reaction following immunisation could be useful to public health departments.

1.0.4 Chapter 5: Outbreak Investigation

The final chapter discusses work conducted while on secondment at the Western Sydney public health department. My work here consisted of assisting in investigating a measles outbreak, which affected 17 people from March to April on 2017. In the chapter, I discuss the epidemiology of the outbreak and the public health resources committed to containing the outbreak, particularly through contact tracing. I conclude that responding to measles outbreaks pose a large burden on public health departments, particularly
through extra staff hours. However, literature has shown that this effort does result in a significant reduction in secondary measles cases.

1.1 Requirements

This section outlines how I fulfilled the requirements for the Masters of Philosophy (Applied Epidemiology) program.

1.1.1 Field Projects

<table>
<thead>
<tr>
<th>Field Projects</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public health surveillance system establishment</td>
<td>Building a syndromic surveillance system using telephone and internet data</td>
</tr>
<tr>
<td>and evaluation</td>
<td>(Chapter 2)</td>
</tr>
<tr>
<td>Public health data analysis</td>
<td>Investigating influenza-like illness using telephone and internet data</td>
</tr>
<tr>
<td></td>
<td>(Chapter 3)</td>
</tr>
<tr>
<td>Epidemiological study</td>
<td>Investigating the utility of conducting surveillance on a telephone helpline</td>
</tr>
<tr>
<td></td>
<td>for a range of syndromes</td>
</tr>
<tr>
<td></td>
<td>(Chapter 4)</td>
</tr>
<tr>
<td>Field investigation of a public health problem</td>
<td>A measles outbreak in western Sydney, Australia from March 2017 - May 2017</td>
</tr>
<tr>
<td></td>
<td>(Chapter 5)</td>
</tr>
</tbody>
</table>
### 1.1.2 Additional Requirements

<table>
<thead>
<tr>
<th>Literature Review</th>
<th>A literature review was completed for each of the projects listed above</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication (Appendix C)</td>
<td>Hartley, M., Biggs, J., Leonessa, C., Glass, K., Forewarning healthcare practitioners - how telephone and digital data can predict illness in the community. Submitted to the Communicable Diseases Intelligence journal on 15 May, 2017, undergoing revisions.</td>
</tr>
</tbody>
</table>
Hartley, M., Araco, M., Biggs, J., Leonessa, C., Glass, K., Predicting and monitoring influenza-like illness using telephone and digital data. Prepared by M. Hartley, presented by M. Araco at Health Informatics Conference; 6 – 9 August, 2017, Brisbane, Australia  
Hartley, M., Araco, M., Biggs, J., Leonessa, C., Glass, K., Forewarning GPs – how healthdirect helpline data can predict illness in the community. Presented at RACGP Conference; 26 – 28 October, 2017, Sydney, Australia |
| Poster Presentations (Appendix D) | Hartley, M., Biggs, J., Leonessa, C., Glass, K., Forewarning healthcare practitioners: How telephone and digital data can predict illness in the community. Presented at Communicable Diseases Conference 2017; 26 – 28 June, Melbourne, Australia  
Hartley, M., Biggs, J., Leonessa, C., Glass, K., Building a syndromic surveillance system using telephone and digital data. Presented at Communicable Diseases Conference 2017; 26 – 28 June, 2017, Melbourne, Australia |
1.1.3 Teaching

<table>
<thead>
<tr>
<th>Teaching to first year cohort</th>
<th>Wright, A., Hartley, M., McEwen, S., Sheel, M. <em>Critical Thinking: what is wrong with this study?</em>; Presented at MAE courseblock, March 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson from the field</td>
<td>Hartley, M. <em>Lesson from the field 2: Introduction to R programming language</em>; Conducted via teleconference; 1 February, 2017</td>
</tr>
</tbody>
</table>

1.1.4 Coursework

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Name</th>
<th>✔</th>
</tr>
</thead>
<tbody>
<tr>
<td>POPH8914</td>
<td>Issues in Applied Epidemiology</td>
<td></td>
</tr>
<tr>
<td>POPH8917</td>
<td>Public Health Surveillance</td>
<td></td>
</tr>
<tr>
<td>POPH8916</td>
<td>Outbreak Investigation</td>
<td></td>
</tr>
<tr>
<td>POPH8915</td>
<td>Methods in Applied Epidemiological Research</td>
<td></td>
</tr>
<tr>
<td>POPH8915 Analysis of Public Health Data</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 2

Surveillance: Building a syndromic surveillance system using telephone and internet data

2.1 Introduction

This chapter introduces a syndromic surveillance system based on the telephone and online data sources at Healthdirect Australia, and an evaluation of this system. We introduce the data sources used, present some descriptive epidemiology and evaluate using the US Centers for Disease Control and Prevention guidelines.

Healthdirect Australia is a national, government-owned, not-for profit organisation that delivers a range of telephone and digital health services across Australia. This includes the healthdirect helpline which is a telephone-based health advice line and receives over 700,000 calls per year and the healthdirect online symptom checker which receives over 500,000 visits per year. The healthdirect helpline is available 24 hours a day, and is answered by registered nurses. Consumers of the helpline can call, answer questions about their symptoms and receive health advice about further action, which ranges from calling an ambulance to providing self-care at home. The healthdirect online symptom checker is available online and provides a similar service, where consumers answer questions about their symptoms and receive health advice.

Both telephone triage data and internet data are currently being reviewed by many public health departments worldwide as potential sources for disease surveillance. Researchers have compared telephone triage data with GP presentations [1-4], ED surveillance [5-7], lab notifications [1, 4, 7, 8], antiviral prescriptions [7] and FluTracking [8]. Results vary, but most researchers have found good correlation between telephone triage systems and other surveillance systems. Some studies have found a greater degree of correlation by introducing a lag time of 1 – 2 weeks [2, 8, 9] however others have shown no increase in correlation by introducing a lag time [1, 5].

Epidemiologists are now also using internet data for syndromic surveillance. Studies in Sweden [10] and the US [11] have counted the number of clicks on health advice website such as Healthlink (United States), healthdirect (Australia) or Vardguiden.se (Sweden). Other researchers use information from the whole of the web, by utilising keywords in search queries [12] and Google’s keyword ads [13]. Both of these reported
good correlation with comparison data sets. An even more sophisticated method was developed by Google FluTrends [10, 14] which used an automated method of discovering influenza-related search queries. The Google Flu Trends method required no prior knowledge of influenza, and is not limited to pre-determined keywords, but rather tested up to 50 million possible search queries for the best fit with a probability model based on laboratory data. There is currently no published research on using a symptom checker-like tool for syndromic surveillance.

Internet search data can be more volatile than more traditional methods of surveillance. This was demonstrated by Google Flu trends, which predicted twice as many flu cases as the American Centres for Disease Control and Prevention in 2013, and consistently overestimated the prevalence of flu prior to that [14]. This unreliability caused Google to discontinue Google Flu Trends, though further studies have shown that a combination of Google Flu Trends and other data sources can provide more accurate predictions of influenza activity [14].

In Australia, there are many data sources currently contributing to influenza surveillance, including data from general practitioners (GPs), emergency departments (EDs), paediatric wards, ambulances, laboratories and self-reported community levels of influenza-like illness through FluTracking. These data sources are described in "A brief overview of influenza surveillance systems in Australia, 2015" [15].

The data source with the best specificity is notification data available through the National Notifiable Diseases Surveillance System (NNNDSS) since inclusion in the data requires a positive result on a laboratory influenza test. However, this data source does not include all the untested influenza cases in the community, which represents a significant portion of all cases, and there is a lag between a case presenting to a medical centre and reporting of results. Both ED and GP data capture more cases of influenza-like illness (ILI), but are less specific. FluTracking data are self-reported and consequently include people with ILI who do not seek medical attention. This makes it both more sensitive and less specific than the ED, GP and NNDSS data sources, though has the fewest cases per year. The healthdirect helpline also captures people with ILI who do not seek other medical attention and has a large records each year. Similar to FluTracking, it is consequently more sensitive and less specific than the ED, GP and NNDSS data sources. The data are updated daily, so the lag between a case calling and reporting of results is a maximum of 24 hours.

Healthdirect Australia data sources could provide a new way of tracking community-level incidence of illness, since they provide information about people who are symptomatic but do not visit any healthcare providers. Healthdirect Australia research has shown that forty percent of callers to the helpline do not contact another health care provider, as seen in Figure 2.1 [16]. This critical piece of information is useful in monitoring diseases that can affect a large proportion of the population but are largely self-limiting. In order to contribute to disease surveillance in Australia, we used these data to build a syndromic surveillance system using telephone and digital data. Throughout the remainder of this chapter, we consider how to build a syndromic surveillance system in general, using influenza-like illness as an example.
2.1 Introduction

The Commonwealth Department of Health receives a data extract from the healthdirect helpline each week, but ongoing surveillance of this data was discontinued in 2015. In 2013, a Masters of Applied Epidemiology student evaluated Healthdirect Australia telephone data with respect to influenza-like illness. She made the following main recommendations:

1. That Healthdirect Australia ILI data should continue to be reported in the National Influenza Surveillance Scheme report during the influenza season.
2. The ratio of Healthdirect Australia ILI patients who are assigned a final disposition of “Attend ED immediately” to Healthdirect Australia ILI patients who were assigned a final disposition of “Provide self/home care” should be incorporated into the National Influenza Surveillance Scheme report as a measure of severity.
3. All states and territories utilising Healthdirect Australia should be provided with ILI data for their jurisdictions during the influenza season in a timely manner.
4. All states utilising Healthdirect Australia should incorporate the Healthdirect Australia ILI data for their state into their influenza surveillance.
5. The objectives identified in this study for this surveillance system should be adopted by the Commonwealth Department of Health.

Based on this previous research and developmental needs, Healthdirect Australia identified 4 objectives that the syndromic surveillance system would need to achieve in order to support Australia’s health protection mandate and inform early warning systems. These are:

1. To use Healthdirect Australia data to contribute to and complement public health surveillance in Australia by providing validated information about community cases of syndromes.
2. To provide information about syndromes of high public health interest.
3. To forewarn of any emerging public health concerns.
4. To report to the Commonwealth, state health departments and the public in a form that is timely and relevant.
Healthdirect Australia data has the ability to provide information about the following indicators:

1. **Incidence**: This indicator describes the distribution of a syndrome throughout Australia. This can be further investigated by describing of patterns in age groups, by indigenous status, and incidence throughout the year. The data can also be used to compare the incidence in different years. This will provide information about the level of disease in the community, and highlight any high-risk groups.

2. **Severity**: Surveillance concerning the severity of a disease helps describe the impact the disease has on the community, health care workers and any high-risk groups. The proportion of callers that are advised to visit the emergency department is a marker of severity that can be monitored.

3. **Geographical variations**: Syndrome incidence and severity can vary greatly across Australia, so investigating the geographic variation helps provide information which is more relevant to local health departments and health care providers. Surveillance into geographic variation will also build a more detailed picture of the distribution of a syndrome.

4. **Seasonal Thresholds and Signal Identification**: The Healthdirect Australia data in conjunction with cumulative sum algorithms can be used to determine both thresholds for the start and finish of seasonal syndromes (e.g. influenza-like illness) and alerts for unusual activity to help detect outbreaks (e.g. gastroenteritis).

### 2.2 Data Sources

**Healthdirect Helpline Data**

The healthdirect helpline data has three fields that could be used for identifying possible cases of a syndrome. These data fields are called *assessment presenting problem/symptom*, *patient guideline* and *patient question yes*. Some examples of these three fields are given in Table 2.1, and are also described in the following three paragraphs.

The *assessment presenting problem/symptom* is a free text field which is completed at the beginning of the call by the nurse. It is the free text description of what the caller describes as his/her initial reason for calling. The triage system then recommends a *patient guideline* or set of questions that are relevant to the patient’s problem. Each set of questions is identified by a title which is the main symptom of the patient. For example, for a patient inquiring about a headache, the nurse might choose the guideline titled “Headache” or “Head Injury”, while for a patient inquiring about a breathing difficulties the nurse might choose the guideline titled “Breathing problems” or “Wheezing or asthma” or “Colds and Flu”, depending on the exact nature of the query and symptoms. This method ensures that exactly the same set of questions are asked of every patient who calls with those symptoms.

The *patient question yes* is the final question that the nurse asks before the program provides him/her with advice for the patient – this advice is known as a *final disposition*. For example, in the “Colds and Flu” guideline, if the patient answers yes to Question 4
### §2.2 Data Sources

<table>
<thead>
<tr>
<th>assessment presenting problem/ symptom</th>
<th>patient guideline</th>
<th>patient question yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>fever/ 38.9 (PA)/ barking cough/ lying down/ worried about continuing fever</td>
<td>Seen Doctor/ Health Care Provider [Paediatric]</td>
<td>Previously assessed by Doctor or Medibank Health Solutions AND [1] Symptoms have not changed AND [2] improvement expected by time of call</td>
</tr>
<tr>
<td>Croup cough/ awoke with croup cough/ when coughing struggles to breathe/ ribs sucking in with each breath</td>
<td>Croup (Paediatric)</td>
<td>The ribs are pulling in with each breath (retractions)</td>
</tr>
<tr>
<td>fever 38C 5/60 (ax)/ cough</td>
<td>Cough (Paediatric)</td>
<td>Patient also has mild cold symptoms</td>
</tr>
<tr>
<td>HEADACHE FLU LIKE SYMPTOMS FEVER</td>
<td>COLDS AND FLU</td>
<td>DOES THE INDIVIDUAL HAVE ANY OF THE FOLLOWING SYMPTOMS? SWELLING AROUND AN EYE, CONFUSION, STIFFNESS OF THE NECK, HEADACHE</td>
</tr>
</tbody>
</table>

Table 2.1: Examples of data fields in the healthdirect helpline data. *Assessment presenting problem/ symptom* is a free text field entered by the nurse, *patient guideline* is the clinical guideline used by the nurse and *patient question yes* is the final question which prompts the guideline to produce outcome advice. Rows in all lower case are examples from before the guideline change in July 2015, rows in all capitals are examples from after the guideline change.

“Does the individual have any of the following symptoms? [bleeding under the skin, pin-point red, brown, or purple spots in the skin which do not fade when pressed on]” then the final disposition is to call 000. However, in the same guideline, if the patient question yes is Question 28 “What has been the response to any self-care measures tried – no self-care measure or medications have been tried” then the final disposition is to remain at home and provide self-care.

**Healthdirect Online Symptom Checker Data**

There are several possible case definitions using the Healthdirect Australia online data. The healthdirect online symptom checker allows users to click on their main symptom and then follow a series of questions which leads them to a final disposition, similar to the helpline. We can investigate users who choose a symptom, which is similar to the guideline title in the healthdirect helpline. Secondly, we can also track clicks on influenza-related pages on the healthdirect website. Examples of these two possible data fields are in Table 2.2.
Table 2.2: Examples of data fields in healthdirect online symptom checker data and pages visited. The “symptom” field is chosen by the user at the beginning of the healthdirect online symptom checker to describe the “symptom that is bothering you the most”. The ILI related pages are web page on the healthdirect australia website, with the addresses described above.

<table>
<thead>
<tr>
<th>Symptom</th>
<th>ILI related page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colds and Flu</td>
<td>healthdirect.gov.au/colds-and-flu</td>
</tr>
<tr>
<td>Feeling sick or unwell</td>
<td>healthdirect.gov.au/colds-and-flu-symptoms</td>
</tr>
<tr>
<td>Fever</td>
<td>healthdirect.gov.au/influenza-a-h1n1-or-swine-flu</td>
</tr>
</tbody>
</table>

### 2.3 Methods

#### 2.3.1 Case Definition

**Healthdirect Helpline**

There are many options for defining a case definition for the healthdirect helpline data using the three fields patient guideline, assessment presenting problem/symptom or patient question yes. A case definition using the patient guideline has been used by several other countries, including the USA [1], the UK [17] and Canada [5]. Similarly, a keyword search using the assessment presenting problem/symptom has been used by studies in Ireland [2], Sweden [18] and the Public Health Real-time Emergency Department Surveillance System (PHREDSS) in NSW [19]. We chose to use just the patient guideline field and used a correlation method to determine which guidelines titles to include. For a description of how the guideline titles were chosen, please refer to Section 3.3.2.

We validated the case definition using a correlation method since validating data against the WHO definition, as has been used elsewhere, was not possible here. Due to the scant information available for each call or visit, we are unable to determine which cases are actually influenza or influenza-like illness according to the WHO definition. Therefore there is no gold standard against which to compare the Healthdirect Australia data. Instead this case definition was validated by determining the correlation between the Healthdirect Australia data and six other data sources. The fit was measured by Spearman’s rank correlation coefficient, Pearson’s correlation coefficient and a cross-correlation function was calculated. Further details of how the case definition was developed is discussed in Chapter 3. For more details of the results of this analysis, please refer to Section 3.4.2.
Table 2.3: Case definition for influenza-like illness calls to the healthdirect helpline. There was a systematic change in all guidelines on 1 July 2015, so the case definition includes guideline titles from both before and after the change. Guidelines used before the change are in lower case and guidelines used after are in UPPER CASE.

Table 2.4: Case Definition for the healthdirect online symptom checker.

The final case definition, which we have used for all the following analysis is in Table 2.3.

**Healthdirect Online Symptom Checker**

The online data for the healthdirect website is made up of the healthdirect online symptom checker data and clicks to health pages. For this research, we chose to investigate the healthdirect online symptom checker only, since that was the data that match the healthdirect helpline system the most closely. The healthdirect online symptom checker has fewer guideline titles than the healthdirect helpline system, so determining a case definition was less difficult. We trialled 4 different case definitions, but the only definition that demonstrated the seasonal shape of a typical influenza year was the symptom of *Colds and Flu*. Therefore the final case definition used in the following analysis is in Table 2.4.

**Adjusting for Healthdirect Online Symptom Checker Errors**

During routine surveillance of the healthdirect online symptom checker in August of 2016, we identified an unusual dip in the rate of influenza activity, which was likely
Estimating the true rate of influenza-related activity in the symptom checker from August 26 - September 7

Figure 2.2: This demonstrates the method used for estimating the percentage of influenza-related visits to the healthdirect online symptom checker. The green represents the usual way of calculating the percentage, which is the total number of cold and flu visits divided by the total number of visits. The blue is the total number of cold and flu visits divided by the number of visits in this subgroup. The orange is the estimate which is a linear function of the blue.

due to a technological change. This change affected some users attempting to access the healthdirect online symptom checker. From August 26 – September 7, a new structure for the healthdirect online symptom checker data was trialled but was proved to be faulty. This new structure affected the page which allowed consumers to select the symptom that troubled them the most. The error caused some consumers to log on and count in the “visit” denominator, but then be unable to click on the symptom “colds and flu,” which caused the percentage of cold and flu visits to drop dramatically. However, the “colds and flu” algorithm is accessible both by choosing from a list of symptoms, and choosing from a list of topics. Consequently, patients were still able to access the colds and flu topic by selecting their topic (but not by selecting their symptom).

In order to adjust for the period in which the healthdirect online symptom checker was faulty, we identified a denominator that showed the effect of fewer people being able to progress via the method of selecting their symptom. To do this, we looked at the subgroup of people who had clicked on a topic, instead of all people who had visited the home page. We then compared the percentage of all cold and flu visits divided by the number of visits to the subgroup and the percent of all cold and flu visits divided by all visits (Figure 2.2). The subgroup correlated very well with the whole sample, with a correlation coefficient of 0.93 for the 2 months prior to the outage and 0.96 in the two
§2.3 Methods

Percentage of influenza-related visits to the Healthdirect symptom checker

Figure 2.3: Incidence of Influenza-like Illness queries for the healthdirect online symptom checker as a percentage of all queries.

months after. Consequently, we estimated the percentage of flu calls as a linear function of the percentage of flu calls in the subgroup.

2.3.2 Indicators

We identified four indicators that would allow us to assess the disease surveillance tool. The methods for obtaining these indicators are described below.

1. Incidence: The incidence of a syndrome is calculated by extracting the number of calls or visits that meet the case definition each day and also the total number of calls or visits each day to determine the percentage of calls or visits. This percentage is similar to a percent positive, which is reported by other surveillance systems such as the Australian Sentinel Practices Research Network (ASPREN). Using the percentage positive in this way removes the effect of annual fluctuations in call or visit volumes. We can also stratify the incidence by age and indigenous status.

2. Severity: A severe case is defined as a user who was advised to dial 000 or visit the emergency department. We represent the severity of ILI in two ways, both of which describe an aspect of the severity of the season. These methods are: A) severe ILI calls as a percentage of all ILI calls; and B) the total number of severe ILI calls.
3. **Geographical variations**: We identified differences in the incidence of influenza in different regions of Australia by stratifying the incidence data by state. We extracted the number of calls that meet the ILI definition per day, and the total number of calls per day to find the percentage in each state each day. Note that Queensland and Victoria are not part of the healthdirect helpline, and consequently are not included in this representation.

4. **Seasonal Thresholds and Signal Identification**: We use 3 different formulae to determine seasonal thresholds and signal identification, both of which are a modification of the cumulative sum (CuSum). These 3 methods are called the Hutt-wagner method, which identifies unusual behaviour, the Cowling method and the PHREDSS method, both of which identify the start and finish of the flu season. The CuSum formula finds an index, $S_i$, by calculating the difference between an observation $X_i$ and a reference value, $k$. This is an iterative process where the CuSum Index, $S_i$, is updated using the CuSum index of the previous day, $S_{i-1}$. More details of the CuSum are in Chapter 3 and a derivation of the formula is in Appendix A.

2.3.3 **Attributes of the surveillance system**

We used the United States Centers for Disease Control and Prevention surveillance system attributes as a foundation for our surveillance system. These attributes are: simplicity, flexibility, acceptability, sensitivity, predictive value positive, representativeness and timeliness. We considered how best to achieve a good results for each of these attributes while determining how to build the surveillance system. These can be used as an evaluation tool in future studies.

2.4 **Results**

2.4.1 **Indicators**

2.4.2 **Descriptive Epidemiology**

There are more female callers (58.9%) than male callers (41.1%) to the healthdirect helpline, and the age group with the largest number of calls is the 0 – 4 age group with 1,813,454 calls, which makes up 32.4% of all calls to the helpline. The two states with the largest number of calls are NSW and WA, due mostly to the large population in these two states. Victoria and Queensland are not members of Healthdirect Australia. The small number of people calling from these states are likely to be either on holidays in one of the participating states at the time of calling, or live in a border town (for example, Albury/ Wodonga). The most number of calls occurred in 2013, though the number of calls remains similar across the years. The healthdirect helpline began on July 8, 2008 and this thesis covers data until October 10, 2017. Consequently, 2008 and 2017 have fewer calls because the data does not cover the whole year.

**Incidence - Healthdirect Helpline**

The incidence of influenza-like illness related calls to the healthdirect helpline, is seen Figure 2.4. Influenza-like illness demonstrates a seasonal trend with a peak around July
§2.4  Results

<table>
<thead>
<tr>
<th>Age and Sex of calls</th>
<th>number</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>5,593,100</td>
<td>100%</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2,297,550</td>
<td>41.1%</td>
</tr>
<tr>
<td>Female</td>
<td>3,294,190</td>
<td>58.9%</td>
</tr>
<tr>
<td>Unknown</td>
<td>1360</td>
<td>0.02%</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 – 4</td>
<td>1,813,454</td>
<td>32.4%</td>
</tr>
<tr>
<td>5 – 9</td>
<td>384,785</td>
<td>6.9%</td>
</tr>
<tr>
<td>10 – 14</td>
<td>187,511</td>
<td>3.4%</td>
</tr>
<tr>
<td>15 – 19</td>
<td>241,013</td>
<td>4.3%</td>
</tr>
<tr>
<td>20 – 24</td>
<td>373,681</td>
<td>6.7%</td>
</tr>
<tr>
<td>25 – 29</td>
<td>400,405</td>
<td>7.2%</td>
</tr>
<tr>
<td>30 – 34</td>
<td>379,843</td>
<td>6.8%</td>
</tr>
<tr>
<td>35 – 39</td>
<td>299,157</td>
<td>5.3%</td>
</tr>
<tr>
<td>40 – 44</td>
<td>239,376</td>
<td>4.3%</td>
</tr>
<tr>
<td>45 – 49</td>
<td>199,721</td>
<td>3.6%</td>
</tr>
<tr>
<td>50 – 54</td>
<td>186,214</td>
<td>3.3%</td>
</tr>
<tr>
<td>55 – 59</td>
<td>171,030</td>
<td>3.1%</td>
</tr>
<tr>
<td>60 – 64</td>
<td>162,204</td>
<td>2.9%</td>
</tr>
<tr>
<td>65 – 69</td>
<td>146,157</td>
<td>2.6%</td>
</tr>
<tr>
<td>70+</td>
<td>408,469</td>
<td>7.3%</td>
</tr>
<tr>
<td>Unknown</td>
<td>81</td>
<td>0.001%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location and Year of calls</th>
<th>number</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>245,388</td>
<td>4.4%</td>
</tr>
<tr>
<td>NSW</td>
<td>2,653,406</td>
<td>47.4%</td>
</tr>
<tr>
<td>NT</td>
<td>84,181</td>
<td>1.5%</td>
</tr>
<tr>
<td>QLD*</td>
<td>12,037</td>
<td>0.2%</td>
</tr>
<tr>
<td>SA</td>
<td>804,725</td>
<td>14.4%</td>
</tr>
<tr>
<td>TAS</td>
<td>197,411</td>
<td>3.5%</td>
</tr>
<tr>
<td>VIC*</td>
<td>15,507</td>
<td>0.3%</td>
</tr>
<tr>
<td>WA</td>
<td>1,418,989</td>
<td>25.4%</td>
</tr>
<tr>
<td>Unknown</td>
<td>161,456</td>
<td>2.9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>number</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008**</td>
<td>188,674</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>536,823</td>
<td>3.3%</td>
</tr>
<tr>
<td>2010</td>
<td>552,427</td>
<td>9.6%</td>
</tr>
<tr>
<td>2011</td>
<td>596,184</td>
<td>10.7%</td>
</tr>
<tr>
<td>2012</td>
<td>684,584</td>
<td>12.2%</td>
</tr>
<tr>
<td>2013</td>
<td>705,340</td>
<td>12.6%</td>
</tr>
<tr>
<td>2014</td>
<td>648,316</td>
<td>11.6%</td>
</tr>
<tr>
<td>2015</td>
<td>611,696</td>
<td>10.9%</td>
</tr>
<tr>
<td>2016</td>
<td>591,006</td>
<td>10.6%</td>
</tr>
<tr>
<td>2017**</td>
<td>478,050</td>
<td>8.5%</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 2.5: Age, Sex, Location and Year of calls to the healthdirect helpline.

* Queensland and Victoria are not members of Healthdirect Australia.
** 2008 and 2017 are not complete years

of each year and a trough around January of each year. The year with the highest percentage of flu-related calls was 2009, during the H1N1 pandemic.

The rate of influenza-related calls to the helpline is different across age groups. In general, children under the age of ten are over-represented for influenza-like illness, making up nearly sixty percent of all ILI calls, as opposed to thirty percent of all calls. In 2010 – 2017, the rate of influenza-related calls in the 0 – 5 age group was the highest, with 12% of calls in that age group being influenza-related. In all other age groups, the percentage of influenza-related calls is below 5% of all calls. However, in 2009, the percentage of influenza-related calls increased in all age groups, with the largest increase in the 10 – 60 age groups. These age groups saw almost a three fold increase in the percentage of influenza related calls, with the percentage increasing from below 3% of calls to between 8.5% and 10% of calls.
<table>
<thead>
<tr>
<th>Guideline Title</th>
<th>number</th>
<th>percent</th>
<th>Guideline Title</th>
<th>number</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vomiting (Paediatric)</td>
<td>134,167</td>
<td>7.4%</td>
<td>Vomiting (Paediatric)</td>
<td>9,416</td>
<td>1.2%</td>
</tr>
<tr>
<td>Fever (Paediatric)</td>
<td>92,932</td>
<td>5.1%</td>
<td>Headache (Paediatric)</td>
<td>6,523</td>
<td>0.8%</td>
</tr>
<tr>
<td>Cough (Paediatric)</td>
<td>82,909</td>
<td>4.6%</td>
<td>Abdominal Pain - Female (Paediatric)</td>
<td>5,848</td>
<td>0.7%</td>
</tr>
<tr>
<td>Colds (Paediatric)</td>
<td>78,513</td>
<td>4.3%</td>
<td>Trauma - Head (Paediatric)</td>
<td>5,791</td>
<td>0.7%</td>
</tr>
<tr>
<td>Trauma - Head (Paediatric)</td>
<td>67,626</td>
<td>3.7%</td>
<td>Abdominal Pain - Male (Paediatric)</td>
<td>5,684</td>
<td>0.7%</td>
</tr>
<tr>
<td>Total calls</td>
<td>1,813,454</td>
<td></td>
<td></td>
<td>813,309</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Guideline Title</th>
<th>number</th>
<th>percent</th>
<th>Guideline Title</th>
<th>number</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication Questions - Adult</td>
<td>15,960</td>
<td>0.7%</td>
<td>Medication Questions - Adult</td>
<td>11,637</td>
<td>1.6%</td>
</tr>
<tr>
<td>Abdominal Pain</td>
<td>12,544</td>
<td>0.6%</td>
<td>Chest Pain</td>
<td>7209</td>
<td>1.0%</td>
</tr>
<tr>
<td>Chest Pain</td>
<td>10,269</td>
<td>0.5%</td>
<td>Abdominal Pain</td>
<td>4985</td>
<td>0.7%</td>
</tr>
<tr>
<td>Abortion less than 20 weeks</td>
<td>7,436</td>
<td>0.3%</td>
<td>Dizziness / Vertigo</td>
<td>4929</td>
<td>0.7%</td>
</tr>
<tr>
<td>Headache</td>
<td>6,338</td>
<td>0.3%</td>
<td>Limb Pain</td>
<td>3727</td>
<td>0.5%</td>
</tr>
<tr>
<td>Total calls</td>
<td>2,249,427</td>
<td></td>
<td></td>
<td>716,830</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.6: Guidelines that received the most number of calls in four age groups.

Incidence - Healthdirect Online Symptom Checker

The healthdirect online symptom checker started in October 2014 and use has been steadily increasing since its inception, from 1,685 visits in the first week to 30,419 visits in September 2017. Influenza-related activity in the healthdirect online symptom checker shows a similar seasonal trend to the helpline, as seen in Figure 2.3. However, there have been several discrepancies with the data, caused by technical difficulties, traffic rerouting and marketing. This is visible in the large dip in influenza-related visits in August of 2016, and also in the unusual shape of the influenza season in 2017.

We were able to identify a technological error in August of 2016, which is the likely cause of the large dip in the influenza-related visits. However, we were unable to identify a single cause of the unusual shape of the activity in 2017. From January to July 2017, the percentage of influenza-related visits stays below 4%, when the usual range is from 5% – 10%. Then from July 7, 2017 the rate of influenza-related visits increases at a very high
Figure 2.4: Incidence of influenza-like illness calls to the healthdirect helpline, as a percentage of all calls.

rate to a peak of 19.3% on September 8, 2017. This is not consistent with the influenza-related activity in the healthdirect helpline (Figure 2.6).

Severity

The two methods of describing the severity of influenza described in the methods section demonstrate different aspects of the disease epidemiology. Method A demonstrates that the severe ILI calls as a percentage of all ILI calls is largely constant in every year (Figure 2.7). In Method B which uses the number of severe ILI calls we see that the number of severe flu calls follows a seasonal pattern. The largest number of severe calls occurred in 2009.

Geographic variation

The seasonal impact of influenza is clear in each state and territory in Australia. However, the seasonal effect is much less pronounced in the Northern Territory, with the peak of influenza-related activity occurring in September at 6.6% of all calls in the Northern Territory. By contrast, the peak of flu season in the other states occurred in September or August, with between 8.8% and 12.6% of calls. (Figure 2.8).
2.4.3 Consideration of Surveillance System Attributes

In order to create a high quality surveillance system utilised by stakeholders we considered the seven attributes of surveillance systems described by the Centers for Disease Control and Prevention. Table 2.8 describes the considerations for each attribute and any steps taken while designing the system in relation to that attribute.
§2.4 Results

Percentage of influenza-related encounters: healthdirect helpline and online symptom checker

Figure 2.6: Incidence of influenza-related visits in both the healthdirect online symptom checker and the healthdirect helpline. The unusually low activity in the healthdirect online symptom checker in early 2017, followed by a small, sharp spike in late September 2017 does not match the influenza activity in the healthdirect helpline, even though the activity is well matched in previous years.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>System Strengths and Limitations</th>
<th>Steps Taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplicity</td>
<td>Data already routinely collected daily, cleaned and saved into a database.</td>
<td>Reporting process will be automated and will include accessing the routinely collected data, running analysis including health alerts and updating an online platform.</td>
</tr>
<tr>
<td>Attribute</td>
<td>Continuation of CDC Attributes</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>System Strengths and Limitations</strong></td>
<td><strong>Steps Taken</strong></td>
</tr>
<tr>
<td>Flexibility</td>
<td>Adding new syndromes limited by ability of the data sources to identify cases.</td>
<td>Created list of all existing guideline titles in the data set.</td>
</tr>
<tr>
<td></td>
<td>Identifiable cases are limited to one of 130 predetermined guideline titles, or a very short free text field.</td>
<td>Created a list of possible new syndromes to add to the system.</td>
</tr>
<tr>
<td></td>
<td>New syndromes that are identifiable by guideline titles can be added to surveillance system.</td>
<td></td>
</tr>
<tr>
<td>Acceptability</td>
<td>This is a new system with a new data source, so in order to encourage use we need to make the reporting simple, easy to use and straightforward to integrate into routine state and territory reporting.</td>
<td>Formed expert advisory group HATSS consisting of senior members of NSW Ministry of Health, WA Department of Health and Commonwealth Department of Health to advise on how to make the system as acceptable to users as possible.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presented proposed surveillance system to special interest groups such as the National Influenza Surveillance Committee for feedback.</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>This system is not aiming to detect all cases of influenza-like illness.</td>
<td>Investigated three different methods of detecting unusual activity including the method used by the NSW Emergency department surveillance.</td>
</tr>
<tr>
<td></td>
<td>However, this system does have the ability to detect an outbreak.</td>
<td>Demonstrated that the system has the ability to detect unusual activity.</td>
</tr>
</tbody>
</table>
## 2.4 Results

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Continuation of CDC Attributes</th>
<th>Steps Taken</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predictive Value Positive</strong></td>
<td>Cannot calculate exactly since there is no way of determining how many callers had the disease in question (for example, how many callers with influenza-like symptoms had a positive laboratory influenza result).</td>
<td>To determine an approximate predictive value positive we calculated the correlation between ILI calls to the HDA data and 6 data sets: Lab notifications of influenza in Australia Lab notifications of influenza in NSW ED surveillance of ILI in NSW ED surveillance of ILI in WA FluTracking surveillance of ILI in Australia Sentinel GP surveillance (ASPREN) of ILI in Australia. Results showed very good correlation between the healthdirect helpline data and all other data sources, including lab notifications of influenza which implies that the predictive value positive is high, as shown in Chapter 3.</td>
</tr>
<tr>
<td><strong>Representativeness</strong></td>
<td>This system has the potential to have excellent representativeness due to its large sample size of 5,593,100 calls and the ability to be easily accessed in rural and remote Australia. Data quality for the system is very good, due to the data being automatically populated by the program used in the call centres.</td>
<td>Completed descriptive epidemiology of the data set to determine how representative it is of the Australian population.</td>
</tr>
<tr>
<td><strong>Timeliness</strong></td>
<td>Data already routinely collected daily</td>
<td>Platform will be updated daily. Analysis including any health alerts of outbreaks will be completed daily.</td>
</tr>
</tbody>
</table>
Method A: shows the percentage of all influenza-related calls that are classed as severe. This means calls with a guideline title related to either cough or colds and flu which is given the advice to call 000 or go to the emergency department. This percentage is constant throughout the year. Method B shows the total number of severe influenza-related calls to the healthdirect helpline.

2.5 Discussion

The aim of this project is to determine the viability of the two Healthdirect Australia data sources to contribute to surveillance in Australia by investigating influenza-like illness. Each data source was considered for viability by investigating how well the data source
2.5 Discussion

Figure 2.8: Influenza-like illness in each state and territory in each month of 2016.

could describe aspects of the disease and the CDC attributes. The four aspects considered were: incidence, severity, geographical variation and seasonal thresholds or signal identification. In this section we will discuss each of these aspects for both data sources.

2.5.1 Indicators

Incidence - Telephone

We identified a seasonal trend in all years. A study published by Newall and Scuffham showed that the rate of hospitalisation for influenza-like illness is highest in children aged under 5 years and adults aged over 60, however the rate of GP visits is highest in the 15 – 44 age group\[20\]. For the healthdirect helpline data, callers in the 0 – 5 age group have the highest percentage of influenza-related calls. This is likely due to the fact that older callers are more likely to call about other health issues, for example abdominal pain. The number of calls and the age profile of calls in 2009 differs from all other years in the dataset, when there was a large increase in callers between the ages of 10 and
<table>
<thead>
<tr>
<th>Year</th>
<th>Start Date (Month)</th>
<th>Finish Date (Month)</th>
<th>Peak Date (Month)</th>
<th>Cowling Index with cut-off at 270</th>
<th>PHREDDS Index with cut-off at 35</th>
<th>NSW Health Index of Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>May 24</td>
<td>Sept. 18</td>
<td>July 28</td>
<td>May 1</td>
<td>Aug. 12</td>
<td>June 14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>July 6</td>
<td>July 24</td>
<td>July 15</td>
<td>Aug. 8</td>
<td>Oct. 12</td>
<td>Sept. 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>June 13</td>
<td>Aug. 21</td>
<td>July 17</td>
<td>May 24</td>
<td>Sept. 4</td>
<td>July 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>May 31</td>
<td>Aug. 31</td>
<td>July 15</td>
<td>May 20</td>
<td>Aug. 16</td>
<td>July 12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>June 10</td>
<td>Sept. 15</td>
<td>July 12</td>
<td>June 9</td>
<td>Sept. 20</td>
<td>Aug. 29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>June 14</td>
<td>Oct. 2</td>
<td>Aug. 27</td>
<td>June 6</td>
<td>Sept. 23</td>
<td>Aug. 20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>June 12</td>
<td>Oct. 10</td>
<td>Sept. 5</td>
<td>June 29</td>
<td>Oct. 7</td>
<td>Aug. 20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.7: Start and finish dates for influenza-like illness using Cowling Index, PHREDSS Index and the NSW Emergency Department Surveillance System. Note that the NSW Emergency Department only began publishing the start and finish of the flu season in 2013.

This could be due to the epidemiology of the A (H1N1) virus which had a younger median age of infection compared with A (H3N2) \[21\]. The younger median age of infection does not fully explain the increase seen in the healthdirect helpline however, since the median age only changed by 8 – 13 years, for example from a median age of 31 to 18 in Western Australia, and the change in the healthdirect helpline was greater than that. Another possible explanation is that people in the 10 – 70 age range were more affected by worry and anxiety, which has been documented in several university students in Stanford \[22\], at UNSW \[23\] and by the public in Sydney \[24\]. If this were true, it might cause people in this age group to call the helpline for less severe illness, and consequently artificially inflate the attack rate in this age group.
Figure 2.9: Moving 7-day average of Hutwagner, Cowling and PHREDSS Indices operating on the percentage of ILI calls as a proportion of all calls to the healthdirect helpline. The intersection of the index (black solid line) with the cut-off (blue dot-dash line) indicates the start and finish of the flu season for the Cowling and PHREDSS indices, and a threshold for signalling unusual behaviour for the Hutwagner index.
Incidence - Healthdirect Online Symptom Checker

There was a strong seasonal trend visible in the influenza-like illness visits to the healthdirect online symptom checker. However, the dip in influenza-related visits in August 2016, the lower incidence in the first half of 2017, followed by the very sharp increase in incidence shows that this data is more volatile. The sharp dip in August 2016 could be explained by the faults in the checker and has been adjusted for in the analysis. However, the odd shape of the influenza activity in 2017 could be due to a number of different factors, including a larger number of pages linking to the healthdirect online symptom checker from mental health related content, symptom-specific marketing, age groups-specific marketing or marketing on platforms with particular age profiles such as Facebook.

The factors listed above would affect the profile of the 2017 influenza curve since we are reporting influenza-related symptom checker visits as a proportion of all visits. If the number of visits for other symptoms increases, the proportion of influenza-related visits decreases. When Healthdirect Australia directs traffic from mental health related pages or conducts marketing for other symptoms, the proportion of influenza-related clicks decreases. Age specific marketing or marketing on platforms with particular age profiles could change the shape of the influenza curve since the proportion of people clicking on influenza-related illness is higher in the younger age groups than in the older. For example, if Healthdirect Australia were to conduct marketing on a website targeting young parents, this would increase the proportion of influenza-related clicks.

This demonstrates that the healthdirect online symptom checker data source is prone to greater variability than healthdirect helpline data since there are more factors influencing who clicks on the healthdirect online symptom checker and when. These factors include symptom-specific marketing, the proportion of visitors that are redirected from other sites, the healthdirect online symptom checker’s place in a google search for various symptoms and also technical difficulties, such as broken links. However, there is great potential for using this data, particularly in conjunction with the healthdirect helpline data. There is some evidence that combining data sources in this way helps to improve the reliability of the predictions [9].

Severity

The severity of influenza-like illness, as measured by the proportion of callers advised to visit the emergency department did not change from year to year. Although there are more influenza cases during the winter, as seen by the seasonal trend in the number of severe ILI cases, the proportion of cases being advised to visit the emergency department remains the same throughout the year.

Geographic variation

The effect of the tropical climate on influenza-like illness was visible the healthdirect helpline data, since the seasonal trend in the Northern Territory is much less pronounced
than in other states and territories. This is consistent with influenza activity in the Northern Territory as observed through laboratory surveillance.

**Seasonal Thresholds and Signal Identification**

The three indexes investigated demonstrate that the healthdirect helpline data has the ability to detect unusual activity and also to identify the start and finish of the influenza season in Australia. In particular, for identifying unusual activity, we can set an alarm threshold which is above the maximum of all years except for 2009. This threshold would have triggered on May 27, 2009, seven days after the second case of A (H1N1) influenza was detected in Australia. This is very important since identifying unusual activity is one of the goals of surveillance, and is included in the "sensitivity" attribute. More details about the strengths and weaknesses of each of these indexes is provided in Sections 3.4.1 and 3.5.1 of this thesis.

**2.5.2 Consideration of Surveillance System Attributes**

This system demonstrates excellent representativeness due to both the large sample size and the proportion of users that do not seek other health care, estimated at forty percent [16]. Furthermore, this system is one of only 2 which gives an indication of the community level of disease and can provide insight into the broader impact of influenza in Australia. The other major strength of the system is the timeliness, since the data are updated daily and therefore can be reported contemporaneously. Since these data are already collected daily as part of standard operating procedure, the system has excellent simplicity, data quality and stability.

The system is largely constrained by the coarseness of the data, leading to the two weaknesses of flexibility and sensitivity. The flexibility of the system is limited since adding new syndromes is only possible if the symptoms of the syndrome match one of the predetermined list of guideline titles or are identifiable in the short free text field. However, there are a large range of syndromes already identified as possibilities for surveillance. This system’s other weakness is sensitivity because it does not have the ability to detect all cases of influenza as only a caller’s main symptom is recorded. Consequently, it is possible that a caller with influenza-like illness is listed under a different symptom - for example fever- and is not captured in the definition. However, in order to test whether the system was sensitive enough, we investigated the correlation between the healthdirect helpline and six different data sources. This is described in more detail in Chapter 3. The correlation between the healthdirect helpline and lab confirmed cases of influenza was good, with an average correlation of 0.79 (range in yearly correlations 0.49 – 0.93). Correlation with other data sources was even better, particularly with the NSW Emergency room surveillance with an average correlation of 0.96 (range in yearly correlations 0.92 – 0.99). Consequently we can conclude that the healthdirect helpline data is sensitive enough to accurately represent community incidence of influenza.

**2.6 Implementation**

The next important steps are to present the data from the syndromic surveillance system in a way that is accessible to public health professionals. These could include local health
departments, jurisdictions, national surveillance groups such as OzFoodNet, universities and other research or surveillance institutions.

To this end, we have created a syndromic surveillance dashboard, which is interactive, modifiable and updated daily. The dashboard is currently viewed by staff at Healthdirect Australia and we are planning on distributing the dashboard more broadly. Healthdirect Australia data has the potential to report on syndromes as diverse as ILI, gastroenteritis, acute fever and rash, adverse reactions to vaccines, acute mental health episodes and drug and alcohol related illness. This dashboard was also introduced to the expert advisory board via a youtube video. For a demonstration of this dashboard, please go to https://www.youtube.com/watch?v=MLgK9OPU3Ao We aim to publish this dashboard more broadly via the healthdirect web page.

2.7 Limitations

The healthdirect helpline data has very limited information about symptoms of a case and only describes the main symptom of the caller. This means that many cases could be misclassified, causing both callers without ILI to be identified as cases and cases with ILI not to be identified. We were also unable to obtain more information or to test against a gold standard definition since we could not contact the callers. To take this into account, we assumed that these errors were equally likely to occur and consequently in a large data source such as the healthdirect helpline they would cancel each other out.

The internet data was even more coarse than the healthdirect helpline data, meaning this data set is also prone to misclassification errors. We therefore make the same assumption, that the two kinds of misclassification are equally likely to occur and would cancel each other out in a large data set. The online symptom checker data was also more volatile and subject to change. As mentioned in the discussion section, some of the unusual changes in the influenza-related activity could be due to known technical issues, but no clear explanation could be found for other changes. In particular, the technical error that occurred between August 26 and September 7, 2016 was at the same time as the peak of influenza activity in that year, which could have a large impact on the correlations.

Finally, we were unable to conduct a complete evaluation of the surveillance system since it has not yet been established. For this reason, we were unable to interview key stakeholders about their experiences with the healthdirect helpline surveillance system and discussion about aspects such as acceptability, flexibility and simplicity are based on the comments of a few members of an expert advisory group. The opinions of these members may differ from the experience of stakeholders once the surveillance system is established.

2.8 Public Health Implications

This surveillance system has the potential to greatly improve knowledge of community level disease, especially for syndromes such as influenza-like illness. Once built, the
Lessons Learnt

Building this system was an excellent learning opportunity for me because it was largely outside of my expertise. In particular, I learnt about how project management, stakeholder engagement and concept-selling are important to a scoping project like this one. I also learnt about how to use the work of other researchers to build my own methods instead of exactly copying their methods.

Project management was an integral part of getting this project running and I was lucky to be able to attend a project management workshop to get a short overview of how to be an effective project manager. What struck me immediately was that ideally a project needs to be planned from start to finish before the commencement of any work. In particular, planning out any time limitations, especially those that are outside of your control will help to ensure that you complete steps in the most efficient way. This came up particularly when requesting data and building the software platform for the surveillance system. In real-time, we could not build a sample platform until the data analysis was complete, however there are many aspects of building the platform that could begin before the analysis is complete. This included steps like determining which software to use, obtaining quotes and applying for funding. If all these steps are completed before data analysis is complete, it ensures that you can move directly from data analysis to uploading the analysis results onto the platform. Similarly, the data
requests had to occur long before I actually wanted to use the data since it took up to 3 months between sending the data request and receiving the data.

Stakeholder engagement was also an integral part of this process since we were aiming to build a system that would be utilised by the stakeholders. In particular, I learnt how important it is to approach senior members of the jurisdictional health departments through the correct avenue instead of via email or word of mouth. This meant that the members of our expert advisory group had permission from their supervisors to be part of the group, and consequently had more time to contribute to the advisory group. Members were also better able to leverage advice from within their departments with official approval.

Finally, I learnt about concept-selling in order to secure funding to build the platform for the surveillance system. I had not anticipated the platform to be so costly, but soon realised that the freeware solutions that we thought might work didn’t look as professional as we would have liked. Consequently we created a proof of concept to demonstrate the scope of the system. This was a new way of securing funding for me, and a great introduction to a new way of working on a project.

Analytically, I also learnt about the value of creating a new method instead of adapting a method that doesn’t quite fit the research aims. I was looking for a way of creating and then validating a case definition for influenza-like illness using the data fields available. Other researchers had created case definitions for the healthdirect helpline data, and initially I intended to adapt their methods. However the methods described were not applicable to healthdirect helpline data because we had no way of determining if a caller truly had the disease. Consequently we could not compare our case definition with a "gold standard" case definition. Instead, I was advised by the National Influenza Surveillance Committee to try a different technique not used in any of the papers I had researched. On reflection, trying new and novel methods or adapting methods from other fields for use in your project can be a powerful research tool, and something I should consider more often.
Chapter 3

Data Analysis: Investigating influenza-like illness using telephone and internet data

3.1 Introduction

This chapter describes in detail the data analysis that was conducted in order to build the syndromic surveillance system in Chapter 1. As with Chapter 2, this chapter focusses on influenza-like illness and influenza-related health care contacts. There are three aspects to the data analysis: signal identification, determining a case definition and comparing Healthdirect Australia data with other data sources. Signal identification describes methods used to identify unusual activity. Determining a case definition describes the method of choosing between seven different case definitions. The final section compares the healthdirect helpline data with six other data sources to determine the degree of correlation between influenza-related activity in the data sources.

Influenza is a viral disease that is easily transmitted from person to person and causes significant morbidity and mortality annually. It is estimated to affect 5 to 10 percent of the adult population and twenty to thirty percent of children annually [25]. Influenza has the greatest health impact on people aged 65 and over, with average mortality rates in Australia (2003 – 2008) estimated as 13.0 deaths per 100,000 in people 65 and over, while in people under 65, the rate is estimated at 0.4 deaths per 100,000. There are between 18,000 – 207,000 hospitalisations for influenza and pneumonia in Australia each year as well as 310,000 GP visits [20]. Influenza is also estimated to cost the Australian healthcare system around $115 million annually due to visit to GPs and hospitalisations [20]. The indirect costs of influenza, particularly due to productivity losses are estimated to contribute an additional 60% - 400% or $70 - $460 bringing the total to between $185 - $460 million annually [26].

In temperate climates, influenza activity peaks during the winter. In Australia, this seasonal peak is typical in New South Wales, Tasmania, Victoria, Australian Capital Territory, South Australia, and the southern regions of Queensland and Western Australia. In tropical climates such as the Northern Territory and the northern regions of Western Australia and Queensland, influenza can occur at any point during the year [27]. There have been a number of theories postulated to explain the seasonality of flu, falling into three main categories: changes in host resistance to infection such as fluctuations in
melatonin or circulating vitamin D; changes in survival of the virus due to ambient
temperature, relative humidity and vapor pressure; and changes in host behaviour, such
as more time spent indoors and grouping of susceptible children in schools [28].

There is an vaccine for influenza that includes currently circulating influenza strains,
which community members are recommended to have annually. The WHO has issued
an bi-annual report (once each for the northern and southern hemispheres) since 1973
with recommendations for composition of the vaccine [29] based on the analysis of
samples sent to five laboratories in Atlanta, Beijing, London, Tokyo and Melbourne.
These laboratories receive samples from public and private laboratories in their regions,
and analyse the samples to determine their similarity with strains included in the vaccine
and other circulating strains and also their resistance to antiviral treatment [30]. Some
samples are then further selected for gene sequencing to determine genetic diversity and
the genetic basis of antigenic changes. This information then allows the laboratories to
recommend suitable influenza strains for the next annual vaccine.

Seasonal influenza has a large impact on the health system. The increased rates
of hospital visits is one of the main causes of emergency department crowding [31]
and has been shown to have a significant impact on hospital mortality [32]. Consequently it is helpful to determine the start and finish of the flu season in order to assist
hospitals, doctors and public health departments with staffing, antiviral medication,
public health messaging and other preventative or containment strategies. Influenza
has also caused a number of pandemics. These pandemics have ranged in severity
from the Spanish flu in 1918 which killed an estimated 50 – 100 million people [33]
to the 2009 H1N1 pandemic which caused an estimated 100,000 – 200,000 deaths [34, 35].

For these reasons, influenza surveillance groups monitor for potential pandemics as
well as for circulating viruses that are particularly severe, or that are sufficiently different
to the strains in the vaccine to cause illness in vaccinated people. The Healthdirect
Australia surveillance system aims to contribute to disease surveillance by provid-
ing validated information about the community level of disease. For users to have
confidence in Healthdirect Australia data, it is important to demonstrate its ability to
accurately report on influenza-like illness. This ensures that users can trust that an in-
crease in influenza-like illness calls at Healthdirect Australia will occur simultaneously
or slightly before an increase in influenza-related activity.

We compare our data to the NSW Public Health Real-Time Emergency Department
Surveillance Surveillance System (NSW ED), the WA Emergency Department surveil-
ance (WA ED), Flutracking, the Australian GP surveillance system (called Australian
Sentinel Practices Research Network or ASPREN), and lab notifications of influenza from
the National Notifiable Diseases Surveillance System (NNDSS) both in NSW (NSW Lab)
and all of Australia (Aus Lab). We also investigate the correlation between the media
attention to the H1N1 pandemic in 2009 and the number of calls to the healthdirect
helpline. This aspect of the study was identified by the expert advisory group as being
of particular importance to describe the effect of media reporting on influenza activity.
The change in influenza activity in 2009 was very large, but it is unclear whether that
change was due to a true increase in influenza activity or due to behaviour change
influenced by media activity. Investigating this relationship is particularly important
for the healthdirect helpline, since a common criticism of the helpline data is that it varies greatly depending on marketing, media attention and other factors. Describing the relationship between media activity and influenza-related calls to the healthdirect helpline helps to determine the reliability of the Healthdirect Australia data.

This chapter addresses three related data analysis projects which all contribute to the aim of the research. The research aim is

to investigate the validity and reliability of the Healthdirect Australia data and to investigate its potential to contribute to influenza-like illness surveillance.

The three projects are:

1. identifying unusual activity and the start and finish of the influenza season.
2. comparing Healthdirect Australia data with seven other data sources using a time series analysis.
3. determining a case definition for the healthdirect helpline data by comparing the effect that different case definitions have on the correlation with other data sources.

3.2 Data Sources

The Healthdirect Australia data sources used in this section are the healthdirect helpline and the healthdirect online symptom checker. The healthdirect helpline is a telephone-based health advice service in which a registered nurse provides health advice to participants over the phone. The service has been running since July 2008. Demographic details and characteristics of symptoms are captured in a de-identified, line-listed format. From this, we can extract the number of calls with symptoms that correspond to an influenza-like illness case definition each day. For a description of the case definition for influenza-like illness and more details of how these records are extracted, see section 3.4.2.

The healthdirect online symptom checker is a tool on the healthdirect website which captures demographic details and characteristics of symptoms from participants online. The healthdirect online symptom checker has been online since November 2014. The clinical decision making algorithms are based upon clinical guidelines and are closely related to those used by the nurses on the healthdirect helpline. As with the healthdirect helpline, we can extract the number of healthdirect online symptom checker users with influenza-like illness daily. For a description of how the case definition for influenza-like illness for the healthdirect online symptom checker was determined, see section 3.4.2.

Both the healthdirect helpline and the healthdirect online symptom checker are also able report the total number of encounters each day. This means we can determine the denominator, which allows us to estimate influenza-like illness as a percentage of all encounters. This is a similar method used in the National Sentinel GP Surveillance network, ASPREN, that reports positive flu tests as a percentage of all flu tests. Using this percentage has proven to be useful particularly with the healthdirect online symptom
checker checker data due to the significant increase of its use since inception. Consequently using the percentage of visits removes the effect of the changing denominator.

Healthdirect Australia data were compared with seven different data sources, detailed in Table 3.1. The data were requested from each data source administrator, and influenza-like illness was defined according to each data source administrator, with the exception of the media data source. As a proxy for media coverage around swine flu in Australia, we calculated the number of newspaper articles published that contained the words "swine flu". This was determined by searching the media database ProQuest’s Australia and New Zealand collection, which contains all articles published by 409 papers in these two countries. Search results were further modified by restricting the document type to newspaper, the article type to "news" and searching only between 1 January 2009 - 31 December 2009. The exact search criteria were (swine flu) AND stype.exact(“Newspapers”) AND at.exact(“News”) AND pd(20090101 − 20091231). We searched for the words “swine flu” as opposed to “H1N1” due to research findings that the media coverage during that time used the words “swine flu” almost exclusively [36]. This search had a total of 7,782 articles. Dates for the pandemic phases in Australia were extracted from the Commonwealth Department of Health and Ageing report on the response to the 2009 pandemic [37].

We estimated the percentage of the total sample size captured by each data source (Table 3.1). This was estimated in different ways according to the data source. The percentage of laboratory influenza test that are captured by the NNDSS is approximately 100%, since there is mandatory reporting of all positive laboratory tests of all notifiable diseases, including influenza [38]. However, there is no requirement to report the total number of tests requested, which means that the Australian laboratory notifications data source only receives notifications of positive lab results. This represents 1 - 41% of all tests, estimated using the percentage of positive tests in the NSW laboratory during the same time period. The NSW laboratory data source has information on both how many positive tests and how many total tests were requested, allowing us to make the above estimate. The percentage of ILI visits to GPs captured by ASPREN was estimated using the number of GPs reporting to ASPREN [39] and the total number of GPs [40]. Estimates of the percentage of influenza-related visits to the emergency departments that are captured by the NSW and WA emergency department surveillance systems were calculated using the total number of records each year and the Australian Institute for Health and Wellbeing (AIHW) report [41]. Finally the percentage of media articles about H1N1 influenza or swine flu that were captured by the ProQuest database search was estimated using the number of publications in the ProQuest database (409) and the total number of publications in Australia and NZ (593) as listed by Wikipedia, accessed on 7 November, 2017.

3.3 Methods

3.3.1 Signal Identification

We investigate three different methods, one used by New South Wales Emergency department surveillance [19], one formulated by Hutwagner et al [42] and one formu-
Table 3.1: Information on data sources used in the analysis in this chapter, including their name, location, years of data, type of data reported, approximate percentage of sample reported and approximate size. ¹ type of data refers to number, percent positive (PP) or both ² details on estimates for the percentage of the sample captured by the surveillance system are in the data sources section of this chapter (Section 3.2) ³ National Notifiable Diseases Surveillance System (NNDSS)
originally formulated to identify any changes in an outcome of a continuous production process, for example to determine if a machine was producing products of the correct weight or height [44]. There are many different types of CuSum depending on the type of data. We have chosen the Poisson CuSum formulated by Lucas, which is used to detect an increase in counts [45] and is appropriate if the distribution of the counts is a Poisson distribution.

Lucas’ original formula for Poisson CuSum is [45]

\[ S_i = S_{i-1} + X_i - k \] (3.1)

This is an iterative process where the CuSum Index, \( S_i \), is updated using the CuSum index of the previous day, \( S_{i-1} \). The rest of the CuSum is calculated using the observation value, \( X_i \) (in our case the number of flu cases in the \( i \)th week) and \( k \) as a reference value. Lucas gave a formula for defining the value of \( k \). Other researchers have chosen to standardise \( X_i \), which then enables them to choose a value for \( k \) which relates to a Z-score.

The three different CuSums investigated were: A) the Public Health Real-Time Emergency Department Surveillance System (PHREDSS), which detects unusual activity in emergency department and ambulance surveillance in New South Wales; B) the Hutwagner system which detects unusual activity [42]; and C) the Cowling system which detects the start and finish of the flu season Cowling et al. [43]. These methods differ in the way that they define the reference value, \( k \).

In order to derive the formulae for these three methods, we begin with the equation above (Equation 3.1). The PHREDSS system is normalised in two different steps in order to make the index independent both of the level and variability of incidence. To see a derivation of the PHREDSS method, see Appendix A.1. As opposed to the PHREDSS index, the Hutwagner and Cowling systems first normalise \( X_i \) before calculating the index, which means that reference value \( k \) now has a concrete interpretation. For example, if \( k = 1 \) all values more than one standard deviation away from the mean CuSum change the CuSum, and all those within one standard deviation do not. All of these methods avoid the complex formula for determining \( k \) that Lucas used for his original CuSum definition which also means that the value of \( k \) does not have to change between different diseases or data sets.

The normalisation used to derive the Hutwagner and Cowling methods is demonstrated in Equation 3.2. Here, \( S_i \) is the CuSum Index and is updated using the CuSum index of the previous day, \( S_{i-1} \), the reference value, \( k \) and the normalisation of \( X_i \) as seen in Equation 3.2. The other variables in the equation are \( \bar{X}_{i} \), the expected value of \( X_i \), \( s_i \), the standard deviation of \( X \) and \( N \), the sample size. As stated above, \( \bar{X}_{i} \), \( s_i \) and \( N \) are different for the two different techniques. For a derivation of this formula and to show how it matches those given by Hutwagner et al and Cowling et al, see Appendix A.2.
Hutwagner’s formulation determines if a flu season is unusual by comparing each day’s number of flu cases with the number of flu cases on the same day in the previous seven years. However, one very unusual year, such as 2009, can cause the index to increase so dramatically that it remains high for several years. The index cannot fall back to zero because the formula is designed such that the observation needs to be one standard deviation below the average in order to bring the index back down. It is not enough for the calls to return to usual levels. In order to counteract this problem, we reset the index to zero on the summer solstice of every year (in the Southern Hemisphere, this is around December 21). We chose this date because disease outbreak data has shown that the probability of a pandemic flu outbreak in the middle of summer is very low. Furthermore, choosing the summer solstice day allows comparison between hemispheres.

Cowling’s formulation compares the day’s flu count with the counts of the previous weeks to see if the count has increased over a short period of time. This will identify the sharp increase or decrease in cases which is characteristic of the start and finish of the flu season. We compare the number of flu cases with the average number of flu cases on the same day in the previous seven weeks. We have set a proposed threshold which would indicate the start and finish of each flu season. This threshold was calculated by trialling different values and choosing the value that most closely matched the dates published by the NSW Ministry of Health for the start and finish of the flu season.

We use the percentage of all calls that are due to ILI in each method, since this provides more consistent results, and also meant the index was then independent of call volumes, allowing better comparison between years.

### 3.3.2 Case Definition

The data fields in the healthdirect helpline data that could be used to create a case definition are described in the Surveillance chapter, in Section 2.3.1. These fields are the assessment presenting problem/symptom, patient guideline and patient question yes. We determined that the patient guideline is the best choice for the case definition in Healthdirect Australia data for two main reasons. Firstly, the guideline titles used by the healthdirect helpline are similar to those used in the healthdirect online symptom checker, allowing comparison between the two data sources. Secondly, this case definition is likely to be very sensitive and capture most of the cases, though it is unlikely to be very specific. High sensitivity is important due to the role this data plays for influenza surveillance in Australia, which is to provide broad picture of the community level of influenza-like illness in Australia. Laboratories in Australia already conduct surveillance on positive laboratory tests of influenza, which provide a picture of influenza activity that is specific but not sensitive. Consequently, these data can better contribute to influenza surveillance in Australia by using a case definition with high sensitivity but low specificity.

To determine which guidelines would be included in the case definition, we used the
data sources described in section 3.2 and compared seven possible case definitions. These case definitions are in Table 3.2. For each case definition, we extracted the data using the given case definition, then compared the case definition with each data source using the Pearson’s, Spearman’s and cross correlation functions. The analysis was completed for each year individually and across a number of different comparison options unique to each data set (for example, comparing to just metropolitan Western Australia or the whole state). We calculated both of Pearson’s and Spearman’s correlation coefficients because it was unclear whether the relationship between the two data sources was linear (in which case Pearson’s correlation would be more appropriate) or monotonic (where Spearman’s would be more appropriate). We calculated both measures and reported a range of correlation coefficients. The cross-correlation function calculates Pearson’s correlation coefficient but shifted forward and backward in time by \( x \) steps. Consequently the reported lag in weeks for each data set was calculated using Pearson’s correlation coefficient, assuming a linear relationship.

The case definitions under investigation all proved to have relatively high correlations, although the poor performance of the (Fever) definition demonstrates the broad nature of the guideline titles. This is because callers who are given the guideline title Fever have a fever but also any number of other symptoms of which we have no knowledge. This is a limitation of the system which does not allow nurses to choose multiple symptoms, and consequently it is impossible to choose only the callers whose symptoms include both fever and flu-like symptoms.

The poor performance of the (Fever) case definition meant that we removed the (Fever), (Colds and Flu) OR Fever, (Cough OR Fever) and (Colds and Flu) OR Cough OR Fever definitions from consideration. Even though the (Colds and Flu) OR Cough OR Fever definition had the best or second best correlation across all data sets, we know that it contains all of the fever records which correlate poorly with ILI and are very noisy. For this reason, we have chosen the (Colds and Flu) OR Cough definition which had best or second best correlation in half the data sets.

### 3.3.3 Comparison of Data Sources

In this section we compared the healthdirect helpline data with each of the data sources described in Table 3.1. The case definition used was determined by the results of the investigation described in 3.3.2.

The data were analysed using Pearson’s correlation coefficient (similar to [1][3]), and Spearman’s rank correlation coefficient (similar to [2][5][8]) and a cross correlation function in R (version 3.4). The Person’s correlation function is appropriate for data sets that meet the following criteria: the data are either interval or ratio measure and are in related pairs, there is an absence of outliers, data is approximately normally distributed, the data sets are linearly related, and the data are homoscedastic (i.e. the variance is constant. The cross correlation function funds the Pearson’s correlation between two data sets shifted in time and consequently has the same assumptions. The Spearman’s correlation function is appropriate when the normality, linearity and homoscedasticity assumptions do not hold. The only assumptions for the Spearman’s test are that the data are ordi-
### Table 3.2: Proposed case definitions for influenza-like illness calls to the healthdirect helpline. Guideline titles in lower case are "old" guidelines from before a system change on July 1, 2015, guideline titles in UPPER CASE are "new", from after the system change. Both old and new guideline titles are included in the case definition, since the investigation includes dates from 2008 – 2017, both before and after the change.

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Guideline Titles Included</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(Colds and Flu)</em></td>
<td>Colds (Paediatric) Colds and Flu, Flu-like symptoms Colds and Flu, H1N1 Swine Flu</td>
</tr>
<tr>
<td><em>(Cough)</em></td>
<td>Cough (Adult) Cough (Paediatric) Cough INFANT Cough INFANT Cough CHIL AND TODDLER Cough INFANT</td>
</tr>
<tr>
<td><em>(Fever)</em></td>
<td>Adult Fever Fever (Paediatric) Fever ADULT AND CHILD Fever INFANT AND TODDLER Fever INFANT</td>
</tr>
<tr>
<td><em>(Colds and Flu)</em> AND <em>(Cough)</em></td>
<td>Colds (Paediatric) Colds and Flu, Flu-like symptoms Colds and Flu, H1N1 Swine Flu Colds and Flu, H1N1 Swine Flu Cough (Adult) Cough (Paediatric) Cough INFANT</td>
</tr>
<tr>
<td><em>(Colds and Flu)</em> OR <em>(Fever)</em></td>
<td>Colds (Paediatric) Colds and Flu, Flu-like symptoms Colds and Flu, H1N1 Swine Flu Adult Fever Fever (Paediatric) Fever ADULT AND CHILD Fever INFANT AND TODDLER</td>
</tr>
<tr>
<td><em>(Cough)</em> OR <em>(Fever)</em></td>
<td>Cough (Adult) Cough (Paediatric) Adult Fever Fever (Paediatric) Cough INFANT Cough INFANT FEVER ADULT AND CHILD FEVER INFANT AND TODDLER</td>
</tr>
<tr>
<td><em>(Colds and Flu)</em> OR <em>(Cough)</em> OR <em>(Fever)</em></td>
<td>Colds (Paediatric) Colds and Flu, Flu-like symptoms Colds and Flu, H1N1 Swine Flu Colds and Flu, H1N1 Swine Flu Cough (Adult) Cough (Paediatric) Adult Fever Fever (Paediatric) Cough INFANT FEVER ADULT AND CHILD FEVER INFANT AND TODDLER</td>
</tr>
</tbody>
</table>

The degree to which the data sets conform to Pearson’s assumptions varies by year.
and data set. This means that in some years or data sets, the Pearson’s test would be more appropriate but in other years the Spearman’s would be. This is clearly seen in the difference between the Pearson’s and Spearman’s correlations - if the numbers are similar than the assumptions hold, but if they are quite different then the assumptions do not hold for that year. The cross-correlation function has no reliable equivalent for non-linear but monotonic relationships, so we calculate the cross correlation coefficient, assumption the Pearson assumptions hold. Due to this limitation, we calculate all three correlation coefficients and calculate their average in each year.

The correlation was calculated for each year individually which produced three correlation coefficients for each year, and a range of correlation coefficients for each data source. For the cross correlation function, the data sources were shifted in time from $-10$ to $+10$ weeks. The time lag in weeks that produced the best correlation was recorded for each year.

### 3.4 Results

#### 3.4.1 Signal Analysis

All three index methods clearly demonstrate 2009 as an unusual year. When Hutwagner’s formula is applied to the healthdirect helpline data, the results indicate that 2009 is a very unusual year, but also shows a significant index of increase in 2012 (Figure 3.1). We propose a cut-off of 215, which is higher than the peak of all years except 2009. The cut-off of 215 would have triggered on May 27, 2009.

The other two methods are both useful for estimating the start and finish of the flu season. The PHREDDS formula shows the highest index of increase in the pandemic year, 2009 (Figure 3.1). We proposed a threshold value of 35 to determine the start and finish of the flu season. The Cowling Index shows the highest index of increase in 2015 of 524.06, though the index in 2009 is very close at 521.96 (Figure 3.1). The proposed cut-off is 270, which makes the season in 2010 very short, but keeps the season length reasonable in the other years. Dates for the start and finish of the flu season according to both the PHREDDS and Cowling indexes are in Table 3.3, along with the start and finish of the flu season calculated by the NSW emergency department.

#### 3.4.2 Case Definition

Correlation varied between case definitions and data sources, but overall trends highlighted a case definition that was most likely to correlate well across all data sources. The case definitions with the highest mean correlation were (Colds and Flu) OR (Cough) with FluTracking and NSW ED data sources and (Colds and Flu) OR (Cough) OR (Fever) with ASPREN, NSW Lab and WA ED. The highest correlation in the Aus lab data set occurred with case definition (Colds and Flu) OR (Fever). Case definitions (Fever) performed poorly across all data sets. Results are displayed in Table 3.4.
### 3.4 Results

Table 3.3: Start and finish dates for influenza-like illness for Cowling Index, PHREDDS Index and the NSW Emergency Department Surveillance System. Note that PHREDDS only began publishing the start and finish of the flu season in 2013.

<table>
<thead>
<tr>
<th>Year</th>
<th>Cowling Index with cut-off at 270</th>
<th>PHREDDS Index with cut-off at 35</th>
<th>NSW Health Index of Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>Start: May 24&lt;br&gt;Finish: July 28</td>
<td>May 1&lt;br&gt;Aug. 12&lt;br&gt;June 14</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>Start: July 6&lt;br&gt;Finish: July 24</td>
<td>Aug 8&lt;br&gt;Oct. 12&lt;br&gt;Sep. 22</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>Start: June 13&lt;br&gt;Finish: July 17</td>
<td>May 24&lt;br&gt;Sep. 4&lt;br&gt;July 3</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>Start: May 31&lt;br&gt;Finish: July 15</td>
<td>May 20&lt;br&gt;Aug. 16&lt;br&gt;July 12</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>Start: June 10&lt;br&gt;Finish: July 12</td>
<td>June 9&lt;br&gt;Sep. 20&lt;br&gt;Aug. 29</td>
<td>June 26&lt;br&gt;Sep. 17&lt;br&gt;Aug. 20</td>
</tr>
<tr>
<td>2014</td>
<td>Start: June 14&lt;br&gt;Finish: Aug. 27</td>
<td>June 6&lt;br&gt;Sep. 23&lt;br&gt;Aug. 20</td>
<td>July 1&lt;br&gt;Sep. 12&lt;br&gt;Aug. 13</td>
</tr>
<tr>
<td>2015</td>
<td>Start: May 27&lt;br&gt;Finish: Aug. 28</td>
<td>May 26&lt;br&gt;Oct. 6&lt;br&gt;Aug. 24</td>
<td>June 26&lt;br&gt;Sep. 23&lt;br&gt;Aug. 19</td>
</tr>
<tr>
<td>2016</td>
<td>Start: June 12&lt;br&gt;Finish: Sept. 5</td>
<td>June 29&lt;br&gt;Oct. 7&lt;br&gt;Aug. 20</td>
<td>June 26&lt;br&gt;Sep. 27&lt;br&gt;Aug. 28</td>
</tr>
</tbody>
</table>

3.4.3 Comparison of Data Sources

We analysed over 5 million healthdirect helpline calls for ILI between July 2008 and December 2016 and over 1 million visits to the healthdirect online symptom checker for ILI between November 2014 and December 2016. These data showed a very high correlation with other data sources for ILI. The helpline data correlated most strongly with emergency department surveillance. ASPREN and FluTracking also had high correlation with the healthdirect helpline, though correlation with laboratory data varied. The online data correlated well with the other data sources, particularly GP and Emergency department surveillance data (Figure 3.3 and Table 3.5).
Figure 3.1: Moving 7-day average of Hutwagner, Cowling and PHREDSS Indices operating on the percentage of ILI calls as a proportion of all calls to the healthdirect helpline. The intersection of the index (black) with the cut-off (blue) indicates the start and finish of the flu season for the Cowling and PHREDSS indices, and a threshold for signalling unusual behaviour for the Hutwagner index.
### §3.4 Results

The years that produced the highest and lowest correlation coefficients in the healthdirect helpline data varied, though 2013 had consistently low correlation with all data sources. The mode lag between healthdirect helpline data and other surveillance data was 0 weeks for every data source except the Australian lab data. The lag in the healthdirect online symptom checker data was zero weeks except for the two lab data sources (Table 3.5).

The number of media articles published had three distinct peaks on April 30, May 30 and June 9. The number of calls received also had three distinct peaks on 29 April, 15 June and 20 July (Table 3.6). Figure 3.4 shows the percentage of calls, the number of articles published and the phases of the pandemic.

<table>
<thead>
<tr>
<th></th>
<th>(Colds and Flu)</th>
<th>(Cough)</th>
<th>(Fever)</th>
<th>(Colds and Flu) OR (Cough)</th>
<th>(Colds and Flu) OR (Fever)</th>
<th>(Cough) OR (Fever)</th>
<th>(Colds and Flu) OR (Cough) OR (Fever)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASPREN</strong></td>
<td>0.81 (0.64 – 0.94)</td>
<td>0.84 (0.65 – 0.94)</td>
<td>0.18 (0.66 – 0.95)</td>
<td>0.85 (0.66 – 0.95)</td>
<td>0.84 (0.69 – 0.93)</td>
<td>0.87 (0.69 – 0.95)</td>
<td>0.87 (0.69 – 0.95)</td>
</tr>
<tr>
<td><strong>Aus Lab</strong></td>
<td>0.77 (0.51 – 0.92)</td>
<td>0.77 (0.47 – 0.94)</td>
<td>0.40 (0.02 – 0.95)</td>
<td>0.79 (0.49 – 0.93)</td>
<td>0.82 (0.60 – 0.94)</td>
<td>0.81 (0.57 – 0.95)</td>
<td>0.81 (0.55 – 0.95)</td>
</tr>
<tr>
<td><strong>Flu-Tracker</strong></td>
<td>0.79 (0.64 – 0.90)</td>
<td>0.77 (0.58 – 0.89)</td>
<td>0.38 (−0.08 – 0.81)</td>
<td>0.79 (0.62 – 0.89)</td>
<td>0.72 (0.49 – 0.86)</td>
<td>0.73 (0.51 – 0.86)</td>
<td>0.76 (0.57 – 0.87)</td>
</tr>
<tr>
<td><strong>NSW ED</strong></td>
<td>0.92 (0.86 – 0.97)</td>
<td>0.96 (0.92 – 0.98)</td>
<td>0.09 (−0.29 – 0.90)</td>
<td>0.96 (0.92 – 0.99)</td>
<td>0.87 (0.77 – 0.93)</td>
<td>0.91 (0.84 – 0.96)</td>
<td>0.95 (0.92 – 0.98)</td>
</tr>
<tr>
<td><strong>NSW Lab</strong></td>
<td>0.78 (0.60 – 0.94)</td>
<td>0.77 (0.56 – 0.94)</td>
<td>0.39 (−0.10 – 0.91)</td>
<td>0.79 (0.58 – 0.95)</td>
<td>0.81 (0.54 – 0.95)</td>
<td>0.80 (0.57 – 0.95)</td>
<td>0.81 (0.61 – 0.96)</td>
</tr>
<tr>
<td><strong>WA ED</strong></td>
<td>0.86 (0.67 – 0.94)</td>
<td>0.91 (0.80 – 0.97)</td>
<td>0.25 (−0.25 – 0.87)</td>
<td>0.91 (0.82 – 0.96)</td>
<td>0.83 (0.63 – 0.95)</td>
<td>0.89 (0.73 – 0.96)</td>
<td>0.92 (0.78 – 0.97)</td>
</tr>
</tbody>
</table>

Table 3.4: Correlations for each of the seven case definitions (columns) and six data sources (rows). Correlations calculated were pearson’s, spearman’s and the cross correlation function for each year individually. Results are displayed as the mean correlation and the range of correlations. The case definition with the highest mean and range are coloured in orange and the second highest mean and range are shaded in light orange.

The number of calls received also had three distinct peaks on 29 April, 15 June and 20 July (Table 3.6). Figure 3.4 shows the percentage of calls, the number of articles published and the phases of the pandemic.
Data Analysis: Influenza-like illness using telephone and internet data

Figure 3.2: Percentage of influenza-like illness in the New South Wales Emergency Department surveillance system and the healthdirect helpline from 2008 – 2016. This visually demonstrates the excellent correlation between the two data sources.

### 3.5 Discussion

#### 3.5.1 Signal Analysis

The three different signal analysis methods vary in their effectiveness for surveillance. The Hutwagner index clearly highlights a year with unusually high rates of influenza-like illness calls. On the other hand, the Cowling and PHREDDS indices give an indication of the start and finish of the flu season, but it is difficult to determine the accuracy of this information.

The Hutwagner index clearly demonstrates that 2009 is an unusual year, with 2012, 2015 and 2016 showing slightly elevated levels. Setting the alarm threshold at just above the maximum index of 2012 means that the alarm would have triggered on May 27, 2009. This would have been five days after Australia had instigated the “contain” phase, and seven days after the second case was confirmed in Australia. In the case of 2009, public health departments were alerted to the possibility of a pandemic due to activity overseas so it is not possible to determine whether the alert would have given early warning of unusual activity.

The Cowling and PHREDDS indices give an indication of the start and finish of the flu season and we chose cut-offs such that the dates determined were relatively comparable to start and finish dates published by the NSW department of health. Even though these dates seem somewhat arbitrary, they are used for public health
Table 3.5: Summary of the correlation between healthdirect helpline and healthdirect online symptom checker data and other data sources, including the range in yearly correlations, the mean correlation, mode of the lag between data sources, and the years of the highest and lowest correlations. Correlations calculated were pearson’s, spearman’s and the cross correlation function for each year individually.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Range</th>
<th>Mean</th>
<th>Mode of weeks lag</th>
<th>Year of highest</th>
<th>Year of lowest</th>
<th>corr. in 2015</th>
<th>corr in 2015</th>
<th>weeks lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASPREN</td>
<td>0.65 – 0.95</td>
<td>0.85</td>
<td>0</td>
<td>2015</td>
<td>2013</td>
<td>0.95</td>
<td>0.95</td>
<td>0</td>
</tr>
<tr>
<td>Aus Lab</td>
<td>0.49 – 0.93</td>
<td>0.79</td>
<td>1</td>
<td>2009</td>
<td>2013</td>
<td>0.90</td>
<td>0.92</td>
<td>1</td>
</tr>
<tr>
<td>Flu-Tracker</td>
<td>0.62 – 0.89</td>
<td>0.79</td>
<td>0</td>
<td>2010</td>
<td>2011</td>
<td>0.74</td>
<td>0.77</td>
<td>0</td>
</tr>
<tr>
<td>NSW ED</td>
<td>0.92 – 0.99</td>
<td>0.96</td>
<td>0</td>
<td>2014</td>
<td>2009</td>
<td>0.97</td>
<td>0.97</td>
<td>0</td>
</tr>
<tr>
<td>NSW Lab</td>
<td>0.58 – 0.95</td>
<td>0.79</td>
<td>both 0 and 1</td>
<td>2012</td>
<td>2013</td>
<td>0.80</td>
<td>0.87</td>
<td>1</td>
</tr>
<tr>
<td>WA ED</td>
<td>0.82 – 0.96</td>
<td>0.91</td>
<td>0</td>
<td>2009</td>
<td>2013</td>
<td>0.95</td>
<td>0.95</td>
<td>0</td>
</tr>
</tbody>
</table>

In summary, all three indices function as expected, and both the Hutwagner and Cowling indices fulfil one of the purposes identified in the introduction to this chapter. The two purposes of the indices were: to identify any unusual activity; and to predict the start and finish of the flu season. The Hutwagner index can identify unusual activity, through retrospectively identifying the start of the 2009 influenza pandemic. The public health utility of this index is clear, especially if it allows a public health response to increased influenza activity to occur earlier than otherwise. The Cowling and PHREDDS indices both predict the start and finish of the flu season, but the Cowling index is more preparedness, including staffing and messaging, particularly in New South Wales. However, it is concerning that the PHREDDS Index relates only to the previous year, not the previous seven years. Any unusual activity in one year will have a large effect on the PHREDDS Index the next year. This is seen in 2010 where the the start of the flu season according to the PHREDDS index was predicted to be August 8, 2010 even though there had been a period of high flu activity since the beginning of June. This is because 2010 is compared to the same time in 2009, which had unusually high and unusually early influenza activity. The PHREDDS formula could be greatly improved by using an average mean and standard deviation of all years’ data excluding the 2009 pandemic year to standardise.
Figure 3.3: Correlation between healthdirect helpline data and six other data sources: ASPREN, Aus Lab, FluTracking, NSW ED, NSW Lab and WA ED. Spearman’s, Pearson’s and cross correlation function were calculated for each year. The box chart shows the range across the years and three different types of correlation function.

reliable than the PHREDDS since it takes more years’ worth of data into account. The PHREDDS index is not as reliable as the Cowling and fills the same role, so it is the least useful. Consequently, the Cowling Index could be used at Healthdirect Australia to help predict call volumes, as well as helping public health departments to prepare for the influenza season. For these reasons, both the Hutwager Index and the Cowling Index contribute to the public health to both pandemic and annual influenza and will be used for influenza surveillance at Healthdirect Australia.

3.5.2 Comparison of Data Sources

The results of the correlation investigation indicate that both the healthdirect helpline and healthdirect online symptom checker ILI data correlate very well with other ILI data, particularly that from emergency department surveillance.

Surveillance systems that had a syndromic definition (EDs, GPs and FluTracking) had the best correlation with both Healthdirect Australia data sources. The lower correlation
Number of articles published, percentage of ILI calls to the Healthdirect helpline and phases of public health intervention in 2009

Figure 3.4: Comparison between the phases of pandemic action declared by the Australian Government, the number of news articles published and the percentage of influenza-like illness related calls to the healthdirect helpline. The number of articles and percentage of calls are presented here as a moving 7-day average in order to smooth the effect of day of the week, meaning that the exact dates and numbers reported in Table 3.6 are not visible on this plot.

with FluTracking and GPs compared to EDs could be due to the smaller sample sizes of these data sources. The healthdirect online symptom checker and the phone data have similar correlations to the six other data sets in 2015 and 2016. However, as there are only two years of online data, correlation with further years would need to be investigated before drawing strong conclusions for this data source.

The year of best correlation occurred across five different years, and the worst correlation occurred in 2013 for four out of six data sources. This could be due to the nature of the correlation function, which is strongly influenced by its tails. For example, 2009 provided the best fit in two data sources but the worst fit in one. Data sources matched at the peak in 2009, which is significantly larger than the mean and disproportionately increased or decreased the estimate of correlation. By contrast, the helpline data for 2013 had a flat or double peak, (Figure 3.2) which would lead to lower correlation due to the nature of the correlation function. This flat peak was seen across all data sources, as is
<table>
<thead>
<tr>
<th>Events</th>
<th>Date</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>First reports from Mexico</td>
<td>24 April</td>
<td></td>
</tr>
<tr>
<td>First article published in Australia</td>
<td>25 April</td>
<td></td>
</tr>
<tr>
<td>Delay phase in Australia initiated</td>
<td>27 April</td>
<td></td>
</tr>
<tr>
<td>WHO pandemic phase 4 initiated</td>
<td>27 April</td>
<td></td>
</tr>
<tr>
<td>WHO Pandemic Phase 5 initiated</td>
<td>29 April</td>
<td></td>
</tr>
<tr>
<td>First peak in calls</td>
<td>27 April</td>
<td>10.4% of all calls were influenza related</td>
</tr>
<tr>
<td>First peak in media</td>
<td>30 April</td>
<td>201 articles published</td>
</tr>
<tr>
<td>Thermal imaging starts in airports</td>
<td>30 April</td>
<td></td>
</tr>
<tr>
<td>First case in Australia (Queensland)</td>
<td>8 May</td>
<td></td>
</tr>
<tr>
<td>Second case in Australia (New South Wales)</td>
<td>20 May</td>
<td></td>
</tr>
<tr>
<td>Contain phase in Australia initiated</td>
<td>22 May</td>
<td></td>
</tr>
<tr>
<td>Second peak in media</td>
<td>30 May</td>
<td>139 articles published</td>
</tr>
<tr>
<td>Third peak in media</td>
<td>9 June</td>
<td>152 articles published</td>
</tr>
<tr>
<td>WHO pandemic phase 6 initiated</td>
<td>11 June</td>
<td></td>
</tr>
<tr>
<td>Second peak in calls</td>
<td>14 June</td>
<td>33.1% of all calls were influenza related</td>
</tr>
<tr>
<td>Protect phase in Australia initiated</td>
<td>17 June</td>
<td></td>
</tr>
<tr>
<td>First death in Australia</td>
<td>19 June</td>
<td></td>
</tr>
<tr>
<td>Third peak in calls</td>
<td>19 July</td>
<td>28.8% of all calls were influenza related</td>
</tr>
<tr>
<td>WHO lower pandemic alert</td>
<td>10 August</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.6: Summary of relevant events during the 2009 A(H1N1) pandemic. The green are events related to media coverage of the pandemic, and the peaks in the number of articles published about the pandemic. The orange events pertain to the pandemic phases declared by the Australian government. The blue events are those pertaining to the influenza-like illness related calls to the healthdirect helpline.

demonstrated in Figure 3.2 by the NSW Emergency Department data source.

A lag of zero weeks for all data sources except the laboratories shows that an increase in healthdirect helpline calls or healthdirect online symptom checker visits occurs at the same time as an increase in visits to the ED, GP and self-reported ILI on FluTracking. On the other hand, the increase in calls to the healthdirect helpline occurs one week earlier...
than a positive influenza test, which is what we would expect given the additional time it takes to run the test for influenza in the lab.

The comparison between the number of news articles published and the number of calls to the healthdirect helpline helps to quantify the effect of the media on people’s opinions and actions. Forty percent of the American public follow health-related news, and news related to influenza is one of the most followed [46]. There was much speculation surrounding the media response to the 2009 pandemic, and researchers have questioned whether the media response to the pandemic was more “hype” than factual reporting. Several papers have looked at a wide variety of media sources, including TV news stories in Australia [36], Australian newspapers [47], UK newspapers [48], European newspapers [49], and a systematic review which looked at newspapers, radio, TV, online, twitter, facebook, youtube and blogs and [50]. These papers all sought to examine the type of language used while reporting the H1N1 pandemic, and whether that affected the public’s response to the pandemic. Other papers have shown that anxiety increased during the pandemic by looking at a cohort of university students in Stanford [22], in UNSW [23] or in the public in Sydney [24]. It is clear that influenza and health-related news are issues of concern for many Australians, and this research has demonstrated the effect of that concern on the health care system.

3.6 Limitations

The limitations of this study encompass data limitations and assumptions of the statistical tests. As mentioned in Chapter 2, the data set is very coarse, and the only information about a person’s symptoms is their main complaint or reason for calling the helpline. This could lead to misclassification both of not identifying callers with ILI and people without ILI being classed as cases. However, we assumed that these types of misclassification were equally likely to happen, and in a data set as large as the healthdirect helpline, would cancel each other out.

There were also limitations in the comparison data sets which led to several assumptions. These limitations included data sets changing size with the addition of reporting sites or people, data sets without age data causing an inability to age-standardise, and each data set aggregating data with different temporal and geographical cells. To take changing data sizes into account, we compared the percentage of ILI related calls instead of the raw number and for the different temporal and geographical aggregations, and we matched the healthdirect helpline data with the temporal and geographical aspects of each data set as closely as possible. As with the previous limitation, we assumed that the differences in data sets would have only a small effect on the correlation due to the size of the data.

The inability to age-standardise some data sets led to two further assumptions: 1) that data sets had the same age profile or 2) that the shape of the ILI incidence does not change with age in each data set. The healthdirect helpline data set has a greater proportion of callers under 5 than in the Australian population, so it is unlikely the age profile of the data sets matched. However, the shape of ILI incidence in the healthdirect helpline data
Data Analysis: Influenza-like illness using telephone and internet data

does not change with age, except in 2009. Even though the incidence of ILI is different in each age group, we assumed that each age group has the same pattern of incidence across the year, or that it is a scalar transformation of the same function. Consequently, a different age profile will also be a scalar transformation of the same function. Both the Spearman’s and Pearson’s statistical tests are invariant under transformations of scale and compare the shape of ILI incidence in two data sets. In other words, if the pattern of ILI incidence is the same across age groups then a different age profile will not affect the Spearman’s and Pearson’s correlation coefficients.

3.7 Public Health Implications

We propose presenting the healthdirect helpline and healthdirect online symptom checker data from the syndromic surveillance system in a way that is accessible to local health departments, jurisdictions, national surveillance groups such as OzFoodNet, universities and other research or surveillance institutions. This will increase the provision of more accurate surveillance data that can help public health departments manage surge capacity, workforce, immunisations, public health messaging and other public health related activities.

These results will be displayed on a secure surveillance dashboard, enabling public health professionals to view data, download plots and tables for insertion into reports, and receive alerts when syndromes reach a threshold. Data will include information about the incidence of disease, the severity, temporal and spatial variation and any unusual activity alerts that could be an indicator of an outbreak.

This is the first time that a national telephone health line in Australia has reported on its contribution to early detection of ILI when compared with emergency departments and sentinel GP surveillance. Other studies comparing telephone triage data to lab confirmed influenza have shown similar results in the USA [1, 4, 7] and Australia [8] or comparing Emergency Department surveillance in Canada [5, 6] and the USA [4, 7]. There have also been studies comparing telephone triage data to GP Presentations in Ireland [2], the USA [7] and the UK [3]. All studies reported high levels of correlation between telephone triage data and a more conventional data source, with the telephone triage data leading by at least one week.

The high degree of correlation between the HDA data and other data sources show that the HDA data are consistent and trustworthy. We conclude that the Healthdirect Australia data can contribute to both routine disease surveillance and event based surveillance of influenza-like illness in Australia.

3.8 Lessons Learnt

This was a large piece of work that was the focus of my first year as an MAE student, and there were many lessons learnt over the year. I used one of the lessons in the teaching block in second year as a case study in double-checking your data before starting analysis. However the two major lessons were 1) what does “best” mean and 2)
having more data sources does not necessarily improve prediction.

I began this project with the aim of finding the best case definition and the best algorithm for signal identification. However I soon discovered that this was poorly defined and, especially for signal identification, unnecessary. There were many reasons that this problem was poorly defined: there was no clear gold standard with which to compare, there were multiple years each of which had its own best fit and there were many aspects to consider. For example with the case definition, the best definition could defined as the one which produces the highest correlation or the highest mean correlation or the smallest range in correlations or any number of other indicators. Secondly, since there was no gold standard and case definitions performed differently across the data sources, it was unclear how to weight the evidence given by different data sources. For example, do you choose the definition that performed well in the majority of data sources or do you weight the evidence by some factor such as the size of the data source.

In the case of the signal identification, there are so many different methods with various strengths and weaknesses that investigating all of them and evaluating which was the best was a huge task. Furthermore, it was a task that was unnecessary: for public health action the best method is not required, but rather a method that does the job. In fact, a simpler solution is often better so that your work can be broadly understood by all interested parties and also used or built upon by others.

The second lesson that I learnt while researching this chapter was that more data did not necessarily make the research any better or the conclusions any stronger. In fact, in the case definition and comparison between data sources investigation, more data made the investigation more convoluted and contributed to the problem above of how to define "best." Each data source I received had its own set of internal qualifiers, which we had to take into account in the analysis. Instead of being able to add each data set to the existing analysis, changes had to be made in order to take the internal qualifiers into account. Consequently each data set required analysis built specifically instead of being able to use the analysis already completed. For example, the definition of a week changed between data sets - some had week ending Friday, some week ending Sunday and one changed from Friday to Sunday halfway through. Similarly, some data sets were state-specific and some Australia-wide meaning that each data set required a slightly different query of the healthdirect helpline data in order to match as closely as possible.

These lessons are all artefacts of working with several very large data sets where you have the luxury of picking and choosing ways to cut the data or different kinds of indicators. However, with the push towards bigger and more detailed data sets, these lessons will stand me in good stead for further work in public health.
Chapter 4

Epidemiology Chapter: Investigating the utility of conducting surveillance on a telephone helpline for six syndromes

4.1 Introduction

This chapter expands on the work on syndromic surveillance for influenza-like illness discussed in Chapters 3 and 4. We identified 6 other syndromes and investigated the utility of continued surveillance of these. The syndromes are acute mental health events (AMHE), suicide ideation, adverse event following immunisation (AEFI), alcohol or drug related health events (ADRE) and gastroenteritis. For each, we created a case definition, investigated syndrome incidence between 2008 and 2017, and commented on the viability and utility of ongoing monitoring of the syndrome.

There are several telephone surveillance systems in the US [1, 4], the UK [3, 51], Sweden [18], Canada [6, 52] and Ireland [2] that monitor influenza-like illness, but few that monitor trends of other syndromes. Both the UK [51] and Canada [6] have proposed conducting surveillance for up to nine other syndromes or syndrome categories. The reasons for conducting this surveillance is both to provide early warning of outbreaks and to monitor ongoing trends of illness. There has been particular interest in monitoring syndromic gastroenteritis because of the large impact that both viral and bacterial gastroenteritis has on the community.

There are varying degrees of surveillance carried out for each of these syndromes. Gastroenteritis has the most detailed and established surveillance systems, with laboratory surveillance being conducted on a number of pathogens that cause gastroenteritis being on the Nationally Notifiable Diseases Surveillance System, for example salmonellosis, listeriosis and shiga toxin-producing E.Coli. There is also a provision for gastroenteritis to be notified locally under syndromes such as: gastroenteritis in an institution; foodborne or waterborne illness in two or more related cases; food poisoning; and gastrointestinal illness cluster [53]. There is also surveillance being undertaken by the NSW Emergency Departments, the WA Emergency Department and the GP sentinel surveillance system, ASPREN. The OzFoodNet group collects and summarises information on gastroenteritis in Australia.
Surveillance of adverse events following immunisation is completed through both the Office of Product Review and the Therapeutic Goods Administration. Surveillance is passive, available through the database of adverse event notifications and is reviewed by the Advisory Committee on the Safety of Medicines. Since 2003, an annual report has been prepared by the National Centre for Immunisation Research and Surveillance in conjunction with the Therapeutic Goods Administration and published by the Communicable Diseases Intelligence [54]. Some researchers have investigated the effectiveness of continuous active surveillance of adverse events following immunisation using other technologies such as SMS [55] or online surveys such as VaxTracker [56] but this is not systematic nor widespread.

Surveillance for the remaining three syndromes of suicide ideation, acute mental health events and alcohol and drug related events is more limited. The NSW Health department monitors visits to the emergency department due to these syndromes through the Public Health Rapid, Emergency, Disease and Syndromic Surveillance system. The Australian Paediatric Surveillance unit conducts surveillance on child mental health problems. There are support agencies such as Lifeline, Beyond Blue and the Alcohol and Drug Foundation that collect limited information about these syndromes through keeping records of the people who access their programs. However, this surveillance is not systematic, is limited in scope and is not nationally analysed or reported.

Due to the extensive volume of calls for gastroenteritis we conducted further research to assess the plausibility of conducting surveillance for gastroenteritis using healthdirect helpline data.

Gastroenteritis is estimated to affect between 19% – 28% of the population annually in developed countries such as the UK [57] and the Netherlands [58, 59]. Some studies have found much higher rates of illness, though this could be due to differences in syndrome definition [58, 60, 61]. This rate appears to have been decreasing in the past 30 years, with older studies showing a much higher prevalence of gastroenteritis diseases in the community [60, 61] and hospitalisations in the US for gastroenteritis having decreased by 20% between 1979 to 1995 [62]. Mortality due to gastroenteritis is rare in developed countries, but diarrhoeal disease remains the second largest cause of mortality in children under the age of five in low and middle income countries [63] and is also associated with reduced growth, reduced cognitive development and reduced vaccine efficacy in these children [64].

The epidemiology of gastroenteritis is a topic of extensive research. In the developed world, viruses are the most common cause of gastroenteritis and Norwalk-like illness the most common identified cause of gastroenteritis [59, 62]. Evidence on the seasonality of gastroenteritis varies by pathogen and country, and researchers often focus on the epidemiology of particular gastrointestinal illnesses. The Tecumseh study conducted in the United States found that 36.6% of gastroenteritis occurs in winter, 28.3% in autumn and only 35% in spring and summer combined [60]. Furthermore there is some evidence that viral illnesses, including Norwalk-like illness, are more likely to occur in the colder, drier months [62, 65]. On the other hand, other investigations have shown that peak activity of rotavirus, another common viral gastrointestinal illness, varies according
to location \[66\] and a study conducted in Australia found that the risk of contracting infectious gastroenteritis was highest in the summer months \[67\].

There is some evidence that gastroenteritis calls to a telephone helpline correlates with actual gastroenteritis activity in GPs, emergency departments and the community. A study in Ontario found a high degree of correlation between telephone helpline data and emergency department visits \[68\]. Similarly, researchers in the Baltimore-Washington area linked calls to a nurse health advice line with visits to a GP clinic \[69\]. They were then able to calculate the sensitivity, specificity and positive predictive value of a gastroenteritis syndrome in their telephone data, using the GP visits as a gold standard. Gastroenteritis was found to have a high sensitivity and specificity. Finally, researchers in the UK found that they were able to map the movement of gastroenteritis illness geographically using telephone data \[17\].

Several researchers have also conducted retrospective studies into the suitability of telephone helpline data in their country for the surveillance of gastroenteritis outbreaks, with varying results. A study in Sweden \[18\] demonstrated that the telephone data retrospectively identified four out of nine known water-borne diarrhoeal outbreaks in Sweden. These outbreaks were all very large with more than 350 known cases, with an the attack rate in the affected county of at least 30 per 100,000. By contrast, a retrospective study in the United Kingdom concluded that telephone data were not able to detect known outbreaks of cryptosporidiosis \[70\]. However, a later UK study on the same data concluded that they were able to give 4 weeks’ early warning of the start of norovirus activity in hospitals and the community \[71\].

In summary, this chapter presents some preliminary results of surveillance through telephone data of 6 syndromes. For the gastroenteritis syndrome, we present results about the seasonality of gastroenteritis calls, the degree of matching between telephone data and other data sources and also the ability of the telephone data to detect outbreaks in small geographical regions.

### 4.2 My Role

I completed all aspects of this project for the mental health events, suicide ideation, adverse event following immunisation, alcohol or drug related health events and gastroenteritis syndromes. This included developing a research question, creating a data analysis plan, analysing and interpreting data and writing the final report.

### 4.3 Data Sources

The data source used in this section is the healthdirect helpline. This is the same data source that was used in Chapters 2 and 3. This section is a short summary of the information presented in the Data Sources section of each of those chapters.

The nurse triage helpline is a telephone-based health advice service in which a registered nurse provides health advice to participants over the phone. The service has
been running since July 2008. Demographic details and characteristics of symptoms are captured in a de-identified, line-listed format. There are three fields that have information about the symptoms of callers. These are assessment presenting problem/symptom, patient guideline and patient question yes. We chose to use the patient guideline field for constructing case definitions. The patient guideline field is the symptom which most closely matches the caller’s complaint, selected from a pre-determined list. Once chosen, the patient guideline then corresponds to a set of questions about the symptom and the answers to these questions will determine the final advice given. For a more detailed description, examples of these three fields, and justification for using the patient guideline see section 2.2.

Using the patient guideline field, we can extract the number of calls each day with symptoms that correspond to each syndrome and the total number of calls. This allows us to estimate the incidence of each syndrome as a percentage of all calls. This is similar to surveillance systems such as the National Sentinel GP Surveillance network, ASPREN, that reports the percent of flu tests that are positive (percent positive or PP).

We use also use 3 different data sources in this chapter to test the case definition for the gastroenteritis syndrome. These are the Australian Sentinel Practices Research Network (ASPREN), the NSW Emergency Department (NSW ED) and the WA Emergency Department (WA ED) (Table 4.1. This was only possible for the gastroenteritis syndrome, since we did not have comparison data sources for the other syndromes investigated in this chapter. Each data source was requested from the relevant data source administrator. For each comparison data source, we adopted the case definition for gastroenteritis used by that source’s administrator or research team.

We estimated the percentage of the total sample size captured by each data source (Table 4.1). This was estimated in different ways according to the data source. The percentage of gastroenteritis-related visits to GPs captured by ASPREN was estimated using the number of GPs reporting to ASPREN [39] and the total number of GPs [40]. Estimates of the percentage of gastroenteritis-related visits to the emergency departments that are captured by the NSW and WA emergency department surveillance systems were calculated using the total number of records each year and the Australian Institute for Health and Wellbeing (AIHW) report [41].

4.4 Methods

4.4.1 Case Definitions

In this section we discuss the case definitions for gastroenteritis, suicide ideation, acute mental health events, drug and alcohol events and adverse event following immunisation. All case definitions except for gastroenteritis were formulated by selecting all guideline titles related to the syndrome, and then checking for any inconsistencies. In particular, in July of 2015 there was a systematic change of the healthdirect helpline service, and all the question flows used by the nurses were evaluated for refinements. Guideline titles changed during this time as well to match the new question flows, so all case definitions were monitored for inconsistencies in activity in between June and August of 2015. Case
Constructing the case definition for the *gastroenteritis* syndrome was more detailed and followed a method similar to that described in the Data Analysis Chapter (Section 3.3.2). As discussed above, this was only possible for the gastroenteritis syndrome since this was the only syndrome for which we were able to obtain comparison data. We compared the incidence of gastroenteritis-related calls to the helpline with gastroenteritis-related visits to the NSW Emergency Department, WA Emergency Department and the sentinel GP surveillance system, ASPREN and tested four possible symptoms: “abdominal pain”, “diarrhoea”, “constipation”, “vomiting” as shown in the first four rows of Table 4.2. These symptoms were compared individually to the three comparison data sets using the Spearman’s, Pearson’s and cross correlation coefficients. After those with poor correlation were removed, combinations of symptoms were tested for improved fit compared to single symptoms using the same correlation functions. We defined a poor correlation as the average correlation lying between $-0.3$ and $0.3$ in most data sets. For more information about this method of determining a case definition, see section 3.3.3. We also investigated whether we could detect outbreaks on a smaller geographical scale, whether gastroenteritis had a seasonal pattern, and the strengths and weaknesses of monitoring both the total volume of gastroenteritis calls and the gastroenteritis calls as a percentage of all calls.
### Table 4.2: Proposed case definitions for gastroenteritis calls to the helpline.

Guidelines in lower case are "old" guidelines from before a system change on July 1, 2015, guidelines in upper case are "new", from after the system change.

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Guideline Titles Included</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abdominal Pain</strong></td>
<td>Abdominal Pain, Pregnant &gt;20 Weeks</td>
</tr>
<tr>
<td></td>
<td>Abdominal Pain - Female (Paediatric)</td>
</tr>
<tr>
<td></td>
<td>Abdominal Pain - Male (Paediatric)</td>
</tr>
<tr>
<td></td>
<td>Abdominal Pain / Discomfort</td>
</tr>
<tr>
<td><strong>Diarrhoea</strong></td>
<td>Diarrhoea (Paediatric)</td>
</tr>
<tr>
<td></td>
<td>Diarrhoea / Change in Bowel Habits</td>
</tr>
<tr>
<td><strong>Constipation</strong></td>
<td>Constipation (Paediatric)</td>
</tr>
<tr>
<td></td>
<td>Constipation / Rectal Symptoms</td>
</tr>
<tr>
<td><strong>Vomiting</strong></td>
<td>Vomiting (Paediatric)</td>
</tr>
<tr>
<td></td>
<td>Vomiting of Blood (Paediatric)</td>
</tr>
<tr>
<td></td>
<td>Nausea / Vomiting</td>
</tr>
<tr>
<td><strong>Diarrhoea OR Vomiting</strong></td>
<td>Diarrhoea (Paediatric)</td>
</tr>
<tr>
<td></td>
<td>Diarrhoea / Change in Bowel Habits</td>
</tr>
<tr>
<td></td>
<td>Vomiting (Paediatric)</td>
</tr>
<tr>
<td></td>
<td>Vomiting of Blood (Paediatric)</td>
</tr>
<tr>
<td></td>
<td>Vomiting - Paediatric</td>
</tr>
<tr>
<td></td>
<td>Nausea / Vomiting</td>
</tr>
</tbody>
</table>

### 4.4.2 Spatial Investigation

In order to determine if we could detect geographic variations in gastroenteritis activity, we chose a 12-week period in time which had the highest recorded percentages of gastroenteritis activity in Australia. With higher percentages of gastroenteritis-related calls overall, there was a better chance of having enough people in each geographical cell to see patterns in activity. To conduct the spatial investigation, we used the postcode field in the healthdirect helpline data and then mapped that to the Australian Bureau of Statistics’ (ABS) “Statistical Local Area” or SLA. If a postcode mapped to several SLAs, we chose the SLA that had the greatest percentage by area of a postcode. The mapping of postcode to SLA was downloaded from the ABS website. Maps were drawn in R, using a shapefile (the geographical data) downloaded from www.gadm.org. We used the Level
### 4.5 Results

#### 4.5.1 Gastroenteritis

**Case Definition**

The symptoms *Abdominal Pain* and *Constipation* matched poorly with all three comparison data sets. The average correlation was between \(-0.3\) and \(0.3\) in 2 out of 3 data sets for abdominal pain and all 3 data sets for constipation (Table 4.4). The average correlation for the symptoms of diarrhoea and vomiting was above \(0.3\), so we also considered a further case definition of *Diarrhoea OR Vomiting*. The analysis showed that the combined definition had the best average correlation with NSW Emergency Department data and the second best average correlation with the ASPREN and WA Emergency Department data.

There were three case definitions that performed very similarly in all three data sets. These case definitions were *Diarrhoea, Vomiting* and *Diarrhoea OR Vomiting*. Although *Diarrhoea* had the best fit in both the ASPREN and WA Emergency Department data, *Diarrhoea OR Vomiting* had the best fit in the NSW Emergency Department Data, which

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Guideline Titles Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>acute mental health events</td>
<td>Anxiety: Severe / Panic&lt;br&gt;Anxiety: Mild to Moderate&lt;br&gt;Depression / Mood Disorders&lt;br&gt;Depression: Post natal&lt;br&gt;Confusion - Delirium (Paediatric)&lt;br&gt;Psychosis&lt;br&gt;Aggression and Behavioural Disturbance</td>
</tr>
<tr>
<td>suicide ideation</td>
<td>Suicide Concerns (Paediatric)&lt;br&gt;Suicide and / or Homicidal Behaviour</td>
</tr>
<tr>
<td>adverse event following immunisation</td>
<td>Immunisation Reactions (Paediatric)&lt;br&gt;Immunisation Reactions H1N1&lt;br&gt;Immunisation Reactions H1N1 [Paediatric]</td>
</tr>
<tr>
<td>alcohol or drug related events</td>
<td>Acute Alcohol Intoxication&lt;br&gt;Withdrawal Symptoms&lt;br&gt;Substance Abuse: Diagnosed / Suspected&lt;br&gt;Substance Abuse (Paediatric)</td>
</tr>
</tbody>
</table>

Table 4.3: Proposed case definitions for *acute mental health events, suicide ideation, adverse event following immunisation* and *alcohol or drug related events*. Guidelines in lower case are "old" guidelines from before a system change on July 1, 2015, guidelines in upper case are "new", from after the system change.

2 data in the R (SpatialPolygonsDataFrame) format.
Correlation Mean and (Range) for each case definition and data source

<table>
<thead>
<tr>
<th></th>
<th>Abdominal Pain</th>
<th>Constipation</th>
<th>Diarrhoea</th>
<th>Vomiting</th>
<th>Diarrhoea OR Vomiting</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS-PREN</td>
<td>0.24 (−0.21–0.48)</td>
<td>0.13 (−0.22–0.51)</td>
<td>0.42 (0.11 − 0.76)</td>
<td>0.32 (0.14 − 0.75)</td>
<td>0.38 (0 − 0.77)</td>
</tr>
<tr>
<td>NSW ED</td>
<td>0.37 (0 − 0.72)</td>
<td>0.07 (−0.35–0.42)</td>
<td>0.66 (0.34 − 0.85)</td>
<td>0.68 (0.32 − 0.85)</td>
<td>0.74 (0.50 − 0.86)</td>
</tr>
<tr>
<td>WA ED</td>
<td>0.27 (−0.32–0.71)</td>
<td>0.18 (−0.19–0.49)</td>
<td>0.55 (0.19 − 0.82)</td>
<td>0.43 (0.06 − 0.77)</td>
<td>0.52 (0.09 − 0.87)</td>
</tr>
</tbody>
</table>

Table 4.4: Correlations for each of the four case definitions (columns) and three data sources (rows). Correlations calculated were Pearson’s, Spearman’s and the cross correlation function for each year individually. Results are displayed as the mean correlation and the range of correlations. The case definitions with the highest mean and range are coloured in orange and the second highest mean and range are shaded in light orange.

had the strongest correlations in general, and the second best fit in the other two data sets. For these reasons, we chose to use the Diarrhoea OR Vomiting definition for analysis.

However, since all three definitions have merit, this case definition should be reviewed when more data becomes available. In particular, the case definition should be reviewed in 2 to 3 years or if the researchers can identify and use another comparison data source, such as laboratory tests for all gastrointestinal illnesses, pharmaceutical purchases of over the counter gastroenteritis symptom relief medications, or emergency department surveillance in other states.

Comparison with other data sources

Gastroenteritis-related calls to the helpline showed some seasonality, but the effect did not occur every year. The peak of gastroenteritis activity occurred between October - December in each of 2013, 2014, 2015 and 2016 however from 2008 – 2012 there was not clear seasonality (Figure 4.1).

Gastroenteritis activity in the healthdirect helpline data is similar to the activity seen in the New South Wales Emergency Department data, with an average correlation of 0.74 (range of yearly correlation coefficients 0.50 – 0.86). The other data sources do not match as well, with an average correlation of 0.38 for the sentinel GP surveillance system and 0.52 for the WA Emergency Departments, as seen in Figure 4.3.

Percentage versus Number

The lowest percentage of gastroenteritis calls occurred during May to July of 2009. This is very likely due to the 2009 H1N1 influenza pandemic, which caused a large increase in the number of influenza-related calls to the healthdirect helpline. This decrease is not seen in the number of gastroenteritis calls, which steadily rises during 2009, as seen in Figure 4.1. The influence of influenza activity is also visible in 2011 and 2012 where
Figure 4.1: The number and percentage of calls related to gastroenteritis. This is presented as a moving 7-day average in order to smooth the effect of day of the week.
Figure 4.2: Comparison between gastroenteritis-related calls to the healthdirect helpline and visits to the NSW emergency department, each as a percentage of all calls or visits. This is presented as a moving 7-day average in order to smooth the effect of day of the week.

there is a trough in the percentage of calls related to gastroenteritis in the winter of 2011 and 2012, even though the number of gastroenteritis calls remains mostly constant. The clearly defined peaks in gastroenteritis activity by percentage in October - December of 2013 – 2016 are still visible in the number of gastroenteritis calls, though they are not as clearly defined.

Spatial Investigation

We investigated the spatial spread of gastroenteritis-related calls to the healthdirect helpline at the state and statistical local area (SLA) level from 11 June, 2017 to 27 August, 2017. In states with a larger population, such as New South Wales, we can see that gastroenteritis activity increases slowly during the time period to a peak in mid September (Figure 4.4). However, states with a smaller population, particularly the Northern Territory have a small number of calls, meaning there is more noise in the trend data. The maps of the statistical local areas in New South Wales (Figure 4.6) and in the Sydney region (Figure 4.6) show increasing rates of gastroenteritis-related calls each week (Figure 4.6). However, there is no obvious pattern of spatial transmission in either map.
4.5.2 Adverse Event Following Immunisation (AEFI)

The *adverse event following immunisation* or AEFI syndrome stayed steady at approximately 11.7 calls per day or 0.9 percent of all calls, except from March - April 2010, as seen in Figure 4.7. During this time, the number of calls rose to a maximum of 164 calls on April 21, 2010. The average number of calls from in 2009 was 11.75 or 0.9% each day. This increase was driven mostly in Western Australia, which increased from 4.3 calls per day to 26.4 calls per day. Children aged 0 – 5 years had the highest incidence of adverse reactions with an average of 43.3 calls per day. However the proportional increase was similar across all age groups, with around a 4 fold increase in calls per day in every age group. The difference between the number of AEFI-related calls occurring on a weekday compared to a weekend increased dramatically. Before March of 2010, there were 1.6 AEFI-related calls on a weekday for every 1 during the weekend, a ratio of 1.6 : 1. This increased to a ratio of 2.9 : 1 during March and April of 2010, as shown in Table 4.5. This is visible in plot 4.8, where the number of calls drops substantially every weekend or public holiday, especially from April 1-4, which was the Easter long weekend in 2010.
4.5.3 Suicide Ideation and Acute Mental Health Events

The *suicide ideation* syndrome has been steadily rising over the past 10 years, from a mean of 2.0 calls per day (0.2 percent of all calls) in 2008, to 6.8 calls (0.4% of all calls) in 2017, as seen in Figure 4.9. The *acute mental health events* syndrome is also a very small percentage of calls, with a mean of 15.1 calls, or 0.9% of all calls (Figure 4.10). However, the number and percentage of calls for *acute mental health events* has remained constant during the 9 years of the service. The most common symptom is anxiety, followed by depression, psychosis, and confusion.

4.5.4 Alcohol and Substance Abuse Events

The *alcohol or drug related events* syndrome is an interesting example of the effect of guideline changes on syndromic surveillance. From July 2008 - 14 July 2015, the alcohol and drug guidelines were separate from other toxic substance ingestion. Consequently we could measure the rate of calls for *alcohol or drug related events* at 6.1 calls per day,
or 0.36% of all calls. After 14 July, 2015, the guideline title changed to become *Ingestion of Toxic Substances*, which included alcohol, drugs and poisons. When the guidelines were combined, we could no longer differentiate between the calls that were drug or alcohol related and the poisons calls. The mean of *Ingestion of Toxic Substances* increased to 17.76 calls per day or 1.09% of all calls. On July 29, 2016, a further change was made to the guideline: this was to include a question about ingesting a small amount of dirt or animal faeces. Training was provided to nurses, so that nurses would begin using the "Ingestion of Toxic Substances" for people calling and requesting advice about eating dirt or animal faeces. After this change, the rate of calls to the *Ingestion of Toxic Substances* guideline increased further to 30.3 calls per day, or 1.9% of all calls.
Discussion

4.6.1 Gastroenteritis

Comparison of Data Sources

The data set that correlated best with the healthdirect helpline data was the New South Wales Emergency Department data, which also had the highest correlation for influenza-like illness. A factor contributing to this result is the size of the NSW ED data set compared to the other two, with NSW recording around 2.2 million visits as compared to the WA ED which records around 700,000 visits and the ASPREN system which has 200 GPs reporting. A smaller data set would lead to more noise and would decrease the correlation.
Percentage of adverse event following immunisation calls to the healthdirect helpline

Figure 4.7: The percentage of calls related to adverse event following immunisation syndrome. This is presented as a moving 7-day average in order to smooth the effect of day of the week.

**Percentage versus Number**

The gastroenteritis syndrome provides an interesting example of the advantages and disadvantages of using a percentage based indicator for surveillance. This is clearest in 2009, when the percentage of gastroenteritis calls appears to decline drastically (Figure 4.1). However, we know from the Chapter 3 that this year had a huge increase in the number of influenza-related calls, peaking on 14 June, 2009 at 33.1% and again on 19 July at 28.8% of all calls. This would explain the decrease in the percentage of calls due to gastroenteritis, which reached a trough on 19 July, 2009 at 5.0% of all calls - the same day as the second peak of influenza activity. This effect is also visible in most years, with a trough of gastroenteritis activity in July of 2011, 2012, 2013 and 2014.

However, investigating just the number of gastroenteritis related calls makes it much more difficult to identify increases in activity, and to compare between years. For example, 2012 and 2013 appear to have the most gastroenteritis activity according to the number of gastroenteritis calls, but the total number of calls to the helpline was highest in 2012 and 2013, which would contribute to the higher number of gastroenteritis calls. Using both the percentage and number of calls for surveillance provides a broader understanding of the situation.
Percentage of adverse event following immunisation calls to the healthdirect helpline from 1 November, 2009 to 30 May, 2010

Figure 4.8: The percentage of calls related to adverse event following immunisation syndrome from 1 November 2009 to 30 May, 2010. This is plotted daily, so that the effect of the weekend days is clearly visible.

Spatial Investigation

The results of the spatial investigation into gastroenteritis-related calls to the healthdirect helpline demonstrate that detecting outbreaks of gastroenteritis activity in the healthdirect helpline data is not possible unless that outbreak is very large. For states with smaller populations, particularly the Northern Territory, the activity can change from a small percentage to a large percentage week to week. This means that it is unclear whether there is a signal in the data, and whether a large percentage of gastroenteritis calls in the Northern Territory is due to an outbreak or to chance. In states with larger populations, such as New South Wales, there is a sufficient number of calls to ensure that a high percentage of gastroenteritis calls is not due to chance.

When we reduce the geographical cell size to Statistical Local Areas (SLAs) we see a similar result. Some SLAs are have large enough populations to show a gradual increase in activity, but most SLAs will have more volatile behaviour from week to week. However, the map does indicate increased activity in broader regions of the state. For larger states like New South Wales, calls could be mapped into 3 or 4 regions, which would have large enough populations.

This result is similar to what has been reported in other studies. Researchers in Sweden were able to identify outbreaks of gastroenteritis using telephone data, but only in outbreaks that were larger than 350 known cases, with an attack rate in the affected county between 30 and 21,311 per 100,000 [18]. Researchers in the United Kingdom
### §4.6 Discussion

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Number (percentage) per day</td>
<td>11.7 (0.9%)</td>
<td>19.6 (1.4%)</td>
<td>49.9 (3.6%)</td>
<td>13.3 (0.9%)</td>
</tr>
</tbody>
</table>

**Distribution of calls by State by percentage (average number of calls per day)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Western Australia</td>
<td>4.3</td>
<td>4.4</td>
<td>26.4</td>
<td>4.1</td>
</tr>
<tr>
<td>New South Wales</td>
<td>4.3</td>
<td>9.3</td>
<td>14.9</td>
<td>5.5</td>
</tr>
<tr>
<td>South Australia</td>
<td>0.9</td>
<td>1.4</td>
<td>3.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Australian Capital Territory</td>
<td>1.6</td>
<td>1.6</td>
<td>2.2</td>
<td>0.8</td>
</tr>
<tr>
<td>Tasmania</td>
<td>0</td>
<td>0.8</td>
<td>1.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Northern Territory</td>
<td>1</td>
<td>1.5</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>Other/ Unknown</td>
<td>1</td>
<td>1.5</td>
<td>3.0</td>
<td>1.1</td>
</tr>
</tbody>
</table>

**Distribution of calls by Age (Average number of calls per day)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 4</td>
<td>10.5</td>
<td>17.2</td>
<td>43.3</td>
<td>11.9</td>
</tr>
<tr>
<td>5 – 9</td>
<td>0.4</td>
<td>0.9</td>
<td>3.4</td>
<td>0.3</td>
</tr>
<tr>
<td>10 – 19</td>
<td>0.4</td>
<td>0.5</td>
<td>1.3</td>
<td>0.8</td>
</tr>
<tr>
<td>20 – 29</td>
<td>0.1</td>
<td>0.2</td>
<td>0.6</td>
<td>0.1</td>
</tr>
<tr>
<td>30 – 39</td>
<td>0.1</td>
<td>0.2</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>40 – 49</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>50 – 59</td>
<td>0.1</td>
<td>0.1</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>60 – 69</td>
<td>0.1</td>
<td>0.2</td>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td>70+</td>
<td>0.1</td>
<td>0.1</td>
<td>0.6</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Weekday and Weekend distribution (average number of calls per day)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekday</td>
<td>13.2</td>
<td>23.2</td>
<td>63.5</td>
<td>15.3</td>
</tr>
<tr>
<td>Weekend</td>
<td>8.3</td>
<td>12.5</td>
<td>21.9</td>
<td>8.8</td>
</tr>
</tbody>
</table>

Table 4.5: Characteristics of adverse event following immunisation syndrome before, during and after the unusual increase in March-April of 2010. Characteristics include mean number and percentage of calls, the age distribution of calls, the distribution of calls by state and proportion of calls on the weekday compared to the weekend.
Figure 4.9: The percentage of calls related to suicide ideation syndrome. This is presented as a moving 7-day average in order to smooth the effect of day of the week.

Furthermore, both the Swedish and UK studies had a higher average population per geographical unit, improves the signal to noise ratio. In Sweden, the study was conducted through municipalities, of which Sweden has 290, and with Sweden’s population of 9.9 million this corresponds to an average of 34,140 people per geographical unit. The study in the UK used 152 “primary care trusts”, which have an average of 425,000 people per geographical unit. By contrast, Australia has 1353 SLAs and a population of about 24.1 million: which is an average of 17,810 people per geographical unit. Consequently each geographical unit in Australia will have low power to detect a signal in the data.

In conclusion, spatial surveillance of gastroenteritis activity to the helpline could be useful if we were able to reduce the amount of noise in each cell, or increase the signal size. This could be occur if: we decreased the number of geographical cells; there was a very large and widespread outbreak; or if the proportion of Australians who used the helpline increased. If none of these criteria are met, evaluating the spatial distribution of calls will be of limited use.
§4.6 Discussion

4.6.2 Adverse Reactions to Vaccines

The spike in calls related to an adverse event following immunisation occurred at the same time as an increase in adverse events following immunization (AEFI) reported by the Therapeutic Goods Administration (TGA). [72, 73] The TGA collects data on adverse events following immunisation from the public, health professionals and industry. A report was written on this increase in AEFI which found there was a higher frequency of "early fever responses [and this is] associated with substantially higher rates of trivalent influenza vaccine- associated febrile convulsions in children 6 months to less than 5 years of age, particularly in WA." This increase occurred from 8 March 2010 - 22 April 2010, which is the period that the trivalent influenza H1N1 vaccine was made available, and resulted in an estimated 9 cases per 1,000 doses in WA and 5 cases per 1,000 doses in other jurisdictions. This was a large increase from the usual rate of between 0.08 – 0.17 per 1,000 doses. The early fever responses occurred within 4 – 24 hours of vaccination, which means that on the weekends, the number of reactions dropped as vaccines were not administered.

Many attributes of the outbreak described by the TGA are also demonstrated in the healthdirect helpline data. The number of AEFI calls increased from 8 March, 2010 - 25 April, 2010, and returned immediately back to baseline levels from 25 April, 2010. The increase occurred mostly on weekdays, which is consistent with a 4 – 24 hour reaction after vaccination. Western Australia saw the largest proportional increase in calls, with nearly six times as many calls per day. The TGA was unable to identify why Western Australia would be more affected than the other states and territories, though
possible explanations include a higher level of influenza vaccination per population in Western Australia, a higher level of reporting or more use of the brand that caused the majority of the reactions. There were four influenza vaccines being used in Australia in 2010, but the Seqirus (previously bio CSL) FLUVAX/FLUVAX JR had a higher rate of febrile convulsions than the other three. Consequently, if GPs in Western Australia favoured this brand, that would lead to higher rates of adverse reactions in that state. Unfortunately, data on brand preferences between states was unavailable.

The final attribute of this outbreak described by the TGA was that it disproportionately affected children under the age of 5 [72]. In the healthdirect helpline data, children under the age of 5 had the largest number of calls per day but there was a similar increase proportionally across all age groups. In other words, there was a 4 fold increase in the number of calls per day in all age groups, not just children under the age of 5. The difference between the findings of the TGA and the healthdirect helpline could be due to the reaction being more severe in children. If the reaction were more severe in children, this would lead to more children than adults seeing GPs or being admitted to hospital for an adverse reaction. Consequently, we would see an equal increase in calls across all age groups but an disproportionate increase in GP and hospital visits for children under the age of five. In the TGA investigation of the 2010 spike, all of the patients with febrile convulsions that were reported had visited a healthcare professional. This would lead to the TGA seeing a disproportionate increase in reactions for children under the age of 5. Consequently, the difference in age distribution between the healthdirect helpline and the TGA report could be due to an increase in reactions across all age groups, but a more
There was also a small increase in calls in January and February of 2010, particularly in New South Wales. This is likely due to the Seqirus (previously bio CSL) FLU-VAX/ FLUVAX JR, which was introduced in Australia for people aged 10 and over from September 30, 2009 and for children between 6 months and 10 years from December 2009. The increased rates of AEFI in early 2010 are likely due to people who have elected to have the vaccination before the commencement of the government funded trivalent influenza vaccination program, which began on March 8 in WA. A quarterly report published by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) demonstrates an increase in febrile reactions in the first quarter of 2010, though it is unclear whether the majority of these reactions occurred in March or whether the rate of reactions was elevated in January and February as well [74]. However, in both the CSIRO report and the healthdirect helpline data, as soon as the vaccine was discontinued on April 22, 2010, rates of reactions dropped to 2009 rates. This analysis demonstrates that the event was recognisable early through the healthdirect helpline data.

4.6.3 Suicide Ideation and Acute Mental Health Events

Neither suicide ideation nor acute mental health events (AMHE) demonstrated any notable events, seasonal patterns or attributes that would assist in surveillance. The percentage of suicide ideation calls has increased slightly throughout the nine years of data. Though there is limited data about the rates of suicide ideation, the Australia Bureau of Statistics reported a slight increase from 10.2 suicide deaths per 100,000 in 2006 to 12.6 deaths per 100,000 in 2015, which is consistent with our results.

The acute mental health events syndrome consist of callers with symptoms of mental health illness, with anxiety and depression being the two most common symptoms. This is consistent with the National Survey of Mental Health and Well-being, conducted by the Australian Bureau of Statistics in 2009. This survey found that anxiety disorders were the most common, followed by affective disorders, of which depression makes up the largest part. Since the most recent mental health survey was conducted in 2007, it is unclear whether the rate of mental illness has changed over the past ten years, however the the rate of mental illness remained unchanged from 1997 to 2007.

4.6.4 Alcohol and Substance Abuse Events

Alcohol and substance abuse events demonstrate the difficulty in conducting surveillance with a dataset that changes in order to optimise triage. The three changes in the data occurred first when alcohol and illegal substances was combined with ingestion of all substances, and then again when the guideline was updated to include ingestion of dirt or animal faeces. It is unclear how nurses were classifying the dirt/ faeces calls before the change, since “Foreign Body Ingestion or Insertion” and “Symptom with no matching algorithm” remained unchanged before and after the change on July 29, 2016. It is likely that nurses were previously considering calls about eating dirt or animal faeces as non-clinical calls, and were giving advice without conducting triage. Consequently prior to the change these calls would not be included in the nurse triage database at all, and after the change they significantly increased the percentage of callers using the guideline "Ingestion of Toxic Substances."
4.7 Public Health Implications

The syndromes investigated in this chapter demonstrate some of the many opportunities for surveillance, but also the difficulties. This ranges from the suicide ideation syndrome which does not appear to be useful, to the adverse event following vaccination syndromes which retrospectively identified a known outbreak.

Of the 6 syndromes, the syndrome with the least utility at this stage are suicide ideation and acute mental health event. Both of these syndromes do not demonstrate any seasonality, nor can we identify any periods of time with increased risk of suicide or acute mental health calls. Consequently, there are no actionable conclusions that could be used for public health intervention. Although the suicide ideation syndrome appears to have been steadily increasing, it is unclear why this has occurred, or what public health action would result from this information. Finally, we did not identify any unusual spikes in activity during the ten years of surveillance. Continued surveillance on these syndromes is unlikely provide useful or actionable information.

The alcohol or drug related events syndrome was similar to the acute mental health events and suicide ideation syndromes in many ways. Alcohol or drug related events did not demonstrate any seasonality, any periods of time with increased risk or any spikes of unusual activity. However, this syndrome did show the effect of changes in guidelines and how a change could affect surveillance. There were two unusual increases in the number of calls to this syndrome, but both were not due to increased number of alcohol or drug related calls but rather to the incidental inclusion of other types of calls into the syndrome due to guideline changes. The first change was due to the inclusion of poisons and the second, the inclusion of dirt and animal faeces. This demonstrates that any unusual activity needs to be fully investigated by a person well versed in the healthdirect helpline system at Healthdirect Australia before being reported further or used for public health action in order to avoid misreporting.

The syndrome that demonstrated the most utility was adverse event following immunisation, since this syndrome detected a large spike in activity. This spike in activity was consistent with a spike in febrile convulsions, which was detected by the TGA, and led to removing the vaccine from the market. In order to contribute to this public health action, Healthdirect Australia needs to partner with the Therapeutic Goods Administration, public health departments and other government bodies. This would mean that if Healthdirect Australia were to identify another outbreak, it would be able to notify the appropriate government bodies for public health action.

The final syndrome investigated was gastroenteritis, and this also demonstrated some of the complications with surveillance. All public health departments conduct surveillance activities for gastroenteritis, and there is also a national entity, OzFoodNet, which is responsible for investigating potential foodborne disease. Our research showed that the pattern of gastroenteritis calls to the healthdirect helpline is quite volatile and not easy to predict. There is a degree of seasonality, with a peak in gastroenteritis activity occurring mostly in autumn, but not every year. Consequently, surveillance on the overall level of gastroenteritis is not particularly useful, since public health actions as a result of this information are limited. Furthermore, we are unable to identify outbreaks
Lessons Learnt

This work was a continuation of work completed for Chapters 2 and 3, so many of the lessons learnt applied to this work as well. However, there were 3 further lessons learnt while researching and writing this chapter. Firstly, to use a smaller “sample” data set to check time-consuming sections of code, secondly to have several backups, particularly when working with these computationally intensive codes and thirdly that some results can be unsurprising but still inform further research.

The work on geographical regions presented in this chapter was computationally very slow, and also meant that an error in the code would often freeze and crash the computer. This was very frustrating, since my computer would often be largely unusable for up to one hour at a time. The most frustrating aspect was that sometimes a simple error such as a missed parenthesis would essentially cause a wasted hour. In retrospect, I should have utilised a very small data set - or created a very simple dummy data set - to test my code. Once I was confident it worked, I could then run it overnight or during lunchtime when my computer was not being used.

Furthermore, crashes sometimes caused the file to be completely erased - where a saved file would be rewritten with a completely blank document. This first time this occurred it was terrifying! Due to complications with saving to the cloud and running the code, I had only daily backups, instead of having every previous version. After losing a whole days worth of work due to a mysterious crash, I learnt better version control and backups.

Finally, some results of the investigation were what we predicted, but completing the research for them was still useful to inform further research. In particular, we predicted that the investigation into the smaller geographical cells for gastroenteritis would be unable to detect outbreaks. However, having completed the research we now have a better indication of why this might be the case, how big a sample size you would need to detect signals and characteristics of a syndrome which would increase or decrease the
chances of detecting outbreaks on a small geographical scale. The other unsurprising yet very helpful result was the changing guidelines in the alcohol or drug related events syndrome. We predicted that there would be too few calls to detect any unusual activity, which was the case. However, this syndrome gave us a very clear example of what might go wrong in a surveillance system if it were entirely automated, or if it were monitored by someone who did not understand the intricacies of the system.

The result of this is both good and bad. This means that sometimes research can be useful even if you are showing no effect. On the other hand, this also means that in order to determine the utility of a syndrome, you have to test it extensively. So adding a syndrome to this surveillance system will be a time-consuming process, and discounting syndromes as not useful for surveillance will be even more time consuming.
Outbreak Investigation: A measles outbreak in western Sydney, Australia from March 2017 - May 2017

5.1 Introduction

In this chapter, I report on my secondment to the Western Sydney Public Health department on a part-time basis from 23 March, 2017 to 19 May, 2017 to assist in a measles investigation. My role was to assist with the investigation in any way possible, including contact tracing, public outreach, data entry, data analysis and coordinating a report of the public health actions as a response to the measles investigation. This chapter describes the outbreak, characteristics of the cases, and the public health response to the outbreak and comments on the lessons learnt.

Measles has been notifiable to the National Notifiable Diseases Surveillance Scheme (NNDSS) since 1944 due to its high morbidity, mortality and its highly infectious nature. Prior to the introduction of the measles vaccination worldwide, measles caused an estimated 2.6 million deaths each year and major epidemics of measles occurred every 2 - 3 years [75]. Australia achieved measles elimination on 20 March, 2014. To achieve this status, countries must show an interruption of endemic measles virus transmission for a period of at least 36 months as well as a well-performing surveillance system and supportive genotyping evidence.

The outbreak occurred in South-west Sydney from 9 March, 2017 to 2 May, 2017 and was the largest measles outbreak in New South Wales since February, 2014.

The index case of this outbreak was a 23 year old male who returned from Indonesia to his home in Liverpool. Onset of illness was on March 2 and the case was reported to the Parramatta public health unit by a GP on March 8 following a positive laboratory test. Prior to the laboratory test, he presented at a medical center on 4 March, 7 March and 8 March, an emergency department on 6 March and a chemist on 4 March. Contact tracing was commenced on 8 March and included phoning all people who were in the same room as the man, and and those who were in the room in the following half hour.
Following the index case, a further 17 cases of measles were notified to one of 4 different local district public health departments in Sydney. These cases had onset of illness between March 17 and April 15 and all contracted the disease locally since none reported travelling overseas during their incubation period. However, none of the cases reported being in the same location as the initial patient, though three did report being in a similar area at a similar time.

This chapter describes the outbreak, the characteristics of the measles cases, the public health actions and comments on lessons learnt both personally and from a public health perspective. This piece of work was used by the New South Wales Ministry of Health to evaluated the time, cost and resources required to respond to this kind of measles investigation.

5.2 Methods

Cases of measles were notified to the public health department through routine passive surveillance from laboratories, GPs and emergency departments. Further cases continued to be notified through passive surveillance and also were identified through contact tracing. The corresponding local public health departments conducted interviews with cases to identify possible contacts and then contacted all contacts of each case. The case definition for the outbreak was defined as a case that meets the Australian national notifiable diseases case definition for measles, as stated above, with disease onset between 9 March, 2017 and 2 May, 2017 and located in Sydney, New South Wales during the exposure period.

The Australian case definition of measles is:
Laboratory definitive evidence; or clinical evidence and epidemiological evidence.

Laboratory definitive evidence is defined as:
- isolation of measles virus;
- or detection of measles virus by nucleic acid testing or;
- or detection of measles virus antigen;
- or IgG seroconversion or a significant increase in antibody level or a fourfold or greater rise in titre to measles virus except if the case has received a measles-containing vaccine eight days to eight weeks before testing;
- or detection of measles virus-specific IgG antibody confirmed in an approved reference laboratory except if the case has received a measles-containing vaccine eight days to eight weeks before testing.

Clinical evidence is defined as:
- a generalised maculopapular rash last three or more days
- and a fever of 38 degrees C or more at the time of rash onset
- and one of: cough or coryza or conjunctivitis or koplak spots.

Epidemiological evidence is defined as:
- contact between two people involving a plausible mode of transmission at a time when one of them is likely to be infectious (approximately five days before to four days after
rash onset and the other has an illness that starts within 7 - 18 (usually 10) days after this contact; and one cases in the chain of epidemiologically linked cases (which may involve many cases) is laboratory confirmed.

All cases were interviewed by public health officers over the phone. Data collected included locations visited while infectious, locations visited while in the incubation period, all known contacts visited during the infectious period including family members, work colleagues and friends. We collected data on the case’s age, sex, vaccination status and ethnicity in an excel spreadsheet. We also recorded all public health interventions performed, including the numbers of telephone calls, faxes, clinics and media releases. I performed descriptive analysis on the data, which included investigating the epidemiology curve, the age and sex distribution, vaccination status and the ethnicity of the cases and possible transmission pathways. I also completed analysis on the public health interventions performed, including summarising the numbers of people identified, the number contacted, the number identified as susceptible, locations identified, and advocacy actions performed.

5.3 Results

There were 23 measles cases notified to the NSW Ministry of Health from 1 January, 2017 - 2 May, 2017. The first 6 of these cases were acquired internationally and do not fit the case definition for the outbreak, as seen in Figure 5.1. Case 6 was identified as the index case for the outbreak, with disease contracted in Indonesia. A further 17 cases met the case definition for the measles outbreak as they were located in Sydney, NSW during their infectious period and their case onset was between 9 March, 2017 and 2 May 2017.

Data collected by public health officers during the investigation shows that 7 of the 17 cases identified as Aboriginal and 3 as Pacific Islander, as is seen in Figure 5.2. The median age of the cases was 21 and 3 cases were under 12 months old, which is below the recommended age for measles vaccinations at 12 and 18 months. A further 5 cases were under the age of 10, though 4 of these were in the same household, and 2 cases were under 20, both in the same household. The remaining 7 cases were between 20 and 60 years of age, as seen in Table 5.1 and Figure 5.3. There were 10 unvaccinated cases, 3 of those because they were under the age of 1. As seen in Table 5.2, 1 case had received 1 measles- containing vaccine and the remaining 6 cases were unsure of their vaccination status. All of the cases that were genotyped were infected with measles subtype D8. Eight of the cases had an unknown infection source, while the remaining 9 were infected by another case in the outbreak, as seen in Figure 5.4. Ten of the cases attended an emergency department or hospital, and the remaining 7 only attended GP clinics, though 5 of those belonged to one family unit.

Public health actions in response to the outbreak included contacting 7,582 people over 57 locations around Sydney. Of these contacts, 6,125 were identified by Church X as having potentially attended a service which was also attended by 2 infectious measles cases. A “location” was defined as any indoor area where a case with confirmed measles spent more than five minutes while infectious. The infectious period was calculated as
24 hours prior to symptom onset until 4 days after the onset of the rash. There were also several locations and contacts in Rockhampton that were connected to the outbreak in Sydney but were investigated by the Rockhampton public health unit. These contacts are not included in this summary. This was due to a contact of a measles case who flew to Rockhampton the day of onset of symptoms. He subsequently received a positive laboratory result for measles in Rockhampton.
Birth year and vaccination status of measles cases in NSW between 1 January 2017 and 2 May 2017

Figure 5.3: Birth year and vaccination status of measles cases in Sydney, NSW with case onset between 1 January 2017 and 2 May 2017

<table>
<thead>
<tr>
<th>Age Group in years</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2</td>
<td>3</td>
</tr>
<tr>
<td>2-10</td>
<td>5</td>
</tr>
<tr>
<td>11-20</td>
<td>2</td>
</tr>
<tr>
<td>21-30</td>
<td>2</td>
</tr>
<tr>
<td>31-40</td>
<td>3</td>
</tr>
<tr>
<td>41-50</td>
<td>1</td>
</tr>
<tr>
<td>51-60</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.1: Age of measles cases in Sydney, NSW with case onset between 1 January 2017 and 2 May 2017.

<table>
<thead>
<tr>
<th>Vaccination Status</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Yes - Partially</td>
<td>1</td>
</tr>
<tr>
<td>Yes</td>
<td>0</td>
</tr>
<tr>
<td>Unknown - can’t recall</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 5.2: Vaccination status of measles cases in Sydney, NSW with case onset between 1 January 2017 and 2 May 2017

Public health actions included inquiring about vaccination status either through investigating Australian Immunisation Records or by speaking to the contact directly,
Outbreak Investigation: A measles outbreak in Sydney, 2017

telephoning contacts, running clinics to deliver the measles, mumps and rubella (MMR) vaccine or to give normal human immunoglobulin (NHIG) injections, arranging for MMR or NHIG injections, sending letters, hanging posters in shopping centres, sending fax alerts and sending media alerts. Details for these actions, as well as the number of contacts involved are provided in Table 5.3. Through contact tracing, public health officers identified 389 people who were susceptible to measles infection, 180 of which were identified through searching the Australian Immunisation Register for students and teachers who attended the same High School as a case.

5.4 Discussion

This outbreak investigated 17 measles cases, and found a high incidence in Aboriginal and Torres Strait Islander populations, and in people aged under 10. Of these, 10 people were unvaccinated, 1 had received only 1 measles-containing vaccine and the rest were unsure of their vaccination status. There was also a high proportion of cases with unknown infection pathways. Ten of the cases visited an emergency department, and the remaining 7 visited a GP clinic. Public health interventions reached 7,582 people, including 6,125 identified by Church X as potentially having attended a service with 2 infectious measles cases and 793 who attended the same high school as 2 measles cases. Further investigation such as telephone calls and searching on the Australian Immunisation Register identified 389 people who were susceptible to measles infection. However, with only 17 cases in the outbreak, conclusions should be made with caution.

The incidence of measles was higher in Aboriginal and Torres Straight Islander people than expected. However, 5 of these cases belonged to one family group, which skewed the results. The Australian Immunisation Records reports on the proportion of children with a Medicare number that have had a vaccination through an immunisation provider, including general practitioners, public immunisation clinics and others [76]. These are published quarterly online for three cohorts: cohort 1 are children aged 12 − 15 months in that quarter, cohort 2, 24 − 27 months and cohort 3, 60 − 63 months. Rates of vaccination coverage for measles, mumps and rubella is only recorded in Cohorts 2 and 3 since measles vaccination occurs at 12 and 18 months. Current estimates in New South Wales are that the vaccination rate in both cohorts is higher in those of Aboriginal and Torres Straight Islander heritage than those without [77,78]. Consequently, it is likely that the higher incidence of measles cases in Aboriginal and Torres Straight Islander people found in this outbreak is due to chance.

The public health actions in response to this outbreak were extensive and varied. The most time consuming actions were running the clinics, telephoning contacts and investigating the vaccination status of contacts. Many people were unable to recall if they had been vaccinated once or twice, and did not have an entry on the Australian Immunisation Records. If the public health officer called within three days of contact, the officer would suggest a measles, mumps and rubella (MMR) injection. However, if they called later than three days and the contact did not have any risk factors requiring normal human immunoglobulin (NHIG), jurisdictions varied on whether they classified these cases as “susceptible” or not. Consequently the number of susceptible people


<table>
<thead>
<tr>
<th>Public health action or</th>
<th>Number</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locations in Sydney identified</td>
<td>57</td>
<td>A location was defined as any indoor area where a case with confirmed measles spent more than five minutes while infectious. There were also several locations and contacts in Rockhampton, managed by the Rockhampton public health unit and not included in these results.</td>
</tr>
<tr>
<td>Contacts identified</td>
<td>7,582</td>
<td>Of this total, 6,125 were contacts identified by Church X as having potentially attended a service which was also attended by two infectious cases.</td>
</tr>
<tr>
<td>Vaccination status investigated</td>
<td>1,457</td>
<td>Of this total, 793 were investigated through the AIR since contacts were identified by High school Y as attending the same school as an infectious case. The remainder of the contacts were telephoned.</td>
</tr>
<tr>
<td>Contacts telephoned</td>
<td>664</td>
<td>Each contact was attempted to be called three times. If required, a telephone interpreter was used.</td>
</tr>
<tr>
<td>Contacts identified as susceptible through contact tracing</td>
<td>389</td>
<td>Of this total, 180 were identified as susceptible through searching the Australian Immunisation Records (AIR) of students and teachers at High School Y.</td>
</tr>
<tr>
<td>Measles, Mumps Rubella (MMR) vaccine given or arranged</td>
<td>210</td>
<td>Of this total, 95 were at High School Y.</td>
</tr>
<tr>
<td>Normal Human Immunoglobulin (NHIG) injection given</td>
<td>20</td>
<td>This included 3 pregnant women, 4 babies under the age of 1, 1 immunocompromised person and 12 unrecorded reasons.</td>
</tr>
<tr>
<td>Letter sent due to contact being uncontactable by telephone</td>
<td>52</td>
<td>After 3 attempts to call over 48 hours, contacts were sent a letter.</td>
</tr>
<tr>
<td>Media releases by NSW Ministry of Health</td>
<td>9</td>
<td>These media releases were then further published in a number of papers and received TV coverage.</td>
</tr>
<tr>
<td>Clinics run</td>
<td>3</td>
<td>This consisted of two clinics where only NHIG was administered and one clinic where only MMR was administered.</td>
</tr>
<tr>
<td>GP and ED alerts faxed to Western Sydney and South Western Sydney</td>
<td>4</td>
<td>These alerts were sent to all general practitioners and emergency departments registered in the western Sydney and south western Sydney districts.</td>
</tr>
</tbody>
</table>

Table 5.3: Public health actions in response to the measles cases in Sydney.
Figure 5.4: Probable infection pathways. Squares indicate family groups, solid lines indicate an epidemiological link between cases, dotted lines indicate an unknown infection source.
could be higher or lower than the 389 people identified.

The measles outbreak appeared to affect mostly people under the age of 40, particularly children under the age of 10. Three of these were children under the age of one who could not yet have been vaccinated. A further 5 cases were aged between 1 and 10, however this result is again skewed by 4 unvaccinated children under the age of 10 in the same family group, all of whom contracted the disease. There were also 7 cases aged between 10 and 40, with birth years between 1980 – 2001. This age group has previously been identified as a having a lower rate of measles vaccinations. A serosurvey conducted in 2001 identified that immunity was lowest in adults born between 1994 – 1998 and 1974 – 1980 [79], and a similar study conducted only in Victoria reported lowest immunity rates in those born between 1975 – 81 [80]. This was also reported to have contributed to 2 large outbreaks in young adults in Victoria in 1999 [81] and 2001 [82]. However, these are only 7 cases of 17, and this result could be due to chance.

A large challenge in this outbreak was the inability to establish a clear epidemiological link between all cases and secondary cases. It is unknown how 6 of the seventeen cases contracted the disease since they do not report being in the same area as any previous cases during their incubation period. A further three cases (Cases 1, 2 and 3) report being in the same general area (the town centre for cases 1 and 2, the hospital for case 3) as the initial case but not in the same room or at the same time. Furthermore, contact tracing investigations discovered secondary cases only in family groups and one secondary case at High School X. In all other instances, there were no known secondary cases identified through contact tracing of hospitals, GP clinics, Most notably, there were no secondary cases from Church X, despite the two infectious cases attending a very large church service with up to 4,000 people in attendance. Public health departments were unable to provide MMR to any of those contacts since public health departments were not advised of this contact before the 72 hour deadline, so any susceptible contacts were likely to catch the disease.

5.5 Public Health Implications

This is the largest measles outbreak to have occurred in New South Wales since February, 2014. Since then, measles elimination in Australia was declared by the WHO on 20 March, 2014. In order to achieve this status, countries must show an interruption of endemic measles virus transmission for a period of at least 36 months as well as a well-performing surveillance system and supportive genotyping evidence. Since the index case was acquired overseas and measles transmission was interrupted after 2 months, Australia still meets these requirements.

The resources required to conduct the measles investigation and maintain this status are immense. A study conducted western Sydney in 2011 estimated costs to a public health department of about 2,488 AUD, or 1,701 USD for one case of measles with 75 contacts [83]. This included the cost of staff hours, telephone, mail and pathology in the public health department and did not include the cost of the MMR or NHIG vaccinations, visits to the hospital or GP or costs incurred by the patients. This is similar
to a study conducted in the Netherlands, which attributed approximately 1078 USD per case to the public health costs associated in investigating and containing an outbreak of 2,700 cases [84]. A further 606 USD per case was attributed to hospitalisation costs, GP visits and vaccinations. Other studies have attributed costs to public health departments equivalent to 13,485 USD for a single case in Kentucky, United States [85], 17,196 USD for each case in an outbreak of 22 cases in China [86], to 142,452 USD for a single case in Iowa, United States [87]. The large range in costs associated per case could be caused by the number of contacts per case, more extensive contact tracing and the expensive or extensive pathology tests.

These studies have all demonstrated that measles investigations are expensive and require a lot of staff time at a public health department. However, contact tracing and public health interventions, as well as vaccination have been shown to reduce the overall cost to the health care system. A systematic review of 67 studies found significant health and economic benefits associated with measles and rubella vaccination [88].

In the Australian setting, the cost of hospitalisation, GP visits and pathology tests for people infected with measles would be significantly higher if there was no public health response to a measles outbreak. A study in Korea of a measles elimination program estimated that a change from a single MMR vaccination to a two-dose MMR campaign with additional catch-up of measles and rubella at 7 – 16 years would reduce the number of measles cases from 178,560 to 5,936 [89]. This, together with the reduction of rubella and mumps cases would constitute a cost-benefit ratio of 1.27 with a savings of 51.6 billion KRW (South Korean Won). A further study demonstrated that public health intervention strategies were effective in reducing the number of secondary measles cases in an outbreak and the probability of an uncontrolled outbreak [90]. These strategies included contact tracing, providing both MMR and NHIG post-exposure prophylaxis, and recommending voluntary isolation and quarantine, and caused a reduction of 0.28 secondary cases per case. This demonstrates the importance of contact tracing, and the role that it plays in identifying people at risk, monitoring the spread of the outbreak, providing information and advice to cases and contacts and reducing the chance of an uncontrolled outbreak.

This study contributes to the literature on the effort associated with the public health response to a measles outbreak. This work includes contact tracing, telephone, posting and emailing contacts, running MMR and NHIG clinics and pathology testing. These activities all cost the public health department, mostly through additional staff hours required. Despite these costs, the investigation provides an overall saving to the health care system, due to the reduction of secondary cases.

5.6 My Role

I assisted in investigating this outbreak through the Parramatta public health department and also liaising with the New South Wales Ministry of Health to coordinate the public health responses after the outbreak had finished. In particular, I assisted in contact tracing and public outreach, particularly to with regards to Church X. This involved calling contacts of cases, following up contacts that required either MMR or NHIG, and
recording details of contacts and contact tracing outcomes.

After the outbreak concluded, I coordinated a report of public health actions that were taken across local health districts in response to the outbreak. This included collating the public health responses undertaken at Parramatta and also conducting telephone interviews with Liverpool, Penrith and Camperdown public health units. I then presented these results to the New South Wales Ministry of Health in the form of an spreadsheet and this dissertation.

5.7 Lessons Learnt

Working in the Parramatta public health unit was invaluable for my MAE experience since it gave me an opportunity to observe an outbreak investigation and also the usual work of a public health department. In particular I learned the value of ready-made action plans, the necessity of a fast-moving response and the important role public health departments play in informing people so as to ease public fears.

When I arrived at the Parramatta public health unit on 23 of March, the measles cases 1, 2 and 3 had only just been confirmed about an hour previously. However, since there was an action plan in place, all the documents, work plans and methods were ready to be deployed immediately. This enabled the public health department to begin contact tracing within minutes of receiving notification of measles. Since measles contract tracing is quite involved and also time critical, this preparation saved a huge amount of time but also allowed the public health department to reach potentially vulnerable people in time.

The necessity to complete a very rapid response also took me by surprise, though now makes complete sense. I was informed about the outbreak by email at 9am on Wednesday morning and by 10.15am I was on a train out to Parramatta and speaking to contacts by 11.30am. This speed enabled us to get NHIG treatments to vulnerable contacts within the requisite 6 days post contact - even starting 1 day later would have made this impossible.

Finally it was very interesting to hear about the necessity to keep the public informed, not because we believed there would be many people who had contact with a case but to assuage their fears. It is important to be seen to be taking action and to be keeping the public informed in the event of a public health emergency, both politically and for the mental health of the people.
Teaching

I completed two teaching tasks
Teaching
Appendix: Deriving the Cumulative Sum Formula

In this appendix we derive the three formulae used for signal analysis in Section 3.

A.1 NSW Emergency Department CuSum

The Public Health Real-time Emergency Department Surveillance System (PHREDSS) used by New South Wales chooses to set the reference value $k$ equal to the number of presentations seven days previously (one week ago on the same day of the week) and then adjusts for both seasonal variability and intra-week variability in a two-step process, as follows.

In step one, we make the index, $S_i$, independent of the level of syndrome incidence. The index is divided by the mean observation of the previous 365 days to create $\bar{S}_i$. This serves to normalise the magnitude of the event.

$$\bar{S}_i = \frac{S_i \cdot 365}{X_i + X_{i-1} + \ldots + X_{i-364}} \quad (A.1)$$

We then further standardise by making the index, $S_i$, independent of the variability of the syndrome incidence. This is done creating an estimator called the difference, which is defined as $D_i = \bar{S}_i - \bar{S}_{i-1}$. We then take the standard deviation of the differences for the previous 365 days. This serves to remove the effect of any variation in standard incidence over the year.

$$\hat{S}_i = \frac{\bar{S}_i \cdot 365}{D_i + D_{i-1} + \ldots + D_{i-364}} \quad (A.2)$$

The final index, $\hat{S}_i$ is what is referred to as the "PHREDSS Index" and is used to create the plots and results in Section 3.4.
A.2 Hutwagner and Cowling

To derive both the Hutwagner and Cowling indices, we standardise using the Student’s t-statistic as we have a sample size of 7 (either 7 weeks or 7 days) so we are not able to use the usual standardisation formula. Therefore, in order to standardise \( X_i \) we must use the following formula.

\[
Z_i = \frac{X_i - \bar{X}_i}{s_i/\sqrt{N}}
\]  
(A.3)

Where \( s \) is the standard deviation of the sample, \( N \) is the sample size and \( \bar{X} \) is the mean of the sample.

We begin again with Lucas’ Poisson CuSum formula (Equation (3.1)) and insert the standardisation formula, Equation (A.3):

\[
S_i = S_{i-1} + \frac{X_i - \bar{X}_i}{s_i/\sqrt{N}} - k
\]  
(A.4)

\[
= S_{i-1} + \frac{\sqrt{N} (X_i - \bar{X}_i) - ks}{s_i}
\]  
(A.5)

This is the formula given by Hutwagner. Note however that Hutwagner defines \( N \) as the total number of cases over the past 7 years. However, my derivation indicates that \( N \) is the sample size, which is 7.

The other variables are defined as:

Hutwagner:
\( \bar{X} \) is the mean of the number of cases on the same day of the week in the same Epiweek of the past seven years (including the observation)
\( s_i \) is the standard deviation of number of cases on the same day of the week in the same Epiweek of the past seven years (including the observation)

Cowling:
\( \bar{X} \) is the mean of the number of cases on the same day of the week for the previous seven weeks (including the observation)
\( s_i \) is the standard deviation of number of cases on the same day of the week for the previous seven weeks (including the observation)
Appendix: Writing for a non-scientific audience

During my time at Healthdirect Australia, I wrote many documents for a non-scientific audience. I have chosen 2 documents to demonstrate the breadth of writing for a non-scientific audience that I undertook. These documents are:

- Syndromic Surveillance Standard Operating Procedures

A copy of these two documents is included in this Appendix.
Syndromic Surveillance Standard Operating Procedure

<table>
<thead>
<tr>
<th>NAME</th>
<th>TITLE</th>
<th>SIGNATURE</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>M. Hartley MAE Scholar, Healthdirect Australia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authoriser</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Effective Date: 
Review Date: 

READ BY

<table>
<thead>
<tr>
<th>NAME</th>
<th>TITLE</th>
<th>SIGNATURE</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. PURPOSE
This document details the working, rationale and goals of the syndromic surveillance system at Healthdirect Australia.

2. INTRODUCTION
Healthdirect Australia runs the nurse triage helpline, which receives over 700,000 calls per year and the online symptom checker, which receives over 600,000 visits per year. Both of these data sources provide information about symptoms of users and can be used for syndromic surveillance for syndromes such as influenza-like illness and gastroenteritis.

3. SCOPE
This document details the responsibilities, operations, quality control, stakeholders and limitations of the surveillance system.

4. DEFINITIONS

5. RESPONSIBILITIES

5.1 Healthdirect Australia
The responsibilities of Healthdirect Australia include updating the portal, monitoring the outbreak alerts, communicating with stakeholders and quality control.

Reporting Team
The reporting team is responsible for updating the data from Medibank, running the automated query and sending a jpeg of the updated plot to the publishing team.

Clinical Governance Team
The clinical governance team looks after safely and quality of the presented information. They validate the data and confirm any alerts before they are sent further.

Publishing Team
The publishing team is responsible for updating the webpage with data and plots provided to them by the reporting team.

Service Operations/Management
The service operations/management team manages operations of the contact centres and digital products.

5.2 Steering Committee
The responsibilities of the Healthdirect Australia Telephone Syndromic Surveillance (HATSS) steering committee include overseeing the operation of the system and advising on actions for outbreak alerts.
6. **SPECIFIC PROCEDURE**

6.1 **Methods:**

A flow chart of the methods is seen in Figure 1.

**To conduct regular surveillance and update the flu page:**

Data from Medibank are added daily to the Healthdirect Australia database through an automatic process. Twice weekly, on Monday and Thursday, a query is run by the reporting team. This query counts both the total number of calls each day and the calls that match the case definition each day. The query updates an excel spreadsheet which contains the plots to be published onto the Healthdirect Australia website each day. These plots get sent as a jpeg image via email to the publishing team. The publishing team then update these plots on the website.

**To update the Flumeter:**

A Tableau dashboard will be sent to the reporting and publishing teams at Healthdirect Australia by Medibank. This includes analysis regarding the flu meter. The flu meter update will be checked for validity by team members in the clinical governance and reporting teams.

If the team members decide to update the flu meter, he/she will inform the publishing team via email. The flu meter will then be updated.

**To generate a flu alert:**

A Tableau dashboard will be sent to the reporting and publishing teams at Healthdirect Australia by Medibank. This includes analysis regarding a flu alert. The flu alert will be checked for validity by team members in the clinical governance and reporting teams.

If the team member decides that the flu alert is valid, he/she will assess the flu impact together with relevant members of the clinical governance and communications teams. This group will engage with the public health departments, service providers, national surveillance groups and other interested parties. Healthdirect will work together with other public health entities to implement public health actions. This will occur in accordance with the Healthdirect Australia Emergency Response Plan (HERP). An evaluation of the response will be completed once the response is completed.
Figure 1: Flow chart of operational steps in syndromic surveillance.
6.2 Published plots:

There is 1 plot and the flumeter published to the healthdirect webpage at https://www.healthdirect.gov.au/flu-trends

Figure 2: Influenza-related calls to the Healthdirect Australia telephone helpline, as a percentage of all calls. The grey shading is two standard deviations around the mean of 2010-2016, to indicated the typical flu range. The blue line indicates the influenza-line illness activity in the current year.

Figure 3: The flumeter gives an indication of the current flu risk. The level on the flu meter is automatically calculated using the percentage of ILI calls to the helpline and previous years’ data. This is calculated using the cutoff points of: 0-6% minimal; 6-8% low; 8-10% moderate; 10-12% high; 12% + intense. These cutoff points are demonstrated in the following Figure (Figure 4)

*Current flu risk: Minimal, as of 28 November 2016.*
Figure 4: This plot demonstrates what the FluMeter would have been from 2008 – 2016, using the cutoffs for each risk level. These are: low 6%, minimal 6-8 %, medium 8-10%, high 10-12% and intense 12% +. This is a measure of influenza like illness calls as a percentage of all calls.

There are 3 figures published via Tableau which describe the incidence, severity and age and sex distribution of influenza like-illness in the community. There is also an overview panel which describes whether the flumeter should be updated, and whether to instigate an outbreak alert. These figures and panel are shown below.

Figure 5: Incidence of influenza-like illness calls to the Healthdirect Australia telephone helpline, as a percentage of all calls. The grey shading is the highest and lowest recorded percentage in any year from 2012 – 2016. The blue line indicates the influenza-line illness activity in the current year.
Figure 6: This demonstrates the severity of influenza-like illness calls to the Healthdirect Australia telephone helpline. The pink section shows the percentage of flu calls advised to attend ED: the red, the percentage advised to call 000: the dark grey section the percentage advised to see a doctor, the light grey the percentage advised to provide home care and the very light grey is all other advice given.

Figure 7: The age and sex distribution of callers with influenza-like illness symptoms. Age groups are 0-4, 5-15, 16-44, 45-64 and 65+. 
Figure 8: This figure demonstrates the flu meter levels and also the flu alert. The flu meter levels should be changed according to the level of the “all states” percentage. A flu alert is generated if the level reaches “intense.”

7. **QUALITY CONTROL**

   The syndromic surveillance system will undergo both systematic and once-off quality control. The systematic check will consist of the syndromic surveillance officer confirming the counts of the automatic system for one day that is randomly chosen. These systematic checks will also be supplemented with once-off checks of any health alerts or unusual activity.

8. **STAKEHOLDER ENGAGEMENT**

   An advisory group has been established with senior members from the Commonwealth Department of Health, the NSW Ministry of Health and the WA Department of Health. This group serves as a steering committee and has provided advice regarding creating the syndromic surveillance system.

   We will create a steering committee, which will oversee the operation of the system both for routine surveillance and outbreaks. This will include confirming any outbreak alerts, communicating with public health departments, contributing to state or national public health actions and other relevant tasks.

9. **RISKS**

   The risks involved in running and publishing the syndromic surveillance system are mainly reputational risks. This could occur if the published data is incorrect meaning that public health departments, media and disease surveillance groups might act on or report incorrect data.

10. **LIMITATIONS**

    The limitations of the syndromic surveillance system include data that is broad, meaning that syndrome definitions are not specific. For this reason, syndromic surveillance is subject to error due to patients calling with relevant symptoms that do not have the disease under surveillance. Both telehealth and internet data can be effected by the media, with more people calling or visiting the services directly after a media release.
11. FORMS/TEMPLATES TO BE USED

12. INTERNAL AND EXTERNAL REFERENCES

12.1 Internal References

12.2 External References

13. CHANGE HISTORY

<table>
<thead>
<tr>
<th>SOP no.</th>
<th>Effective Date</th>
<th>Significant Changes</th>
<th>Previous SOP no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial Version</td>
<td></td>
<td>NA</td>
</tr>
</tbody>
</table>

Adapted from CTRG Template SOP Version 2.1
© Copyright: The University of Oxford 2009
Flu season and trends in Australia

It’s more common to catch the flu, or experience flu-like symptoms, in the colder months of the year (April to October). This page provides information about current flu risk levels and trends across Australia based on calls to the healthdirect helpline. It also tells you how to help prevent colds and flu.

Someone with the flu (influenza) may show symptoms that include a cough, sore throat, sinusitis or fever.

Healthdirect Australia collects flu data based on calls to the healthdirect helpline (1800 022 222).

Both the flu risk meter (left) and the graph below show the percentage (%) of flu-related calls.

Here’s how to read our flu risk meter:

- **Minimal** – less than 6 in 100 calls (<6%)
- **Low** – 6 to 8 in every 100 calls (6-8%)
- **Moderate** – 8 to 10 in every 100 calls (8-10%)
- **High** – 10 to 12 in every 100 calls (10-12%)
- **Intense** – greater than 12 in every 100 calls (>12%)

Current flu risk

As of 19 September 2017, the flu risk is moderate across Australia.

This is a good time to get the annual flu vaccination.

Flu trend and forecast

The flu risk usually stays low during the summer months and rises to a moderate or high risk at the peak of the flu season. The risk is only intense in years of unusual flu activity, such as in 2009 when a new influenza strain emerged.

How to read the flu trend graph

When the blue line is within the grey shaded area, it’s a typical flu season. If the blue line rises above the shaded area this may be the first sign of an increase in flu-like illness in Australia.

The data collected from the healthdirect helpline is provided to Australian health departments on a regular basis during the flu season. Results are also published in the Australian Influenza Surveillance Report.

How to avoid the flu

In addition to following good hygiene practices, flu vaccination is a recommended protection against influenza. The best time to get it is early autumn, before the flu season.

Other ways to help prevent flu include:

- keeping surfaces clean
- washing your hands regularly
- avoiding sharing cups and cutlery
- covering your mouth and nose when coughing or sneezing.

Check your symptoms

If you are still concerned about your cold or flu, why not use healthdirect’s Symptom checker to get advice on when to seek medical attention.

The Symptom checker guides you to the next appropriate healthcare steps, whether it's self-care, talking to a health professional, going to a hospital or calling triple zero (000).

Last reviewed: September 2017

24 hour health advice and information you can count on ☎ 1800 022 222

We are a government-funded service, providing quality, approved health information

The following appendix is a copy of the paper submitted to the Communicable Diseases Intelligence on 15 May, 2017. This paper includes research from Chapters 2 and 3. It is currently undergoing revision before publication.
Forewarning healthcare practitioners – how telephone and digital data can predict illness in the community

Abstract - max 250 words

Introduction:

Digital and telehealth data have the potential to improve the timeliness and accuracy of disease surveillance, contributing to an appropriate and timely public health response. We show how these data can forewarn health professionals of rises in influenza-like illness (ILI) in the community.

Methods:

Healthdirect’s online symptom checker and telephone helpline capture demographic details and characteristics of symptoms from users. We compare healthdirect data on ILI with emergency department (ED), laboratory, Flutracking and general practice (GP) data using cross-correlation functions.

Results:

Correlation for the helpline was strongest with ED data (range in yearly correlations from 0.82-0.99), strong with GP data (0.65 – 0.95) and Flutracking data (0.62-0.89), but yearly correlations with laboratories varied (0.49-0.95). Correlation for the online symptom checker was highest with GP data (0.94) and emergency departments (0.94 NSW, 0.92 WA) and high with laboratories (0.88 in Australia, 0.86 NSW) and Flutracking (0.81). Correlation was highest when the helpline and the symptom checker led the laboratory data by one week and the other data sets by zero weeks.

Discussion:

Our analysis demonstrates that both the helpline and symptom checker data are reliable indicators of ILI incidence. An increase in use is likely to occur simultaneously with an increase in visits to GPs and EDs, allowing forewarning of a rise in ILI activity. A surveillance system based on these data will allow health practitioners to receive timely and accurate estimates of the level of ILI in the community.

Introduction:

The last decade has seen a rapid development and adoption of digital technologies that has changed the way we perceive and use health care. Data collected from these technologies can improve the timeliness and accuracy of disease surveillance, contributing to an appropriate and timely public health response.
Healthdirect Australia (HDA) is a government-owned organisation that delivers a range of telehealth and digital services across Australia. The main services include the healthdirect nurse triage helpline (HDAH), which receives over 700,000 calls per year and the online symptom checker which receives over 650,000 visits per year.

Approximately 40% of calls to the HDAH do not visit any other health care provider (1), so these data provide an immediate indicator of syndrome prevalence and a unique insight into community-level disease. This is useful for syndromes with high incidence that are largely self-limiting yet are highly contagious and can impact the health of others particularly the vulnerable.

Australia has many sources of data for ILI disease surveillance, including sentinel general practice networks, emergency departments, laboratories and specialist surveillance or research. However, all surveillance systems from GPs, hospitals or ambulances report only a percentage of the total number of visits in Australia and reports are weekly. In contrast, Healthdirect Australia has data on every call or symptom checker visit, and these data are updated daily.

This paper describes the process of building a syndromic surveillance system using the healthdirect helpline and online symptom checker data. In particular we describe the data analysis required in testing and validating the data, using influenza-like illness (ILI) as an example. The aim of the paper is to determine the validity of the data for ongoing disease surveillance and public health action.

Methods:

To create a syndromic surveillance system that is useful to stakeholders we convened an expert group consisting of senior members from the Commonwealth and jurisdictional health departments. The Group’s role included offering guidance to increase the acceptability of the system and helping to facilitate access to six different data sources which are detailed in Table 1.

The healthdirect nurse triage helpline collects demographic details and characteristics of symptoms from participants as part of routine data collection. The online symptom checker is a digital tool which is available to the public via their computer or mobile device. The tool provides tailored and symptom-specific information and advice to users depending on responses to questions about their symptoms. These data are uploaded to the main database daily and can be queried to identify syndromes. ILI cases for the Healthdirect data were defined as calls triaged to either a “cough” or “colds and flu” clinical guideline for the helpline and “colds and flu” symptom for the online symptom checker. The data were analysed using cross correlation functions and the correlation was calculated for each year individually and for the full time period.

This project was granted ethics approval from the ANU Human Research Ethics Committee number 2016_599, NSW Population and Health Services Research Ethics Committee number 2016_07_646 and the Government of WA Department of Health Human Research Ethics Committee number 2016.39.
Table 1: Summary of data sources analysed including details about the years and locations sampled, type of data, percentage of sample reported and sample size.

<table>
<thead>
<tr>
<th>Name</th>
<th>Data Source</th>
<th>Location</th>
<th>Years</th>
<th>Number/Percent Positive (PP)</th>
<th>percentage of sample reported</th>
<th>Average Yearly Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDA NT</td>
<td>Healthdirect Australia helpline</td>
<td>All states except VIC and QLD</td>
<td>2008-2016</td>
<td>Number and PP</td>
<td>100%</td>
<td>530,000 - 700,000</td>
</tr>
<tr>
<td>HDA SC</td>
<td>Healthdirect online symptom checker</td>
<td>Australia</td>
<td>Nov 2014-2016</td>
<td>Number and PP</td>
<td>100%</td>
<td>Increasing: (340,000 - 650,000)</td>
</tr>
<tr>
<td>Aus Lab</td>
<td>Positive Influenza Laboratory Tests</td>
<td>Australia</td>
<td>2009-2015</td>
<td>Number</td>
<td>varies from 1% - 41%* of all tests</td>
<td>Increasing: (13,000 - 99,000)</td>
</tr>
<tr>
<td>NSW Lab</td>
<td>Positive Influenza Laboratory Tests</td>
<td>NSW</td>
<td>2009-2016</td>
<td>Number and PP</td>
<td>100% of all tests</td>
<td>Increasing (13,000 - 260,000)</td>
</tr>
<tr>
<td>ASPREN</td>
<td>ILI visits to Sentinel GP (syndromic)</td>
<td>Australia</td>
<td>2008-2016</td>
<td>PP</td>
<td>0.7% of all GP visits†</td>
<td>unknown</td>
</tr>
<tr>
<td>FluTracking</td>
<td>ILI self-reported to FluTracking (syndromic)</td>
<td>Australia</td>
<td>2008-2015</td>
<td>Number and PP</td>
<td>85% minimum response rate in sample in 2015†</td>
<td>Increasing (3,700 - 26,500)</td>
</tr>
<tr>
<td>WA ED</td>
<td>ILI visits to Emergency Department (syndromic)</td>
<td>WA</td>
<td>2008-2016</td>
<td>Number and PP</td>
<td>90.9% of all ED visits in WA in 2015‡</td>
<td>Increasing (460,000 – 730,000)</td>
</tr>
<tr>
<td>NSW ED</td>
<td>ILI visits to Emergency Department (syndromic)</td>
<td>NSW</td>
<td>2008-2015</td>
<td>Number and PP</td>
<td>81.5% of all ED visits in NSW in 2015‡</td>
<td>Increasing (1,900,000 2,200,000)</td>
</tr>
</tbody>
</table>

(*), estimate, based on the percentage of all positive tests at the NSW laboratories during that time period. The Aus Lab data source only receives notifications of positive lab results, but no data on how many tests were requested. The NSW Lab data source receives information on both how many positive tests and how many total tests were requested.

(†), estimated by calculating the number of GPs reporting to ASPREN (3) and the total number of GPs (4).

(‡), estimated using the total number of records each year and the AIHW report (5).
**Results:**

We analysed over 5 million healthdirect helpline calls for ILI between July 2008 and December 2016 and over 1 million visits to the online symptom checker for ILI between November 2014 and December 2016. These data showed a very high correlation with other data sources for ILI. The strongest correlation was with emergency department surveillance both in NSW and WA. Sentinel GP surveillance (ASPREN) and self-assessed community reporting (FluTracking) also had high correlation with the Healthdirect helpline, though correlation with laboratory data varied. The online data correlated well with the other data sources, particularly GP and Emergency department surveillance data (Figure 1 and Table 2).

The years that produced the highest and lowest correlation coefficients in the telephone data varied, though 2013 had consistently low correlation with all data sources. The mode lag between healthdirect helpline data and other surveillance data was 0 weeks for every data source except the Australian lab data. The lag in the online symptom checker data was zero weeks except for the two lab data sources.

**Discussion:**

These results indicate that both the healthdirect helpline and symptom checker ILI data correlate very well with other ILI data, particularly that from emergency department surveillance.

Surveillance systems that had a syndromic definition (EDs, GPs and FluTracking) had the best correlation with both healthdirect data sources. The lower correlation with FluTracking and GPs compared to EDs could be due to the smaller sample sizes of these data sources. The online symptom checker and the phone data have similar correlations to the six other data sets in 2015 and 2016. However, as there are only two years of online data, correlation with further years would need to be investigated before drawing strong conclusions for this data source.

The year of best correlation occurred across five different years, and the worst correlation occurred in 2013 for four out of six data sources. This could be due to the nature of the correlation function, which is strongly influenced by its tails. For example, 2009 provided the best fit in two data sources but the worst fit in one. Data sources matched at the peak in 2009, which is significantly larger than the mean and disproportionately increased or decreased the estimate of correlation. By contrast, 2013 in the helpline data had a flat or double peak, as seen in Figure 2, which could lead to lower correlation with data that did not show a similar shape at the peak.

A lag of zero weeks for all data sources except the laboratories shows that an increase in helpline calls or symptom checker visits occurs at the same time as an increase in visits to the ED, GP and self-reported ILI on FluTracking. On the other hand, the increase in calls to healthdirect occurs one week earlier than a positive influenza test, which is what we would expect given the additional time it takes to run the test for influenza in the lab.
Figure 1: The box plot shows the range in correlations between Healthdirect Australia telephone data and other data sources. The range in correlations for each data source is calculated by using Spearman’s, Pearson’s and the cross correlation function individually for each year of data, making either 21 and 24 individual correlation calculations. Data sources are GP Sentinel surveillance (ASPREN), laboratory confirmed influenza in Australia (Aus Lab) and NSW (NSW Lab), Emergency department surveillance in NSW (NSW ED) and WA (WA ED) and FluTracking.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ASPREN</td>
<td>0.65-0.95</td>
<td>0.85</td>
<td>0</td>
<td>2015</td>
<td>2013</td>
<td>0.95</td>
<td>0.95</td>
<td>0</td>
</tr>
<tr>
<td>Aus Lab</td>
<td>0.49 – 0.93</td>
<td>0.79</td>
<td>1</td>
<td>2009</td>
<td>2013</td>
<td>0.90</td>
<td>0.92</td>
<td>1</td>
</tr>
</tbody>
</table>

Correlation between healthdirect helpline and six other data sources
Table 2: Summary of the correlation between Healthdirect Australia telephone and symptom checker data and other data sources, including the range in yearly correlations, the mean correlation, mode of the lag between data sources, and the years of the highest and lowest correlations.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Correlation Range</th>
<th>Mean</th>
<th>Lag</th>
<th>Year of Highest Correlation</th>
<th>Year of Lowest Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FluTracking</td>
<td>0.62 – 0.89</td>
<td>0.79</td>
<td>0</td>
<td>2010</td>
<td>2011</td>
</tr>
<tr>
<td>NSW ED</td>
<td>0.92 – 0.99</td>
<td>0.96</td>
<td>0</td>
<td>2014</td>
<td>2009</td>
</tr>
<tr>
<td>NSW Lab</td>
<td>0.58 – 0.95</td>
<td>0.79</td>
<td>Both 0 and 1</td>
<td>2012</td>
<td>2013</td>
</tr>
<tr>
<td>WA ED</td>
<td>0.82 – 0.96</td>
<td>0.91</td>
<td>0</td>
<td>2009</td>
<td>2013</td>
</tr>
</tbody>
</table>

Table 2: Summary of the correlation between Healthdirect Australia telephone and symptom checker data and other data sources, including the range in yearly correlations, the mean correlation, mode of the lag between data sources, and the years of the highest and lowest correlations.

Figure 2: The black line shows the proportion of calls for influenza like illness as a percentage of all calls to the healthdirect helpline in Australia from 2008 – 2017. The healthdirect helpline is utilised in every state and territory except for Victoria and Queensland.

Implementation and Further Work

We propose presenting the data from the syndromic surveillance system in a way that is accessible to members. These could include local health departments,
jurisdictions, national surveillance groups such as OzFoodNet, universities and other research or surveillance institutions.

These results will be displayed on a secure surveillance dashboard, enabling users to view data, download plots and tables for insertion into reports, and receive alerts when syndromes reach a threshold. Data will include information about the incidence of disease, the severity, temporal and spatial variation and any unusual activity alerts. Healthdirect Australia data has the potential to report on syndromes as diverse as ILI, gastroenteritis, acute fever and rash, adverse reactions to vaccines, acute mental health episodes and drug and alcohol related illness.

This is the first time that a national telephone line in Australia has reported on its contribution to early detection when compared with emergency departments and sentinel GP surveillance. Other studies in Australia have compared telehealth data with lab confirmed influenza with similar results (6). High correlation between telehealth data and lab confirmed influenza has also been observed in the USA (7) and the UK (8). There have also been studies comparing telehealth data to Emergency Department surveillance in Canada (9,10) and to GP presentations in Ireland (11), America (7) and the UK (8). All studies reported high correlation between telehealth data and a more conventional data source, with the telehealth data leading by at least one week.

The high degree of correlation between the HDA data and other data sources show that the HDA data are consistent and trustworthy. These data can now be used to contribute to routine disease surveillance by reporting elevated levels of influenza like illness to emergency departments, public health departments, diseases surveillance groups, researchers, public and other interested parties. This information could be used for staffing, infection control, research, public health messaging, surveillance and other public health actions. We conclude that the proposed Healthdirect Australia syndromic surveillance system can contribute to routine disease surveillance of influenza like illness in Australia.

Acknowledgements:
We would like to acknowledge the contribution of the following people to this paper: members of the HATSS advisory group, consisting of Melissa Irwin (NSW Ministry of Health), Robin Gilmour (NSW Ministry of Health), Christina Bareja (Commonwealth Department of Health), Rachael Corvisy (Commonwealth Department of Health), Donna Mak (WA Department of Health), and the data custodians of our data sources: Craig Dalton (FluTracking), Andrew Puljic (WA Emergency Department Data Collection), Melissa Irwin (NSW Public Health Real-Time Emergency Department Surveillance System), Nigel Stocks (ASPREN), Christina Bareja (National Notifiable Diseases Surveillance System), Robin Gilmour (NSW Laboratory Notifications System).

Authors:
Ms. Mica Hartley, MAE scholar, ANU and Healthdirect Australia
Dr. Janice Biggs, Research and Evaluation Lead, Healthdirect Australia
Mr. Carlo Leonessa, Knowledge and Data Insights Lead, Healthdirect Australia
Dr. Kathryn Glass, Associate Professor, ANU
References:

1. Tran D, Gibson A, Randall D, Havard A, Jorm L. Mapping the outcome of calls to the healthdirect helpline. 2015;(June).
Appendix D

Appendix: Presentations

Over the two years of my MAE program, I presented at 2 conferences, 1 young talent night and prepared a presentation for a colleague to present at 1 further conference. I also presented 2 posters at 1 further conference. These presentations were are in Table D.1. A copy of both posters constitute the following pages in this appendix.
### Oral Presentations

<table>
<thead>
<tr>
<th>Date</th>
<th>Conference</th>
<th>Presentation Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 – 9 August, 2017</td>
<td>Health Infomatics Conference (HIC), Brisbane, Australia</td>
<td>Hartley, M., Araco, M., Biggs, J., Leonessa, C., Glass, K., Predicting and monitoring influenza-like illness using telephone and digital data.</td>
</tr>
<tr>
<td>26 – 28 October, 2017</td>
<td>Royal Australian College of General Practitioners (RACGP) Conference, Sydney, Australia</td>
<td>Hartley, M., Araco, M., Biggs, J., Leonessa, C., Glass, K., Forewarning GPs - how healthdirect helpline data can predict illness in the community.</td>
</tr>
</tbody>
</table>

### Poster Presentations

<table>
<thead>
<tr>
<th>Date</th>
<th>Conference</th>
<th>Presentation Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>26 – 28 June</td>
<td>Communicable Diseases Conference 2017, Melbourne, Australia</td>
<td>Hartley, M., Biggs, J., Leonessa, C., Glass, K., Building a syndromic surveillance system using telephone and digital data.</td>
</tr>
</tbody>
</table>

Table D.1
Forewarning healthcare practitioners: How telephone and digital data can predict illness in the community

Introduction
There is an opportunity to use data collected from telephone and digital health to improve the timeliness and accuracy of disease surveillance contributing to a more appropriate and timely public health response.

Aim
To determine if telephone and digital health data can accurately predict influenza-like illness (ILI) activity in the community and the impact on emergency departments (EDs), General Practices (GPs), and Laboratories.

Methods

Choose data sources

Healthdirect Australia
- Telephone helpline: 700,000 calls/year
- Online symptom checker: 500,000 visits/year

Comparison
- FluTracking
- GPs
- EDs
- Labs

Calculate incidence
- Incidence of ILI cases.
- Aggregated by week.

Calculate correlation
- Between Healthdirect Australia data and each comparison data source.
- Use 3 tests: Pearson’s, Spearman’s and cross correlation.

Results

Conclusions:
- Correlation between data sources is very good.
- Increase in calls/visits occurs simultaneously with an increase in ED and GP visits.
- Increase in calls/visits occurs one week before an increase in lab notifications.

Recommendations
- Build a surveillance system using these data.

Conclusions:
- Telephone and internet data can provide advance warning of increases in ILI.
- Both can contribute to public health surveillance.
- Stakeholders can receive trustworthy, timely and accurate surveillance data.
- Could have many uses, including to inform workforce management, disease surveillance and infection control.

References

Figure 1: Correlation between Healthdirect Australia telephone helpline data and six other data sources: ASPREN, Aus Lab, FluTracker, NSW ED, NSW Lab and WA ED. Spearman’s, Pearson’s and cross correlation functions were calculated for each year. The box chart shows the range across the years and different correlation functions.

Figure 2: Comparing the percentage of influenza-like illness calls to the Healthdirect Australia helpline and the visits to the NSW Emergency Department from 2008 – 2018.

Authors:
Mica Hartley
Janice Biggs
Carlo Leonessa
Kathryn Glass
Introduction

Health-seeking behaviour is rapidly changing with the prevalence of online health information and a focus on consumer-centric healthcare.

Aim

To evaluate a proposed syndromic surveillance system based on telephone and digital data, using influenza-like illness syndrome as an example.

Methods

Sample size of surveillance systems in Australia

For every 10 people who call the Healthdirect helpline, after the call

3 visit an ED
2.5 visit a GP
4 provide self care only
0.5 other*

*other = visiting dentist, mental health practitioner, other healthcare provider, call is not completed, unknown

Data sources

- Telephone helpline
  700,000 calls/ year updated daily
- Online symptom checker
  500,000 visits/ year updated in real time

Data Collected

- Demographic
- Geographic
- Symptoms
- Advice received
- Date and Time

Results, Recommendations and Conclusions

Strengths

Representativeness:
- Represents community level of disease.
- Forty per cent of callers do not visit any other health care provider.
- Large sample size.

Timeliness:
- Telephone data is daily.
- Online data in real time.

Weaknesses

Flexibility:
- New syndromes are limited by the ability to identify cases, due to coarse data.
- However, a large number of syndromes are possible.

Sensitivity:
- Cannot detect all callers or visitors with disease, due to coarse data.
- However, can identify unusual activity.

Recommendations and Conclusions

- Recommend building a surveillance system using these data.
- Proposed surveillance system will have excellent representativeness and timeliness.
- Stakeholders can receive trustworthy, timely and accurate surveillance data.
- Could have many uses, including workforce management, disease surveillance and infection control.

Prototype of the proposed surveillance system using telephone and digital data.


