Statistical approaches to the evaluation of the impact of vaccination programs: a case study exploring rotavirus vaccination in New South Wales

Nicole Mealing

A thesis submitted in fulfilment of the requirements for the

Degree of Doctor of Philosophy

Faculty of Health

The University of Technology Sydney

2018

Certificate of Original Authorship

I, Nicole Mealing, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Health at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by an Australian Government Research Training Program Scholarship.

Signature of Student: Production Note: Signature removed prior to publication. Date: 22/03/2018

Acknowledgements

First and foremost I'd like to thank my supervisors, Professor Andrew Hayen and Associate Professor Anthony Newall. It is a proud moment to see this thesis come together, so thank you for your expertise, direction and feedback over my candidature. Special mention must also go to Dr Robin Turner, not just a good friend but a knowledgeable and supportive biostatistician who offered her skills whenever needed. Thank you especially for your input over this past year, I shall miss our café sessions.

Obtaining data for this thesis proved difficult and time consuming with multiple potential data sources approached. Sincere thanks to Sarah Thackway and her team in the Centre for Epidemiology and Evidence at the NSW Ministry of Health who provided the NSW emergency department and hospital data used in this thesis. These data came from the Emergency Department Data Collection and the Admitted Patient Data Collection maintained by the NSW Ministry of Health. Many thanks to Terrence Flood and Glenys Conner from the Pathology West network who provided the rotavirus pathology data used in this thesis.

This research began at the University of New South Wales and was completed at the University Technology Sydney. It was supported by an Australian Government Research Training Program Scholarship, formerly known as an Australian Postgraduate Award. I was fortunate to attend local and international conferences along with a mathematical modelling course with support from the University of New South Wales Australia Postgraduate Research Support Scheme, University of Technology Sydney's Health Services and Practice Research Development Award and as part of my involvement in an Australian Research Council linkage project grant (LP120200043).

There were many friends who were part of this journey in big and small ways. To Therese, thank you for all the support from start to finish and for editing this thesis. To Cameron, cheers for your advice on extracting the laboratory data into a workable dataset. To Ellen, you were right. Thank you for your friendship, support and the many adventures, even though I didn't listen to you. To Saritha, your patience is incredible. This thesis would not have been completed 'so soon', or possibly even at all, if it wasn't for you. To Mum and Vanessa, for their encouragement and support along the way, especially during an unexpected move. To Bellevue Manor, for always making me feel welcome no matter what state I turned up in. To Eve, it was a big help to share the end journey with you. To my friends from Sydney Summer Series, the Sydney Bush Walkers and the UTS Outdoor Adventure Club, thank you for the many adventures that got me away from the skyscrapers and helped me keep my sanity over these long years.

Finally, thank ye to Brooner. It would have been tougher without you.

Table of Contents

List of Tables

List of Figures

List of Boxes

List of Abbreviations

ACT	Australian Capital Territory
AGE	Acute Gastroenteritis
AIC	Akaike's Information Criterion
APDC	Admitted Patient Data Collection
AR	Autoregressive
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
CI	Confidence Interval
ED	Emergency Department
EDDC	Emergency Department Data Collection
GP	General Practitioner
HBeAg	Hepatitis B Envelope Antigen
HBsAg	Hepatitis B Surface Antigen
ICD-9	International Classification of Diseases Ninth Revision
ICD-10	International Classification of Diseases Tenth Revision
ICD-10-AM	International Classification of Diseases Tenth Revision Australian Modification
MA	Moving Average
NSW	New South Wales
NT	Northern Territory
PCV7	7-valent Pneumococcal Conjugate Vaccine
QQ-plot	Quantile-quantile plot
Qld	Queensland
RCT	Randomised Controlled Trial
RSV	Respiratory Syncytial Virus

RV	Rotavirus
SA	South Australia
SNOMED CT	Systemised Nomenclature of Medicine Clinical Terms
US	United States
Vic	Victoria
WA	Western Australia

Abstract

Vaccination programs can provide an effective means to control infectious diseases at a population level. Evaluating the impact of these programs after implementation can be complicated by limitations of routine surveillance systems and lack of routine testing to confirm diagnosis, as well as natural fluctuations in disease rates over time. This thesis provides a structured explanation of statistical methods and how they can be used to address the epidemiological challenges in assessing changes in burden of disease as a result of vaccination programs. It explores statistical approaches to the evaluation of vaccination programs at a population level, using rotavirus vaccination in New South Wales (NSW), Australia, as a case study.

A summary of the key features that need to be considered when trying to detect any changes in the burden of infectious diseases due to vaccination is given. An assessment of the impact of the Australian rotavirus vaccination program, introduced in 2007, on hospitalisations and emergency department (ED) presentations of children aged under five years for all-cause gastroenteritis in NSW from July 2001 to June 2013 was conducted. Previously published methods that separate these hospitalisations and ED presentations into those due to rotavirus and those that are not are compared. A simulation study is used to explore these methods in controlled scenarios to determine the most appropriate method for these data.

The Australian rotavirus vaccination program had an almost immediate impact and led to a fifty and sixteen percent reduction in the rate of hospitalisations and ED presentations for acute gastroenteritis within the first 2.5 years. These declines were mostly attributed to a decline in the size of the seasonal peak. The methods to determine rotavirus cases from all-cause

gastroenteritis cases had different strengths and limitations and the derived estimates varied. No robust method was identified from the simulation study for our data. Each method that relied on using weekly counts of positive rotavirus laboratory tests to estimate rotavirus-attributable cases underestimated the true number of rotavirus cases when their assumptions held.

The evaluation of the effectiveness of vaccination programs requires the use of rigorous statistical methods to ensure the robustness and validity of findings. Appropriate statistical methods that account for temporal trends are needed to provide a detailed understanding of any changes in disease burden observed. While this thesis focused on rotavirus disease burden in NSW, Australia, many of the concepts discussed are applicable to other infectious diseases.

Chapter 1: Introduction

Vaccines are used to help develop immunity to targeted infectious diseases [1]. The first vaccine was developed in the late 18th Century to protect against smallpox by using material from the pustules of cowpox disease [2]. Since then a number of vaccines have been developed to control infectious diseases, including major diseases such as polio, yellow fever, diphtheria, measles, mumps, rubella, typhoid, rabies, rotavirus, and hepatitis B [1].

Vaccination programs can provide an effective means to control infectious diseases at a population level by inducing immunity in the target population through vaccination. They reduce the spread of infection by giving direct protection to the individuals who receive the vaccine. They can also provide indirect protection to persons who have not received the vaccine, through a process known as herd immunity [3]. Herd immunity is a form of indirect protection that occurs when there are sufficient individuals immune to an infectious disease in a population that it reduces the spread of that infectious disease to individuals who are not immune and are susceptible to the disease. The effectiveness of vaccination programs at a population level will depend on many factors including the uptake of the vaccine in the target population, efficacy of the vaccine and the transmissibility of the infectious disease.

It is important to assess the impact a vaccination program has on the burden of disease after it is implemented to determine the impact of the program. This may be an assessment of the change in total burden of that disease, in regards to cost, mortality, morbidity, and other indicators. It could also be an assessment of the disease burden in specific healthcare settings. For example, any change in the total number of hospitalisations due to the disease of interest. This assessment may reveal herd immunity effects (in non-targeted groups) or vaccine-induced shifts in circulating strains or serotypes of the pathogen. Evaluation of the impact of these programs can be complicated by factors such as the limitations of routine surveillance systems and lack of routine testing to confirm diagnosis of many infectious diseases, as well as natural fluctuations in disease rates over time. Sometimes it is not possible to directly monitor a vaccine preventable disease specifically due to a lack of confirmation of disease (e.g. rotavirus). In such circumstances non-specific disease burden, of which the specific disease of interest is a subset, is often studied instead, which can provide an estimate of the impact of the vaccination program. For example, acute gastroenteritis is often studied to estimate the impact of rotavirus vaccination programs.

Varying statistical methods are used in the literature to assess the impact of vaccination programs at a population level, even when analysing the same pathogens. As we will outline in Chapter 2, many studies have important limitations, for example, not considering pre-existing temporal trends in disease burden prior to implementation of a vaccination program when assessing the program's impact. This inadvertently attributes all of an observed decline (or change) to vaccination efforts; however some of an observed decline may be the result of those pre-existing temporal trends.

This thesis explores statistical approaches to the evaluation of vaccination programs at a population level. It focuses on the use of routinely collected health databases to estimate rates of infectious diseases in the population when vaccination status is unknown. It uses rotavirus vaccination in New South Wales (NSW), Australia, as a case study.

The aims of this thesis are:

- to provide an outline of the broad classes of statistical methods that can be used to estimate changes in burden of disease over time that best account for the characteristics of infectious diseases (Chapters 2 and 4);

- to estimate the impact of the Australian rotavirus vaccination program on acute gastroenteritis hospitalisations and emergency department (ED) presentations (non-specific disease) in children under five years of age in NSW, Australia, using appropriate statistical methods (Chapter 4);
- to directly compare statistical methods that can separate hospitalisations and ED presentations in children under five years of age in NSW for acute gastroenteritis into those that are due to rotavirus and those that are not (Chapters 5 and 6);
- and to estimate the fraction of acute gastroenteritis hospitalisations and ED presentations in children under five years of age in NSW that may be due to rotavirus, and investigate whether this fraction has changed following the introduction of the rotavirus vaccine (Chapters 5 and 6).

1.1 Outline of the thesis

Chapter 2 provides a summary of the key features of infectious diseases and the limitations of routine surveillance and health databases that affect our ability to detect changes in the burden of diseases as a result of vaccination efforts. It highlights some of the main statistical methods published in the literature to assess the impact of these vaccination programs on incidence of disease. It covers five infectious diseases that have established infant immunisation programs and presents a case study review of the statistical methods used to assess the rotavirus vaccination program impact in Australia.

Chapters 3 to 7 focus on rotavirus disease burden in NSW, Australia. Chapter 3 gives an introduction to rotavirus disease and to the Australian rotavirus vaccination program. It describes the three datasets that will be used for the analyses in the remaining chapters of the thesis: hospitalisations, ED presentations and positive rotavirus laboratory results.

Chapter 4 defines in detail the components to fully assess the impact of vaccination programs. Separate methods are explained to i) estimate any changes in temporal trends due to vaccination efforts, and ii) quantify the changes in disease burden if applicable. It then applies these statistical methods to analyse the impact of the Australian rotavirus vaccination program on the disease burden of acute gastroenteritis within hospitals and EDs in children younger than five years of age in the Australian state of NSW.

Chapter 5 directly compares estimates of the fraction of acute gastroenteritis ED presentations and hospitalisations of children under five years of age in NSW that are attributable to rotavirus when four previously published methods are applied. These estimates are calculated over the entire period of observation, before and after vaccination program introduction. The varied results observed and the uncertainty around the true numbers of rotavirus hospitalisations and ED presentations means that it is not clear which of these four methods perform best.

Chapter 6 conducts a simulation study to explore the methods from Chapter 5 in controlled scenarios in order to determine the most appropriate method(s) for estimating the true rotavirus disease burden in NSW. The varied estimates of the fraction of acute gastroenteritis ED presentations and hospitalisations of children under five years of age in NSW stratified by before and after introduction of the Australian rotavirus program are reported for each method considered in the simulation study. No robust method was identified from the simulations conducted and hence we were unable to conclude which method is appropriate for our data. Therefore we did not investigate whether these fractions have changed through rotavirus vaccination efforts. Chapter 7 provides a thesis summary and conclusion.

1.2 Ethics

Ethical approval for this study was obtained from the Human Research Ethics Advisory (HREA) Panel at the University of New South Wales (HC15349) and ratified by the University of Technology Sydney Human Research Ethics Expedited Review Committee (ETH16-0749).

Chapter 2: Assessing the impact of vaccination programs on burden of disease: underlying complexities and statistical methods

Mealing N, Hayen A, Newall AT. Assessing the impact of vaccination programs on burden of disease: underlying complexities and statistical methods. Vaccine 2016. 34(27):3022-9. doi: 10.1016/j.vaccine.2016.04.014. [1]

[1] Nicole Mealing searched the literature, reviewed published articles and drafted this manuscript. Anthony Newall and Andrew Hayen assisted with the literature search and critically revised the draft manuscript.

2.1 Chapter overview

This published review paper [4] provides a summary of the key features of infectious diseases and the limitations of routine surveillance and health databases that affect our ability to detect changes in the burden of diseases as a result of vaccination efforts. It covers five infectious diseases that have established infant immunisation programs: hepatitis B, meningococcal, pneumococcal, rotavirus and varicella disease. It highlights some of the main statistical methods published in the literature to assess the impact of these vaccination programs on incidence of disease and/or related to healthcare use. It then presents a case study review of the statistical methods used in the literature to assess the rotavirus vaccination program impact in Australia. Throughout the article, advice is provided about which statistical methods to apply in certain circumstances (e.g. in the presence of temporal trends) and the limitations of some of the statistical methods used are discussed.

2.2 Introduction

Vaccination programs can provide an effective means to control infectious diseases. They can reduce the spread of infection by conferring direct protection to vaccinated individuals and may also provide herd protection to the population [3]. Before receiving licensure, new vaccines undergo rigorous testing through randomised controlled trials (RCTs), which typically estimate the efficacy of the vaccine against disease [5]. Both the efficacy of a vaccine and its (non-trial) uptake in the target group play a role in the effectiveness of a vaccination program, defined as the change in disease in the population as a result of the vaccination program [6].

Mathematical models can be used to predict the impact of a vaccination program prior to its implementation [7]. These models typically use the vaccine efficacy from RCTs and make assumptions about the efficacy against infection/transmission, vaccine uptake, infection transmission rates and contact patterns [8]. Model predictions are uncertain for many reasons including data limitations, model simplification and the generalisability of trial results to different populations. Therefore, to understand the impact more fully, it is also important to monitor the observed effects of vaccination programs at a population level (in both vaccinated and unvaccinated groups) after implementation. The assessment can be complicated by various factors, such as data availability and unrelated variations in circulation of the infection and resulting disease.

There are four main study designs that assess the effectiveness of implemented vaccination programs [6]. The most common designs are case control studies, in which the vaccination status of infectious cases and their controls are compared over time [9], and ecologic studies, which examine changes in disease burden over time [10]. This thesis focuses on the use of routinely collected health databases to estimate rates of infectious diseases in the population when vaccination status is unknown (since vaccination status is seldom linked to these databases). Even when the vaccination status of individuals is known, studies that measure the overall impact of the program, instead of efficacy, are still important. This chapter outlines the broad classes of statistical methods that can be used to estimate changes in burden of disease over time that best account for the characteristics of infectious diseases.

2.3 Estimation of burden of disease

The ability to identify cases using routinely collected health datasets (e.g. hospitalisations, emergency department (ED) presentations, infectious disease laboratory notifications, and death notifications) is important for monitoring program impact. It may not always be possible to identify all cases using these datasets because of data limitations or the absence of data. For example, not all cases of an infectious disease seek healthcare and are therefore not recorded in a routinely collected system. In other instances, there can be multiple, unlinked admissions for chronic conditions (e.g. hepatitis B), such that a measure of incidence of new infections cannot be provided.

Laboratory testing is the gold standard to confirm diseased cases when clinical confirmation is not specific. However, not all patients may be referred for testing (e.g. in the case of community acquired pneumonia), which complicates the identification of the causative agent for diseased cases. Further, the extent to which identification of cases in a dataset is representative of the underlying disease burden is affected by multiple factors including the type and sensitivity of the test, and the frequency of testing. In addition, the apparent incidence of disease may also be affected by changes to testing patterns over time (e.g. the frequency and eligibility of testing).

Another means to identify cases is through diagnostic codes in administrative datasets, such as International Classification of Diseases (ICD) codes or billing codes [11, 12]. These codes are limited in their use for categorisation purposes, are subject to changes over time and may rely heavily on laboratory testing, which has its own limitations as described above [11] (Table 2.1). For example, the rotavirus ICD-10-AM (Australian Modification) code A08.0 (Rotaviral enteritis) has been shown to be highly specific (98.4%, 95% CI:

Table 2.1*: Key features of some infant immunisation programs that may affect assessment of vaccination impact.*

Infectious disease targeted	Population affected by disease	Seasonality of disease	Confirmation of disease	Efficacy of the vaccine	Spectrum of infectious disease targeted by current vaccines	Herd protection from vaccination
Rotavirus [13-16]	Disease occurs primarily in children under the age of five.	Seasonal trends, typically peaking during the late winter months in temperate climates.	Laboratory confirmation uncommon. Non-specific gastroenteritis within routinely collected data.	High for rotavirus specific disease.	Multiple strains that are the same each year. Monovalent and pentavalent vaccines available. Serotype replacement has not been observed.	Observed after implementation for younger age groups in many countries.
Pneumococcal [14, 17-21]	Invasive disease most common in young children and the elderly. Otitis media (non-invasive disease) primarily in young children.	None known.	Laboratory confirmation common for invasive disease and uncommon for non-invasive disease.	High for vaccine-type invasive disease. Low/moderate for all-cause non-invasive disease (otitis media and pneumonia).	Multiple serotypes that vary over time. Vaccines cover only a sub-set of types (7, 10, or 13). Serotype replacement has been observed in many countries after vaccination.	Observed after implementation for invasive disease across all ages in many countries.
Hepatitis B [14]	Acute infection is highest in young adults. Initial infection is usually asymptomatic in infants. The risk of an initial infection becoming chronic is greatest in those infected as infants.	None known.	Laboratory confirmation common.	High for pre-exposure in adults. High for chronic infection in infants born to HBeAg-positive mothers. Moderate/high for acute infections in infants born to HBsAg or HBeAg mothers.	Double-stranded DNA virus that has multiple genotypes. A recombinant DNA virus is available. Mutations of the hepatitis B genome have been observed from passive and active immunisation.	None observed.
Meningococcal [14, 22, 23]	Disease occurs primarily in children under the age of five; a secondary peak occurs at 15-24 years of age.	Seasonal trends, typically peaking during the winter and spring months.	Usually diagnosed by clinical observation and confirmed by laboratory testing.	Short-term and high for serogroups A and C in older children and adults. Not effective for group C polysaccharide vaccine in children <2 years of age. No data on efficacy of serogroup Y and W135 polysaccharides.	Multiple strains that vary by age group and are the same each year. Vaccines cover a sub-set of serogroups (A, C, Y, W135). Meningococcal C vaccine is recommended for routine immunisation. Serotype replacement has been observed in some studies and not in others.	Observed after implementation for group C meningococcal disease across all ages in some countries.
Varicella [14, 24-29]	Disease occurs primarily in pre-school and early school aged children; and usually by mid adulthood.	Seasonal trends, typically peaking in the winter and spring months, in temperate and some tropical climates. Periodic outbreaks with a cycle of 2 to 5 years have been observed.	Usually diagnosed by clinical observation.	High for varicella. Very high for severe varicella.	Single serotype virus. Vaccine covers a single strain.	Observed for older children and adults in several countries.

HBeAg: Hepatitis B envelope antigen, HBsAG: Hepatitis B surface antigen.

97.5-99.1%) but only moderately sensitive (61.6%, 95% CI: 58.0-65.1%) at identifying all-cause acute gastroenteritis that is due to rotavirus [13]. These estimates were from a single centre and, although testing and coding practices are likely to vary across centres, they highlight the limitation of this code in identifying rotavirus cases. One strategy here may be to also evaluate other ICD codes for acute gastroenteritis hospitalisations where the causative organism is unknown (e.g. ICD-10-AM codes A09 and A08.4) but which are likely to include undiagnosed rotavirus caused disease (see Section 2.4.7 for further discussion on non-specific disease).

2.4 Statistical methods for assessing disease burden over time

There are natural fluctuations in the rates of infectious diseases that can complicate whether observed changes in these rates are a result of vaccination efforts or a natural variation of the data series. These fluctuations may have an underlying pattern to them (e.g. a decreasing trend) or be irregular (i.e. random noise).

Regression methods are widely used to understand trends in data over time. Since the outcome is typically counts or rates of disease burden, a basic analysis would fit a count regression model, such as Poisson regression, with disease burden regressed against time. This models how the logarithm of the expected number of diseased cases (or rates if an offset is included) changes linearly with equal increment increases in time and assumes that the variance of the number of cases is equal to the mean and that the resulting residuals are independent over time [30]. These assumptions may not always be satisfied and alternative options are discussed in Section 2.4.4.

Counts cannot be assumed to be independent when analysing data that has been collected over time, since observations close together in time are likely to be correlated. Autocorrelation is typically used to summarise these serial effects and is a measure of this correlation structure on a standardised scale [31]. If autocorrelation is neglected then the model will bias the estimate or underestimate the uncertainty of the estimate (i.e. the width of the confidence interval). Methods to capture any autocorrelation present are discussed in Sections 2.4.1, 2.4.2 and 2.4.3. Periodicity is the special case where there is an evident systematic pattern to the rates of disease that is correlated with time (e.g. repeating peaks and troughs) and is discussed in Section 2.4.2.

Our ability to identify temporal trends in a data series will be influenced by the surveillance units of the available data (e.g. weekly, monthly, annual). For example, annual data cannot identify seasonality. The time period for which data are available, both before and after the introduction of the vaccination program, is also crucial and the minimum amount of time that is reasonable will depend on the infectious disease being studied and will influence the statistical method applied. For example, varicella has periodic outbreaks so will require a longer time period to capture the underlying temporal trends than other infectious diseases that fluctuate less over time (Table 2.1). Evaluating potential changes in rare outcomes (e.g. deaths attributable to rotavirus in Australia) will require a long length of analytical time in order to detect any changes in the underlying trends, since these will be less discernible.

By modelling the periodicity, autocorrelation and random noise (see Sections 2.4.1-2.4.4) we are able to identify the underlying trend in burden of disease. Then one can test whether this trend has changed as a result of vaccination efforts or not, independently of the other components of the data series (see Section 2.4.5).

2.4.1 Modelling autocorrelation

Time series methods can be used to model autocorrelation in the data series. The simplest models use autoregressive (AR) and moving average (MA) processes [32]. An AR process models the temporal dependence of the data by performing a regression analysis of the disease burden with its values at previous points in time. The number of past observations included determines the order of the AR process. A MA process smooths the data series by calculating a weighted average (usually the most recent data points are given the highest weighting) over a fixed-length window of consecutive time points. The weighted average is repeatedly calculated by shifting this window forward one increment at a time (dropping the earliest time point and including the next consecutive time point). The number of observations to be averaged determines the order of the MA process.

A combination of AR and MA terms can be used to model the temporal dependence of burden of disease, which is referred to as an Autoregressive Moving Average (ARMA) model. If the data are not stationary (i.e. if the mean or variance changes over time) more complicated models, such as an Autoregressive Integrated Moving Average (ARIMA) can be used [32]. An ARIMA model first differences the data before applying the AR and MA terms. For example, Höhle et al. [33] used a multivariate time series model to assess the effectiveness of the varicella vaccination program in Germany (Table 2.2). An ARMA (2,1) model was used to capture the remaining autocorrelation in the data series after first adjusting for seasonality and secular trends.

The presence of autocorrelation in a data series can be visually inspected using sample autocorrelation and partial autocorrelation plots [32]. Partial autocorrelation is the correlation of the rates of disease burden at one point in

Table 2.2: *Key features of non-Australian publications that assess the impact of vaccination programs on burden of disease that are discussed throughout the manuscript.* [33-36]

	Study			
Key features of study	Höhle	Simonsen	Do Carmo	Kupek
Infectious Disease	Varicella	Influenza	Rotavirus	Varicella
Year vaccination program introduced	2004	First introduced in 1945. Vaccine updated annually.	2006	2002
Data sources	Paediatric and GP consultations to a convenience sample of 1176 primary-care, private-practice physicians in Germany (2005-2008).	Adults aged 65 years or over with a death notification for pneumonia or influenza in the United States multiple-cause-of-death database (1968-2001).	Children (<5 years) with a death notification for AGE in Brazil (2002-2009). Children (<5 years) with a hospitalisation for AGE (2002-2009) in Brazilian hospitals that contribute to the electronic Hospital Information System (this constitutes ~70% of all Brazilian hospitalisations).	Varicella notifications in the state of Santa Catarina, Brazil (1997-2007).
Identification of diseased cases	*Observational:* Patients identified with a clinical picture compatible with varicella (presence of skin exanthema and concomitant presentation of papules, blisters, pustules or crusts).	*Observational*: Deaths with a pneumonia, influenza or all-cause mortality ICD-9 (pre-1999) or ICD-10 (1999 onwards) code (exact codes not specified). Influenza deaths with an influenza ICD-9 code (487).	*Observational*: Hospitalisations and deaths with a primary diagnosis field containing any of the AGE ICD-10 codes for diarrheal disease of any cause: A00-A03, A04, A05, A06.0-A06.3, A06.9, A07.0-A07.2, A07.9, A08-A09. No changes occurred to the number of hospitals that reported to this system during the study period.	*Observational:* Notifications for varicella where cases i) had a positive laboratory test result, ii) presented with signs and symptoms compatible with varicella, iii) had an epidemiological link to a case, or iv) any combination of these.
Statistical method to assess disease burden over time	Generalised linear model with log link. The AIC was used to determine the covariates in the final model.	Regression model (parameter estimation method not stated). 5-year age groups were age standardised to adjust for ageing and mortality rates identified using ICD-10 codes were weighted to adjust for changes in reporting from ICD-9 to ICD-10. Observed mortality compared to model predicted mortality is shown in Figure 1 and was good.	Interrupted Poisson regression, which excluded the year the national immunisation for rotavirus was introduced (2006). Scaling was applied to the standard error of the rate ratio to remove the over-dispersion present. Model fit not reported.	Ordinary linear regression. Model fit not reported.
Surveillance Units	Monthly.	Monthly.	Monthly.	Annual.

Testing for changes as a result of vaccination efforts	Tested for a significant linear trend in the mean number of varicella cases per physician i) over time (in months), and ii) over vaccine coverage (in percent). Only data post introduction of the vaccination program was included in these models.	Tested for an interaction with year and age using a linear regression model for the logarithm of excess mortality. No interactions observed. Excess mortality was determined using serfling-like methods by first modelling influenza-attributed mortality cases and then identifying any months where the observed mortality exceeded the upper 95% confidence interval of the predicted mortality.	Tested for an interaction between a binary indicator variable (pre- or during vaccination program era) and the underlying trend of diarrheal i) deaths and ii) hospitalisations over time. Interactions were observed. In addition, the observed counts in the vaccination era (2007-2009) were compared to the expected counts when projecting the pre-vaccination era (2002-2005) trends forward in time using Poisson regression.	Tested for an interaction between a binary indicator variable (pre- or during vaccination program era) and the underlying trend of the annual incidence ratio (of varicella incidence in the state capital (Florianópolis) compared to annual incidence in the rest of the state of Santa Catarina).
Assessment of temporal trends	A linear term for month captured secular trends, two cosinor terms with frequencies of 12 and 6 months modeled seasonality and an ARMA (2,1) model was used to capture the remaining autocorrelation in the data series.	A spline function was used to remove secular trends and then a cosinor term was used to model seasonality (May to November only).	Seasonality was adjusted for by including a binary indicator variable for each month (i.e. 11 indicator variables) and secular trends by including a linear term for calendar year in the model.	A linear term for time (in years) was included to capture secular trends.
Identification of indirect effects	Explored herd immunity effects by simultaneously analysing data from vaccinated and unvaccinated age cohorts using a multinomial logistic regression model (declines observed).	All age cohorts were eligible for the vaccination program. Mortality rates were analysed for each individual age cohort (trends varied by age cohort).	Explored herd immunity effects by analysing data from unvaccinated age cohorts (declines observed).	Explored herd immunity effects by analysing data from unvaccinated age cohorts (declines observed but not statistically significant).
Key results	There was a declining (secular) trend in the mean number of varicella cases per physician in Germany in the years following the introduction of the varicella vaccination program.	Mortality benefits of influenza vaccination were not as large as previously published.	Large reductions in both diarrhoea-related deaths and diarrhoea-related hospitalisations in Brazil following the introduction of rotavirus vaccination.	There was noticeable varicella incidence reduction observed after the introduction of varicella vaccination program in an already declining trend of varicella incidence in Santa Catarina, Brazil.

AGE: Acute gastroenteritis, AIC: Akaike's Information Criterion, ARMA: Autoregressive Moving Average, GP: General Practitioner, ICD-9: International Classification of Diseases Ninth Revision, ICD-10: International Classification of Diseases Tenth Revision, PCV7: 7-valent Pneumococcal Conjugate Vaccine, RSV: Respiratory Syncytial Virus, US: United States.

time with the rates of disease burden of the same disease at other points in time and adjusted for the values of all time points in between (i.e. shorter lags). For a series with little to no autocorrelation, we would expect both of these plots to have no discernible pattern and no statistically significant values (particularly of the lower order lags).

2.4.2 Modelling periodicity

The most well-known example of periodicity is seasonality, where asystematic pattern in the rates of disease is correlated with a component of the calendar year. For example, meningococcal disease, rotavirus and varicella tend to have a peak in incidence in the winter months each year [15, 22, 25, 27, 28] (Table 2.1). Periodicity in the data series is typically identified through plots of the data series over time and visual inspection of the sample autocorrelation and partial autocorrelation functions plots [37]. The simplest approach to model periodicity is to include a combination of sine and cosine functions (i.e. a cosinor term) as an additive term in a regression model [31]. Simonsen et al. [34] used this method to model seasonality while assessing the impact of influenza vaccination on mortality in the United States (Table 2.2). A sine and cosine term were included in the model, which assumed a sinusoidal pattern to the mortality rate (and hence dependency between the observations) that was steady and symmetric. Predicted mortality was plotted against time with observed mortality to assess model fit. If periodicity is not accounted for the standard errors will not be correct and incorrect conclusions may be drawn when assessing if there has been a change in disease burden over time.

If the periodic structure is more complicated (for example, if the timing of the winter peak shifts each year) then more flexible models can be adopted [31, 38, 39]. For example, a generalised additive model adds a (unconstrained) smoothed function of time to the regression model, which aims to capture all the periodic variation in the data (example given in Section 2.5) [31].

2.4.3 Alternative methods to model the correlation structure

Sometimes more basic models will be adequate to model the correlation structure of the data series; however it is difficult to give rules on when this is appropriate. In some instances, specifying the periodicity and underlying trend in the data is sufficient to capture any autocorrelation present. In other cases, the inclusion of fixed or random effects terms can adequately specify the correlation structure. For example, do Carmo et al. [35] fitted a Poisson regression model to assess the impact of the Brazilian rotavirus vaccination program on national diarrhoea related mortality and acute gastroenteritis hospitalisations (Table 2.2). Secular trends were modelled using a linear term for calendar year and seasonality by including a binary indicator variable for each month. These binary variables are fixed effects, meaning that the rates are assumed to be different each month. Random effects could also have been fitted that instead constrained all the monthly estimates to follow some distribution (e.g. Gaussian). If these methods are used then justification needs to be given explaining that these more basic models sufficiently account for any autocorrelation and periodicity. This is often done by checking that the residuals are independent over time (e.g. inspecting residual plots of the standardised Pearson residuals) [30, 32, 37]. Do Carmo et al. do not report if they checked for remaining autocorrelation [35] (Table 2.2).

2.4.4 Modelling disease burden over time using an alternative link function

As discussed at the start of Section 2.4, a Poisson regression model uses a log-link function, meaning that it models how the logarithm of expected counts or rates of disease burden changes linearly with equal increment increases in time (i.e. the link function relates the linear predictor to the mean response). Over-dispersed Poisson regression and Negative Binomial regression models also use the log-link function and may be more appropriate than a Poisson regression model depending on the mean and variance relationship of the data [30]. In addition, there are other link functions that can be used instead of the log-link [40] and there can be debate about which is most appropriate [41] and how to check the adequacy of a chosen link function [42]. The adequacy of a link function may be checked informally (e.g. by examining model diagnostics such as the standardised Pearson residuals against the fitted values) or formally (e.g. Pregibon's goodness of link test), with the appropriate checks to perform dependent on the statistical methods being applied [42, 43]. If the assumptions of the Poisson regression model are not met or the model (that includes all the components of the correlation structure of the data series) does not converge or the log-link assumption does not fit well then a different link function should be considered. For example, linear regression models are a good fit to the Poisson regression model when the counts are not too small or if they are thought to be proportional with time. They are also easier to adjust for complex autocorrelation and periodicity, and the mean and variance are no longer constrained to be equal, all of which assist with model convergence. Whichever link function is selected, it is important to check the adequacy of the model fit to ensure that it is appropriate [30, 37, 40].

2.4.5 Testing for a change in trend as a result of vaccination efforts

After identification of the underlying trend in incidence of disease, independently of any other patterns evident in the data series, one can test whether this trend has changed as a result of vaccination efforts or not. This is often achieved by including a binary indicator variable in the model, which represents before or during the vaccination program era, and then testing for an interaction (i.e., effect modification) between it and the underlying trend. For example, Kupek et al. [36] fitted a linear regression model to the annual incidence ratio of varicella cases to non-cases and tested for an interaction with introduction of the varicella zoster virus vaccine in 2002 in Brazil (Table 2.2). Furthermore, the introduction of a vaccination program can cause shifts in the magnitude of and timing of the peak of a seasonal trend, which may be of interest (e.g. to understand the number of hospital wards required). Again an interaction between the binary variable (pre- and during vaccination program era) and underlying seasonality can be tested to ascertain if there has been a change in the underlying seasonality as a result of vaccination efforts. An alternative approach to including a binary indicator variable is to use vaccination coverage data for individual age cohorts over time [34].

A major limitation in this type of analysis is the inability to attribute causation of the vaccination program to observed declines in rates of disease. One can potentially infer causation, particularly if other data sources and evidence from RCT trials give support to causation (e.g. the vaccine has shown a high efficacy against the outcome) and there is high vaccine uptake. However, if this is the first study to note a decline post-implementation of a vaccination program, more caution is needed in inferring attribution than if a similar finding has been found in a number of other locations already.

Chapter 4 will expand on this sub-section and provide more details on using statistical methods to test for a change in trend as a result of vaccination efforts.

2.4.6 Analysing the indirect effects of vaccination programs

Some vaccination programs result in herd protection, although, these effects can be difficult to predict in advance of implementation, such as for rotavirus vaccination, where a herd impact was not incorporated into many of the pre-implementation modelling assessments [44]. Studies that assess population impact do not typically have access to vaccination status, so they often separate the population into 'vaccinated' or 'unvaccinated' cohorts based on their eligibility for the vaccination program. Then the impact on 'unvaccinated' cohorts is usually assessed using the same statistical methods (e.g. Poisson regression) that were applied to the 'vaccinated' age cohorts [14, 35, 36] (Table 2.1, Table 2.2). Herd immunity effects have been observed for diseases such as varicella, meningococcal disease, invasive pneumococcal disease and rotavirus (Table 2.1). Reductions in disease as a result of herd protection to 'unvaccinated' cohorts are likely to be smaller and therefore harder to detect in routinely collected datasets.

Some vaccination programs may only target a subset of the multiple strains or serotypes of some infectious disease (examples in Table 2.1). When serotype replacement is present, serotype specific data needs to be analysed in order to understand how the vaccination program impacts on serotypes covered by the vaccine and those that are not, which may lead to reductions (i.e. cross-protection) or increases (i.e. replacement) as a result of the vaccination program.

This can impact on future decisions regarding replacement of the vaccine with a new vaccine that covers these serotypes.

2.4.7 Modelling the proportion of non-specific burden of disease that is vaccine preventable

As discussed in Section 2.3, sometimes we can only monitor non-specific disease due to a lack of specific diagnosis codes or laboratory confirmed data. However, it may be of interest to estimate the proportion of non-specific disease burden that is vaccine preventable and whether this proportion has changed over time. This can provide an indirect estimate of the impact of the program on the disease burden. To do this an additional step is required to estimate the non-specific disease burden (e.g. community acquired pneumonia) attributable to the infectious agent (e.g. influenza) over time. In effect this is done by estimating a baseline unrelated to the infectious agent and then examining the 'excess' above this baseline which is estimated to be attributed to the infection of interest [45-47]. The proportion of the disease that is due to the infection can be estimated in a number of ways [45-47]; however, one common method is to examine the relationship (over time) between an infection specific variable (e.g. pathology data) and the non-specific burden. For example, Davey et al. [48] (Table 2.3) examined trends in non-admitted ED presentations for all-cause acute gastroenteritis to estimate the rotavirus-attributable fraction of these presentations using data on rotavirus positive laboratory tests. Chapter 5 will compare four methods to estimate the proportion of acute gastroenteritis hospitalisations and ED presentations of children under five years of age in NSW that are attributable to rotavirus.

***Table 2.3**: Key features of all unique published studies that assess the impact of a national rotavirus vaccination program introduced in 2007 on burden of disease in Australia* [16, 48-53]

Key features of study	Study						
	Lambert (2009)	Buttery (2011)	Clarke (2011)	Macartney (2011)	Dey (2012)	Pendleton (2013)	Davey (2015)
Data sources	Qld RV notifications (2006-2008). Qld laboratory test results (2000-2008).	Children (<18 years) hospitalised for AGE from one major paediatric hospital in each of NSW and Vic and the corresponding laboratory test results (2001-2009). Qld hospitalisations (2001-2009). Qld laboratory testing data (2000-2010). NSW ED presentations for AGE during the winter-spring RV season (2001-2008).	Children (<6 years) hospitalised in SA for AGE (2005-2010).	Children (<18 years) hospitalised for AGE (2001-2009) at a major paediatric hospital in NSW and the corresponding laboratory test results.	Hospitalisations coded as RV or all-cause AGE in NSW, Vic, Qld, WA and SA (2004-2010) or NT (2001-2010).	Hospitalisations for RV, AGE or acute bronchiolitis in children (<5 years) in NSW, ACT or NT (1998-2009).	Non-admitted ED presentations for all-cause AGE in children (<5 years) from 50 NSW public hospital EDs (2003-2011). Positive RV tests from 10 NSW public laboratories (2003-2011).
Identification of diseased cases	*Pathology*: Positive RV antigen (test changed in 2005). *Observational*: Notification to Qld Health for RV.	*Pathology*: Positive RV antigen. (Changes in test not specified.) *Observational*: Hospitalisations with primary diagnosis field containing any of the AGE ICD-10-AM codes: A08.0, A08.2, A08.3, A08.4, A09, A09.0, A09.9.	*Observational*: Hospitalisations with a RV (A08.0) ICD-10-AM code, or an AGE ICD-10-AM code (A01 to A09, including A08.0).	*Pathology*: Positive RV antigen (test changed in 2006).	*Observational*: All hospitalisations with a RV (A08.0) ICD-10-AM code or any other AGE not coded as RV (ICD-10-AM: K52 and A01 to A09, excluding A08.0).	*Observational*: All RV, AGE and acute bronchiolitis associated hospitalisations identified through ICD-10-AM coding (exact codes not specified).	*Pathology*: Positive RV antigen. (Changes in test not specified) *Observational*: Non-admitted ED presentations with an AGE code identified through ICD-9, ICD-10-AM or SNOMED CT.
Statistical method to assess disease burden over time	Descriptive statistics.	Descriptive statistics.	Descriptive statistics.	Descriptive statistics.	Descriptive statistics. CIs reported for incidence per 100,000 population.	Descriptive statistics. CIs reported for incidence per 10,000 child years.	Ordinary linear regression. Model fit was checked using QQ-plots and partial autocorrelation function plots.
Surveillance Units	Yearly (monthly presented in figures).	Yearly (monthly presented in figures).	Two 24-month periods.	Yearly (monthly presented in figures).	Yearly (monthly presented in figures).	Yearly (monthly presented in figures).	Weekly.
Assessment of temporal trends	No adjustment for seasonality or autocorrelation reported. (Figure 2 shows seasonality.)	No adjustment for seasonality or autocorrelation reported. (Figures show seasonality.)	Seasonality adjusted by comparing a 24 month review period immediately before and a 24 month review period a year after the introduction of the vaccination program.	No adjustment for seasonality or autocorrelation reported. (Figure 1 shows seasonality.)	No adjustment for seasonality or autocorrelation reported. (Figure 1 shows seasonality.)	No adjustment for seasonality or autocorrelation reported. (Figures show seasonality.)	Non-parametric cubic spline and a linear term for week were used to adjust for baseline secular trends and observed seasonal variation in AGE ED presentations. Model fit, including residual autocorrelation, was described.

Testing for changes as a result of vaccination efforts	Presented the percentage change in the number of RV notifications, tests performed and positive tests in each year after the introduction of the vaccination program, as compared with the average beforehand. No tests of significance or CIs reported.	Presented the number of RV hospitalisations and positive test results, in each year after the introduction of the vaccination program, as compared with the average beforehand. Statistical significance statements are included, but no method is stated for the calculation.	Presented the percentage reduction in RV and AGE hospitalisations, and the proportion of AGE hospitalisations attributed to RV during a 24 month period prior to and a 24 month period after the introduction of the vaccination program. No CIs reported. P-values presented but no method is stated for the calculation.	Presented the number of RV hospitalisations in each year after the introduction of the vaccination program, as compared with the average beforehand. No tests of significance or CIs reported.	Presented the change in the rate of RV and non-RV coded hospitalisations in each year after the introduction of the vaccination program, as compared with the average beforehand, and stratified by Indigenous status. No tests of significance performed.	Presented the change in incidence of RV, AGE and acute bronchiolitis coded hospitalisations in each year after the introduction of the vaccination program, as compared with the average beforehand. No tests of significance performed.	Tested for an interaction between a binary indicator variable (pre- or during vaccination program era) and the underlying trend of i) AGE, ii) RV-attributable, and iii) non-RV attributable ED presentations over time. Non-admitted AGE ED presentations attributable to RV was determined using the weekly rate of RV positive laboratory tests over time in a generalised additive model. Interactions were observed.
Identification of indirect effects	Explored herd immunity effects by analysing data from unvaccinated age cohorts (declines observed).	Explored herd immunity effects by analysing data from unvaccinated age cohorts (declines observed).	Explored herd immunity effects by analysing data from unvaccinated ae cohorts (declines observed).	Explored herd immunity effects by analysing data from unvaccinated age cohorts (declines observed).	Explored herd immunity effects by analysing data from unvaccinated age cohorts (declines observed).	Explored herd immunity effects by analysing data from unvaccinated age cohorts (declines observed).	All age cohorts included were eligible for the vaccination program and data was analysed by year of age (declines observed).
Key results	RV notifications and the proportion of positive RV tests were reduced following vaccination efforts. Minimal reduction in the number of laboratory tests was observed.	Hospital admissions for both RV and non-RV-coded AGE were reduced in all states following vaccination efforts. Marked reductions in AGE ED presentations and short-stay unit admissions were also observed.	Hospital admissions for both RV and AGE were reduced in children aged less than 6 years following vaccination efforts. Children less than two years demonstrated the greatest reduction in hospitalisations.	Hospital admissions for RV were reduced and the number of laboratory tests performed increased following vaccination efforts.	Hospital admissions for both RV and non-RV AGE were reduced in children aged less than 5 years following vaccination efforts. Decreases in hospitalisations of Indigenous children were smaller than those for the general population.	The average incidence of both RV and AGE were reduced in children aged less than 12 months following vaccination efforts.	The average weekly rate of ED presentations for all-cause AGE and the proportion of AGE presentations attributable to RV were reduced following vaccination efforts.

ACT: Australian Capital Territory, AGE: Acute Gastroenteritis, CI: Confidence interval, ED: Emergency Department, ICD-10-AM: International Classification of Diseases 10th revision Australian Modification, NSW: New South Wales, NT: Northern Territory, QQ-plot: Quantile-Quantile plot, Qld: Queensland, RV: Rotavirus, SA: South Australia, SNOMED: Systemised Nomenclature of Medicine Clinical Terms, Vic: Victoria, WA: Western Australia.

2.5 Statistical methods used in the literature for assessing disease incidence over time

As an example, we reviewed all the statistical methods used in the literature to assess the impact of the rotavirus vaccination program in Australia. There have been seven unique studies published to date that assess the impact of the Australian rotavirus immunisation program and all reported a decline in rotavirus disease burden after the introduction of the program nationally in 2007 [16, 48-53] (Table 2.3). Six of these studies compared descriptive statistics in the pre- and post- vaccination program introduction only. These six studies all had access to monthly rates, as shown by the figures included, but only presented results for annualised rates. Furthermore, the figures showed seasonality; however five studies did not report any methods to control for seasonality or autocorrelation. The sixth restricted the analytic period to a 24-month window prior to the introduction of the program and a 24-month window one year after its implementation (stating this was to minimise the effects of seasonality evident in the data) [51]. The seventh used ordinary linear regression and weekly rates to assess the impact of the rotavirus vaccine on ED presentations [48]. The correlation structure of the data was accounted for using a generalised additive model, where a non-parametric cubic spline and a linear term for week were included to adjust for baseline secular trends and observed seasonal variation in the acute gastroenteritis ED presentations. Model fit was checked by assessing the quantile-quantile plots (QQ-plots) (the linearity indicated that the assumption of normally distributed residuals is reasonable) and partial autocorrelation function plots (minor autocorrelation remained).

Similarly, a systematic review of the impact of varicella vaccination programs on the incidence of herpes-zoster found a mixture of different methods were applied including simple descriptive statistics, Poisson regression, negative Binomial regression, linear regression and chi-squared tests for trend [54].

2.6 Discussion

The statistical assessment of changes in burden of infectious disease as a result of vaccination programs using observational or ecologic study designs has not previously been discussed in depth to our knowledge. Our article does not cover all methods or all challenges; however it provides an overview of the core techniques to meet the main epidemiological challenges.

Our review of rotavirus vaccination in Australia showed variation in the methods used in studies for assessing the impact of vaccination programs (Table 2.3), as has also been seen in a systematic review of herpes zoster virus [54]. The choice of the methods to be used in a study will be driven by the data available. However, a number of these statistical approaches had methodological limitations, (e.g. five of the seven studies only used descriptive statistics, even though seasonality was present, and additional (monthly) data were available) (Table 2.3), suggesting that further clarification of the most appropriate statistical methods in assessing the impact of vaccination programs is required. For some analyses there are often multiple methods that can be used, but it is important for authors to justify the method chosen. A detailed discussion of statistical methods for the prospective detection of infectious disease outbreaks is given in Unkel et al. [55], which covers some of the concepts discussed in our article but not others.

The underlying patterns in rates of infectious diseases must be considered so that the most appropriate statistical method is chosen. Changes in descriptive statistics from the pre- to post-vaccination program introduction era are not appropriate in the presence of temporal trends, since this method could result in inadvertently attributing a decline to the program that is really the result of other underlying patterns over time. However, this method has often been used. Linear and Poisson regression models can account for autocorrelation if appropriate methods and variables are included in the model and the model assumptions are satisfied.

All Australian studies were able to show a decline in the incidence of rotavirus disease, despite six of these seven studies not accounting for the seasonality (Table 2.3). This is likely due to the fact that both rotavirus vaccines available in Australia are highly efficacious, with immediate and high uptake, and rotavirus was reasonably common [56]. This means that the decline in rotavirus-attributable health care encounters was likely of a higher magnitude than the inherent variation and temporal tendencies of the rotavirus series. Using complex methods can lead to a more detailed understanding in any changes in disease burden observed. For example, it may identify shifts in peak incidence rates, which informs burden on health care facilities.

It is important for governments and policy makers to evaluate the effectiveness of vaccination programs after their introduction. This assessment requires the use of rigorous statistical methods to ensure the robustness and validity of findings. We provide a structured explanation of statistical methods and how they can be used to address the epidemiological challenges in assessing changes in burden of disease as a result of vaccination programs.

2.7 Chapter summary

This chapter covered a number of infectious diseases and highlights that the underlying complexities in estimating the burden of diseases is consistent across varying pathogens. The remainder of this thesis will focus on rotavirus and acute gastroenteritis (of which, rotavirus is a major cause). However, a number of the issues discussed and methodologies applied throughout the following chapters are applicable to the assessment of diseases other than rotavirus as well.

The next chapter gives a background of the Australian rotavirus vaccination program and outlines the datasets that will be used for the analyses in the remaining chapters of this thesis.

Chapter 3: Datasets for estimation of rotavirus and acute gastroenteritis disease burden in New South Wales, Australia using routinely collected data

3.1 Chapter overview

The remaining chapters (Chapters 4-6) in this thesis, excluding the concluding chapter, will focus on the burden of rotavirus in NSW, a common virus causing acute gastroenteritis in young children. This chapter gives a summary of rotavirus disease and the Australian rotavirus vaccination program. It outlines the datasets that will be used for all subsequent analyses in this thesis. These analyses will focus explicitly on rotavirus and acute gastroenteritis hospitalisations and ED presentations in NSW in children aged less than five years old.

3.2 Rotavirus disease

Rotavirus is the leading cause of severe dehydrating diarrhoea in young children and infants throughout the world [57]. The World Health Organization estimated that 215,000 children under five years of age died from rotavirus worldwide in 2013 [58]. India, Nigeria, Pakistan and the Democratic Republic of Congo accounted for approximately half of these deaths [58].

Rotavirus is primarily spread by faecal-oral transmission [59]. While it can affect people of all ages, it typically affects infants and young children under five years of age [15]. Temperate climates usually experience seasonal variation in rotavirus infections, with the highest rates occurring in late winter [60].

Infection can be asymptomatic, which is often the case for infants under 3 months of age [15]. First infections for those aged 3 months or older are associated with mild to severe diarrhoea [61, 62]. Other rotavirus symptoms include fever, nausea and vomiting. Most children are infected in their first few years of life and their first infection is the most likely to cause severe diarrhoea and dehydration [61, 62]. Infection often reduces the probability and severity of future infections [61].

3.3 Rotavirus vaccines

The first rotavirus vaccine to be licensed was RotaShield in the United States of America in mid-1998 for universal use in infants [63]. It was not licensed in any other country and was voluntarily withdrawn from the market by the manufacturer before the end of 1999 due to its apparent association with intussusception, which was later confirmed [64]. In 2006 two oral, live-attenuated vaccines, RotaTeq and Rotarix, were licenced for use in Australia, the United States of America and several other countries [15, 65]. Rotarix is a human monovalent (i.e. single-strain) vaccine that requires two doses at two and four months of age [66]. RotaTeq is a pentavalent human-bovine vaccine that requires three doses at two, four and six months of age [67]. Both vaccines had similar RCT efficacy against rotavirus [66, 67].

3.4 Rotavirus in Australia

Rotavirus is a seasonal disease in temperate Australia that typically peaks in late winter (August) [14]. In the northern tropical and arid regions of Australia, particularly in remote Indigenous communities, there is no consistent seasonal pattern and epidemics do occur [68, 69].

Population level rotavirus disease burden is difficult to determine in Australia, as described partially in Section 2.3. Rotavirus can only be confirmed through laboratory testing as symptoms are non-specific [70]. However, laboratory testing is not routinely performed in Australia [48, 50]. Rotavirus cases are often coded as acute gastroenteritis in hospitals and other settings since rotavirus is one of several viruses that cause this disease [13].

Acute gastroenteritis refers to illnesses of the gastrointestinal tract that result in acute diarrhoea. It may be coupled with additional non-specific symptoms such as nausea, vomiting, abdominal pain and fever. The average rate of hospital admission for acute gastroenteritis in Australia for children under five years of age was 14.9 (95% CI: 14.7-15.1) per 1,000 children under five years of age from July 1998 to June 2003 [56]. This annual rate was higher for children less than one year of age (21.0 per 1,000 population; 95% CI: 20.4-21.6), and aged one to two years old (24.8 per 1,000 population; 95% CI: 24.2-25.5) [56].

As mentioned in Section 2.3, a means to identify rotavirus cases is through diagnostic codes in administrative datasets. The average annual rate of hospital admission for rotavirus as a principal diagnosis in Australia in children under five years of age was 259.9 per 100,000 children less than five years of age from July 2005 to June 2007 [14]. This average rate was 11.4, 0.4, 0.3 and 1.5 per 100,000 population for persons aged 5-14, 15-24, 25-59, and 60 and over, respectively [14]. Prior to the introduction of the rotavirus vaccine in Australia, Indigenous infants and children were three to five times more likely to be hospitalised for rotavirus than non-Indigenous infants and children [69].

One estimate, prior to the introduction of the two rotavirus vaccines, indicated that 52% of hospitalisations, 42% of ED presentations and 39% of general practitioner visits for acute gastroenteritis were attributable to rotavirus in children under five years of age [56]. On average, there was 1 death per year in Australia attributed to rotavirus prior to the introduction of the vaccination program [71]. Since introduction of the Australian rotavirus vaccination program marked declines in acute gastroenteritis hospitalisations and ED presentations have been observed [16, 48-53].

3.5 Rotavirus vaccination in Australia

Both RotaTeq and Rotarix were licensed for use in Australia in early 2006 and were available from the private market [72]. Vaccination against rotavirus was included on the Australian National Immunisation Program from 1 July 2007 for infants born from 1 May 2007 [65]. The exception to this program introduction was in the Northern Territory where rotavirus funded vaccination began in October 2006 due to the high hospitalisation rates, particularly in Indigenous infants and children, in that territory [16]. Each Australian state and territory chose which vaccine (RotaTeq or Rotarix) to introduce in their state or territory. Rotarix was chosen in the Australian Capital Territory, New South Wales, Northern Territory, Tasmania and Western Australia. RotaTeq was chosen for the remaining states Queensland, South Australia and Victoria. In May 2009 Western Australia changed to RotaTeq. No catch up vaccination program was offered due to upper age limits on vaccine doses [65].

Uptake of the vaccines was immediate and approximately 80% of infants under 12 months of age had received a full rotavirus vaccine course in Australia by 2008 [73], 85% by 2010 [74] and 90% by 2013 [75].

3.6 Data sources

There are a number of population level surveillance datasets available to monitor infectious disease at a population level in Australia, for example hospital and ED data. These routinely collected data usually code rotavirus presentations as gastroenteritis. This thesis will use combinations of hospital, ED and laboratory data from the Australian state of NSW to investigate its aims in children under five years of age. This aggregated age group was chosen because the majority of rotavirus disease burden occurs in under five year olds and to ensure sufficient counts to analyse weekly data, which is important for assessing temporal trends.

3.6.1 Hospital data

Hospital data were obtained from the NSW Ministry of Health from the Admitted Patient Data Collection (APDC) [76]. This is a routinely collected dataset of all admitted patient services provided by public hospitals, public psychiatric hospitals, multi-purpose services, private hospitals and private day procedure centres in NSW. Reporting to this collection is a requirement under the Health Administration Act 1982 for public hospitals, and the Private Hospitals and Day Procedures Centres Act 1988 and Section 44(1) Health Insurance Act 1973 for private hospitals. The APDC excludes all services provided on non-admitted patients as well as stillbirths, hospital boarders, posthumous organ procurement, community residential clients and residential aged care clients in beds funded directly by the Australian Government. Each row of this data collection is an episode of care in hospital. An episode of care ends when a patient is discharged, transferred (within or between hospitals) or dies. Therefore, one period of hospitalisation can have many episodes of care.

Diagnosis information is coded by trained clinical information managers who choose diagnoses from the International Classification of Diseases using the tenth revision Australian modification (ICD-10-AM) [77]. There is an ICD-10-AM code specifically for rotavirus, A08.0 Rotaviral enteritis. Hospitalisations typically only receive the A08.0 code if there is confirmation of rotavirus infection, but not all patients with acute gastroenteritis are tested. This code has been shown to underreport the true burden of rotavirus hospitalisations in a review of all acute gastroenteritis presentations at The Children's Hospital, Westmead, Sydney from 2000-2009 [13]. The ICD-10-AM codes for acute gastroenteritis (A01-A09, K52) should capture all primary cases of rotavirus [52].

The final hospital dataset used in this thesis included weekly counts of all children under the age of five years that were hospitalised in NSW with a primary diagnosis of acute gastroenteritis (ICD-10-AM: A01-A09, K52) from 1 July 2001 to 30 June 2013. Data provided prior to 1 July 2001 were excluded since it used ICD-9-AM coding. Only hospitalisations for which the primary diagnosis was acute gastroenteritis were included.

3.6.2 Emergency department data

Emergency department data used in this thesis were obtained from the NSW Ministry of Health from the NSW Emergency Department Data Collection (EDDC) [76]. This is a routinely collected dataset of presentations to an ED in NSW; but not all hospitals are included in the collection. There are approximately 150 EDs in NSW, and when the EDDC formally began in July

1996 there were 46 EDs providing data. In 2010 the number of reporting EDs had increased to around 90, but this varies over time. All major EDs participate in the EDDC. A small number of participating hospitals have had limited completeness of diagnosis entry for some years or have missed reporting periods altogether.

Unlike the hospital data, the ED data has diagnoses recorded by medical, nursing or clerical personnel at the point of care. These personnel are not trained in clinical coding and the diagnoses are selected by keyword searching or tables of a limited set of diagnoses. There are several different computer programs used in NSW EDs. Different programs use different classifications to record the diagnosis, including ICD-9, ICD-10, or Systemised Nomenclature of Medicine Clinical Terms (SNOMED CT) [11, 78]. As a result there may be variation in diagnosis coding practices across EDs and some disease categories are not available in some programs but may be in others. The NSW Ministry of Health provides a data quality statistic, Good Diagnosis Coding, for each hospital and year. Good Diagnosis Coding status is assigned when i) each monthly total ED visit count varies from the average monthly count by less than 50%, ii) monthly diagnosis completion is greater than 60%; and iii) annual ICD-9, ICD-10 or SNOMED CT diagnosis completion is greater than 75%.

One cannot distinguish between acute gastroenteritis due to rotavirus and acute gastroenteritis that is not due to rotavirus in the ED data [48]. Firstly, the data captured by this system is primarily focused on presenting symptoms more so than on actual diagnoses. Secondly, these are non-admitted ED presentations and laboratory test results may not be commonly performed nor available before the end of the ED presentation.

The final ED dataset included weekly counts of all non-admitted children under the age of five years that presented to a NSW ED department with a primary diagnosis for all cause gastroenteritis from 1 July 2001 to 30 June 2013. Only those public hospitals that provided data and with a positive Good Diagnostic Coding status across the entire study period were included (n=43). These hospitals account for over 70% of all ED presentations in NSW.

A highly sensitive definition of acute gastroenteritis was used based on recommendations from a group at the NSW Ministry of Health that had recently conducted a study on rotavirus disease in NSW EDs over time [48]. This definition aimed to identify all primary diagnoses for gastroenteritis or symptoms of gastroenteritis (e.g. diarrhoea, nausea, vomiting). The following codes were used to identify all cause gastroenteritis ED presentations across the three different diagnosis classification systems used by EDs in NSW:

- ICD-9: 001-003.0, 003.9-005.0, 005.2-009, 078.82, 535.0, 535.5, 536.2, 558, 564.5, 578.0, 787.0, 787.91,787.7
- ICD-10: A00-A05.0, A05.2-A09, K29.0-K29.1, K29.7, K52-52.9, K59.1-K59.2, R11, K92.0
- Systemized Nomenclature of Medicine concept identifiers, as used by the NSW Ministry of Health, are listed in Appendix A.

3.6.3 Pathology data

Pathology data were obtained from Pathology West and contained the test result of all rotavirus tests carried out by the 26 laboratories within their network within the Western Sydney region. This is one of the largest pathology networks in the state of NSW. These laboratory tests were performed across multiple health care providers (i.e. emergency departments, general practitioners, hospitals, etc.). Even though it is not routinely performed on all acute gastroenteritis cases, this data source provides an oversight to the amount of rotavirus circulating in the community at any given time.

The rotavirus test changed from enzyme immunoassays to Rapid Antigen Testing in June 2011. The data management system changed at the end of 2009. There were some zero weekly counts of positive test results, which occurred primarily in the months of January to June.

The final pathology dataset included weekly counts of both positive and negative rotavirus test results for children under the age of five years performed within the Pathology West network from 1 July 2001 to 30 June 2013.

3.6.4 NSW Population data

Weekly population estimates of the total number of children aged under five years in NSW were calculated by linear interpolation of the summation of mid-year estimated resident populations for 0, 1, 2, 3 and 4 year olds produced by the Australian Bureau of Statistics [79]. These population estimates were used to convert the weekly hospital and emergency department counts into weekly rates by dividing weekly counts by the respective interpolated weekly NSW population. There were an estimated 434,943 children under five years of age living in NSW at the start of the study period, 1 July 2001. This estimated resident population was lower at the time of the 2006 census and higher at the 2011 census. With linear interpolation, there were an estimated 443,715 children aged under five years when the national rotavirus program commenced on 1 July 2007 and 486,633 children at the end of the study period, 30 June 2013.

3.7 Software

All analyses were carried out using SAS version 9.4 and SAS University editions 9.3 and 9.4 software [80].

3.8 Chapter summary

The following three chapters of this thesis focus on analysing the impact of Australian rotavirus vaccination program on hospitalisations and ED presentations in children under five years of age in NSW. The final datasets described in this chapter will be used in the analyses contained in chapters 4, 5 and 6.

Chapter 4: Impact of the Australian rotavirus vaccination program on acute gastroenteritis burden in NSW

4.1 Chapter overview

This chapter begins by expanding on sub-section 2.4.5 and details how to assess the impact of vaccination programs and test for a change in temporal trends in more detail. Separate methods are explained to i) estimate any changes in temporal trends due to vaccination efforts, and ii) quantify any identified changes in disease burden. Concepts are explained in the introduction section of this chapter (Section 4.2) using hypothetical scenarios. The chapter then applies appropriate statistical methods to analyse the impact of the Australian rotavirus vaccination program on the disease burden of acute gastroenteritis within hospitals and EDs in children under five years of age in NSW, Australia. Chapters 5 and 6 will investigate methods to estimate the rotavirus specific disease burden of these hospitalisations and ED presentations.

4.2 Introduction

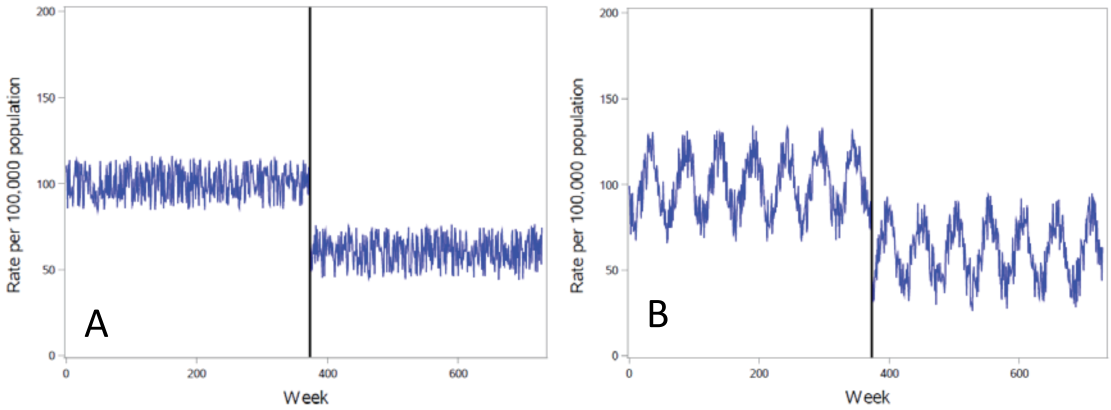
For health policy making, there is often a focus on the burden of an infectious disease with the goal to reduce this disease burden. This interest may be in the total burden of that disease or in the disease burden in particular healthcare settings (e.g. total number of hospitalisations due to that disease). The other key components of interest about an infectious disease are whether there is a general increasing or decreasing trend in the disease burden over time (i.e. the underlying trend) and whether there are any repeating periodic surges in diseased cases (i.e. periodicity). This knowledge can allow health policy makers to understand the disease burden and allocate resources appropriately (where available) as well as work on strategies to reduce this disease burden.

With the mass roll out of the Australian rotavirus program from 1 July 2007 any assessment of its impact is essentially a before and after observational study. This is because there is no equivalent group that did not receive the intervention after 1 July 2007 (i.e. there is no control group) to compare directly with the intervention group. In this scenario, how should the impact of this vaccination program be measured? This is one of the fundamental questions this thesis will answer.

If the number of rotavirus cases was stable (i.e. not increasing or decreasing in both the pre-vaccination and the vaccination era and the magnitude and frequency of periodicity was constant in both time periods), then a comparison of the mean number or rate of rotavirus cases prior to vaccination efforts can be compared with the mean number or rate in the vaccination era using an independent samples t-test (or equivalent Poisson test) [81]. This is equivalent to fitting a simple linear regression model with a binary explanatory variable that represents before or during the vaccination program era. For example, Figure 4.1 shows two hypothetical scenarios where it is appropriate to perform such a test to evaluate the impact of a vaccination program. Panel A shows a stable disease burden over time where an immediate reduction (i.e. step reduction) in disease burden (e.g. the number of hospitalisations) occurs with introduction of a vaccination program. Panel B is similar to Panel A but with additional periodicity that remains constant in the size and timing of the peaks over time.

In this simple hypothetical scenario, the reduction in the mean disease burden can be attributed to the vaccination program (e.g. mean weekly rate reduction of 40 hospitalisations per 100,000 population). As discussed in Section 2.4.5, a major limitation in this type of analysis is the inability to attribute

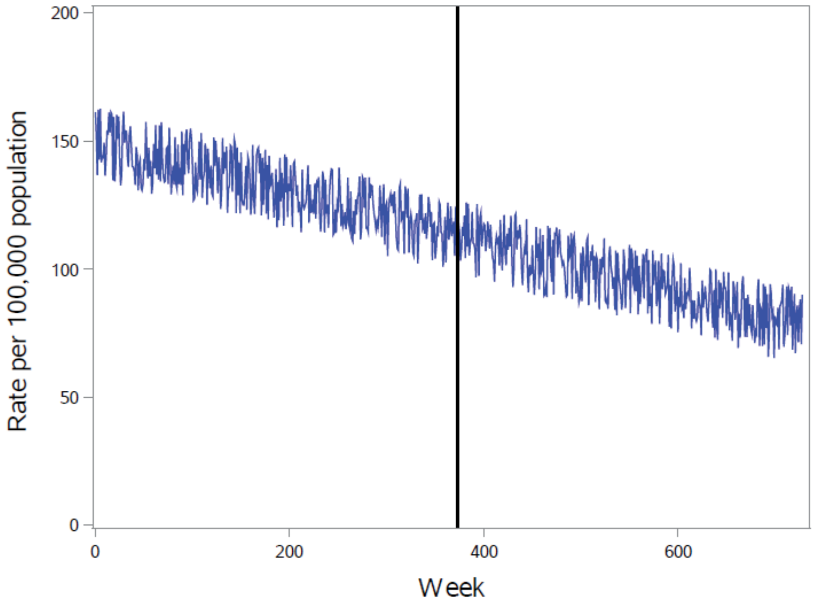
Figure 4.1: *Hypothetical scenarios where an independent samples t-test can be used to test for an impact of the vaccination program. Panel A shows a stable disease burden over time where an immediate reduction in disease burden occurs with introduction of the vaccination program. Panel B is similar to Panel A but with additional periodicity that remains constant in the size and timing of the peaks over time. The vertical line represents introduction of the vaccination program.*



causation of the vaccination program to observed declines in rates of disease. There may be other reasons for this reduction that coincide with the timing of the introduction of the vaccination program that need to be considered.

This simple comparison of the mean rate in the pre-vaccination program era to the mean rate in the vaccination program era is not appropriate when the disease burden is not stable across time. For example, assume that the rate of rotavirus hospitalisations was 150 per 100,000 population in the first week of a fourteen year study period and there was a consistent, decreasing trend in this rate in the pre-vaccination era of 0.1 per 100,000 population per week that was not affected by vaccination efforts (introduced at the end of the seventh year) (Figure 4.2) and that there were steady random fluctuation in these disease rates over time. If the mean weekly rate of rotavirus hospitalisations in the pre-vaccination era of 131 per 100,000 population (95% CI: 130-133 per 100,000 population) is compared with the mean weekly rate of rotavirus hospitalisations

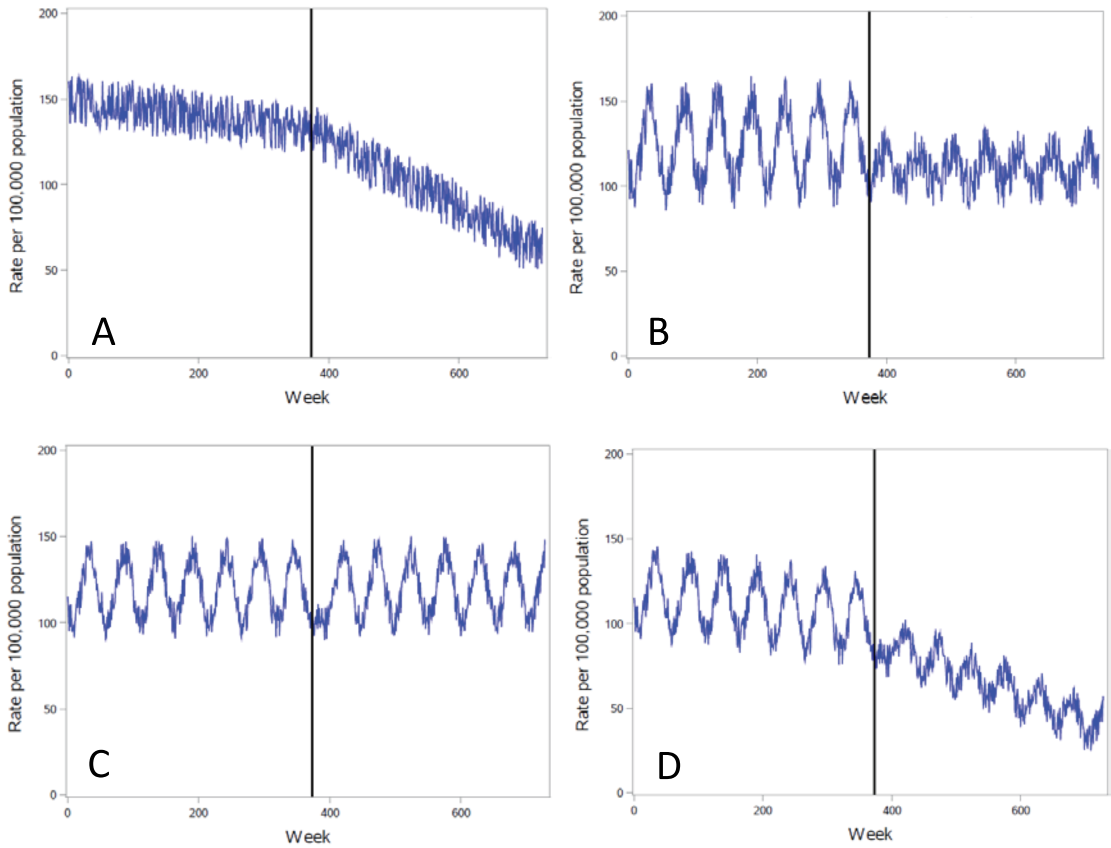
Figure 4.2: *Hypothetical scenario where the rate of rotavirus hospitalisations was 150 per 100,000 population in the first week and there was a decreasing weekly trend in this rate of 0.1 per 100,000 population that was not affected by vaccination efforts. The vertical line represents introduction of the vaccination program.*



in the vaccination era of 96 per 100,000 population (95% CI: 94-97 per 100,000 population), then a reduction is unduly attributed to the vaccination program of 36 per 100,000 population fewer rotavirus hospitalisations per week ($p<0.001$).

While this is an extreme example, it is easy to extrapolate that if you add non-constant periodicity as well, then these simple methods are not appropriate. Despite this, a review of the seven unique published articles that have analysed the impact of the Australian rotavirus program (Section 2.5 and Table 2.3) showed that five of these studies only compared the mean annual rate (or percentage) of disease burden in the pre-vaccination era to the observed rate in each year after introduction of the Australian rotavirus vaccination program and some did not perform a statistical test [16, 49, 50, 52, 53]. However, each of these five studies published figures that showed non-steady disease burden was

Figure 4.3: *Hypothetical scenario where vaccination efforts cause the underlying trend to change (Panel A), the periodicity to change (Panel B), the periodicity to shift (Panel C), or all of the above (Panel D). The vertical line represents introduction of the vaccination program.*



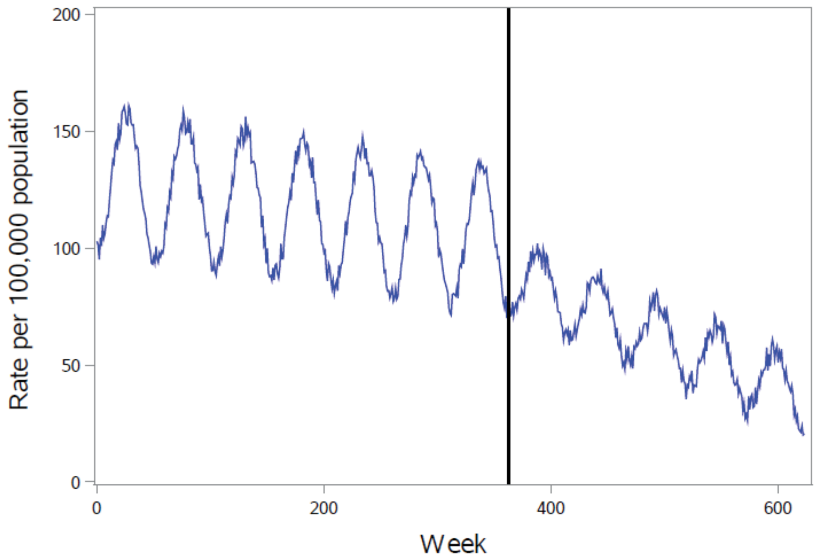
apparent in the datasets, particularly with changes in the size of the seasonal peaks.

When a vaccination program is introduced there is potential for both the underlying trend and the periodicity to be affected, which can directly affect the burden of disease estimated to be prevented. For example, an already declining trend could decline faster (Figure 4.3, Panel A) or the magnitude of the seasonal peaks may decline (Figure 4.3, Panel B) or seasonal peaks may shift (Figure 4.3, Panel C) or a combination of all three (Figure 4.3, Panel D). There are many ways in which a vaccination program can affect the underlying trend and periodicity. The influence of the vaccination program could have no or a negative effect as well. For example, a seasonal peak in disease burden may remain constant or increase in size after the vaccination program is introduced due to vaccine-induced shifts in circulating strains or serotypes of the pathogen. These positive or negative effects may be observed in the disease burden of a population not targeted by the vaccination program also. For example, a steady trend in disease burden in a population not targeted by the vaccination program may change to an increasing trend after the vaccination program is introduced due to a shift in the age of occurrence of the disease. On top of any changes to the periodicity and underlying trend there may be a step reduction in the disease burden as was shown in Figure 4.1.

Statistical models can be used to test formally if there has been a shift in the disease burden or if either the periodicity or underlying trend has changed after introduction of a vaccination program. This requires investigating interactions between a variable that represents the introduction of the vaccination program (either as a binary indicator variable for prior and during the vaccination program or as a continuous variable representing vaccination coverage over time) and variables representing i) linear trend and ii) periodicity (e.g. cosinor terms). Since the outcome variables are typically rates or counts of disease burden over time then these models need to account for any remaining autocorrelation as well, as outlined in Section 2.4.

A statistically significant interaction implies that the vaccination program has had an impact. The coefficient estimates will indicate if this is a positive or negative impact of the vaccination program on this outcome over time. If any of these interactions are statistically significant then it is difficult for these models to quantify the specific effects of the vaccination program on the burden of disease, independent of the pre-existing underlying trend or periodicity before

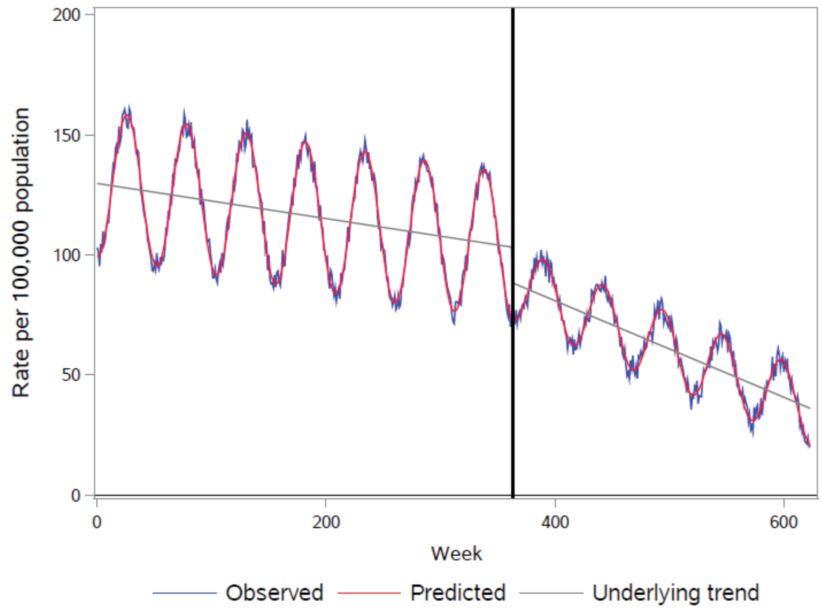
Figure 4.4: *Hypothetical scenario where vaccination efforts cause the underlying trend to decline faster and the size of the seasonal peaks to halve. The vertical line represents introduction of the vaccination program.*



introduction of the vaccination program. For example, assume that the rate of rotavirus hospitalisations was 100 per 100,000 population in the first week of a study period and that this time point coincided with the off-peak season, that there was a consistent, decreasing trend in this rate in the pre-vaccination era of 0.075 per 100,000 population per week and that the periodicity was consistent over the pre-vaccination era with a rate of 60 per 100,000 population more hospitalisations per week in the peak week (Figure 4.4). Assume that with introduction of the vaccination program the decreasing trend is stronger (0.2 per 100,000 population fewer hospitalisations per week) and that the size of the seasonal peaks is halved (only 30 per 100,000 population per week higher in the peak week). The random fluctuations in the rates of hospitalisations are assumed to be stable over time and from a uniform distribution over the interval of -5 to +5 hospitalisations per 100,000 population per week.

A statistical model can be fitted and the statistically significant interactions and direction of the coefficients will imply that the vaccination

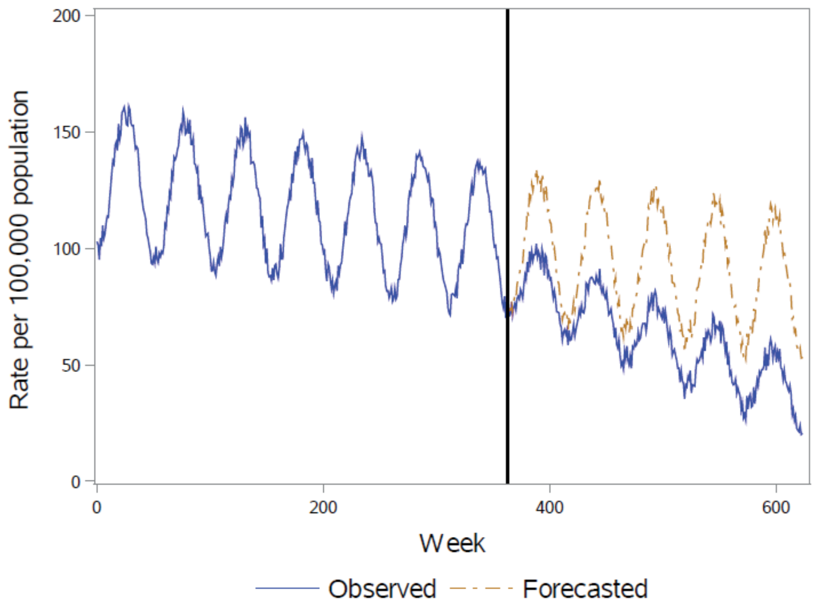
Figure 4.5: *Hypothetical scenario from figure 4.4 that shows the change in underlying trend and seasonality due to vaccination efforts. The vertical line represents introduction of the vaccination program.*



program had a positive impact on reducing the burden of disease. The coefficients from an appropriate model will identify that the maximum height of the seasonal peaks has reduced from approximately 60 to 30 hospitalisations per 100,000 population and that the underlying trend has reduced from approximately 0.075 to 0.2 fewer hospitalisations per 100,000 population per week (Figure 4.5). Additional coefficients for the intercept and introduction of the vaccination program will also be obtained. All of these model coefficients, however, cannot compute how many fewer hospitalisations there are in the vaccination program era that are due to the introduction of the vaccination program because there is no data available at the same point in time on the rate of rotavirus hospitalisations with and without the vaccination program.

An option for quantifying the effect on the number of hospitalisations (i.e. the disease burden of interest in our example) is to create what the disease burden may have been in the "vaccination program era" if a vaccination program had not

Figure 4.6: *Hypothetical scenario from figure 4.4 that forecasts a "no vaccination program" scenario into the vaccination era using pre-vaccination era data only. The vertical line represents introduction of the vaccination program.*



been introduced. That is, to estimate a hypothetical "no vaccination program" disease burden into the vaccination program era from the historical data and then compare these estimates with the observed disease burden in the vaccination program era (i.e. the intervention group). This can be done by forecasting, which predicts the outcome outside the time period of observation. The observations in the pre-vaccination program era are used to predict into the vaccination program era, ignoring the observed data available in the vaccination program era. The mean rate of the forecasted disease burden can then be compared with the observed mean rate in the vaccination era. Continuing the seasonal example, the mean weekly forecasted rate in the first year after the vaccination program was introduced was 101.5 (95% CI: 95.3-107.6) and the mean weekly observed rate was 83.1 (95% CI: 79.9-86.4) (Figure 4.6).

This simple example assumed that the periodicity and underlying trend were consistent in each of the pre-vaccination and vaccination program eras and that there was no remaining autocorrelation after accounting for the underlying trend and periodicity evident in the data series, which is unrealistic in practice. For more complicated patterns of disease burden over time, as is the case for rotavirus and acute gastroenteritis disease burden in NSW, two different models are needed to fully assess the impact of a vaccination program: one model to estimate any changes in temporal trends (by exploring interactions), and another model to quantify those changes in disease burden due to vaccination efforts (by forecasting a "no vaccination program" scenario into the vaccination era). This is because each of these models handle the autocorrelation differently. Fitting a combined, single model is possible, but would have a rather complex mean structure and would only inform better predictions if components, such as the autocorrelation, do not change with introduction of the vaccination program.

Time series methods are best used to account for autocorrelation and to forecast outside the time period of observation [32, 37]. When fitting a model to estimate any changes in temporal trends the variables representing underlying trend, periodicity (there may be many), introduction of the vaccination program and the corresponding interaction terms need to be included in the initial model. Then time series terms (moving average and autoregressive terms) can be included in the model to capture any remaining autocorrelation in the residuals [32]. The autocorrelation is assumed to be consistent across the whole study period. If the autocorrelation is modelled first then some or all of the effects of interest (underlying trend and periodicity) may be captured by the autocorrelation terms and therefore these effects cannot be assessed directly. Since the autocorrelation is assumed to be consistent across the whole study period, these models will not produce the most accurate forecasted outcomes.

Instead, time series models can be used to forecast the outcome by using moving average and autoregressive terms to fully capture the autocorrelation in the pre-vaccination era only [32]. The series may also be differenced if it is not stationary already [37]. Any underlying trend and periodicity will be captured by the model but may not be modelled explicitly in order to increase the accuracy of the forecast. Therefore, this forecasting method cannot be used to assess any changes in the temporal trends due to vaccination efforts, and hence why two separate models are required.

4.2.1 Chapter aims

This chapter will apply the statistical approaches that were outlined using hypothetical scenarios in the introduction to this chapter (Section 4.2) to assess the impact of the Australia rotavirus vaccination program in NSW. Specifically, it will assesses the impact of the Australian rotavirus vaccination program in NSW on acute gastroenteritis burden in hospitals and EDs in children under five years of age in NSW. It will account for the temporal components of the data and investigate i) any changes in temporal trends due to vaccination efforts, and ii) quantify the changes in disease burden identified.

4.3 Methods

4.3.1 Outcome variable to analyse the impact of the Australian vaccination program in NSW

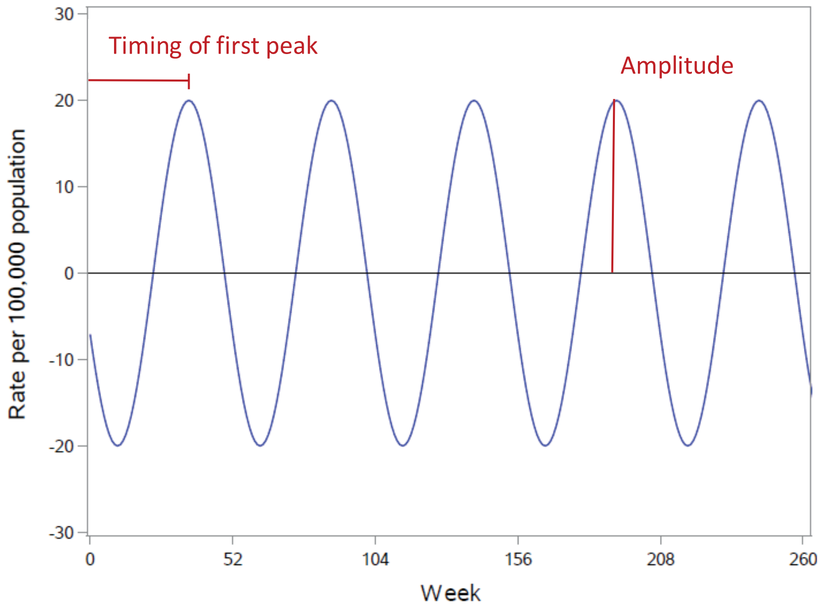
It is difficult to estimate the specific burden of rotavirus hospitalisations and ED presentations in NSW. The ED data for non-admitted presentations in NSW cannot distinguish between acute gastroenteritis due to rotavirus and acute gastroenteritis that is not due to rotavirus. Some studies have used A08.0 coded hospitalisations to monitor the effects of the rotavirus vaccination program in Australia [51, 52], but these analyses will underestimate the true rotavirus burden [13]. Further, attributing the relative decline in A08.0 coded hospitalisations to vaccination efforts assumes that laboratory testing patterns before and after introduction of the vaccination program did not change.

Rotavirus cases are predominantly coded as acute gastroenteritis. In this chapter we assessed the impact of the Australian rotavirus vaccination program by evaluating acute gastroenteritis disease burden over time, before and after vaccination program introduction. The study population was all children aged under five years in NSW for July 2001 to July 2013, as described in Chapter 3. Weekly rates of ED presentations and hospitalisations for all cause gastroenteritis were transformed using the natural logarithm before carrying out analyses. These log-transformed rates were analysed to ensure that the predicted and forecasted rates were non-negative, since rates of healthcare utilisation cannot be negative.

4.3.2 Assessing changes in temporal trends due to vaccination efforts

The log-transformed rates of hospitalisations and ED presentations in week t were modelled using multivariable time series regression models [32]. Initially each model included a linear term for t to denote the underlying trend, cosinor terms to represent seasonality (see below), and a binary indicator variable called program for pre- and post- introduction of the Australian rotavirus vaccination program. Interactions between the underlying trend and the seasonal variables with the introduction of the vaccination program were also included. Backward stepwise regression was used to select the best model from these variables. Variables were removed if their p-value was greater than 0.05 and their removal did not change the adjusted R^2.

Figure 4.7: *Sinusoidal curve demonstrating annual seasonality.*



The cosinor model assumes a sinusoidal pattern, which is symmetric [31]. It consists of transforming t into a linear combination of a sine and cosine term. Taken together, these two terms give a linear representation of a sine curve, which can be used to describe a seasonal pattern with an amplitude and phase. The height of the seasonal peak is double the amplitude and the phase indicates the timing of the seasonal peak for a fixed frequency (e.g. annual seasonality). An example of a cosinor term representing annual seasonality for these weekly data is $\beta_c \cos\left(\frac{2\pi}{52}t\right) + \beta_s \sin\left(\frac{2\pi}{52}t\right)$. The coefficients of these sine and cosine terms (i.e. β_c and β_s) have no meaningful interpretation on their linear scale, but can be transformed into the amplitude $\left(\sqrt{\beta_s^2 + \beta_c^2}\right)$ and phase $\left(\text{ArcTan}\left(-\frac{\beta_s}{\beta_c}\right)\right)$. The constant $\frac{2\pi}{52}$ defines an annual frequency of the cosinor peaks and troughs. A generic formula for this cosinor term is $\beta_c \cos\left(\frac{2\pi k}{52}t\right) + \beta_s \sin\left(\frac{2\pi k}{52}t\right)$, where k specifies the number of peaks and troughs per year.

Figure 4.7 shows an example of an annual sinusoidal pattern. It is possible to combine multiple cosinor terms of different periodicity if there are multiple seasonal patterns in the data. Here three, six and twelve monthly seasonal patterns (i.e. $k = 4, 2$ and 1, respectively) are considered in the initial model to capture the annual seasonality evident in the observed data and any additional seasonal patterns or potential shifts in seasonality.

The initial model of the logarithm of the rate of hospitalisations or ED presentations in week t, denoted by Y_t, for acute gastroenteritis in children under five years of age, including interactions of underlying trend and periodicity with introduction of the rotavirus vaccination program, is:

$$\begin{aligned}
\log Y_t = {} & \beta_0 + \beta_1 \text{program} + \beta_2 t + \beta_3 \text{program} * t + \\
& + \beta_4 \cos\left(\frac{2\pi}{52} t\right) + \beta_5 \sin\left(\frac{2\pi}{52} t\right) + \beta_6 \cos\left(\frac{2\pi}{26} t\right) \\
& + \beta_7 \sin\left(\frac{2\pi}{26} t\right) + \beta_8 \cos\left(\frac{2\pi}{13} t\right) + \beta_9 \sin\left(\frac{2\pi}{13} t\right) \\
& + \beta_{10} \text{program} * \left\{ \cos\left(\frac{2\pi}{52} t\right) + \sin\left(\frac{2\pi}{52} t\right) \right\} \\
& + \beta_{11} \text{program} * \left\{ \cos\left(\frac{2\pi}{26} t\right) + \sin\left(\frac{2\pi}{26} t\right) \right\} \\
& + \beta_{12} \text{program} * \left\{ \cos\left(\frac{2\pi}{13} t\right) + \sin\left(\frac{2\pi}{13} t\right) \right\}
\end{aligned}$$

where β_0 is the intercept and β_1 to β_{12} are the coefficients for the variables described above. For ease of interpretation of the interaction coefficients, the interaction terms can be multiplied out and the above formula can be re-written as:

$$\begin{aligned}
\log Y_t = {} & \beta_0 + \beta_1 \text{program} + \beta_2 t + \alpha_1 t_{\text{pre}} + \alpha_2 t_{\text{post}} \\
& + \alpha_3 \cos\left(\frac{2\pi}{52} t\right)_{\text{pre}} + \alpha_4 \sin\left(\frac{2\pi}{52} t\right)_{\text{pre}} + \alpha_5 \cos\left(\frac{2\pi}{52} t\right)_{\text{post}} \\
& + \alpha_6 \sin\left(\frac{2\pi}{52} t\right)_{\text{post}} + \alpha_7 \cos\left(\frac{2\pi}{26} t\right)_{\text{pre}} + \alpha_8 \sin\left(\frac{2\pi}{26} t\right)_{\text{pre}}
\end{aligned}$$

$$+\alpha_9 \cos\left(\frac{2\pi}{26}t\right)_{\text{post}} + \alpha_{10} \sin\left(\frac{2\pi}{26}t\right)_{\text{post}} + \alpha_{11} \cos\left(\frac{2\pi}{13}t\right)_{\text{pre}}$$

$$+\alpha_{12} \sin\left(\frac{2\pi}{13}t\right)_{\text{pre}} + \alpha_{13} \cos\left(\frac{2\pi}{13}t\right)_{\text{post}} + \alpha_{14} \sin\left(\frac{2\pi}{13}t\right)_{\text{post}}$$

where t_{pre} is equal to *t* in the pre-program era and zero in the program era, t_{post} is zero in the pre-program era and equal to t from introduction of the vaccination program, $\cos\left(\frac{2\pi k}{52}t\right)_{\text{pre}}$ and $\sin\left(\frac{2\pi k}{52}t\right)_{\text{pre}}$ are equal to the cosinor term $\cos\left(\frac{2\pi k}{52}t\right)$ and $\sin\left(\frac{2\pi k}{52}t\right)$ for a given value of k in the pre-program era ($k = 1, 2$ and 4) and zero in the program era, and $\cos\left(\frac{2\pi k}{52}t\right)_{\text{post}}$ and $\sin\left(\frac{2\pi k}{52}t\right)_{\text{post}}$ are zero in the pre-program era and equal to the cosinor term $\cos\left(\frac{2\pi k}{52}t\right)$ and $\sin\left(\frac{2\pi k}{52}t\right)$ for a given value of k in the vaccination program era ($k = 1, 2$ and 4).

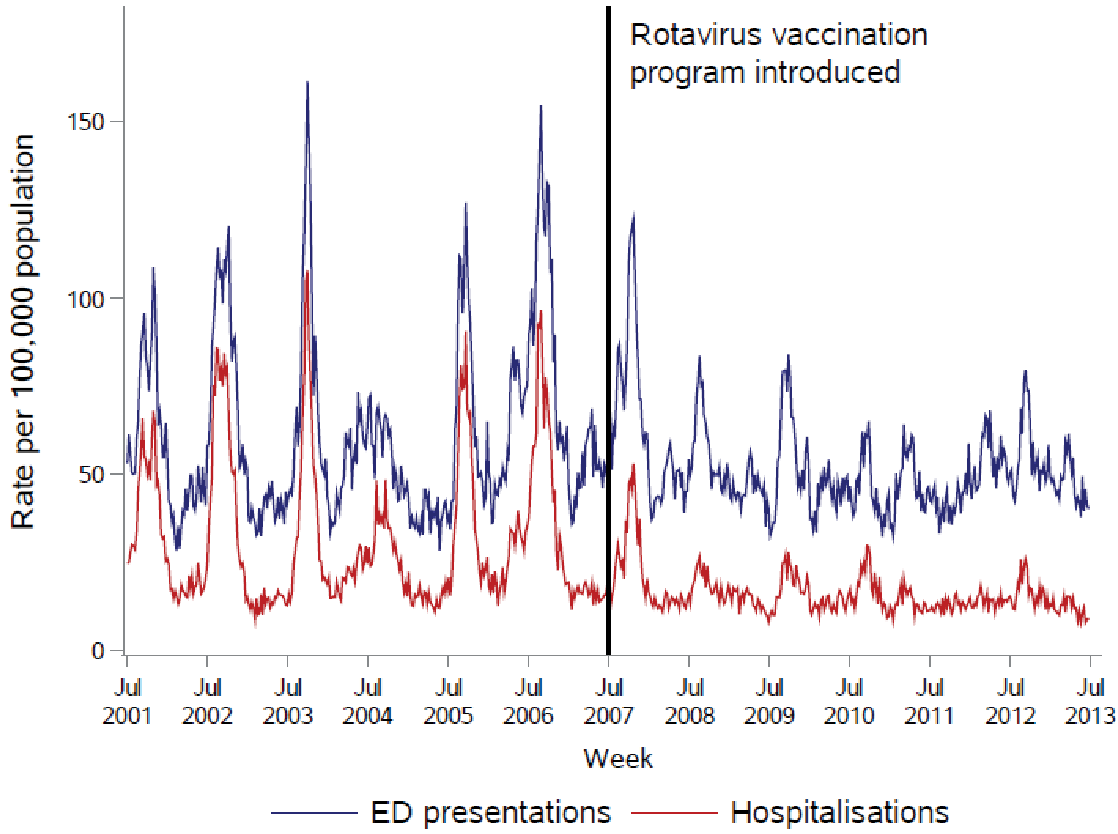
After the best model was selected from the initial variables an ARMA(p, q) model with p autoregressive (AR) and q moving-average (MA) terms was included to capture any remaining autocorrelation in the residuals. This model is a linear combination of past values and past errors in the residuals [37]. The final model with ARMA(p, q) terms was selected based on i) no evidence of remaining autocorrelation in the residuals, ii) reasonable residual normality diagnostics (of the QQ-plot and histogram), and iii) minimum Akaike's Information criterion (AIC). The log-likelihood is a measure of model fit and the AIC is equal to -2 times the log-likelihood plus twice the number of parameters in the model [82]. It penalises the addition of new variables since the addition of any variable to the model will improve the log-likelihood, even if it is not associated with the outcome. The lower the AIC the better the model fit [82].

P-values for interaction terms were obtained using the Likelihood Ratio Test [40].

4.3.3 Quantifying changes in the burden of disease due to vaccination efforts

An ARIMA(p, d, q) model was fitted to the pre-vaccination program introduction data (1 July 2001 to 30 June 2007) [32], where p was the order of the AR component of the model, q was the order of the MA component of the model and d was the level of differencing (refer to Section 2.4.1). The final model was selected based on i) no remaining autocorrelation in the residuals, ii) reasonable residual normality diagnostics (of the QQ-plot and histogram), and iii) stable forecasted confidence limits (i.e. not tending to infinity). This model was used to forecast forward in time from 1 July 2007 to 30 June 2013 to project what may have occurred if no vaccination program had been introduced on 1 July 2007 based on the trends in the outcome variable prior to vaccination efforts. The weekly forecasted rates and limits of the 95% confidence interval were exponentiated to transform them back onto the original scale of rate per 100,000 population. A small adjustment to the forecasted values of $\frac{\text{standard error}^2}{2}$ was required when applying this transformation [31]. The further into the vaccination program era the "no vaccination program" disease burden is forecasted, the greater the uncertainty in that forecast. Since the decline in acute gastroenteritis hospitalisations and ED presentations is almost immediate and sustained, we compare the exponentiated mean annual rates of forecasted hospitalisations and ED presentations to the observed mean annual rates for the first two full calendar years after introduction of the vaccination program only (i.e. 2008 and 2009).

Figure 4.8: *Weekly rates of NSW hospitalisations and non-admitted ED presentations for all cause acute gastroenteritis in children under the age of 5 years in NSW, Australia, 2001-2013. The vertical line represents introduction of the vaccination program.*



4.4 Results

4.4.1 Crude estimates of acute gastroenteritis disease burden in NSW

During the study period there were a total of 66,436 NSW hospitalisations for children under the age of five years with a primary diagnosis for acute gastroenteritis. There were 160,294 ED presentations for this same age group with a primary symptom of gastroenteritis. Figure 4.8 displays the weekly distribution of acute gastroenteritis for these two surveillance datasets in children under five years of age over the study period, July 2001 to June 2013. There is an obvious seasonal peak each year and the size of these peaks appears to decline from 2008, which is after the rollout of the national rotavirus vaccination program started in July 2007.

In the pre-vaccination program era (mid-2001 to mid-2007), the mean weekly rate of hospitalisations and ED presentations was 31.6 (95% CI: 29.2-33.9) and 62.0 (95% CI: 59.1-64.9) per 100,000 children under five years of age, respectively (Table 4.1). In the vaccination era (mid-2008 to mid-2013), this mean weekly rate reduced to 15.2 (95% CI: 14.7-15.8) and 49.9 (95% CI: 48.7-51.2) per 100,000 population, respectively.

The mean annual rate of acute gastroenteritis hospitalisations and ED presentations prior to introduction of the Australian rotavirus vaccination program were 1665.2 and 3245.7 per 100,000 children under five years of age in

Table 4.1: *Descriptive statistics of the weekly rates of NSW hospitalisations and ED presentations for acute gastroenteritis per 100,000 children aged less than five years by year.*

	Hospital admissions (Rate per 100,000 population)			ED presentations (Rate per 100,000 population)		
Year	Mean (95% CI)	Min	Max	Mean (95% CI)	Min	Max
2002	35.8 (28.9-42.7)	13.2	85.7	63.5 (55.8-71.3)	28.4	120.6
2003	29.2 (22.2-36.2)	8.8	107.9	59.2 (50.5-68.0)	32.5	161.7
2004	25.8 (23.6-28.1)	14.2	48.5	54.4 (51.4-57.3)	33.8	73.5
2005	29.7 (23.6-35.8)	11.0	90.6	55.0 (47.7-62.3)	28.5	127.2
2006	38.6 (32.6-44.7)	12.8	96.7	78.1 (69.8-86.3)	36.3	154.8
2007	22.6 (19.6-25.6)	11.6	53.0	65.4 (59.6-71.3)	35.8	122.1
2008	16.3 (15.2-17.4)	11.1	26.7	52.4 (49.6-55.3)	37.1	83.6
2009	16.4 (15.1-17.7)	8.1	28.0	51.5 (47.9-55.0)	32.8	84.1
2010	15.6 (14.2-17.0)	9.3	29.7	45.4 (43.3-47.4)	32.3	65.2
2011	13.3 (12.6-14.0)	8.4	20.5	45.1 (43.0-47.2)	32.5	64.2
2012	15.4 (14.5-16.4)	10.8	25.6	55.0 (52.6-57.4)	42.6	79.7
Jul 2001-Jun 2007	31.6 (29.2-33.9)	8.8	107.9	62.0 (59.1-64.9)	28.4	161.7
Jul 2008-Jun 2013*	15.2 (14.7-15.8)	8.0	29.7	49.9 (48.7-51.2)	32.3	84.1

CI: Confidence Interval
* The period July 2007 to June 2008 was not included to allow time for infants to receive the rotavirus vaccine.

NSW, respectively. The annual rates in 2008 were 849.7 and 2733.0 per 100,000 population, respectively. The annual rates in 2009 were 853.9 and 2680.0 per100,000 population, respectively. These observed rates will be compared to the forecasted rates for the same years.

4.4.2 Assessing changes in the underlying temporal trends due to vaccination efforts

The final hospitalisations model included annual and bi-annual seasonality with the magnitude of the seasonal peak declining after introduction of the Australian rotavirus vaccination program (Table 4.2). The estimated average seasonal maximum in the pre-vaccination era was 64.0 hospitalisations per 100,000 population and the non-seasonal minimum was 15.1 hospitalisations per 100,000 population. After introduction of the vaccination program, the average seasonal maximum reduced to 22.6 hospitalisations for acute gastroenteritis per 100,000 children aged under five years and the non-seasonal trough was 12.6 per 100,000 population. There was no increasing or decreasing underlying trend detected in the acute gastroenteritis disease burden after adjusting for the seasonality present.

Figure 4.9 shows the observed and model predicted weekly rates of NSW hospitalisations. The underlying trend and periodicity components estimated by the model are shown in Figure 4.10. A figure that combines these two figures as well as figures that show the residual normality diagnostics and residual correlation diagnostics are located in Appendix B.

The final ED presentations model included an underlying trend and there was borderline evidence that this trend changed with vaccination efforts from a small increasing trend to a small decreasing trend (Table 4.2, p=0.05). In the pre-vaccination era the rate of ED presentations for acute gastroenteritis increased on

Table 4.2: *Model output from multivariable time series regression models assessing temporal trends due to vaccination efforts in the weekly rates of NSW hospitalisations and ED presentations for acute gastroenteritis per 100,000 children aged less than five years.*

	Log(Hospitalisations)			Log(ED presentations)		
	Coefficient	P-value	P-value [a]	Coefficient	P-value	P-value [a]
Intercept	3.25	<0.0001		4.01	<0.0001	
Program	-0.50	<0.0001		0.39	0.06	
t_{pre}	-			0.00019	0.6	0.05
t_{post}	-			-0.00098	0.02	
sin_12m $_{pre}$ [b]	0.63	<0.0001	<0.0001	0.36	<0.0001	0.0002
sin_12m $_{post}$ [b]	0.21	0.0004		0.08	0.08	
cos_12m $_{pre}$ [c]	0.13	0.02		0.10	0.01	
cos_12m $_{post}$ [c]	-0.04	0.5		0.04	0.3	
sin_6m $_{pre}$ [d]	0.03	0.4	0.05	-		
sin_6m $_{post}$ [d]	-0.02	0.6		-		
cos_6m $_{pre}$ [e]	-0.27	<0.0001		-		
cos_6m $_{post}$ [e]	-0.15	<0.0001		-		
sin_6m [d,f]	-			-0.02	0.2	<0.0001
cos_6m [e,f]	-			-0.16	<0.0001	
AR_1 [g,h]	0.55	<0.0001		0.67	<0.0001	
AR_2 [g]	0.18	0.0001		-		
AR_3 [g]	0.12	0.002		-		
AR_5 [h]	-			0.13	0.0001	
MA_2 [h]	-			-0.23	<0.0001	
MA_3 [h]	-			-0.17	<0.0001	
MA_21 [h]	-			0.08	0.04	
MA_50 [h]	-			0.17	<0.0001	

ED: Emergency Department; AR: Autoregressive; MA: Moving Average.

[a] P-value for interaction using the Likelihood Ratio Test

[b] sin_12m is the sine component of the annual cosinor term, i.e. $\sin\left(\frac{2\pi}{52}t\right)$

[c] cos_12m is the cosine component of the annual cosinor term, i.e. $\cos\left(\frac{2\pi}{52}t\right)$

[d] sin_6m is the sine component of the bi-annual cosinor term, i.e. $\sin\left(\frac{2\pi}{26}t\right)$

[e] cos_6m is the cosine component of the bi-annual cosinor term, i.e. $\cos\left(\frac{2\pi}{26}t\right)$

[f] The interaction of the bi-annual cosinor term and program introduction was not statistically significant and therefore not included in the final model for the ED data.

[g] The ARMA model included in the final hospitalisations model included AR 1, 2 and 3 terms.

[h] The ARMA model included in the final ED presentations model included AR 1 and 5 terms and MA 2, 3, 21 and 50 terms.

Figure 4.9: *Observed and predicted weekly rates of NSW hospitalisations from the model shown in Table 4.2.*

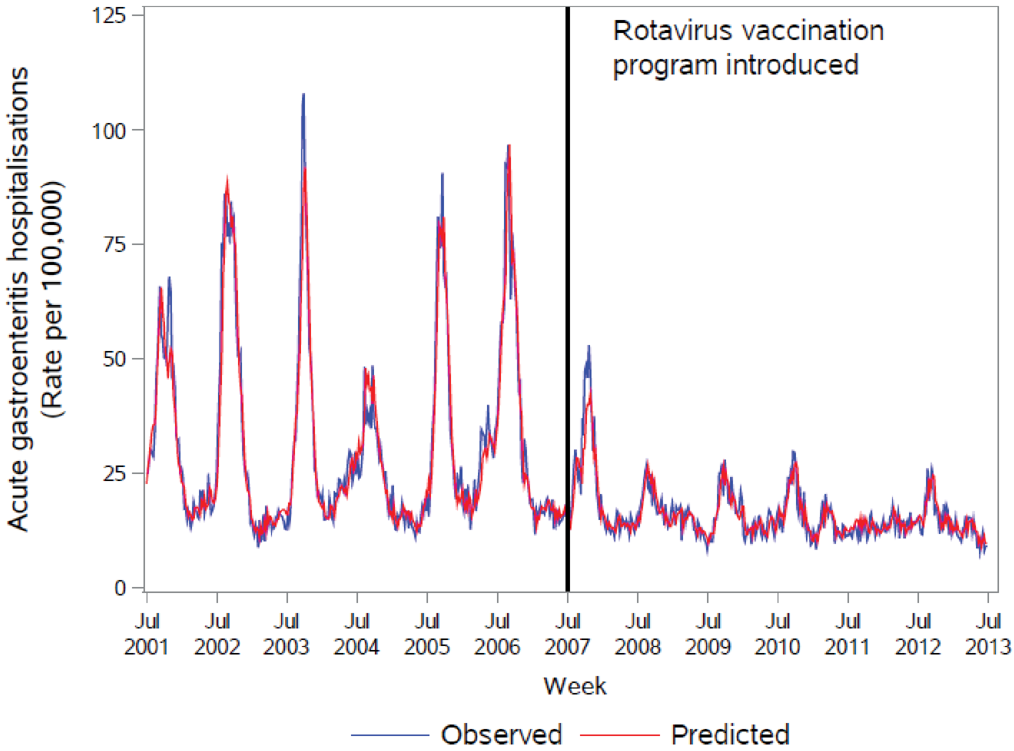


Figure 4.10: *Estimated periodicity and underlying trend in weekly rates of NSW hospitalisations from the model shown in Table 4.2.*

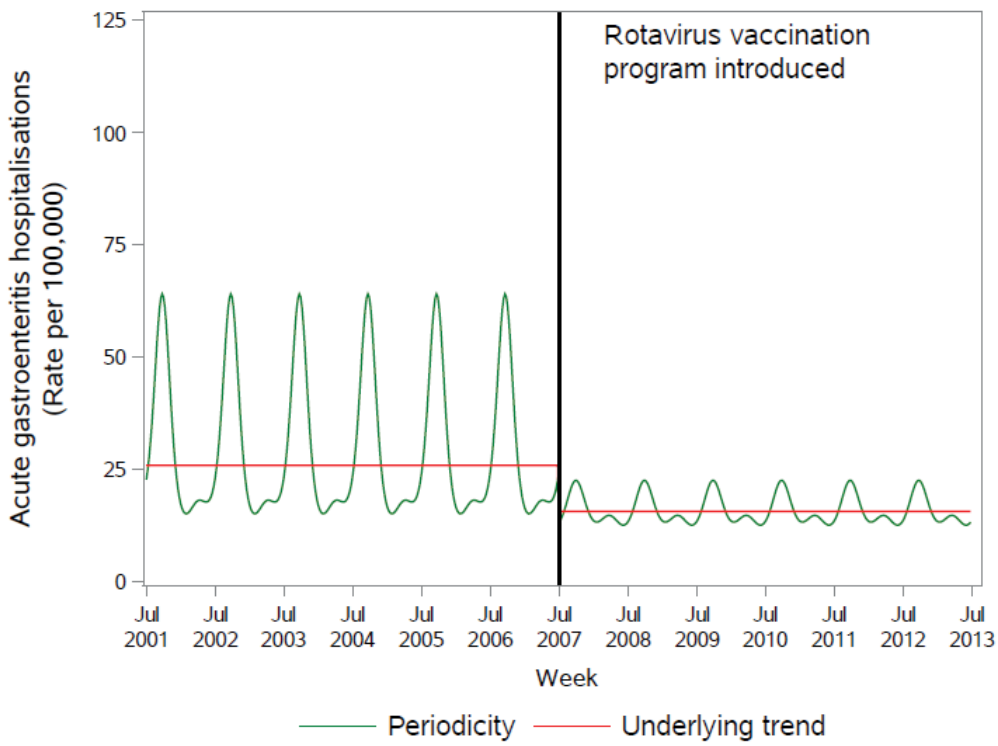


Figure 4.11: *Observed and predicted weekly rates of NSW ED presentations from the model shown in Table 4.2.*

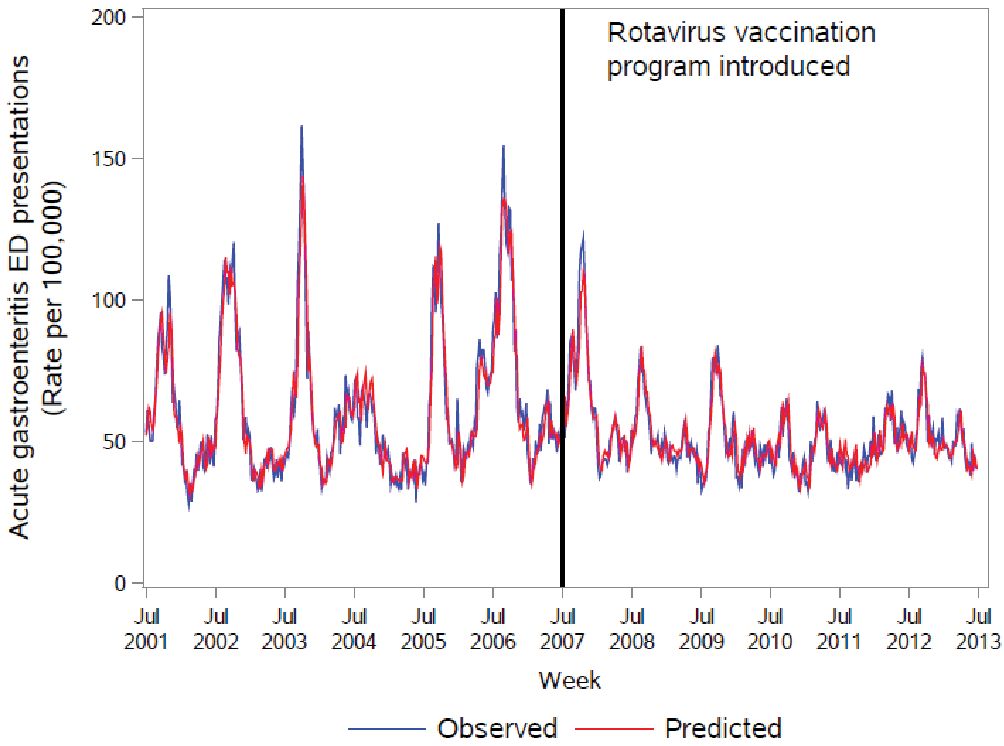
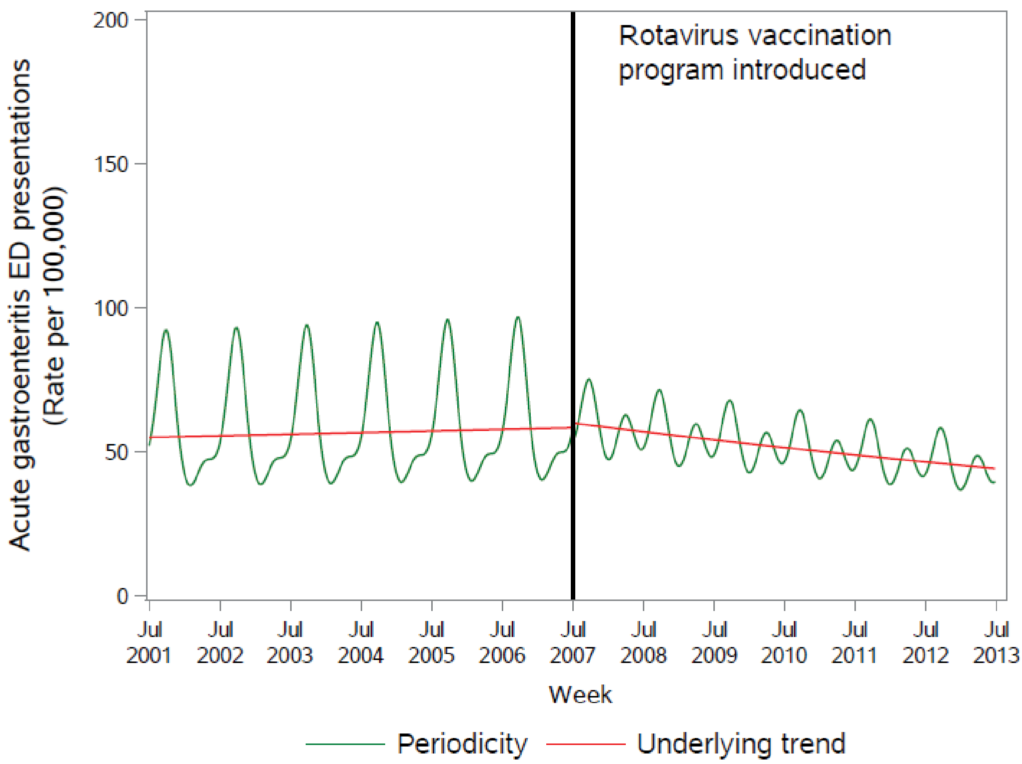


Figure 4.12: *Estimated periodicity and underlying trend in weekly rates of NSW ED presentations from the model shown in Table 4.2.*



the multiplicative scale at a rate of 52.01 ED presentations per 100,000 population per year (Table 4.2). In the vaccination program era the rate of ED presentations for acute gastroenteritis decreased on the multiplicative scale at a rate of 51.9 ED presentations per 100,000 population per year. There was annual seasonality with the magnitude of the seasonal peak declining after introduction of the Australian rotavirus vaccination program ($p<0.0002$). There was bi-annual seasonality that did not change post program implementation ($p<0.0001$). The estimated difference between the average peak and trough (i.e. the seasonal magnitude) in the pre-vaccination era was 53.8 hospitalisations per 100,000 population. After introduction of the vaccination program, this difference reduced to 37.2 hospitalisations for acute gastroenteritis per 100,000 children aged under five years of age.

Figure 4.11 shows the observed and model predicted weekly rates of NSW ED presentations. The underlying trend and periodicity components estimated by the model are shown in Figure 4.12. A figure that combines these two figures as well as figures that show the residual normality diagnostics and residual correlation diagnostics are located in Appendix C.

Table 4.3: *Mean weekly observed rates and forecasted "no vaccination program" rates of acute gastroenteritis hospitalisations per 100,000 children aged less than five years by year.*

Year	Observed Rates (95% CI)	Forecasted Rates (95% CI)
2008	16.3 (15.2-17.4)	31.2 (26.2-36.2)
2009	16.4 (15.1-17.7)	31.6 (26.6-36.5)

CI: Confidence Interval.

4.4.3 Quantifying changes in the burden of disease

The final ARIMA model to forecast the hospitalisations data for "no vaccination program introduced" in the vaccination program era had differencing of order 52 weeks, autoregressive terms for 1, 2, 3, 22 and 26 weeks and a moving average term for 52 weeks. The residual normality diagnostics were reasonable and there was no remaining autocorrelation in the residuals (Appendix D). Figure 4.13 shows the forecasted weekly rates of hospitalisations per 100,000 population for the first three seasons after introduction of the vaccination program. The uncertainty around the forecasts was greatest around the peak of each season and lowest in the off-peak season. Table 4.3 summarises the mean weekly observed and model forecasted rates of acute gastroenteritis hospitalisations for the calendar years 2008 and 2009. The observed rates were approximately half the forecasted rates. There was no obvious shift in the peak timing of the rotavirus season after introduction of the national vaccination program (Figure 4.14).

The final ARIMA model to forecast the ED presentations data for "no vaccination program introduced" in the vaccination program era had differencing of order 52 weeks, autoregressive terms for 1, 2, 5, and 10 weeks and a moving average term for 52 weeks. The residual normality diagnostics were reasonable

Table 4.4: *Mean weekly observed rates and forecasted "no vaccination program" rates of acute gastroenteritis ED presentations per 100,000 children aged less than five years by year.*

Year	Observed (95% CI)	Forecasted (95% CI)
2008	52.4 (49.6-55.3)	62.3 (56.7-68.0)
2009	51.5 (47.9-55.0)	62.3 (56.7-68.0)

CI: Confidence Interval.

Figure 4.13: *Observed and forecasted weekly rates of NSW hospitalisations per 100,000 children aged under five years of age (July 2001- December 2010). The vertical line represents introduction of the vaccination program.*

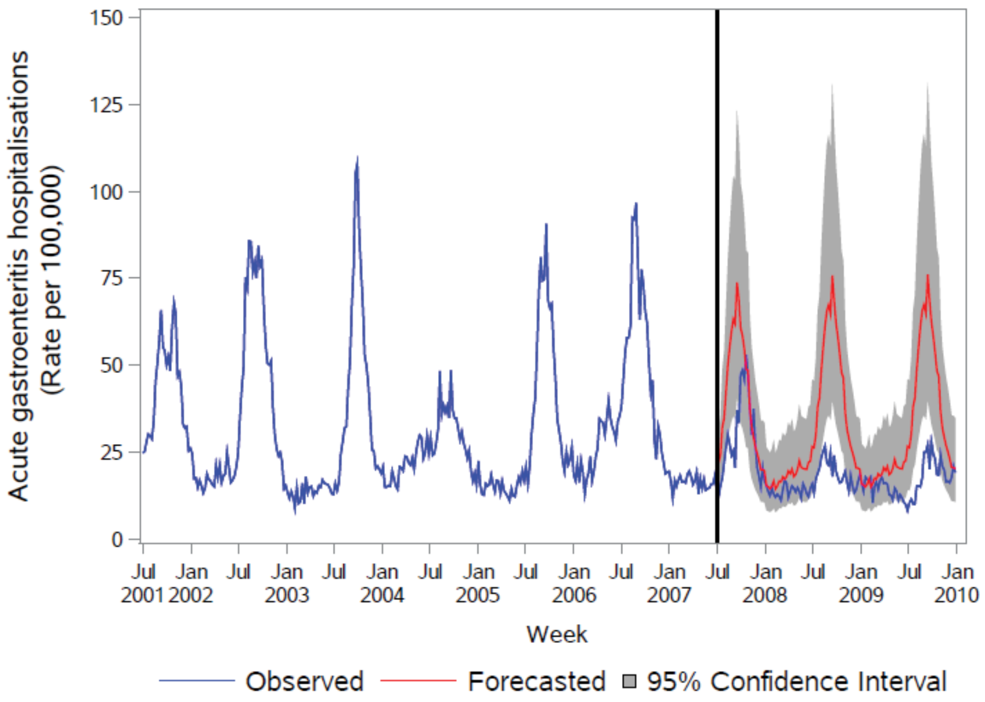


Figure 4.14: *Observed and forecasted weekly rates of NSW hospitalisations per 100,000 children aged under five years of age (July 2007- July 2013).*

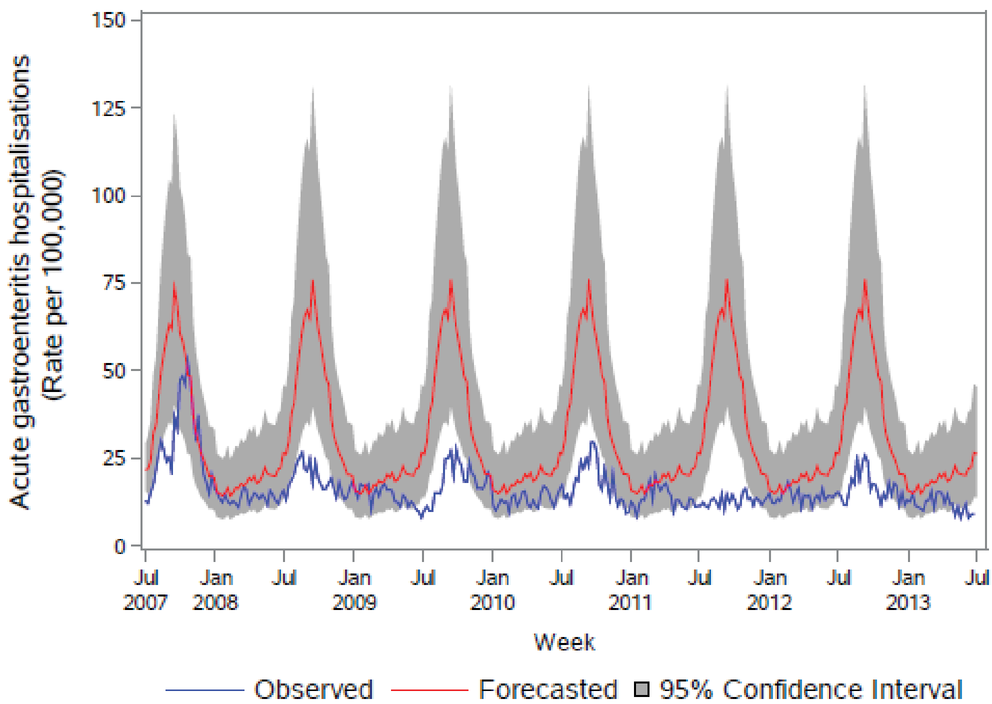


Figure 4.15: *Observed and forecasted weekly rates of NSW ED presentations per 100,000 children aged under five years of age (July 2001- December 2010). The vertical line represents introduction of the vaccination program.*

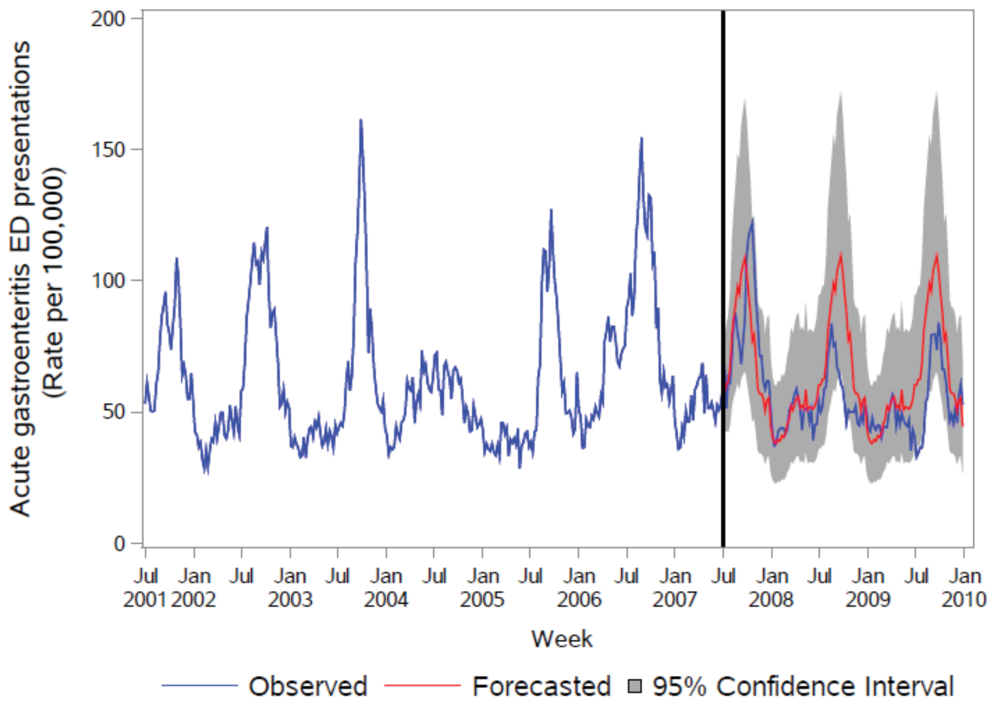
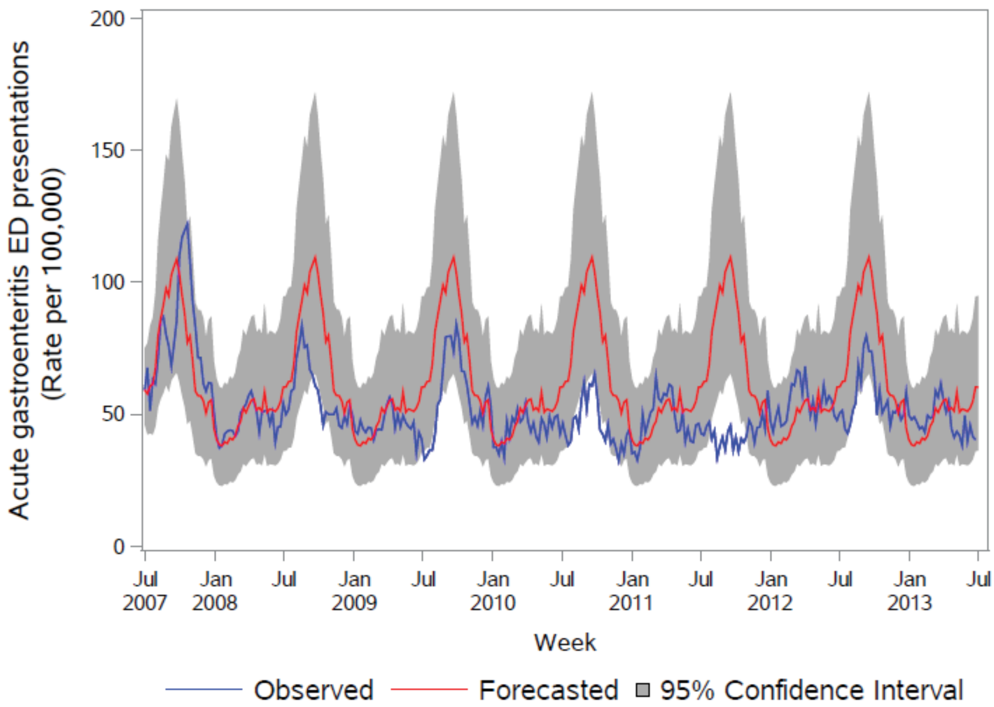


Figure 4.16: *Observed and forecasted weekly rates of NSW ED presentations per 100,000 children aged under five years of age (July 2007- July 2013).*



and there was no remaining autocorrelation in the residuals (Appendix E). Figure 4.15 shows the forecasted weekly rates of ED presentations per 100,000 population for the first three seasons after introduction of the vaccination program. The uncertainty around the forecasts was greatest around the peak of each season and lowest in the off-season. Table 4.4 summarises the mean weekly observed and model forecasted rates of acute gastroenteritis ED presentations for the calendar years 2008 and 2009. The observed rates were approximately 10 per 100,000 population fewer than the forecasted rates in both years. There is no obvious shift in the peak timing of the rotavirus season after introduction of the national vaccination program (Figure 4.16).

4.5 Discussion

This chapter explained in detail why two separate models were used to fully analyse the impact of the Australian rotavirus vaccination program on i) hospitalisations, and ii) ED presentations in children aged under five years in NSW. One statistical model was used to assess any changes in temporal trends due to vaccination efforts by exploring interactions with the vaccination program introduction and these temporal trends. Another statistical model was used to quantify the changes in disease burden identified by forecasting “no vaccination program” disease burden into the vaccination era and compared the forecasts to the observed disease burden in the vaccination era. It was found that, on average, the Australian rotavirus vaccination program has halved the weekly rate of acute gastroenteritis hospitalisations and led to a sixteen percent reduction in the weekly rate of acute gastroenteritis ED presentations of children under five years of age in NSW in the years 2008 and 2009 (Tables 4.3 and 4.4). The large decline in hospitalisations was attributed to the substantial decline in the magnitude of the periodicity (i.e. the difference between the seasonal peak and trough). The decline in the ED presentations was attributed to both a statistically significant decline in the magnitude of the periodicity and borderline evidence of a change in the direction of the small underlying trend from increasing to decreasing (Table 4.2).

These reductions attributed to the rotavirus vaccination program are similar to the crude results presented in Table 4.1. A comparison of the (crude) mean weekly rates in the pre-program era to the mean weekly rates in the program era finds a 52% reduction in hospitalisations and a 20% reduction in ED presentations (Table 4.1). The (crude) mean weekly rates of hospitalisations and ED presentations for acute gastroenteritis in the pre-program era (July 2001 to June 2007) were 31.6 (95% CI: 29.2-33.9) and 62.0 (95% CI: 59.1-64.9) per 100,000 children under five years of age in NSW, respectively (Table 4.1). Even though these averages did not take into consideration pre-existing temporal trends in the pre-vaccination program era, they were similar to the model forecasted values in 2008 of 31.2 (95% CI: 26.2-36.2) and 62.3 (95% CI: 56.7-68.0), respectively, and 2009 of 31.6 (95% CI: 26.6-36.5) and 62.3 (95% CI: 56.7-68.0) per 100,000 population, respectively (Tables 4.3 and 4.4). This is because there was no strong underlying trend present in either data series before introduction of the rotavirus vaccination program.

Five out of seven studies in the review of unique articles that analysed the impact of the Australian rotavirus program used only a basic analytic approach to assess program impact (Section 2.5, Table 2.3) [16, 49, 50, 52, 53]. They compared the mean annual rate, number or percentage of disease burden (hospitalisations, laboratory tests, ED presentations, rotavirus notifications) in the pre-program era to the observed rate, number or percentage in individual years after program implementation. This simple method draws the same conclusions as the time series regression models used in this study. However, this can only be known after applying appropriate methods that consider temporal trends first. The (crude) mean annual rates of hospitalisations and ED presentations for acute gastroenteritis in the pre-program era (July 2001 to June 2007) were 1665.2 and 3245.7 per 100,000 children under five years of age in NSW, respectively. The annual rate in 2009 was 853.9 and 2680.0 per 1000,000 population, respectively. This equates to a reduction of 48.7% and 17.4%, respectively. If there had been a strong underlying trend present in either data series before implementation of the rotavirus vaccination program then this basic method would not be appropriate. Published articles that assess the impact of rotavirus vaccination programs outside of Australia have also used this basic method [10, 83-86].

In addition, our times series regression models were able to enumerate the size of the seasonal peaks and troughs and their timing and to assess whether the underlying trend or periodicity had changed through vaccination efforts using statistical tests, which the simple analytic approach is unable to do. We found that, on average, the decline in the magnitude of the periodicity was from 48.9 to 10.0 in the hospital data and 53.8 to 37.2 in the ED data per 100,000 children less than five years of age in NSW. Further, the ED data changed from an increasing trend on the multiplicative scale at a rate of 52.01 ED presentations per 100,000 population per year in the pre-program era to a decreasing trend of 51.9 ED presentations per 100,000 population per year on the multiplicative scale.

The sixth Australian study, Clarke et al. [51], tried to account for the temporal trends evident in their data by comparing the rate of hospitalisations over a 24 month period immediately prior to program implementation to a 24 month period commencing one year after introduction of the vaccination program. This analytic approach can still over or underestimate the true impact of the vaccination program if there is an underlying trend in the pre-program era, as explained in Section 4.2 and shown in Figure 4.2. However, it will be less of an over or underestimate than for the method of the five studies discussed above, since Clarke et al. [51] included fewer years of data than the other studies.

The seventh study, Davey et al. [48], used a generalised additive model to apportion ED presentations in NSW for acute gastroenteritis as those due to rotavirus and those that are not. After doing so, they used a basic analytic approach that compared the mean weekly model predicted rate of ED presentations in the pre-program era to the mean weekly model predicted rate of ED presentations in the program era. The same issues apply in this case when underlying trends are present.

Four of these seven studies have analysed the impact of the Australian rotavirus vaccination program on acute gastroenteritis hospitalisations or ED presentations in Australia. Davey et al. analysed ED presentations for acute gastroenteritis in children under five years of age in NSW from 2003 to 2011. Although their estimate of the mean weekly rate of ED presentations in the pre-program era at 70.4 per 100,000 population is different to this study (62.0 per 100,000 population), the percentage reduction in all-cause acute gastroenteritis was similar and found to be 18.3 (95% CI: 11.6-25.0) [48]. The discrepancy in rates is likely due to the number of hospitals included in each study. There were fifty hospitals included in their study and only 43 in this study. Our study period was longer (2001 to 2013) and required that all hospitals included had provided data and with a Good Diagnosis Coding status (assigned by the NSW Ministry of Health) across the entire study period.

Clarke et al. [51] found a 48.4% reduction in the number of hospitalisations of children under five years of age with any diagnosis of acute gastroenteritis (primary diagnosis or otherwise) in South Australia when comparing hospitalisations from May 2005 to April 2007 with hospitalisations from May 2008 to April 2010. This is similar to the finding in this study of approximately 50%, even though our study only analysed hospitalisations with a primary diagnosis of acute gastroenteritis and our definition of acute gastroenteritis included the ICD-10-AM code K52 and theirs did not. Dey et al. [52] analysed hospitalisations of children under five years of age in Australia with any diagnosis of acute gastroenteritis. They found only a 35.9% reduction between the mean annual number of hospitalisations in the pre-program era (July 2001 to June 2007) and the 2009-10 financial year (i.e. July 2009 to June 2010) [52]. The differences between our estimate and these national estimates may reflect differences across the states and territories of Australia.

The fourth study, Pendleton et al. [53], analysed acute gastroenteritis hospitalisations of children under five years of age in New South Wales, the Australian Capital Territory and the Northern Territory, where Rotarix was administered. It is not stated in the published article which ICD-10-AM codes were used to select acute gastroenteritis hospitalisations nor if they were from the primary diagnosis field only or any diagnosis field. When comparing the mean annual rate in the pre-vaccination program era (July 1998 to June 2006) to the 2008-09 financial year (i.e. July 2008 to June 2009), the percentage reductions in children aged 0, 1, 2, 3 and 4 years were 36.8, 54.2, 51.1, 47.9 and 48.2, respectively [53].

Some international articles that have analysed the impact of a rotavirus vaccination program have acknowledged that separate models are needed to i) assess changes in disease burden and ii) test if there has been a change [35, 87, 88]. However, they have not explicitly stated what they were testing or why other aspects were not tested. Do Carmo et al. [35] used a Poisson regression model to assess the impact of the Brazilian rotavirus vaccination program introduced in 2006 on monthly rates of all-cause diarrhoea related hospitalisations in children under five years of age in Brazil (2002 to 2009). The regression model first adjusted for seasonality by including eleven binary indicators for month and it adjusted for secular trends by including a term for year of admission. This model was used to forecast the expected monthly rates in the absence of a vaccination (2007 to 2009) using pre-vaccination program data (2002 to 2005), which were then compared to the observed rates (2007 to 2009). This is similar to our approach but uses a different method to account for seasonality and underlying trend. No information was presented on model fit or residual autocorrelation. Do Carmo et al. then fitted a Poisson regression model to the pre-program and program eras that adjusted for periodicity and underlying trend in the same way as their previous model and included a binary indicator for introduction of the vaccination program (pre-/post-introduction). This binary indicator was used to calculate rate ratios between the pre-program and program eras and the corresponding p-value was used to determine if the percentage reduction was statistically significant or not. While a statistically significant binary indicator does indicate a program impact, a non-statistically significant binary indicator does not rule out a program impact. This is because any underlying trend and periodicity may contain effects of the vaccination program. By simply adjusting for the underlying trend and periodicity but only assessing the binary indicator these effects cannot be attributed to the vaccination program if they exist. No information was provided in the do Carmo et al. article on the coefficients of the periodicity or underlying trend nor if there was an interaction between them and the introduction of the vaccination program other than that there was a significant declining trend in children under one year of age in the pre-vaccination program era. This approach, where only a binary indicator for introduction of the vaccination program is used to test for a program impact, is also seen in other articles that assess the impact of a rotavirus vaccination program [87, 89].

When adjusting for periodicity and underlying trends in disease burden over time it is important not to overfit the model (i.e. have too many parameters compared to the number of data points). Our study had over 300 weekly rates in each of the pre-program and program eras and at most 16 parameters in a single model, which is a reasonable number of data points for these 16 parameters. Do Carmo et al. [35] had 48 monthly data points in the pre-vaccination era and 13 parameters (linear term for year of admission and binary indicators for month of admission of their monthly data), which could lead to model over fitting. Hungerford et al. [87] had a similar analytical approach to the do Carmo et al. study; however they used a negative binomial regression model instead of a Poisson regression model to assess monthly rates of acute gastroenteritis hospitalisations in children under fifteen years of age at one children's hospital in the UK. With thirteen years of monthly data in the pre-vaccination era and the same 13 parameters in their model Hungerford et al. are less likely to have overfitted their model than do Carmo et al.

Time series forecasting is generally used to forecast for a short time into the unknown future [37]. Examples include forecasting future monthly numbers of airline passengers and future daily stock closing prices [37]. Typically when a new data point becomes available (e.g. today's daily stock closing price) the model is updated and new forecasts for future data points are estimated. This study uses time series forecasting methods to create an estimate of what the disease burden may have been in the vaccination program era if a vaccination program had not been introduced. No new data points are available after introduction of the vaccination program. Consequently, the chances of our forecasted values representing the truth in the vaccination program era if no vaccination program had been introduced decreases the further into the future the forecast is made (i.e the uncertainty increases the further into the program).

Therefore, it is important to consider this uncertainty when determining the number of forecasted values to compare with the observed values in the vaccination program era.

In this study, the decline in acute gastroenteritis hospitalisations and ED presentations was almost immediate and is sustained. Hence we chose to compare only the first two full calendar years after introduction of the vaccination program. In situations where the impact of the vaccination program (rotavirus or otherwise) is not immediate or not sustained more forecasted rates would need to be compared to the observed rates to quantify changes in the burden of disease. In this study, the benefit of forecasting beyond two full calendar years was to visually identify any shifts in peak incidence that may not have been picked up in the model that assessed changes in temporal trends. Our short time frame comparison showed a large decline in hospitalisations for acute gastroenteritis disease burden and a smaller decline in the ED presentations for acute gastroenteritis (Tables 4.3 and 4.4). When forecasting it is also important to consider the number of data points available to make these forecasts as well. Our forecasts, which were computed until mid-2013 for the visual comparison, are unlikely to be reliable in 2013 since this involved forecasting six years of data (July 2007 to July 2013) from six years of data (July 2001 to July 2007).

The current study used a cosinor model to capture the periodicity in the hospital and ED presentations data. This model assumes a sinusoidal seasonal pattern which is symmetric both in the horizontal (time) and vertical (rate of acute gastroenteritis disease burden) axes. This assumption may be too restrictive for these data; yet the figures that overlay the modelled seasonality on the observed data look reasonable (Appendix Figures B.1 and C.1). An alternative when this plot does not look reasonable is to include a combination of a cosinor function and another function (e.g. a sawtooth function) to better capture the periodicity [31].

A strength of this study was having access to weekly data and carrying out analyses in weeks. This allowed us to better capture the temporal trends in the data. Six of the seven studies that analysed the impact of the Australian rotavirus program analysed their data annually, even though they had access to monthly data [16, 51-53]. This study analysed the data on the log-scale and used a generalised linear model that assumed the residuals were normally distributed. The exponential of the linear combination of coefficients on the log-normal scale gave multiplicative coefficients on the original scale (rate per 100,000 population). This made the coefficients harder to interpret but ensured that no negative rates or confidence limits were outputted by the models.

A limitation of this study was not having access to incident cases of acute gastroenteritis hospitalisations. It is possible to remove those hospitalisations with a hospital transfer code in order to be closer to incident cases. However, we only had access to this information within hospitals and not across hospitals as we did not have access to linked data in NSW which provides this information. Therefore, we did not exclude any transfers in case it introduced differential bias. The ED data consists of non-admitted presentations to the ED and therefore are likely to be incident cases of acute gastroenteritis.

In conclusion, it is important to assess the impact a vaccination program has on the burden of disease after it is implemented. To do this it is essential to consider pre-existing temporal trends in the pre-vaccination program introduction era and to consider any changes to the underlying trend and periodicity. A full analysis requires fitting statistical models that: i) estimate any changes in temporal trends due to vaccination efforts, and ii) quantify the changes in disease burden if applicable.

4.6 Chapter summary

This chapter assessed the impact of the Australian rotavirus vaccination program on acute gastroenteritis disease burden in children aged under five years in NSW hospitals and EDs. It specifically looked for any changes in underlying trend, periodicity and disease burden using appropriate methods. It found that, on average, the Australian rotavirus vaccination program has halved the weekly rate of acute gastroenteritis hospitalisations and led to a sixteen percent reduction in the weekly rate of acute gastroenteritis ED presentations within the first 2.5 years of the program. For both hospitalisations and ED presentations there was a statistically significant decline in the magnitude of the seasonal peaks. There was borderline evidence of a change in the long-term trend of acute gastroenteritis ED presentations after introduction of the Australian rotavirus vaccination program (from increasing to decreasing).

These analyses, of the acute gastroenteritis burden in NSW hospitals and ED presentations, do not answer all the questions on the impact of the rotavirus vaccination program in NSW. As described in Section 2.4.7, it can be of interest to estimate the percentage of (non-specific) acute gastroenteritis burden that is potentially vaccine preventable (i.e. rotavirus-attributable) and whether this percentage has changed over time. The next chapter will estimate the rotavirus specific acute gastroenteritis disease burden.

Chapter 5: An investigation of statistical methods to estimate the number of acute gastroenteritis hospitalisations and emergency department presentations in NSW that are due to rotavirus

5.1 Chapter Overview

The previous chapter assessed the impact of the Australian rotavirus vaccination program on the (non-specific) disease burden of acute gastroenteritis hospitalisations and ED presentations in children under five years of age in NSW. This chapter investigates statistical methods to separate hospitalisations and ED presentations for acute gastroenteritis into those that are due to rotavirus and those that are not.

5.2 Introduction

Determining how many of the acute gastroenteritis hospitalisations and ED presentations analysed in Chapter 4 that are attributable to rotavirus is complex. Despite the challenges outlined in Chapter 2, it is still beneficial to estimate the proportion of acute gastroenteritis that is due to rotavirus. An estimate of rotavirus burden was calculated prior to the introduction of the national rotavirus program in Australia to provide information on the numbers and costs associated with rotavirus before the vaccines were introduced [56]. Estimates of rotavirus-attributable illness are critical in pre-implementation economic evaluations, which (often) use vaccine efficacy estimates to predict the impact on this rotavirus specific burden [90, 91]. Economic evaluation was used to help determine if rotavirus vaccination should be added to the national immunisation schedule [92]. In addition, estimates of rotavirus disease burden can be used to evaluate the benefits associated with rotavirus prevention post-vaccination program introduction by comparing the burden before and after introduction of national immunisation. These estimates can also provide information about the (reduction in or shift in) peak season burden, which may inform future seasonal epidemics of rotavirus.

This goal of estimating the amount of non-specific disease burden attributable to an infectious agent over time is not unique to acute gastroenteritis. For example, studies have tried to determine the proportion of all respiratory illness that is attributable to influenza [45-47], due to the similar issues around a lack of laboratory confirmation. Typically this is done by estimating a baseline unrelated to the infectious agent and then examining the 'excess' above this baseline which is estimated to be attributed to the infection of interest [45-47]. Since rotavirus circulates year round, peaking in late winter, it is difficult to establish a baseline of "no rotavirus". One approach to address this is to examine the relationship over time between an infection specific variable (e.g. pathology data) and the non-specific burden. Different statistical methods have been used to do this for influenza and rotavirus [45, 48, 93].

A number of methods have been used in publications to estimate the rotavirus attributable acute gastroenteritis burden, some before and others before and after introduction of a rotavirus vaccination program [48, 51-53, 56, 93]. However, it is not clear what the strengths and weaknesses of each method are. This chapter provides a direct comparison of the estimates produced from using four different methods to determine the rotavirus disease burden in hospitals and EDs in under five year olds in NSW, Australia (July 2001 to June 2013).

5.3 Methods

Four methods are used to estimate rotavirus associated hospitalisations and ED presentations in under five year olds in NSW over the entire study period (July 2001 to June 2013): (i) selecting rotavirus specific ICD-10-AM coded hospitalisations only [51-53] (ii) applying an off-season attributable fraction method [56], (iii) fitting a linear regression method [93], and (iv) using a generalised additive model [48].

5.3.1 Rotavirus coded hospitalisations

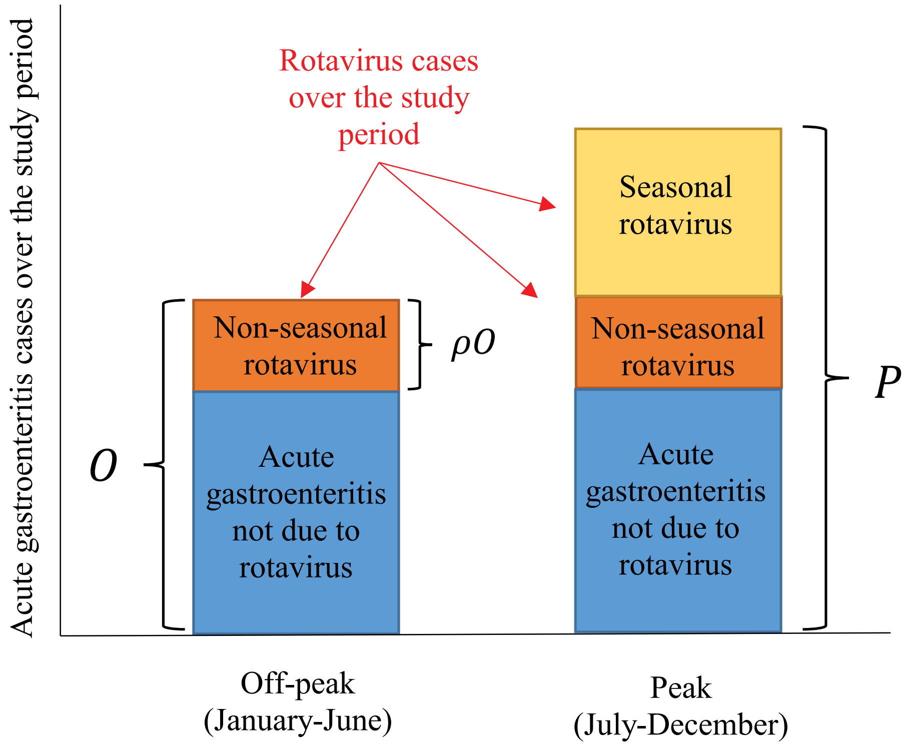
An analysis of hospitalisations with an A08.0, rotaviral enteritis, ICD-10-AM coded hospitalisation will be presented. This approach has been used by a number of studies to give a lower bound on the true number of rotavirus hospitalisations [51-53]. However, it is likely to underestimate the true number of hospital rotavirus cases and there is no such code available for the ED data.

5.3.2 Off-season attributable fraction method

A previously published method [56] estimated the rotavirus fraction of Australian ED presentations, hospital admissions and general practitioner visits for acute gastroenteritis on the basis of the seasonal pattern of incidence. Using this method, the total numbers of i) rotavirus attributed hospitalisations and ii) ED presentations over the entire study period (July 2001 to July 2013 in children less than five years of age in NSW were estimated using the formula:

$$Y = P + O(2\rho - 1)$$

Figure 5.1: *Breakdown of the attributable fraction method showing the components of acute gastroenteritis cases and rotavirus cases over the study period.*



where Y is the estimated total number of rotavirus hospitalisations or rotavirus ED presentations, P is the total number of acute gastroenteritis cases during the peak seasons (July to December) over the entire study period, O is the total number of acute gastroenteritis cases during the off-peak seasons (January to June) over the entire study period, and ρ is the proportion of rotavirus cases in the off-peak season (Figure 5.1).

This method assumed that rotavirus is the only seasonal disease that causes acute gastroenteritis, that the number of non-seasonal rotavirus cases in the off-peak season also occurred in the peak season, and that the percentage of rotavirus-attributable acute gastroenteritis hospitalisations in the off-peak season was the same as the proportion of rotavirus-attributable acute gastroenteritis ED presentations in the off-peak season (i.e. ρ is constant). Galati et al. [56] estimated ρ by auditing all medical records in children under five years of age that could be linked to a rotavirus pathology test result within two days of admission for acute gastroenteritis at two major children's hospitals in Australia. The estimated off-peak season rotavirus fraction (i.e. ρ) in children under the age of five years in Australia was 25.3%. This value for the off-season rotavirus fraction will be used in this study as well.

There are no data available on the off-peak season rotavirus fraction since introduction of the Australian rotavirus vaccination program and it is possible that the off-season rotavirus fraction has declined through vaccination efforts. However, the extent of the decline is unknown. Therefore, two sensitivity analyses were performed where the off-peak season rotavirus fraction was assumed to be i) zero (i.e. $\rho = 0$), which is the most extreme decline in this fraction, and ii) half that of the pre-vaccination era rotavirus fraction (i.e. $\rho = 0.1265$), which may identify a skewed distribution in the estimated rotavirus cases using this method.

5.3.3 Linear regression model

A previously published multivariable linear regression model [93] estimated the rotavirus fractions of hospital admissions, ED presentations, outpatient visits, general practitioner visits and National Health Service nurse-led health helpline calls in parts of England and Wales for acute gastroenteritis on the basis of positive laboratory results for different causative agents (e.g. rotavirus, norovirus, e.coli, etc.).

Using this method, the number of rotavirus attributed hospitalisations and ED presentations (separate models) in week t, denoted Y_t, in children under five years of age in NSW was modelled using the formula:

$$Y_t = \beta L_t + \alpha$$

where L_t is the number of positive rotavirus laboratory results in week t, β is the regression coefficient used to estimate the number of rotavirus-attributable cases associated with each laboratory report, and α is the constant (i.e. intercept) representing the number of non-rotavirus attributable acute gastroenteritis cases each week.

This model assumed that the number of hospitalisation for each positive laboratory result and the number of ED presentations for each positive laboratory result (i.e. the β from each model) did not vary over time. This model had fewer terms than that used by Harris et al. [93] since they had access to laboratory test results for many causative agents and this study only had access to rotavirus laboratory results. Those acute gastroenteritis events for non-rotavirus causative agents were captured in the constant term (i.e. α). The estimated weekly number of rotavirus attributable hospitalisations and ED presentations for acute gastroenteritis was obtained by multiplying the weekly number of positive rotavirus laboratory results by the β from each model. The corresponding numbers of non-rotavirus attributed acute gastroenteritis cases was obtained by calculating the difference between the model predicted numbers of acute gastroenteritis hospitalisations and ED presentations and these estimated numbers of rotavirus attributed hospitalisations and ED presentations, respectively.

5.3.4 Generalised additive model

A previously published generalised additive model [56] approach estimated the rotavirus fraction of NSW ED presentations for acute gastroenteritis on the basis of positive rotavirus laboratory test results. Using this method, the number of rotavirus attributed hospitalisations and ED presentations (separate models) in week t, denoted Y_t, in children under five years of age in NSW was modelled using the formula:

$$Y_t = \beta L_t + \gamma * t + f(\,t\,) + \alpha$$

where L_t is the number of positive rotavirus laboratory results in week t, β is the regression coefficient used to estimate the number of rotavirus-attributable acute gastroenteritis hospitalisations or ED presentations, γ is the regression coefficient for the long-term linear trend over time, $f(\,t\,)$ is a non-parametric cubic spline to account for remaining autocorrelation present, and α is the constant (i.e. intercept) representing the residual number of non-rotavirus attributable acute gastroenteritis hospitalisations or ED presentations.

This generalised additive model assumed that the number of hospitalisations for each positive laboratory report and the number of ED presentations for each positive laboratory report (i.e. the β from each model) did not vary over time. It follows that the estimated weekly number of rotavirus attributable hospitalisations or ED presentations was obtained by multiplying the weekly number of positive rotavirus laboratory results by the β. The corresponding numbers of non-rotavirus attributed acute gastroenteritis cases was obtained by calculating the difference between the model predicted numbers of acute gastroenteritis hospitalisations and ED presentations and these estimated numbers of rotavirus attributed hospitalisations and ED presentations, respectively.

The non-parametric cubic spline $f(t\,)$ is a piecewise polynomial curve. It divides the data along the time axis into regions and fits a polynomial within each of these regions, which is not restricted to be linear. Where each curve joins is called a knot. The larger the number of knots used, the more flexible the curve. The non-parametric cubic spline included in this model allows for six knots per year (i.e. 72 knots in total), which is the number used by Davey et al. In total this generalised additive model uses 73 degrees of freedom, with one additional degree of freedom required to fit the linear component of time. The appropriate number of knots per year may vary depending on the data. As a sensitivity analysis, the results for 2 and 3 knots per year will also be presented (i.e. degrees of freedom of 25 and 37, respectively).

5.4 Results

During the study period, July 2001 to July 2013, there were 66,436 hospitalisations and 160,294 ED presentations in NSW for children under the age of five years with acute gastroenteritis. There were a total of 2,076 positive laboratory results in the Pathology West Network for the same time period and population. Figure 5.2 displays the weekly number of positive rotavirus laboratory tests in children under five years of age over the study period, July 2001 to June 2013. There was an obvious seasonal peak each year in the pre-vaccination era, however some years appeared to have no obvious seasonal peak after the rollout of the national rotavirus vaccination program in July 2007.

5.4.1 Estimates of rotavirus hospitalisations using the A08.0 rotavirus code

There were a total of 9,479 rotavirus coded hospitalisations over the study period. These hospitalisations made up 14.3% of the total number of acute

Figure 5.2: *Weekly numbers of positive rotavirus laboratory results in the Pathology West Laboratory Network in children under the age of 5 years, 2001-2013.*

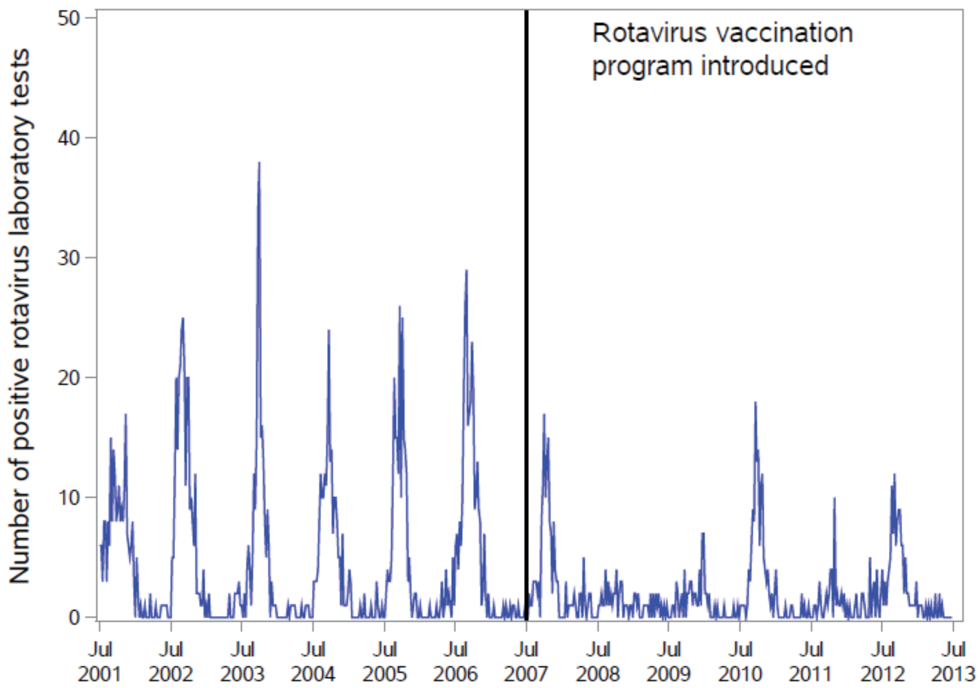


Figure 5.3: *Weekly numbers of hospitalisations with a primary rotavirus coded (A08.0) diagnosis in children under the age of 5 years in NSW, Australia, 2001-2013.*

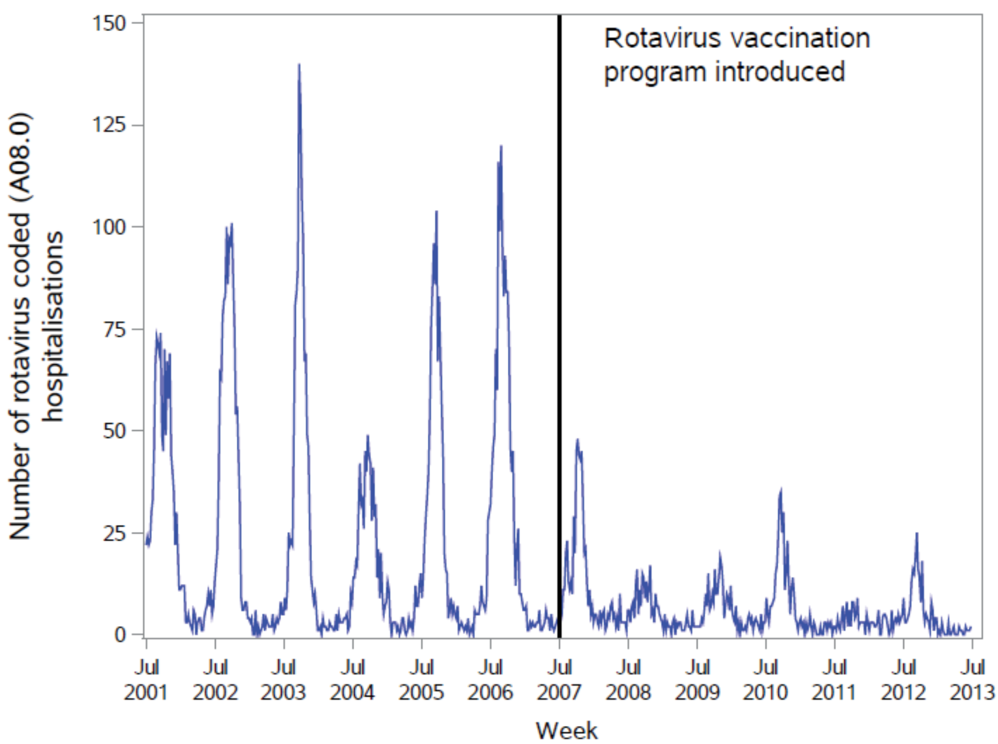


Table 5.1: *Estimated total numbers of rotavirus-attributed hospital and ED cases over the study period in children under the age of five years and as a percentage of the total number of respective acute gastroenteritis encounters using the off-season attributable fraction method.*

Off season rotavirus fraction		Hospitalisations		ED Presentations	
Pre-vaccination era (%)	Vaccination era (%)	Estimated rotavirus cases	Percentage of AGE cases	Estimated rotavirus cases	Percentage of AGE cases
25.3	25.3	33259	50.1	60789	37.9
25.3	12.65	30767	46.3	51963	32.4
25.3	0	28275	42.6	43137	26.9

AGE: Acute gastroenteritis; ED: Emergency Department.

gastroenteritis hospitalisations over the study period. The seasonal peak evident in the pre-vaccination era is smaller in the vaccination era (Figure 5.3).

5.4.2 Estimates of rotavirus-attributed hospitalisations and ED presentations using the off-season attributable fractions method

Among children aged less than five years, there were a total of 22,207 hospitalisations for acute gastroenteritis during the off-peak seasons and 44,229 during the peak seasons across the whole study period. There were 66,603 ED presentations for acute gastroenteritis during the off-peak seasons and 93,691 during the peak seasons. When the off-season rotavirus fraction was assumed to be 23.5% over the entire study period the total number of rotavirus-attributed hospitalisations was estimated to be 33,259 (50.1%) and the total number of rotavirus-attributed ED presentations was estimated to be 60,789 (37.9%) using the off-season attributable fraction method (Table 5.1). These estimates declined to 30,767 (46.3%) and 51, 963 (32.4%), respectively, when the off-season rotavirus fraction in the vaccination era was assumed to be half the off-season

Table 5.2: *Output from the linear regression model.*

	Hospitalisations		ED presentations	
	Coefficient	P-value	Coefficient	P-value
Intercept	66.7 (63.4-70.1)	<0.0001	213.3 (207.8-218.9)	<0.0001
Positive laboratory result	11.88 (11.34-12.41)	<0.0001	12.88 (12.01-13.76)	<0.0001

ED: Emergency Department.

rotavirus fraction in the pre-vaccination era (i.e. 12.65%). The smallest estimates were obtained when the off-season rotavirus fraction was assumed to be zero in the post-vaccination era (28,275 (42.6%) and 43,137 (26.9%), respectively).

5.4.3 Estimates of rotavirus-attributed hospitalisations and ED presentations using the linear regression model

Using the linear regression method of Harris et al. [93], the estimated number of rotavirus-attributed acute gastroenteritis hospitalisations associated with each laboratory report was 11.88 (95% CI: 11.34-12.41) (Table 5.2). The estimated number of non-rotavirus attributed acute gastroenteritis hospitalisations per week was 66.74 (95% CI: 63.36-70.11). Figure 5.4 shows the estimated weekly number of acute gastroenteritis hospitalisations broken down into rotavirus and non-rotavirus hospitalisations using this method. The total number of rotavirus hospitalisations over the study period was estimated to be 24,659 (95% CI: 23,547-25,770), which was 37.1% (95% CI: 35.4-38.8%) of the total number of predicted acute gastroenteritis hospitalisations (Table 5.3).

The estimated number of rotavirus attributed acute gastroenteritis ED presentations associated with each positive laboratory report was 12.88 (95% CI: 12.01-13.76) (Table 5.2). The estimated number of non-rotavirus attributed acute gastroenteritis ED presentations per week was 213.33 (95% CI: 207.79-218.87).

Table 5.3: *Estimated total numbers of rotavirus attributed hospital and ED cases over the study period in children under the age of five years and as a percentage of the total number of respective acute gastroenteritis encounters when using linear regression models or generalised additive models.*

Healthcare setting	Method	Estimated rotavirus cases (95% CI)	Percentage of AGE cases (95% CI)
Hospitalisations			
	Linear regression	24,659 (23,547-25,770)	37.1 (35.4-38.8)
	GAM with 6 knots	13,396 (12,853-13,940)	20.2 (19.3-21.0)
	GAM with 3 knots	19,126 (18,390-19,861)	28.8 (27.7-29.9)
	GAM with 2 knots	21,370 (20,517-22,223)	32.2 (30.9-33.5)
ED presentations			
	Linear regression	26,749 (24,924-28,574)	16.7 (15.5-17.8)
	GAM with 6 knots	15,911 (14,933-16,890)	9.9 (9.3-10.5)
	GAM with 3 knots	22,906 (21,937-24,175)	14.3 (13.5-15.1)
	GAM with 2 knots	25,409 (23,961-26,858)	15.9 (14.9-16.8)

AGE: Acute gastroenteritis; ED: Emergency Department; GAM: Generalised Additive Model.

Figure 5.5 shows the estimated weekly number of acute gastroenteritis ED presentations broken down into rotavirus and non-rotavirus presentations using this method. The total number of rotavirus ED presentations over the study period was estimated to be 26,749 (95% CI: 24,924-28,574), which was 16.7% (95% CI: 15.5-17.8%) of the total number of predicted acute gastroenteritis ED presentations (Table 5.3).

5.4.4 Estimates of rotavirus-attributed hospitalisations and ED presentations using the generalised additive models

Using the generalised additive model of Davey et al. [48] with six knots per year, the estimated number of rotavirus-attributable acute gastroenteritis hospitalisations associated with each laboratory report was 6.45 (95% CI: 6.19-

Figure 5.4: *Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis hospitalisations using a linear regression model for children under five years of age.*

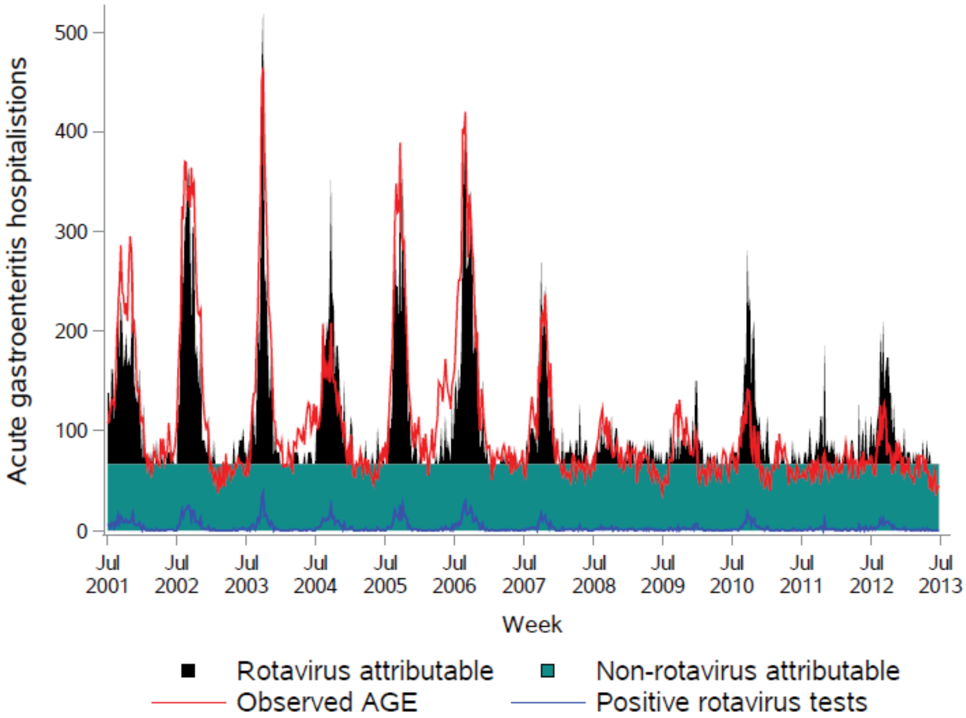


Figure 5.5: *Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis ED presentations using a linear regression model for children under five years of age.*

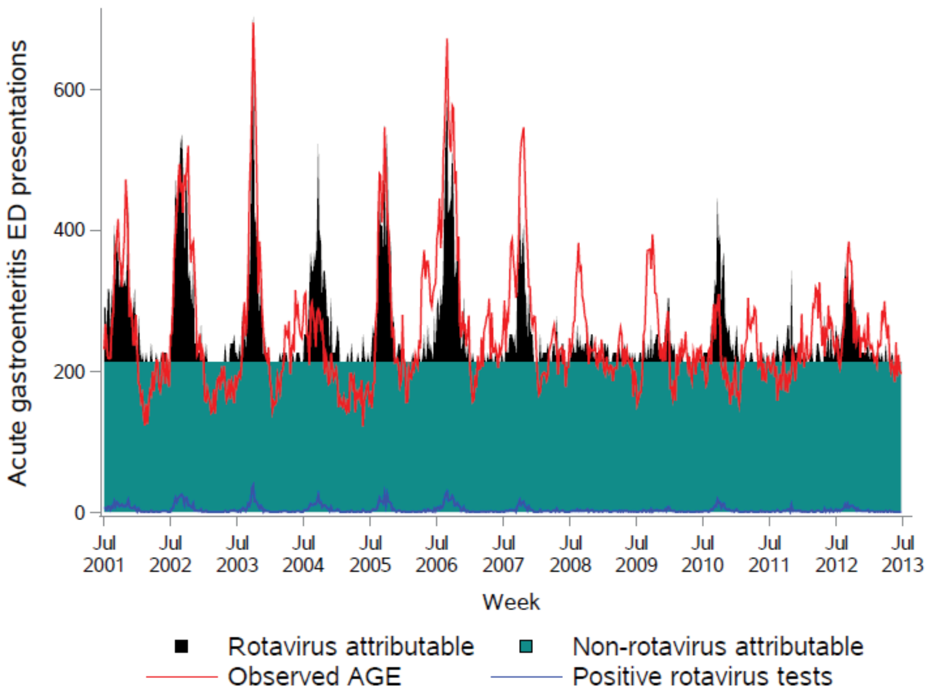


Table 5.4: *Output from the generalised additive models.*

Knots per year	Variable	Hospitalisations		ED presentations	
		Coefficient	P-value	Coefficient	P-value
6	Intercept	121.5 (118.4-124.6)	<0.0001	232.3 (226.8-237.9)	<0.0001
	Positive laboratory result	6.45 (6.19-6.71)	<0.0001	7.66 (7.19-8.13)	<0.0001
	Time (linear)	-0.12 (-0.13- -0.11)	<0.0001	-0.005 (-0.019-0.009)	0.4
	Spline(time)*	0.994	<0.0001	0.994	<0.0001
3	Intercept	105.8 (101.6-110.0)	<0.0001	213.2 (205.9-220.4)	<0.0001
	Positive laboratory result	9.21 (8.86-9.57)	<0.0001	11.03 (10.42-11.64)	<0.0001
	Time (linear)	-0.10 (-0.11- -0.09)	<0.0001	0.020 (0.002-0.038)	0.03
	Spline(time)*	1.000	<0.0001	1.000	<0.0001
2	Intercept	99.7 (94.8-104.5)	<0.0001	206.3 (198.1-214.5)	<0.0001
	Positive laboratory result	10.29 (9.88-10.70)	<0.0001	12.24 (11.54-12.94)	<0.0001
	Time (linear)	-0.09 (-0.10- -0.08)	<0.0001	0.029 (0.009-0.050)	0.006
	Spline(time)*	1.000	<0.0001	1.000	<0.0001

ED: Emergency Department.

* Non-parametric cubic spline of time (the number of weeks since the start of the study)

6.72) (Table 5.4). There was a small negative long-term trend of 6.1 (95% CI: 5.7 -6.6) fewer hospitalisations per year as time increased (p<0.001). Figure 5.6 shows the estimated weekly number of acute gastroenteritis hospitalisations broken down into rotavirus and non-rotavirus hospitalisations using this method. The total number of rotavirus hospitalisations over the study period was estimated to be 13,396 (95% CI: 12,853-13,940), which was 20.2% (95% CI: 19.3-21.0%) of the total number of predicted acute gastroenteritis hospitalisations (Table 5.3).

The sensitivity analyses where the number of knots per year was either two or three instead of six showed that the estimated number of rotavirus-attributable hospitalisations increased as the number of knots per year decreased

Figure 5.6: *Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis hospitalisations using a generalised additive model with 6 knots per year.*

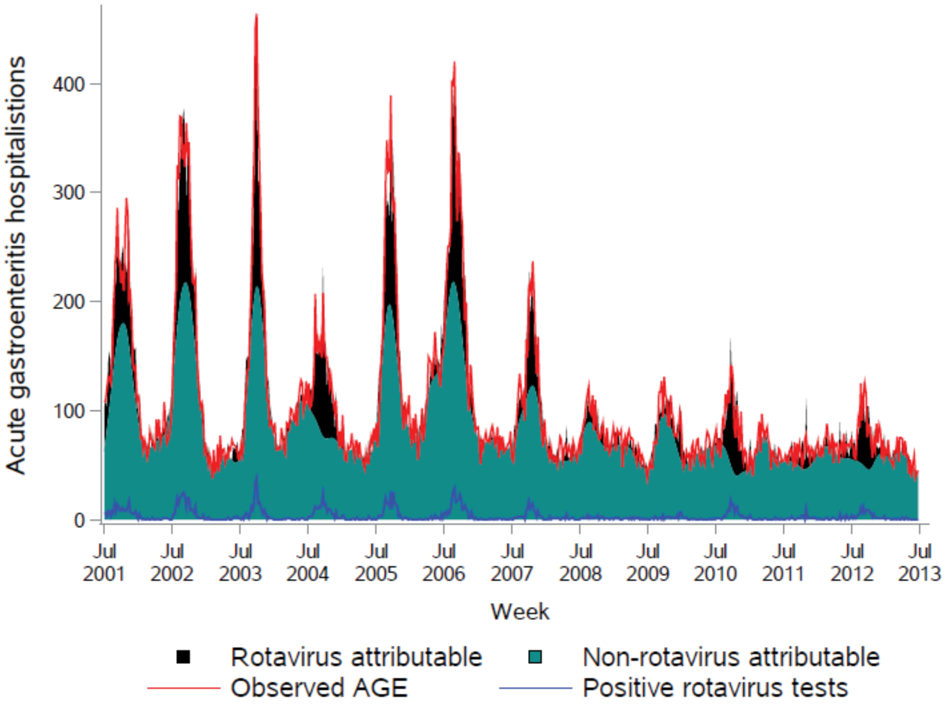


Figure 5.7: *Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis hospitalisations using a generalised additive model with 2 knots per year.*

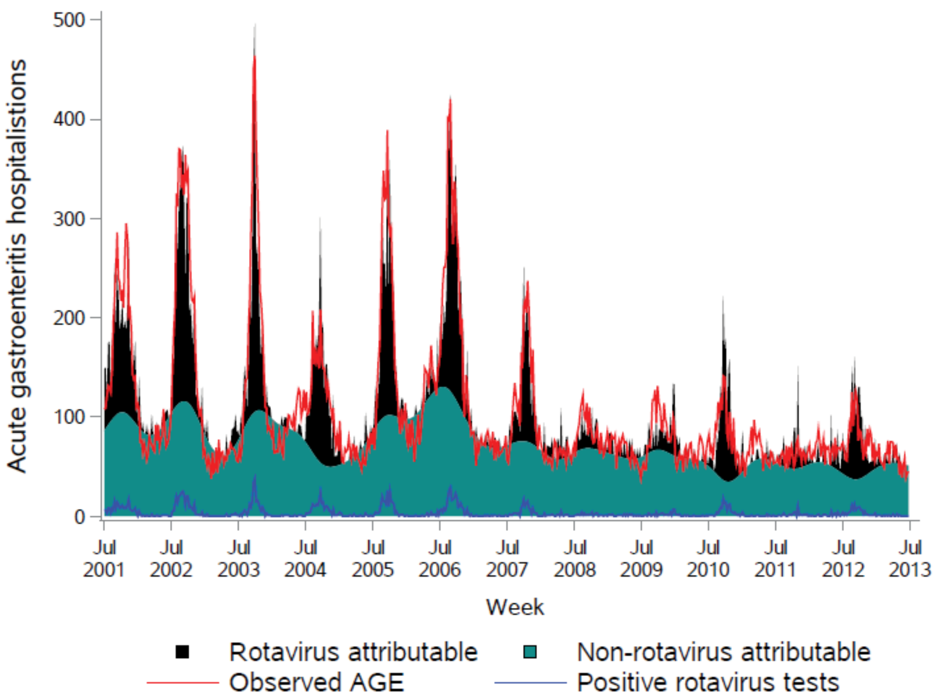


Figure 5.8: *Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis ED presentations using a generalised additive model with 6 knots.*

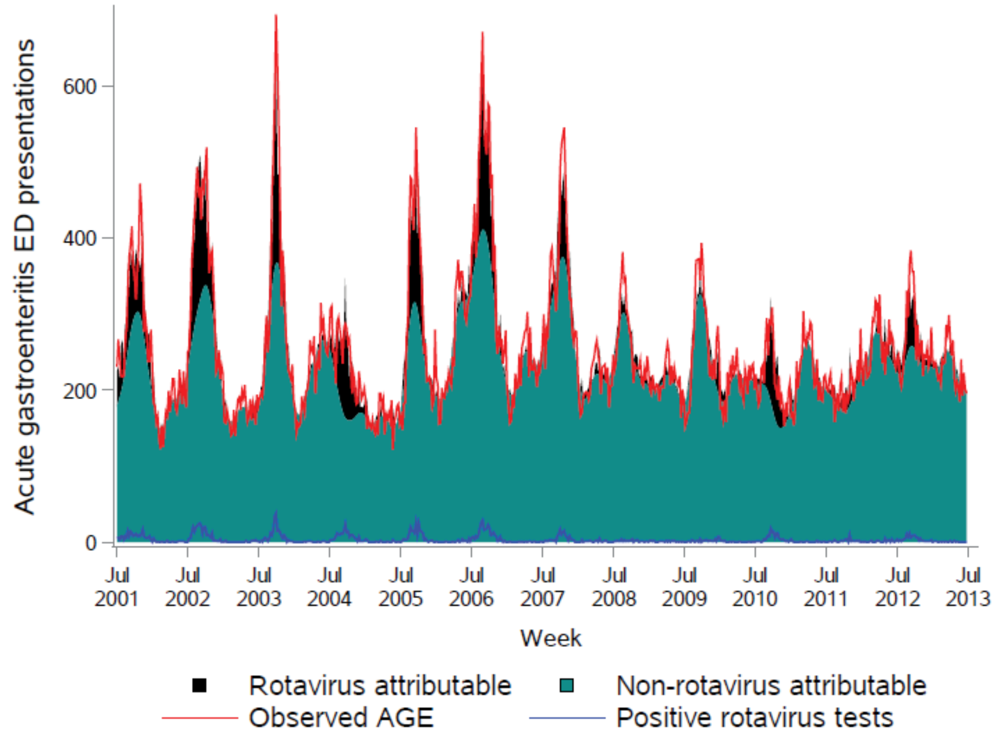
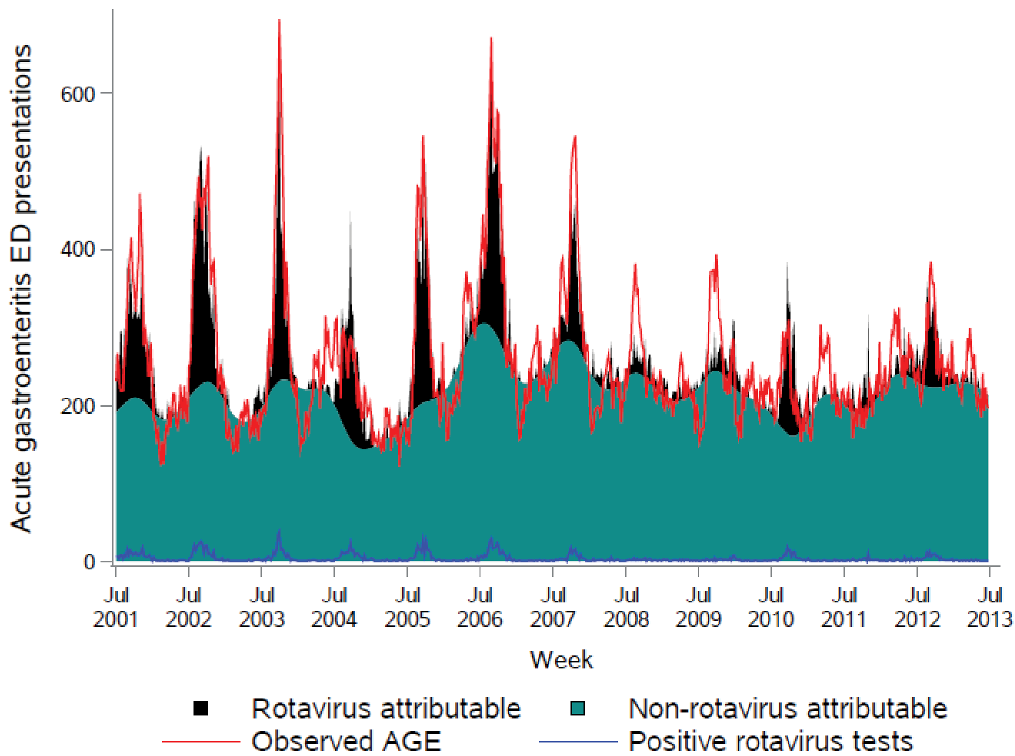


Figure 5.9: *Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis ED presentations using a generalised additive model with 2 knots.*



(Table 5.3). Figure 5.7 shows the estimated weekly number of acute gastroenteritis hospitalisations broken down into rotavirus and non-rotavirus hospitalisations when only 2 knots per year were used.

The estimated number of rotavirus-attributed acute gastroenteritis ED presentations associated with each laboratory report was 7.66 (95% CI: 7.19-8.14) (Table 5.4) when fitting the model with six knots per year. There was a small, negative long-term trend of 0.26 fewer ED presentations per year as time increased (95% CI: 1.0 fewer to 0.5 more ED presentations per year, p=0.4). Figures 5.8 and 5.9 show the estimated weekly number of acute gastroenteritis ED presentations broken down into rotavirus and non-rotavirus presentations using this method for six knots and two knots, respectively. The total number of rotavirus ED presentations over the study period was estimated to be 15,911 (95% CI: 14,933-16,890) from the model with six knots per year, which was 9.9% (95% CI: 9.3-10.5%) of the total number of predicted acute gastroenteritis hospitalisations (Table 5.3). When only two knots per year were used, these estimates increased to 25,409 (95% CI: 23,961-26,858) and 15.9% (95% CI: 14.9-16.8), respectively.

5.5 Discussion

Rotavirus burden estimates have been used previously to evaluate the costs and benefits associated with rotavirus prevention. The four methods to categorise acute gastroenteritis hospitalisations and ED presentations into those that are due to rotavirus and those that are not due to rotavirus produced rather varied results. The proportion of rotavirus coded (A08.0) hospitalisations from all acute gastroenteritis hospitalisations of 14.3% should underestimate the true proportion of rotavirus-attributed acute gastroenteritis hospitalisations, since it requires confirmation of rotavirus disease to assign this code and this code has been found to be highly specific but only moderately sensitive in practice in Australia [13]. All other estimates of the proportion of rotavirus-attributed hospitalisations were higher than this percentage (Range: 20.2% to 50.1%). The estimated proportion of rotavirus-attributed ED presentations ranged from 9.9% to 37.9%. For both healthcare settings (i.e. hospital and ED), the off-peak season attributable fraction method produced the highest estimates of the number of rotavirus cases, even in the most conservative scenario where it was assumed that there were no off-peak season rotavirus cases in the vaccination program era. The generalised additive model with six knots per year produced the smallest estimates (excluding the estimate using rotavirus-coded hospitalisations) and these were less than half the estimates produced from using the off-season attributable fraction method (Tables 5.1 and 5.3).

Each method has strengths and limitations. The off-season attributable fraction method is straightforward to apply. It only requires an estimate of the percentage of rotavirus cases in the off-peak season on top of the absolute number of acute gastroenteritis presentations separated by peak and off-peak season. However, this estimate may not be readily available. The estimate of 25.3% used in the pre-vaccination era for this study, which comes directly from Galati et al.'s study [56], should be fairly accurate for the hospital data used in this study. To date, no estimate of the off-season rotavirus proportion in Australia is available for the vaccination era and it is possible that this value has changed through vaccination efforts.

This study produced results for three scenarios using the off-season attributable fractions method, where the off-season rotavirus percentage was 0, 12.65 and 23.5, which resulted in the estimated percentage of rotavirus cases over the entire study period ranging from 42.6 to 50.1 of hospitalisations and

26.9 to 37.9 of ED presentations (Table 5.1). In addition, the assumption that rotavirus is the only seasonal disease that causes acute gastroenteritis and that the number of non-seasonal rotavirus cases in the off season also occur in the peak season is unlikely to be accurate. Harris et al. [93] found that a number of pathogens that cause infectious intestinal disease exhibit seasonal patterns that peak at different times of the year. E. coli and salmonella peaked in the summer weeks, rotavirus and astrovirus peaked in winter and early spring and cryptosporidium peaked both in late spring and in summer. Another limitation to this method is that it does not take into account the uncertainty in the (estimated) off-peak season fraction used in the calculations.

The linear regression model and generalised additive models used weekly counts of laboratory confirmed rotavirus as an explanatory variable to estimate the acute gastroenteritis hospitalisations and ED presentations that are attributed to rotavirus. This assumes that all of the rotavirus activity in children under five years of age in NSW is explained by these positive laboratory counts, which likely depends on the quality of these data. Further, the final rotavirus estimates involved multiplying the weekly positive rotavirus laboratory counts by the laboratory coefficient (β). This assumes that the ratio of positive rotavirus laboratory results to rotavirus-attributed hospitalisations and ED presentations is constant over time, which may not be the case. It is possible that this ratio was influenced by potential changes in the rate of laboratory testing over time or it may vary with seasonal variation throughout the year (e.g. the rate may be different in the off-season weeks as compared to the peak weeks of rotavirus burden). It is also possible that this ratio of the number of rotavirus cases associated with every positive laboratory results has changed with introduction of the national rotavirus vaccination program. If this is the case, then there is additional uncertainty in these estimates of rotavirus attributed acute gastroenteritis, as was seen when the off-season rotavirus fraction was varied in the vaccination program era for the off-season attributable fractions method (Tables 5.1).

The generalised additive model aimed to determine β independent of the temporal trends evident in the data. It is possible that the model with six knots per year over fitted the model and resulted in too low an estimate of the total number of rotavirus-attributed healthcare encounters, since the estimate of rotavirus attributable hospitalisations was close to the rotavirus-coded A08.0 estimate and only a small portion of the seasonal peaks were attributed to rotavirus (Figures 5.6-5.9). Determining the correct number of knots to use when specifying a generalised additive model is difficult.

The linear regression method did not try to determine β independent of the temporal trends in the data. It assumed that the seasonal variation and autocorrelation in the acute gastroenteritis hospitalisations and ED presentations was fully explained by the positive rotavirus laboratory results. It is not clear which of these two approaches for accounting for temporal trends is most appropriate when trying to determine the absolute numbers of acute gastroenteritis cases that are due to rotavirus.

The estimates in this study produced using the off-season attributable fraction method are consistent with those published by Galati et al. [56]. They found that 52% of hospitalisations and 42% of ED presentations for acute gastroenteritis in children under five years of age in Australia are due to rotavirus (Galati 2006). This study found these estimates to be 50.1% and 37.9%, respectively, when assuming that the proportion of rotavirus cases in the off-season is constant over the study period and equal to 25.3% (as used in Galati et al.).

The methods that fitted statistical models (i.e. linear regression and generalised additive models) had higher estimates of the percentage of rotavirus attributable healthcare encounters than this study. Harris et al. found that 51.4% (95% CI: 49.5-53.2%) of hospitalisations (1995-2003) and 21.5% (95% CI: 16.8-26.1%) of ED presentations (in 2004) for all cause gastroenteritis in children under five years of age in England and Wales were attributed to rotavirus when fitting a model with the same coefficients as this study (i.e. only an intercept and number of positive rotavirus laboratory tests) [93]. The estimates from this study were 37.1% (95% CI: 35.4-38.8) and 16.7% (95% CI: 15.5-17.8%), respectively. Possible explanations for this are the amount of ED data available in Harris et al.'s study (2004 only) and the fact that the hospitalisation estimates are based on data from five hospitals in the Greater London area only. Davey et al. found that 19.3% of the acute gastroenteritis ED presentations over their study period (2003-2011) in children under five years of age in NSW were attributed to rotavirus [48], but this study estimated it to be only 9.9% (95% CI: 9.3-10.5%) when using the same number of knots per year. One possible explanation is that Davey et al. had access to more ED and laboratory data than this study (more details given in the following paragraph). Their data may have had fewer zero counts and more pronounced seasonal peaks than our data, which may explain why they identified a stronger association between positive rotavirus results and rotavirus cases than this study.

A limitation of this study was only having access to laboratory data from one pathology network in NSW, whereas Davey et al. had access to 10 public laboratory networks in NSW. As a result the positive rotavirus pathology data in this study may not be representative of the positive rotavirus laboratory data in the whole state of NSW. The seasonal patterns exhibited in the weekly positive laboratory counts are similar to those seen in the rotavirus-coded hospitalisations (Figures 5.1 and 5.2), with both data sources showing peaks in 2007, 2010 and

2012 after introduction of the vaccination program. This study only compared the methods currently used in the published literature to determine the proportion of acute gastroenteritis hospitalisations and ED presentation in NSW that are attributable to rotavirus. There are a number of alternative methods that could have been used to calculate this proportion and future studies may wish to consider these.

A major strength of this study was the ability to directly compare the absolute number of acute gastroenteritis hospitalisations and ED presentations attributed to rotavirus from the same data for the four different methods. This gave a better understanding of the true uncertainty in estimating the proportion of acute gastroenteritis hospitalisations and ED presentations that is due to rotavirus than using one method. It is unclear from the results which method performs best. It is possible to rewrite the three statistical models compared in this chapter using sub-models of a generalised model and then compare their model fit using AIC or likelihood ratio tests [40]. However, this comparison would not identify which method was the most accurate representation of the true numbers of rotavirus hospitalisations and ED presentations in NSW.

5.6 Chapter Summary

This chapter provided a direct comparison of the estimated number of rotavirus-attributable hospitalisations and ED presentations in children under five years of age in NSW when applying four different methods to the same data. The results were diverse across these methods, demonstrating the importance of methodological choices. It is not clear which method was closest to the true numbers of rotavirus hospitalisations and ED presentations. The next chapter will use simulated data to explore these methods in controlled scenarios to determine the most appropriate method(s) for estimating the true rotavirus disease burden. It will explore how these methods perform when the ratio of positive rotavirus laboratory results to rotavirus-attributed cases changes over time.

Chapter 6: Simulations to explore methods for estimating the true rotavirus disease burden

6.1 Chapter Overview

The previous chapter directly compared four different methods to estimate the number of rotavirus-attributable hospitalisations and ED presentations in children under five years of age in NSW. The results varied across these methods and it is not clear which method provides the best estimates of the true numbers of rotavirus hospitalisations and ED presentations in NSW for this age group. This chapter uses simulations to explore these four methods in controlled scenarios to determine the most appropriate method for estimating the true rotavirus disease burden for our data. In addition, it explores the assumptions around the four methods and how the methods perform when their assumptions are not met.

6.2 Introduction

The true rotavirus disease burden of hospitalisations and ED presentations in children under the age of five years for acute gastroenteritis in NSW is not known. That is, we do not know how many of the hospitalisations and ED presentations for children under five years of age in NSW each week are due to rotavirus and how many are attributed to non-rotavirus diseases. Four methods were compared in Chapter 5 to estimate the percentage of acute gastroenteritis ED presentations and hospitalisations in children under five years of age in NSW that were due to rotavirus (Jul 2001 to Jul 2013). The results of the methods varied considerably and it is difficult to conclude which method is most appropriate to estimate the true percentage of rotavirus disease burden within these acute gastroenteritis hospitalisations and ED presentations.

Some of the methods used positive laboratory results to infer rotavirus activity. These methods assumed that the relationship between observed positive laboratory results and rotavirus cases was fixed and did not vary over time. It is not clear how sensitive or robust these methods are when this relationship does not meet this assumption. It is also not clear how these statistical models perform when the laboratory data is sparse or the relationship is not strong. The remaining method assumed that there was no seasonal component to non-rotavirus cases throughout the whole year. Harris et al. [93] found that a number of non-rotavirus pathogens that cause infectious intestinal disease exhibit seasonal patterns in England and Wales that peak at different times of the year, including during peak rotavirus season. It is possible that there is seasonality in non-rotavirus cases, including in the peak-season, in Australia as well. If this is the case, we do not know how much of the seasonal peaks each winter are due to rotavirus cases and how much are due to non-rotavirus cases. This chapter will use a simulation study to compare the methods presented in Chapter 5 when the true distributions of rotavirus and non-rotavirus cases are known.

Simulation studies use computing intensive, step-by-step procedures to explore the approximate behaviour of a system [94, 95]. They are an important tool for investigating systems that are difficult or impossible to study using real data. They were first introduced in the fields of meteorology and nuclear physics, and today they are used in many fields including public health [96]. Among other purposes, simulation studies can be used to evaluate the robustness of a statistical method under ideal and non-ideal conditions and to offer clarification on the strengths and weaknesses of competing methods. Simulation studies are used to generate datasets that conform to a set of known values or relationships that are specified by the investigators. The accuracy of the model computed parameter estimates or results are compared to their specified known values to determine how adequately the models perform under the specific conditions and to determine if some methods offer greater accuracy than others.

In this chapter we will develop scenarios to assess the appropriateness and accuracy of the statistical methods presented in Chapter 5 in relation to known rotavirus attributed acute gastroenteritis cases, non-rotavirus attributed acute gastroenteritis cases and positive rotavirus laboratory counts over time that we specify for each scenario. The scenarios developed try to reflect the complexities in these data signals that we see in practice. There are many features of the three data series (weekly rotavirus cases, weekly non-rotavirus cases and weekly positive laboratory results) that needed to be specified when defining each scenario. For example:

- Is there seasonality in any or all series? If yes, how high and wide are the peaks?
- Should there be a non-seasonal component in any of the series? If yes, how large?
- Is the ratio of the number of rotavirus cases per positive laboratory result in a given week constant over the study period? If not, does it vary with introduction of the vaccination program, or by season (peak-season, off-season)?
- Is there a increasing or decreasing trend prior to vaccination efforts?

After the scenarios are defined, independent simulated datasets are generated from the known truth of each scenario. The statistical methods from Chapter 5 and additional methods are applied to each of the random samples and their estimates of the percentage of acute gastroenteritis cases attributed to

Box 6.1: *Summary of the simulation study process carried out in this chapter.*

1. Define a set of assumptions about the nature of rotavirus cases, non-rotavirus cases and positive laboratory results over time. This is referred to in the text as a scenario.
2. Randomly generate a dataset according to these assumptions. This is referred to in the text as a simulation.
3. Apply a statistical method to separate the acute gastroenteritis cases over time into those that are due to rotavirus and those that are not. Calculate and retain the percentage of rotavirus attributed acute gastroenteritis cases separately for the pre-program and program era. If the method generated a model coefficient(s) to calculate these percentages then retain this value(s) also.
4. Repeat step 3 for all statistical methods to be compared.
5. Repeat steps 2 to 4 for 1000 simulations.
6. Report the median of the 1000 simulations and the lower and upper 2.5th percentiles of the pre-program and program era percentages and model coefficient estimates, where applicable.
7. The assumptions in step 1 are modified and steps 2 to 6 are repeated using the new assumptions for all 16 scenarios.
8. The results across methods are compared to the assumptions of their corresponding scenario to evaluate if there is a robust method(s).

rotavirus, stratified by pre- and post- vaccination program introduction are collated. The specific details are given in the following three sections (Sections 6.3-6.5). An overview of the process is itemised in Box 6.1. The objective of this chapter is to determine if any method(s) is robust to the attributes we think these disease signals may potentially have. That is, the objective is to identify a method(s) that is able to correctly estimate the true percentage of rotavirus attributed acute gastroenteritis cases both when the assumptions of that method are met and when there are small, but deliberate, variations in model parameters (that may or may not affect the assumptions of that method).

6.3 Scenarios

The first step in this simulation study was to assign the known truth for a number of scenarios, which allocated rotavirus and non-rotavirus disease burden and corresponding positive laboratory results over time. Sixteen scenarios were constructed to explore the robustness and uncertainty of the four methods applied in chapter 5 that estimated the rotavirus disease burden in each of the hospital and ED datasets. These scenarios were chosen to try to reflect the complexities in these data signals that we see in practice, to highlight the strengths and weaknesses of the various methodological approaches, and to demonstrate the sensitivity in parameter specification for the various methods. These sixteen scenarios do not have equal weight in determining which method(s) perform best. Nor do they cover all components of the various methods or all possible permutations of the rotavirus, non-rotavirus and positive laboratory data signals. A method will be considered robust if it contains the true percentage of rotavirus attributed acute gastroenteritis cases in its error margin (defined in sub-section 6.5.10) i) for scenarios where its assumptions are met and ii) for scenarios that are similar to the complexities exhibited in our hospital and ED acute gastroenteritis data.

A summary of the scenarios is given in Table 6.1 and the details are below. A visual image of the defined known truth for each of the sixteen scenarios in numerical order is given in Figures 6.1 to 6.16, respectively. The

Table 6.1: *Summary of the sixteen scenarios explored in this chapter.*

	Pre-program introduction era					Vaccination program era			
Scenario	Magnitude of seasonal rotavirus each year [1]	Number of non-seasonal rotavirus counts each week	Number of non-seasonal non-rotavirus counts each week	Magnitude of non-seasonal rotavirus each year [1]	Ratio of rotavirus cases per positive laboratory result in a given week	Percentage reduction in rotavirus cases [2]	Percentage reduction in positive laboratory results [2]	Ratio of rotavirus cases per positive laboratory result	Other
1	250	25	75	-	12	75	75	12	
2	250	25	75	-	12	75	75	4	
3	250	25	75	-	peak-season: 12 off-season: 6	75	75	peak-season: 5 off-season: 3	Ratio varies by season and program introduction
4	250	100	75	-	12	75	75	12	
5	250	0	100	-	6	75	75	6	
6	250	0	100	-	6 [a]	75	75	6 [a]	Positive laboratory results lag by one week
7	250	25	75	-	Independent [b,c]	75	40-70 [c]	Independent [b,c]	Rotavirus and laboratory seasonality independent
8	250	25	75	-	Independent [d,c]	75	40-70 [c]	Independent [d,c]	Rotavirus and laboratory seasonality independent
9	Various [e]	25	75	-	6	50	50	6	
10	Various [e]	25	75	-	Various [e]	50	50	Various [e]	
11	200	25	100	25	6	75	75	6	
12	200	25	100	50	6	75	75	6	Non-rotavirus cases seasonal in winter and summer.
13	250 [f]	25 [g]	75 [g]	-	6 [g]	75	75	6 [g]	Decreasing linear trend in rotavirus cases
14	250 [f]	25	75	-	6	75	75	6	Decreasing linear trend in magnitude of seasonal RV
15	Various [e]	25	75	Various [e]	Various [e]	75	50	Various [e]	
16	Various [e]	25	75	Various [e]	Various [e]	75	50	Various [e]	Non-rotavirus cases seasonal in winter and summer.

RV: Rotavirus

[1] The magnitude is the difference between the seasonal peak and trough.

[2] As compared to the pre-program introduction era.

[a] With a one week delay.

[b] Timing of seasonal peak and width coincide, magnitudes are independent.

[c] Varies by week.

[d] Timing of seasonal peak coincides, magnitude and width of seasonal components are independent.

[e] Varies by year.

[f] Value for the first year. A decreasing trend per week applies from the first week of the study.

[g] Value for the first week. A decreasing trend per week applies from the first week of the study.

black shaded area in each figure is the true rotavirus disease burden over time and the green area is the true non-rotavirus disease burden. The total of these two areas is the acute gastroenteritis disease burden. The blue signal is the number of positive rotavirus laboratory results each week.

The study period for all scenarios was twelve years of weekly data with a rotavirus vaccination program introduced after six years, in week 313 (beginning of year 7) of 624 weeks (end of year 12). The middle 26 weeks of each year are defined to be the peak-season and the first and last 13 weeks of each year the off-season. We assumed each of the three data series of interest (weekly counts of rotavirus attributed acute gastroenteritis cases, weekly counts of non-rotavirus attributed acute gastroenteritis cases and weekly counts of positive rotavirus laboratory results each week), which are count data, were Poisson distributed.

For a given week, t, the number of rotavirus cases, X_t, non-rotavirus cases, Y_t, and positive laboratory results, Z_t, were generated from Poisson distributions with means of μ_t, φ_t and ω_t, respectively. That is,

$$X_t \sim Poisson(\mu_t)$$
$$Y_t \sim Poisson(\varphi_t)$$
$$Z_t \sim Poisson(\omega_t).$$

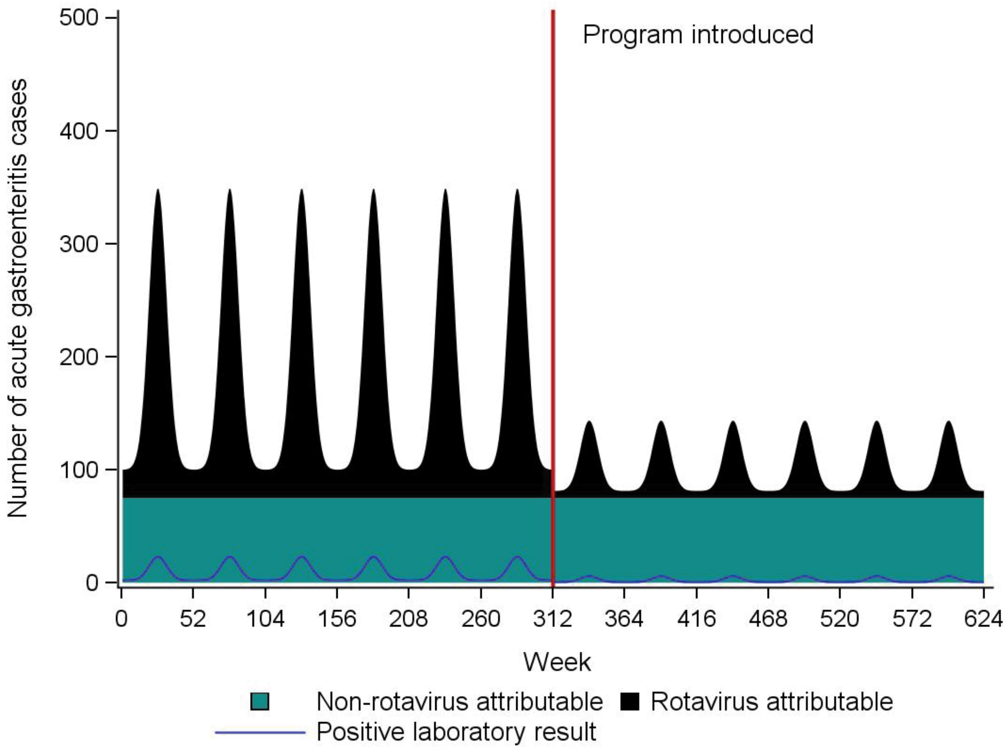
The mean of each Poisson distribution needs to be specified for each week (i.e. the μ_t, φ_t and ω_t). This was done in the scenario development (Sub-sections 6.3.1 to 6.3.16). Then 1000 random samples were drawn from the known truth of each scenario (details in Section 6.4). Multiple methods to separate acute gastroenteritis cases into those that are due to rotavirus and those that are not due to rotavirus were applied to each of the 1000 random samples and their estimates of the percentage of acute gastroenteritis cases attributed to rotavirus, stratified by pre- and post- program introduction were collated (details in Section 6.5).

6.3.1 Scenario One

Scenario 1 was developed with the features of the hospital data used in this thesis in mind (Figure 6.1). First the parameters for the pre-program era (weeks 1 to 312) were specified. It was assumed that there was no seasonality in the non-rotavirus cases and that these were constant at 75 cases per week and did not change with introduction of the vaccination program in week 313. The rotavirus cases were assumed to have a non-seasonal and seasonal component. The magnitude of the seasonal peak (i.e. the difference between the seasonal peak and trough) was set at 250 rotavirus cases in the pre-program era and the peak occurred in the middle weeks of each year. This annual seasonality was symmetric and fit using the probability density function of a standard normal curve with 52 evenly spaced increments between -4 and 4. It was transformed such that the peak equated to 250. The number of non-seasonal rotavirus cases was 25 per week. It was assumed that the total number of rotavirus cases (seasonal and non-seasonal) decreased by 75% immediately after the introduction of the vaccination program. Therefore the magnitude of the seasonal peak reduced to 62.5 rotavirus cases and the number of non-seasonal rotavirus cases reduced to 6.25 in the program era. The numbers of seasonal and non-seasonal rotavirus cases each week were added together to give the total number of rotavirus cases per week over the study period. Finally, it was assumed that every positive laboratory report represented twelve rotavirus cases in a given week over the entire study period. Therefore, the laboratory data was set at one twelfth of the rotavirus cases and they also declined by 75% through vaccination efforts.

Recall that these numbers of rotavirus cases, non-rotavirus cases and positive laboratory reports in each week define the means for the Poisson distributions of X_t, Y_t and Z_t, respectively (i.e. the μ_t, φ_t and ω_t, respectively). Therefore it is acceptable that these values are not integers (but they must be non-negative) even though the observed data are integers. This is important so that we can control the true value of the ratio of rotavirus cases to positive laboratory results each week and the decline due to vaccination efforts. The counts produced from these distributions when the 1000 simulations are generated will yield integers and the ratio and percentage declines due to vaccination efforts will vary by simulation but on average these will be 12 and 75%, respectively.

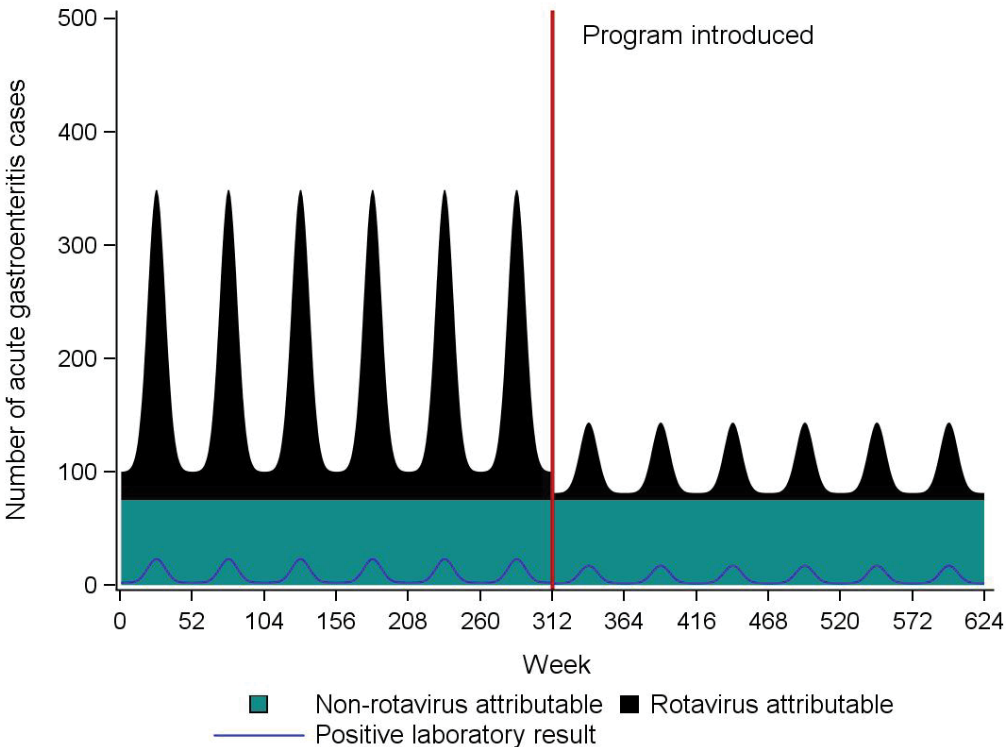
Figure 6.1: *Scenario 1: ratio of weekly rotavirus cases per positive laboratory result is constant over the study period.*



6.3.2 Scenario Two

To explore how the various methods perform when the ratio of weekly rotavirus cases per positive laboratory result in the same week is allowed to vary between the pre-program and program eras, Scenario 2 was defined to be the same as Scenario 1 (Sub-section 6.3.1) except that the ratio of the number of rotavirus cases for every positive laboratory result was reduced from 12 to 4 per week in the program era (Figure 6.2).

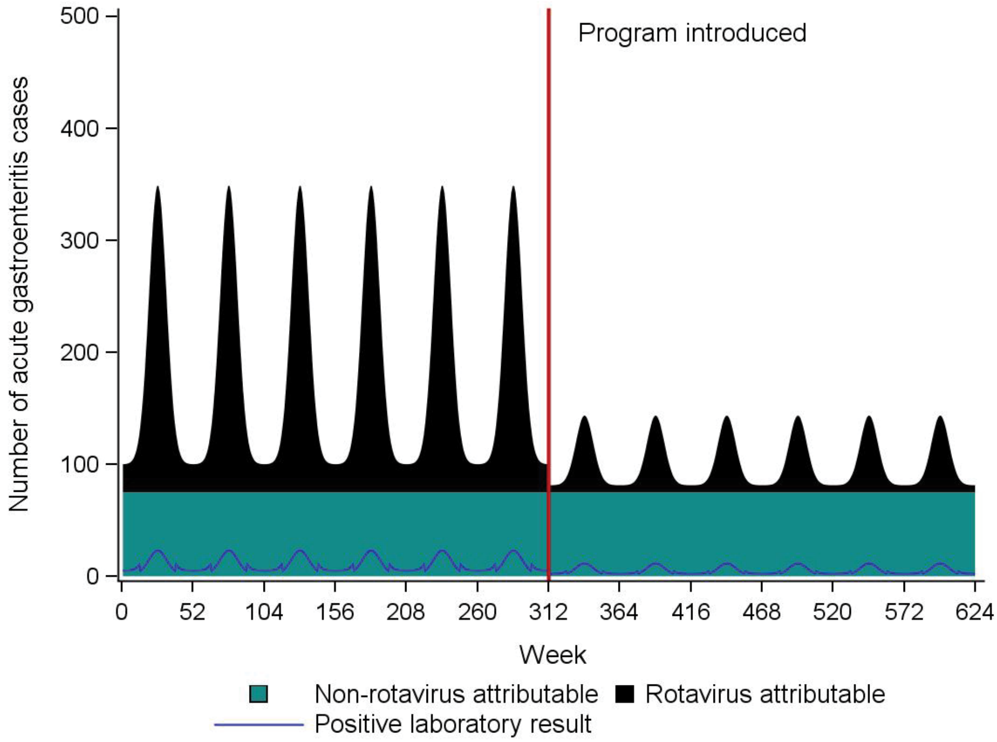
Figure 6.2: *Scenario 2: ratio of rotavirus cases per positive laboratory result varies by program introduction (pre/post).*



6.3.3 Scenario Three

This scenario explored the performance of the various methods when the ratio of weekly rotavirus cases per positive laboratory result varied between the pre-program and program eras and in peak-season and off-season. Scenario 3 was the same as Scenario 1 (Sub-section 6.3.1) except that the ratio of weekly rotavirus cases per positive laboratory result was 12 during the peak-season and 6 during the off-season in the pre-program era (Figure 6.3). It was 5 during the peak-season and 3 during the off-season in the program era.

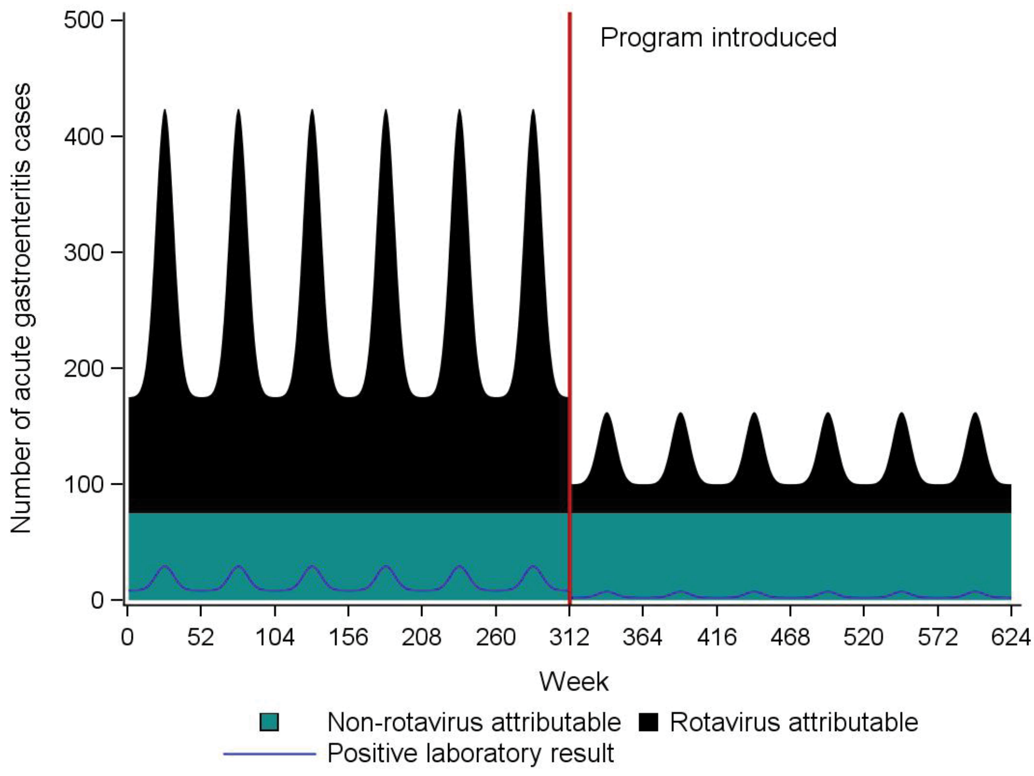
Figure 6.3: *Scenario 3: ratio of weekly rotavirus cases per positive laboratory result varies by season (peak/off-peak) and program introduction (pre/post).*



6.3.4 Scenario Four

This scenario explored an increase in the number of non-seasonal rotavirus cases, which is observed in the ED data (see Chapter 4 and Figure 4.8). Scenario 4 was the same as Scenario 1 (Sub-section 6.3.1) except that the number of non-seasonal rotavirus counts each week in the pre-program era was increased to 100 from 25 (Figure 6.4). Since there was a 75% reduction in rotavirus cases due to vaccination efforts, the mean number of non-seasonal rotavirus counts in the program era was 25 per week.

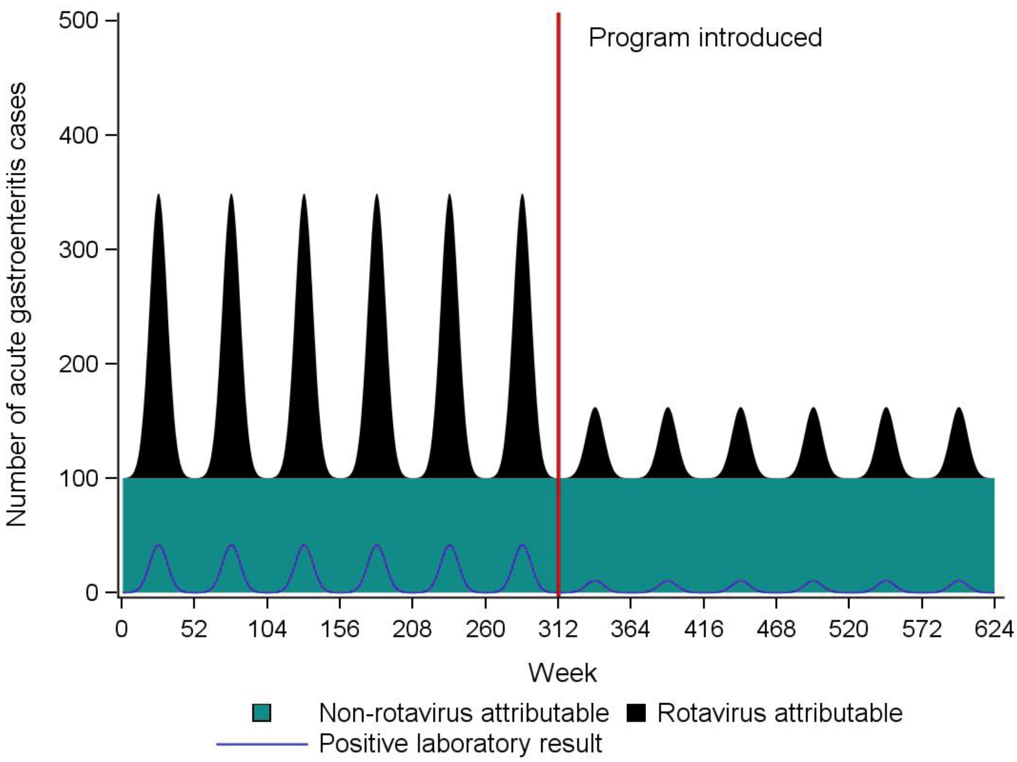
Figure 6.4: *Scenario 4: increase in non-seasonal rotavirus cases.*



6.3.5 Scenario Five

This scenario contrasted how the various methods perform when there is no non-seasonal rotavirus counts as compared with the previous scenarios. Scenario Five was similar to Scenario 1 (Sub-section 6.3.1), and the differences were i) there were no non-seasonal rotavirus counts over the entire study period, ii) the number of non-seasonal non-rotavirus cases each week increased from 75 to 100 cases per week, and iii) the ratio of weekly rotavirus cases per positive laboratory result declined from 12 rotavirus cases per one positive laboratory result to 6 rotavirus cases for every one laboratory result per week over the entire study period (Figure 6.5).

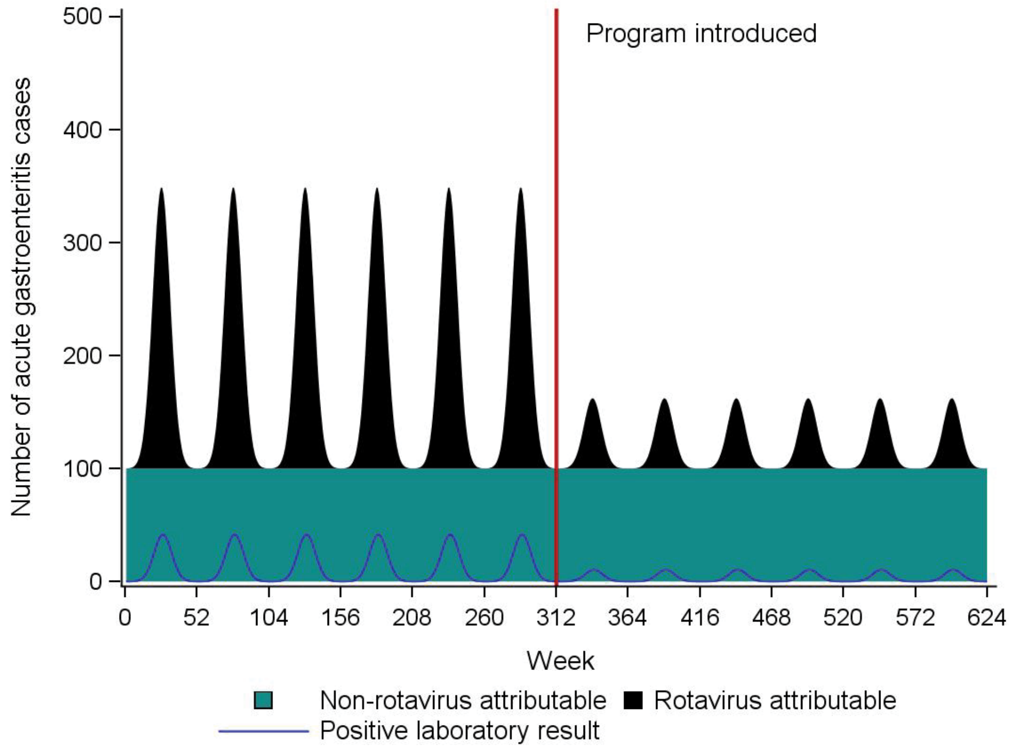
Figure 6.5: *Scenario 5: no non-seasonal rotavirus cases.*



6.3.6 Scenario Six

This scenario investigated if a delay in rotavirus laboratory results affected any of the methods that used laboratory data to identify rotavirus-attributed cases. Scenario 6 was exactly the same as Scenario 5 (Sub-section 6.3.5) except that the positive laboratory results were lagged by one week (Figure 6.6). That is, the number of rotavirus cases in week t, for example, was six times the number of positive laboratory results in week $t + 1$.

Figure 6.6: *Scenario 6: Scenario 5 with positive laboratory results lagged by one week.*



6.3.7 Scenario Seven

This scenario explored how the various methods performed when there was no fixed relationship between positive laboratory results and rotavirus cases in a given week, but where their seasonal pattern was the same. That is, it explores how the methods perform when the number of laboratory results is not a multiplicative factor of rotavirus cases. Scenario 7 is the same as Scenario 1 in all aspects except the laboratory data (Figure 6.7). The number of positive laboratory results each week was independent of the weekly number of rotavirus cases in both the pre- and program era. The timing and width of the seasonal peak in the laboratory data was the same as the timing and width of the seasonal peak in the rotavirus cases; however the magnitudes were independent. That is, there was correlation between the laboratory and rotavirus data). The mean number of non-seasonal positive laboratory results was 5 per week in the pre-program era and 3 per week in the program era. The magnitude of the seasonal peak in positive laboratory results was 55 in the pre-program era and 15 in the program era. This resulted in a percentage reduction in positive laboratory results of 40 in the seasonal trough and 70 at the seasonal peak, as compared with the pre-program laboratory data.

6.3.8 Scenario Eight

This scenario also explored how the various methods performed when there was no fixed relationship between positive laboratory results and rotavirus cases in a given week. Scenario 8 was exactly the same as Scenario 7 (Sub-section 6.3.7) except that the width of the seasonal peak of the positive laboratory counts over the study period was halved (Figure 6.8).

Figure 6.7: *Scenario 7: no relationship between weekly positive laboratory results and rotavirus cases (seasonal peaks and widths coincide, seasonal magnitudes vary).*

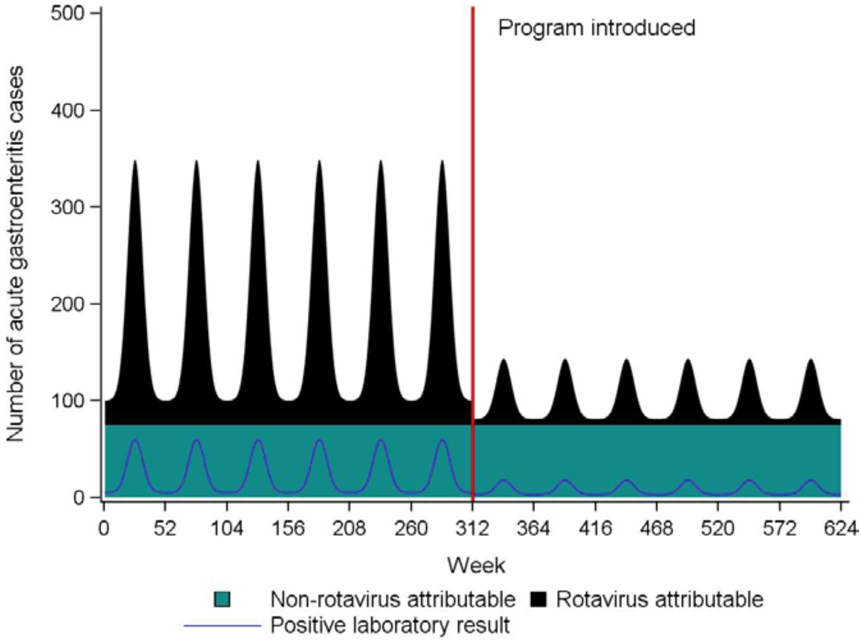
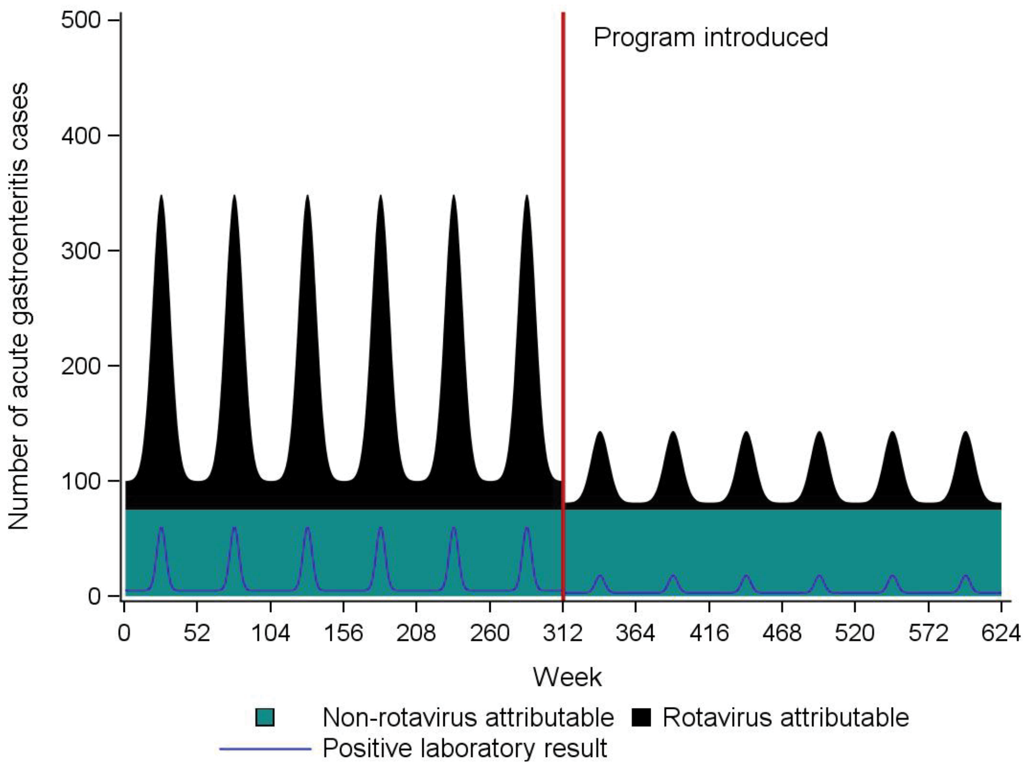


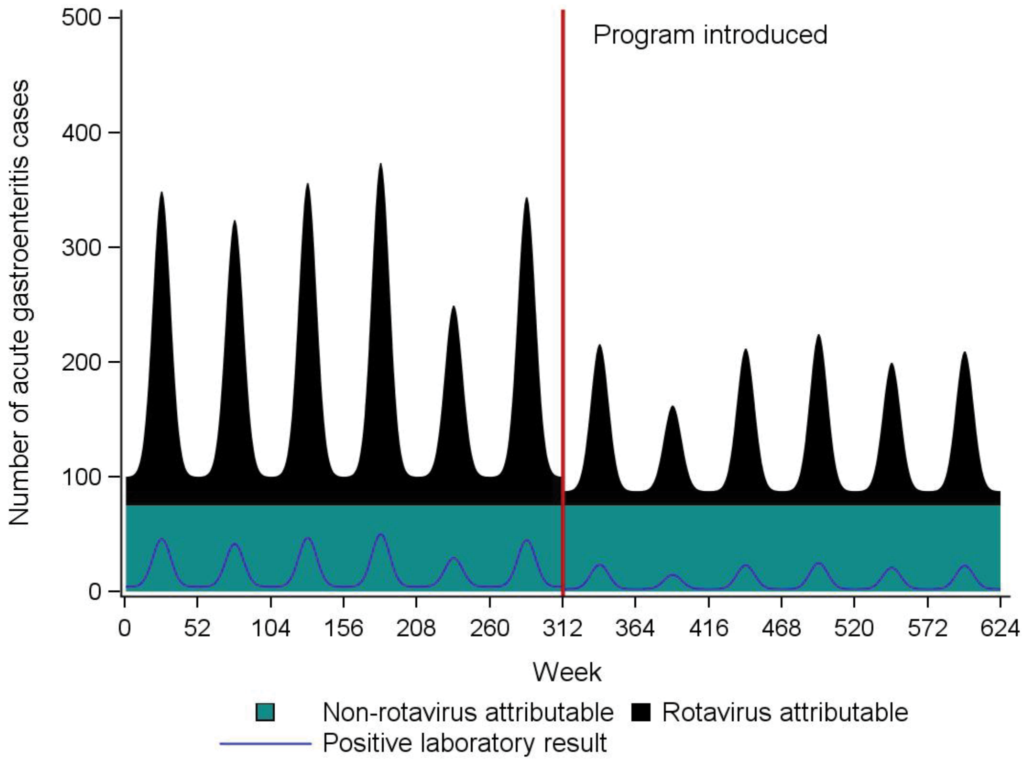
Figure 6.8: *Scenario 8: no relationship between weekly positive laboratory results and rotavirus cases (seasonal peaks coincide, seasonal widths and magnitudes vary).*



6.3.9 Scenario Nine

This scenario investigated how the methods compare when the magnitude of the rotavirus seasonal peak varied by year. Scenario 9 was the same as Scenario 1 (Sub-section 6.3.1) except that i) the magnitude of the seasonal rotavirus peak in the pre-program years one to six were 250, 225, 257.5, 275, 150 and 245, respectively, (these were 100%, 90%, 103%, 110%, 60% and 98% of 250, which is the magnitude of the seasonal peak in Scenario 1) ii) there was a 50% reduction in rotavirus cases after implementation of the vaccination program, iii) there were six rotavirus cases per positive laboratory result in a given week, and iv) the order of the 50% reduced magnitude of seasonal peaks in the program era were shuffled; for years seven to twelve they were 128.75, 75, 125, 137.5, 112.5 and 122.5, respectively (Figure 6.9).

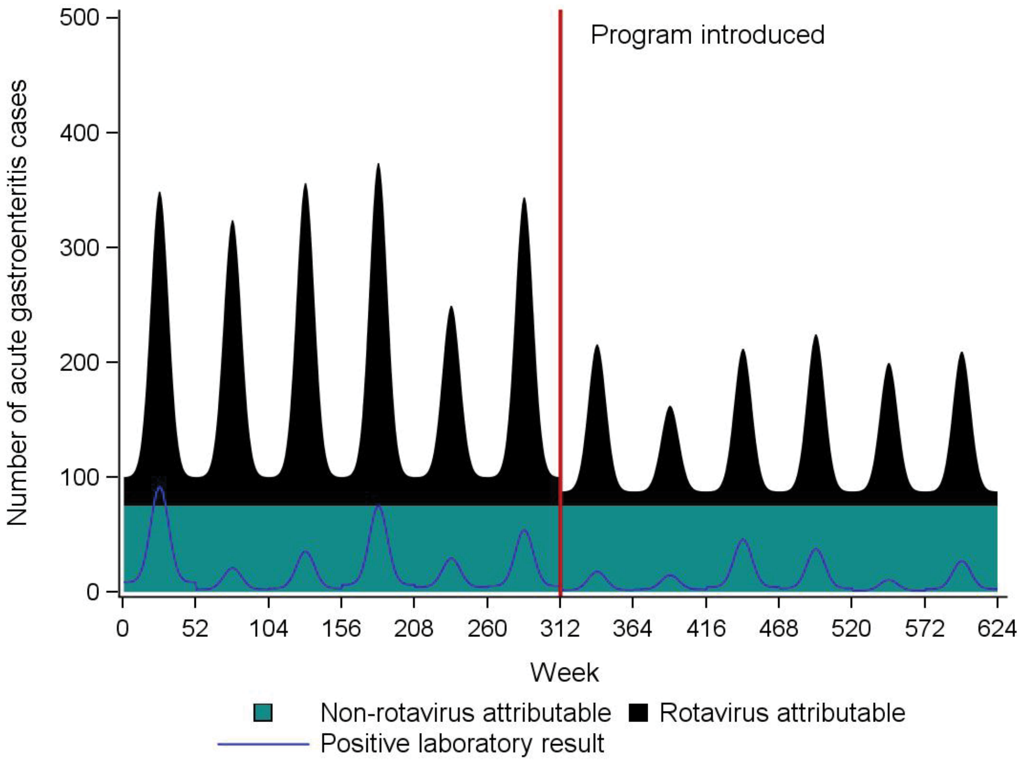
Figure 6.9: *Scenario 9: varying heights in peaks of rotavirus seasonality. Ratio of weekly rotavirus cases per positive laboratory result is constant over the study.*



6.3.10 Scenario Ten

This scenario considered the effect of varying the ratio of the number of rotavirus cases per positive rotavirus laboratory result in a given week by year. Scenario 10 was the same as Scenario 9 (Sub-section 6.3.9) with the addition that the ratio of the number of rotavirus cases per positive rotavirus laboratory result varied by year (Figure 6.10). The ratio in order of year was 3, 12, 8, 4, 6, 5, 8, 6, 3, 4, 12 and 5, respectively.

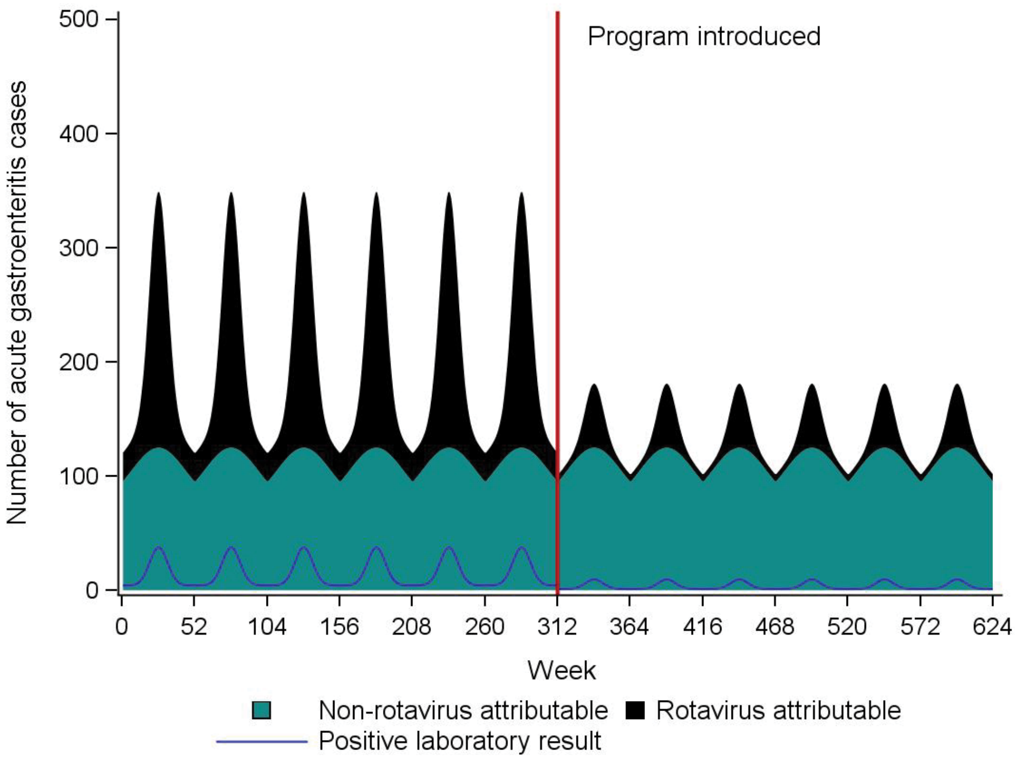
Figure 6.10: *Scenario 10: varying heights in peaks of rotavirus seasonality. Ratio of weekly rotavirus cases per positive laboratory result varies by year.*



6.3.11 Scenario Eleven

This scenario examined the effect of winter seasonality in the non-rotavirus cases on the estimates produced for each method. Scenario 11 was similar to Scenario 1 (Sub-section 6.3.1) apart from i) the magnitude of the seasonal rotavirus peak was reduced from 250 to 200 cases in the pre-program era, ii) the number of non-seasonal rotavirus cases was increased from 75 to 100 in the pre-program era, iii) seasonality in the non-rotavirus cases was introduced with the timing and width of the seasonal peak equal to that of the rotavirus seasonal peak and the magnitude equal to 25 non-rotavirus cases, and iv) the positive laboratory results were one sixth instead of one twelfth of the weekly rotavirus cases (Figure 6.11).

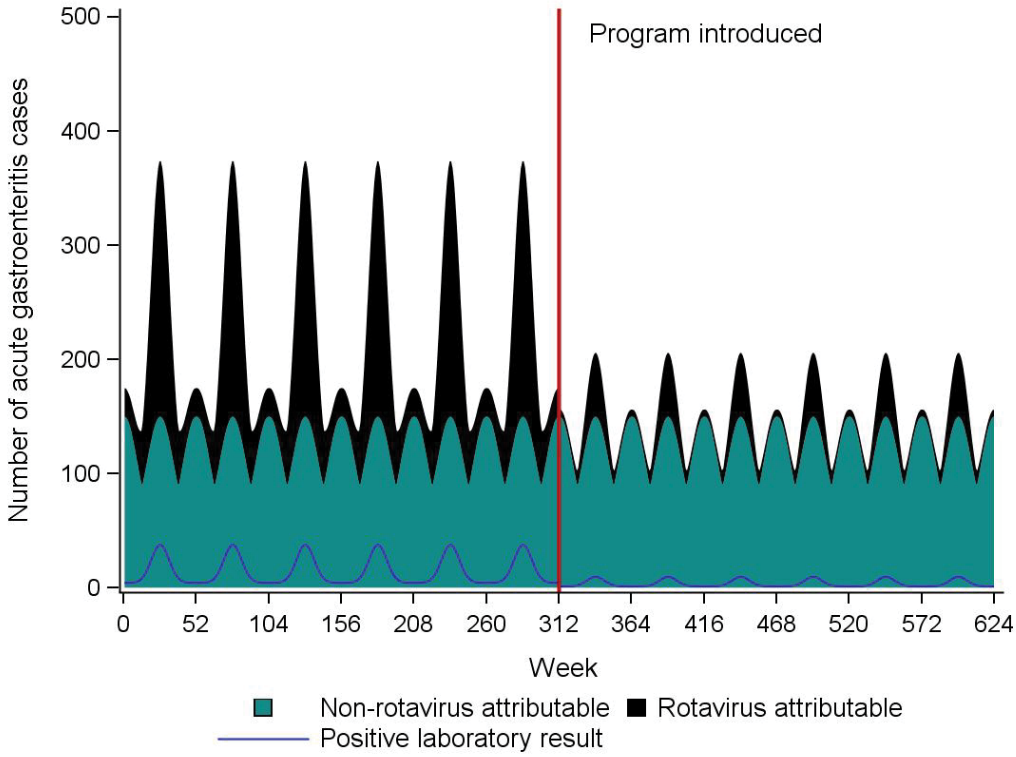
Figure 6.11: *Scenario 11: seasonality in non-rotavirus cases with the seasonal non-rotavirus peak coinciding with the seasonal rotavirus peak.*



6.3.12 Scenario Twelve

To further explore the effect of seasonality in the non-rotavirus cases summer seasonality was introduced. Scenario 12 was exactly the same as Scenario 11 (Sub-section 6.3.11) except the magnitude of the non-rotavirus peak was increased from 25 to 50 cases and the frequency was doubled (i.e. there was a winter peak and a summer peak, both of equal magnitude) (Figure 6.12).

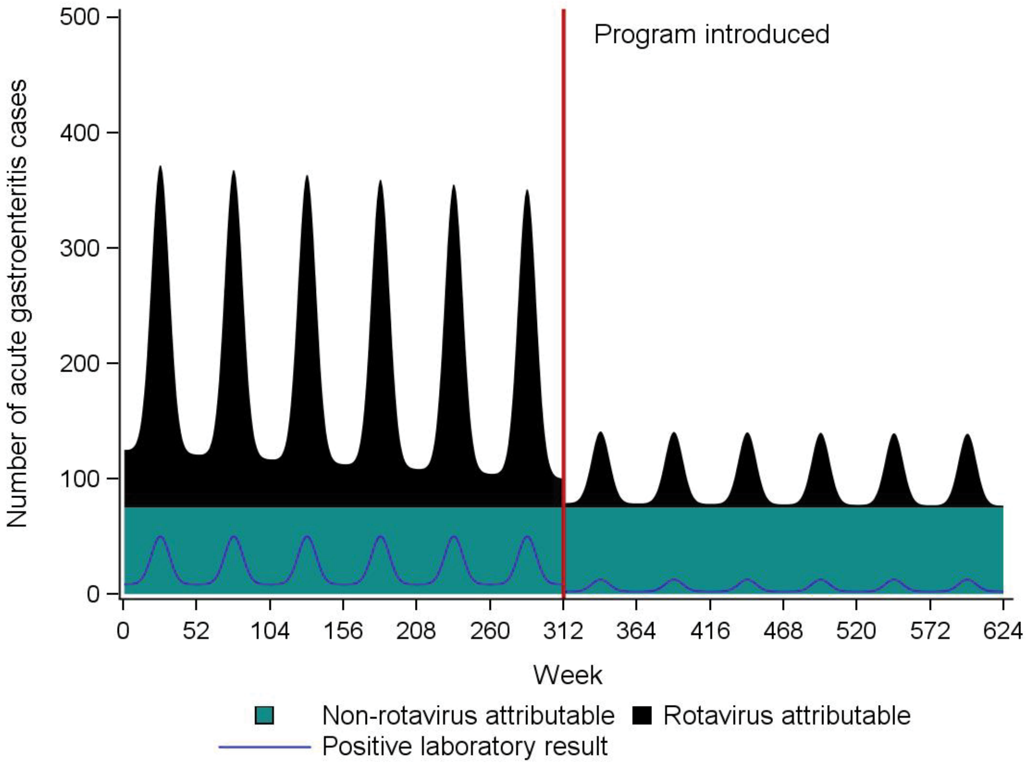
Figure 6.12: *Scenario 12: summer and winter seasonality in non-rotavirus cases.*



6.3.13 Scenario Thirteen

This scenario investigated the effect of a decreasing linear trend on the results produced by the various methods. This was the same as Scenario 1 (Sub-section 6.3.1) except that the rotavirus cases decreased at 0.08 cases per week and the ratio of weekly rotavirus counts per positive laboratory result was six (Figure 6.13).

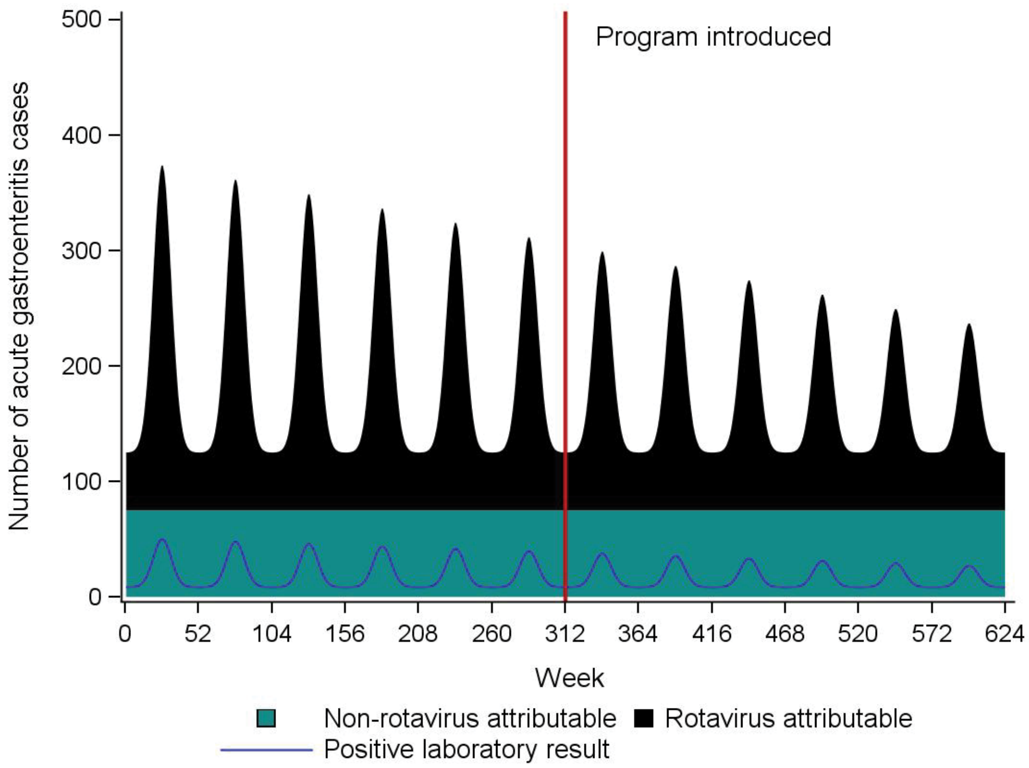
Figure 6.13: *Scenario 13: decreasing linear trend in rotavirus cases.*



6.3.14 Scenario Fourteen

This scenario considered the effect of a decreasing linear trend in the magnitude of the rotavirus seasonal peak. Scenario 14 was the same as Scenario 13, however the magnitude of the seasonal rotavirus peak decreased by 12.5 cases per year and there was no other decreasing trend (Figure 6.14).

Figure 6.14: *Scenario 14: decreasing linear trend in peaks of seasonal rotavirus cases and no effect of vaccination efforts.*



6.3.15 Scenario Fifteen

Scenario 15 was a combination of Scenarios 10 and 11 with the addition of varying magnitudes in the seasonal non-rotavirus peak to observe how the methods perform in a more complex setting (Figure 6.15). The number of non-seasonal rotavirus cases was 25 in the pre-program era. The number of non-seasonal non-rotavirus cases was 75 across the whole study period. There was a 75% decline in rotavirus cases and a 50% decline in laboratory reports after introduction of the vaccination program. The magnitudes of the seasonal rotavirus peak in order of year were 212.5, 225, 125, 275, 150, 245, 125, 43.75, 75, 87.5, 37.5, and 37.5, respectively. The magnitudes of the seasonal non-rotavirus peak in order of year were 25, 21.5, 50, 27.5, 20, 45, 25, 17.5, 25, 25, 37.5 and 20, respectively. The number of rotavirus cases per positive laboratory result in a given week changed by year. These ratios were 3, 12, 8, 4, 6, 5, 1.5, 6, 4, 2.3, 3 and 2.5 in order of year.

Figure 6.15: *Scenario 15: combination of Scenarios 10 and 11 with varying heights in peaks of non-rotavirus seasonality.*

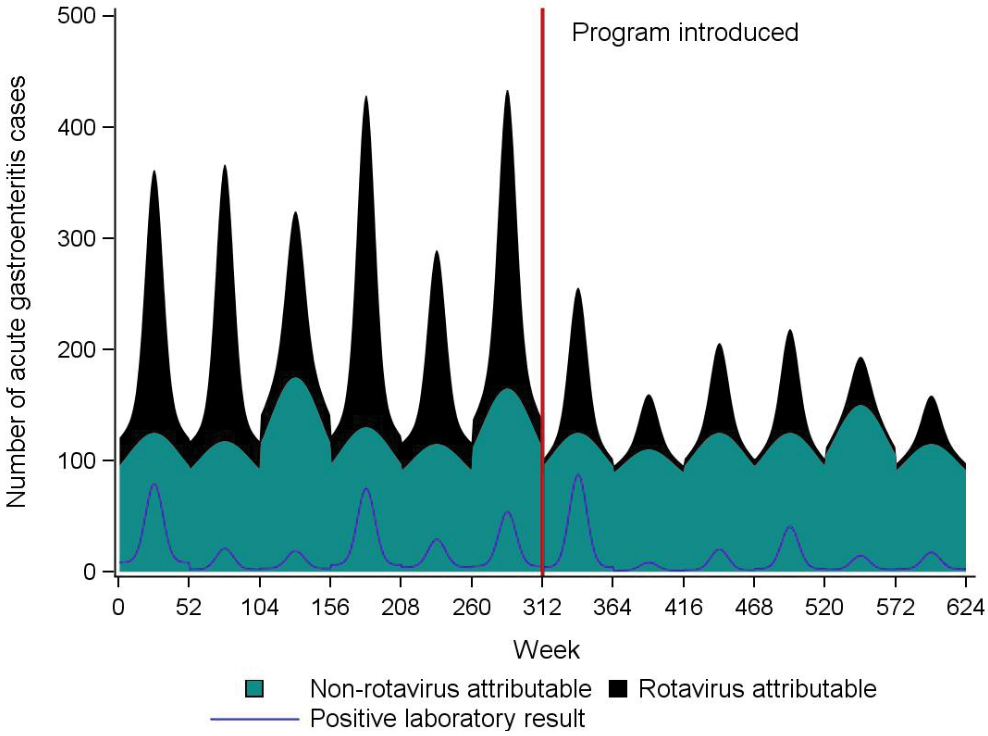
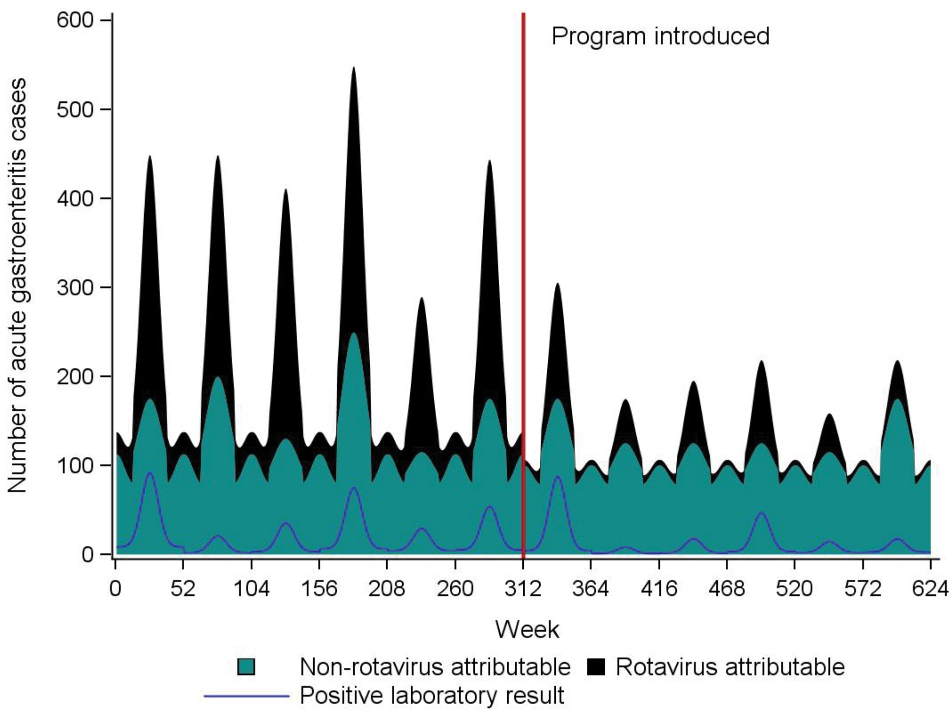


Figure 6.16: *Scenario 16: combination of Scenarios 10 and 12 with varying heights in peaks of non-rotavirus seasonality.*



6.3.16 Scenario Sixteen

Scenario 16 was a combination of Scenarios 10 and 12 with the addition of varying magnitudes in the seasonal non-rotavirus peak to generate another complex setting (Figure 6.16). The number of non-seasonal rotavirus cases was 25 in the pre-program era. The number of non-seasonal non-rotavirus cases was 75 across the whole study period. There was a 75% decline in rotavirus cases and a 50% decline in laboratory reports after introduction of the vaccination program. The magnitudes of the seasonal rotavirus peak in order of year were 250, 225, 257.5, 275, 150, 245, 125, 43.75, 64.5, 87.5, 37.5, and 37.5, respectively. The magnitudes of the seasonal non-rotavirus peak in the peak-season in order of year were 60, 75, 32.5, 100, 25, 60, 60, 30, 30, 30, 25 and 60, respectively. The magnitude of the seasonal non-rotavirus peak in the off-season was 32.5 cases in the pre-program era and 20 in the program introduction era (i.e. there was a decline in the magnitude of the summer seasonal peak that coincided with vaccination efforts). The number of rotavirus cases per positive laboratory result in a given week changed by year. These ratios were 3, 12, 8, 4, 6, 5, 1.5, 6, 4, 2, 3 and 2.5 in order of year of the study.

6.4 Simulations

For each scenario, 1000 random simulations were generated from the known truth of that scenario and stored. This provided a sufficient number of simulations for all results to be stable across the sixteen scenarios. For each simulation 1,872 values were drawn from 1,872 Poisson distributions, 3 per week over the entire study period (624 weeks). The three values per week determined the number of rotavirus cases, non-rotavirus cases and positive laboratory counts that week for that simulation. The means of the 3 Poisson distributions for a given week were the true values of the number of rotavirus cases, non-rotavirus cases and positive laboratory results for that week as defined by the scenario (Section 6.3). For example, in the first week of scenario 1 the number of rotavirus cases is drawn from a Poisson distribution with mean 25.1. This is because the number of non-seasonal rotavirus cases that occur each week in the pre-program era for Scenario 1, regardless of seasonality, is 25 cases per week. According to the seasonal rotavirus curve fitted (details in Sub-section 6.3.1) there is 0.1 seasonal rotavirus cases in week one. The number of non-rotavirus cases in week 1 is drawn from a Poisson distribution with mean 75 and the number of positive laboratory results is drawn from a Poisson distribution with mean 2.1 (i.e. 25.1/12) since 12 rotavirus cases are attributed to each positive laboratory result according to the definition of Scenario 1. Similarly, in weeks 26 and 27 of scenario 1 the number of rotavirus cases is drawn from a Poisson distribution with mean 273.6, the number of non-rotavirus cases is drawn from a Poisson distribution with mean 75, and the number of positive laboratory results is drawn from a Poisson distribution with mean 22.8.

The starting seeds to generate the random samples were randomly assigned before scenario development and are listed in Appendix F, Table F.1. These 16,000 samples are independent datasets by scenario. All the statistical methods that will be defined in Section 6.5 were applied to each of the 1000 random samples generated for a given scenario. Their estimates of the percentage of acute gastroenteritis cases attributed to rotavirus, stratified by pre- and post-program introduction were collated (details in sub-section 6.5.10).

6.5 Statistical analyses

First four methods from Chapter 5 were each applied to the same 1000 simulations for each of the sixteen scenarios. These methods were the off-season attributable fraction method, linear regression model, generalised additive model with six knots per year, and generalised additive model with two knots per year. The objective in each case was to calculate the percentage of acute gastroenteritis cases that was due to rotavirus separately for the pre-program era and for the program era. Summary statistics were calculated to compare these methods by scenario (details in sub-section 6.5.10). In addition, interaction terms for vaccination program introduction and season were included in the statistical models (linear regression and generalised additive models) and these were also applied to the same 1000 simulations for each of the sixteen scenarios. Full details are given in the following sub-sections, 6.5.1 to 6.5.9.

6.5.1 Off-season attributable fraction method

For each scenario, the percentage of acute gastroenteritis cases identified as rotavirus in the pre-program era for simulation i from 1000 simulations, $R_{\text{pre},i}$, was calculated using the formula:

$$R_{\text{pre},i} = \frac{P_{\text{pre},i} + O_{\text{pre},i}\left(2\rho_{\text{pre}} - 1\right)}{P_{\text{pre},i} + O_{\text{pre},i}}$$

where $P_{\text{pre},i}$ is the total number of acute gastroenteritis cases during peak season in the pre-program era and $O_{\text{pre},i}$ is the total number of acute gastroenteritis cases during off-peak season in the pre-program era for simulation i. The percentage of rotavirus cases in the off-peak season in the pre-program era, ρ_{pre}, was calculated directly from the known truth of that scenario (i.e. it did not vary by simulation, only by scenario). In practice, this is determined by medical audits.

Similarly, the percentage of acute gastroenteritis cases identified as rotavirus in the vaccination program era for simulation i was:

$$R_{\text{post},i} = \frac{P_{\text{post},i} + O_{\text{post},i}(2\rho_{\text{post}} - 1)}{P_{\text{post},i} + O_{\text{post},i}}$$

where $P_{\text{post},i}$ is the total number of acute gastroenteritis cases during peak season in the program era and $O_{\text{post},i}$ is the total number of acute gastroenteritis cases during off-peak season in the program era for that simulation. The percentage of rotavirus cases in the off-peak season in the program era, ρ_{post}, was calculated directly from the known truth of that scenario.

6.5.2 Linear regression model with no interaction term

For each simulation, a simple linear regression model was fitted with the weekly number of positive rotavirus laboratory results as the only explanatory variable included in the model. This is the same linear regression model described in Chapter 5, Section 5.3.3. The results from the linear regression model for simulation i were used to determine the percentage of acute gastroenteritis cases identified as rotavirus in the pre-program era, $R_{\text{pre},i}$, and program era, $R_{\text{post},i}$, with the following formulae:

$$R_{\text{pre},i} = \frac{\sum_{j=1}^{312} \beta_{1,i} L_{it}}{\sum_{j=1}^{312} \left[\beta_{1,i} L_{it} + \alpha_i\right]} \quad ; \quad R_{\text{post},i} = \frac{\sum_{j=313}^{624} \beta_{1,i} L_{it}}{\sum_{j=313}^{624} \left[\beta_{1,i} L_{it} + \alpha_i\right]}$$

where $\beta_{1,i}$ is the regression coefficient used to estimate the ratio of the number of weekly rotavirus cases per positive laboratory result, L_{it} is the number of positive rotavirus laboratory results in week t, and α_i is the constant (i.e. intercept) representing the number of non-rotavirus attributable acute gastroenteritis admissions each week for simulation i. It follows that $\beta_{1,i} L_{it}$ is the estimated number of rotavirus attributed acute gastroenteritis cases in week t of simulation i and $\beta_{1,i} L_{ij} + \alpha_i$ is the model predicted number of acute gastroenteritis cases in week t of simulation i. The weeks 1 to 312 are the pre-program era and the weeks 313 to 624 are the program era. Hence, the numerators of the formulae, from left to right, are the sum of all rotavirus cases estimated by the model in the pre-program era and program era, respectively. The denominators, from left to right, are the sum of the model predicted numbers of acute gastroenteritis cases in the pre-program era and program era, respectively.

6.5.3 Generalised additive models with no interaction term

For each scenario and simulation, two generalised additive models were fitted with the weekly number of positive rotavirus laboratory results and both a linear term and a non-parametric cubic spline for the number of weeks since the start of the study. One generalised additive model allowed for six knots per year to fit the spline and the other model only two knots per year. These are the same generalised additive models described in Chapter 5, Section 5.3.4, but the model with 3 knots per year was omitted to reduce computation time (since its model estimates will lie between the model with six knots per year and the model with two knots per year as seen in Chapter 5). The results from each generalised additive model for simulation i were used to determine the percentage of acute gastroenteritis cases identified as rotavirus in the pre-program era for simulation i from 1000 simulations, $R_{\text{pre},i}$, and the percentage of acute gastroenteritis cases identified as rotavirus in the program era for simulation i from 1000 simulations, $R_{\text{post},i}$, with the following formulae:

$$R_{\text{pre},i} = \frac{\sum_{j=1}^{312} \beta_{1,i} L_{it}}{\sum_{j=1}^{312} [\beta_{1,i} L_{it} + \gamma_i * t + f_i(\,t\,) + \alpha_i]};$$

$$R_{\text{post},i} = \frac{\sum_{j=313}^{624} \beta_{i} L_{it}}{\sum_{j=313}^{624} [\beta_{1,i} L_{it} + \gamma_i * t + f_i(\,t\,) + \alpha_i]}$$

where $\beta_{1,i}$ is the regression coefficient used to estimate the ratio of weekly rotavirus cases per positive laboratory result, L_{it} is the number of positive rotavirus laboratory results in week t, γ_i is the regression coefficient for the long-term linear trend over time, $f_i(\,t\,)$ is a non-parametric cubic spline to account for remaining autocorrelation present, and α_i is the constant (i.e. intercept) representing the residual number of non-rotavirus attributable acute gastroenteritis cases for simulation i. $\beta_{1,i} L_{it}$ is the estimated number of rotavirus attributed acute gastroenteritis cases in week t of simulation i and $\beta_{1,i} L_{it} + \gamma_i * t + f_i(t\,) + \alpha_i$ is the model predicted number of acute gastroenteritis cases in week t of simulation i. The weeks 1 to 312 are the pre-program era and the weeks 313 to 624 are the program era. Hence, the numerators of the formulae, are the sum of all rotavirus cases estimated by the model in the pre-program era and program era, respectively. The denominators are the sum of the model predicted numbers of acute gastroenteritis cases in the pre-program era and program era, respectively.

6.5.4 Linear regression model with an interaction with program introduction

To explore the performance of the linear regression method when the ratio of weekly rotavirus cases per positive laboratory result in the same week is allowed to vary between the pre-program and program eras, the simple linear regression model from sub-section 6.5.2 was re-fitted with the addition of an interaction with introduction of the vaccination program in week 313. The explanatory variables in the model are now the weekly number of positive rotavirus laboratory results, a binary indicator variable called program to represent introduction of the vaccination program (it is one in the program era and zero in the pre-program era) and the multiplication of these two variables. These covariates can be re-worked to obtain the desired model coefficients directly. The predicted number of acute gastroenteritis cases in week t of simulation i, $\hat{Y}_{it}$, is modelled using the formula:

$$\hat{Y}_{it} = \beta_{1,i}L_{\text{pre},it} + \beta_{2,i}L_{\text{post},it} + \delta_i * \text{program} + \alpha_i$$

where $L_{\text{pre},it}$ is the number of positive laboratory results in week t in the pre-program era and zero in the program era, $L_{\text{post},it}$ is zero in the pre-program era and equal to the number of positive laboratory results in week t in the program era for simulation i. The variable program is the binary indicator representing introduction of the vaccination program. It is zero in the pre-program era and one in the program era. For the coefficients, $\beta_{1,i}$ is the estimate of the ratio of weekly rotavirus cases per positive laboratory result in the pre-program era, $\beta_{2,i}$ is the estimate of the ratio of weekly rotavirus cases per positive laboratory result in the program era, α_i is the constant representing the number of non-rotavirus attributable acute gastroenteritis cases each week in the pre-program era, and $\alpha_i + \delta_i$ is the number of non-rotavirus attributable acute gastroenteritis cases each week in the program era for simulation i. It follows that $\beta_{1,i}L_{\text{pre},it}$ is the estimated number of rotavirus attributed acute gastroenteritis cases in week t in the pre-program era and $\beta_{2,i}L_{\text{post},it}$ is the estimated number of rotavirus attributed acute gastroenteritis cases in week t in the program era for simulation i.

The results from the linear regression model with interaction with program introduction for simulation i are used to determine the percentage of acute gastroenteritis cases identified as rotavirus in the pre-program era, $R_{\text{pre},i}$, and program era, $R_{\text{post},i}$, with the following formulae:

$$R_{\text{pre},i} = \frac{\sum_{j=1}^{312} \beta_{1,i}L_{\text{pre},it}}{\sum_{j=1}^{312}\left[\beta_{1,i}L_{\text{pre},it} + \alpha_i\right]} \quad ; \quad R_{\text{post},i} = \frac{\sum_{j=313}^{624} \beta_{2,i}L_{\text{post},it}}{\sum_{j=313}^{624}\left[\beta_{2,i}L_{\text{post},it} + \delta_i + \alpha_i\right]}.$$

Again, the numerators of the formulae, from left to right, are the sum of all rotavirus cases estimated by the model in the pre-program era and program era, respectively. The denominators, from left to right, are the sum of the model predicted numbers of acute gastroenteritis cases in the pre-program era and program era, respectively.

6.5.5 Generalised additive models with an interaction with program introduction

To explore the performance of the generalised additive models when the ratio of weekly rotavirus cases per positive laboratory result in the same week is allowed to vary between the pre-program and program eras, the basic six knots per year and two knots per year generalised additive models presented in sub-section 6.5.3 were re-fitted with the addition of an interaction with introduction of the vaccination program in week 313. The explanatory variables in the model are now the weekly number of positive rotavirus laboratory results, a binary indicator variable called program to represent introduction of the vaccination program (it is one in the program era and zero in the pre-program era), multiplication of these two variables, and both a linear term and a non-parametric cubic spline for the number of weeks since the start of the study. The first three covariates are multiplied out in the same way as for the linear regression model with an interaction with program introduction. The predicted number of acute gastroenteritis cases in week t of simulation i, $\hat{Y}_{ij}$, is modelled using the formula:

$$\hat{Y}_{it} = \beta_{1,i} L_{\text{pre},it} + \beta_{2,i} L_{\text{post},it} + \gamma_i * t + f_i(t\,) + \delta_i * \text{program} + \alpha_i$$

where $L_{\text{pre},it}$ is the number of positive laboratory results in week t in the pre-program era and zero in the program era, $L_{\text{post},it}$ is zero in the pre-program era and equal to the number of positive laboratory results in week t in the program era for simulation i. The variable program is the binary indicator representing introduction of the vaccination program. It is zero in the pre-program era and one in the program era. $\beta_{1,i}$ is the estimate of the ratio of weekly rotavirus cases per positive laboratory result in the pre-program era, $\beta_{2,i}$ is the estimate of the ratio of weekly rotavirus cases per positive laboratory result in the program era, γ_i is the regression coefficient for the long-term linear trend over time, $f_i(\ t\ )$ is a non-parametric cubic spline to account for remaining autocorrelation present, α_i is the residual number of non-rotavirus attributable acute gastroenteritis cases each week in the pre-program era, and $\alpha_i + \delta_i$ is the residual number of non-rotavirus attributable acute gastroenteritis cases each week in the program era for simulation i. It follows that $\beta_{1,i}L_{\text{pre},it}$ is the estimated number of rotavirus attributed acute gastroenteritis cases in week t in the pre-program era and $\beta_{2,i}L_{\text{post},it}$ is the estimated number of rotavirus attributed acute gastroenteritis cases in week t in the program era for simulation i.

The results from each generalised additive model for simulation i are used to determine the percentage of acute gastroenteritis cases identified as rotavirus in the pre-program era, $R_{\text{pre},i}$, and program, $R_{\text{post},i}$, with the following formulae:

$$R_{\text{pre},i} = \frac{\sum_{j=1}^{312} \beta_{1,i}L_{\text{pre},it}}{\sum_{j=1}^{312}\left[\beta_{1,i}L_{\text{pre},it} + \gamma_i * t + f_i(\ t) + \alpha_i\right]}\ ;$$

$$R_{\text{post},i} = \frac{\sum_{j=313}^{624} \beta_{2,i}L_{\text{post},it}}{\sum_{j=313}^{624}\left[\beta_{2,i}L_{\text{post},it} + \gamma_i * t + f_i(\ t\ ) + \delta_i + \alpha_i\right]}\ .$$

Again, the numerators of the formulae are the sum of all rotavirus cases estimated by the model in the pre-program era and program era, respectively. The denominators are the sum of the model predicted numbers of acute gastroenteritis cases in the pre-program era and program era, respectively.

6.5.6 Linear regression model with an interaction with season

To explore the performance of the linear regression method when the ratio of weekly rotavirus cases per positive laboratory result in the same week is allowed to vary between peak-season and off-season, the simple linear regression model presented in 6.5.2 was re-fitted with an interaction with season (peak-season and off-season). This is equivalent to the formulas presented in Section 6.5.4 where the simple linear regression includes an interaction with introduction of the vaccination program except the binary indicator for program introduction is replaced with the binary indicator for season (called season and it is zero in the off-season and one in the peak-season). These results will only be presented for scenario 3.

The results from the linear regression model with interaction with season for simulation i are used to determine the percentage of acute gastroenteritis cases identified as rotavirus in the pre-program era, $R_{pre,i}$, and program era, $R_{post,i}$, with the following formulae:

$$R_{\text{pre},i} = \frac{\sum_{j=1}^{312}\left[\beta_{1,i}L_{\text{peak},it} + \beta_{2,i}L_{\text{off},it}\right]}{\sum_{j=1}^{312}\left[\beta_{1,i}L_{\text{peak},it} + \beta_{2,i}L_{\text{off},it} + \varepsilon_i * \text{season} + \alpha_i\right]}$$

$$R_{\text{post},i} = \frac{\sum_{j=313}^{624}\left[\beta_{1,i}L_{\text{peak},it} + \beta_{2,i}L_{\text{off},it}\right]}{\sum_{j=313}^{624}\left[\beta_{1,i}L_{peak,ij} + \beta_{2,i}L_{\text{off},it} + \varepsilon_i * \text{season} + \alpha_i\right]}.$$

where $L_{\text{peak},it}$ is the number of positive laboratory results in week t during peak-season and zero during the off-season, $L_{\text{off},it}$ is zero during peak-season and equal to the number of positive laboratory results in week t during the off-season for simulation i. The variable season is a binary indicator that is zero during the off-peak season and one during the peak-season. $\beta_{1,i}$ is the estimate of the ratio of weekly rotavirus cases per positive laboratory result during the peak-season, $\beta_{2,i}$ is the estimate of the ratio of weekly rotavirus cases per positive laboratory result during the off-season, α_i is the number of non-rotavirus attributable acute gastroenteritis cases each week during the off-season, and $\alpha_i + \varepsilon_i$ is the number of non-rotavirus attributable acute gastroenteritis cases each week during the peak-season. It follows that $\beta_{1,i}L_{\text{peak},it}$ is the estimated number of rotavirus attributed acute gastroenteritis cases in week t in peak-season and $\beta_{2,i}L_{\text{off},it}$ is the estimated number of rotavirus attributed acute gastroenteritis cases in week t in the off-season for simulation i.

Again, the numerators of the formulae are the sum of all rotavirus cases estimated by the model in the pre-program era and program era, respectively. The denominators are the sum of the model predicted numbers of acute gastroenteritis cases in the pre-program era and program era, respectively.

6.5.7 Generalised additive models with an interaction with season

To explore the performance of the generalised additive models when the ratio of weekly rotavirus cases per positive laboratory result in the same week is allowed to vary between peak-season and off-season, the basic generalised additive models presented in sub-section 6.5.3 were re-fitted with an interaction with season (peak-season and off-season). This is equivalent to the formulas presented in Section 6.5.5 where the binary indicator for program introduction is replaced with the binary indicator for season (called season and it is zero in the off-season and one in the peak-season). These results will only be presented for scenario 3.

The results from each generalised additive models with interaction with season for simulation i are used to determine the percentage of acute gastroenteritis cases identified as rotavirus in the pre-program era, $R_{pre,i}$, and program era, $R_{post,i}$, with the following formulae:

$$R_{\text{pre},i} = \frac{\sum_{j=1}^{312}\left[\beta_{1,i}L_{\text{peak},it} + \beta_{2,i}L_{\text{off},it}\right]}{\sum_{j=1}^{312}\left[\beta_{1,i}L_{\text{peak},it} + \beta_{2,i}L_{\text{off},it} + \gamma_i * t + f_i(\,t) + \varepsilon_i * \text{season} + \alpha_i\right]}\ ;$$

$$R_{\text{post},i} = \frac{\sum_{j=313}^{624}\left[\beta_{1,i}L_{\text{peak},ij} + \beta_{2,i}L_{\text{off},ij}\right]}{\sum_{j=313}^{624}\left[\beta_{1,i}L_{\text{peak},ij} + \beta_{2,i}L_{\text{off},ij} + \gamma_i * t + f_i(\,t\,) + \varepsilon_i * \text{season} + \alpha_i\right]}\,.$$

where $L_{\text{peak},it}$ is the number of positive laboratory results in week t during peak-season and zero during the off-season, $L_{\text{off},it}$ is zero during peak-season and equal to the number of positive laboratory results in week t during the off-season for simulation i. The variable season is a binary indicator that is zero during the off-peak season and one during the peak-season. $\beta_{1,i}$ is the estimate of the ratio of weekly rotavirus cases per positive laboratory result during the peak-season, $\beta_{2,i}$ is the estimate of the ratio of weekly rotavirus cases per positive laboratory result during the off-season, γ_i is the regression coefficient for the long-term linear trend over time, $f_i(\,t\,)$ is a non-parametric cubic spline to account for remaining autocorrelation present, α_i is the residual number of non-rotavirus attributable acute gastroenteritis cases each week during the off-season, and $\alpha_i + \varepsilon_i$ is the residual number of non-rotavirus attributable acute gastroenteritis cases each week during the peak-season. It follows that $\beta_{1,i}L_{\text{peak},it}$ is the estimated number of rotavirus attributed acute gastroenteritis cases in week t in peak-season and $\beta_{2,i}L_{\text{off},it}$ is the estimated number of rotavirus attributed acute gastroenteritis cases in week t in the off-season for simulation i.

Again, the numerators of the formulae are the sum of all rotavirus cases estimated by the model in the pre-program era and program era, respectively. The denominators are the sum of the model predicted numbers of acute gastroenteritis cases in the pre-program era and program era, respectively.

6.5.8 Linear regression model with an interaction with season and program introduction

To explore the performance of the linear regression method when the ratio of weekly rotavirus cases per positive laboratory result in the same week is allowed to vary between peak-season and off-season and between pre-program and program era, the simple linear regression model presented in 6.5.2 was re-fitted with an interaction with season (peak-season and off-season) and program (pre- and program introduction). These results will only be presented for scenario 3.

The formulae present a similar theme to the previous sub-sections. In brief, the results from the linear regression model with interaction with season and program introduction for simulation i are used to determine the percentage of acute gastroenteritis cases identified as rotavirus in the pre-program era, $R_{\text{pre},i}$, and program era, $R_{\text{post},i}$, with the following formulae:

$$R_{\text{pre},i} = \frac{\sum_{j=1}^{312}\left[\beta_{1,i}L_{\text{pre,peak},it} + \beta_{2,i}L_{\text{pre,off},it}\right]}{\sum_{j=1}^{312}\left[\beta_{1,i}L_{\text{pre,peak},it} + \beta_{2,i}L_{\text{pre,off},it} + \varepsilon_i * \text{season} + \alpha_i\right]}$$

$$R_{\text{post},i} = \frac{\sum_{j=313}^{624}\left[\beta_{3,i}L_{\text{post,peak},it} + \beta_{4,i}L_{\text{post,off},it}\right]}{\sum_{j=313}^{624}\left[\beta_{3,i}L_{\text{post,peak},it} + \beta_{4,i}L_{\text{post,off},it} + \delta_i + \varepsilon_i * \text{season} + \alpha_i\right]}.$$

where $\beta_{1,i}L_{\text{pre,peak},it}$ is the estimated number of rotavirus cases in week t during peak-season in the pre-program era and zero otherwise, $\beta_{2,i}L_{\text{pre,off},it}$ is the estimated number of rotavirus cases in week t during the off-season in the pre-program era and zero otherwise, $\beta_{3,i}L_{\text{post,peak},it}$ is the estimated number of rotavirus cases in week t during peak-season in the program era and zero otherwise, $\beta_{4,i}L_{\text{post,off},it}$ is the estimated number of rotavirus cases in week t during the off-season in the program era and zero otherwise for simulation i.

Again, the numerators of the formulae are the sum of all rotavirus cases estimated by the model in the pre-program era and program era, respectively. The denominators are the sum of the model predicted numbers of acute gastroenteritis cases in the pre-program era and program era, respectively.

6.5.9 Generalised additive models with an interaction with season and program introduction

To explore the performance of the generalised additive methods when the ratio of weekly rotavirus cases per positive laboratory result in the same week is allowed to vary between peak-season and off-season and between pre-program and program era, the basic generalised additive models presented in 6.5.3 was re-fitted with an interaction with season (peak-season and off-season) and program (pre- and program introduction). These results will only be presented for scenario 3.

The formulae present a similar theme to the previous sub-sections. In brief, the results from the generalised additive models with interaction with season and program introduction for simulation i are used to determine the percentage of acute gastroenteritis cases identified as rotavirus in the pre-program era, $R_{pre,i}$, and program era, $R_{post,i}$, with the following formulae:

$$R_{\text{pre},i} = \frac{\sum_{j=1}^{312}\left[\beta_{1,i}L_{\text{pre,peak},it} + \beta_{2,i}L_{\text{pre,off},it}\right]}{\sum_{j=1}^{312}\left[\beta_{1,i}L_{\text{pre,peak},it} + \beta_{2,i}L_{pre,off,ij} + \gamma_i * t + f_i(\, t\,) + \varepsilon_i * \text{season} + \alpha_i\right]};$$

$$R_{\text{post},i} = \frac{\sum_{j=313}^{624}\left[\beta_{3,i}L_{\text{post,peak},it} + \beta_{4,i}L_{\text{post,off},it}\right]}{\sum_{j=313}^{624}\left[\beta_{3,i}L_{\text{post,peak},it} + \beta_{4,i}L_{\text{post,off},it} + \gamma_i * t + f_i(\, t) + \delta_i + \varepsilon_i * \text{season} + \alpha_i\right]}.$$

where $\beta_{1,i}L_{\text{pre,peak},it}$ is the estimated number of rotavirus cases in week t during peak-season in the pre-program era and zero otherwise, $\beta_{2,i}L_{\text{pre,off},it}$ is the estimated number of rotavirus cases in week t during the off-season in the pre-program era and zero otherwise, $\beta_{3,i}L_{\text{post,peak},it}$ is the estimated number of rotavirus cases in week t during peak-season in the program era and zero otherwise, $\beta_{4,i}L_{\text{post,off},it}$ is the estimated number of rotavirus cases in week t during the off-season in the program era and zero otherwise for simulation i.

Again, the numerators of the formulae are the sum of all rotavirus cases estimated by the model in the pre-program era and program era, respectively. The denominators are the sum of the model predicted numbers of acute gastroenteritis cases in the pre-program era and program era, respectively.

6.5.10 Summary statistics for each scenario

For each method, the median percentage of acute gastroenteritis cases estimated to be due to rotavirus in the pre-program era and in the program era from the 1000 simulations is reported for each scenario. The lower and upper 2.5$^{\text{th}}$ percentiles are given as a measure of variability in these estimates. In addition, the model coefficients from all regression models (linear regression and general additive models, with and without interaction terms) are summarised by reporting the median, 2.5th percentile and 97.5th percentile of the estimated model coefficients from the 1000 simulations for each scenario.

6.5.11 Application of methods to the acute gastroenteritis hospitalisations and ED presentations of children under five years of age in NSW (July 2001 to June 2013)

Finally, each of the methods described in Sections 6.5.1 to 6.5.5 were applied to the hospital and ED data used previously in this thesis to estimate the percentage of acute gastroenteritis hospitalisations and ED presentations due to rotavirus in children under five years of age in NSW (July 2001 to June 2013) in the pre-program era and in the program era. The coefficients from the statistical models are also reported.

6.6 Results

Sixteen thousand independent simulations were generated, 1000 for each scenario, by drawing from Poisson distributions as described in Section 6.4. These simulations gave an integer count of the number of rotavirus cases, non-rotavirus cases and positive laboratory results for each week of the study. For example, recall that the number of rotavirus cases in week 1 of scenario 1 was defined to have a *Poisson(25.1)* distribution (Section 6.4). The first simulation drew the value 20 as the number of rotavirus cases in week 1 and the second simulation drew the value 27 (results not shown). On average, across the 1000 simulations for Scenario 1, the number of rotavirus cases in week one will be 25.1. In addition, the first simulation assigned 4 as the number of positive laboratory counts in week 1 (from a *Poisson(2.1)* distribution) and the second simulation assigned the value 2. On average, across the 1000 simulations for Scenario 1, the number of positive laboratory results in week 1 of Scenario 1 will be 2.1. Therefore, the observed number of rotavirus cases per positive laboratory result in week 1 was 5 in simulation 1 and 13.5 in simulation 2. On average, this value will be 12 as was defined by scenario 1 (Sub-section 6.5.1).

For illustration, the fifteenth random simulation from each scenario is shown in Appendix G, Figures G.1 to G.16. Tables 6.2 to 6.17 show the summary results of scenarios one to sixteen, respectively. The true percentage of rotavirus cases in the pre-program era varied from 40.7 to 70.2 in the pre-program era and from 16.0 to 55.6 in the program-era.

The summary results of the off-season attributable fractions method from the 1000 simulations of the sixteen scenarios included the true percentage of rotavirus cases within its 2.5th to 97.5th percentile interval for both the pre-program and program eras in all scenarios except for scenarios 11, 15 and 16, which was three of the four scenarios that included seasonality in the non-rotavirus cases (Tables 6.12, 6.16 and 6.17).

The generalised additive model with six knots per year under-reported the percentages of acute gastroenteritis cases attributed to rotavirus in every scenario except the last one (Scenario 16; Table 6.17), however this was only for the program era percentage when fitting the model without a program interaction term. The corresponding pre-program percentage was half the true value. The under-reporting for the other fifteen scenarios was often over 50%.

The coefficients that represent the number of inferred rotavirus cases for each positive laboratory result in a given week were consistently highest for the linear regression model as compared to the two generalised additive models, both

Table 6.2: *Scenario 1 results from 1000 simulations: Ratio of rotavirus cases per positive laboratory result is constant over the study period.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**57.5**	**25.3**	**12**	**12**	**12**
Off-season attributable fractions	57.5 (57.0, 58.1)	25.3 (24.4, 26.2)	-	-	-
No interaction					
Linear regression	51.9 (50.5, 53.3)	21.3 (19.5, 23.1)	10.5 (10.1, 11.0)	-	-
GAM, 6 knots per year	13.6 (10.6, 16.4)	5.99 (4.66, 7.24)	2.84 (2.25, 3.40)	-	-
GAM, 2 knots per year	44.3 (41.9, 46.6)	19.5 (17.8, 21.3)	9.24 (8.75, 9.78)	-	-
Interaction with program introduction					
Linear regression	49.6 (47.3, 51.7)	15.5 (13.9, 17.2)	-	10.3 (9.81, 10.9)	7.34 (6.54, 8.24)
GAM, 6 knots per year	15.4 (11.9, 18.9)	2.10 (0.31, 3.91)	-	3.20 (2.49, 3.93)	0.99 (0.15, 1.85)
GAM, 2 knots per year	45.9 (43.3, 48.5)	12.7 (10.9, 14.4)	-	9.58 (9.00, 10.2)	6.01 (5.22, 6.87)

GAM: Generalised additive model

Table 6.3: *Scenario 2 results from 1000 simulations: ratio of weekly rotavirus cases per positive laboratory result varies by program introduction (pre/post).*

Method	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**57.5**	**25.3**	**-**	**12.0**	**4.00**
Off-season attributable fractions	57.5 (57.0, 58.1)	25.3 (24.3, 26.3)	-	-	-
No interaction					
Linear regression	48.2 (45.9, 50.3)	41.0 (39.4, 42.8)	8.39 (7.98, 8.81)	-	-
GAM, 6 knots per year	10.7 (8.57, 13.0)	14.2 (11.4, 17.0)	2.24 (1.81, 2.69)	-	-
GAM, 2 knots per year	34.2 (32.2, 36.2)	45.0 (42.7, 47.7)	7.13 (6.72, 7.52)	-	-
Interaction with program introduction					
Linear regression	49.7 (47.3, 52.0)	20.9 (19.3, 22.7)	-	10.4 (9.70, 11.0)	3.31 (3.04, 3.61)
GAM, 6 knots per year	15.5 (12.0, 18.9)	5.40 (2.65, 8.34)	-	3.22 (2.53, 3.94)	0.85 (0.42, 1.32)
GAM, 2 knots per year	46.0 (43.2, 48.6)	19.1 (17.4, 21.0)	-	9.60 (8.92, 10.3)	3.03 (2.73, 3.33)

GAM: Generalised additive model

Table 6.4: *Scenario 3 results from 1000 simulations: ratio of weekly rotavirus cases per positive laboratory result varies by season (peak and off-peak) and program introduction (pre/post).*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**57.5**	**25.3**		**peak-season: 12** **off-season: 6**	**peak-season: 5** **off-season: 3**
Off-season attributable fractions	57.5 (57.0, 58.1)	25.3 (24.3, 26.2)	-	-	-
No interaction					
Linear regression	64.1 (61.8, 66.4)	45.9 (43.1, 48.9)	10.4 (9.85, 10.9)	-	-
GAM, 6 knots per year	11.3 (8.78, 13.9)	9.42 (7.43, 11.5)	1.95 (1.51, 2.37)	-	-
GAM, 2 knots per year	47.3 (44.0, 50.2)	39.6 (36.7, 42.5)	8.11 (7.57, 8.61)	-	-
Interaction with program introduction					
Linear regression	64.5 (60.6, 68.4)	22.4 (20.3, 24.5)	-	11.1 (10.3, 11.8)	4.61 (4.15, 5.09)
GAM, 6 knots per year	14.0 (10.6, 17.6)	3.69 (1.35, 6.15)	-	2.40 (1.83, 3.02)	0.76 (0.28, 1.27)
GAM, 2 knots per year	55.6 (51.7, 59.6)	18.9 (16.8, 21.2)	-	9.55 (8.83, 10.3)	3.89 (3.44, 4.35)

GAM: Generalised additive model

Table 6.5: *Scenario 4 results from 1000 simulations: increased non-seasonal rotavirus cases.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**70.2**	**37.1**	**12**	**12**	**12**
Off-season attributable fractions	70.2 (69.9, 70.5)	37.1 (36.4, 37.9)	-	-	-
No interaction					
Linear regression	63.0 (61.4, 64.7)	29.7 (27.7, 32.2)	10.4 (9.97, 10.8)	-	-
GAM, 6 knots per year	11.0 (8.14, 14.0)	5.78 (4.28, 7.36)	1.87 (1.39, 2.38)	-	-
GAM, 2 knots per year	46.5 (43.8, 49.6)	24.5 (22.5, 26.5)	7.96 (7.48, 8.47)	-	-
Interaction with program introduction					
Linear regression	55.0 (52.0, 58.2)	17.7 (15.5, 20.0)	-	9.40 (8.88, 10.0)	5.73 (5.03, 6.46)
GAM, 6 knots per year	12.1 (8.82, 16.0)	1.86 (-.17, 3.97)	-	2.07 (1.52, 2.73)	0.60 (-.06, 1.29)
GAM, 2 knots per year	48.9 (45.5, 52.2)	13.7 (11.5, 16.1)	-	8.37 (7.80, 8.93)	4.44 (3.72, 5.15)

GAM: Generalised additive model

Table 6.6: *Scenario 5 results from 1000 simulations: no non-seasonal rotavirus cases.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**43.4**	**16.1**	**6**	**6**	**6**
Off-season attributable fractions	43.4 (42.6, 44.1)	16.1 (15.1, 17.1)	-	-	-
No interaction					
Linear regression	41.4 (40.4, 42.3)	15.0 (13.9, 16.1)	5.67 (5.45, 5.89)	-	-
GAM, 6 knots per year	19.5 (16.5, 22.8)	7.22 (6.10, 8.43)	2.69 (2.29, 3.14)	-	-
GAM, 2 knots per year	39.1 (37.6, 40.4)	14.5 (13.4, 15.6)	5.41 (5.16, 5.63)	-	-
Interaction with program introduction					
Linear regression	41.0 (39.7, 42.1)	13.0 (11.9, 14.1)	-	5.67 (5.41, 5.91)	4.86 (4.40, 5.34)
GAM, 6 knots per year	21.3 (17.8, 24.7)	3.13 (1.53, 5.09)	-	2.94 (2.47, 3.42)	1.17 (0.59, 1.88)
GAM, 2 knots per year	39.7 (38.1, 41.1)	11.7 (10.5, 12.9)	-	5.49 (5.22, 5.75)	4.38 (3.90, 4.86)

GAM: Generalised additive model

Table 6.7: *Scenario 6 results from 1000 simulations: Scenario five with positive laboratory results lagged by one week.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**43.4**	**16.1**	**6**	**6**	**6**
Off-season attributable fractions	43.4 (42.7, 44.1)	16.1 (15.1, 17.0)	-	-	-
No interaction					
Linear regression	41.1 (40.1, 42.0)	14.8 (13.8, 15.9)	5.63 (5.39, 5.86)	-	-
GAM, 6 knots per year	18.9 (15.9, 21.9)	7.02 (5.89, 8.22)	2.62 (2.23, 3.04)	-	-
GAM, 2 knots per year	38.6 (37.1, 39.9)	14.3 (13.2, 15.4)	5.33 (5.09, 5.59)	-	-
Interaction with program introduction					
Linear regression	40.5 (39.3, 41.8)	12.8 (11.7, 13.9)	-	5.61 (5.35, 5.87)	4.79 (4.35, 5.29)
GAM, 6 knots per year	20.7 (17.3, 24.1)	3.01 (1.20, 4.85)	-	2.86 (2.39, 3.34)	1.12 (0.44, 1.80)
GAM, 2 knots per year	39.2 (37.7, 40.7)	11.5 (10.3, 12.7)	-	5.42 (5.15, 5.69)	4.30 (3.83, 4.80)

GAM: Generalised additive model

Table 6.8: *Scenario 7 results from 1000 simulations: no relationship between weekly positive laboratory results and rotavirus cases (seasonal peaks and widths coincide, seasonal magnitudes vary).*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**57.5**	**25.3**			
Off-season attributable fractions	57.5 (57.0, 58.1)	25.3 (24.3, 26.3)	-	-	-
No interaction					
Linear regression	57.1 (55.9, 58.5)	31.6 (30.1, 33.3)	4.42 (4.27, 4.57)	-	-
GAM, 6 knots per year	24.6 (21.3, 28.2)	15.0 (12.9, 17.1)	1.98 (1.70, 2.24)	-	-
GAM, 2 knots per year	50.4 (48.7, 52.2)	30.8 (29.2, 32.6)	4.05 (3.89, 4.21)	-	-
Interaction with program introduction					
Linear regression	53.5 (51.8, 55.2)	24.8 (23.0, 26.9)	-	4.29 (4.13, 4.46)	3.27 (3.00, 3.58)
GAM, 6 knots per year	27.9 (24.0, 32.4)	5.40 (2.23, 8.48)	-	2.24 (1.92, 2.60)	0.71 (0.30, 1.13)
GAM, 2 knots per year	51.9 (49.9, 53.8)	22.0 (20.1, 24.4)	-	4.16 (3.98, 4.35)	2.91 (2.62, 3.24)

GAM: Generalised additive model

Table 6.9: *Scenario 8 results from 1000 simulations: no relationship between weekly positive laboratory results and rotavirus cases (seasonal peaks coincide, seasonal widths and magnitudes vary).*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**57.5**	**25.3**			
Off-season attributable fractions	57.5 (57.0, 58.0)	25.3 (24.3, 26.4)	-	-	-
No interaction					
Linear regression	42.2 (40.7, 43.7)	22.3 (20.9, 23.8)	4.99 (4.75, 5.23)	-	-
GAM, 6 knots per year	11.7 (10.1, 13.3)	8.11 (6.95, 9.32)	1.54 (1.33, 1.74)	-	-
GAM, 2 knots per year	29.0 (27.4, 30.5)	20.1 (18.7, 21.6)	3.81 (3.58, 4.04)	-	-
Interaction with program introduction					
Linear regression	34.9 (33.2, 36.5)	18.2 (16.4, 20.2)	-	4.59 (4.33, 4.85)	3.46 (3.10, 3.87)
GAM, 6 knots per year	12.4 (10.7, 14.2)	4.40 (2.58, 6.63)	-	1.63 (1.40, 1.87)	0.84 (0.49, 1.26)
GAM, 2 knots per year	29.7 (27.9, 31.3)	15.1 (13.3, 17.1)	-	3.90 (3.65, 4.16)	2.86 (2.51, 3.26)

GAM: Generalised additive model

Table 6.10: *Scenario 9 results from 1000 simulations: varying heights in peaks of rotavirus seasonality. Ratio of rotavirus cases per positive laboratory result is constant over the study period.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**56.3**	**39.2**	**6**	**6**	**6**
Off-season attributable fractions	56.3 (55.7, 56.9)	39.2 (38.4, 39.9)	-	-	-
No interaction					
Linear regression	52.4 (51.1, 53.8)	35.5 (33.7, 37.3)	5.52 (5.31, 5.74)	-	-
GAM, 6 knots per year	20.6 (17.3, 23.7)	14.4 (12.1, 16.5)	2.19 (1.84, 2.52)	-	-
GAM, 2 knots per year	48.6 (46.7, 50.4)	33.8 (31.9, 35.8)	5.18 (4.96, 5.40)	-	-
Interaction with program introduction					
Linear regression	52.1 (50.2, 54.0)	33.7 (31.9, 35.6)	-	5.55 (5.30, 5.81)	5.17 (4.82, 5.51)
GAM, 6 knots per year	23.0 (18.9, 26.8)	10.1 (7.32, 13.3)	-	2.45 (2.01, 2.87)	1.55 (1.12, 2.03)
GAM, 2 knots per year	49.7 (47.6, 51.9)	31.0 (29.0, 33.1)	-	5.30 (5.04, 5.58)	4.75 (4.40, 5.10)

GAM: Generalised additive model

Table 6.11: *Scenario 10 results from 1000 simulations: varying heights in peaks of rotavirus seasonality. Ratio of rotavirus cases per positive laboratory result varies by year.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**56.3**	**39.2**			
Off-season attributable fractions	56.3 (55.8, 56.8)	39.2 (38.4, 40.0)	-	-	-
No interaction					
Linear regression	40.1 (38.7, 41.5)	25.0 (23.7, 26.4)	3.46 (3.31, 3.61)	-	-
GAM, 6 knots per year	21.2 (18.3, 24.3)	14.7 (12.6, 16.9)	1.91 (1.64, 2.20)	-	-
GAM, 2 knots per year	43.4 (41.4, 45.3)	30.2 (28.5, 32.0)	3.93 (3.72, 4.10)	-	-
Interaction with program introduction					
Linear regression	37.1 (35.3, 39.0)	24.8 (23.1, 26.7)	-	3.36 (3.17, 3.54)	3.22 (2.98, 3.48)
GAM, 6 knots per year	23.2 (19.4, 27.2)	10.9 (7.80, 14.0)	-	2.10 (1.75, 2.46)	1.41 (1.03, 1.82)
GAM, 2 knots per year	44.4 (42.0, 46.6)	27.8 (25.9, 30.1)	-	4.01 (3.78, 4.23)	3.62 (3.34, 3.93)

GAM: Generalised additive model

Table 6.12: *Scenario 11 results from 1000 simulations: seasonality in non-rotavirus cases with the seasonal non-rotavirus peak coinciding with the seasonal rotavirus peak.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**43.3**	**16.0**	**6**	**6**	**6**
Off-season attributable fractions	47.5 (46.9, 48.1)	22.2 (21.3, 23.0)	-	-	-
No interaction					
Linear regression	43.6 (42.4, 44.8)	16.2 (15.0, 17.4)	6.05 (5.79, 6.32)	-	-
GAM, 6 knots per year	14.0 (11.2, 17.0)	5.17 (4.12, 6.29)	1.94 (1.56, 2.34)	-	-
GAM, 2 knots per year	40.5 (38.6, 42.3)	15.0 (13.8, 16.1)	5.61 (5.31, 5.91)	-	-
Interaction with program introduction					
Linear regression	43.8 (41.9, 45.6)	16.6 (15.1, 18.1)	-	6.06 (5.76, 6.37)	6.20 (5.59, 6.87)
GAM, 6 knots per year	15.5 (12.4, 19.1)	2.02 (0.05, 4.02)	-	2.16 (1.72, 2.63)	0.75 (0.02, 1.53)
GAM, 2 knots per year	40.9 (38.8, 42.8)	13.5 (11.9, 15.2)	-	5.66 (5.35, 5.98)	5.06 (4.43, 5.73)

GAM: Generalised additive model

Table 6.13: *Scenario 12 results from 1000 simulations: summer and winter seasonality in non-rotavirus cases.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**40.7**	**14.7**	**6**	**6**	**6**
Off-season attributable fractions	40.5 (39.9, 41.1)	14.3 (13.5, 15.2)	-	-	-
No interaction					
Linear regression	40.5 (39.4, 41.6)	14.6 (13.4, 15.7)	5.97 (5.72, 6.24)	-	-
GAM, 6 knots per year	19.8 (16.4, 23.1)	7.08 (5.87, 8.35)	2.91 (2.42, 3.39)	-	-
GAM, 2 knots per year	40.5 (38.6, 42.2)	14.6 (13.4, 15.7)	5.97 (5.67, 6.28)	-	-
Interaction with program introduction					
Linear regression	40.5 (38.9, 42.2)	14.4 (13.0, 15.8)	-	5.96 (5.67, 6.28)	5.89 (5.26, 6.60)
GAM, 6 knots per year	20.9 (17.1, 24.8)	4.92 (2.83, 7.20)	-	3.08 (2.50, 3.64)	2.01 (1.18, 2.95)
GAM, 2 knots per year	40.5 (38.7, 42.4)	14.5 (12.8, 16.3)	-	5.97 (5.65, 6.32)	5.92 (5.20, 6.70)

GAM: Generalised additive model

Table 6.14: *Scenario 13 results from 1000 simulations: decreasing linear trend in rotavirus cases.*

Method	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**60.3**	**22.6**			
Off-season attributable fractions	60.3 (59.9, 60.8)	22.6 (21.5, 23.7)	-	-	-
No interaction					
Linear regression	62.4 (61.0, 63.7)	29.3 (27.5, 31.2)	5.52 (5.32, 5.73)	-	-
GAM, 6 knots per year	22.3 (18.9, 25.9)	10.9 (9.12, 12.7)	2.00 (1.69, 2.31)	-	-
GAM, 2 knots per year	56.5 (54.1, 59.0)	27.6 (25.8, 29.4)	5.07 (4.82, 5.30)	-	-
Interaction with program introduction					
Linear regression	60.9 (58.5, 63.5)	23.4 (21.2, 25.4)	-	5.46 (5.21, 5.71)	4.30 (3.87, 4.73)
GAM, 6 knots per year	24.8 (20.7, 29.1)	4.13 (1.49, 6.87)	-	2.23 (1.86, 2.60)	0.76 (0.27, 1.26)
GAM, 2 knots per year	57.9 (55.3, 60.7)	20.1 (17.9, 22.4)	-	5.19 (4.92, 5.46)	3.71 (3.27, 4.16)

GAM: Generalised additive model

Table 6.15: *Scenario 14 results from 1000 simulations: decreasing linear trend in peaks of seasonal rotavirus cases and no effect of vaccination efforts.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**60.9**	**55.6**	**6**	**6**	**6**
Off-season attributable fractions	61.0 (60.5, 61.4)	55.6 (55.1, 56.1)	-	-	-
No interaction					
Linear regression	53.7 (51.9, 55.6)	48.3 (46.2, 50.4)	5.25 (5.02, 5.46)	-	-
GAM, 6 knots per year	17.1 (14.3, 20.1)	15.6 (12.9, 18.3)	1.68 (1.40, 1.96)	-	-
GAM, 2 knots per year	49.4 (47.3, 51.6)	45.1 (42.7, 47.4)	4.86 (4.65, 5.09)	-	-
Interaction with program introduction					
Linear regression	54.5 (52.1, 57.0)	45.9 (43.2, 48.7)	-	5.37 (5.09, 5.64)	4.95 (4.63, 5.26)
GAM, 6 knots per year	20.1 (16.4, 24.2)	11.6 (8.02, 15.5)	-	1.99 (1.61, 2.35)	1.25 (0.86, 1.67)
GAM, 2 knots per year	51.4 (48.7, 54.1)	41.5 (38.7, 44.4)	-	5.06 (4.76, 5.33)	4.48 (4.15, 4.81)

GAM: Generalised additive model

Table 6.16: *Scenario 15 results from 1000 simulations: combination of Scenarios 10 and 11 with varying heights in peaks of non-rotavirus seasonality.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**41.7**	**19.3**			
Off-season attributable fractions	46.7 (46.1, 47.3)	25.3 (24.4, 26.1)	-	-	-
No interaction					
Linear regression	29.4 (28.1, 30.7)	21.2 (20.3, 22.1)	3.11 (2.97, 3.26)	-	-
GAM, 6 knots per year	13.1 (11.0, 15.1)	12.7 (10.6, 14.6)	1.58 (1.32, 1.80)	-	-
GAM, 2 knots per year	28.4 (26.8, 29.7)	27.4 (26.1, 28.7)	3.41 (3.24, 3.58)	-	-
Interaction with program introduction					
Linear regression	30.3 (28.8, 31.8)	14.6 (13.6, 15.7)	-	3.65 (3.45, 3.85)	1.82 (1.68, 1.98)
GAM, 6 knots per year	16.7 (13.6, 19.9)	7.39 (5.13, 9.68)	-	2.02 (1.64, 2.39)	0.92 (0.63, 1.20)
GAM, 2 knots per year	35.8 (33.9, 37.7)	16.9 (15.4, 18.2)	-	4.32 (4.07, 4.56)	2.09 (1.91, 2.27)

GAM: Generalised additive model

Table 6.17: *Scenario 16 results from 1000 simulations: combination of Scenarios 10 and 12 with varying heights in peaks of non-rotavirus seasonality.*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Truth	**43.9**	**19.8**			
Off-season attributable fractions	55.8 (55.3, 56.3)	32.6 (31.8, 33.5)	-	-	-
No interaction					
Linear regression	40.6 (39.1, 42.1)	29.3 (28.2, 30.4)	4.11 (3.95, 4.28)	-	-
GAM, 6 knots per year	19.9 (17.3, 22.4)	19.8 (17.3, 22.5)	2.30 (2.01, 2.60)	-	-
GAM, 2 knots per year	38.9 (37.1, 40.7)	38.8 (37.2, 40.7)	4.51 (4.31, 4.72)	-	-
Interaction with program introduction					
Linear regression	40.8 (39.1, 42.7)	21.2 (20.1, 22.4)	-	4.74 (4.51, 4.98)	2.46 (2.31, 2.62)
GAM, 6 knots per year	24.5 (20.9, 28.4)	12.1 (9.16, 15.2)	-	2.85 (2.42, 3.32)	1.41 (1.06, 1.78)
GAM, 2 knots per year	47.7 (45.6, 50.1)	24.2 (22.5, 25.9)	-	5.54 (5.27, 5.84)	2.81 (2.59, 3.02)

GAM: Generalised additive model

when interaction terms were and were not included in the models. A higher model coefficient did not guarantee the highest estimated percentage of rotavirus attributed gastroenteritis cases (for example, Scenario 2; Table 6.3). The results of the generalised additive models with two knots per year were often similar to the results of the corresponding linear regression models.

In the scenarios where the ratio of rotavirus cases per positive laboratory result was related and constant over the study period the statistical models (i.e. linear regression and generalised additive models) under-reported the true percentage of rotavirus attributed acute gastroenteritis cases in both the pre-program and program eras (Scenarios 1, 4, 5, 6, 9; Tables 6.2, 6.5, 6.6, 6.7, 6.10). This under-reporting was most noticeable when there were a large number of non-seasonal rotavirus cases (Scenario 4; Table 6.5). In all of these cases the model coefficients under estimated the true ratio of positive laboratory results to rotavirus cases. The exception to this under-reporting of the percentages was in those scenarios where there was seasonality in non-rotavirus cases in the peak-season (Scenarios 11 and 12; Tables 6.12 and 6.13). The only other scenarios where the summary statistics of the linear regression model or generalised additive model with two knots per year included the true percentage in their 2.5th to 97.5th percentile interval were i) Scenarios 3, 5 and 7 (Tables 6.4, 6.6, 6.8), however this was only for one method and for one of either the pre-program or program era true percentages, and ii) Scenario 13 for models that included an interaction term for program introduction (Table 6.14).

In the scenarios where the ratio of rotavirus cases per positive laboratory result was related but varied by program introduction (pre-/post- introduction), season (peak-/off-season) or year (Scenarios 2, 3, 10, 15, 16; Tables 6.3, 6.4, 6.11, 6.16, 6.17) the linear regression model and generalised additive models under- and over-estimated the true percentages. A lag of one week in the laboratory results did not affect the results when no other features changed (Scenarios 5 and 6; Tables 6.6 and 6.7).

Sometimes the statistical models found a difference in the ratio of rotavirus cases per positive laboratory result in a given week (i.e. the model coefficients) between the pre-program and program era even when it did not exist (for example, Tables 6.2 and 6.5). Figures 6.17 and 6.18 show the observed numbers of positive laboratory counts versus the rotavirus cases in the same week by pre- and post-program era from the fifteenth simulation of Scenario 1 and Scenario 2, respectively. These represent scenarios where the ratio of rotavirus cases per positive laboratory result was constant (Scenario 1) or varied by program era (Scenario 2).

Table 6.18 presents the results of the models that contain an interaction with season (peak- and off-season) alone or an interaction with season and program era (pre-/program-era) for Scenario 3. The models were a poor fit and the coefficients in the off-peak season were low. The true numbers of positive laboratory counts versus the rotavirus cases in the same week by pre- and post-program era and season (peak-season and off-season) of Scenario 3 (ratio varies by program era and season) is shown in Figure 6.19. The corresponding observed numbers from the fifteenth simulation of Scenario 3 are presented in Figure 6.20.

Tables 6.19 and 6.20 display the observed percentages of rotavirus attributed acute gastroenteritis cases in the pre-program and program era for each method when applied to our hospital and ED data, respectively. The percentages varied; ranging from 22.1 to 56.7 in the pre-program era and from 9.57 to 38.1 in the program era for the hospital data (Table 6.19). The estimated percentages ranged from 13.4 to 43.5 in the pre-program era and from 5.0 to 31.8 in the program era for the ED data (Table 6.20). The off-season attributable fractions

Figure 6.17: *Observed numbers of positive laboratory counts versus the rotavirus cases in the same week by pre- and post-program era from simulation 15 of scenario 1.*

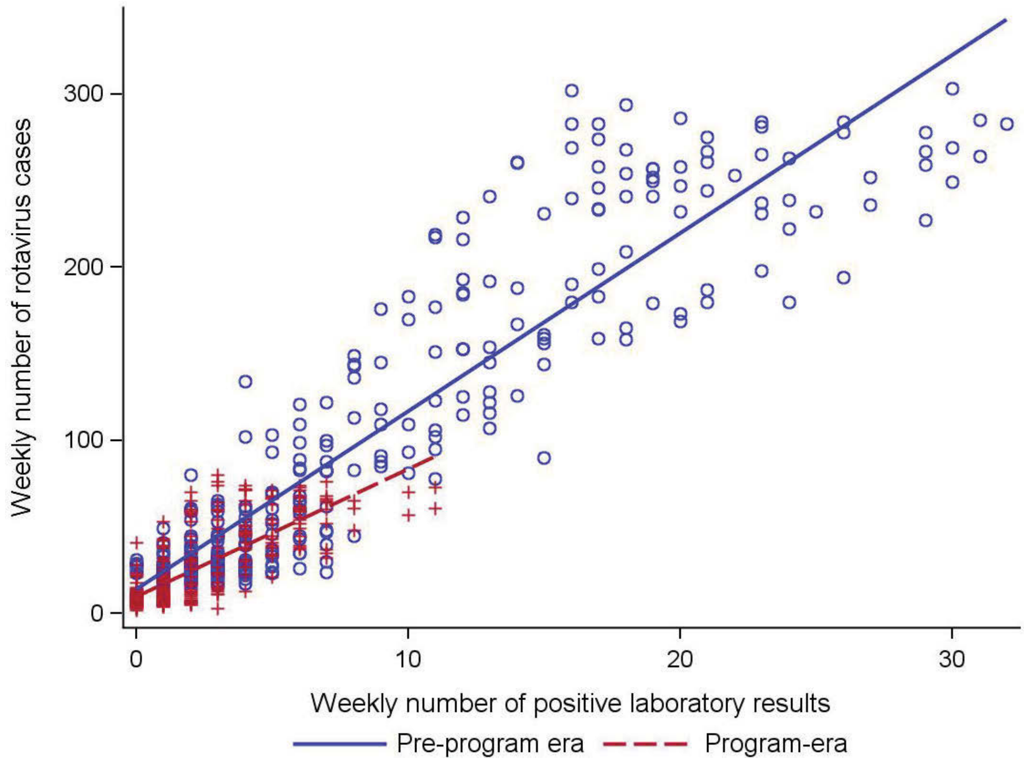


Figure 6.18: *Observed numbers of positive laboratory counts versus the rotavirus cases in the same week by pre- and post-program era from simulation 15 of scenario 2.*

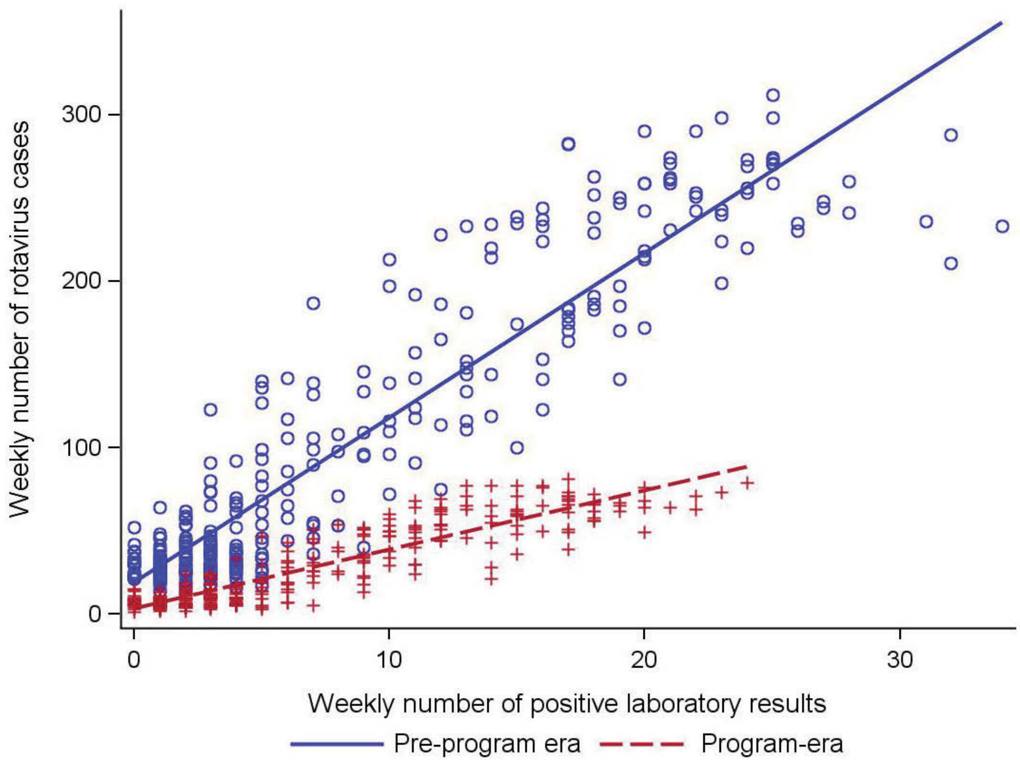


Figure 6.19: *True numbers of positive laboratory counts versus the rotavirus cases in the same week by pre-/post-program era and season for scenario 3.*

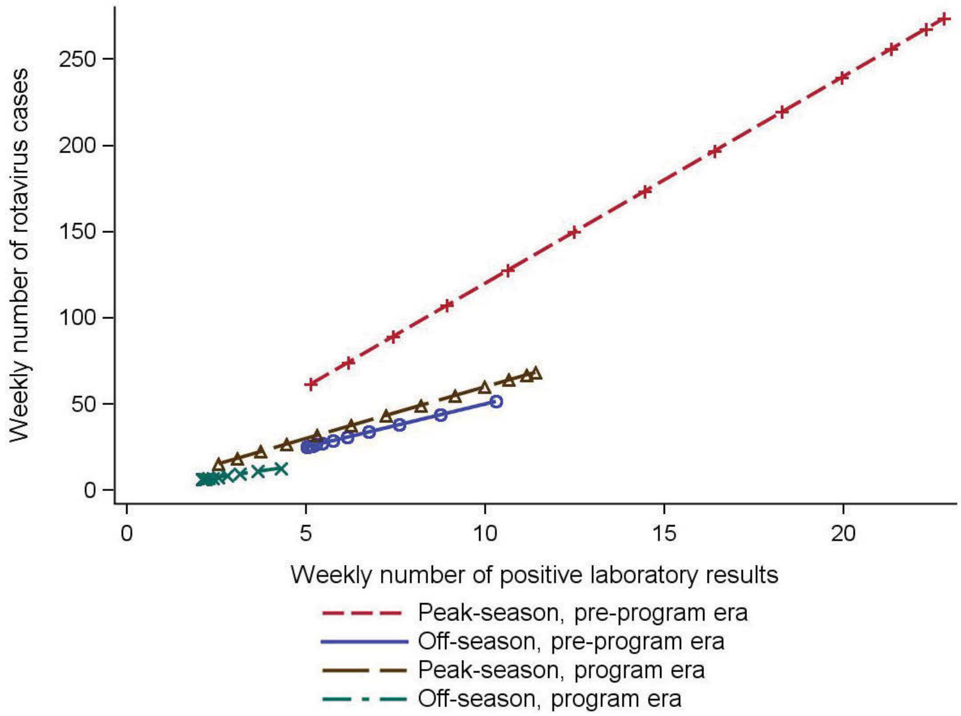


Figure 6.20: *Observed numbers of positive laboratory counts versus the rotavirus cases in the same week by pre-/post-program era and season from simulation 15 of scenario 3.*

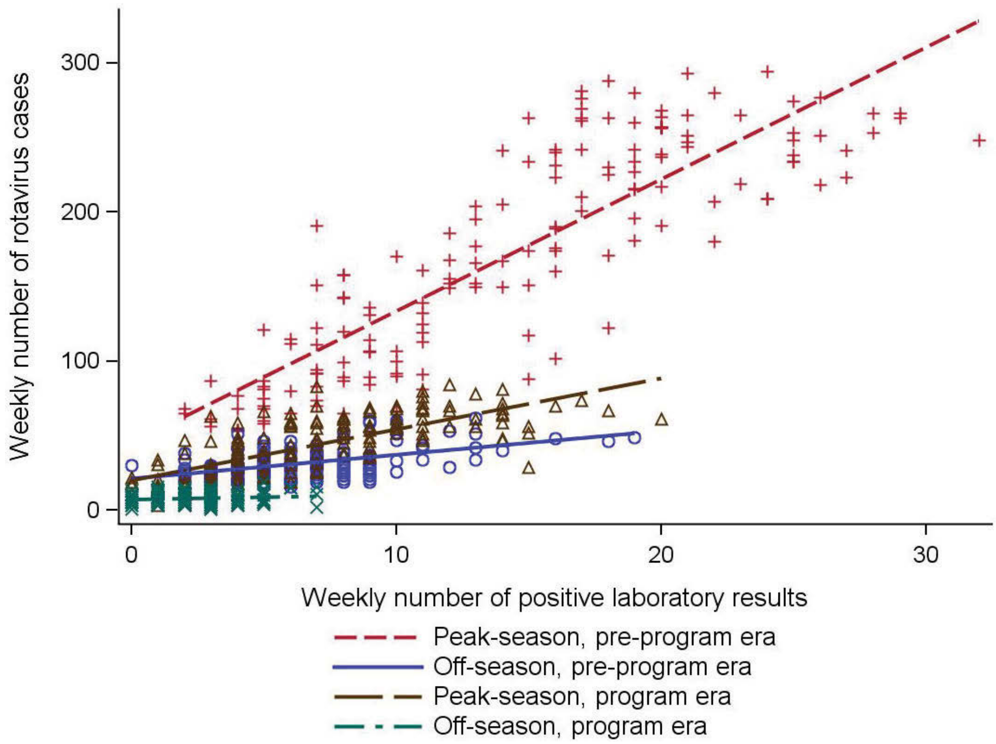


Table 6.18: *Scenario 3 results from 1000 simulations with seasonal interactions incorporated into the statistical models: ratio of weekly rotavirus cases per positive laboratory result varies by season (peak and off-peak) and program introduction (pre/post).*

	Percentage identified as rotavirus Median (2.5th, 97.5th percentile)		Ratio of rotavirus cases per positive laboratory result Median (2.5th, 97.5th percentile)					
					Pre-program era		Program era	
Method	Pre-program era	Program era	Peak season	Off-peak season	Peak season	Off-peak season	Peak season	Off-peak season
Truth	**57.5**	**25.3**			**12**	**6**	**5**	**3**
Interaction with season								
Linear regression	44.9 (42.3, 47.5)	14.5 (12.3, 16.7)	10.3 (9.65, 11.0)	3.08 (2.64, 3.56)	-	-	-	-
GAM, 6 knots per year	10.6 (8.02, 13.3)	-1.1 (-3.5, 1.18)	2.85 (2.27, 3.42)	-.23 (-.75, 0.25)	-	-	-	-
GAM, 2 knots per year	28.8 (25.9, 31.4)	-4.0 (-7.1, -1.2)	7.77 (7.17, 8.36)	-.85 (-1.5, -.25)	-	-	-	-
Interaction with program introduction and season								
Linear regression	28.5 (25.9, 30.9)	37.6 (34.1, 41.1)	-	-	8.88 (8.22, 9.57)	0.91 (0.20, 1.61)	3.09 (2.49, 3.61)	1.12 (0.14, 2.09)
GAM, 6 knots per year	10.3 (7.76, 13.1)	14.4 (10.6, 18.7)	-	-	3.57 (2.76, 4.44)	0.02 (-.55, 0.55)	1.16 (0.58, 1.79)	-.77 (-1.7, 0.05)
GAM, 2 knots per year	27.7 (25.1, 30.2)	36.4 (32.8, 40.1)	-	-	8.61 (7.91, 9.34)	0.87 (0.15, 1.50)	3.14 (2.58, 3.65)	0.80 (-.12, 1.74)

GAM: Generalised additive model

Table 6.19: *Observed percentages and model coefficients for the hospital data by method.*

	Percentage identified as rotavirus		Ratio of positive laboratory results to rotavirus cases		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Off-season attributable fractions	56.7	max 38.1, min 17.2 *	-	-	-
No interaction					
Linear regression	45.4	25.8	11.9	-	-
GAM, 6 knots per year	22.1	16.6	6.45	-	-
GAM, 2 knots per year	35.3	26.6	10.3	-	-
Interaction with program introduction					
Linear regression	41.5	16.6	-	12.1	6.45
GAM, 6 knots per year	24.6	9.57	-	7.17	3.71
GAM, 2 knots per year	38.4	14.6	-	11.2	5.66

GAM: Generalised additive model

[*] Max is when the off-season percent is assumed to be the same as in the pre-program era of 25.3. Min is when the off-season percent is assumed to be zero.

Table 6.20: *Observed percentages and model coefficients for the ED data by method.*

	Percentage identified as rotavirus		Ratio of positive laboratory results to rotavirus cases		
Method	Pre-program era	Program era	Constant	Pre-program era	Program era
Off-season attributable fractions	43.5	max 31.8, min 8.73 *	-	-	-
No interaction					
Linear regression	22.0	10.6	12.9	-	-
GAM, 6 knots per year	13.4	6.15	7.66	-	-
GAM, 2 knots per year	21.3	9.84	12.2	-	-
Interaction with program introduction					
Linear regression	24.0	8.27	-	13.7	10.3
GAM, 6 knots per year	14.1	4.96	-	8.07	6.18
GAM, 2 knots per year	22.3	7.24	-	12.8	9.03

GAM: Generalised additive model

* Max is when the off-season percent is assumed to be the same as in the pre-program era of 25.3. Min is when the off-season percent is assumed to be zero.

method gave the highest estimates and the generalised additive model with six knots per year gave the lowest estimates of the percentages.

6.7 Discussion

In this chapter we defined sixteen scenarios to test the robustness of methods from Chapter 5 that aimed to determine those acute gastroenteritis hospitalisations and ED presentations in children under five years of age in NSW that were due to rotavirus and those that were not. Some variations to the statistical models (linear regression and generalised additive models) were included that allowed the ratio of rotavirus cases to laboratory results to vary by program introduction or season by including interaction terms. These sixteen scenarios do not have equal weight in determining which method(s) performed best. The statistical models required an additional data source (positive laboratory results) to infer rotavirus burden. As a result more scenarios were needed to understand the robustness of these models than for the off-season attributable fractions method. Nor do these sixteen scenarios cover all components of the various methods or all possible permutations of the rotavirus, non-rotavirus and positive laboratory data signals.

There was no robust method identified by the sixteen scenarios investigated. The generalised additive model with six knots per year, with and without interactions terms, never captured the true percentage of rotavirus attributed acute gastroenteritis cases within its 2.5th to 97.5th percentile intervals in both the pre-program and program eras in any scenario. The generalised additive model with two knots per year, with and without a program interaction, contained the true percentages of rotavirus cases in the pre-program and program eras in its 2.5th to 97.5th percentile interval for Scenario 12 only (summer and winter seasonality in non-rotavirus disease)(Table 6.13). The linear regression model was successful in Scenarios 11 (winter seasonality in non-rotavirus disease) and 12, for both the no interaction and program interaction models, and Scenario 13 (decreasing linear trend in rotavirus disease over time) for the model with program interaction only (Tables 6.12, 6.13, 6.14). There were only six unique scenarios (Scenarios 1, 11, 12, 13, 15, 16) that tested the robustness of the off-season attributable fractions method (the other scenarios were all similar to these scenarios in terms of the parameter specifications and assumptions of the off-season attributable fractions method). This method was only successful in including the true percentages in its summary statistics for half of these six scenarios. It failed to capture the true percentages in the pre-program and program eras in its 2.5th to 97.5th percentile interval for three of the four scenarios that included seasonality in the non-rotavirus cases (Scenarios 11, 15, 16; Tables 6.12, 6.16, 6.17). It was successful in the fourth scenario because the seasonal pattern in the off-season was the same in the peak-season (Scenario 12; Table 6.13). None of the methods explored in this chapter was able to capture both of the true rotavirus percentages (pre-program and program eras) in the complex scenarios 15 and 16.

The results for the off-season attributable fractions method are not surprising given its simple definition of no seasonality in non-rotavirus cases. The results for the statistical methods are more complex to untangle. Recall that the linear regression model and generalised additive models with no interaction term assume that the number of rotavirus cases is a multiple of positive rotavirus laboratory results and that this is constant over time. Scenario 1 assumes this to be the case for the ratio (the true value is set at 12), yet the results of all the statistical models under-reported the true ratio of the number of rotavirus cases per positive laboratory results in a given week and hence under-estimated the true percentages of rotavirus-attributable acute gastroenteritis in the pre-program and program eras. Part of the issue is likely because it is difficult for the ratio of rotavirus cases per positive laboratory result in a given week to actually be constant over time, given the count based (i.e. integer) nature of these data. For example, the first simulation of Scenario 1 drew the value of 20 for the number of rotavirus cases in that week, which is not a multiple of 12. Another issue is with the number of non-seasonal rotavirus cases: when there was no non-seasonal rotavirus circulating the statistical models under-reported less (Scenarios 5, 6). This may be because the non-seasonal rotavirus cases in the off-season are hard to distinguish from random noise by the models.

In the scenarios where this assumption did not hold and the ratio of rotavirus cases per positive laboratory result varied by program introduction (pre-/post- introduction), season (peak-/off-season) or year (Scenarios 2, 3, 10, 15, 16; Tables 6.3, 6.4, 6.11, 6.16, 6.17) the linear regression model and generalised additive models with no interaction term either under- or over-estimated the true percentages, sometimes by a large amount. Relaxing the model assumption so that the ratio of rotavirus cases to positive laboratory results could vary by program era (pre-/post) but was constant within each era, by including a program interaction, sometimes improved the estimates (e.g. Scenario 2, Table 6.3). The inclusion of a seasonal interaction term with or without a vaccination program interaction did not. Table 6.18 shows a poor fit for models that contain an interaction with season (peak- and off-season) and program era (pre-/program-era) for Scenario 3 when this ratio was defined to vary by season and program introduction. The inability of the statistical models to estimate the correct ratio of the number of rotavirus cases per positive laboratory results in this case appears to be influenced by a lack of variability in these two variables in the off-season and program era (Figures 6.19 and 6.20).

A higher model coefficient of the estimate of the ratio of the number of rotavirus cases per positive laboratory result did not guarantee the highest estimated percentage of rotavirus attributed gastroenteritis cases (for example, Scenario 2, Table 6.3). This is due to the fact that the denominator for calculating the percentage of rotavirus attributed acute gastroenteritis cases for the statistical models is the predicted number of acute gastroenteritis cases of the model (and the observed acute gastroenteritis cases for the off-season attributable fractions method). This means that, even if the estimates of the statistical models contain the true percentage of rotavirus attributed cases they may not capture all the true numbers of rotavirus cases (or they may infer too many).

The coefficients that represent the number of inferred rotavirus cases for each positive laboratory result in a given week were consistently highest for the linear regression model as compared to the two generalised additive models and also the closest to the true values, both when interaction terms were and were not included in the models. The coefficients were lowest for the generalised additive model with six knots per year. This implies that how a method accounts for the autocorrelation when determining the absolute numbers of rotavirus cases is important, and that the generalised additive models over-fitted the data. In the linear regression model case, the estimate of the ratio(s) was not calculated independently of seasonal variation or autocorrelation. When the coefficient(s) obtained from the model was multiplied by the positive laboratory results to give the number of estimated rotavirus cases per week this approach allocated the seasonal variation and autocorrelation in the positive laboratory results to rotavirus cases and any other seasonality or autocorrelation to non-rotavirus disease cases. In the case of the generalised additive models, the ratio(s) was estimated after accounting for autocorrelation through a non-parametric spline. The amount that it accounted for increased as the number of knots per year increased. This means that, when the coefficient(s) obtained from the model was multiplied by the positive laboratory results to give the number of estimated rotavirus cases per week this approach was allocating all the autocorrelation identified by the model to non-rotavirus cases; however some of this needs to be allocated to the rotavirus cases as well. This essentially highlights, again, the issue that regression models are typically used to understand associations but not estimate absolute numbers. Both statistical approaches underestimated the true percentages of rotavirus cases in simple scenarios where their assumptions held (e.g. Scenarios 1 and 5; Tables 6.2, 6.6). The noteworthy underreporting for the majority of the results produced by the generalised additive model with six knots per year infers that the results published by Davey et al. [48] are likely to have underestimated the true numbers of rotavirus cases in NSW EDs, and hence underestimate the absolute decline in rotavirus due to vaccination efforts.

The linear regression model and generalised additive model with two knots per year were successful at identifying the true percentages of rotavirus attributed acute gastroenteritis cases in Scenarios 11 and 12. Given that these models under-reported the true percentages of rotavirus attributed acute gastroenteritis cases for the basic scenarios 1 and 5, it is likely that they were successful here because the seasonal non-rotavirus cases offset the non-seasonal rotavirus cases. Any changes to the number of non-seasonal rotavirus cases or magnitude of the seasonal non-rotavirus cases in Scenarios 11 and 12 will likely impact on these results and they may no longer capture the true percentages.

There was a large range in the estimated percentages of rotavirus attributed acute gastroenteritis cases in the pre-program and program eras when the different methods were applied to the hospital and ED data for children under five years of age in NSW. The original intention of this simulation study was to identify a robust method(s) to separate the hospital and ED acute gastroenteritis cases into those that were due to rotavirus and those that were not. After application of the robust method I was then going to carry out a full impact analysis using the two methods applied in Chapter 4. That is, we had planned to apply a multivariable time series regression model to assess any changes in temporal trends of rotavirus disease burden in NSW in children under five years of age due to rotavirus vaccination efforts. Another model, an ARIMA model, would have been used to quantify the changes in rotavirus disease burden identified by forecasting “no vaccination program” rotavirus disease burden into the vaccination era and comparing the forecasts to the observed rotavirus disease burden in the vaccination era. However, given the results of this simulation study, where no robust method was identified, we do not know which scenario is closest to our observed hospital and ED data, and therefore we decided not to do this. Instead, we rely on the results from Chapter 4 where we assessed the impact of the rotavirus vaccination program on acute gastroenteritis hospitalisations and ED presentations in children under five years of age in NSW (July 2001 to June 2013) as our best estimate of the impact of the Australian rotavirus vaccination program on disease burden of hospitalisations and ED presentations in under five year olds in NSW.

These results imply that other methods or data are needed to be able to accurately separate our non-specific acute gastroenteritis data into rotavirus and non-rotavirus disease under a range of plausible conditions. Harris et al. had access to positive laboratory results in England and Wales for ten pathogens that cause acute gastroenteritis, including rotavirus [93]. The estimated percentage of rotavirus attributed acute gastroenteritis hospitalisations in children under five years of age was 51.4% (95% CI: 49.5-53.2%) when only positive rotavirus laboratory results were included in the simple linear regression model. When the individual positive laboratory results for all ten pathogens were included in the model, the estimate of rotavirus attributed gastroenteritis was similar (50.8%, 95% CI: 48.6-52.7%). This may suggest that if their model with only positive rotavirus laboratory results as a covariate underestimates the true percentage of rotavirus attributed acute gastroenteritis (as may be the case for our NSW data) then the model with ten covariates may underestimate the true percentage as well. This will likely depend on how well the other covariates predict their diseases, the seasonal variation in each of those diseases and their respective positive laboratory test results. Another approach may be to try and estimate the percentages of rotavirus attributed acute gastroenteritis burden from laboratory data where negative results are also considered. This will require testing a representative sample of infants and children presenting with acute gastroenteritis and that the data is available in small enough units (e.g. weekly data) to carry out an impact analysis on the identified rotavirus and non-rotavirus cases.

This simulation study has several limitations. To calculate estimates using the off-season attributable fractions method in these simulations we needed to define the proportion of rotavirus cases in the off-season in both the pre-program era and program era. We did this by assigning the off-season percentages from the known truth of each scenario. In practice, this is determined by medical audits and there will be some uncertainty in the estimates. This study could have included some scenarios to show this. This study could have also included other scenarios to show how differences in the width and timing of the rotavirus seasonal peak affect the results produced by the various methods and other alterations to the scenario assumptions. However, it is difficult to cover all permutations of the rotavirus, non-rotavirus and positive laboratory counts over time. Our study tried to cover all the key scenarios needed to understand how the methods perform when their assumptions are and are not met. This study used the Poisson distribution to generate the number of rotavirus cases, non-rotavirus cases and positive laboratory counts for any given week. It could have used a Negative Binomial distribution or an over-dispersed Poisson distribution to generate this count data instead, which would have changed the variation in these data. However, it is unlikely that the choice of the distribution for generating these count data would have any significant impact on the performance of the models being compared in this chapter. This study used the median and 2.5th upper and lower percentiles to compare methods within each of the sixteen scenarios to try and determine the most precise method for estimating the proportion of proportion of acute gastroenteritis hospitalisations and ED presentations that were attributable to rotavirus in children aged under five years. Other measures of reporting, such as the mean squared error and coverage, could have been used instead and would have provided further details on the bias and variance of the results for a more detailed comparison.

This study has several strengths. It applied all the methods to the same datasets, instead of each method to an independent data set, which is typical of simulations studies. It was important to use this approach to eliminate the within sample variability of the generated datasets and therefore provide a sensitive means to detect any differences between methods [95]. There were sufficient simulations generated (1000) for each scenario to understand the underlying empirical distributions of the results from each method. This study investigated whether small additions to the statistical models (e.g. interaction terms) improved the performance of these statistical methods. Although no robust method was identified in this simulation study, this technique of exploring appropriate methods using a simulation study can be used for other diseases, not only rotavirus, to understand their performance when identifying specific disease from non-specific disease. The same methods explored in this study will not be appropriate for all diseases. For example, the off-season attributable fractions method will not be appropriate for non-specific diseases that have multiple diseases driving its seasonal peak. However, it may be possible to adapt this method for the disease of interest.

6.8 Chapter summary

This chapter investigated multiple methods to separate acute gastroenteritis cases into those that are due to rotavirus and those that are not. There was no robust method identified across the sixteen scenarios investigated. This technique of exploring appropriate methods using a simulation study can be used for other diseases, not only rotavirus, to understand their performance when identifying specific disease from non-specific disease. The same methods explored in this study will not be appropriate for all diseases but it may be possible to adapt this method for the disease of interest. The next chapter, the final chapter, gives a thesis summary.

Chapter 7: Thesis Summary

7.1 Chapter summary

This concluding chapter summarises the key results of this thesis and provides suggestions for future research and recommendations for future analyses that assess the impact of vaccination programs.

7.2 Summary of thesis results

A summary of the results of this thesis is given below and separated by the aims of this thesis.

7.2.1 Thesis Aim 1: To provide an outline of the broad classes of statistical methods that can be used to estimate changes in burden of disease over time that best account for the characteristics of infectious diseases

This thesis provided a structured explanation of statistical methods and how they can be used to address the epidemiological challenges in assessing changes in burden of disease as a result of vaccination programs. It began with a summary in Chapter 2 of the key features of infectious diseases and the limitations of routine surveillance and health databases that affect our ability to detect changes in the burden of diseases as a result of vaccination programs.

The literature review in Chapter 2 showed that there is variation in the methods used in studies for assessing the impact of vaccination programs and that the full details and justification for the choice of method is not always stated. This suggests that further clarification of the most appropriate statistical methods in assessing the impact of vaccination programs is required, which this thesis provides. Changes in descriptive statistics from the pre-program to post-vaccination program eras are often used to analyse the impact of vaccination programs, such as rotavirus, where there is an obvious and strong decline. Chapters 2 and 4 explained why it is not appropriate to use this method in the presence of temporal trends, since this method could result in inadvertently attributing a decline to the program that is really the result of other underlying patterns over time.

Chapter 4 explained in detail why two methods are needed to carry out a full analysis of the impact of a vaccination program on burden of disease. One method is needed to assess associations between changes in disease burden and introduction of the vaccination program and another method is needed to quantify changes in disease burden if an association was identified.

7.2.2 Thesis Aim 2: To estimate the impact of the Australian rotavirus vaccination program on acute gastroenteritis hospitalisations and emergency department presentations (non-specific disease) in children under five years of age in New South Wales, Australia, using appropriate statistical methods

Chapter 4 modelled the rates of hospitalisations and ED presentations in children under five years of age in NSW, Australia, using multivariable time series regression models to identify changes in temporal trends due to vaccination efforts. For both hospitalisations and ED presentations there was a statistically significant decline in the magnitude of the seasonal peaks, which resulted in a decline in the overall hospital and ED burden. There was borderline evidence of a change in the long-term trend of acute gastroenteritis ED presentations after introduction of the Australian rotavirus vaccination program (from increasing to decreasing). Next an ARIMA model was fitted to the data from the pre-vaccination program era to forecast into the program era and quantify the changes in disease burden. It found that, on average, the Australian rotavirus vaccination program has halved the weekly rate of acute gastroenteritis hospitalisations and led to a sixteen percent reduction in the weekly rate of acute gastroenteritis ED presentations in NSW.

These results were similar to the conclusions drawn when analysing the impact of the rotavirus vaccination program in NSW using changes in descriptive statistics from the pre- to post-vaccination program introduction era, which other Australian and international studies have applied (Chapter 4). However, this can only be known after applying appropriate methods that consider temporal trends. If there had been a strong underlying trend present in the data before implementation of the rotavirus vaccination program then the descriptive methods would not be appropriate and may have led to an over or under estimation of the impact of the vaccination program.

Using complex methods, as used in this thesis, can lead to a more detailed understanding of any changes in disease burden. The general approach suggested in this thesis can be applied when analysing the impact of a vaccination program for diseases other than rotavirus as well. Using these complex methods is particularly important for assessing diseases where the possible decline due to vaccination efforts is not as obvious as it is for rotavirus. For example, when assessing declines in all-cause pneumonia hospitalisations as a results of pneumococcal vaccination [97]. Our multivariable times series regression models were able to enumerate the size of the seasonal peaks and troughs and their timing and to assess whether the underlying trend or periodicity had changed through vaccination efforts using statistical tests, which the descriptive analytic approach is unable to do. We found that, on average, the decline in the magnitude of the periodicity was from 48.9 to 10.0 in the hospital data and 53.8 to 37.2 in the ED data per 100,000 children less than five years of age in NSW. Further, the ED data changed from an increasing trend on the multiplicative scale at a rate of 52.01 ED presentations per 100,000 population per year in the pre-program era to a decreasing trend of 51.9 ED presentations per 100,000 population per year on the multiplicative scale in the program era.

7.2.3 Thesis Aim 3: To directly compare statistical methods that can separate hospitalisations and ED presentations in children under five years of age in NSW for acute gastroenteritis into those that are due to rotavirus and those that are not

The analyses in Chapter 4 assessed the impact of the rotavirus vaccination program in NSW on acute gastroenteritis disease burden (non-specific disease), of which rotavirus is a subset, and did not analyse rotavirus hospitalisations and rotavirus ED presentations specifically. Chapters 2 and 3 explained why it is difficult to obtain accurate rates of rotavirus disease burden in NSW due to lack of confirmation of rotavirus disease. Chapter 5 explained why it is still important to estimate the true rotavirus disease burden for both pre-implementation decision making and post-implementation evaluations.

Chapter 5 applied four previously published methods to separate the NSW hospitalisations and ED presentations for acute gastroenteritis in children under five years of age into those that are due to rotavirus and those that are not. The varied results observed and the uncertainty around the true numbers of rotavirus hospitalisations and ED presentations meant that it was not clear which of the four methods performed best. A simulation study was conducted in Chapter 6 to evaluate the robustness of these statistical methods under ideal and non-ideal conditions (i.e. where the assumptions of the methods did or did not hold) and to offer clarification on the strengths and weaknesses of the competing methods. Additional methods were included in the simulation study to evaluate whether relaxing the assumptions of the statistical models increased their accuracy.

Results of the simulations depended on the assumptions made but gave a clearer understanding of the uncertainty of each estimate. Surprisingly, there was no robust method identified across the sixteen scenarios investigated. The generalised additive model with six knots per year is not an appropriate method to be used for the purposes of separating acute gastroenteritis hospitalisations and ED presentations in NSW into those that are due to rotavirus and those that are not. There are potential scenarios where the other methods considered are appropriate. However, we do not know which scenario is closest to the true percentage of rotavirus disease in our observed hospital and ED data in NSW.

7.2.4 Thesis Aim 4: To estimate the fraction of acute gastroenteritis hospitalisations and ED presentations in children under five years of age in NSW that may be due to rotavirus, and investigate whether this fraction has changed following the introduction of the rotavirus vaccine

Chapter 5 applied four previously published methods to separate the NSW hospitalisations and ED presentations for acute gastroenteritis in children under five years of age into those that are due to rotavirus and those that are not. The estimated percentage of rotavirus attributed acute gastroenteritis hospitalisations over the entire study period ranged from 20.2 to 50.1 depending on the method used. Likewise, the estimated percentage for rotavirus attributed acute gastroenteritis ED presentations ranged from 9.9 to 37.9.

Chapter 6 calculated these percentages (of acute gastroenteritis hospitalisations and ED presentations in children under five years of age in NSW) attributed to rotavirus separately in the pre-program and vaccination program eras. The percentages were calculated using the methods from Chapter 5 as well as additional methods that included interaction terms in the statistical models. The percentages varied across methods; ranging from 22.1 to 56.7 in the pre-program era and from 9.57 to 38.1 in the program era for the hospital data. The estimated percentages ranged from 13.4 to 43.5 in the pre-program era and from 5.0 to 31.8 in the program era for the ED data. The lowest percentages were consistently produced by the generalised additive model with six knots per year and the highest percentages by the off-season attributable fractions method.

Since there was no robust method identified in Chapter 6 and the results varied in Chapters 5 and 6 when applied to the hospital and ED data by method, we were unable to conclude which method is appropriate for our data to estimate the true fraction of acute gastroenteritis hospitalisations and ED presentations in children under five years of age in NSW that is due to rotavirus. Therefore we did not investigate whether this fraction has changed through rotavirus vaccination efforts. Instead, we concluded that the analyses carried out in Chapter 4, which analysed the impact of the rotavirus vaccination program on acute gastroenteritis disease burden in NSW, are the most appropriate for our data.

7.3 Strengths and limitations

This thesis provided a structured explanation of appropriate statistical methods that can be used to address the epidemiological challenges in assessing changes in burden of disease as a result of vaccination programs. The analyses focused on rotavirus disease burden in NSW, Australia. Nonetheless, many of the statistical approaches and considerations discussed in this thesis can be applied to other infectious diseases as well. The multivariable time series regression model and ARIMA model used to fully analyse the impact of the rotavirus vaccination program on acute gastroenteritis disease burden is not restricted to this disease. The advantages of these methods will be most noteworthy when applied to diseases where there is no obvious decline due to vaccination efforts or there is an underlying trend prior to introduction of the vaccination program.

The methods explored in this thesis to separate non-specific acute gastroenteritis hospitalisations and ED presentations into rotavirus and non-rotavirus attributed disease are specific to these diseases (Chapters 5 and 6). These methods may be appropriate for other diseases but may need to be adapted for the disease of interest. For example, similar regression methods have been used for separating pneumonia into those due to influenza and those that are not [45]. However, the off-season attributable fraction method would need adapting since it would be unreasonable to assume all of the seasonal rise in pneumonia is due to influenza. Nonetheless, the technique of exploring appropriate methods using a simulation study can be used for any disease to understand their performance when identifying specific disease from non-specific disease.

This study analysed the aggregated weekly counts of children under five years of age hospitalised or presenting to the ED for acute gastroenteritis in NSW from 1 July 2001 to 30 June 2013 (Chapter 4). These data collections capture almost all hospitalisations and over 70% of ED presentations in NSW (Chapter 3). This provided sufficient power to analyse weekly data, which is important for assessing temporal trends. The impact of vaccinating infants against rotavirus in their first six months of life was seen almost immediately in all other age groups under five years of age due to herd immunity. Therefore, it was appropriate to look at aggregated counts of under five year olds. However, this may not be the case for other diseases. In such situations it would be necessary to assess individual vaccinated cohorts through time after introduction of the vaccination program or to analyse the impact in single year ages. If this results in insufficient power to perform these analyses on weekly data then monthly data should be considered.

This study did not try to separate out the impact of herd immunity from the direct impact (to those vaccinated) after introduction of the national rotavirus vaccination program. To assess the impact of unvaccinated individuals within the vaccinated cohorts we would have needed to have access to linked records to assess vaccination status. In unvaccinated cohorts (i.e. older ages) the same methods presented in this thesis could be used to determine the indirect impact of the rotavirus vaccination program on these unvaccinated cohorts. Again, if there is insufficient data to analyse weekly data then monthly data should be considered.

We did not use a Poisson regression model to carry out the impact analysis of rotavirus vaccination on acute gastroenteritis hospitalisations and ED presentations even though these are count data. This is because it is difficult to account for autocorrelation with a Poisson model, which is necessary when analysing the impact of vaccination efforts on disease burden over time. Rotavirus is a common disease and therefore the use of a Gaussian model was reasonable.

Despite using a simulation study, this work was unable to estimate the true fraction of rotavirus attributed acute gastroenteritis hospitalisations and ED presentations in under five year olds in NSW. Other methods or data are needed to confidently separate non-specific acute gastroenteritis data into rotavirus and non-rotavirus disease. Future work could try to estimate the percentages of rotavirus attributed acute gastroenteritis cases from laboratory data where negative rotavirus test results are also considered. This will require testing a representative sample of infants and children presenting with acute gastroenteritis and that the data is available in small enough units (e.g. weekly data) to carry out an impact analysis on the identified rotavirus and non-rotavirus cases.

7.4 Recommendations

Rigorous statistical methods are required to ensure the robustness and validity of findings. Future analyses of the impact of vaccination programs should consider:

- Separate methods are required to:
    - estimate any changes in temporal trends due to vaccination programs, and
    - estimate any changes in disease burden.
- It is not appropriate to use simple pre- and post-program introduction methods in the presence of temporal trends.
- Autocorrelation needs to be considered when assessing diseases over time
- It is important for authors to justify the method chosen and discuss any limitations
- Using appropriate complex methods leads to a more detailed understanding in any changes in disease burden observed.

7.5 Thesis conclusion

This thesis provided a structured explanation of statistical methods and how they can be used to address the epidemiological challenges in assessing changes in burden of disease as a result of vaccination programs. While it focused on rotavirus disease burden in NSW, Australia, many of the concepts discussed are applicable to other infectious diseases. Future studies that assess the impact of vaccination programs on burden of disease should consider the lessons learned from the approaches outlined in this thesis to ensure that all important factors are considered in their evaluations.

References

1. Plotkin SL, Plotkin SA: A short history of vaccination. In: Vaccines. 6th edn. Edited by Plotkin SA, Orenstein WA, Offit P. Philadelphia: Elsevier-Saunders; 2013: 1-13.
2. The History of Vaccines. A project of The College of Physicians of Philadelpia, http://www.historyofvaccines.org [last accessed: 29/8/2017].
3. Anderson RM, May RM: Vaccination and herd immunity to infectious diseases. Nature 1985, 318(6044):323-329.
4. Mealing N, Hayen A, Newall AT: Assessing the impact of vaccination programmes on burden of disease: Underlying complexities and statistical methods. Vaccine 2016, 34(27):3022-3029.
5. Greenwood M, Yule GU: The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. Proceedings of the Royal Society of Medicine 1915, 8(Sect Epidemiol State Med):113-194.
6. Weinberg GA, Szilagyi PG: Vaccine epidemiology: Efficacy, effectiveness, and the translational research roadmap. The Journal of Infectious Diseases 2010, 201(11):1607-1610.
7. Newall AT, Reyes JF, Wood JG, McIntyre P, Menzies R, Beutels P: Economic evaluations of implemented vaccination programmes: key methodological challenges in retrospective analyses. Vaccine 2014, 32(7):759-765.
8. Nokes DJ, Anderson RM: The use of mathematical models in the epidemiological study of infectious diseases and in the design of mass immunization programmes. Epidemiology and Infection 1988, 101(1):1-20.

9. Clemens JD, Shapiro ED: Resolving the pneumococcal vaccine controversy: are there alternatives to randomized clinical trials? Reviews of Infectious Diseases 1984, 6(5):589-600.

10. Curns AT, Steiner CA, Barrett M, Hunter K, Wilson E, Parashar UD: Reduction in acute gastroenteritis hospitalizations among US children after introduction of rotavirus vaccine: analysis of hospital discharge data from 18 US states. The Journal of Infectious Diseases 2010, 201(11):1617-1624.

11. International Statistical Classification of Diseases and Related Health Problems. 10th edn. Geneva: World Health Organization; 1992.

12. Nigrovic LE, Fine AM, Monuteaux MC, Shah SS, Neuman MI: Trends in the management of viral meningitis at United States children's hospitals. Pediatrics 2013, 131(4):670-676.

13. Jayasinghe S, Macartney K: Estimating rotavirus gastroenteritis hospitalisations by using hospital episode statistics before and after the introduction of rotavirus vaccine in Australia. Vaccine 2013, 31(6):967-972.

14. Chiu C, Dey A, Wang H, Menzies R, Deeks S, Mahajan D, Macartney K, Brotherton J, Jardine A, Quinn H et al.: Vaccine preventable diseases in Australia, 2005 to 2007. Communicable Diseases Intelligence 2010, 34 Supp:S1-S167.

15. Clark HF, Offit PA, Parashar UD: Rotavirus vaccines. In: Vaccines. 6th edn. Edited by Plotkin SA, Orenstein WA, Offit PA. Philadelphia: Elsevier-Saunders; 2013: 669-687.

16. Macartney KK, Porwal M, Dalton D, Cripps T, Maldigri T, Isaacs D, Kesson A: Decline in rotavirus hospitalisations following introduction of Australia's national rotavirus immunisation programme. Journal of

Paediatrics and Child Health 2011, 47(5):266-270.

17. Dias R, Canica M: Invasive pneumococcal disease in Portugal prior to and after the introduction of pneumococcal heptavalent conjugate vaccine. FEMS Immunology and Medical Microbiology 2007, 51(1):35-42.

18. Huang SS, Platt R, Rifas-Shiman SL, Pelton SI, Goldmann D, Finkelstein JA: Post-PCV7 changes in colonizing pneumococcal serotypes in 16 Massachusetts communities, 2001 and 2004. Pediatrics 2005, 116(3):e408-e413.

19. Klugman KP, Black S, Dagan R, Malley R, Whitney CG: Pneumococcal conjugate vaccine and pneumococcal common protein vaccines. In: Vaccines. 6th edn. Edited by Plotkin SA, Orenstein WA, Offit PA. Philadelphia: Elsevier-Saunders; 2013: 504-541.

20. Pilishvili T, Lexau C, Farley MM, Hadler J, Harrison LH, Bennett NM, Reingold A, Thomas A, Schaffner W, Craig AS et al: Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. The Journal of Infectious Diseases 2010, 201(1):32-41.

21. Salleras L, Dominguez A, Ciruela P, Izquierdo C, Navas E, Torner N, Borras E: Changes in serotypes causing invasive pneumococcal disease (2005-2007 vs. 1997-1999) in children under 2 years of age in a population with intermediate coverage of the 7-valent pneumococcal conjugated vaccine. Clinical Microbiology and Infection 2009, 15(11):997-1001.

22. Granoff DM, Pelton S, Harrison LH: Meningococcal vaccines. In: Vaccines. 6th edn. Edited by Plotkin SA, Orenstein WA, Offit PA. Philadelphia: Elsevier-Saunders; 2013: 388-418.

23. Hellenbrand W, Elias J, Wichmann O, Dehnert M, Frosch M, Vogel U: Epidemiology of invasive meningococcal disease in Germany, 2002–

2010, and impact of vaccination with meningococcal C conjugate vaccine. Journal of Infection 2013, 66(1):48-56.

24. Bramley JC, Jones IG: Epidemiology of chickenpox in Scotland: 1981 to 1998. Communicable Disease and Public Health 2000, 3(4):282-287.
25. Chang LY, Huang LM, Chang IS, Tsai FY: Epidemiological characteristics of varicella from 2000 to 2008 and the impact of nationwide immunization in Taiwan. BMC Infectious Diseases 2011, 11:352.
26. Deguen S, Chau NP, Flahault A: Epidemiology of chickenpox in France (1991-1995). Journal of Epidemiology and Community Health 1998, 52(Suppl 1):46S-49S.
27. Gershon AA, Takahashi M, Seward JF: Varicella vaccine. In: Vaccines. 6th edn. Edited by Plotkin SA, Orenstein WA, Offit P. Philadelphia: Elsevier- Saunders; 2013: 837-869.
28. Meyer PA, Seward JF, Jumaan AO, Wharton M: Varicella mortality: trends before vaccine licensure in the United States, 1970-1994. The Journal of Infectious Diseases 2000, 182(2):383-390.
29. Socan M, Kraigher A, Pahor L: Epidemiology of varicella in Slovenia over a 20-year period (1979-98). Epidemiology and Infection 2001, 126(2):279-283.
30. Agresti A: Categorical Data Analysis. 3rd edition. New Jersey: John Wiley & Sons; 2013.
31. Barnett A, Dobson A: Analysing seasonal health data. Statistics for biology and health. 1st edn. Heidelberg: Springer; 2010.
32. Brockwell P, Davis R: Introduction to time series and forecasting. 2nd edn. New York: Springer; 2002.
33. Hohle M, Siedler A, Bader HM, Ludwig M, Heininger U, Von Kries R:

Assessment of varicella vaccine effectiveness in Germany: a time-series approach. Epidemiology and Infection 2011, 139(11):1710-1719.

34. Simonsen L, Reichert TA, Viboud C, Blackwelder WC, Taylor RJ, Miller MA: Impact of influenza vaccination on seasonal mortality in the US elderly population. Archives of internal medicine 2005, 165(3):265-272.

35. do Carmo GM, Yen C, Cortes J, Siqueira AA, de Oliveira WK, Cortez-Escalante JJ, Lopman B, Flannery B, de Oliveira LH, Carmo EH et al.: Decline in diarrhea mortality and admissions after routine childhood rotavirus immunization in Brazil: a time-series analysis. PLoS Medicine 2011, 8(4):e1001024.

36. Kupek E, Tritany EF: Impact of vaccination against varicella on the reduction of the disease incidence in children and adolescents from Florianopolis, Brazil. Jornal de Pediatria 2009, 85(4):365-368.

37. Box G, Jenkins G, Reinsel G, Ljung G: Time Series Analysis: Forecasting and Control, 5th edn. New Jersey: Wiley; 2016.

38. Hardelid P, Pebody R, Andrews N: Mortality caused by influenza and respiratory syncytial virus by age group in England and Wales 1999-2010. Influenza and Other Respiratory Viruses 2013, 7(1):35-45.

39. Goldstein E, Viboud C, Charu V, Lipsitch M: Improving the estimation of influenza-related mortality over a seasonal baseline. Epidemiology 2012, 23(6):829-838.

40. Kutner MH, Nachtsheim CJ, Neter J, Li W: Applied linear statistical models. New York: McGraw Hill; 2005.

41. Gay NJ, Andrews NJ, Trotter CL, Edmunds W: Estimating deaths due to influenza and respiratory syncytial virus. JAMA 2003, 289(19):2499-2502.

42. McCullagh P, JA N: Generalized Linear Models, 2nd edn: London

Chapman and Hall; 1989.

43. Pregibon D: Goodness of link tests for generalized linear models. Journal of the Royal Statistical Society Series C (Applied Statistics) 1980, 29(1):15-14.

44. Bilcke J, Beutels P: Reviewing the cost effectiveness of rotavirus vaccination: the importance of uncertainty in the choice of data sources. PharmacoEconomics 2009, 27(4):281-297.

45. Thompson WW, Weintraub E, Dhankhar P, Cheng PY, Brammer L, Meltzer MI, Bresee JS, Shay DK: Estimates of US influenza-associated deaths made using four different methods. Influenza and Other Respiratory Viruses 2009, 3(1):37-49.

46. Newall AT, Viboud C, Wood JG: Influenza-attributable mortality in Australians aged more than 50 years: a comparison of different modelling approaches. Epidemiology and Infection 2010, 138(6):836-842.

47. Serfling RE: Methods for current statistical analysis of excess pneumonia-influenza deaths. Public Health Reports 1963, 78(6):494-506.

48. Davey HM, Muscatello DJ, Wood JG, Snelling TL, Ferson MJ, Macartney KK: Impact of high coverage of monovalent human rotavirus vaccine on emergency department presentations for rotavirus gastroenteritis. Vaccine 2015, 33(14):1726-1730.

49. Lambert SB, Faux CE, Hall L, Birrell FA, Peterson KV, Selvey CE, Sloots TP, Nissen MD, Grimwood K: Early evidence for direct and indirect effects of the infant rotavirus vaccine program in Queensland. The Medical Journal of Australia 2009, 191(3):157-160.

50. Buttery JP, Lambert SB, Grimwood K, Nissen MD, Field EJ, Macartney KK, Akikusa JD, Kelly JJ, Kirkwood CD: Reduction in rotavirus-associated acute gastroenteritis following introduction of rotavirus vaccine

into Australia's National Childhood vaccine schedule. The Pediatric Infectious Disease Journal 2011, 30(Suppl 1):S25-29.

51. Clarke MF, Davidson GP, Gold MS, Marshall HS: Direct and indirect impact on rotavirus positive and all-cause gastroenteritis hospitalisations in South Australian children following the introduction of rotavirus vaccination. Vaccine 2011, 29(29-30):4663-4667.

52. Dey A, Wang H, Menzies R, Macartney K: Changes in hospitalisations for acute gastroenteritis in Australia after the national rotavirus vaccination program. The Medical journal of Australia 2012, 197(8):453-457.

53. Pendleton A, Galic M, Clarke C, Ng SP, Ledesma E, Ramakrishnan G, Liu Y: Impact of rotavirus vaccination in Australian children below 5 years of age: a database study. Human vaccines & immunotherapeutics 2013, 9(8):1617-1625.

54. Ogunjimi B, Van Damme P, Beutels P: Herpes Zoster Risk Reduction through Exposure to Chickenpox Patients: A Systematic Multidisciplinary Review. PloS One 2013, 8(6):e66485.

55. Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N: Statistical methods for the prospective detection of infectious disease outbreaks: a review. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2012, 175(1):49-82.

56. Galati JC, Harsley S, Richmond P, Carlin JB: The burden of rotavirus-related illness among young children on the Australian health care system. Australian and New Zealand journal of public health 2006, 30(5):416-421.

57. Parashar UD, Burton A, Lanata C, Boschi-Pinto C, Shibuya K, Steele D, Birmingham M, Glass RI: Global mortality associated with rotavirus disease among children in 2004. The Journal of Infectious Diseases 2009, 200(Suppl 1):S9-s15.

58. Tate JE, Burton AH, Boschi-Pinto C, Parashar UD: Global, Regional, and National Estimates of Rotavirus Mortality in Children <5 Years of Age, 2000-2013. Clinical Infectious Diseases 2016, 62(Suppl 2):S96-s105.
59. Butz AM, Fosarelli P, Dick J, Cusack T, Yolken R: Prevalence of rotavirus on high-risk fomites in day-care facilities. Pediatrics 1993, 92(2):202-205.
60. Bresee JS, Glass RI, Ivanoff B, Gentsch JR: Current status and future priorities for rotavirus vaccine development, evaluation and implementation in developing countries. Vaccine 1999, 17(18):2207-2222.
61. Velazquez FR, Matson DO, Calva JJ, Guerrero L, Morrow AL, Carter-Campbell S, Glass RI, Estes MK, Pickering LK, Ruiz-Palacios GM: Rotavirus infection in infants as protection against subsequent infections. The New England Journal of Medicine 1996, 335(14):1022-1028.
62. Franco MA, Angel J, Greenberg HB: Immunity and correlates of protection for rotavirus vaccines. Vaccine 2006, 24(15):2718-2731.
63. Centres for Disease Control: Rotavirus vaccine for the prevention of rotavirus gastroenteritis among children. Recommendations of the Advisory Committee on Immunization Practices (ACIP). Morbidity and Mortality Weekly Report 1999, 48(RR-2):1-20.
64. Centres for Disease Control: Withdrawal of rotavirus vaccine recommendation. Morbidity and mortality weekly report 1999, 48(43):1007.
65. Australian Technical Advisory Group on Immunisation: The Australian immunisation handbook. 10th edn (2017 update). Canberra: Department of Health, Australian Government; 2017.
66. Ruiz-Palacios GM, Perez-Schael I, Velazquez FR, Abate H, Breuer T,

Clemens SC, Cheuvart B, Espinoza F, Gillard P, Innis BL et al: Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. The New England Journal of Medicine 2006, 354(1):11-22.

67. Vesikari T, Matson DO, Dennehy P, Van Damme P, Santosham M, Rodriguez Z, Dallas MJ, Heyse JF, Goveia MG, Black SB et al: Safety and efficacy of a pentavalent human-bovine (WC3) reassortant rotavirus vaccine. The New England journal of medicine 2006, 354(1):23-33.

68. Armstrong P: Rotaviral gastroenteritis in the NT: a description of the epidemiology 1995-2001 and future directions for research. Northern Territory Disease Control Bulletin 2001, 8(3):1-5.

69. Schultz R: Rotavirus gastroenteritis in the Northern Territory, 1995-2004. The Medical Journal of Australia 2006, 185(7):354-356.

70. Cortese MM, Parashar UD: Prevention of rotavirus gastroenteritis among infants and children: recommendations of the Advisory Committee on Immunization Practices (ACIP). Morbidity and Mortality Weekly Report 2009, 58(RR-2):1-25.

71. Carlin JB, Chondros P, Masendycz P, Bugg H, Bishop RF, Barnes GL: Rotavirus infection and rates of hospitalisation for acute gastroenteritis in young children in Australia, 1993-1996. The Medical Journal of Australia 1998, 169(5):252-256.

72. Therapeutic Goods Administration: Rotavirus vaccines statement, 18 May 2010. Department of Health, https://www.tga.gov.au/alert/rotavirus-vaccines-statement [last accessed 15/8/2017].

73. Hull BP, Mahajan D, Dey A, Menzies RI, McIntyre PB: Immunisation coverage annual report, 2008. Communicable Diseases Intelligence Quarterly Report 2010, 34(3):241-258.

74. Hull B, Dey A, Menzies R, McIntyre P: Annual immunisation coverage

report, 2010. Communicable Diseases Intelligence Quarterly Report 2013, 37(1):E21-39.

75. Hull BP, Dey A, Beard FH, Menzies RI, Brotherton JM, McIntyre PB: Immunisation coverage annual report, 2013. Communicable Diseases Intelligence Quarterly Report 2016, 40(1):E146-169.

76. Cente for Health Record Linkage (CHeReL): Data dictionaries, http://www.cherel.org.au/data-dictionaries#section1 [last accessed 5/10/2017].

77. Australian Consortium for Classification Development: The international statistical classification of diseases and related health problems, tenth revision, australian modification (ICD-10-AM/ACHI/ACS). 10th edn. Darlinghurst, NSW: Independent Hospital Pricing Authority; 2017.

78. SNOMED CT, https://www.nlm.nih.gov/healthit/snomedct/index.html [last accessed 6/11/2017].

79. Australian Bureau of Statistics 2004, Population projections, by age and sex, New South Wales, cat. no. 3222.0, viewed 17 September 2013 http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3222.02012%20(base)%20to%202101.

80. SAS Institute Inc., SAS Software. Cary, New Carolina: SAS Institute Inc; 2017.

81. Fisher RA: Applications of "Student's" Distribution. Metron 1925, 5:3-17.

82. Akaike H: A new look at the statistical model identification. IEEE Transactions on Automatic Control 1974, 19(6):716-723.

83. Molto Y, Cortes JE, De Oliveira LH, Mike A, Solis I, Suman O, Coronado L, Patel MM, Parashar UD, Cortese MM: Reduction of diarrhea-associated hospitalizations among children aged < 5 Years in Panama following the introduction of rotavirus vaccine. The Pediatric Infectious

Disease Journal 2011, 30(Suppl 1):S16-20.

84. Quintanar-Solares M, Yen C, Richardson V, Esparza-Aguilar M, Parashar UD, Patel MM: Impact of rotavirus vaccination on diarrhea-related hospitalizations among children < 5 years of age in Mexico. The Pediatric infectious disease journal 2011, 30(Suppl 1):S11-15.

85. Yen C, Armero Guardado JA, Alberto P, Rodriguez Araujo DS, Mena C, Cuellar E, Nolasco JB, De Oliveira LH, Pastor D, Tate JE et al: Decline in rotavirus hospitalizations and health care visits for childhood diarrhea following rotavirus vaccination in El Salvador. The Pediatric Infectious Disease Journal 2011, 30(Suppl 1):S6-s10.

86. Enane LA, Gastanaduy PA, Goldfarb DM, Pernica JM, Mokomane M, Moorad B, Masole L, Tate JE, Parashar UD, Steenhoff AP: Impact of Rotavirus Vaccination on Hospitalizations and Deaths From Childhood Gastroenteritis in Botswana. Clinical Infectious Diseases 2016, 62(Suppl 2):S168-174.

87. Hungerford D, Read JM, Cooke RP, Vivancos R, Iturriza-Gomara M, Allen DJ, French N, Cunliffe N: Early impact of rotavirus vaccination in a large paediatric hospital in the UK. The Journal of Hospital Infection 2016, 93(2):117-120.

88. De Oliveira LH, Giglio N, Ciapponi A, Garcia Marti S, Kuperman M, Sanwogou NJ, Ruiz-Matus C, Marinho de Sousa MF: Temporal trends in diarrhea-related hospitalizations and deaths in children under age 5 before and after the introduction of the rotavirus vaccine in four Latin American countries. Vaccine 2013, 31(Suppl 3):C99-108.

89. Wilson SE, Rosella LC, Wang J, Le Saux N, Crowcroft NS, Harris T, Bolotin S, Deeks SL: Population-level impact of Ontario's infant rotavirus immunization program: Evidence of direct and indirect effects. PloS One

2016, 11(5):e0154340.

90. Newall AT, Beutels P, Macartney K, Wood J, MacIntyre CR: The cost-effectiveness of rotavirus vaccination in Australia. Vaccine 2007, 25(52):8851-8860.

91. Pharmaceutical Benefits Advisory Committee: Public summary document: Human Rotavirus Vaccine, lyophilised powder and solvent for oral administration, 1 mL dose, Rotarix® November 2006. Department of Health and Ageing, Australian Government 2006, https://www.pbs.gov.au/ pbs/industry/listing/elements/pbac-meetings/psd/2006-11/pbac-psd-rotarix -nov06, [last accessed 20/9/2017].

92. Pharmaceutical Benefits Advisory Committee: Public summary document: Rotavirus vaccine, live, oral liquid, pentavalent, 2 mL unit dose RotaTeq® November 2006. Department of Health and Ageing, Australian Government 2006, http://www.pbs.gov.au/info/industry/listing/elements/ pbac-meetings/psd/2006-11/pbac-psd-rotavirus-nov06 [last accessed 20/9/2017].

93. Harris JP, Jit M, Cooper D, Edmunds WJ: Evaluating rotavirus vaccination in England and Wales. Part I. Estimating the burden of disease. Vaccine 2007, 25(20):3962-3970.

94. Hallgren KA: Conducting Simulation Studies in the R Programming Environment. Tutorials in Quantitative Methods for Psychology 2013, 9(2):43-60.

95. Burton A, Altman DG, Royston P, Holder RL: The design of simulation studies in medical statistics. Statistics in Medicine 2006, 25(24):4279-4292.

96. Winsberg E: Computer Simulations in Science. In: The Stanford Encyclopedia of Philosophy. Summer 2015 edn. Edited by Zalta EN.

Stanford University; 2015, https://plato.stanford.edu/archives/sum2015/entries/simulations-science [last accessed 23/10/2017].

97. Simonsen L, Taylor RJ, Young-Xu Y, Haber M, May L, Klugman KP: Impact of pneumococcal conjugate vaccination of infants on pneumonia and influenza hospitalization and mortality in all age groups in the United States. mBio 2011, 2(1):e00309-00310.

Appendices

Appendix A

Table A.1: *Systemized Nomenclature of Medicine concept identifiers [78] used to identify ED presentations for all cause gastroenteritis.*

102622004	186152009	236083006	300361008	4834000
105638003	186154005	2367005	300362001	48541000
105639006	186155006	23874000	300363006	4915001
105642000	186156007	23916006	300364000	49206006
105643005	186157003	23971007	300365004	49237006
105644004	186158008	240332005	300366003	49464009
105654000	186159000	240333000	300368002	49781004
105656003	186161009	240334006	300388003	49786009
10679007	186162002	240335007	30140009	50277001
109814008	186163007	240336008	301469002	50874004
11003002	186164001	240337004	301790009	51180003
111817006	186165000	240338009	302168000	51254007
111819009	186166004	240339001	302229004	51290000
111835002	186167008	240340004	302231008	51338006
111839008	186168003	240343002	302769004	52457000
111843007	186169006	240344008	302808000	53372009
111909004	186170007	240345009	30304000	53954009
111910009	186171006	240349003	30493003	54065007
111911008	18624000	240350003	30719001	54597004
111912001	186431008	240351004	307418008	55023005
111938001	18665000	240352006	307421005	55025003
111939009	187267007	240353001	307759003	55184003
11840006	187268002	240354007	308119005	55760004
12463005	187269005	240355008	308151006	55809007
12574004	187270006	240356009	308904008	56301005
128070006	187271005	240358005	3094009	56304002
128333008	187272003	240361006	309755003	56722001
128943002	187273008	240362004	312088007	57419008
129669003	187274002	240363009	312109000	57725006
135830006	187275001	240364003	312121001	58051000
14066009	187276000	240365002	312131008	58265007
14255005	187277009	240366001	312158001	58777003
14384003	187278004	240367005	31437008	58994008
14523006	187279007	240368000	32097002	59253004
1488000	187280005	240369008	32295003	61564002
15387003	187281009	240370009	32527003	61972007
15699003	187282002	240371008	32580004	62005008
15907009	187283007	240372001	332982000	62315008
161538007	187284001	240374000	33687004	63650001
162055004	18773000	240375004	33841007	63722008
162060000	18805001	240661004	34231006	63924002
162065005	18846006	240662006	34335000	64226004
162067002	192015007	240663001	34591007	64581007
167641001	19213003	242004	34923007	64613007
167821009	19547001	24412005	35408001	65154009
167822002	196738004	24526004	359613008	65328007
167823007	196743006	247340001	359651008	65979008

167826004	196746003	24789006	359660000	66107000
167827008	196747007	248123000	359662008	66160001
167828003	196749005	249493007	359671004	66218005
167830001	197010007	249497008	359804008	66301008
167833004	197011006	249498003	360425009	66329006
167835006	197013009	249499006	360427001	67375008
167838008	197014003	249500002	36188001	67915005
167840003	197017005	249501003	36529003	68270003
167841004	197018000	249519007	365442000	68676003
167842006	197019008	249964001	365660005	69776003
167843001	197020002	25319005	36789003	69794009
167845008	197021003	25374005	370518000	69810009
167846009	197024006	25458004	371323006	69980003
167847000	197026008	25497007	371423007	70014009
167848005	197027004	25552000	37224001	70880006
167849002	197028009	25887009	37377008	71085009
167851003	197030006	25898005	37693008	71207007
167852005	197032003	26006005	38205001	71419002
167854006	197034002	266071000	38685005	71572001
167855007	197035001	266072007	387754006	71583005
167856008	197037009	266073002	388759003	72245005
167857004	197038004	266074008	3899003	72547003
16932000	197039007	266075009	39224005	72681001
17551007	197040009	266076005	39341005	72965009
176401000000106	197041008	266077001	396337009	73034009
17960009	197522004	266078006	39747007	73335002
18081009	197533001	266079003	397503006	73817000
18168004	197536009	266080000	397609004	73879007
182182000	197546006	266081001	397683000	7399006
1824008	1985008	266082008	397696004	74361008
18338004	19894004	266083003	398384001	74621002
186088002	207110008	266449001	398570005	75119003
186089005	207111007	266451002	39963006	75333007
186090001	207112000	266453004	402962004	75375008
186091002	207113005	266454005	402964003	76118001
186092009	207114004	267051003	404811004	76623002
186093004	207116002	267060006	406559005	76816000
186094005	20759009	267061005	406623009	77070006
186101007	20958005	26826005	406624003	77524007
186102000	22207007	26838001	406625002	78128001
186103005	22231002	268651002	406626001	78300001
186104004	225586007	271352005	40663003	78420004
186105003	235224000	271353000	407373009	79099006
186107006	235653009	271354006	40835002	7977009
186108001	235654003	271355007	409506009	80774000
186109009	235705002	271356008	409587002	80805005
186111000	235706001	272044004	409966000	80960004
186112007	235707005	27335004	414155001	81020007
186113002	235744005	274080003	414286000	81159000
186116005	235745006	274081004	414529006	82047000
186117001	235746007	275297005	415353009	82051003
186119003	235747003	275371002	415709002	82930004
186121008	235748008	275744000	415822001	83134002
186122001	235750000	275745004	416482004	83852000
186123006	235751001	275746003	418450005	84480002
186124000	235752008	275747007	419219000	84622004
186125004	235753003	275748002	420687005	8579004
186127007	235754009	275786006	421224000	85904008
186128002	235755005	276288002	421429008	86138008
186129005	235756006	27858009	421454008	86500004

186134009	235757002	281102003	421983003	86615009
186135005	235765004	283876006	422400008	8765009
186136006	235791003	283877002	422587007	88905005
186137002	236061007	285344007	42338000	89030000
186138007	236062000	285396005	42359007	89861008
186139004	236063005	286870008	43240000	89933001
186140002	236064004	286974003	43752006	90271007
186141003	236065003	288199005	444673007	90974009
186142005	236066002	28980009	4556007	91173007
186143000	236071009	28986003	45994004	92827005
186146008	236072002	29120000	46609008	95544006
186147004	236073007	2919008	46799006	95545007
186148009	236075000	2946003	47812002	95897009
186149001	236076004	29979000	47941007	9689002
186150001	236077008	300359004	48252005	9814003
186151002	236082001	300360009		

Appendix B

Figure B.1: *Observed rates and estimated predicted rates, periodicity and underlying trend in weekly rates of NSW hospitalisations from the model shown in Table 4.2.*

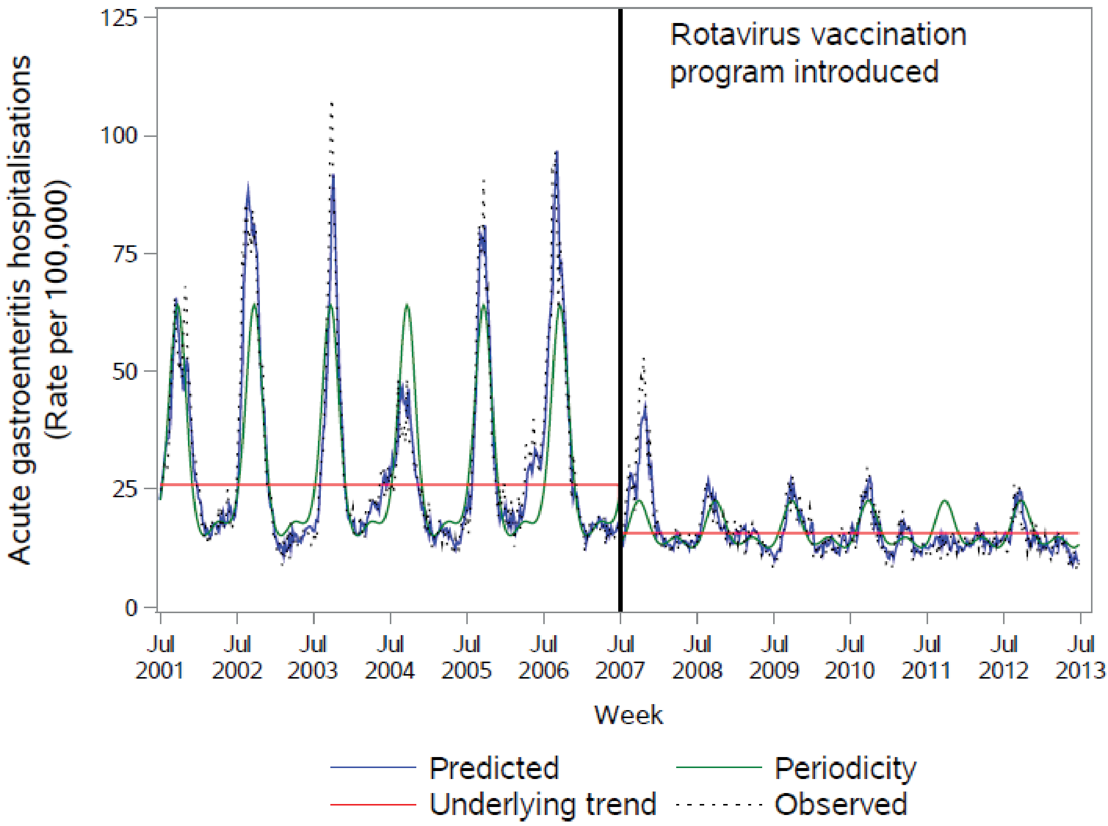


Figure B.2: *Residual normality diagnostics of the hospitalisations model shown in Table 4.2.*

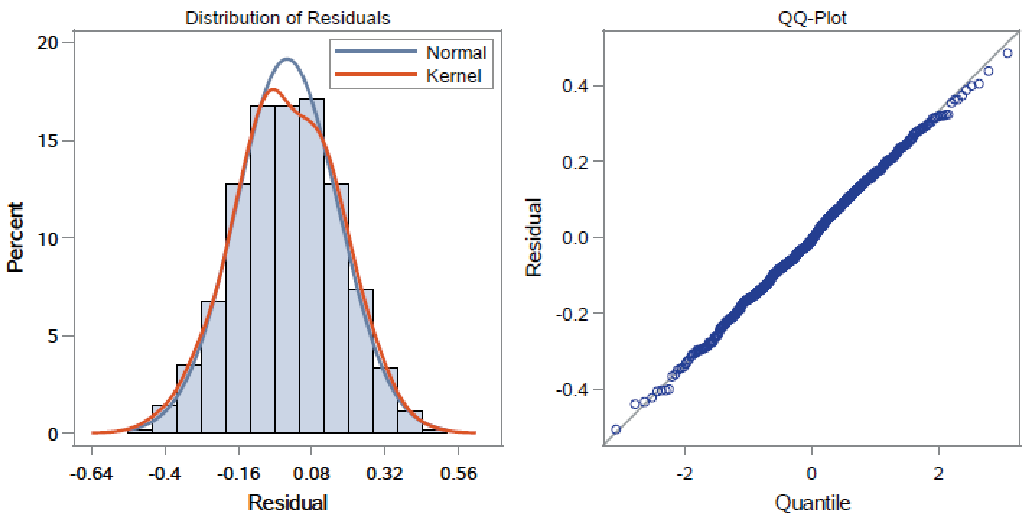
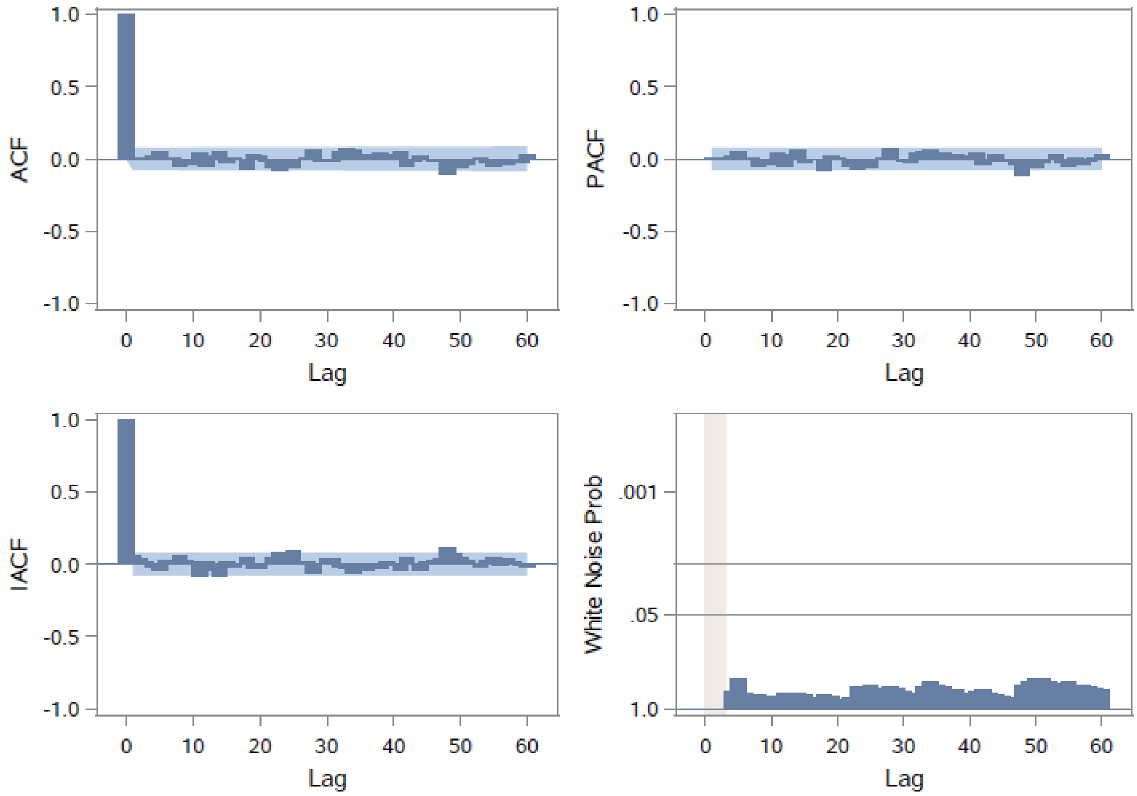


Figure B.3: *Residual correlation diagnostics of the hospitalisations model shown in Table 4.2.*



Appendix C

Figure C.1: *Observed rates and estimated predicted rates, periodicity and underlying trend in weekly rates of NSW ED presentations from the model shown in Table 4.2.*

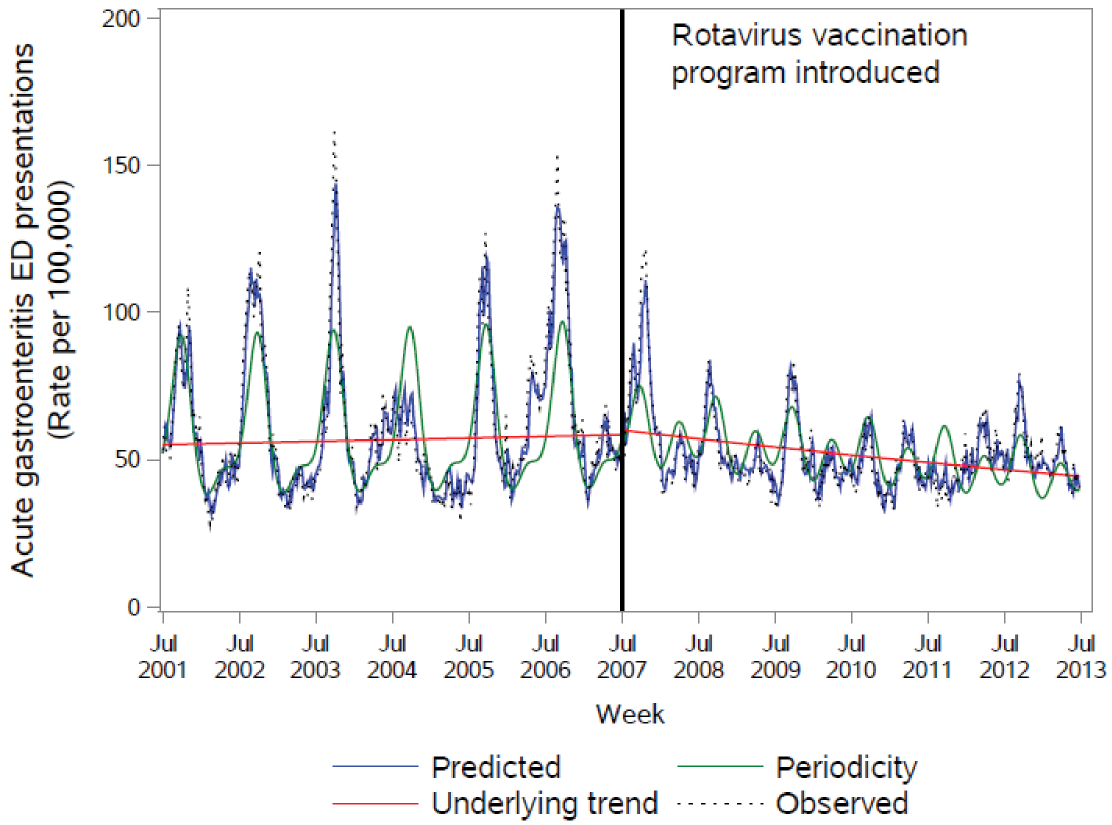
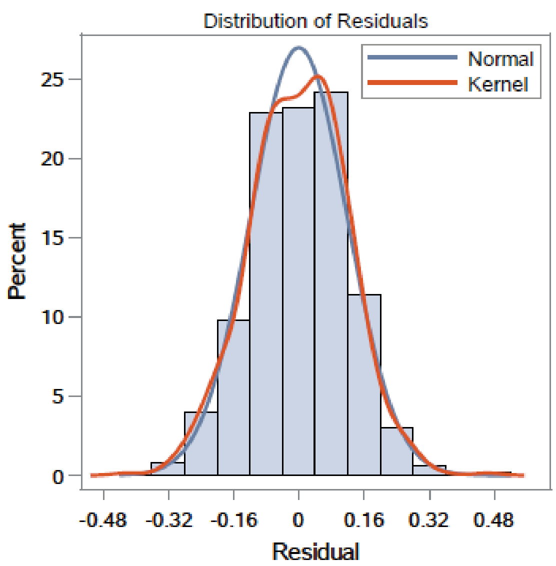


Figure C.2: *Residual normality diagnostics of the ED presentations model shown in Table 4.2.*



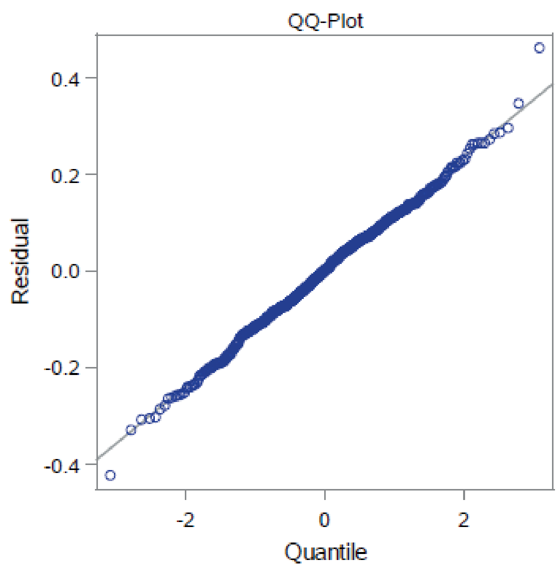
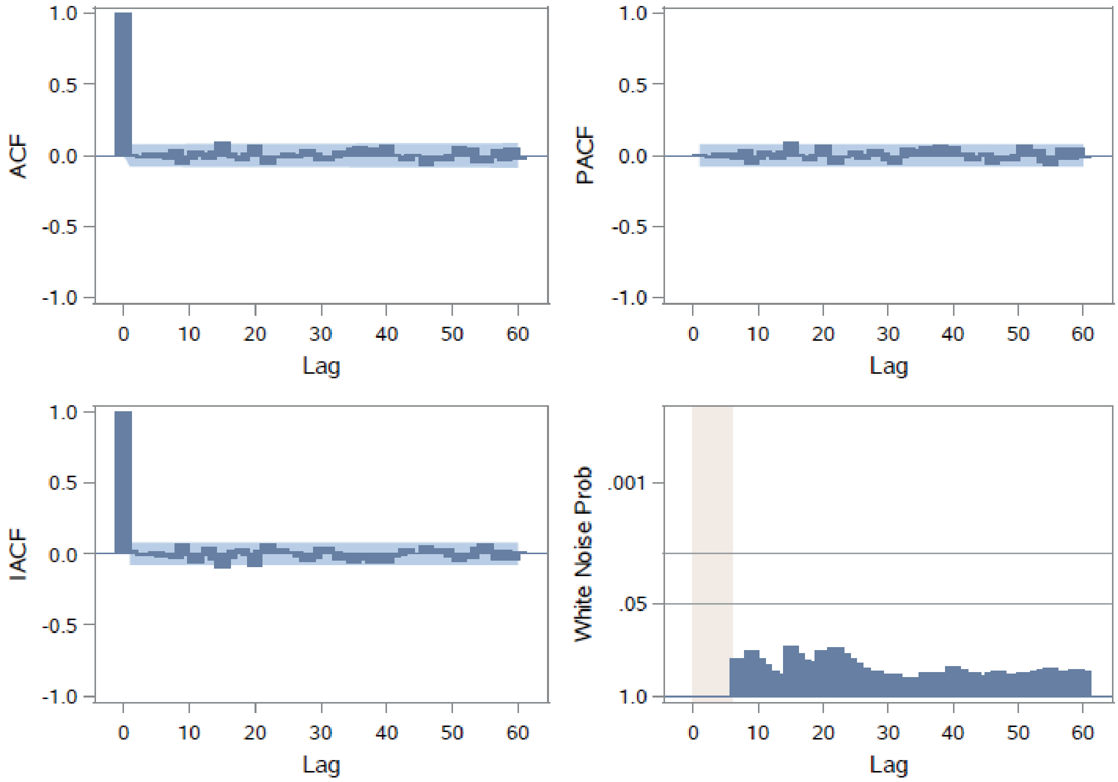


Figure C.3: *Residual correlation diagnostics of the ED presentations model shown in Table 4.2.*



Appendix D

Figure D.1: *Residual normality diagnostics of the hospitalisations forecasting model shown in Table 4.3 and Figures 4.9 and 4.10.*

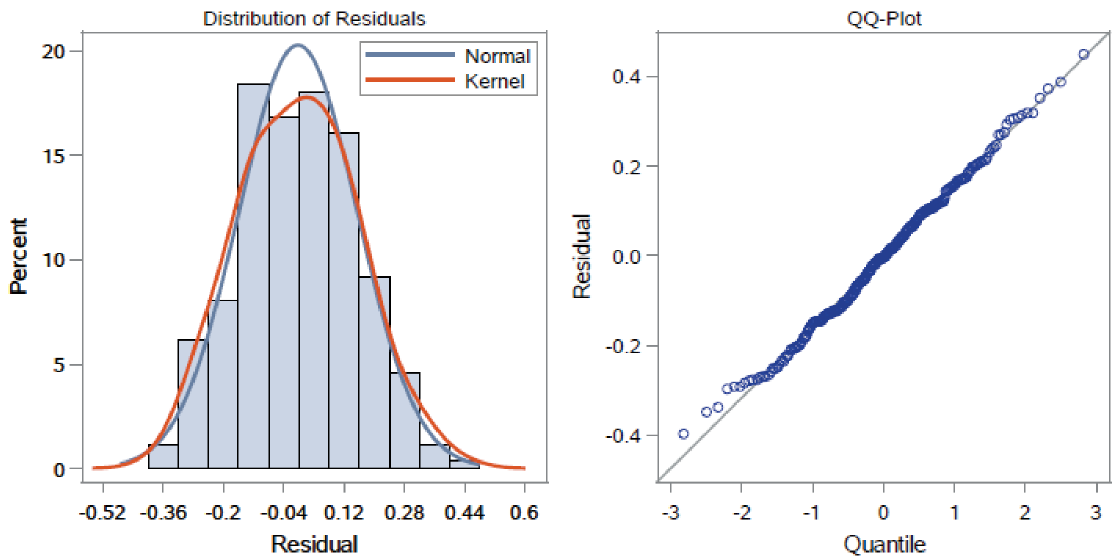
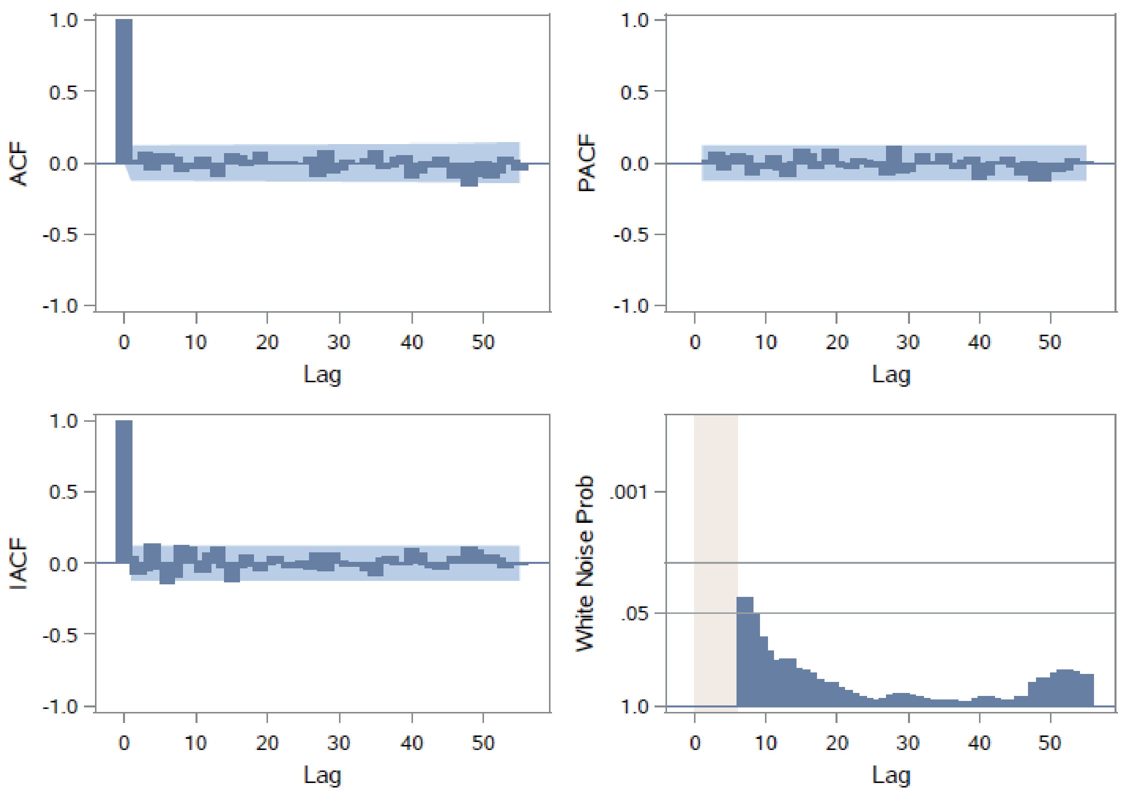


Figure D.2: *Residual correlation diagnostics of the hospitalisations forecasting model shown in Table 4.3 and Figures 4.9 and 4.10.*



Appendix E

Figure E.1: *Residual normality diagnostics of the ED presentations forecasting model shown in Table 4.4 and Figures 4.11 and 4.12.*

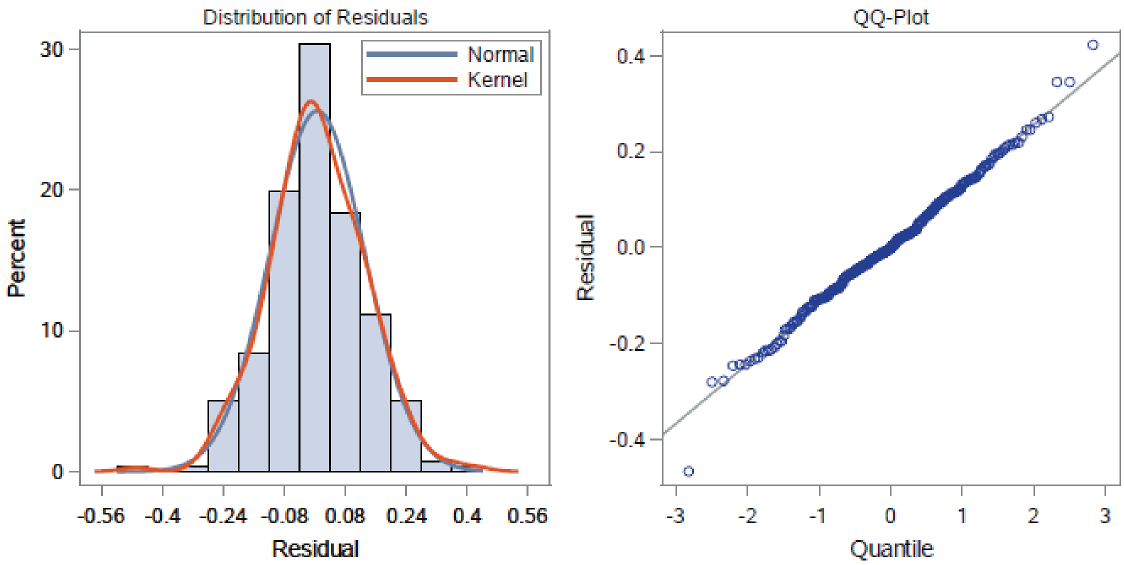
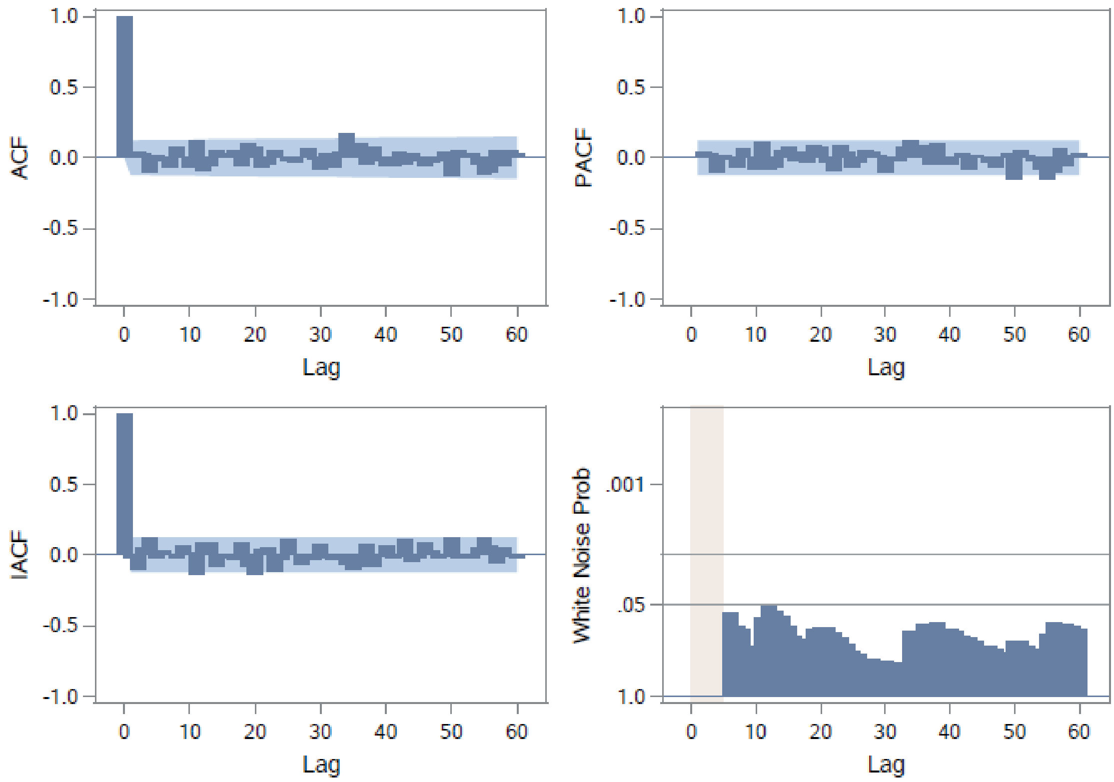


Figure E.2: *Residual correlation diagnostics of the ED presentations forecasting model shown in Table 4.4 and Figures 4.11 and 4.12.*



Appendix F

Table F.1: *Start seeds for the 1000 randomisations by scenario.*

Scenario	Start seed
1	31475
2	9847
3	23458
4	5823
5	8346
6	3286
7	6543
8	82374
9	205
10	3894
11	8271
12	4446
13	17
14	64646
15	4911
16	9823

Appendix G

Figure G.1: *Randomisation 15 from Scenario 1: ratio of weekly rotavirus cases per positive laboratory result is constant over the study period.*

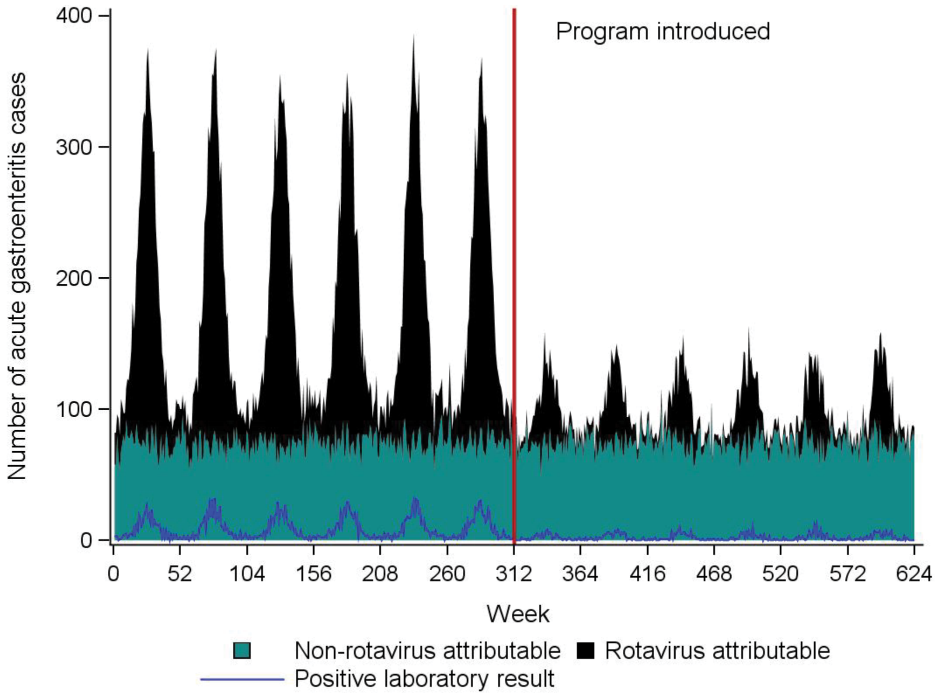


Figure G.2: *Randomisation 15 from Scenario 2: Ratio of rotavirus cases per positive laboratory result varies by program introduction (pre/post).*

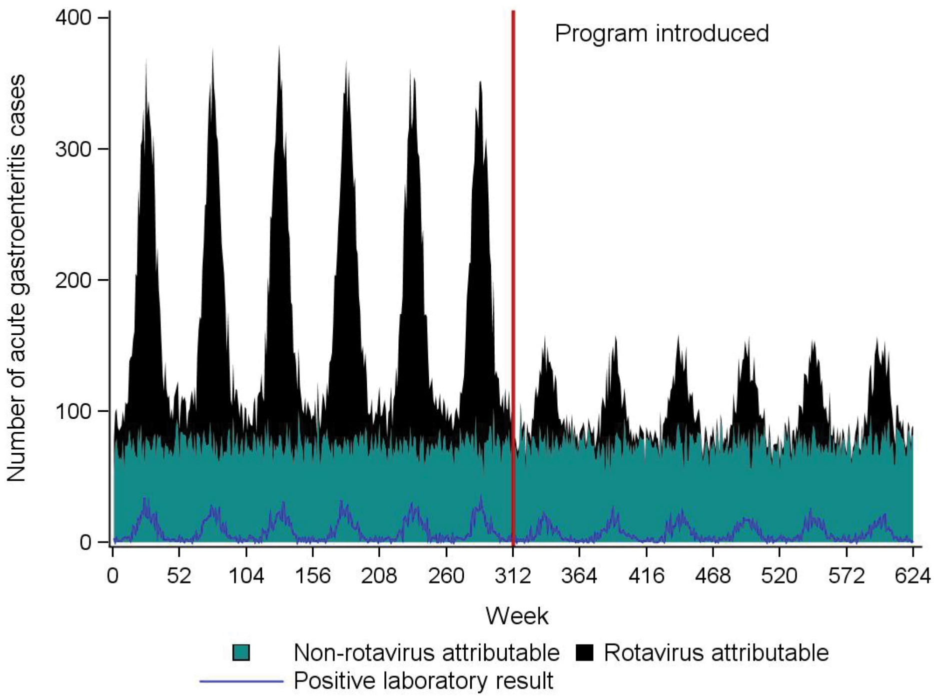


Figure G.3: *Randomisation 15 from Scenario 3: ratio of weekly rotavirus cases per positive laboratory result varies by season (peak/off-peak) and program introduction (pre/post).*

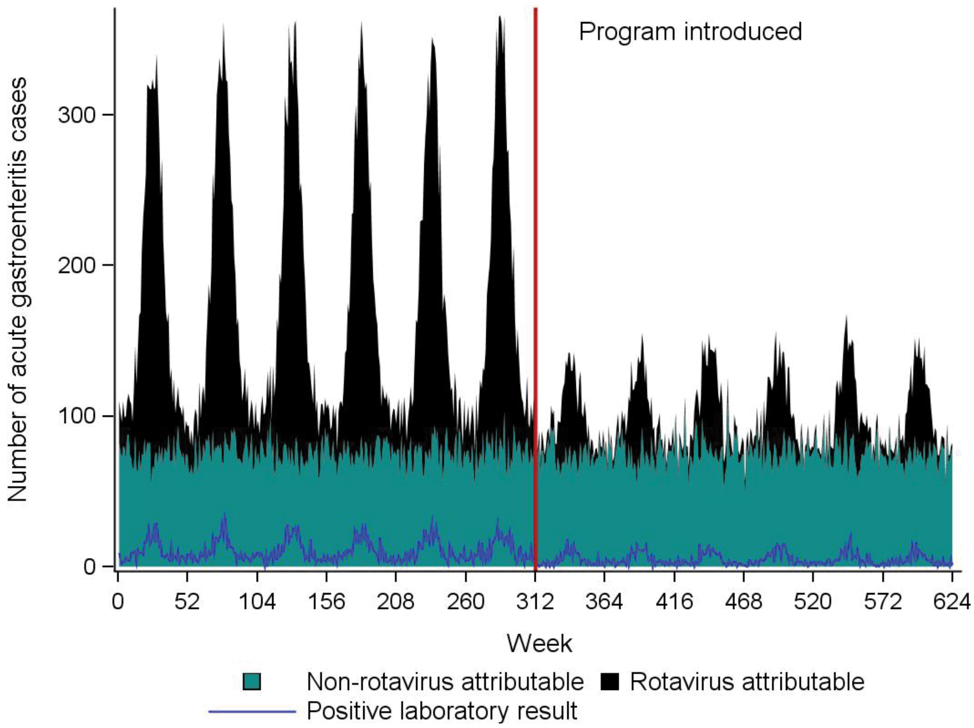


Figure G.4: *Randomisation 15 from Scenario 4: increase in non-seasonal rotavirus cases.*

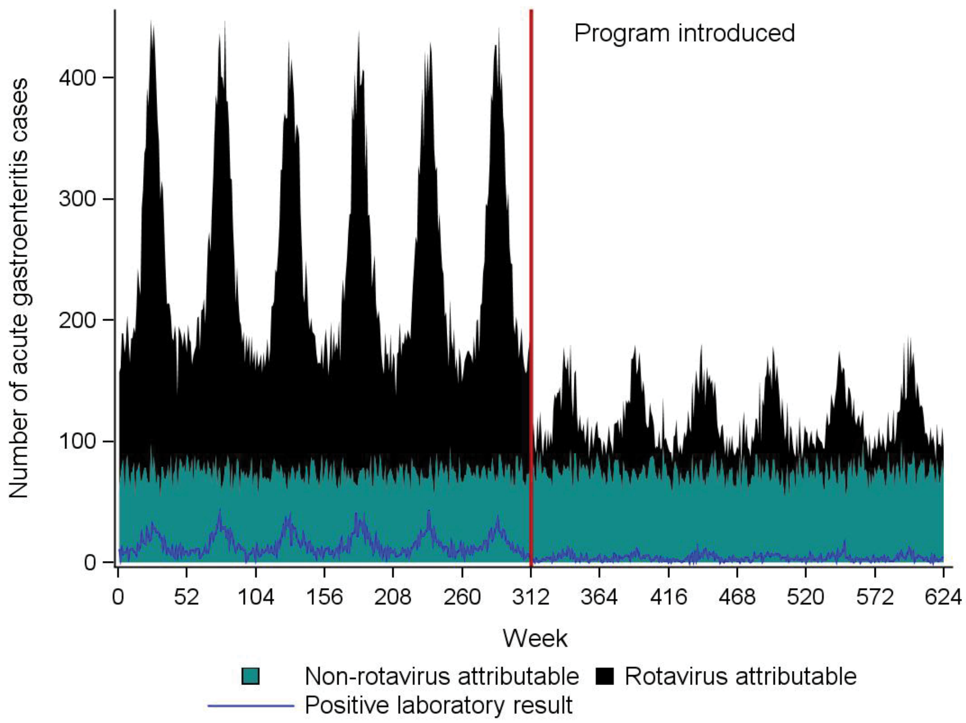


Figure G.5: *Randomisation 15 from Scenario 5: no non-seasonal rotavirus cases.*

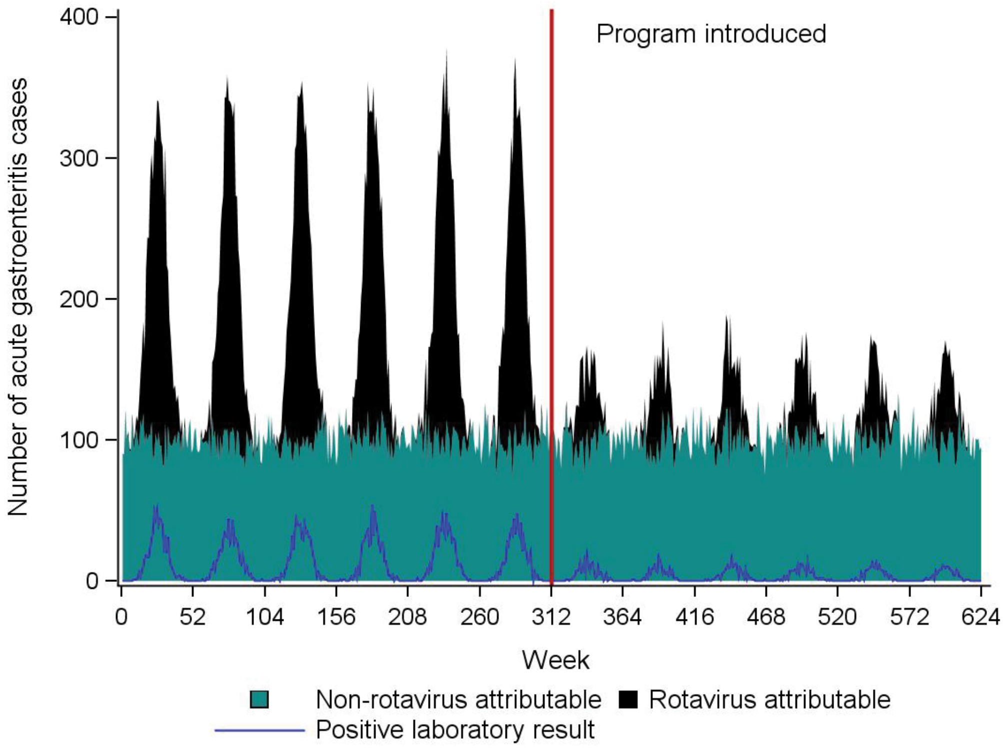


Figure G.6: *Randomisation 15 from Scenario 6: Scenario 5 with positive laboratory results lagged by one week.*

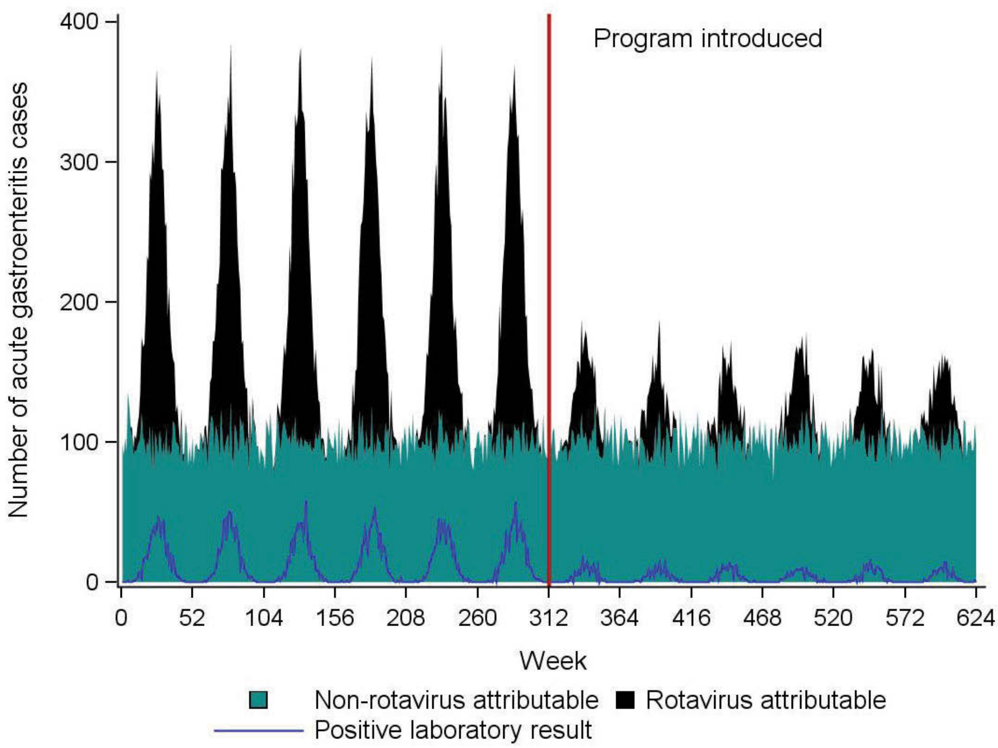


Figure G.7: *Randomisation 15 from Scenario 7: no relationship between weekly positive laboratory results and rotavirus cases (seasonal peaks and widths coincide, seasonal magnitudes vary).*

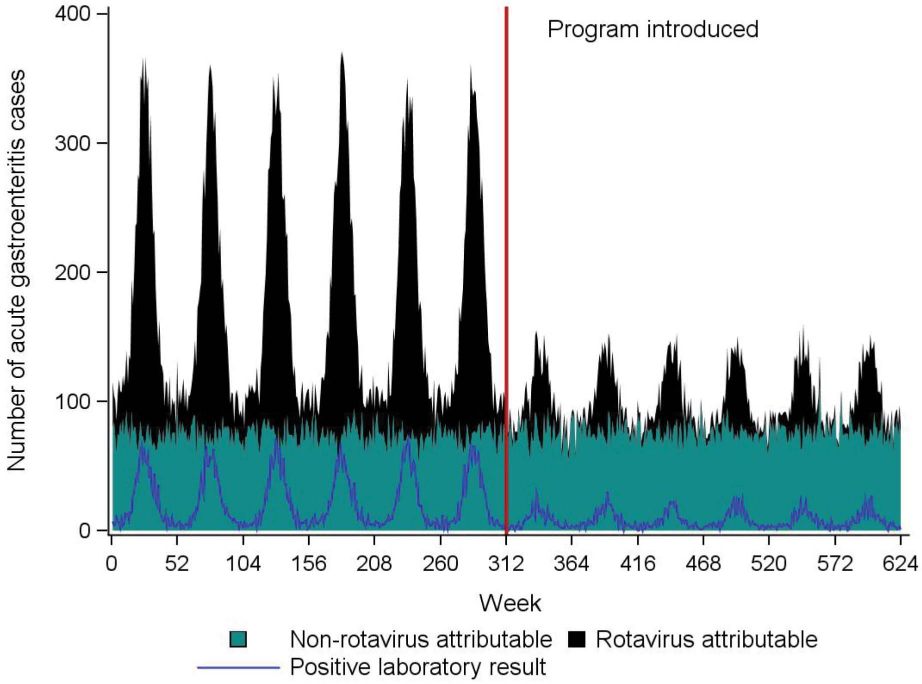


Figure G.8: *Randomisation 15 from Scenario 8: no relationship between weekly positive laboratory results and rotavirus cases (seasonal peaks coincide, seasonal widths and magnitudes vary).*

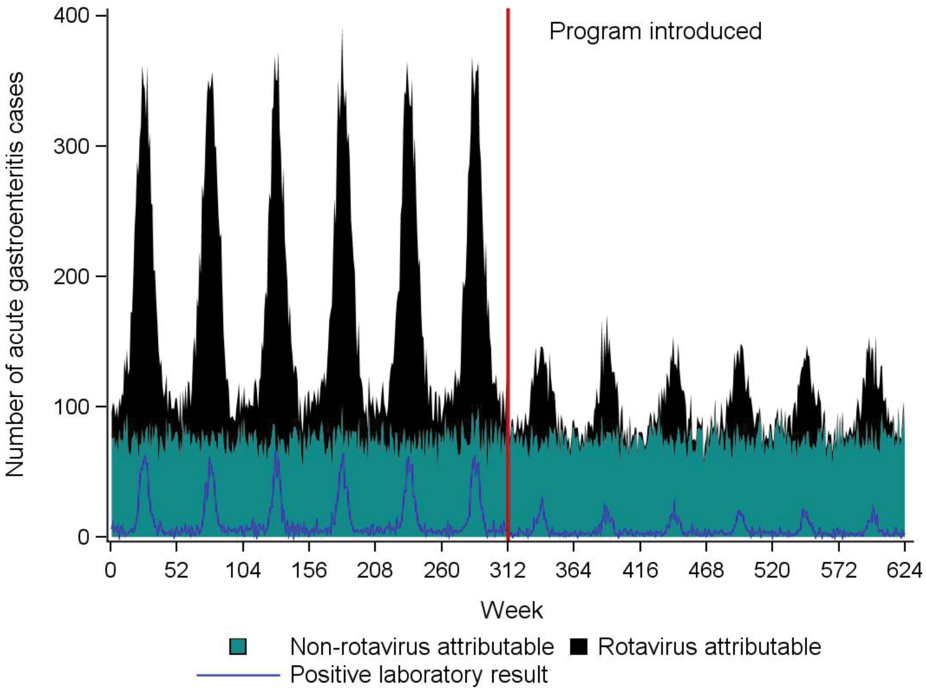


Figure G.9: *Randomisation 15 from Scenario 9: varying heights in peaks of rotavirus seasonality. Ratio of weekly rotavirus cases per positive laboratory result is constant over the study.*

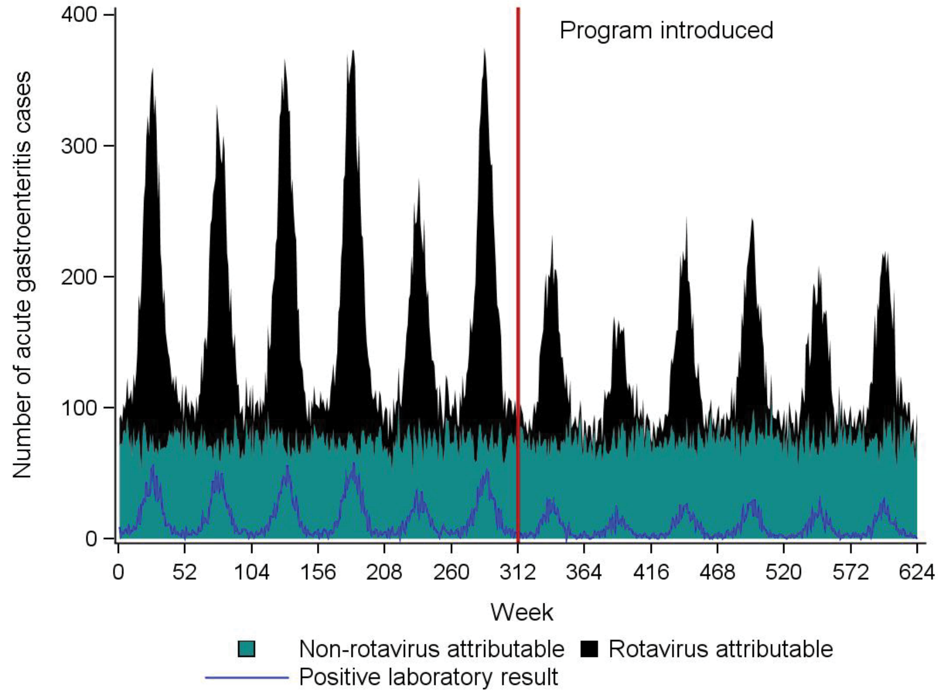


Figure G.10: *Randomisation 15 from Scenario 10: varying heights in peaks of rotavirus seasonality. Ratio of weekly rotavirus cases per positive laboratory result varies by year.*

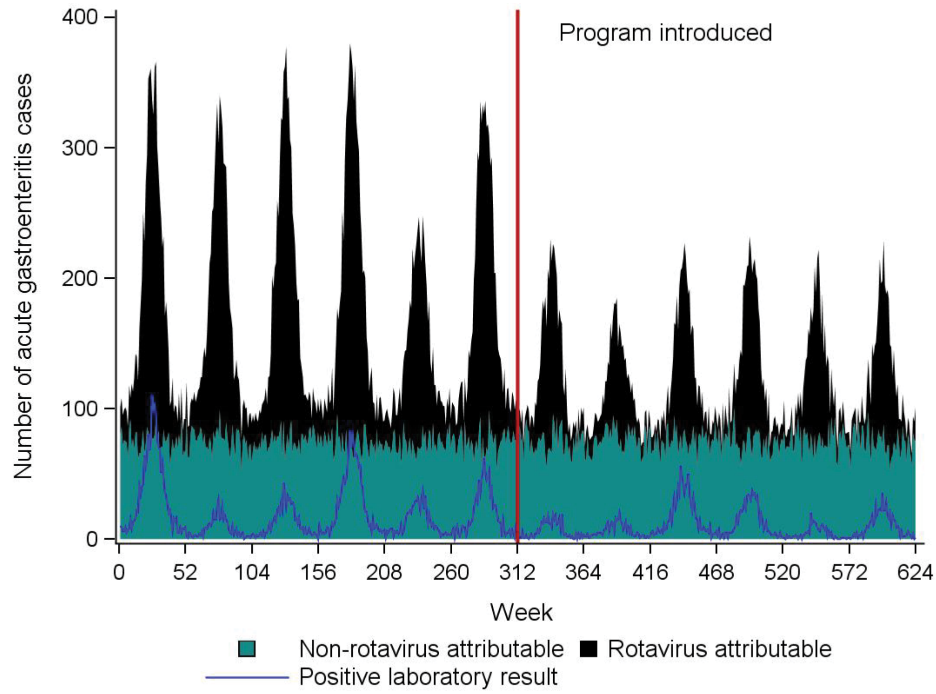


Figure G.11: *Randomisation 15 from Scenario 11: seasonality in non-rotavirus cases with the seasonal non-rotavirus peak coinciding with the seasonal rotavirus peak.*

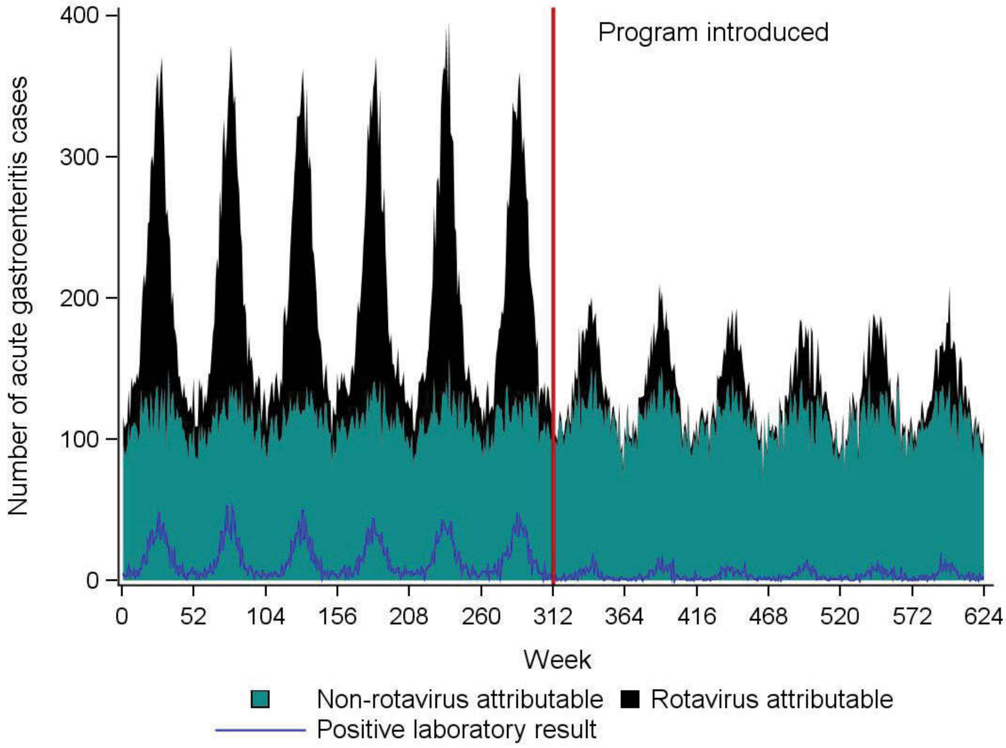


Figure G.12: *Randomisation 15 from Scenario 12: summer and winter seasonality in non-rotavirus cases.*

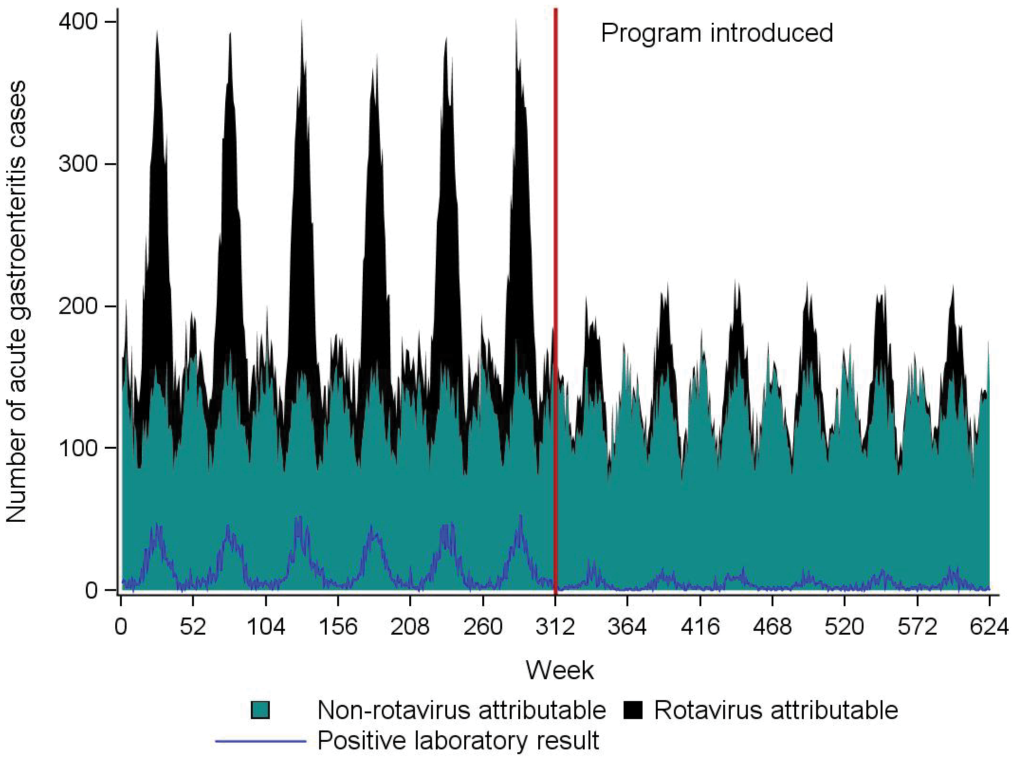


Figure G.13: *Randomisation 15 from Scenario 13: decreasing linear trend in rotavirus cases.*

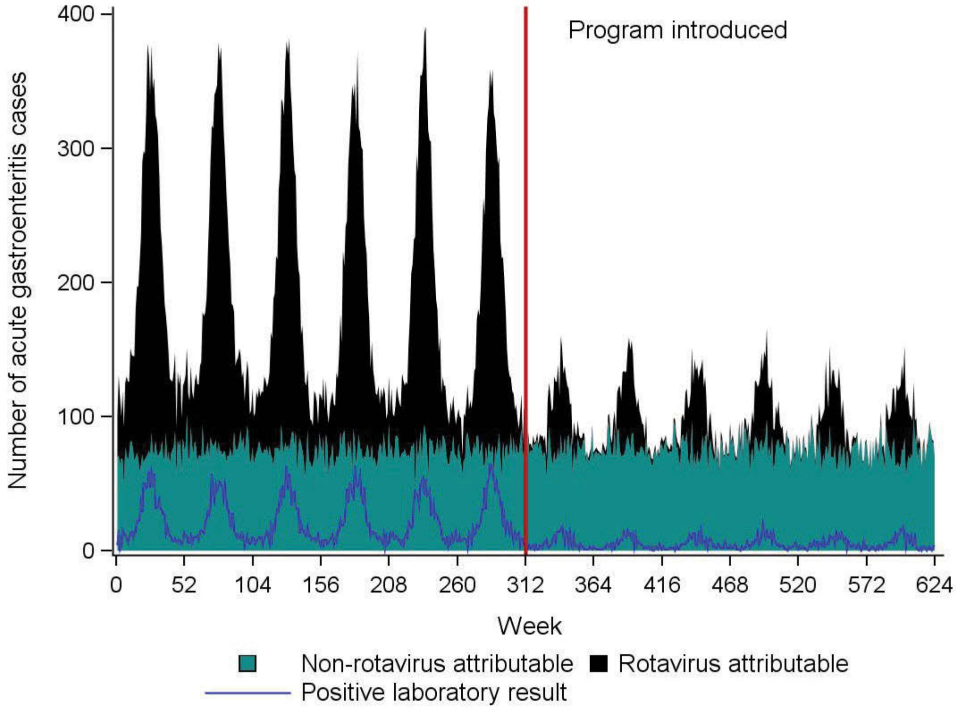


Figure G.14: *Randomisation 15 from Scenario 14: decreasing linear trend in peaks of seasonal rotavirus cases and no effect of vaccination efforts.*

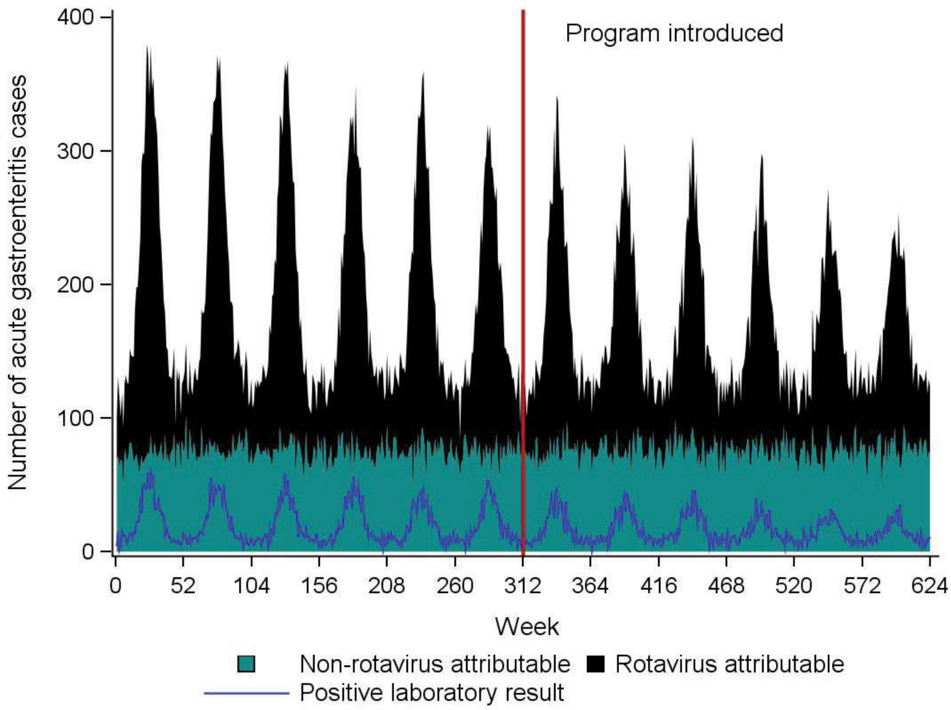


Figure G.15: *Randomisation 15 from Scenario 15: combination of Scenarios 10 and 11 with varying heights in peaks of non-rotavirus seasonality.*

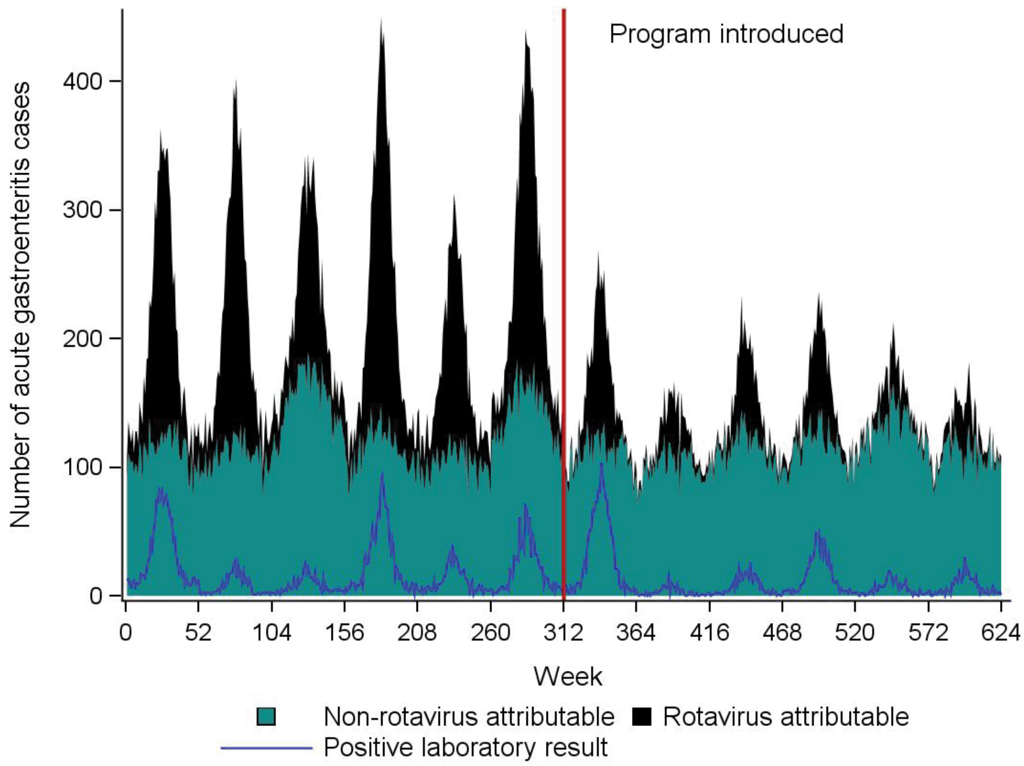


Figure G.16: *Randomisation 15 from Scenario 16: combination of Scenarios 10 and 12 with varying heights in peaks of non-rotavirus seasonality.*