UNIVERSITÀ DEGLI STUDI DI TORINO

DOCTORAL THESIS

Evaluating the impact of meningococcal vaccines with synthetic controls

Author:
Ottavia PRUNAS

Supervisor:
Lorenzo ARGANTE, PhD
Michele TIZZONI, PhD

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

PhD Programme in Complex Systems for Life Sciences
Doctoral School in Life and Health Sciences

February 15, 2021

Transparency Statement

I am a PhD student at the University of Turin, with a scholarship from GlaxoSmithKline Biologicals SA. Lorenzo Argante is an employee of the GSK group of companies. This work was sponsored by GlaxoSmithKline Biologicals SA.

Declaration of Authorship

I, Ottavia PRUNAS, declare that this thesis titled, "Evaluating the impact of meningococcal vaccines with synthetic controls" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

UNIVERSITÀ DEGLI STUDI DI TORINO

Abstract

Doctoral School in Life and Health Sciences

Doctor of Philosophy

Evaluating the impact of meningococcal vaccines with synthetic controls

by Ottavia PRUNAS

Evaluating the impact of vaccines is a critical epidemiological task, particularly challenging for vaccines against diseases presenting low and unpredictable incidence trends. Traditionally, the incidence of meningococcal cases before vaccination is used as a control (pre-post comparison); or, to correct for changes unrelated to vaccination, controls are meningococcal cases in non-vaccine-eligible age groups. However, this is justified only in the absence of indirect vaccine effects. And any control, if arbitrarily selected, has the potential to bias the impact estimate leading to inaccurate conclusions.

In my research project, I investigated the validity of the synthetic control approach, specifically in two infant immunization programmes against serogroup B and C meningococcal disease in England and Brazil, respectively. This data-driven approach uses information from other diseases to evaluate vaccine impact avoiding the potential bias resulting from arbitrary control selection. I used as controls several time series of infectious and non-infectious diseases in infants, and time series of MenB/MenC cases in older unvaccinated age groups.

Results show that the synthetic control method successfully adjusted for underlying trends, outperforming a before-after time trend model. Furthermore, meningococcal B and C cases in unvaccinated age groups were found to be the most predictive controls. However, similar performances were obtained when relying on other non-meningococcal diseases, specifically respiratory diseases and measles, supporting the use of the synthetic control method when indirect effects could not be neglected, and providing suggestive hypotheses of associations among infectious diseases that deserve further investigation.

In conclusion, the synthetic control model is a general and robust method that adjusts for unexplained trends and reduces the risk of bias associated with arbitrarily selected controls, and could be used even when indirect effects are present.

Acknowledgements

I would first like to thank my supervisors, Dr. Lorenzo Argante and Dr. Michele Tizzoni, for their support and precious help during my PhD journey.

I would like to acknowledge Dr. Duccio Medini and Prof. Michele Caselle for their critical feedback on my research project.

I would also like to thank Prof. Daniel Weinberger for his valuable suggestions and Dr. Kayoko Shioda for helping me out with many technical details about my research project.

Besides, I would like to thank my great colleagues Laura, Maddalena, and Simone for supporting me during these three years. Thanks also to my colleague Carlo for sharing with me all the bureaucratic issues of these years.

A warm thanks to all my colleagues from both ISI Foundation and GSK.

Finally, I could not have completed this thesis without the support of my family, my sisters, and of course Alberto.

Contents

List of Figures

List of Tables

List of Abbreviations

MenA	**M**eningococcal serogroup **A**
MenB	**M**eningococcal serogroup **B**
MenC	**M**eningococcal serogroup **C**
MenW	**M**eningococcal serogroup **W**
MenW	**M**eningococcal serogroup **Y**
IMD	**I**nvasive **M**eningococcal **D**isease
NIP	**N**ational **I**mmunization **P**rogramme
OMV	**O**uter **M**embrane **V**esicle
fHbp	factor **H**-binding protein
NadA	**Neisserial** adhesin **A**
NHBA	**N**eisseria **H**eparin-**B**inding **A**ntigen
4CMenB	**F**our-component **M**en**B** vaccine (commercial name **B**exsero)
MATS	**M**eningococcal **A**ntigen **T**yping **S**ystem
ICD-10	**I**nternational **C**lassification of **D**iseases, 10th revision
SINAN	O **S**istema de **I**nformação de **A**gravos de **N**otificação

Chapter 1

Introduction

Evaluating the impact of newly introduced vaccination programmes implies answering the critical question: what would have happened had the vaccine not been introduced?

Addressing this question requires an effective surveillance system providing long and reliable time series of hospitalization, mortality, morbidity, and coverage data. These observational data generally include information over large geographic areas and long time periods both pre- and post-vaccination. At the same time, however, these data can be influenced by other factors unrelated to the vaccination programme, such as changes in healthcare reporting or health in the population, which can either overestimate or underestimate the true effect of the vaccination programme.

Challenges increase when dealing with the real-world evaluation of meningococcal vaccines, because of the low incidence and unpredictable trends of the disease. Measuring the impact of a newly introduced mass immunization program against meningococcus is however of primary importance for public health.

Traditionally, vaccine impact is measured by comparing similar populations with and without a vaccination programme [1, 2]. Most commonly, the same population is compared, before and after vaccination, often using time-trend analyses [3, 4, 5]. This approach could fail to adjust for unexplained trends occurring during the vaccination programme, as it assumes that conditions are identical before and after immunization. Alternatively, external controls could be used to correct for changes unrelated to vaccination. Selecting adequate controls is however challenging and arbitrarily selecting them could introduce a source of bias.

The synthetic control method represents an appealing solution, where controls are not selected a priori. Starting from a large set of control diseases, Bayesian variable selection is employed to select the optimal set of controls based on the similarity with the disease of interest, here meningococcal disease, in the pre-vaccine period [6]. This approach has been successfully applied to evaluate the impact of public health interventions [7].

In my thesis, I investigated the validity of the synthetic control method for assessing the impact of meningococcal immunization programmes. I applied the synthetic control method to two different immunization programmes, against serogroup B and C meningococcal (MenB and MenC) disease in England and Brazil, respectively.

In chapters 2, 3, 4, 5, I provide the building blocks that I used to develop my analysis. In chapter 2, I go through the epidemiology, pathogenicity, and vaccinology of invasive meningococcal disease (IMD). After giving a brief description of the pathogen responsible for this infectious disease, i.e. *Neisseria meningitidis*, I expose the most relevant epidemiological features of IMD related to the spreading on the human population. Among these, I describe how the pathogen spreads, which are the subjects most at risk and its geographical and temporal patterns. Then, I review the vaccines developed to protect against IMD, in particular the MenB and MenC

vaccines. I conclude the chapter with an overview of the two meningococcal vaccination programmes, the MenC in Brazil and the MenB in England, that have been the subject of my doctoral research.

In chapter 3, I provide a general description of the different types of vaccine effects. I then focus on the evaluation of vaccine impact, discussing the challenges related to observational data and the traditional before-after designs used to evaluate this quantity. A before-after design could be inadequate and could mask the true impact when other events occur during the vaccination programme. In meningococcal vaccine impact studies, external controls are sometimes used to adjust for trends unrelated to vaccination, for example, IMD cases in non-vaccine-eligible age groups. These controls could however be affected by indirect effects.

The novelty of the synthetic control approach is that the most suitable controls are selected from a large set of controls based on the similarity with the target disease in the pre-intervention period. In chapter 4, I discuss the main statistical features of the synthetic control model, applied to time series data in a Bayesian setting. I describe the framework of state-space models that incorporate synthetic controls through linear regression, and I give an overview of prior specifications and posterior inference.

Chapter 5 deals with the synthetic control approach specifically developed to fit disease count data. Besides, the chapter provides the main steps followed to investigate the validity of synthetic controls with meningococcal disease. In particular, I describe: a negative control analysis with a comparison between synthetic control and time-trend models; vaccine impact estimates from the synthetic control in comparison with time-trend and models used in previous publications; the most influential meningococcal disease predictors; a sensitivity analysis; and alternative methods to overcome the limitations of the synthetic control model with noisy and sparse control time series.

These steps are put in practice in chapters 6 and 7, where my contribution to the field can be found. In these two chapters, I follow the procedure described in chapter 5 for two case studies of vaccination programmes in Brazil and England, respectively. In both countries, I tested the synthetic control using several infectious and non-infectious diseases and also MenB/MenC cases in non-vaccine-eligible age groups as controls. I also discuss indirect effects and different spatial and temporal aggregation of data to check whether it could improve vaccine impact estimates.

Chapter 8 consists of a final discussion where I highlight similarities and differences in the results among the two meningococcal programmes under study and I summarize the main findings of my research project.

Chapter 2

Meningococcal disease: from epidemiology to vaccinology

In this chapter, I describe the main features of meningococcal disease relevant to my doctoral research. The first section of the chapter deals with the biological and epidemiological features of meningococcal disease. The second section, instead, is an overview of the meningococcal disease vaccinology. I conclude the chapter providing relevant information about the two meningococcal vaccination programmes that were the object of my research: the meningococcal C campaign in Brazil and the meningococcal B campaign in England.

2.1 Biology and epidemiology of Neisseria Meningitidis

Neisseria meningitidis, also known as meningococcus, is an aerobic Gram-negative diplococcus. Humans are the only natural reservoir of *N. meningitidis*, and it usually colonizes the mucosa of the human upper respiratory tract of healthy individuals, normally not causing any symptoms before clearance [8, 9, 10]. *N. meningitidis* can eventually cross the nasopharyngeal epithelium and gain access to the bloodstream, causing life-threatening diseases such as meningitis and sepsis. Disease incidence is uncommon, 0.01–3.6 cases per 100,000 persons globally, but is fatal in 10–15% of cases even if treated with antibiotics, and up to 20% of survivors suffer severe sequelae, including brain damage, deafness, kidney failure, and limb amputation [8, 11]. Disease is a 'dead-end' step in the life cycle of this bacterium with no evolutionary benefit [12].

Meningococci are classified into serogroups, according to the immunologic reactivity of their capsular polysaccharides [8]. Despite not all the strains exhibit a capsule, the pathogenic strains are almost always encapsulated. Indeed, the capsule guarantees their survival in blood, and consequently, it enhances resistance to antibodies and phagocytosis [13]. Among the at least 13 different serogroups, only serogroups A, B, C, W, X, and Y account for the majority of invasive disease cases [8, 14, 15].

Meningococci are traditionally classified by serologic typing systems based on structural differences of capsule (serogroup), major outer membrane porin proteins (serotype), other outer membrane proteins (serosubtype), and lipooligosaccharide (immunotype)[8, 16]. Meningococcal isolates are clustered into clonal complexes with molecular typing methods [12, 17, 18]. Each clonal complex is composed of a group of strains that are supposed to share a common genetic origin [12]. A clonal complex includes several lineages of highly related isolates (clones) that are identical when tested by several typing methods [12].

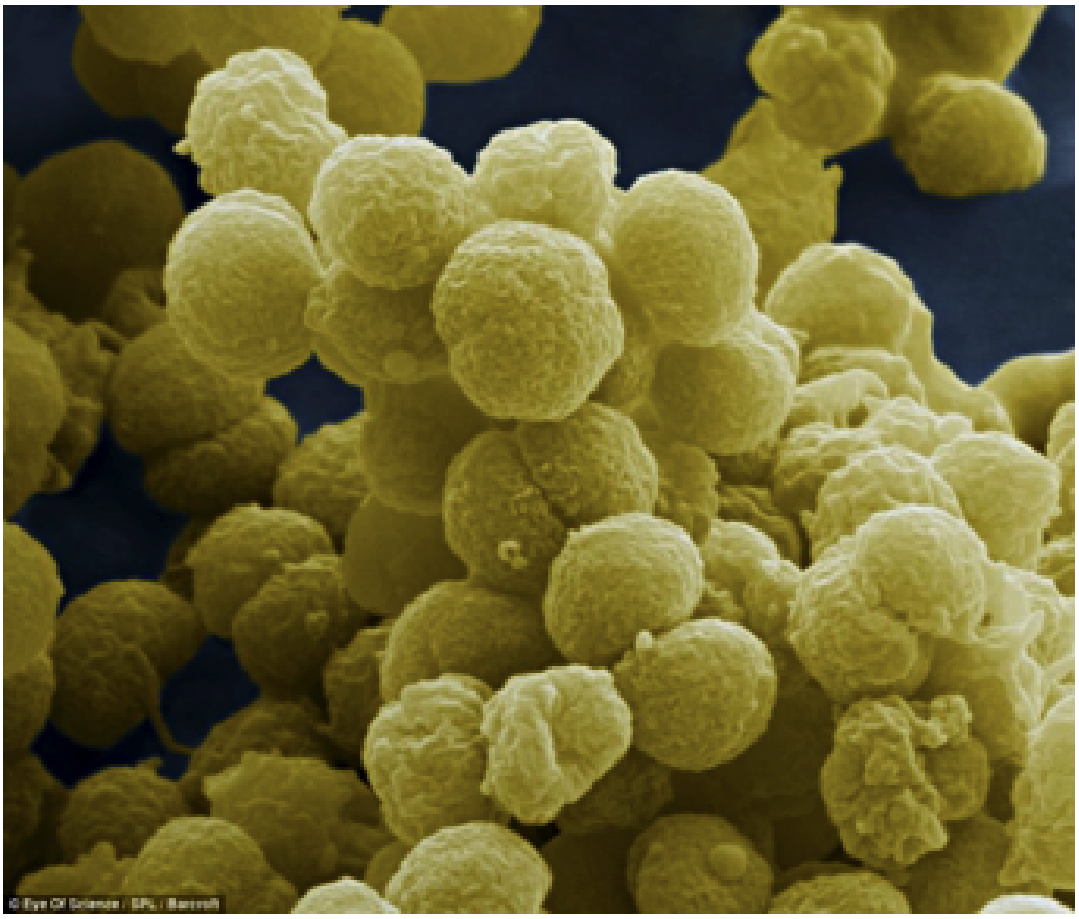

FIGURE 2.1: *Neisseria meningitidis*, gram-negative diplococci that cause meningococcal meningitis [19].

Meningococcal carriage strains are highly heterogeneous as shown by molecular typing [12, 20, 21]. Moreover, meningococci are also capable of exchanging the genetic material responsible for capsule production [8, 12]. They can switch from serogroup B to C or vice versa, and this capsule switching may become a threat to the wide-spread use of vaccines that provide serogroup-specific protection [8].

N. meningitidis lives exclusively in the human upper respiratory tract and the nasopharynx is the site from which meningococci are spread through saliva and respiratory secretions to other individuals [8]. Meningococci overcome host defenses and adhere selectively to non-ciliated columnar cells of the nasopharynx, starting to multiply and colonize the site of mucosal attachment [8]. This attachment is favored by their pili [22]. Besides, capsular polysaccharide, outer membrane proteins, and lipopolysaccharide may also contribute to the attachment of *N. meningitidis* to human cells [22]. In general, meningococcal acquisition is quite common with about 10% of the population being asymptomatic carriers of meningococcus [9, 23]; although rates are variable by age and setting [10]. Carriage prevalence varies with age, being low in infants and growing slowly up to approximately 10% in preadolescents [23, 24]. A sharp peak is then observed at the age of 20 with values around 25%. A subsequently decrease is observed in the elderly with percentages around 10% (Figure 2.2) [23, 24].

In exceptional cases, within 1–10 days after colonization, meningococci can escape their ecological niche, the oropharyngeal mucosa, cross the epithelium and enter the bloodstream, thereby causing invasive meningococcal disease (IMD), most

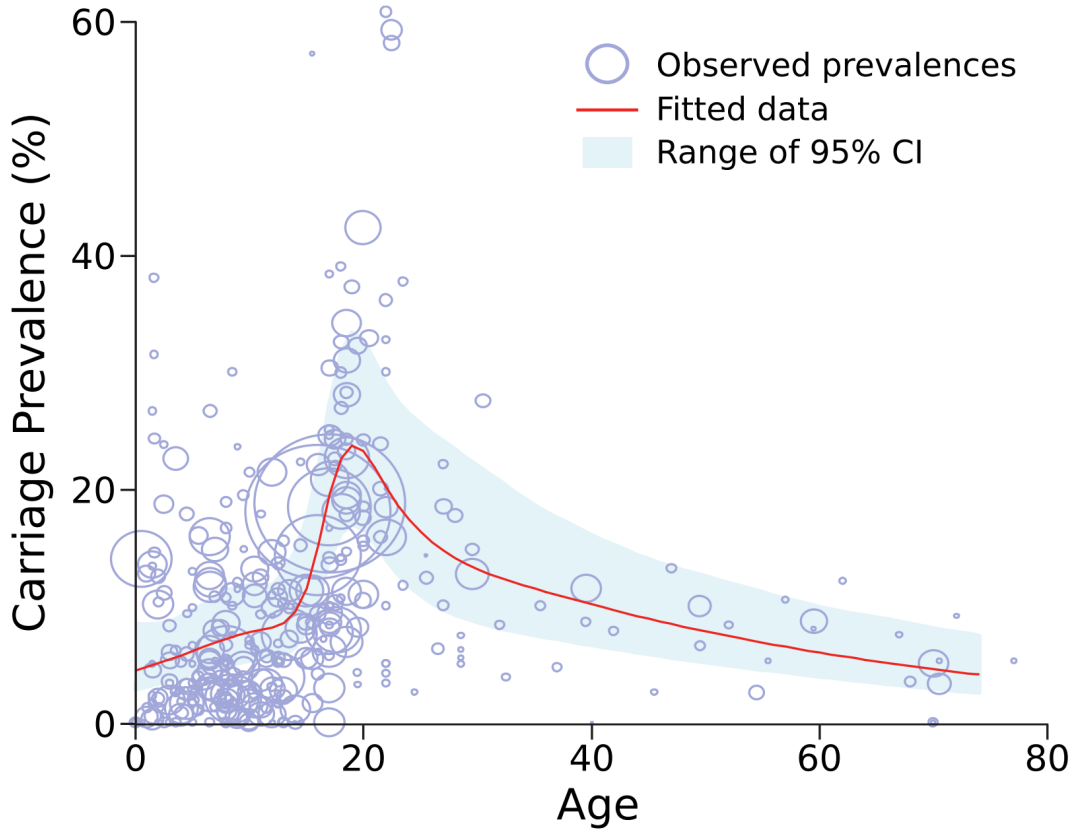

FIGURE 2.2: Meningococcal carriage prevalence plotted against age of individuals. These data were obtained aggregating and fitting data from several carriage studies in Europe (from Christensen et al. [24]).

commonly manifesting as meningitis or sepsis [14, 25]. This process involves upper respiratory epithelial invasion, endothelial cell damage, bloodstream and central nervous system invasion, trigging of an inflammatory cascade [16] (Figure 2.3).

All age groups are susceptible to IMD. The incidence of IMD is highest among children <1 year and adolescents/young adults [11]. Notably, the age distribution of IMD incidence is markedly different from that of asymptomatic carriage (Figures 2.2 and 2.4).

Surveillance data show that incidence and prevalence of the meningococcal serogroups causing invasive disease continually vary both geographically and temporally (Figure 2.5) [11, 26, 27]. MenB is endemic in many countries of the developed world. It is the leading cause of IMD in North America, South America, Australia, North Africa, and Europe, although a decreasing incidence trend is being observed [11, 28]. The incidence and prevalence of MenB naturally fluctuate over time and are currently at an all-time low; the reasons behind this decrease are unknown, among the hypothesis the introduction of a smoking ban in public places in some countries may have played a role [11].

Serogroup C is also common in the developed world, but the incidence of MenC disease has significantly decreased, following the introduction of effective conjugate vaccines against MenC [29]. Currently, it is reported as one of the most prevalent serogroups in Brazil [30], China [31], Russia [32], India, and Niger/Nigeria [33, 34].

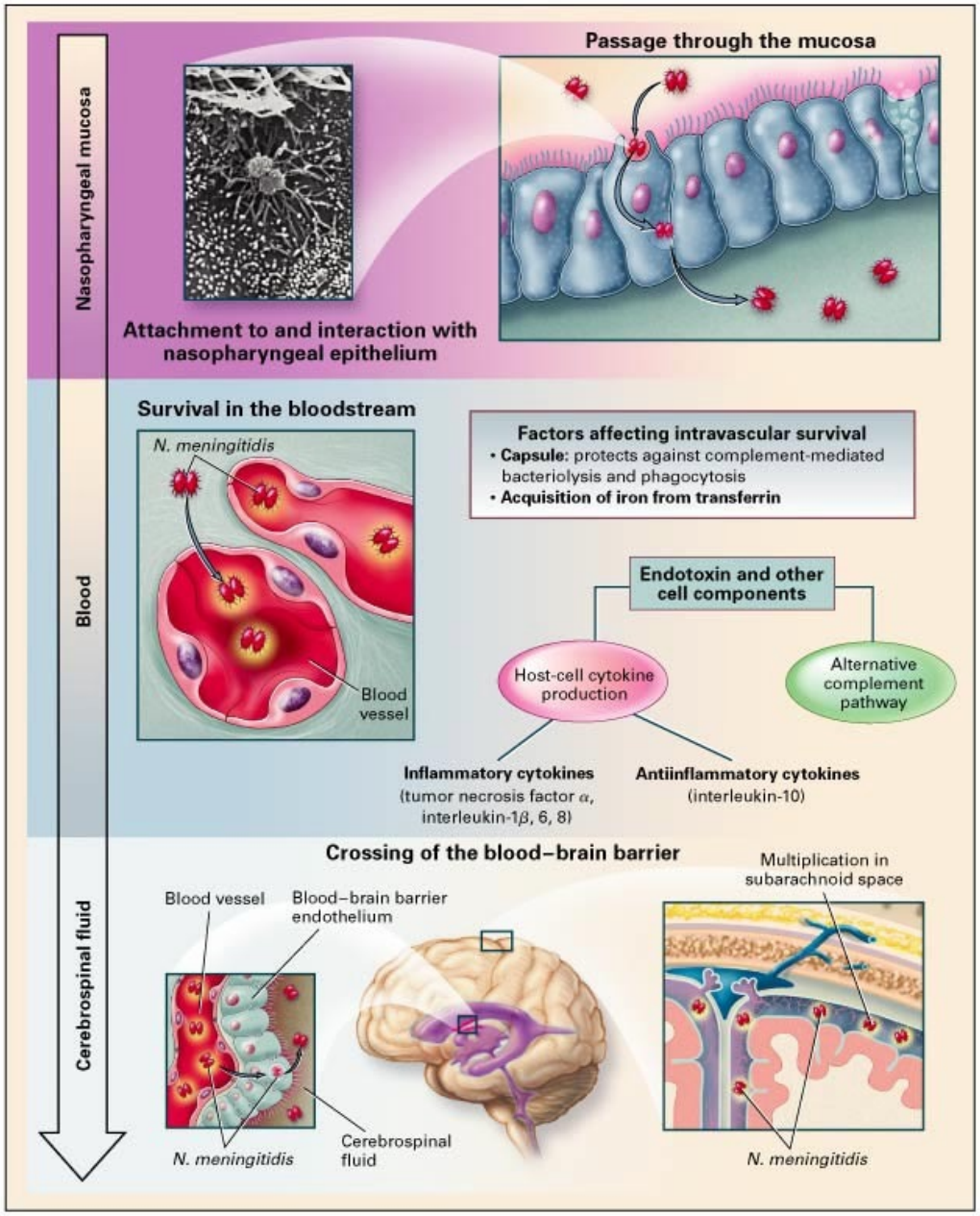

FIGURE 2.3: Colonization of *N. meningitidis* in the nasopharynx. After adhesion of *N. meningitidis* to the nasopharyngeal nonciliated epithelial cells, meningococci may be phagocytosed, thus pass through epithelium and eventually enter the bloodstream. Here they release endotoxins in the form of blebs. These outer-membrane vescicles can cause severe infections and normally they induce an immune-response. In case they are not killed by the immune-system, meningococci can reach the brain and cause meningitis. From Rosenstein *et al* [8].

Meningococcal serogroup A (MenA) historically accounted for 90% of meningococcal disease cases in the meningitis belt [35]. Following mass vaccination campaigns epidemics due to serogroup A have been eliminated [35].

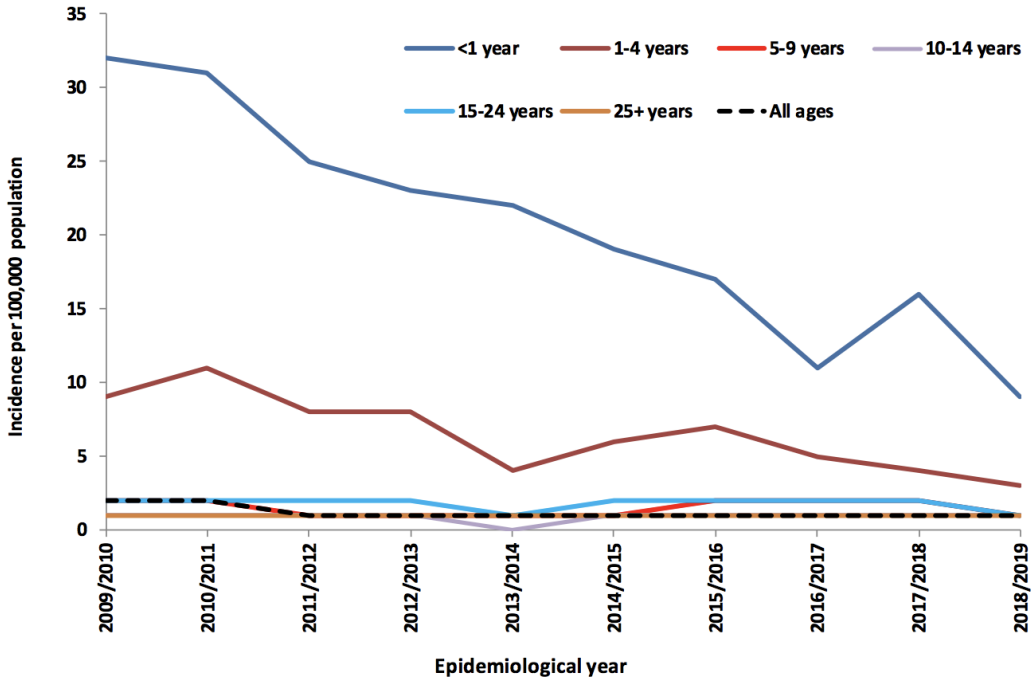

FIGURE 2.4: Incidence of invasive meningococcal disease in England: 2009/2010 to 2018/2019 [36].

Serogroup W (MenW) has historically caused sporadic cases, but after a large MenW outbreak associated with the Hajj pilgrimage in 2000-2002, other large epidemics have been recorded in Africa [37]. Since 2009 MenW cases have been increasing in England and Wales, motivating the introduction of the quadrivalent MenACWY conjugate vaccine in the UK's immunisation programme, with a catch-up campaign targeting 13-18 year-olds since August 2015 [38]. In Japan and Southern Africa (Mozambique) meningococcal serogroup Y (MenY) [39], and meningococcal serogroup W predominated [40], respectively.

Meningococcal disease incidence is strongly seasonal, with higher incidence during the winter in both hemispheres [42]. A typical example of periodical oscillations within a year period is the incidence of IMD caused by serogroups B and C in England and Wales (Figure 2.6). At time scales longer than 1 year, IMD is considered endemic, with sporadic emergence of single cases and occasional small outbreaks, often associated with the emergence of hyper-virulent clones. In 2003, a virulent serogroup B *N. meningitidis* sequence type 269 (ST269) clone emerged in the province of Quebec, Canada [43, 44], which caused a prolonged increase in the incidence of serogroup B invasive meningococcal disease [44].

Moreover, there are slow but consistent fluctuations in incidence as part of secular changes, with hyper-endemic periods that can last for decades. In the African meningitis belt, several cyclical large epidemics have been reported, mostly caused by serogroup A [45].

The mechanism driving these fluctuations remains poorly understood [47, 48].

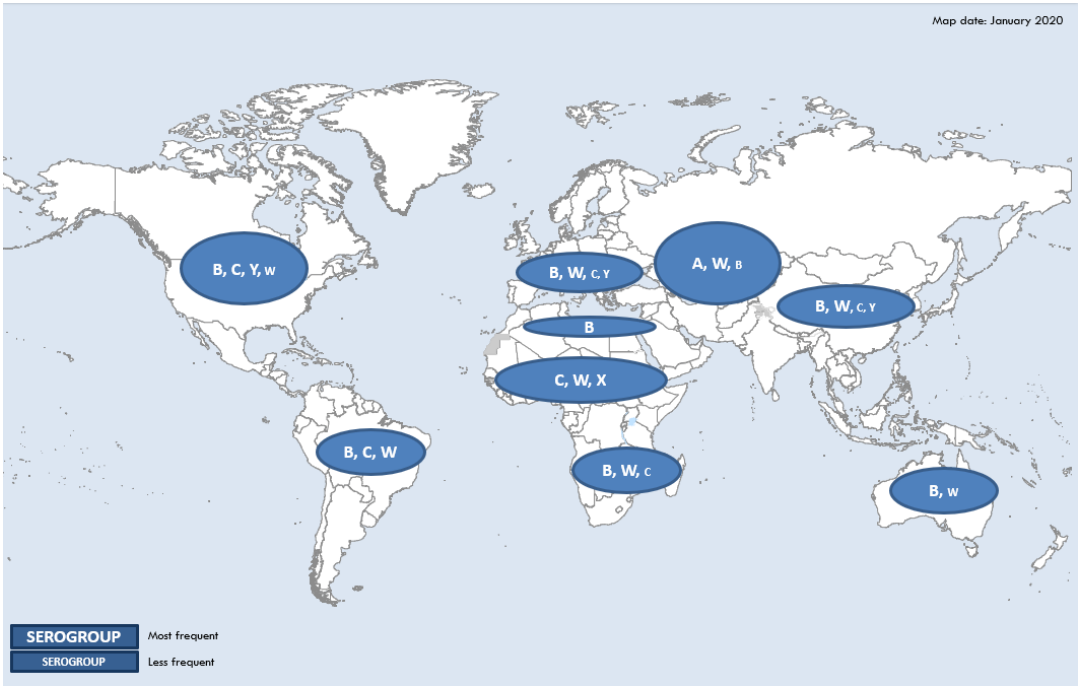

FIGURE 2.5: Serogroup distribution world-wide of Invasive Meningococcal Disease in 2019 [41].

It has been observed that IMD peaks frequently coincide with or closely follow increases in the incidence of influenza and other respiratory virus infections [49, 50, 51, 52], with elevated IMD risk when influenza activity is "epidemic" [47, 53]. This observation could represent a causal link between IMD and respiratory infections [47], where the causal relationship could be that influenza weakens human immune defenses [49]. It has also been shown that influenza can increase susceptibility to IMD directly [54] or indirectly, facilitating meningococcal colonization [55, 56]. Besides, this relationship could also represent shared susceptibility to seasonal environmental exposures or seasonal changes in human behavior (e.g., clustering together of individuals indoors due to cold winter weather, holiday gatherings, etc.) [47].

2.2 Meningococcal vaccines

Vaccination is the only effective prevention measure against IMD [8, 11, 15]. There are different global prevention strategies. Vaccination is implemented either through National Immunisation Programmes (NIPs), or vaccinating only high-risk populations (e.g. conjugate MenACWY in India) or for outbreak control only (e.g. MenB vaccine in Canada; polysaccharide MenA and MenAC vaccines in Russia; polysaccharide vaccines in the African meningitis belt) [11]. To maximize coverage, it is suggested to include vaccination via NIPs; this decision is usually determined by cost-effectiveness analyses [57].

There are three types of vaccines: 1) polysaccharide; 2) conjugate; and 3) protein. In brief, polysaccharide vaccines are composed of pure bacterial cell wall polysaccharide, whereas in conjugate vaccines the capsular polysaccharide is conjugated with a protein carrier (e.g., diphtheria or tetanus toxoids) [11]. Compared to simple polysaccharides vaccines, the conjugation with a protein carrier elicits a longer

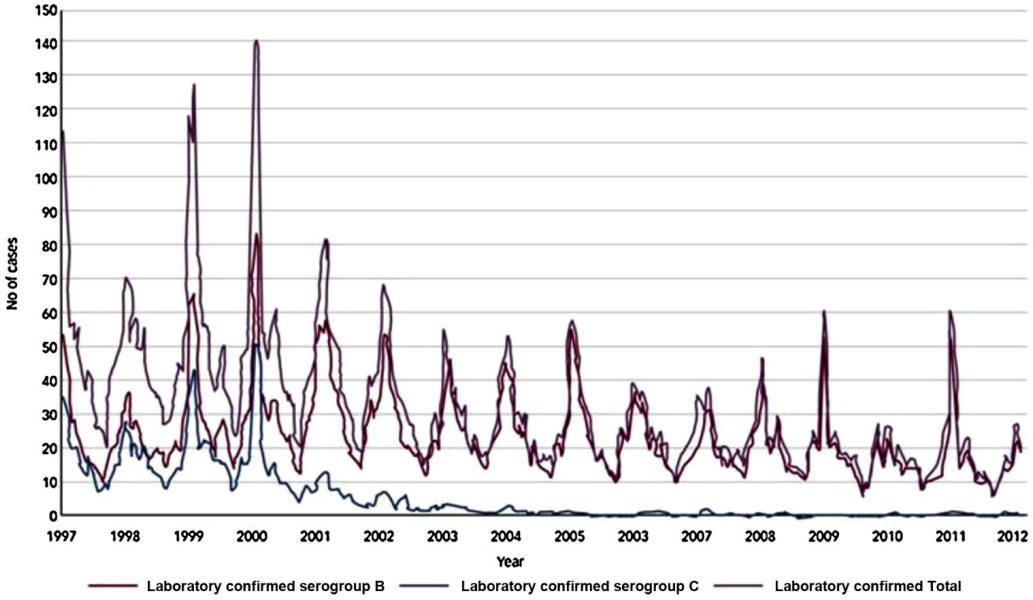

FIGURE 2.6: Laboratory confirmed IMD cases in England and Wales [46].

protection [58] and mucosal immunity. The latter prevents meningococcal colonization and carriage acquisition, thus potentially inducing indirect protection of non-vaccinated individuals at the population level if the vaccine coverage is high enough [11].

Conjugate vaccines have proved to be successful in preventing meningococcal disease worldwide. The MenC conjugate vaccine has been implemented into the NIP in the UK in 1998 [59], significantly reducing both the incidence and carriage of MenC [60]. Another successful example is the introduction of a monovalent MenA conjugate vaccine in Africa, which reduced invasive disease and carrier rates by inducing direct and indirect (herd) protection, respectively [61, 62, 63].

If the use of a polysaccharide or conjugate vaccine is not possible, protein-based vaccines are implemented, which include a particular protein of the pathogen [11]. The most relevant example is MenB, where a protein-polysaccharide conjugate vaccine against serogroup B does not exist. Indeed, the capsular polysaccharide has the same composition of sialic acids found on human tissues, in particular in neuronal cells [64], and thus such vaccines would be poorly immunogenic and could potentially induce an autoimmune response [65].

The first approach that has been adopted to develop MenB vaccines was the use of outer membrane vesicles (OMV). The OMVs contain several molecules, of which the PorA protein (a porin) is considered responsible for generating immunogenicity [67, 68]. OMV-based vaccines have been successfully applied to contain local outbreaks caused by meningococci as in the 1980s in Norway, Cuba, Brazil, France, or in the 2004-2009 epidemic in New Zealand [69].

Following the publication of the first meningococcal genome, reverse vaccinology was used to develop the first vaccine broadly covering for MenB disease [11, 70]. This bottom-up genomic approach was used to deduct the putative bacterial surface-exposed proteins, i.e. the antigens that could be candidates to compose the vaccine. Starting from the complete genome sequence of a virulent strain of serogroup B *N.*

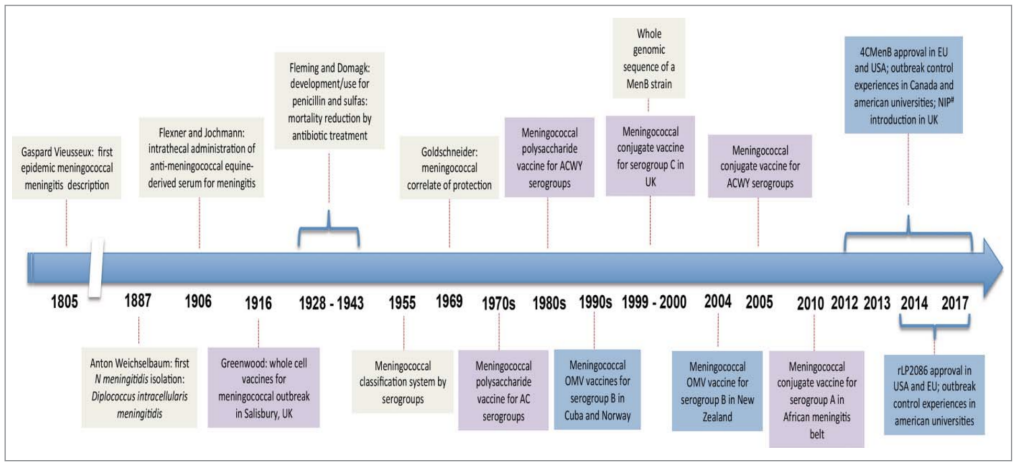

FIGURE 2.7: Timeline of some relevant achiements in meningococcal disease control and prevention. From Villena *et al* [66].

meningitidis, about 600 putative surface-exposed proteins have been predicted using bioinformatics [71]. Selection of the vaccine candidates continued through various steps that consisted of cloning and expression in E. coli, immunization of mice, confirmation of surface exposure, and bactericidal activity. Finally three proteins were selected as antigens for vaccine development [71]: (i) factor H-binding protein (fHbp); (ii) Neisserial adhesin A (NadA), and (iii) Neisseria Heparin-Binding Antigen (NHBA). In addition, the vaccine includes the OMV expressing PorA from the New Zealand strain, PorA P1.4 [70, 71, 72].

Two broadly protective MenB vaccines have been developed using this technique, Bexsero (GSK) and Trumenba (Pfizer), both licensed in the USA for persons between 10 and 25 years of age [73, 74]. The 4CMenB vaccine (Bexsero) was also licensed in Europe, Canada, Australia, and elsewhere for individuals from 2 months of age and older [75, 76, 77]. Besides, it has been introduced in the NIP started in the UK in September 2015.

Figure 2.7 summarizes the relevant achievements in meningococcal disease control and prevention [66].

2.2.1 Meningococcal C campaign in Brazil

Meningococcal disease is endemic in Brazil and sporadic outbreaks have been recorded [78], as for the epidemic of serogroup C disease occurred in the Salvador municipality in 2010 [79]. In Brazil, notification rates of IMD caused by *N. meningitidis* as high as 7.0/100,000 inhabitants were reported for children under two years old during 2009 and 2010 [79, 80]. Between 2002 and 2005, a significant shift from serogroup B, which used to be the most frequent isolate in invasive meningococcal cases (67%), to serogroup C was observed. Thereafter, the circulation of serogroup C increased [78, 79].

The increasing incidence of serogroup C, coupled with recurring outbreaks in different regions [81, 82, 83], prompted to the introduction of meningococcal C conjugate vaccine in the NIP in November 2010 [78, 79]. Brazil was the first country in Latin America to introduce this vaccine in its NIP. The recommended vaccination schedule consisted of priming doses at three and five months, with a booster dose at 12-15 months of age [78, 79, 84].

Brazil has 27 states which belong to the following five regions: North (seven states), Northeast (nine states), Southeast (four states), South (three states), and Center-West (four states). The incidence and serogroup circulation of IMD in all the populations varies across the Brazilian regions. Figure 2.8 shows meningococcal C incidence at a regional level for less than five years old infants during the pre-vaccination period (January 2007 - October 2010). The Southeast region accounts for the highest incidence (1.04 cases per 100,000 population), followed by the Center-west region (0.64 cases per 100,000). The lowest incidence is reported in the North region. Looking at the sub-regional level, São Paulo state, in the Southeast region, has the highest incidence, being the most populated state in Brazil (Figure C.1 in Appendix C shows the incidence of MenC at a state level).

Also in terms of coverage, the country showed heterogeneous responses: three Brazilian regions reached the target vaccination coverage (>95%) in the first year following programme implementation (2011) [78]. The Northeast region achieved this in 2013; the North region achieved vaccination coverage ranging between 80% and 90% during the study period 2011-2013 [78]. These sub-national differences reflect variations in the implementation of health policies, such as immunisation programmes, as a consequence of regional development and a decentralized health system [78]. Besides, vaccine coverage, quality of disease surveillance as well as disease burden influenced by population density and presence of urban clusters could affect the impact of the vaccination programmes at a regional level [78].

Successful reductions of MenC cases following the vaccine introduction have been shown both at a national and regional level [78, 79]. Observed MenC reductions were mostly due to the direct effect of vaccination [79]. To increase MenC impact, including a possible indirect effect on non-vaccinated groups, a gradual introduction of MenC vaccination targeting adolescents has been introduced starting in 2017 [79].

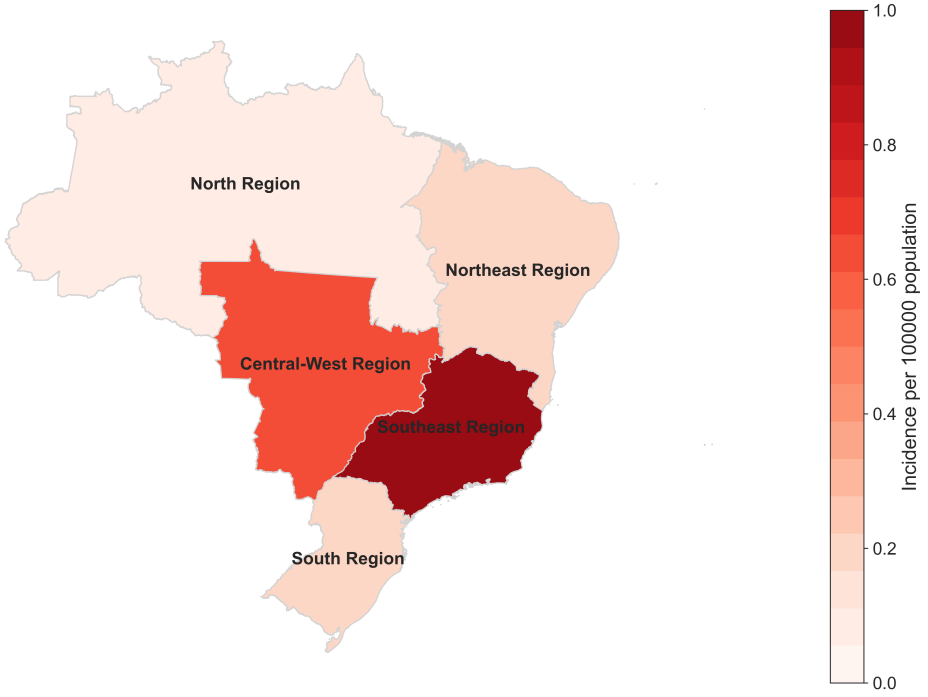

FIGURE 2.8: Meningococcal C incidence per 100,000 population in infants <5 years old at a regional level in Brazil before vaccine introduction (Jan 2007 - Oct 2010).

2.2.2 Meningococcal B campaign in England

In Europe, meningococcal group B is the most prevalent serogroup among childhood meningococcal disease [85]. Indeed, the highest burden of MenB disease occurs during the first 3 years of life [85, 86].

During the past decades, the incidence of MenB in the UK has sharply declined for still unknown reasons (Figure 2.4) [11], among these a possible explanation could be the introduction of a smoking ban in indoor places [11, 85]. In the UK, polysaccharide-protein conjugate vaccines have been effective in preventing meningococcal disease caused by serogroups A, C, W, and Y, thanks to both to the direct and indirect protection conferred through prevention of carriage acquisition among vaccinated adolescents [85, 87]. As discussed in section *2.2*, the development of such a vaccine against MenB is critical given the similarities of MenB polysaccharide capsules with polysialic acid on human neuronal cells cite. A protein-based multicomponent vaccine against MenB (4CMenB, Bexsero) has been licensed in Europe.

For Bexsero, the Meningococcal Antigen Typing System (MATS) has been specifically developed to predict strain coverage by 4CMenB [88, 89]. The MATS allows us to determine whether a certain MenB strain expresses a Bexsero antigen with a minimum degree to be correlated to bactericidal killing. It is sufficient the expression of at least one Bexsero antigen for a strain to be killed. The MATS estimated 73 % strain coverage in England and Wales prior to the introduction of 4CMenB [85, 90].

In September 2015, the UK became the first country to offer 4CMenB to all infants [85]. The vaccine is offered to infants alongside their other routine vaccinations at 2 months, followed by a second dose at 4 months, and a booster dose at 12 months. At the start of the programme, a catch-up vaccination was also offered to children at 12 and 16 weeks of age. From the beginning, high vaccine uptake was observed nationally [85]. Besides, England relies on a near real-time enhanced national surveillance for all laboratory-confirmed cases, which is facilitated by the provision of a single national meningococcal reference unit [85].

A first study providing real-world evidence of the effectiveness of 4CMenB in preventing MenB disease in England has been published 10 months after the programme began [91]. Then, a second study confirmed these results 3 years after the NIP was introduced [85]. In this latter study, it has been observed that the booster at 12-months of age protected against MenB for at least 2 years, which is reassuring towards the initial concerns about swift waning levels of antibodies [92]. To be highlighted, though, that unlike the polysaccharide-protein conjugate vaccines, 4CMenB does not have shown evidence of indirect protection on the carriage in immunized adolescents [85, 93].

Chapter 3

Evaluating the impact of vaccination programmes

Once a new vaccine is introduced in a national immunization program, it becomes critical to evaluate its effectiveness and public health impact. Post-licensure surveillance plays a key role in detecting vaccine effectiveness of various schedules, duration of protection, indirect effects, health disparities, and microbial adaptation while sustaining awareness and identifying potential issues in the implementation of the programme [3, 94, 95]. Post-licensure studies are mainly observational since the real-world effects of a vaccine administered at a population level are difficult to measure in an experimental design [3]. Besides, experimental designs require to administer a placebo to a part of the population, an unethical practice once a vaccine has proven efficacy in the pre-licensure stage [96].

Overall, post-licensure vaccine studies use diverging terms to describe different types of effect [3]. First of all, I aim to illustrate the different effects of vaccination, according to Halloran *et al.* (section *3.1*) [1]. Vaccine effectiveness evaluation comes with several challenges in terms of potential biases which I discuss in the second part of section *3.1*. In section *3.2*, I introduce the three major classes of observational study designs to measure vaccine effects: cohort, case-control, and ecological studies [96]. In sections *3.3* and *3.4*, I focus on vaccine impact both in terms of study designs as well as the critical issues when evaluating this quantity. Finally, in section *3.5*, I discuss the methods applied to estimate the impact of meningococcal vaccines, with examples from immunization programmes implemented worldwide.

3.1 Vaccine effectiveness and impact

The general formula of *vaccine effectiveness* is one minus some measure of relative risk (RR) [1]:

$$VE = 1 - RR \tag{3.1}$$

The relative risk is the ratio between the risk that a certain event occurs in an exposed group over the ratio that the same event happens in an unexposed group. In this setting, the two groups are the vaccinated and unvaccinated individuals respectively, while the event of interest could be the clinical observation of interest or sampling of pathogens in asymptomatic individuals. Depending on the population under study, vaccine effectiveness can be defined differently, as reported in Figure 3.1 [1].

The *direct effectiveness* of a vaccine is measured by comparing vaccinated and unvaccinated individuals belonging to the same population and exposed to the same vaccination programme to cancel the programme-specific effect [1, 97, 98, 99].

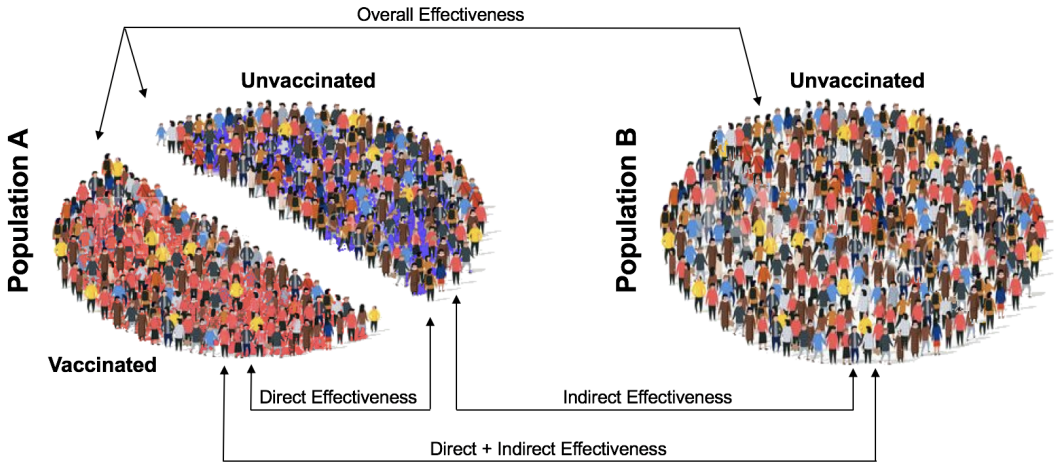

FIGURE 3.1: Schematical representation of groups and population to be compared to compute the different effects of vaccination. Figure rearranged from Halloran *et al* [1].

The *indirect effectiveness* is the population effects of widespread vaccination that is able to reduce transmission and lead to herd immunity [3]. Unvaccinated individuals may thus benefit from those who are vaccinated [1, 95].

The *total effectiveness* is the difference in outcomes between vaccinated individuals in a population with a vaccination programme compared to unvaccinated individuals without a vaccination programme. Here, both direct and indirect effects are taken into account.

The *overall effectiveness* of vaccination compares the average outcomes in the vaccinated communities with those of the unvaccinated communities [1].

The *impact* of vaccination is the overall effectiveness of a vaccination programme [3]. I will use the terms "impact" and "overall effect" of a vaccination programme interchangeably [3, 95].

The vaccine effectiveness can be evaluated on single individuals, communities, or large populations. Here, I will focus on effects at a population level, that are usually evaluated through observational studies.

In general, studies measuring the effectiveness of vaccination programmes rely on three strong assumptions: 1) the individuals/populations being compared must be alike in every way that is relevant for transmission; 2) the probability to be exposed to the infection must be identical in the two groups: vaccinated and unvaccinated; and 3) the probability of being vaccinated must be independent of the probability of developing the disease [3]. However, these assumptions are difficult to ensure in vaccine observational studies at a population level. Several confounding factors could either overestimate or underestimate the effect of vaccination, if they are not properly taken into account.

Indeed, the risk of exposure to the infection, the risk of getting clinically ill if infected, or even the tendency to seek care if ill could introduce a difference between those who are vaccinated and those unvaccinated other than the vaccine [96]. Besides, other sources of bias could be changes in case detection or surveillance methods, the presence of a concomitant event, or even secular trends in baseline transmission or cyclical variations of the target disease [100]. These latter would likely alter the disease's baseline risk before and after the vaccine introduction. On a large scale,

also population characteristics, such as population growth, immigration, changes in age distribution, and changes in risk factors for the disease of interest are likely to be different [3].

3.2 Observational study designs to estimate vaccine effects

In this section, I will give an overview of the observational study designs that are typically applied to evaluate vaccine effects. The three major classes of observational study design are cohort, case-control, and ecological studies. Here, by ecological studies, I will refer to time-trend designs, following the approach of Lipsitch *et al.* [96].

Cohort studies are prospective studies, with two groups followed over a period of time (i.e., one group is vaccinated and the other has received a placebo, and vaccination status is known). Cohort studies are powerful because they have the advantage of being tailored to collect specific exposure data and may be more complete. Among the disadvantages, there is the need for a large sample size.

Case-control studies are instead prospective studies, again the vaccination status is known and the two groups represent vaccinated and unvaccinated people, respectively. To avoid potential biases, the population should be the same both for the cases and the controls (e.g., same geographical area, same age, and sex distribution, etc..). This study is at risk for selection bias (i.e., if proper randomization is not achieved, the sample obtained is not representative of the population intended to be analyzed), with consequent non-comparability between the cases and controls. Negative controls and test-negative designs could address these forms of bias [96].

Time-trend designs are a form of longitudinal ecological study, where the data are collected at a population level. In vaccine impact studies, the vaccinated population is compared with a population without the vaccination programme. Based on the control population, time-trend studies can be in the form of 1) before-after comparison or 2) controlled designs (Figure 3.2) [101, 102].

Both case-control and cohort studies are usually employed to measure the direct effectiveness of a vaccine, while time-trend studies address the impact or overall effectiveness of vaccination programmes [96].

3.3 Vaccine impact

The impact of a vaccination programme answers the critical question: what would have happened in the absence of vaccination? In the majority of applications, the vaccine impact is intended as the overall reduction in disease burden, usually computed by comparing the population with vaccination programme to a reference population not subject to the programme [3]. This reference population could be either the population of interest before vaccination or an external population not influenced by the vaccination programme. The formula defining the vaccine impact is:

$$VI = 1 - IRR \tag{3.2}$$

where IRR is the incidence rate ratio, i.e. the ratio between the number of observed cases in the vaccinated population over the number of predicted cases in the unvaccinated population. As highlighted in section *2.1*, this measure corresponds to the overall effectiveness as defined by Halloran *et al.* [1].

Ecological studies are typically employed in vaccine impact studies at a population level, specifically, time-trend studies [96]. In general, time-trend studies require to make numerous decisions a priori about how to model the data. Among these decisions, I would include selecting which and how many years to analyze in the 'pre' and 'post' vaccine periods, whether to exclude data from the vaccine introduction period, how to control for trends unrelated to the vaccine (including seasonality), which modeling strategy to employ (i.e., a before-after comparison or a controlled design) [103]. For this reason, each methodology stands on its own assumptions and in general, it is always preferable to compare different methods together to increase the reliability of the results and to help in discovering several sources of biases [103]. The two classes of methods, before-after comparison, and controlled design, are presented below.

3.3.1 Before-after comparison

In a before-after comparison, the same population is compared across time, i.e. before and after the intervention. In this case, the pre-period acts as a control. This kind of design is not subjected to between-group differences that could lead to selection bias or unmeasured confounders [102]. Indeed, the population characteristics tend to change only gradually over time and confounding is rarely an issue [101, 102]. Besides, this approach models the underlying trend, thus controlling for within-group characteristics, secular trends, and random fluctuations [102].

This method, however, cannot exclude time-varying confounders, e.g. other events or co-interventions that could affect the outcome [102]. Besides, in most cases, the trend is assumed to continue linearly after the intervention, and this assumption is rarely verified with consequent underestimation or overestimation of vaccine impact.

Time-trend designs have been widely used to assess the impact of vaccination programmes, such as the decline in pneumonia hospitalizations [4], the impact of meningococcal C campaign in Brazil [79], or the effect of rotavirus vaccination on death from childhood diarrhea [5].

3.3.2 Controlled design

In a controlled design, the same outcome in the target group is compared with an external control. Here, the target group and a control group are compared at the same time points, thus avoiding the influence of time-varying factors that could hamper a correct assessment of impact estimates [102].

Following Lipsitch *et al.*, a negative-control design could effectively detect unmeasured confounding and bias in an observational study [103, 104]. Here, the ideal control should be affected by the same set of causal relations as the target, but should not be affected by the exposure of interest, i.e. the vaccination programme [103, 104].

In general, however, choosing the appropriate control group is challenging for several reasons. First, it is unclear how to find a control group that effectively shares the same set of causal factors with the target group, since these are usually unknown. Besides, differences between the target and the control could introduce other sources of bias [102, 103]. Or even, the control group could be partially affected by the intervention, thus the overall effect of the intervention would be underestimated.

Recently, data-driven methods, i.e. the synthetic control model, have been developed to overcome the possible biases due to arbitrarily selecting controls [6, 7, 105]. I will discuss in deep the synthetic control model in chapter 4.

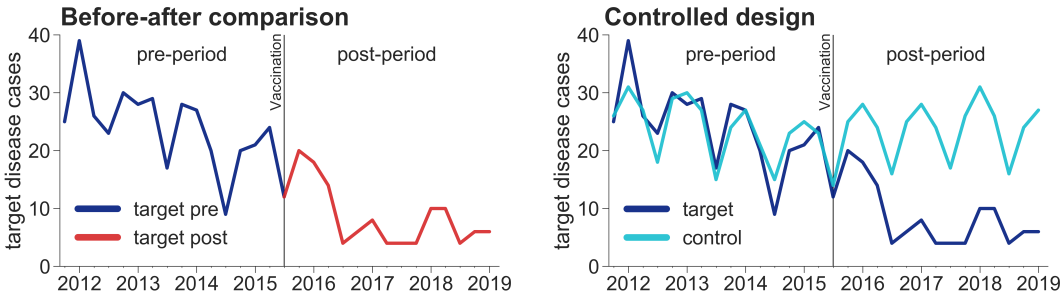

FIGURE 3.2: Schematic representation of time-trend designs. On the left, a before-after comparison where the control group is the population before the intervention. The impact of vaccination is the difference between the same population before and after the intervention. On the right, a controlled design, with an external control being compared with the target of interest. The impact of vaccination is the difference between the two curves in the post-intervention period.

3.4 Challenges in evaluating vaccine impact

Few studies have been carried out about the public health impact of long-standing vaccination programmes at the population level, despite being a causal quantity of policy relevance [3, 95, 96]. Among the challenges encountered, there is the need for time-series of reliable historical data on case notifications, hospitalizations, or on cause-specific mortality [95, 106].

Despite long-term datasets could lack in reporting fidelity (e.g., national public archives undergo substantial modifications in data collection and reporting over the years), they however represent a unique and valuable source of information. Indeed, with long-term data, statistical associations between climate variations and the incidence and dynamics of infectious diseases can be investigated and allow to take into account secular trends in birth rates or vaccination coverages which could influence shifts in dynamical patterns [106]. Also, having long historical time-series is helpful to disentangle the vaccine's effects by other unexplained factors, when immunization programmes are implemented.

Finally, there is also a lack of standardized methods to identify the appropriate control groups [95, 104]. Ideally, the two populations being compared should be the most possible alike, and the only difference should be given by the vaccine. In practice, it is challenging to find such a control population.

3.5 Assessing the impact of meningococcal vaccines

Evaluating the impact of meningococcal vaccines is critical due to the low disease incidence and to IMD unexplained trends. Indeed, several years of observation may be required to obtain statistically significant estimates, during which time natural fluctuations in IMD incidence unrelated to vaccination are likely to occur. This phenomenon inevitably introduces risks of biased estimates and misinterpretation about causal effects [104].

Besides, the peculiar age-and-time pattern of IMD and its relation to geographical locations rely upon the interplay of several possible causative effects such as bacterial transmissibility and virulence, immune system maturation and senescence, level and breadth of mucosal and systemic immunity, frequency and type of person-to-person contacts, social habits like smoking and kissing, and environmental factors

[12, 15, 107]. Explicitly taking into account all these causal factors is extremely challenging, even with sophisticated mathematical models [97, 108, 109, 110, 111, 112, 113].

Finally, also underreporting of IMD cases complicates the efforts to understand its occurrence and burden [114]. In some countries, a lack of specificity regarding the serogroup distribution is present (i.e., invasive bacterial disease cases are often coded as "meningitis due to unspecified bacterium" or "unidentified serogroup"), with consequent difficulties in understanding the chronological trend of the different serogroups [78, 114].

In general, the majority of the studies evaluating the impact of meningococcal vaccination rely on a before-and-after design [100]. For example, the reduction of the incidence of serogroup C IMD in Canada in children aged 1–17 years, following the introduction of the vaccine programme, was computed comparing the disease rates pre- and post-intervention [115]. Similarly, a study conducted in Puglia, Italy, evaluated MenC impact by comparing the incidence rates before and after the vaccination [114]. Among the possible limitations of this approach, there is the lack of adjusting for unexplained trends if other events occur at the same time as the vaccination.

To account for changes unrelated to the meningococcal vaccination, IMD cases in non-vaccine-eligible age groups are used as controls in a few studies. Among the examples, a study assessing the impact of the meningococcal group B vaccine implemented in England in 2015, with MenB incidence in non-vaccine-eligible children of $<$5 years of age used to control for unexplained trends [85, 91]. A similar approach was followed to compute the impact of the MenB vaccine in one region of Quebec, Canada [44].

Selecting IMD cases in unvaccinated age groups as controls guarantees that they probably share most of the causal factors with the target disease vaccinated age groups. However, if the vaccination provides indirect protection to the unvaccinated population, as for conjugate vaccines (chapter 2, section 2.2), this control group would eventually show reduced disease incidence and impact estimates would be no longer reliable. Nevertheless, this would likely happen if the vaccination schedule targets also the adolescents, the segment of the population that usually harbors the highest rates of meningococcal carriage [24, 46, 79]. As a general statement, however, indirect effects cannot be excluded and care is required when evaluating the different methodologies to assess the effect of the vaccination programmes.

3.6 Conclusions

In this chapter, I have reviewed the various definitions of the effects of a vaccination programme, focusing on vaccine impact evaluation. I briefly discussed the main observational study designs, i.e. cohort, case-control, and time-trend studies, in terms of assumptions and potential biases. Time-trend studies are traditionally employed to evaluate the overall effect of a vaccination programme at a population-level, either with a pre-post comparison or with external controls.

In meningococcal studies of vaccine impact, the pre-post comparison is typically applied, despite the potential issues when other events occur at the same time as the intervention. To control for trends unrelated to the vaccination, IMD cases in non-vaccine-eligible age groups are used as controls in a few studies. While this approach ensures that such controls share the same set of causal factors as the target disease, potential biases could be introduced if herd immunity occurs.

In the next chapter, I will discuss the synthetic control method, which avoids the selection of the controls a priori.

Chapter 4

The Synthetic Control model

In this chapter, I provide an introduction to how synthetic controls can be applied to evaluate the causal impact of an intervention. In section *4.1*, I give a general overview of synthetic controls. From the second part of section *4.1* onwards, I then follow the approach developed by Brodersen *et al.* in a Bayesian setting, where the outcome is a time series and the relevant control variables are chosen via Bayesian variable selection [105].

Structural time series models are employed to build the counterfactual, incorporating both the information from the target time series in the pre-intervention and from the control time series in the pre- and post-intervention periods. Structural time series models are described in detail in section *4.2*, focusing on a specific model with trend, seasonal, and regression components. Along this section, I discuss the Bayesian variable selection approach, the posterior inference, and the prediction of the causal impact. In subsection *4.2.1*, I present the spike-and-slab prior used to select the most relevant control time series.

4.1 Introduction to synthetic controls

Synthetic controls have been developed to evaluate the causal impact of an intervention [6, 105, 116]. In several applications, as post-licensure vaccine studies, data are observational and several unobserved confoundings could hamper a reliable assessment of the true impact of interventions (chapter 3, section *3.1*) [104].

To improve the reliability of impact assessment, external controls can be employed to account for the presence of (likely unobserved) confounders and of unpredictable, non-vaccine-related temporal trends in disease incidence [6, 116]. There is however some degree of ambiguity about how to choose these controls, and often choices are made based on subjective measures of affinity between controls and outcome [6].

Data-driven methods, such as synthetic controls, have been introduced to build adequate control groups and to overcome the limitations of arbitrarily selected controls [6]. In the formulation by Abadie *et al.*, the synthetic control is a weighted average of several control units combined together to provide a more robust comparison with the outcome of interest [6]. The rationale is to find the weights for the control units such that the weighted average of the control's outcomes best predicts the outcome of interest in the pre-intervention period [6, 116]. The advantage of such an approach is that it does not require any insight into post-intervention outcomes. Consequently, decisions on study designs are made without any knowledge about how those decisions affect the conclusions of the study [6, 117].

A further contribution to the topic has been proposed by Brodersen *et al.* [105] in a Bayesian setting, where the outcome variable is a time series. In the present setting,

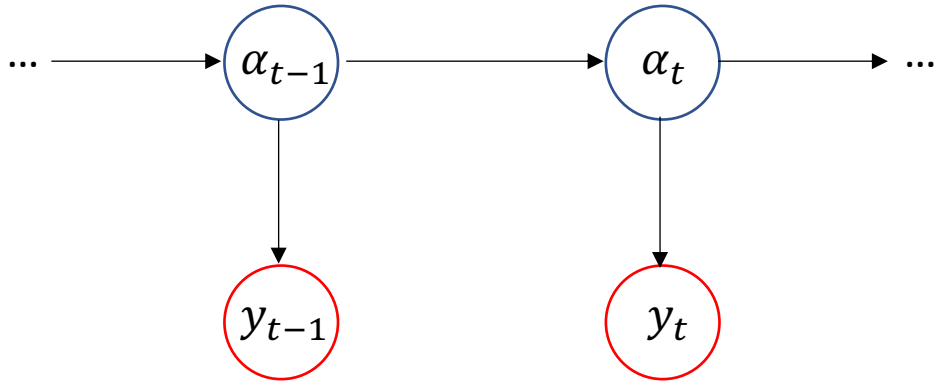

FIGURE 4.1: Diagram of a state space model readjusted by Shumway *et al.* [118]. In this figure, α_t is called the state process, while y_t is the observation equation. The dependence among observations is generated by the states.

the causal impact is the difference between the observed target time series and the time series that would have been observed in the absence of the intervention, i.e. the counterfactual [105].

In the upcoming sections, I will follow the approach by Brodersen *et al.* [105]. Three sources of information are required to build adequate synthetic controls. First, the target time-series in the pre-intervention period [105]. Secondly, the control time-series that are predictive of the target time-series in the pre-intervention period. The selection of the relevant set of control time series is done in the pre-intervention period, while their value for predicting the counterfactual is in their post-intervention temporal evolution [105].

If the control time series are not affected by the intervention, it is often reasonable to assume the relationship between the target and the control series that existed prior to the intervention to continue afterward. Finally, the third source of information lies in the prior knowledge about the model parameters, given the Bayesian framework [105].

These three sources of information are then combined using a state-space time-series model, also defined as structural time series models [105, 119]. In section 4.2, I provide the technical details on structural time series models.

4.2 Bayesian structural time-series models

Structural time-series models are state-space models for time series data [105]. These state-space models are characterized by a latent process α_t, called the state process, and by an observation equation, with y_t being the observations. The state process is assumed to be a Markov process, i.e. future and past are independent conditional on the present, and the observations are independent given the states. The dependence among observations is thus generated by the states (Figure 4.1) [118, 120].

Given y_t an observation at time t, a structural time-series model can be defined in terms of two equations that link the observation y_t to a vector of latent state variables α_t:

$$y_t = Z_t^T \alpha_t + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma_t^2) \tag{4.1}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \qquad \eta_t \sim N(0, Q_t) \tag{4.2}$$

Equation (4.1) is the *observation* equation, which links the observed data y_t with the unobserved latent space α_t; equation (4.2) is the *state* equation and it defines how the latent state evolves over time. The model matrices Z_t, T_t, and R_t typically contain a mix of known values (often 0 and 1), and unknown parameters [119]. The transition matrix T_t is square, but R_t can be rectangular if a portion of the state transition is deterministic [119]. In section *A1* in Appendix A, I provide further details about state space models, specifically the Kalman filter and smoother used to characterize the evolution of the state sequence α.

Ideally, a statistical model should include all relevant information that might help predict the outcome of interest, such as local trends, seasonality, and time-varying factors, these latter accounting for the effects of other unobserved causes otherwise unaccounted for by the model [121]. State space models allow us to flexibly incorporate these multiple sources of variations, with time-varying factors included through the control time series [105].

To avoid a rigid commitment to a fixed set of controls, Bayesian variable selection is employed to choose from among a large set of candidate control variables a priori [105] (left panel in Figure 4.2). With many control variables, each with a fair probability of being irrelevant to modeling the outcome variable, one can give each coefficient a prior distribution with a peak at zero and a long tail [121]. This says that each variable is probably unimportant, but if it has predictive power, it could be large [121]. In practice, a spike-and-slab prior is placed over the coefficients [105, 119, 122]. Spike-and-slab prior combines point mass at zero (the "spike"), for an unknown subset of zero coefficients, to allow shrinkage of small effects to zero with a weakly informative distribution on the complementary set of nonzero coefficients (the "slab") to prevent shrinkage of large effects [105, 119, 122]. The "slab" is usually not completely flat, but rather a Gaussian with a large variance [105, 123].

Once the control's selection is performed, it becomes straightforward to compute the counterfactual time series [105]. A schematic example of the procedure is given in Figure 4.2, with regression, seasonal and random changes components added together and fitted to the pre-intervention period. Their value is projected into the post-intervention period to build the counterfactual. The causal impact is then the difference between the counterfactual and the observed outcome during the post-intervention period [105].

Example in Figure 4.2, with an additional trend component, can be written in state-space form as [119]:

$$\begin{aligned} y_t &= \mu_t + \tau_t + \beta^T x_t + \epsilon_t \\ \mu_t &= \mu_{t-1} + u_t \\ \tau_t &= -\sum_{s=1}^{S-1} \tau_{t-s} + \omega_t \end{aligned} \tag{4.3}$$

where $\eta_t = (u_t, \omega_t)$ contains independent components of Gaussian random noise.

FIGURE 4.2: Schematic example of the synthetic control method. Three key components: 1) the most important predictors combined into a synthetic control, and 2) a seasonal component and 3) random changes are fitted to the outcome of interest in the pre-intervention period and then projected into the post-intervention period to build the counterfactual.

The current level of the trend is μ_t which induces temporal correlation in the outcome [116, 119]; τ_t represents the seasonal component, with a set of S dummy variables with dynamic coefficients constrained to have zero expectation over a full cycle of S seasons [119] and the error component ϵ_t accounts for unexplained variability [116, 119].

Control series are included through linear regression, in state-space form as $\beta^T x_t$. The value of the vector x_t depends on the context: it could represent a set of search queries in economy [119], or a set of diseases in epidemiology [7]. The regression component allows us to obtain counterfactual predictions by building a synthetic control based on a combination of controls not affected by the intervention. The Bayesian variable selection step is obtained by placing a spike-and-slab prior over the coefficients β.

The state-components are assembled independently, with each component providing an additive contribution to y_t. The model of equation (4.3) is fitted to the observed data $y_{1:n}$ in the pre-intervention period, treating the counterfactuals as unobserved random variables [116].

Let ϕ be the set of all model parameters and α the full state sequence, a prior distribution $p(\phi)$ is specified on the model parameters as well as a distribution $p(\alpha_0|\theta)$ on the initial state values. Posterior inference can be carried out as follows: draws of the model parameters ϕ and the state vector α are simulated given the observed data $y_{1:n}$ in the pre-intervention period. Then, posterior simulations are used to simulate from the posterior predictive distribution $p(\bar{y}_{n+1:m}|y_{1:n}, x_{1:m})$, over the counterfactual time series $\bar{y}_{n+1}, ..., \bar{y}_m$ [105]. This step can be done using a Markov chain Monte Carlo (MCMC) sampling scheme which involves only Gibbs sampling steps (details in Appendix A, sections $A.2$ and $A.3$)[105, 116, 119, 123].

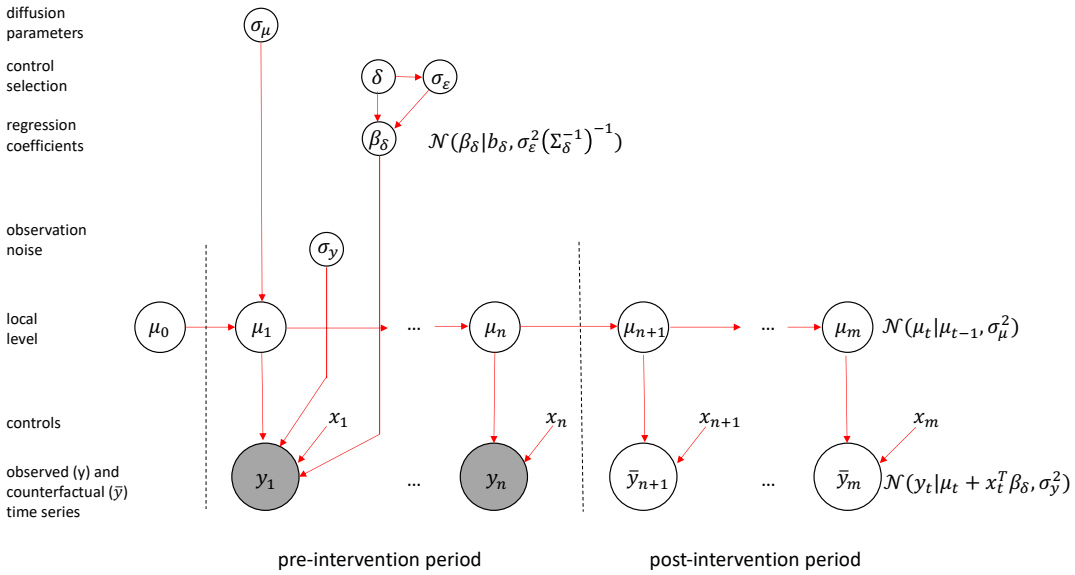

FIGURE 4.3: The observed target $y_{1:n} = (y_1, .., y_n)$ is modeled as the result of a latent plus Gaussian observation noise with error standard deviation σ_y. The state component α_t includes a local level μ_t, and a set of control time-series x_t, scaled by regression coefficients β_δ. State components are assumed to evolve according to independent Gaussian random walks with standard deviation σ_μ (the conditional-dependence arrows are shown for the first time point only) [105]. Of principal interest is the posterior predictive density over the unobserved counterfactual responses $\bar{y}_{n+1}, ..., \bar{y}_m$. Subtracting these from the actual observed data $y_{n+1}, ..., y_m$ yields a probability density over the temporal evolution of causal impact [105].This figure has been readjusted by Brodersen *et al.* [105].

The posterior predictive distribution is conditional on the observed data, in particular on the target time series before the intervention and on the control time series before and after the intervention. The distribution is not conditioned on the estimation of parameters, or the inclusion or exclusion of covariates, all of which have been integrated out. In this way, an arbitrary selection is avoided [105]. The posterior predictive samples are employed to compute the posterior distribution of the pointwise impact $y_t - \bar{y}_t$ for each $t = n{+}1,...,m$ [105].

Figure 4.3 illustrates this process, considering only the trend and regression components.

4.2.1 Additional information on the spike-and-slab prior

Let $\delta = (\delta_1, .., \delta_j)$, where $\delta_j = 1$ if $\beta_j \neq 0$ and $\delta_j = 0$ otherwise; β_δ denotes the nonzero elements of vector β [105, 119]. A spike-and-slab prior (Figure 4.4) is then written as:

$$p(\delta, \beta, \frac{1}{\sigma_\epsilon^2}) = p(\delta)p(\sigma_\epsilon^2|\delta)p(\beta_\delta|\delta, \sigma_\epsilon^2) \tag{4.4}$$

The marginal distribution $p(\delta)$ is called the "spike" and it places positive probability mass at zero. Usually, an independent Bernoulli prior is applied:

$$\delta \sim \prod_{k=1}^{K} \pi_k^{\delta_k}(1 - \pi_k)^{1-\delta_k} \tag{4.5}$$

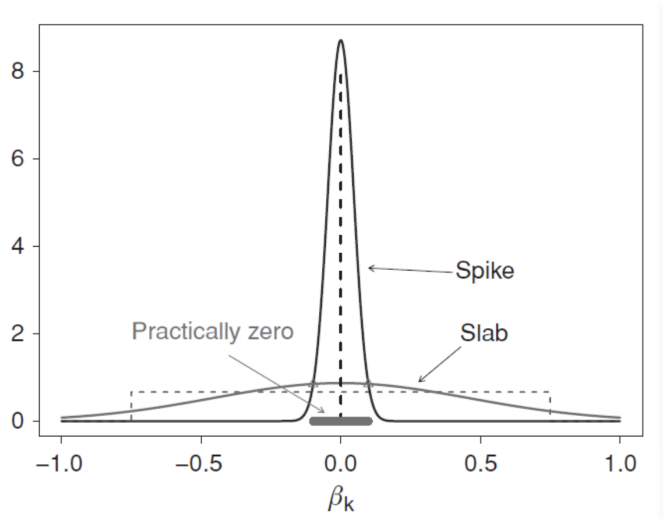

FIGURE 4.4: Representation of a spike-and-slab distribution, with a discrete mixture (π=0.5) of a Dirac delta (the spike) and a Cauchy distribution (the slab). Figure from [126].

For simplicity, all the π_k are assumed to be the same value $\pi = 0.5$, i.e. equal probability over all the control time-series. This value can be changed if a control time series is thought to be more predictive than others.

The slab component is modeled as a conjugate normal-inverse Gamma distribution, with prior parameters only very weakly informative [105, 119]:

$$\beta_\delta | \sigma_\epsilon^2 \sim N(b_\delta, \sigma_\epsilon^2 (\Sigma_\delta^{-1})^{-1}), \tag{4.6}$$

$$\frac{1}{\sigma_\epsilon^2} \sim Ga(\frac{\nu_\epsilon}{2}, \frac{s_\epsilon}{2}), \tag{4.7}$$

Usually, the vector of prior means b is assumed to be zero. The prior parameters ν_ϵ and s_ϵ represent the number of relevant observations and the expected R^2 from the regression, respectively.

The choice of Σ^{-1} is tricky: it could be set to our prior beliefs about the interaction between the β_j, however, this could be implausible when dealing with many features. Conventionally, one option is to use a Zellner prior [105, 119, 124, 125], so $\Sigma^{-1} \sim X^T X$, with X the design matrix.

Setting $\Sigma^{-1} = \frac{g}{n} X^T X$ gives the g observations worth of information on the prior mean b (recall that the likelihood for an ordinary regression model has information matrix $X^T X / \sigma_\epsilon^2$, i.e. it is a measure of how much information is known about a parameter θ) [105, 119]. If X variables are collinear, however, the g-prior becomes improper. To guarantee full rank, $X^T X$ is averaged with its diagonal, obtaining:

$\Sigma^{-1} = \frac{g}{n} w X^T X + (1 - w) diag(X^T X)$, with default values g=1 and w=1/2 [105, 119].

4.3 Conclusions

In this chapter, I have illustrated the principles of the synthetic control method, focusing on the application developed for time-series data. This approach relies on structural time series models, which I described in section 4.2. Along this section,

I have presented an example with a trend, seasonal, and regression components. I have discussed the rationale behind variable selection, which allows us to choose among a large set of control variables. This process is done by placing a spike-and-slab prior over the coefficients of the control variables (details in subsection *4.2.1*).

I have also presented the main features of posterior inference, though at a general level. Further details on posterior definitions and MCMC scheme are available in appendix A, sections *A.2* and *A.3*. In chapter 5, I will describe how synthetic controls can be applied to meningococcal vaccines, defining the main building blocks of my research project.

Chapter 5

Meningococcal vaccine impact evaluation

In this chapter, I describe the methods applied in my research. In section *5.1*, I present the synthetic control approach developed to fit disease incidence data, specifically meningococcal incidence. In section *5.2*, I discuss the control time series employed in my research, including a description of how their importance is evaluated in the model. In section *5.3*, I present the features of a simple time-trend model to be compared with the synthetic control.

Section *5.4* is a brief reminder of how vaccine impact is evaluated. In section *5.5* and *5.6*, I describe the approaches to evaluate the accuracy of synthetic control and time-trend models and to verify the robustness of the synthetic control to the exclusion of specific control variables, respectively. Section *5.7* presents alternative methods to deal with sparse and noisy control time series.

Finally, section *5.8* summarizes the above sections aiming to provide a general approach that I applied to MenB and MenC case studies (chapters 6 and 7).

5.1 Synthetic control model applied to meningococcal incidence data

The synthetic control approach has proven to be a robust and appealing tool to quantify the effect of public health interventions in several studies. It has been applied to assess the impact of pneumococcal conjugate vaccines (PCV) on pneumonia-related hospitalizations in different countries, to evaluate the effect of PCV introduction on pneumonia mortality in children in Brazil, to evaluate maternal acellular pertussis vaccine impact on reducing infant disease burden, and to study the effect of PCV in Brazil comparing low- and high-income populations [7, 127, 128, 129].

To evaluate the impact of public health interventions such as vaccination campaigns, data are typically in the form of counts of disease cases. In this context, a Poisson model is more appropriate, with an observation-level random intercept to take into account overdispersed data. Despite the mathematical complexity, easy implementation is available as an R package *InterventionEvaluat* [130] with Bayesian variable selection performed by the package *poissonBvs* [123].

The target disease cases y_t at time t are modeled as a Poisson process, $y_t \sim Poisson(\lambda_t)$, with mean λ_t [130, 131]:

$$log(\lambda_t) = b_0 + \sum_k c_k * I[month_k = m(t)] + \sum_{j=1}^{p} \beta_j(\delta_j) * x_{jt} + b_{c(t)} \tag{5.1}$$

where t=1,2,..., total number of time points, x_{jt} represents the number of cases of control disease j at time t; m_t is a function that maps a time point to the corresponding calendar month; c_k represents the month k regression coefficient; *I[.]* represents the indicator function; b_0 is an intercept; p is the total number of control diseases included in the analysis; $\beta_j(\delta_j)$ is the regression coefficient for control disease j which is given a spike-and-slab prior distribution (depending on δ_j) in order to allow for data-driven variable selection; δ_j are binary random variables that are equal to 1 if the control disease j is included in the model or equal to 0 if it is excluded; and $b_c(t)$ is an observation specific random intercept, where $b_c(t) \sim N(0, \sigma_b^2)$.

Here, the spike and slab component is defined as follows: the former component is a Dirac spike, i.e. a point mass at zero, while the slab component is specified as a scale mixture of normal distributions, resulting in a Student-t distribution [123]. For posterior inference, an MCMC sampling scheme is used [123] (details in chapter 4, section *4.2* and in appendix A, sections *A2* and *A3*). I collected 10,000 posterior samples after a burn-in period of 5,000 iterations.

5.2 Disease time series used as controls

The control diseases must follow 2 criteria: (1) they must not be affected by the vaccine under study and (2) the relationship between the outcome and the controls have to be constant over time in case the vaccine would not have been introduced.

In the two meningococcal infant immunization programmes under study, i.e. the MenC campaign in Brazil and the MenB campaign in England, the control time series can be distinguished in two groups. The first group includes several infectious/non-infectious diseases in the same country and age group of the target disease, following the same approach of Bruhn *et al.* [7]. As a second group, I included also MenC and MenB cases in non-vaccine-eligible age groups, assuming that they are not impacted by indirect effects of the respective vaccine infant programs during the considered time-lapse [78, 79, 85].

5.2.1 Importance of the control time series

In the synthetic control framework, it is interesting to understand the control time series that most contributed to predicting the target time series, here meningococcal disease.

A quantity of interest is the probability of inclusion, defined as the mean of δ over the MCMC iterations for each control variable (section *5.1*: the binary indicator δ is equal to 1 if the control variable is included in the model, and it is equal to 0 otherwise). Each control time series is assigned a probability of inclusion, whose values range from 0-1, with values closer to 1 indicating that the control variable is included in a larger proportion of the variable combinations that are tested [130]. Following the approach of Bruhn *et al.* [7], I consider a control disease as "selected" by the model if its probability of inclusion is higher than 50%.

Another quantity of interest is the model size. In each MCMC iteration, the synthetic control tests a different combination of control variables. The model size indicates how many controls are selected in any given model. If less than one variable is selected on average, the synthetic control fails to identify suitable control variables [130].

These two quantities, i.e. inclusion probabilities and model size, have been calculated by the model defined in the R package *InterventionEvaluat* [130].

5.3 Time-trend model

I compared the synthetic control model with a simple time-trend model. The time-trend model only adjusts for seasonality and linear trend, assuming that the trend from the pre-vaccine period continues into the post-vaccine period [7, 131]. Results should be consistent with the synthetic control model if there are no unexplained biases in the data [7].

In a time-trend model, time-series cases of meningococcal disease are assumed to follow a Poisson process $y_t \sim Poisson(\lambda_t)$, with mean λ_t [7, 130, 131]:

$$log(\lambda_t) = b_0 + \sum_k c_k * I[month_k = m(t)] + \beta_1 * index_t + b_{c(t)} \tag{5.2}$$

where t = 1,2,..., is the total number of time points; m_t is a function that maps a time point to the corresponding calendar month; c_k represents the month k regression coefficient; $I[.]$ represents the indicator function; b_0 is an intercept; $index_t$ is a time variable with values from 1 to the total number of time points; and $b_c(t)$ is an observation specific random intercept.

5.4 Evaluation of vaccine impact

Vaccine impact is computed by comparing the total number of observed cases (Y_{obs}) and the number of predicted counterfactual cases (Y_{cf}) during the evaluation period. As pointed out in chapter 3, section *3.3*, the general formula for vaccine impact (VI) is:

$$VI = 1 - IRR \tag{5.3}$$

where IRR is the incidence rate ratio, $IRR = Y_{obs}/Y_{cf}$ [7].

I introduced a gap of 1 to 2 years, depending on the age group, between the vaccine introduction and the beginning of the evaluation period, in agreement with previous impact studies (details are displayed in the sections *Data* in chapters 6 and 7) [78, 85]. Besides, vaccine impact estimates obtained using the synthetic control and time-trend models were compared with previous publications [78, 85].

5.5 Negative control analysis

Since the counterfactual is never observed in causal impact studies, any model could be correct, at least in principle. For this reason, it is hard to evaluate the performances of different models.

In my research study, I performed a negative control analysis with the aim of evaluating the prediction accuracy of the synthetic control and time-trend models. Following the assumption of the absence of indirect protection from the vaccine, I used IMD cases in unvaccinated age groups as targets of the two models.

If the models appropriately adjust for trends in the data, there should be no measurable vaccine effect, if the assumption of absence of indirect protection from infant vaccination is met. Consequently, if the model provides good predictions, the counterfactuals will be close to the observed points.

I used the mean absolute error (MAE) to quantify prediction accuracy in the post immunization period (further details in appendix A, section *A.4*).

5.6 Sensitivity analysis: testing the vaccine's indirect effectiveness assumption

Using IMD cases in unvaccinated age groups as controls could bias one of the assumptions of the synthetic control model, i.e. the control diseases must not be affected by the intervention. In the two vaccination campaigns under study, I assumed no indirect protection from the vaccine since the immunization programme targets only infants, a limited fraction of the population to generate measurable herd immunity effects. Besides, infants are characterized by low carriage rates compared to the average population (Figure 2.2 in chapter 2, section *2.1*).

However, if the assumption of the absence of indirect protection from the vaccine is violated, this would bias the vaccine impact estimates towards seeing no effect. Indeed, IMD cases in unvaccinated age groups used as controls would be themselves affected by the intervention leading to underestimated impact results. This condition would affect the reliability of the negative control analysis as well.

To rule out any risk of bias due to indirect effects, I performed a sensitivity analysis by re-running the model excluding from the control set IMD cases in the unvaccinated age groups. I repeated this step also in the negative control analysis. If the vaccine impact results are consistent in both cases, i.e. with and without IMD cases in the control group, I can exclude any bias due to the indirect effect of the vaccination.

Additionally, I quantified the performances of the two models, i.e. with and without IMD cases in the control group, on vaccine-eligible age groups both in terms of goodness of fit (i.e., Deviation Information Criterion, DIC, metric) and prediction accuracy (i.e., MAE metric). Details on DIC and MAE metrics are available in appendix A, section *A.4*.

5.7 Alternative methods with sparse data: the STL+PCA approach

Several vaccine impact studies are often conducted using smaller, subnational datasets, as the impact of meningococcal C vaccines at a regional level in Brazil [78]. The vaccine impact estimates produced by the synthetic control model could be biased, i.e. non-significant, when data are sparse [131]. In these circumstances, the synthetic control model could fail to identify an appropriate set of controls (i.e., the probability of being included in the model would be lower than 50%).

A straightforward approach to reduce the impact of sparseness is to aggregate monthly data into quarterly data [131]. However, this approach would not be a good solution when the pre-intervention data are limited [131]. Besides, using lower resolution time series reduces the number of data points, which makes it challenging to establish relationships between the outcome and the synthetic control [131].

Alongside with testing the synthetic control with different temporal aggregation, I applied an alternative approach, the 'seasonal-trend decomposition plus principal component analysis (STL+PCA)' model (Figure 5.1) [131]. This method has shown to adjust for trends with noisy and sparse control time series [131, 132].

In the STL+PCA method, there is no variable selection. The first step is to extract a long-term trend for each control disease using the seasonal-trend decomposition procedure based on locally weighted scatterplot smoothing (STL)[131, 133]. The second step includes obtaining a "composite" from these extracted trends where principal component analysis (PCA) is performed to reduce the dimensionality of the

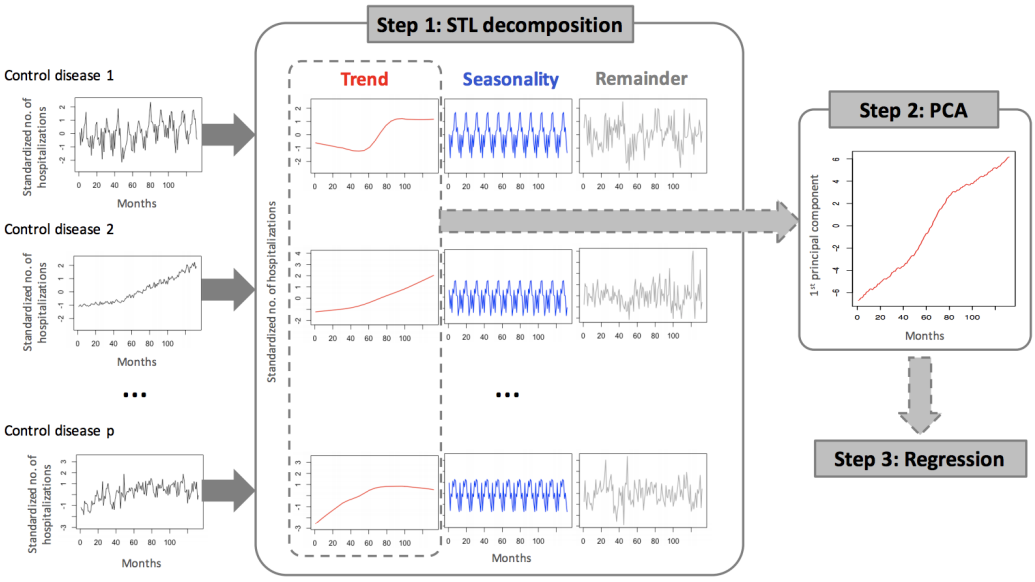

FIGURE 5.1: Diagram of the STL+PCA model. Abbreviations: STL, seasonal-trend decomposition procedure based on locally weighted scatterplot smoothing; PCA, principal component analysis. Figure from [131].

total set of trends [131]. Finally, the first principal component "PC1", which accounts for most of the variance, is used as a covariate in the regression model to build the counterfactual [131].

A relevant disadvantage of the STL+PCA model is that it is no longer possible to interpret relationships between the outcome and control diseases since the original control time series are not directly used as a covariate in the model [131]. The R package *InterventionEvaluat* [130] provides an implementation of the STL+PCA model alongside with the synthetic control and time trend models. Further details are available in Appendix A, section *A.5*.

5.8 Conclusions

The aim of this section is to provide the building blocks of my doctoral research.

The core of my analysis can be then summarized as follows:

1. Negative control analysis: prediction accuracy of synthetic control and time trend models is compared using non-vaccine-eligible age groups as targets.

2. Vaccine impact estimation: impact results obtained with the synthetic control are compared with previous studies and with the time-trend model.

3. Meningococcal disease predictors: the three most relevant predictors of IMD incidence are investigated according to the probability of inclusion.

4. Sensitivity analysis: the indirect protection assumption is tested by excluding the IMD cases in unvaccinated age groups from the set of controls and by re-running the synthetic control model.

5. Alternative methods with sparse data: different control time series' aggregations (both spatial and temporal) and the STL+PCA approach are taken into account to overcome the issue of sparse data.

These building blocks will be applied to the meningococcal C campaign in Brazil and the meningococcal B campaign in England, respectively (chapters 6 and 7), to investigate the validity of synthetic controls in meningococcal vaccine impact studies.

Chapter 6

The impact of the 2010 MenC campaign in Brazil

In chapter 5, I discussed how the synthetic control method can be applied to assess the impact of meningococcal vaccination programmes. In this chapter, I present the results of the meningococcal C campaign in Brazil.

In section *6.1*, I discuss the MenC data collected to perform my analysis. In sections *6.2*, *6.3*, *6.4*, *6.5*, I show the results following the scheme described in chapter 5, section *5.8*, and briefly summarized below. The scheme includes: 1) negative control analysis; 2) vaccine impact estimation; 3) meningococcal disease predictors; 4) sensitivity analysis; and 5) alternative methods with sparse data. Steps 1), 2), 3), 4), are repeated for the Brazilian state with the highest incidence of MenC, São Paulo state. São Paulo state relies on an effective and reliable surveillance system for meningitis and it has been previously used to confirm findings at a national level [79].

In section *6.6*, I address the limitations of the synthetic control model with sparse data (step 5) using MenC cases at a regional level. I show the performances of the synthetic control and the STL+PCA methods (section *5.7*) in evaluating the reduction of MenC cases following the 2010 immunization campaign. These reductions are compared with the results by Moraes *et al.* [78].

6.1 Data

I collected laboratory-confirmed MenC cases at a state, regional and national level, aggregated by age and month of disease onset (January 2007 – December 2013) [79] from Brazilian Notifiable Diseases Information System (SINAN) database [134].

I selected cases reported to SINAN with the following classification: meningococcal meningitis (MM), meningococcemia (MCC), or meningococcemia combined with meningococcal meningitis (MCC+MM) [78]. The MenC vaccine-eligible age groups were <1 year and 1–4 years-old. The evaluation period is from December 2011 to December 2013. The infant immunization programme is described in chapter 2, section *2.2.1*, and it is summarized in Table 6.1.

As control diseases, I used the time series of infectious/non-infectious disease cases from the same age groups of the target disease. Here, a large number of controls was available (see the full list of 36 control time-series in Table B.1 in Appendix B). Control time series were identified by the International Classification of Diseases, 10th Revision (ICD-10) code. At a national level, the control diseases are in the age groups <1year and 1-4-year-olds. Due to the availability of data, at a regional and state level, the control diseases are in the age groups <1 year and 12-23-months.

In addition, I used the time series of MenC cases, in age groups not eligible for the immunization program (the time series are reported in Table B.1 in Appendix B).

Meningococcal C cases are in the age groups: 5-9; 10-14; 15-19; 20-39; 40-59-years-old.

In 2010, following an epidemic of MenC disease, a mass vaccination campaign with a catch-up for adolescents in 2010 was implemented in the city of Salvador. Besides, the whole state of Bahia, whose capital is Salvador, anticipated the infant routine immunization programme to February 2010 [135]. For this reason, data from the state of Bahia were excluded from my study [79].

TABLE 6.1: Details on the Meningococcal C vaccination programmes implemented in Brazil

Country	Target time series	Target age groups	Time range	Vaccine schedule	Start of immunization programme	Start of evaluation period	Data source
Brazil*	Monthly MenC cases	<1y and 1-4y	Jan 2007 - Dec 2013	3 doses at age 3, 5, 12-15 m	Nov 2010	Dec 2011	SINAN[134]

m: months old; SINAN: Sistema de Informação de Agravos de Notificação; y: years old.

* Cases from the city of Salvador excluded from the analysis.

6.2 Negative control analysis on non-vaccine-eligible age groups

To evaluate the prediction accuracy of the synthetic control model and time-trend model, described in sections *5.1* and *5.3* respectively, I performed a negative-control analysis (section *5.5*) using MenC cases in age groups not included in the immunization program as targets of the two models. The post-vaccination cases were not used for fitting the models. We expect to see no vaccine impact in these age groups if the no-herd-immunity assumption is not violated (its validity is evaluated at the end of the section and in section *6.5*).

Both models captured the seasonal behavior of MenC cases in all age groups. However, the synthetic control model predicted MenC cases during the post-vaccination period better than the time-trend, as confirmed by a $\sim$ 40% smaller MAE value for 15-19 years old, a $\sim$ 65% smaller MAE for 10-14 year and a $\sim$ 55% smaller MAE for 5-9-year-olds age groups (Figure 6.1, with blue curves representing the synthetic control and red curves the time-trend models).

The time-trends' lack of accuracy led to over-estimated predictions of post-immunization disease in all age groups, with most of the observed cases (black dots) under the fit. The synthetic control always over-performed the time-trend model in terms of MAE, in all the considered age groups (Figure 6.1), and accurately reproduced long-term non-linear trends in the incidence of diseases, such as the decrease in MenC cases among 5–9 and 10–14-year-old in Brazil since 2012 (Figure 6.1 middle and bottom panels).

To avoid biased estimates due to indirect protection of vaccination, I repeated the analysis this time excluding IMD cases in unvaccinated age groups from the set of controls. Results are shown as cyan curves in the left panels of Figure 6.1. I found that counterfactual predictions are coincident in both cases, i.e. including and excluding MenC cases in unvaccinated age groups, for the 15-19-year-old age group. Instead, I observed small, not statistically significant differences between the two counterfactuals in 5–9 and 10–14-year-olds only in the last year of the study (Figure 6.1 middle and bottom panels).

I repeated these analyses using MenC data from the São Paulo state. Again, the synthetic control well reproduced MenC trends in all age groups, outperforming time-trend models in 5-9 and 15-19-years-old age groups according to MAE (Figure C.2 in Appendix C, top, and bottom panels, respectively). The time-trend model failed to adjust for MenC trends in both age groups, leading to increased trends in the post-vaccine period and to wider credible intervals compared to the synthetic control. MenC trends in 10-14-years-old age groups were comparable with both the models, with a small difference in MAE metrics (MAE equal to 4.14 and to 3.70 for the synthetic control and time-trend models, respectively). In this case, the time-trend model did not predict an increasing trend in the post-vaccination period, though the estimates were characterized by greater uncertainty reflected in wider credible intervals (Figure C.2 in Appendix C, middle panel).

To be consistent with the national case study, I also run the synthetic control model excluding MenC cases from the set of controls. Results are in agreement with national performances, with small, not statistically significant differences between the two counterfactuals in 5–9 and 10–14-year-olds only in the last year of the study (Figure C.2 in Appendix C, cyan curves in middle and bottom panels). Counterfactuals in the 15-19-years-old age group are coincident in both cases.

The counterfactuals including or excluding IMD cases in unvaccinated age groups were comparable in both cases in all age groups at national and São Paulo state scales, and no effect of the vaccine was detected (Figure C.3 in Appendix C, panels A and B), thus supporting the validity of my assumption of absence of indirect protection.

6.3 Vaccine impact in Brazil and in São Paulo state

Reported meningococcal C cases declined in both age groups after the introduction of the immunization campaign (the vaccination start dates are highlighted with black vertical lines in Figure 6.2). Since then, the counterfactuals have started to move away from the observed data. The impact was evaluated starting from one year after vaccination (grey dashed vertical lines).

MenC trends were comparable both at a national level and in Sao Pãulo state only (Figure 6.2, top panels showing MenC trends for Brazil, bottom panels for Sao Pãulo state). The synthetic control model correctly captured the seasonal behavior of IMD incidence in both cases.

In Brazil, I measured a 69% (95% CI: 51%; 80%) vaccine impact on MenC cases in infants (aged less than one year). In children aged 1–4 years old, the impact was 64% (95% CI: 55%; 70%). At a national level, vaccine impact estimates confirmed previous findings by Moraes *et al.* (Table 6.2) [78]. In this study, a generalized least square (GLS) method was employed with a first-order autoregressive (AR1) model to assess the impact [78].

For São Paulo state, impact estimates were: 78% (95% CI: 65%; 85%) for <1 year old children, while for 1-4 year-olds: 67% (95% CI: 58%; 74%).

Finally, I also run the time-trend model on the same data (fits and counterfactual predictions are reported in Figure 6.3). Compared to the synthetic control, the time-trend model did not adjust for the increasing trend in the number of cases among the vaccine eligible age groups observed in Brazil during the pre-vaccination period.

After vaccination, counterfactuals predictions with the time-trend model differed from the synthetic control ones, leading to different impact estimates. In Brazil, the time-trend predicted a monotonous increase in MenC incidence, leading to a higher

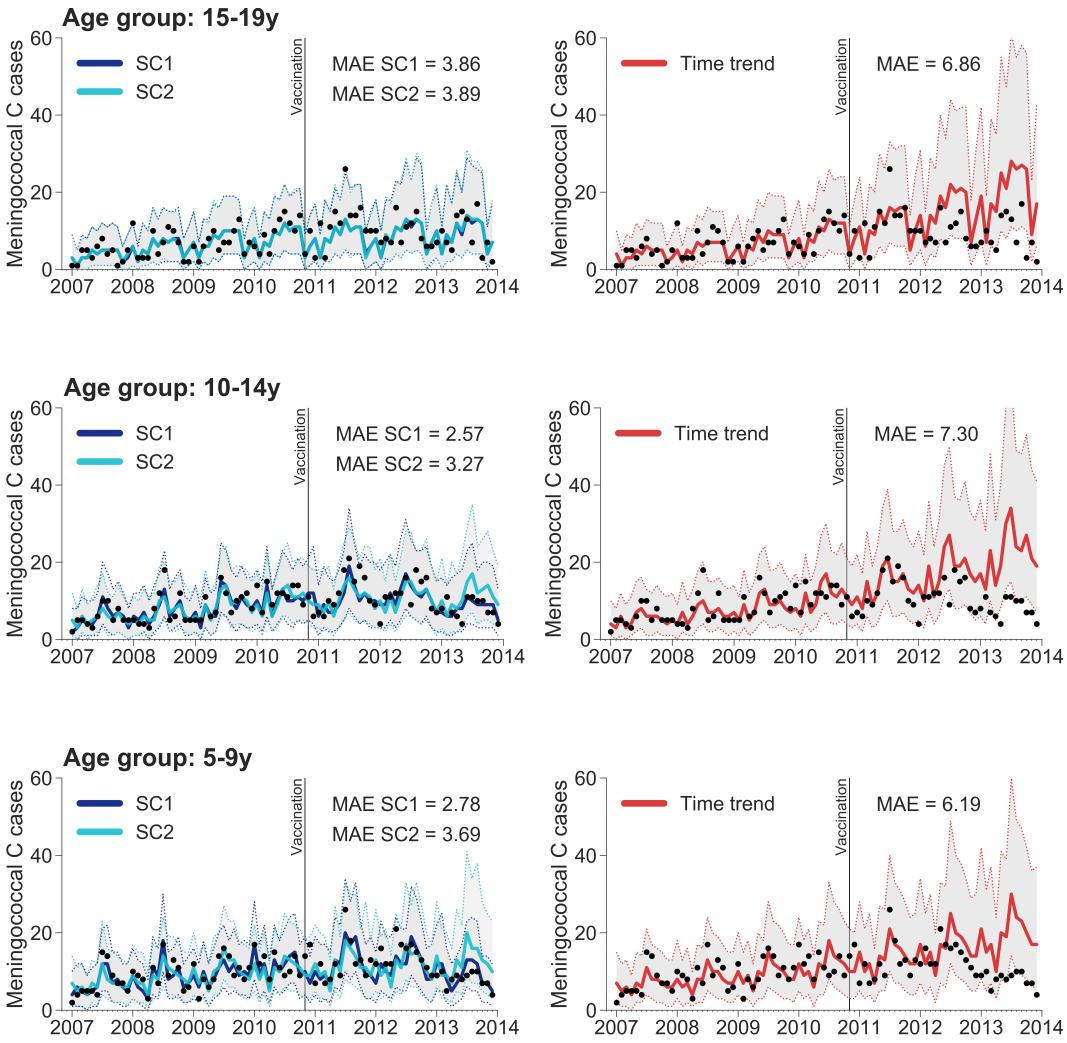

FIGURE 6.1: Predictions generated by synthetic control and time-trend models for MenC (Brazil) disease in the 15–19 (top), 10-14 (middle) and 5-9 (bottom) years old non-vaccinated age groups. In blue, cases predicted with the synthetic control model using all the controls available (SC1) (curve: best estimate; shaded region: 95%CI). In cyan, cases predicted excluding MenC cases in unvaccinated age groups (SC2) (curve: best estimate; shaded region: 95%CI). In red, cases predicted with a time-trend model (curve: best estimate; shaded region: 95%CI). Observed cases are shown as black dots. Vertical black lines divide each plot into two parts: on the left side, data used to fit models; on the right side, data used to evaluate the predictions (MAEs shown).

TABLE 6.2: Vaccine impact estimates in Brazil comparing the synthetic control method with the GLS approach by Moraes *et al*. [78].

Method	Age group	Evaluation period	Impact [95%CI]
SC	<1 y	Dec 2011 - Dec 2013	0.69 [0.51;0.80]
GLS+AR1 [78]	<1 y	Dec 2011 - Dec 2013	0.66 [0.45;0.87]
SC	1-4y	Dec 2011 - Dec 2013	0.64 [0.55;0.70]
GLS + AR1 [78]	1-4 y	Dec 2011 - Dec 2013	0.52 [0.33;0.71]

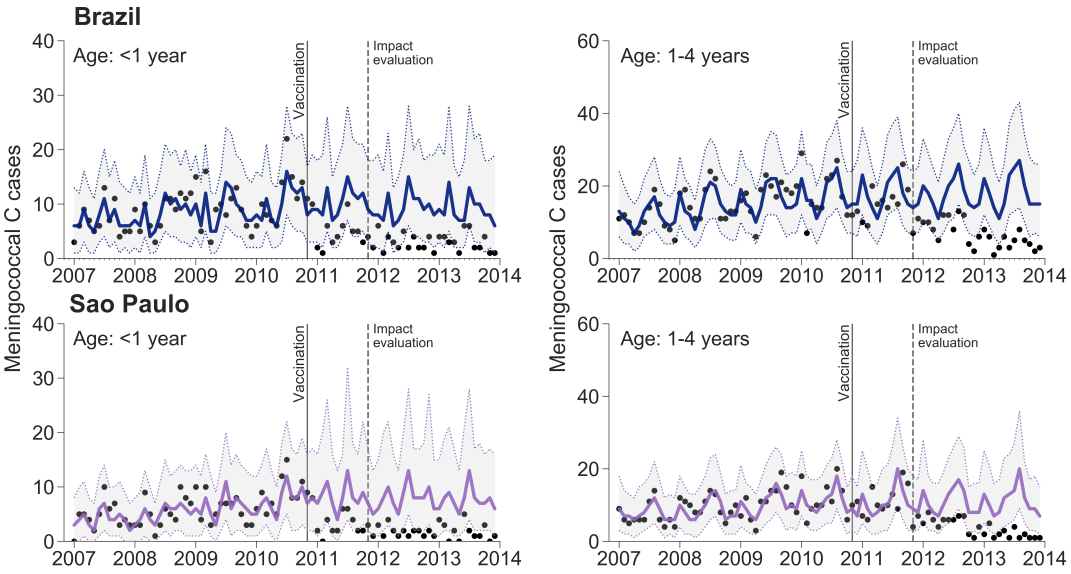

FIGURE 6.2: MenC (Brazil) cases in age groups eligible for vaccination. In blue, cases predicted with the synthetic control model at a national level (top panels), while in purple cases predicted in the Sao Pãulo state (curve: best estimate; shaded region: 95%CI). Observed cases are shown as black dots. Solid black vertical lines indicate the introduction of the vaccination campaign. Dashed grey vertical lines indicate the initial point for measuring impact.

reduction of cases following the vaccination compared to synthetic control. I measured 82% (95% CI: 73%; 88%) and 74% (95% CI: 67%; 83%) vaccine impact in, respectively, infants younger than one year and children 1–4 years old.

A similar increasing trend is observed with São Paulo state (Figure 6.3 bottom plots), with an expected higher reduction of cases in the post-immunization period. Here impact estimates were 87% (95% CI: 78%; 92%) and 79% (95% CI: 69%; 85%) for <1 year-old and 1-4 years-old age groups, respectively.

6.4 Top three selected predictors of meningococcal disease in infants

In this section, I describe the most frequently selected predictors of MenC incidence both at a national and São Paulo state level, according to the probability of inclusion (section *5.2.1*).

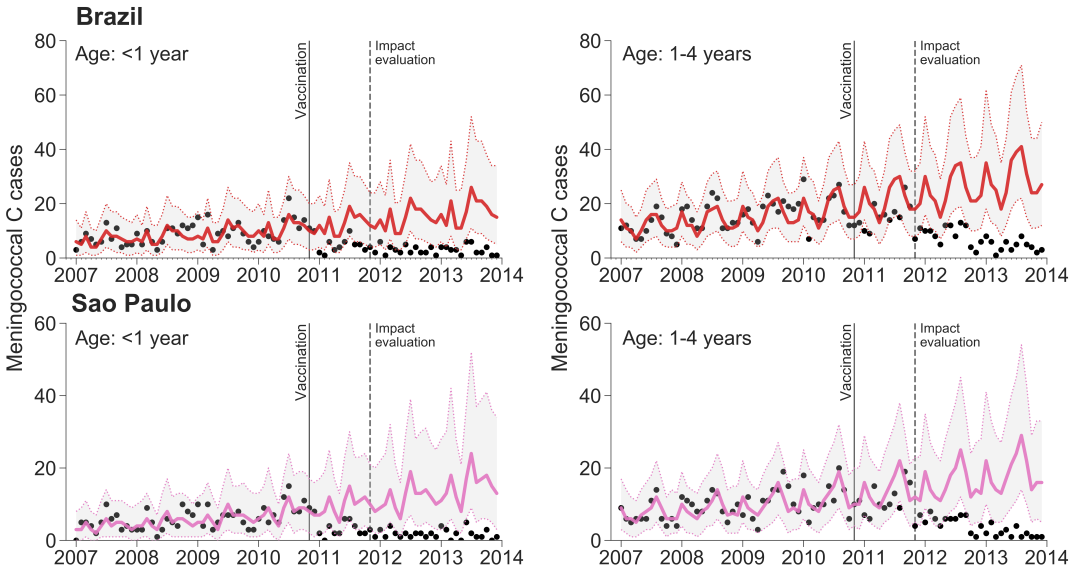

FIGURE 6.3: Time-trend model predicted cases in age groups eligible for vaccination, shown in red for Brazil and in pink for São Paulo state (curve: best estimate; shaded region: 95%CI). Observed cases shown as black dots. Solid black vertical lines indicate the introduction of vaccination campaign. Dashed grey vertical lines indicate the initial point for measuring impact.

Even if the most frequently selected control time series varied by age group, I observed a common general pattern in Brazil, with MenC time-series in non-vaccine-eligible age groups consistently being among the three predictors with the highest probability of being included in the model. The upper panel of Figure 6.4 (*All the controls* panel) displays the three most selected time series to fit pre-immunization MenC cases of age groups <1 and 1–4 years old in Brazil. Among them, I found not only MenC cases in older age groups, but also other infectious and non-infectious disease time series.

The most selected predictor of MenC in the less than one-year-old was "diseases of blood and disorders involving the immune mechanism" (ICD-10 code D50-89, probability of inclusion Prob = 0.57), followed by MenC in adolescents 15–19 years old (Prob = 0.23) and "injury, poisoning and consequences of external causes" (ICD-10 code S00-T98, Prob = 0.19).

For MenC in 1–4-year-old: first "other acute lower respiratory infections" (ICD-10 code J20-J22, Prob = 0.35), then "diseases of the circulatory system" (ICD-10 I00-99, Prob = 0.24) and MenC cases among adults aged 20–39 years (Prob = 0.22). Here, none of the control diseases showed a probability of inclusion higher than 50%, though the model size was still over 1 (Table C.1 in Appendix C).

Conversely, the MenC cases in unvaccinated age groups were not selected to predict meningococcal incidence at the São Paulo state level (Fig. 6.5). The control diseases selected to predict < 1-year-old infants were: "perinatal diseases" (ICD-10 code P00-99, Prob=0.57), "injury, poisoning, and consequences of external causes" (ICD-10 code S00-T98, Prob = 0.48), and "all the controls summed together" (code ach, Prob=0.27). For MenC 1-4 years old children: "Urinary tract infection" (ICD-10 code N39, Prob=0.28), "perinatal diseases" (ICD-10 code P00-99, Prob=0.28) and "injury, poisoning, and consequences of external causes" (ICD-10 code S00-T98, Prob=0.22).

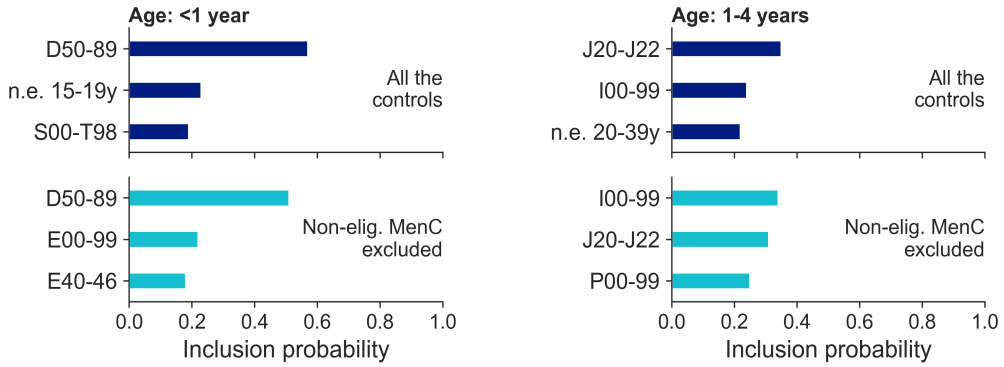

FIGURE 6.4: Top three selected controls with highest probability of inclusion, for <1-year age group and 1-4 years old age group in Brazil (left and right panels, respectively). We report results using all the controls (blue bars) and a subset where MenC cases in non-vaccine-eligible age groups were excluded (cyan bars).
y: years; n.e.: non-eligible age group; *ach*: aggregated variable with all the controls summed together; *E40-46*: malnutrition; *E00-99*: Endocrine, nutritional, metabolic disorders; *J20-J22*: Bronchitis, bronchiolitis and unspecified acute lower respiratory infection; *I00-99*: Diseases of the circulatory system; *P00-99*: Perinatal diseases; *D50-89*: Diseases of blood and blood-forming organs and certain disorders involving the immune mechanism; *S00-T98*: Injury, poisoning and consequences of external causes.

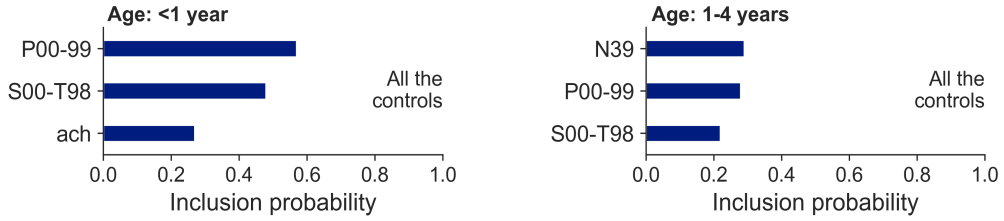

FIGURE 6.5: Top three selected controls with highest probability of inclusion, for <1-year age group and 1-4 years old age group in São Paulo state (left and right panels, respectively). We report results using all the controls (blue bars).
y: years; *ach*: aggregated variable with all the controls summed together; *P00-99*: Perinatal diseases; *S00-T98*: Injury, poisoning and consequences of external causes; *N39*: Urinary tract infection

6.5 Sensitivity of the synthetic control predictions to the exclusion of MenC cases in unvaccinated age groups

MenC controls could be potentially influenced by indirect effects. Removing IMD controls of the same serogroup relaxed my initial assumption, i.e. that indirect effects are negligible (section *5.6*). Hence, I tested if such an assumption was grounded by re-running my analyses using other infectious/non-infectious diseases only as controls. In section *6.2*, I already showed that removing IMD cases from the set of controls did not affect the results, and I repeated the analysis this time with MenC cases in vaccine-eligible age groups used as targets of the synthetic control. I run this

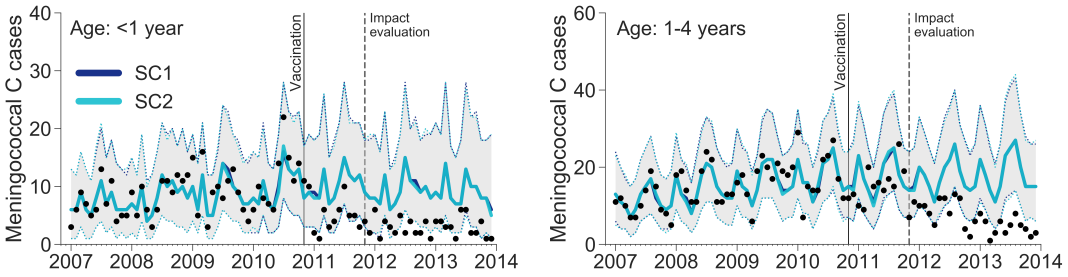

FIGURE 6.6: Predictions generated by the synthetic control model for MenC (Brazil) disease in age groups eligible for vaccination. In blue, cases predicted with the synthetic control method using all the controls available (SC1). In cyan, cases predicted excluding MenC cases in unvaccinated age groups (SC2). Curve: best estimate; shaded region: 95%CI. Observed cases are shown as black dots. Solid black vertical lines indicate the introduction of the vaccination campaign. Dashed grey vertical lines indicate the initial point for measuring impact.

test only at a national level since the IMD cases in unvaccinated age groups were not selected as influential predictors at the São Paulo state level.

Here, reducing the set of controls did not change the goodness of fit (DIC metric computed over the pre-intervention period) and the accuracy of predictions (MAE metric computed over the last three points before and the first point after vaccination) (Figure 6.6 and Table 6.3).

TABLE 6.3: DIC and MAE estimates for each set of controls and age groups eligible for vaccination in Brazil. Best estimates, i.e. lowest DIC and MAE values, are highlighted in bold.

Age group	Set of controls	DIC	MAE
<1 year	*All the controls*	253.92	2.5
	Non-elig. MenC excluded	**253.67**	2.5
1-4 years	*All the controls*	280.19	3.25
	Non-elig. MenC excluded	**279.96**	**2.75**

In both age groups, the counterfactuals predictions were coincident in both cases, with overlapping 95% credible intervals (Figures 6.6). Besides, also impact estimates were robust to the exclusion of IMD cases from the set of controls (Figure 6.7).

Looking at the controls selected with the highest probability, the synthetic control model often selected other respiratory infections. The acute lower respiratory infections (ICD-10 J20-J22) consistently appeared among the top three controls for the 1–4-year-old (Figure 6.4, *Non-elig. MenC excluded* lower panel). J20-J22 refers to bronchitis, bronchiolitis, and other acute lower respiratory infections, including bronchitis and bronchiolitis due to RSV.

6.6 Vaccine impact with regional data

In this section, I tested the performances of the synthetic control and STL+PCA methods on sparse data in the post-vaccine period, i.e. specifically from December 2011 to December 2013 (section 5.7). I used MenC cases at a regional level, which are

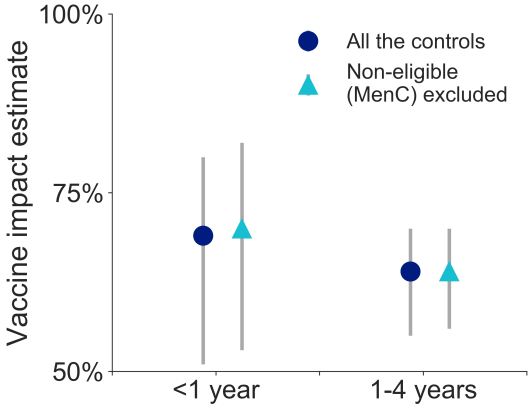

FIGURE 6.7: Vaccine impact estimates when using the full set of controls (blue dots) and a subset where non-vaccine-eligible age groups were excluded (cyan triangles). 95%CIs are shown as grey lines.

sparse in all the regions, with the only exception of the Southeast region. The North region was excluded from the analysis due to several months with zero occurrences.

I found significant reductions of MenC cases using the synthetic control method in the Center-west and Southeast regions for age group $<$ 1-year-old with vaccine impact estimates equal to 83% (95% CI: 40%; 93%) and 77% (95% CI: 49%; 88%), respectively (Figure 6.8, top panel). In infants 1-4 years old, the synthetic control predicted a significant impact of vaccine with estimates of 79% (95% CI: 45%; 93%) for Center-west region, 67% (95% CI: 57%; 74%) for Southeast region and 89% (95% CI: 50%; 99%) for South region.

Results were comparable when running the STL+PCA approach on 1-4 years old infants (Figure 6.8, middle panel). Instead, differently from the synthetic control, the STL+PCA found a significant reduction in the South region for $<$ 1-year-old infants with 89% vaccine impact (95% CI: 63%; 96%).

MenC reduction of cases in the Northeast region was non-significant with both approaches and age groups.

The bottom panel in Figure 6.8 shows the impact estimates reported in the work by Moraes *et al.* [78]. In this study, unidentified IMD serogroup cases were distributed according to the proportion of cases with identified serogroups [78]. The study population was indeed larger compared to the data used in my research, composed not only of confirmed cases of IMD C serogroup but also of "unidentified serogroup" cases, attributable to serogroup C on the basis of proportional distribution [78].

In this study, impact estimates were significant in all regions and age groups, except for the South region in 1-4-year-old infants [78]. Point estimates were in agreement with the synthetic control and STL+PCA methods, while 95% credible intervals varied between the regions.

Finally, I repeated the analysis aggregating data on a quarterly basis, to increase the number of occurrences. In general, I did not find a visible improvement (Figure C.4 in Appendix C).

6.7 Discussion and conclusions

In this chapter, I have presented the first part of the results achieved during my PhD programme.

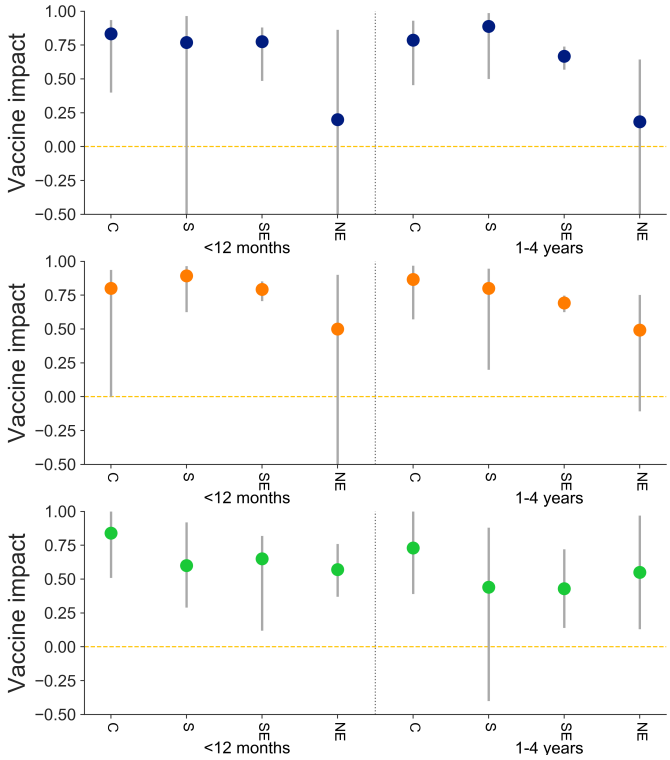

FIGURE 6.8: Vaccine impact estimated with the synthetic control model (top panel), the STL+PCA model (middle panel) and previously published estimates from Moraes *et al* [78] (bottom panel), comparing the observed and predicted number of cases in 2011-2013 by age group and region. 95%CIs are shown as grey lines. Abbreviations: C, center-west; S, south; SE, southeast; NE, northeast.

First, I showed that the synthetic control method outperformed a simple time-trend adjustment in a negative control analysis (i.e., where unvaccinated age groups were used as targets of the synthetic control model). The synthetic control was able to adjust for several unexplained trends, outperforming the time-trend model in terms of prediction accuracy in all the considered age groups (MAE estimates were always lower).

Despite the limitations of using observational data discussed in chapter 3, the synthetic control detected significant reductions of MenC cases in the two considered age groups, with vaccine impact estimates in agreement with previous publications [78]. Besides, results were confirmed running the model on MenC cases of São Paulo state only, which has already been used as a proof of the results' robustness due to its reliable surveillance system [79].

Among the control diseases with the highest probability of being included in the model, MenC cases in unvaccinated age groups and respiratory diseases, such as diseases with the ICD-10 code *J20-J22*, were selected. Since the former controls could be influenced by indirect effects of the vaccine, thus underestimating the impact estimates, I re-run the synthetic control method excluding MenC cases from the set of controls. Results were robust to the exclusion of IMD cases in unvaccinated age groups. It should be highlighted that the MenC campaign introduced in 2010 in Brazil concerned only infants and not adolescents (which usually harbor the highest rates of meningococcal carriage) [79]. This finding is in agreement with Moraes *et al.* and Andrade *et al.*, that found no indirect protection from the vaccination campaign

[78, 79]. Both authors suggested vaccinating adolescents to possibly observe indirect effects on non-vaccinated age groups (to be noted: MenC vaccination targeting adolescents has been introduced starting from year 2017). Besides, a catch-up campaign was introduced in Salvador for the age group 10–24 years showing evidence of herd immunity effects [30]. Data from Salvador were excluded from our study.

Another point which should be discussed is whether the 5–9 age group could be affected by the vaccination campaign. I have performed a few analyses excluding age group 5–9 from the set of controls and I found small, non-significant differences (I highlighted this point where I tested the SC analysis excluding MenC cases in non-vaccine-eligible age groups from the set of controls). Also, to my understanding, only 1 month of 5-year-olds could have been effectively vaccinated (e.g., those born in August 2009 and aged 5 in December 2013, which represented the last time point of my analysis). Considering the high heterogeneity of vaccine coverage (which varied in Brazil's regions, with a few regions having low coverage in the first year of the program), I decided that I could safely use this age group as a control.

Finally, I investigated the reductions of MenC cases following the vaccination programme at a regional level. In certain regions, the MenC cases were extremely low and the synthetic control failed to detect significant impact estimates. Since the sparsity of case counts affected mostly the outcome and the MenC cases used as controls, the performances of the synthetic control and the STL+PCA methods were comparable, with some small exceptions (for example, in the South region the STL+PCA found significant vaccine impact estimates in the $<$1-year-old age group in contrast to the synthetic control). A possible solution could be to take into account the proportion of unknown serogroups to increase the MenC occurrences as done by Moraes *et al.* [78], whereas I found no significant improvement when aggregating the data quarterly.

In this chapter, I have shown that the synthetic control is a robust tool to evaluate the impact of the MenC vaccination campaign in Brazil, despite the challenges given by observational data and the differences in surveillance systems at a national, regional and state levels. Besides, the synthetic control model was able to detect several unexplained trends, while on the contrary, a simple time-trend adjustment failed. Even if the vaccination programme could induce indirect protection in the unvaccinated population, the synthetic control provided consistent results relying on other infectious/non-infectious control diseases only. Finally, the vaccine impact estimates were robust to the exclusion of MenC cases in unvaccinated age groups from the set of controls.

Chapter 7

The impact of the 2015 MenB campaign in England

This chapter presents the results of the meningococcal B vaccination programme in England. I followed the procedure described in chapter 5, summarized in section *5.8*.

In section *7.1*, I discuss the MenB data collected to perform my analysis. Section *7.2* deals with the quality of the synthetic control's predictions compared to a simple time-trend model (step *1*, *Negative control analysis*, in chapter 5, section *5.8*). Sections *7.3*, *7.4*, *7.5*, discuss in order: vaccine impact estimates from the synthetic control and time-trend models and from a previously published study by Ladhani *et al.* [85]; the most influential MenB predictors in infants; and a sensitivity analysis (steps *2*,*3*,*4* in chapter 5, section *5.8*).

Finally, section *7.6* investigates the performances of the synthetic control model aggregating information on an annual level.

7.1 Data

I retrieved MenB cases from the Public Health England (PHE) national surveillance system website [136]. I collected quarterly cases from the last quarter of 2011 to the first quarter of 2019; before this time period data were not stratified by age group.

MenB time series by quarter of year of subjects eligible for the national immunization program were aggregated in two wide age groups (<1 and 1–4 years-old). The previous vaccine impact analyses considered MenB annual cases in finer age groups (i.e., 18–51 weeks and 1-year old age groups) [85, 91]. To be consistent with previous analyses, I derived quarterly MenB cases in the 18–51-weeks and 1-year-old age groups, assuming that the proportion of cases by age group was maintained at a quarterly level (e.g., if cases in the 18–51 weeks age group were 80% of the cases in the <1-year old age group, I used the same proportion to derive time series on a quarterly basis).

The details on the infant immunization programme are described in chapter 2, section *2.2.1*, and summarized in Table 7.1.

As control diseases, I used the time series of infectious/non-infectious disease cases from the same country and age groups of the target disease. Here, the set of available control time series with quarterly data was smaller compared to Brazil. In addition, I used the time series of MenB cases, in age groups not eligible for the immunization program (the time series are reported in Table B.2 in Appendix B).

Meningococcal B cases are in the age groups: 5-9; 10-14; 15-19; 20-24; 25-44; 45-64 and 65+ years-old. All the other control diseases are in the age groups <1 and 1-4-years-old.

TABLE 7.1: Details on the Meningococcal B vaccination programme implemented in England.

Target time series	Target age groups	Time range	Vaccine schedule	Start of immunization programme	Start of evaluation period	Data source
Quarterly MenB cases	18-51w and 1y*	Q4 2011 - Q1 2019	3 doses at age 2, 4, 12 m	Sep 2015	Q4 2016	PHE [136]

*After data augmentation, as described in methods section.

I also collected MenB incidence data on an annual level, from epidemiological years 1998/99 to 2017/18 [136]. Data prior to epidemiological years 2009/10 were available from PHE archives [137] and refer to England + Wales meningococcal B cases in the <1-year and 1-year-old age groups. Again, I have derived MenB cases in age groups 18-51-weeks and 1-year-old to be able to compare the final estimates with both quarterly data and previous publications. The set of control diseases publicly available on an annual basis in the age groups <1 and 1-4-years-old was significantly larger compared to the quarterly data (Table B.3 in Appendix B).

7.2 Negative control analysis on non-vaccine-eligible age groups

I performed a negative-control analysis on IMD cases in age groups not included in the immunization program applying both the synthetic control and a simple before-after time-trend models (step 1, section *5.5*).

The synthetic control model predicted IMD cases during the post-vaccination period better than the time-trend, as confirmed by a $\sim$ 40% smaller MAE value for 15-19 years old, a $\sim$ 20% smaller MAE for 10-14 years and 5-9 years old age groups (see Figure 7.1, with blue curves representing the synthetic control and red curves the time-trend models).

In the 15-19 and 10-14 years old age groups, the time-trends' lack of accuracy led to under-estimated predictions, with most of the observed cases (black dots) far away from the predicted curve. MenB cases in 5-9 and 10-14 years old age groups were mostly explained by their seasonal behavior (middle and bottom plots in Figure 7.1). The synthetic control still over-performed the time-trend model in terms of MAE in both the considered age groups.

In general, the synthetic control accurately reproduced long-term non-linear trends in the incidence of disease, such as the increase in MenB cases among 15-19 years old reported in England during the entire immunization period, compared to negative trends before the immunization (Figure 7.1, top panel).

Since MenB cases in unvaccinated age groups could be influenced by indirect protection of the vaccine, I repeated the analysis, this time excluding IMD cases in unvaccinated age groups from the set of controls. Results are shown as cyan curves in the left panels of Figure 7.1. Counterfactual predictions are coincident in both cases, i.e. including and excluding IMD cases in unvaccinated age groups, for all age groups under study (blue and cyan curves). Besides, impact estimates were non-significant in both age groups (Figure C.5 in Appendix C), confirming the robustness of my assumption of the absence of indirect protection.

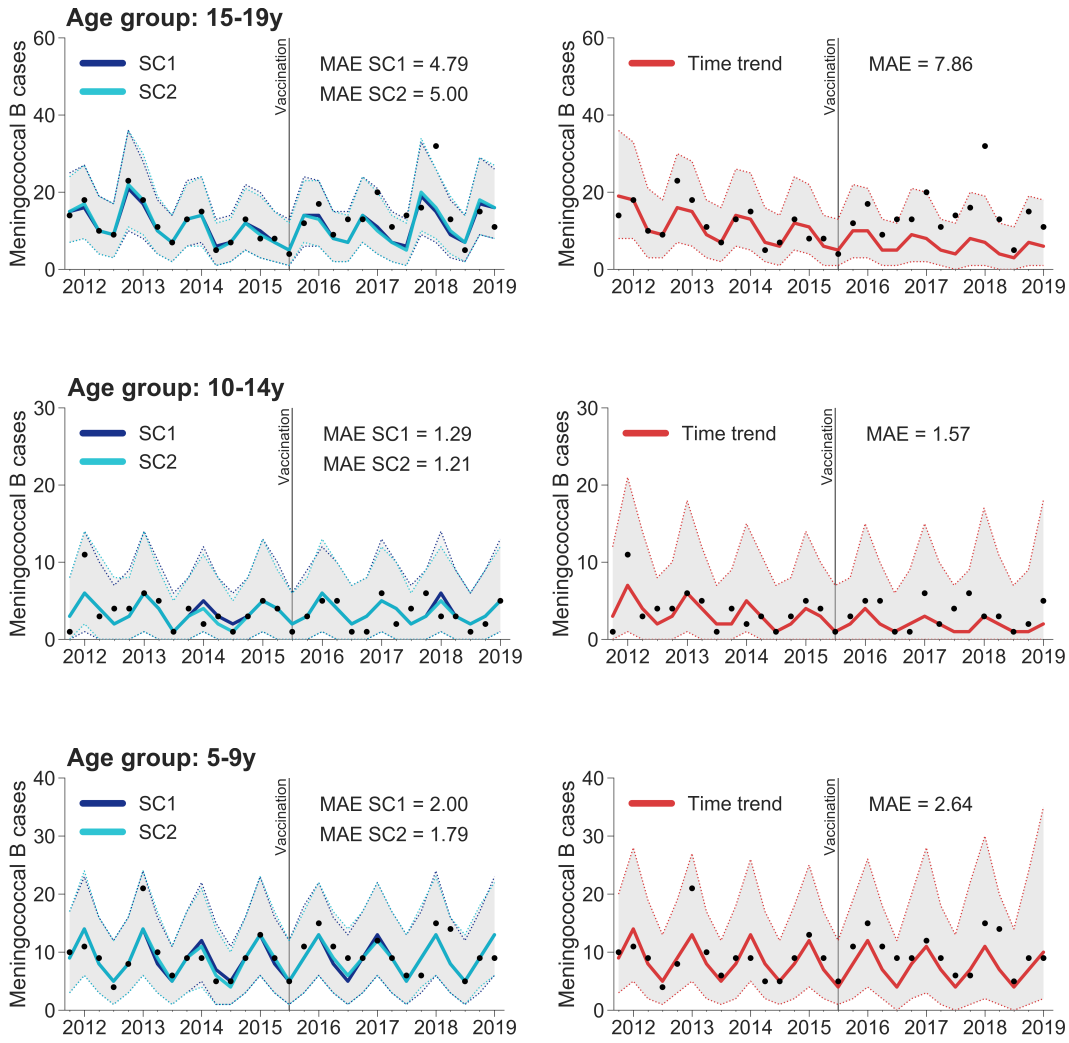

FIGURE 7.1: Predictions generated by synthetic control and time-trend models for MenB (England) disease in the 15–19 (top), 10-14 (middle) and 5-9 (bottom) years old non-vaccinated age groups. In blue, cases predicted with the synthetic control method using all the controls available (SC1) (curve: best estimate; shaded region: 95%CI). In cyan, cases predicted excluding MenB/MenC cases in unvaccinated age groups (SC2) (curve: best estimate; shaded region: 95%CI). In red, cases predicted with a time-trend model (curve: best estimate; shaded region: 95%CI). Observed cases are shown as black dots. Vertical black lines divide each plot into two parts: on the left side, data used to fit models; on the right side, data used to evaluate the predictions (MAEs shown).

7.3 Vaccine impact in England

The synthetic control model well reproduced MenB time-series of cases before the immunization, even in the presence of non-trivial patterns, such as a trend inversion between 2014 and 2015 in both age groups (Figure 7.2).

Reported MenB cases sharply declined in both age groups after the introduction of the immunization campaign (the vaccination start dates are highlighted with black vertical lines in Figure 7.2). Since then, the counterfactual predictions are detached from the observed data.

I evaluated the impact one year after vaccination (grey dashed vertical lines in Figure 7.2). I estimated a 75% reduction (95% CI: 69%; 79%) in infants aged 18-51 weeks, while in the 1-year-old age group the impact was 72% (95% CI: 65%; 79%).

These estimates were in good agreement with the study by Ladhani *et al.* [85]. In this latter case, the impact was assessed with a Poisson regression model on incidence rate ratios. To adjust for changes unrelated to the vaccine, the authors used MenB cases in non-eligible subjects, i.e. infants of <5 years of age, as controls. In table 7.2, the two approaches are compared, with estimates overall in agreement although there is a little difference due to the aggregation at a quarterly and annual level.

TABLE 7.2: Vaccine impact estimates in England comparing the synthetic control method with the approach by Ladhani *et al.* [85].

Method	Age group	Evaluation period	Impact [95%CI]
SC	18-51w	Q4 2016 - Q1 2019	0.75 [0.69;0.80]
SC	1y	Q4 2016 - Q1 2019	0.72 [0.65;0.79]
SC	18w to 1y	Q4 2017 - Q3 2018	0.76 [0.70;0.80]
Poisson regression [85]	18w to 1y	Depends on age group[1]	0.75 [0.64;0.81]

I also run the time-trend model on the same vaccine eligible age groups (fits and counterfactual predictions are reported in Figure 7.3). The time-trend model failed to adjust for the decreasing trend in the number of cases among the vaccine eligible age groups observed in England before 2015.

Counterfactuals predicted with the time-trend model differed from the synthetic control ones, leading to different impact estimates. The time-trend model provided a lower vaccine impact estimate when compared to the synthetic control. In the 18–51 weeks age group, the impact was 59% (95% CI: 28%; 74%); in 1-year-old children, 48% (95% CI: -37%; 76%).

7.4 Top three selected predictors of meningococcal disease in infants

MenB time-series of cases in non-vaccine-eligible age groups were consistently among the three predictors with the highest probability of being included in the model.

As shown in Figure 7.4, *All the controls* upper panel, MenB incidence in the non-vaccine-eligible age groups was predominantly selected among the top controls. In particular, MenB cases in the three-years-old age group were selected with a probability Prob = 0.97 to predict MenB in one-year-old children.

[1] **18-51w**: Sep 2016 - Aug 2018; **1y**: Sep 2017 - Aug 2018

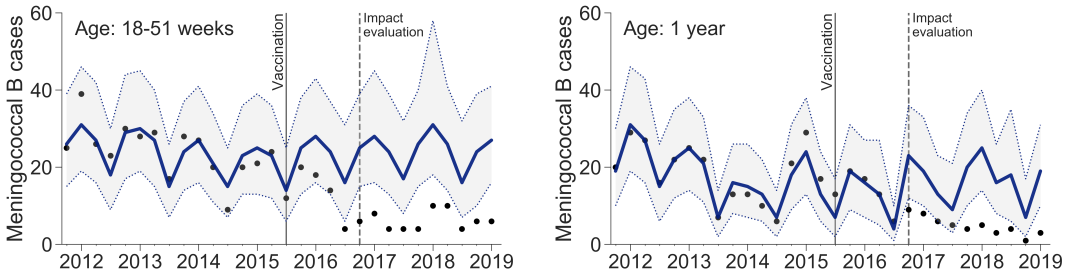

FIGURE 7.2: MenB (England) cases in age groups eligible for vaccination. In blue, cases predicted with the synthetic control method (curve: best estimate; shaded region: 95%CI). Observed cases are shown as black dots. Solid black vertical lines indicate the introduction of the vaccination campaign. Dashed grey vertical lines indicate the initial point for measuring impact.

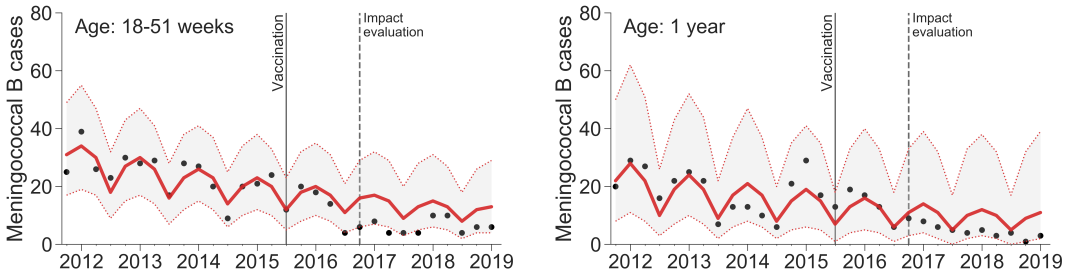

FIGURE 7.3: Time-trend model predicted MenB (England) cases in age groups eligible for vaccination, shown in red (curve: best estimate; shaded region: 95%CI). Observed cases shown as black dots. Solid black vertical lines indicate the introduction of vaccination campaign. Dashed grey vertical lines indicate the initial point for measuring impact.

For the 18–51 weeks age group, posterior inclusion probabilities were lower than 50%. The most selected control time series was MenB cases in 15–19 year-olds with Prob = 0.23. Also, two childhood infectious diseases, measles, and mumps were selected among the top controls for the 18–51 weeks age group and the 1-year-old, respectively, but with a relatively lower probability (respectively, Prob = 0.13 and Prob = 0.08).

7.5 Sensitivity of the results to the exclusion of MenB cases from the controls

MenB cases in the 3-years-old age group had almost 100% of the probability of being included in the model. It is a natural question to understand whether removing such an age group could influence the results. To be even more general, I removed all the MenB cases in unvaccinated age groups from the set of controls and re-run the synthetic control model (section *5.6*). In section *7.2*, I already proved that removing IMD cases from the set of controls did not bias the results.

Here, reducing the set of controls did not change the goodness of fit (DIC metric computed over the pre-intervention period) and the accuracy of predictions (MAE computed over the first 4 points post-vaccination assuming the vaccination was not

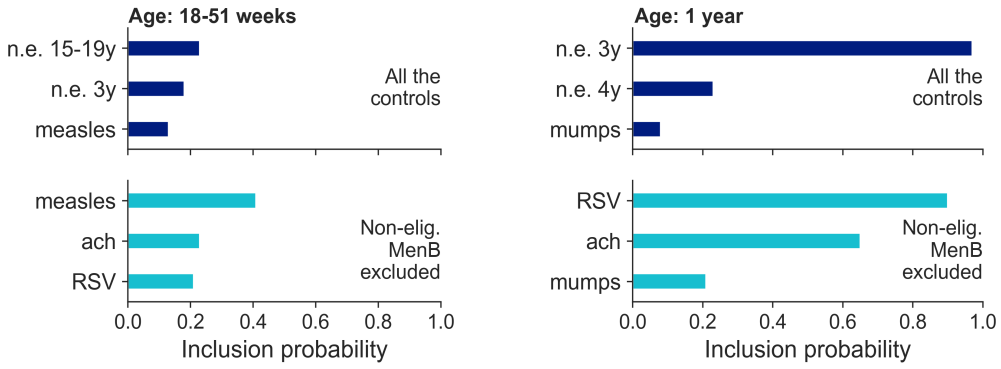

FIGURE 7.4: Top three selected controls with highest probability of inclusion, for 18-51w age group and 1-year-old age group in England (left and right panels, respectively). We report results using all the controls (blue bars) and a subset where MenB cases in non-vaccine-eligible age groups were excluded (cyan bars).
y: years; *n.e.*: non-eligible age group; *ach*: aggregated variable with all the controls summed together; *RSV*: Respiratory Syncytial Virus.

effective yet) in the 18-51-weeks age group (Figure 7.5, left panel and Table 7.3, top row).

Instead, in the 1-year-old age group, excluding MenB cases led to visible worsen DIC and MAE estimates (Table 7.3 bottom row). This difference was reflected in the counterfactuals predictions (Figure 7.5 right panel). Nevertheless, in both age groups, impact estimates were robust, with almost coincident best estimates and overlapping 95%CIs (Figure 7.6).

TABLE 7.3: DIC and MAE estimates for each set of controls and age groups eligible for vaccination in England.

Age group	Set of controls	DIC	MAE
18-51 w	*All the controls*	**98.27**	**3**
	Non-elig. MenB excluded	98.54	3.67
1 yr	*All the controls*	88.26	**0.75**
	Non-elig. MenB excluded	95.22	3

Looking at the controls selected with the highest probability, the synthetic control model often selected other respiratory infections or airborne diseases. Specifically, I found that measles and RSV were frequently selected as predictors of meningococcal disease incidence (Figure 7.4, *Non-elig. MenB excluded* lower panel). In particular, RSV incidence was characterized by a probability of inclusion Prob = 0.90 to predict MenB in the 1-year old.

7.6 Estimation of vaccine impact with annually aggregated data

In this section, I evaluated the performances of the synthetic control model on vaccine-eligible age groups using annual data.

The synthetic control model was able to adjust for the decreasing long-term trend observed in both age groups starting from the epidemiological year 2000/01 (Figure 7.7). In the 18-51-weeks age group, the model precisely reproduced the small

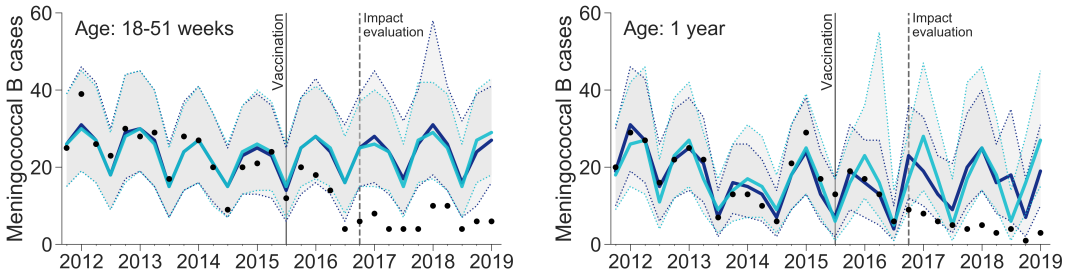

FIGURE 7.5: Predictions generated by the synthetic control model for MenB (England) disease in age groups eligible for vaccination. In blue, cases predicted with the synthetic control method using all the controls available (SC1). In cyan, cases predicted excluding MenB/MenC cases in unvaccinated age groups (SC2). Curve: best estimate; shaded region: 95%CI. Observed cases are shown as black dots. Solid black vertical lines indicate the introduction of the vaccination campaign. Dashed grey vertical lines indicate the initial point for measuring impact.

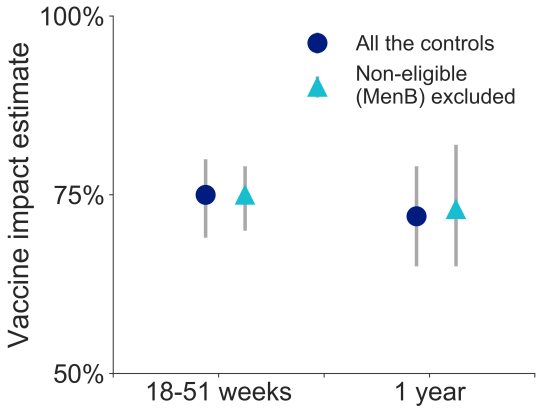

FIGURE 7.6: Vaccine impact estimates when using the full set of controls (blue dots) and a subset where non-vaccine-eligible age groups were excluded (cyan triangles). 95%CIs are shown as grey lines.

increasing trends observed, for example between 1998 to 1999 or between 2010 to 2011. The same precision is observed with 1-year-old age group, except for the last two points before the vaccination start, which are on the edge of the 95% credible intervals (Figure 7.7 right panel).

Both counterfactuals are characterized by a sharp increase, with the observed meningococcal cases out of the 95% credible intervals for all three years after vaccination for age group 18-51-weeks, and for the last two years for the 1-year-old age group (this age group was not vaccinated in 2015/16).

The reduction of MenB cases in the 18-51-weeks age group was 70% (95% CI: 62%, 74%), while for the 1-year-old age group I found a reduction of 59% (95% CI: 45%, 66%). These estimates were slightly lower compared to the results obtained using quarterly data.

MenB cases in unvaccinated age groups were the only controls selected to predict both vaccine-eligible age groups (Figure 7.8, *All the controls* upper panel).

In particular, MenB cases in the 15-19-years-old age group were included in the model with a probability Prob=0.78 to predict 18-51-weeks age group, followed by

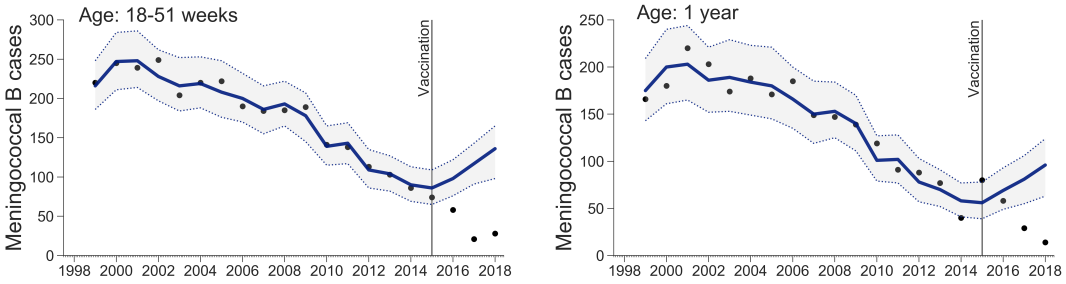

FIGURE 7.7: MenB (England) cases in age groups eligible for vaccination with annual data, starting from the epidemiological year 1998/99 to 2017/18. In blue, cases predicted with the synthetic control method (curve: best estimate; shaded region: 95%CI). Observed cases are shown as black dots. Solid black vertical lines indicate the introduction of the vaccination campaign. Dashed grey vertical lines indicate the initial point for measuring impact.

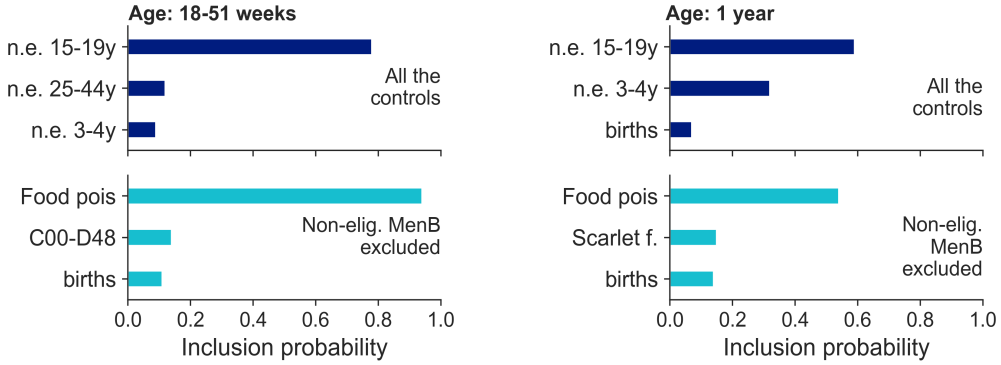

FIGURE 7.8: Top three selected controls with highest probability of inclusion, for 18-51w age group and 1-year-old age group in England using annual data (left and right panels, respectively). We report results using all the controls (blue bars) and a subset where MenB cases in non-vaccine-eligible age groups were excluded (cyan bars). *y*: years; *n.e.*: non-eligible age group; *births*: average number of births in the time period considered (1998-2018); *Food pois.*: Food poisoning; *Scarlet f.*: Scarlet fever; *C00-D48*:Neoplasms.

MenB cases in 25-44-years-old (Prob=0.12). Similarly, in the 1-year-old age group I found MenB cases in the 15-19-years-old age group with a probability Prob=0.59, and then MenB cases in 3-4-years-old (Prob=0.32). Since MenB cases in 15-19 years old could be responsible for the sharp increase observed post-immunization in both age groups, I repeated the analysis excluding MenB cases from the set of controls.

In the 18-51-weeks age group, the curves were almost coincident in the pre-intervention period, while in the post-intervention the curve predicted excluding MenB cases from the controls did not show the same sharp increase. The reduction of cases was indeed lower, with an impact of 60 % (95% CI: 45%, 74%).

Instead, a significant difference is visible in 1-year-old age group (Figure 7.9). Here, removing the MenB cases from the set of controls led to wider 95% credible intervals which included the observed points in the post-intervention period. Consequently, the impact estimate was non significant (VI=52%, 95% CI: -110 %, 72%).

For consistency with section 7.5, I reported the control diseases that were most frequently selected in the absence of MenB cases from the set of controls (Figure 7.8,

Non-elig. MenB excluded lower panel).

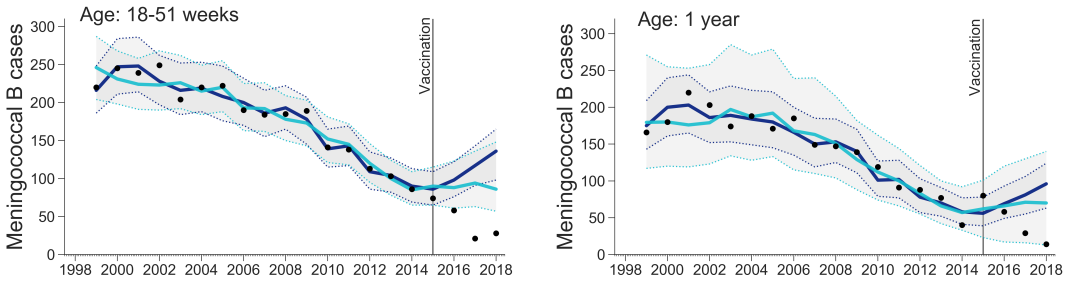

FIGURE 7.9: Predictions generated by the synthetic control model for MenB (England) disease in age groups eligible for vaccination from the epidemiological year 1998/99 to the epidemiological year 2017/18. In blue, cases predicted with the synthetic control method using all the controls available (SC1). In cyan, cases predicted excluding MenB/MenC cases in unvaccinated age groups (SC2). Curve: best estimate; shaded region: 95%CI. Observed cases are shown as black dots. Solid black vertical lines indicate the introduction of the vaccination campaign. Dashed grey vertical lines indicate the initial point for measuring impact.

7.7 Discussion and conclusions

Along with this chapter, I have shown the second part of my results, concerning the MenB campaign in England.

First, the synthetic control method outperformed a simple time-trend adjustment in terms of prediction accuracy in a negative control analysis. Despite the limited number of time-points in the pre-intervention period and the few control diseases available, the synthetic control precisely adjusted for supra-seasonal trends, with lower MAE estimates in all the considered age groups (Figure 7.1).

The synthetic control method was also able to adjust for the decreasing trend observed in vaccine-eligible age groups, with vaccine impact estimates in agreement with a previously published study [85]. In this study, the incidence of MenB cases in unvaccinated infants (i.e., <5 years of age), was used as a control to adjust for trends unrelated to vaccination [85]. I found consistent results, with MenB cases in unvaccinated age groups among the three predictors of meningococcal incidence. In particular, MenB cases in 3-years-old age group were selected with a probability close to 1 to predict infants of 1-year-old (Figure 7.4, *All the controls* upper panel).

As for the time-trend model, I would expect consistent results with the synthetic control if there were no unexplained trends in the data. The time-trend model, however, was unable to adjust for this observed decreasing trend, thus leading to slightly different impact estimates in both age groups (Figure 7.3) with non-significant estimate in 1-year-old infants.

Despite no evidence yet of the impact of MenB vaccine on the carriage and also considering that children do not play a key role in transmission, I could not exclude, however, indirect protection from the vaccine. For this reason, I repeated my analysis excluding this time the MenB cases in unvaccinated age groups.

The set of available infectious/non-infectious diseases was limited (Table B.2 in Appendix B), thus leading to lower performances both in terms of goodness of fit (i.e., DIC) and prediction accuracy (i.e., MAE) in the 1-year-old age group. Despite

these differences, impact estimates were robust to the exclusion of MenB cases in unvaccinated age groups. Instead, I observed no difference in 18-51-weeks age group as expected, since the observed MenB cases were explained by the intercept and the seasonal component only and none of the control diseases had a probability of inclusion higher than 50%.

In this circumstance, the synthetic control model failed to identify an appropriate set of controls (Figure 7.4) and one of the possible explanations could be the low number of both control diseases and pre-vaccine time points [131]. Here, however, the synthetic control model was still able to produce a reliable counterfactual with only the intercept and seasonal components (Figure 7.2, left panel).

I found interesting correlations between RSV incidence and MenB cases in the 1-year-old age group. Associations between RSV incidence and meningococcal disease have been previously highlighted in epidemiological studies [47].

I then assessed the vaccine impact estimates with annually aggregate data. In this case, in addition to MenB cases in the unvaccinated age group, I was able to collect a larger number of control diseases (Table B.3 in Appendix B). I investigated whether the loss of information due to the annual aggregation data was balanced by the larger number of control diseases compared to the quarterly data case study. Here, the synthetic control model accurately reproduced the behavior of MenB incidence in both age groups in the pre-vaccine period, including the sharp decrease observed starting from the year 2000/01 (Figure 7.7).

MenB cases in unvaccinated age groups dominated in predicting IMD incidence in infants, in particular MenB cases in 15-19-years old. In addition, in both age groups, at least one control was selected with a probability of inclusion higher than 50% (Figure 7.8, *All the controls* upper panel). Removing these controls led to identical conditions in MenB cases in the 18-51-weeks age group. Instead, in the 1-year-old age group, the two counterfactuals differed significantly, with larger credible intervals and non-significant impact estimates when removing MenB cases from the set of controls (Figure 7.9). In general, aggregating the data on an annual basis worsened the performances of the synthetic control model, with lower vaccine impact estimates and larger credible intervals compared to the evidence suggested by the quarterly data aggregation and the study by Ladhani *et al.* [85]. Furthermore, the MenB cases in 15-19 years old age group were heavily weighted in the model, and removing them resulted in no vaccine effect in 1-year-old age group, again contradicting previous results.

Several limitations could influence the results presented in this chapter. First, the quality of the quarterly and annual data collected, since these were obtained combining different sources (*Data* section). Secondly, I have made numerous assumptions to obtain MenB cases in finer age groups, specifically 18-51-weeks and 1-year-old age groups, and this process could affect the reliability of the results. Third, I was able to collect a limited set of both control time series and time points pre-vaccination, which could hamper a correct adjustment of unexplained meningococcal B trends unrelated to vaccination.

Despite these limitations, the synthetic control proved to correctly adjust for several unexplained trends both with annual and quarterly data, for example, the sharp decrease in incidence observed in infants starting from the year 2000/01. Besides, the synthetic control model obtained higher prediction performances compared to a simple time-trend model. Finally, the vaccine impact estimates were in agreement with previously published studies [85], with the additional advantage that controls were chosen with data-driven methods.

Chapter 8

Conclusions

My doctoral research was motivated by the challenges in evaluating the true impact of public health interventions. These challenges have been deeply detailed in chapters 2 and 3.

In general, there are several approaches commonly employed to evaluate vaccine impact [132], as presented in chapter 3, section *3.3*. These methods rely on different assumptions about the underlying trend in the data, how they adjust for trends unrelated to vaccination, and how many years to employ before and after vaccination [103, 132]. Since there is not a ground-truth estimate of the vaccine impact under study against which the different models can be compared, there is no way to definitely determine which method is the most suitable [132].

In my research study, I focused my attention on meningococcal vaccines. In this context, low disease incidence, unknown risk factors, and unpredictable supra-seasonal trends could hamper an objective, unbiased and rapid assessment of vaccine impact.

Here, I have shown that synthetic controls 1) adjust for non-trivial changes in IMD incidence; 2) outperform a simple before-after time trend model in a negative control analysis; 3) generalize approaches where controls are arbitrarily selected by the analyst, and 4) do not require the use of controls of the same disease, which may not be available, sufficiently common, or even inappropriate because of indirect effects.

I re-analyzed data from two large infant immunization campaigns in Brazil (against MenC disease) and in England (against MenB). The two settings differed in many aspects as pointed out in chapter 2, including meningococcal serogroup, vaccine type, disease seasonality patterns, age, and socio-demographic features of the population, and public health system. Besides, I relied on observational data collected from national public health authorities, which could lack in reporting fidelity and could undergo substantial modifications during time (details in the *Data* section in both chapters 6 and 7). Also, in England, the data were aggregated on a quarterly basis and the time-points available to build the fit were $\sim 1/3$ compared to Brazil. Nevertheless, the synthetic control method was in good agreement with the original studies, where vaccine impact was assessed employing two different methodologies [78, 85].

I tested the performances of a time-trend model, where disease incidence pre- and post-vaccination was compared (chapter 5, section *5.3*) [3]. I ran both the synthetic control and the time-trend models on non-vaccine-eligible age groups, in a negative control setting. In this case, since I assumed indirect effects negligible, I expected my predictions to be close to the observed points. I found that the time-trend approach failed to explain unexpected changes in trends, observed both in

Brazil and England. Its predictive accuracy was lower when compared to the synthetic control and led to different impact estimates when tested on vaccine-eligible age groups.

Selecting the most suitable control variables is crucial when predicting infectious diseases like IMD, given the large number of possible clinical, epidemiological, social, and environmental factors that could influence its behavior [11]. Some of these factors are probably strictly related to the pathogen studied, while others are reasonably in common with other diseases (chapter 2). For this reason, I included, as candidate controls, infectious and non-infectious disease cases in infants of the same age as those vaccinated, following the approach previously applied to pneumococcal vaccines [7]. Besides, I included MenB and MenC cases in older age groups that were not eligible for the vaccine program, similar to the original analyses run in England and Brazil [78, 85]. I then investigated which control diseases were selected to compose the synthetic control.

MenB and MenC cases in non-eligible age groups were consistently among the most frequently selected controls in England and Brazil, respectively, though in England the probability of inclusion associated with MenB controls was significantly larger compared to what was observed in Brazil. I then examined the non-IMD diseases that were associated with infant IMD with the highest probability. RSV disease and measles were among the best predictors in England for 1-year-old and 18–51 weeks-old infants, respectively. Evidence that IMD could be associated with RSV and measles cases have been reported in epidemiological studies [47, 138]. Also, a recent study found that measles could reduce humoral immune memory, thereby generating potential vulnerability to future infections [139]. Interestingly, I found similar evidence in Brazil, where bronchitis/bronchiolitis and other acute respiratory infections appeared to be good predictors of MenC cases. Both the time series included RSV disease cases [7, 140].

The synthetic control model has some limitations. Using as controls IMD time series of cases in non-vaccine-eligible age groups comes with inherent risks of generating biased impact estimates. Non-vaccine-eligible controls will likely differ in age compared to the target group, and adjustment by stratifying the population by age may, therefore, be required. Meningococcal vaccination may also indirectly protect unvaccinated subjects [60, 111], so a reduced risk of IMD in non-vaccinated age groups would introduce bias and underestimate the vaccine impact [1, 104].

Dynamical models allow to account for such indirect effects [97], but accurate data on carriage prevalence of specific serogroups and between-persons contact rates, strictly needed to parametrize such models, are usually unavailable. However, in both settings investigated, possible indirect effects were negligible due to the low carrier rate in the vaccinated age groups and the low fraction of vaccinated subjects at the time of my study. As a precaution, I nevertheless re-run all the analyses excluding candidate MenB and MenC cases in non-vaccine-eligible age groups from the set of controls. Results were robust to this exclusion.

The risk of incurring biases is reduced compared to other methods, but fully adjusting for any possible confounder remains a challenge [7]. It is therefore always recommended to compare several methods together to increase the robustness of the results and to detect a possible source of bias [103].

In some circumstances, the synthetic control model could fail to identify an appropriate set of controls, i.e. with 18–51-weeks-old infants in England. Since there were no marked trends in the data, the synthetic control model was still able to produce a reliable counterfactual with only the intercept and seasonal components.

In general, noisy and sparse control time series could be responsible for biased results from the synthetic control, thus leading to non-significant vaccine impact estimates [131]. In such situations, I tested both the performances of the STL+PCA model, which represents a solution for data sparsity [131], and I also tried different data aggregations, both spatially (in Brazil, at a national, regional and state level) and temporally (with annual, quarterly, and monthly data) to improve the accuracy of impact estimates.

Both in Brazil and England, I found that aggregating data with lower resolution, i.e. annual or quarter data depending on the context, did not improve the performances (chapter 6, section *6.6*, and chapter 7, section *7.6*). I did not detect, however, a successful and general pattern when dealing with sparsity issues. The STL+PCA was successful in certain cases, but often its results were comparable with the synthetic control ones. A further attempt with zero-inflated Poisson (ZIP) models could be done to address the limitations of sparseness in the number of cases [141, 142].

In conclusion, in my research, I provided a rigorous procedure to investigate the validity of meningococcal vaccine impact studies with synthetic controls (chapter 5, section *5.8*), discussing, among others, predictive accuracy, indirect effects, meningococcal predictors, and different temporal/spatial aggregation to evaluate vaccine impact.

I showed that the synthetic control is a promising approach for estimating the impact of meningococcal immunization programmes. Its general applicability in different contexts allows an objective comparison between vaccines and immunization strategies, offering a valid alternative for public health decision making. Present results suggest that it could be successfully applied to evaluate meningococcal immunization campaigns targeting adolescents and adults, where indirect effects could hamper a correct assessment of the overall impact. I would expect the synthetic control method to successfully adjust for unexplained trends, relying on non-IMD controls only. Besides, the SC approach is quite flexible, and it could be applied to any infectious disease, of course using control time series not affected by the intervention. Conversely, the use of methods exclusively based on the comparison of pre- versus post-immunization IMD incidence data should be counter checked to avoid possible bias. Also, the intuition that IMD could share some causal factors with measles and RSV would deserve further investigation.

Appendix A

Additional information on the synthetic control method

A.1 Main ingredients of the state-space models

The Kalman filter, the Kalman smoother, and Bayesian data augmentation are the key tools for working with state-space models [119, 143]. The Kalman filter is a method developed to estimate the current state of a dynamical system, given the observations so far. In this context, it is applied to estimate the values for the state components α_t. Given a model and a set of observed variables $y_1, .., y_t$, the Kalman filter produces successive predictions $p(\alpha_{t+1}|y_{1:t})$ conditional on the past and concurrent observations. It is actually the forward algorithm. There is also the backward algorithm, i.e. the Kalman smoother, which allows to refine estimates of previous states, in the light of later observations. With the Kalman smoother, the output of the Kalman filter is updated to produce $p(\alpha_t|y_{1:n})$, where n is the length of the time-series, at each value of t. Since all components of the model are Gaussian, all conditional distributions are multivariate normal distributions parametrized by their mean and variance [119].

Finally, the third ingredient is Bayesian data augmentation: $y = y_{1:n}$ and $\alpha = \alpha_{1:n}$ represent the full set of observed and latent data. With Bayesian data augmentation, simulations from $p(\alpha|y)$ are produced using the Durbin and Koopman method [120, 144]. It is not possible to simulate each α_t from $p(\alpha_t|y)$ since the serial correlation between α_t and α_{t+1} must be respected (recall Fig. 4.1). With the simulation smoother developed by Durbin and Koopman [144], it becomes straightforward to simulate random noise with the same covariance as $p(\alpha|y)$ to which the appropriate mean is added, by means of the "state mean smoother" [119, 120, 144].

A.2 Posterior definition for β, σ^2 and δ

For each *t=1,..,n* in the pre-intervention period, let $\tilde{y}_t$ define y_t with only the regression component and with the time series components subtracted away, so $\tilde{y}_{1:n} = (\tilde{y}_1, .., \tilde{y}_n)$. The idea is to simulate from $p(\delta, \beta, \sigma_\epsilon^2|\tilde{y}_{1:n})$. This term can be factorized into $p(\delta|\tilde{y}_{1:n})p(1/\sigma_\epsilon^2|\delta, \tilde{y}_{1:n})p(\beta|\delta, \sigma_\epsilon^2, \tilde{y}_{1:n})$ [105]. Conditional on δ, the joint posterior distribution for β and σ^2 can be derived from standard conjugacy formulas of Bayesian linear regression, (here with sparsity hypothesis over the coefficients) [105, 119, 121]:

$$\beta_\delta|\sigma_\epsilon, \delta, \tilde{y}_{1:n} \sim N(\tilde{\beta}_\delta, \sigma_\epsilon^2(V_\delta^{-1})^{-1}) \qquad 1/\sigma_\epsilon^2|\delta, \tilde{y}_{1:n} \sim Ga(\frac{N}{2}, \frac{S_\delta}{2}) \tag{A.1}$$

with parameters:

$$\begin{aligned} V_\delta^{-1} &= (X^T X)_\delta + \Sigma_\delta^{-1} & \tilde{\beta}_\delta &= (V_\delta^{-1})^{-1}(X_\delta^T \tilde{y}_{1:n} + \Sigma_\delta^{-1} b_\delta) \\ N &= \nu_\epsilon + n & S_\delta &= s_\epsilon + \tilde{y}_{1:n}^T \tilde{y}_{1:n} + b_\delta^T \Sigma_\delta^{-1} b_\delta - \tilde{\beta}_\delta^T V_\delta^{-1} \tilde{\beta}_\delta \end{aligned} \tag{A.2}$$

Then, δ can be obtained by sampling from the marginal distribution $p(\delta|\tilde{y}_{1:n})$, integrating out β and $1/\sigma_\epsilon^2$ because of conjugacy:

$$\delta|\tilde{y}_{1:n} \sim C(\tilde{y}_{1:n}) \frac{|\Sigma_\delta^{-1}|^{\frac{1}{2}}}{|V_\delta^{-1}|^{\frac{1}{2}}} \frac{p(\delta)}{S_\delta^{\frac{N}{2}-1}} \tag{A.3}$$

where $C(\tilde{y}_{1:n})$ is a normalizing constant that depends on $\tilde{y}_{1:n}$ but not on δ. Equation (A.3) is used in a Gibbs sampling algorithm that draws each δ_i given δ_{-i} (the elements of δ other than δ_i). Each full conditional distribution is proportional to equation (A.3), and δ_i can only assume two possible values, i.e. 0 or 1. Thus, the posterior distribution can be simulated efficiently by drawing from $p(\delta|\tilde{y}_{1:n})$, and then drawing from the well known closed form $p(\beta_\delta, \sigma^{-2}|\delta, \tilde{y}_{1:n})$ [105, 119, 145].

A.3 Estimating the model using Markov Chain Monte Carlo

The parameter θ defines the collection of model parameters, other than β and σ^2, while $\phi = (\theta, \beta, \sigma^2, \alpha)$. The complete data posterior distribution is defined as [105, 119, 145]:

$$p(\phi, \alpha|y_{1:n}) \propto p(\phi)p(\alpha_0)\prod_{t=1}^{n}(y_t|\alpha_t, \phi)p(\alpha_t|\alpha_{t-1}, \phi) \tag{A.4}$$

With Markov Chain Monte Carlo, the posterior distribution can be efficiently simulated following the three steps summarized by Scott and collaborators [119]:

- Simulate the latent state α from $p(\alpha|y, \theta, \beta, \sigma_\epsilon^2)$ using the simulation smoother from Durbin and Koopman [144].
- Simulate $\theta \sim p(\theta|y, \alpha, \beta, \sigma_\epsilon^2)$.
- Simulate β and σ_ϵ^2 from a Markov chain with stationary distribution $p(\beta, \sigma_\epsilon^2|y, \alpha, \theta)$.

Repeatedly cycling through the three steps above yields a sequence of draws $\phi_1, .., \phi_n$, from a Markov chain with stationary distribution $p(\phi|y)$, which represents the posterior distribution of ϕ given y. Step 2, i.e. the draw of θ, depends on which state component is selected and it is often trivial. The draw in step 3 is done using a Gibbs sampling algorithm: each element of δ is drawn from its full conditional distribution, obtained by equation (A.3) (recall that each δ has only two possible values). Finally, after one sweep of all the variables in random order, β_δ and σ_ϵ^2 are drawn from their closed form full conditional distributions given in equation (A.1) [105, 119, 145]. This algorithm is known as Stochastic Search Variable Selection (SSVS) [105, 119, 122, 145].

After simulating the model parameter ϕ and the state vector α, the goal is to simulate from the posterior predictive distribution $p(\bar{y}_{n+1:m}|y_{1:n}, x_{1:m})$, to obtain the counterfactual over the post-intervention period [105].

A.4 Metrics used to assess the quality of predictions produced by the synthetic control model

To evaluate the goodness of fit in the pre-intervention period, I used the Deviation Information Criterion (DIC) metric; whereas I used the Mean Absolute Error (MAE) as a measure of prediction accuracy in the post-intervention period.

The definition of all the metrics is presented below, given that Y_i is the observed target time series at the time i and $Ypred_i$ is the predicted target time series at the time i.

- The DIC metric trades off a measure of model adequacy against a measure of complexity and is defined as: $DIC = \overline{D} + p_D$ with θ the set of parameters, $\overline{D} = E[-2logp(Y|\theta)]$ the posterior mean of the deviance and p_D the effective number of parameters.
- The MAE is a measure of the average of the absolute errors between predicted and true values and is defined as: $MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - Ypred_i|$

A.5 Additional information on the STL+PCA model

The seasonal-trend decomposition procedure based on locally weighted scatterplot smoothing (STL method) decomposes the time series into three components: trend, seasonality, and the remaining variation in the data [131, 133]. The control time series j at month/quarter j can then be written as:

$$ln(ControlDisease_{jt}) = T_{jt} + S_{jt} + R_{jt} \tag{A.5}$$

with T_{jt}, S_{jt}, and R_{jt} being the trend, annual seasonal, and remainder components, respectively [131]. The principal component analysis (PCA) creates uncorrelated projections that explain the maximum variability in the data overall [131, 146]. PCA is applied to the extracted trends for the control time series and only the first principal component PC1 is kept to be included in the regression model.

The regression model is specified as follows, with meningococcal cases following a Poisson regression $y_t \sim Poisson(\lambda_t)$, with mean λ_t [130, 131]:

$$log(\lambda_t) = b_0 + \sum_k c_k * I[month_k = m(t)] + \beta_1 * PC1_t + b_{c(t)} \tag{A.6}$$

where t = 1,2,..., is the total number of time points; m_t is a function that maps a time point to the corresponding calendar month; c_k represents the month k regression coefficient; $I[.]$ represents the indicator function; b_0 is an intercept; β_1 is the regression coefficient for the first principal component (PC1); and $b_c(t)$ is an observation specific random intercept.

Appendix B

Time series used as components of the synthetic control model

B.1 Meningococcal C campaign in Brazil

The complete list of monthly control time series used in Brazil is shown in Table B.1. Control time series in Brazil were publicly available [7]. Besides, I collected MenB time series in the non-vaccine-eligible age groups from the SINAN website [134].

Meningococcal C cases are in the age groups: 5-9; 10-14; 15-19; 20-39; 40-59-years-old. All the other control diseases are in the age groups <1 and 1-4-years-old.

TABLE B.1: Time series used as components of the synthetic control in Brazil

Grouping scheme	ICD-10	Description	Exclusions
MenC		Meningococcal C cases in non vaccine-eligible age groups	
ICD-10 chapters			
	C00-D48	Neoplasms	A40.3, B95
	D50-89	Diseases of blood and blood-forming organs and certain disorders involving the immune mechanism	
	E00-99	Endocrine, nutritional, metabolic disorders	
	H00-99_SY	Diseases of the ear and mastoid process	H10, H65, H66
	I00-99	Diseases of the circulatory system	
	K00-99	Diseases of the digestive system	
	L00-99	Diseases of the skin	
	M00-99	Diseases of the musculoskeletal system	
	N00-99	Diseases of the genitourinary system	
	P00-99	Perinatal diseases	
	Q00-99	Congenital malformations, deformations and chromosomal abnormalities	
	R00-99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	
	S00-T99	Injury, poisoning and consequences of external causes	
	U00-99	Codes for special purposes	
	V00-Y99	External causes	
	Z00-99	Factors influencing health status and contact w/ health workers	
Other grouped outcomes			
	A10_B99_nopneumo	Certain infectious and parasitic diseases, except intestinal	A40.3, B95
	B20-24	HIV	
	E10-E14	Diabetes	
	E40-E46	Malnutrition	
	I60-I64	Stroke	
	J20-J22	Bronchitis, bronchiolitis and unspecified acute lower respiratory infection	
	P05-P07	Premature delivery and low birth weight	
	ACH_NOJ	All nonrespiratory hospitalizations	J00–J99, F and O chapters
Specific outcomes			
	A17	Tuberculosis of nervous system	
	A18	Tuberculosis of other organs	
	A19	Miliary tuberculosis	
	A41	Other septicemia	
	B34	Viral infection of unspecified site	
	B96	Other specified bacterial agents as the cause of diseases classified to other chapters	
	B97	Viral agents as the cause of diseases classified to other chapters	
	B99	Other and unspecified infectious diseases	
	K35	Appendicitis	
	K80	Cholelithiasis	
	N39	Urinary tract infection (UTI)	

B.2 Meningococcal B campaign in England

The complete list of quarterly and annual control time series used in England are shown in Tables B.2 and B.3. Control time series in England were publicly available from the PHE website [147] and historical archives [137]. Besides, I collected MenB time series in the non-vaccine-eligible age groups from the PHE website [136].

Meningococcal B cases are in the age groups: 5-9; 10-14; 15-19; 20-24; 25-44; 45-64; 65+ years-old. All the other control diseases are in the age groups <1 and 1-4-years-old.

TABLE B.2: Quarterly time series used as components of the synthetic control in England

Description
Meningococcal B cases in non vaccine-eligible age groups [136, 137]
Respiratory Syncytial Virus (RSV) [148]
Pertussis [149]
Measles and mumps [150]
Meningococcal C, W, Y cases [136, 137]

TABLE B.3: Annual time series used as components of the synthetic control in England

Grouping scheme	ICD-10	Description
MenB		Meningococcal B cases in non vaccine-eligible age groups [136, 137]
Hep A		Hepatitis A [151]
HepC		Hepatitis C [152]
HIV		Human Immunodeficiency Virus [153]
		Measles, mumps and rubella [150]
		Food poisoning [154]
		Malaria [154]
		Cholera [154]
TB		Tubercolosis [154]
		Scarlet Fever [154]
		Whooping cough [154]
births		Yearly average number of births from 1998 to 2018 [155]
MenC and MenY		Meningococcal C and Y [136, 137]
ICD-10 chapters		
	C00-D48	Neoplasms [156]

Appendix C

Supplementary figures

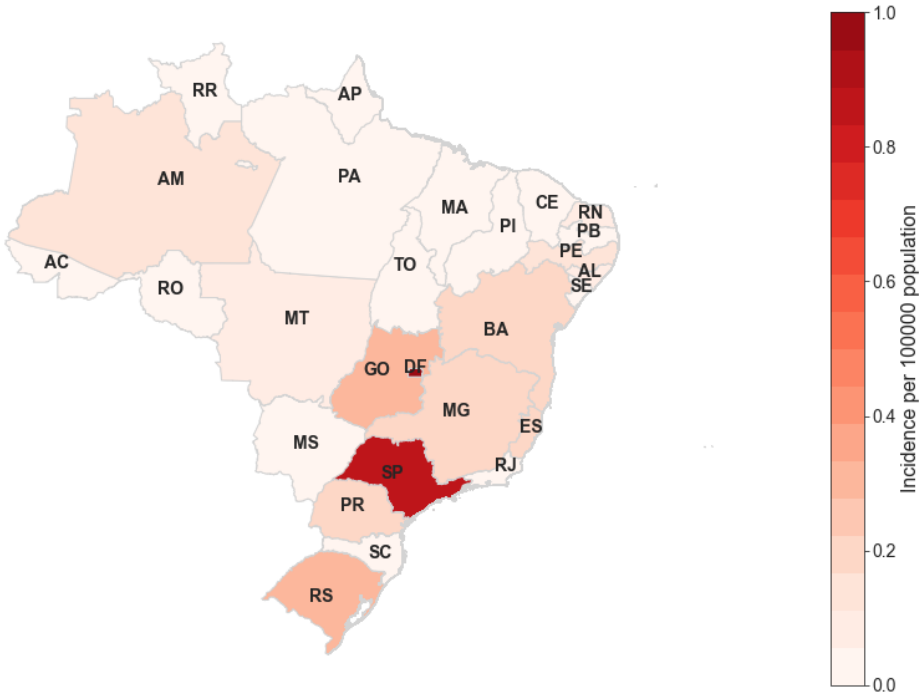

FIGURE C.1: Meningococcal C incidence per 100000 population in infants <5 years old at a state level in Brazil before vaccine introduction (Jan 2007 - Oct 2010).

TABLE C.1: Model size estimates for each set of controls and age groups eligible for vaccination in Brazil.

Age group	Set of controls	Model size
<1 year	*All the controls*	2.31
	Non-elig. MenC excluded	2.43
1-4 years	*All the controls*	1.9
	Non-elig. MenC excluded	1.68

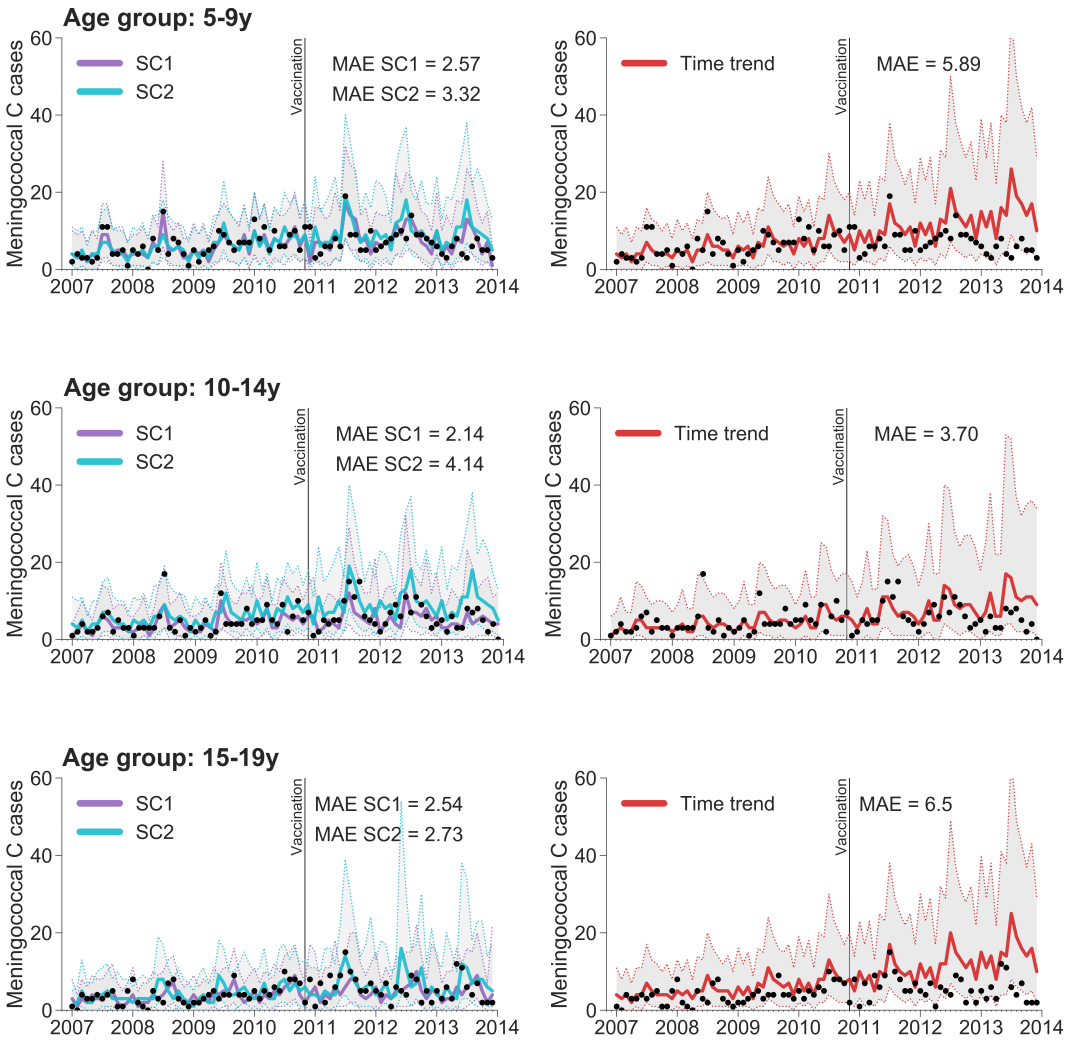

FIGURE C.2: Comparison between predictions generated by synthetic control and time-trend models for MenC disease (São Paulo state, Brazil) in the 15–19 (top), 10-14 (middle) and 5-9 (bottom) years old non-vaccinated age groups. In purple, cases predicted with the synthetic control model using all the controls available (SC1) (curve: best estimate; shaded region: 95%CI). In cyan, cases predicted excluding MenC cases in unvaccinated age groups (SC2) (curve: best estimate; shaded region: 95%CI). In red, cases predicted with a time-trend model (curve: best estimate; shaded region: 95%CI). Observed cases are shown as black dots. Vertical black lines divide each plot into two parts: on the left side, data used to fit models; on the right side, data used to evaluate the predictions (MAEs shown).

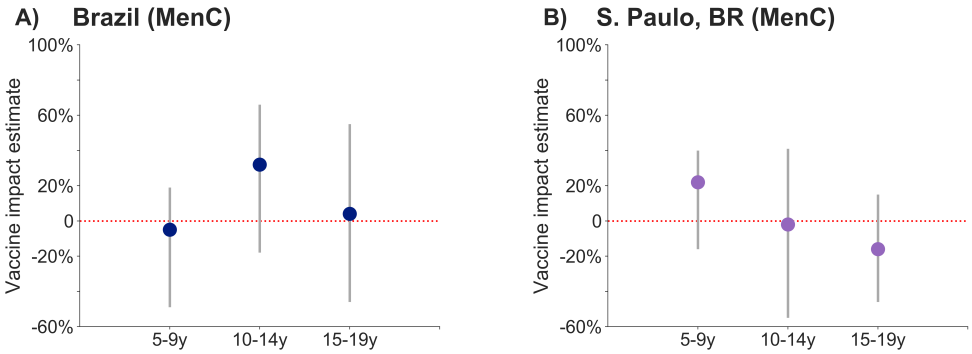

FIGURE C.3: Vaccine impact estimates using MenC cases in non-vaccinated age groups (specifically, 5–9, 10–14 and 15–19 years-old age groups) in Brazil (panel A) and São Paulo (panel B) as targets of the synthetic control model

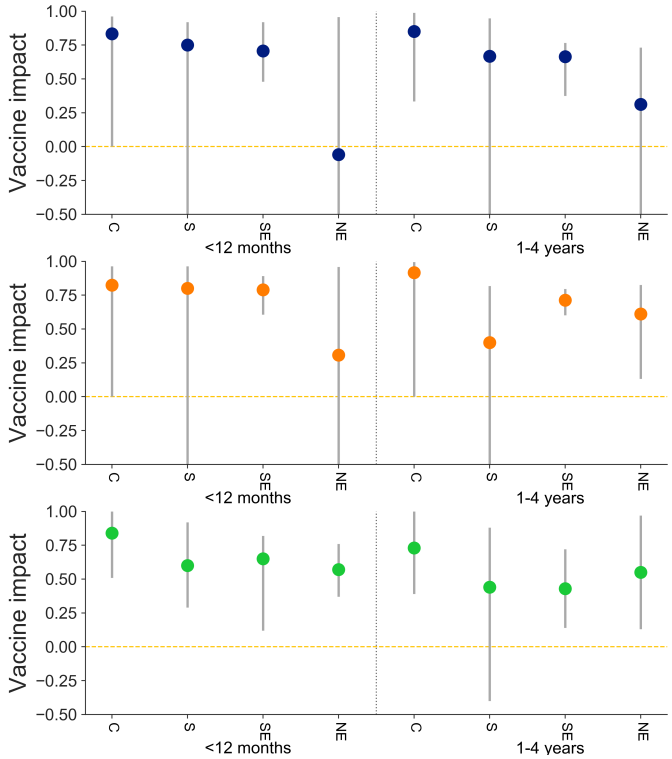

FIGURE C.4: Vaccine impact estimated with the synthetic control model (top panel), the STL+PCA model (middle panel) and previously published estimates from Moraes *et al* [78] (bottom panel), comparing the observed and predicted number of cases in 2011-2013 by age group and region. IMD MenC cases are grouped on a quarterly basis. 95%CIs are shown as grey lines. Abbreviations: C, center-west; S, south; SE, southeast; NE, northeast.

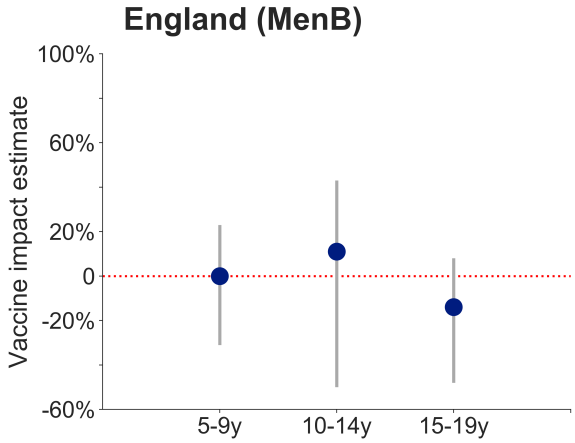

FIGURE C.5: Vaccine impact estimates using MenB cases in non-vaccinated age groups (specifically, 5–9, 10–14 and 15–19 years-old age groups) in England as targets of the synthetic control model

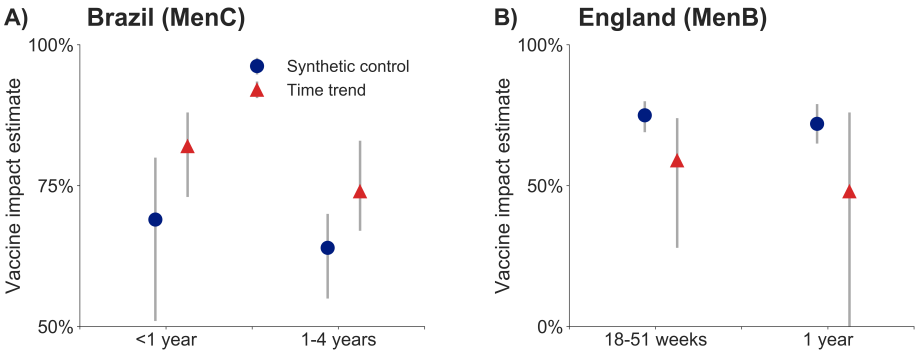

FIGURE C.6: Vaccine impact estimates when using the synthetic control model (blue dots) and the time trend Model (red dots) in Brazil (panel A) and England (panel B).

TABLE C.2: Model size estimates for each set of controls and age groups eligible for vaccination in São Paulo state, Brazil.

Age group	Set of controls	Model size
<1 year	*All the controls*	2.56
1-4 years	*All the controls*	3.00

TABLE C.3: Model size estimates for each set of controls and age groups eligible for vaccination in England with quarterly data.

Age group	Set of controls	Model size
18-51 weeks	*All the controls*	1.00
	Non-elig. MenC excluded	1.01
1 year	*All the controls*	1.65
	Non-elig. MenC excluded	1.97

TABLE C.4: Model size estimates for each set of controls and age groups eligible for vaccination in England with annual data.

Age group	Set of controls	Model size
18-51 weeks	*All the controls*	2.14
	Non-elig. MenC excluded	2.40
1 year	*All the controls*	2.24
	Non-elig. MenC excluded	2.40

Bibliography

[1] M Elizabeth Halloran et al. *Design and analysis of vaccine studies*. Vol. 18. Springer, 2010.

[2] Mr Greenwood and G Udny Yule. *The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general*. 1915.

[3] Germaine Hanquet et al. "Vaccine effects and impact of vaccination programmes in post-licensure studies". In: *Vaccine* 31.48 (2013), pp. 5634–5642.

[4] Carlos G Grijalva et al. "Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: a time-series analysis". In: *The Lancet* 369.9568 (2007), pp. 1179–1186.

[5] Vesta Richardson et al. "Effect of rotavirus vaccination on death from childhood diarrhea in Mexico". In: *New England Journal of Medicine* 362.4 (2010), pp. 299–305.

[6] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program". In: *Journal of the American statistical Association* 105.490 (2010), pp. 493–505.

[7] Christian AW Bruhn et al. "Estimating the population-level impact of vaccines using synthetic controls". In: *Proceedings of the National Academy of Sciences* 114.7 (2017), pp. 1524–1529.

[8] Nancy E Rosenstein et al. "Meningococcal disease". In: *New England journal of medicine* 344.18 (2001), pp. 1378–1388.

[9] Siamak P Yazdankhah and Dominique A Caugant. "Neisseria meningitidis: an overview of the carriage state". In: *Journal of medical microbiology* 53.9 (2004), pp. 821–832.

[10] Ray Borrow et al. "The Global Meningococcal Initiative: global epidemiology, the impact of vaccines on meningococcal disease and the importance of herd protection". In: *Expert review of vaccines* 16.4 (2017), pp. 313–328.

[11] Reinaldo Acevedo et al. "The Global Meningococcal Initiative meeting on prevention of meningococcal disease worldwide: epidemiology, surveillance, hypervirulent strains, antibiotic resistance and high-risk populations". In: *Expert review of vaccines* 18.1 (2019), pp. 15–30.

[12] Muhamed-Kheir Taha et al. "The duality of virulence and transmissibility in Neisseria meningitidis". In: *Trends in microbiology* 10.8 (2002), pp. 376–382.

[13] Philippe De Wals and André Bouckaert. "Methods for estimating the duration of bacterial carriage". In: *International journal of epidemiology* 14.4 (1985), pp. 628–634.

[14] Matthew J Thompson et al. "Clinical recognition of meningococcal disease in children and adolescents". In: *The lancet* 367.9508 (2006), pp. 397–403.

[15] David S Stephens, Brian Greenwood, and Petter Brandtzaeg. "Epidemic meningitis, meningococcaemia, and Neisseria meningitidis". In: *The Lancet* 369.9580 (2007), pp. 2196–2210.

[16] Yih-Ling Tzeng and David S Stephens. "Epidemiology and pathogenesis of Neisseria meningitidis". In: *Microbes and infection* 2.6 (2000), pp. 687–700.

[17] Dominique A Caugant. "Population genetics and molecular epidemiology of Neisseria meningitidis". In: *Apmis* 106.1-6 (1998), pp. 505–525.

[18] Martin CJ Maiden and Matthias Frosch. "Molecular techniques for the investigation of meningococcal disease epidemiology". In: *Molecular biotechnology* 18.2 (2001), pp. 119–134.

[19] https://microbewiki.kenyon.edu/index.php/Neisseria_meningitidis_causing_meningococcal_meningitis.

[20] DA Caugant et al. "Clonal diversity of Neisseria meningitidis from a population of asymptomatic carriers." In: *Infection and Immunity* 56.8 (1988), pp. 2060–2068.

[21] KA Jolley et al. "Carried meningococci in the Czech Republic: a diverse recombining population". In: *Journal of clinical microbiology* 38.12 (2000), pp. 4492–4498.

[22] David S Stephens, Loren H Hoffman, and Zell A McGee. "Interaction of Neisseria meningitidis with human nasopharyngeal mucosa: attachment and entry into columnar epithelial cells". In: *Journal of Infectious Diseases* 148.3 (1983), pp. 369–376.

[23] CL Trotter, NJ Gay, and WJ Edmunds. "The natural history of meningococcal carriage and disease". In: *Epidemiology & Infection* 134.3 (2006), pp. 556–566.

[24] Hannah Christensen et al. "Meningococcal carriage by age: a systematic review and meta-analysis". In: *The Lancet infectious diseases* 10.12 (2010), pp. 853–861.

[25] Dominique A Caugant and Martin CJ Maiden. "Meningococcal carriage and disease—population biology and evolution". In: *Vaccine* 27 (2009), B64–B70.

[26] Yaowen Zhang et al. "Burden of Neisseria meningitidis infections in China: a systematic review and meta–analysis". In: *Journal of global health* 6.2 (2016).

[27] Caroline L Trotter et al. "Impact of MenAfriVac in nine countries of the African meningitis belt, 2010–15: an analysis of surveillance data". In: *The Lancet Infectious Diseases* 17.8 (2017), pp. 867–872.

[28] Shruti Sridhar et al. "Global incidence of serogroup B invasive meningococcal disease: a systematic review". In: *The Lancet infectious diseases* 15.11 (2015), pp. 1334–1346.

[29] https://www.who.int/news-room/fact-sheets/detail/meningococcal-meningitis.

[30] Lara Evellyn do Macedo et al. "Impact of meningococcal C conjugate vaccination programs with and without catch-up campaigns in adolescents: Lessons learned from Bahia, Brazil". In: *Human vaccines & immunotherapeutics* 14.5 (2018), pp. 1131–1137.

[31] Junhong Li et al. "Meningococcal disease and control in China: findings and updates from the Global Meningococcal Initiative (GMI)". In: *Journal of Infection* 76.5 (2018), pp. 429–437.

[32] SV Matosova, KO Mironov, AE Platonov, et al. "Molecular biological monitoring of Neisseria meningitidis in Moscow in the period 2011 to 2015". In: *Epidemiologia I Infekcionnye Bolezni* 2 (2016), pp. 4–9.

[33] Anna Funk et al. "Sequential outbreaks due to a new strain of Neisseria meningitidis serogroup C in northern Nigeria, 2013-14". In: *PLoS currents* 6 (2014).

[34] Cecilia B Kretz et al. "Whole-genome characterization of epidemic Neisseria meningitidis serogroup C and resurgence of serogroup W, Niger, 2015". In: *Emerging infectious diseases* 22.10 (2016), p. 1762.

[35] https://www.cdc.gov/meningococcal/global.html.

[36] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/842368/hpr3819_IMD-ann.pdf.

[37] Lee H Harrison, Caroline L Trotter, and Mary E Ramsay. "Global epidemiology of meningococcal disease". In: *Vaccine* 27 (2009), B51–B63.

[38] H Campbell et al. "Targeted vaccination of teenagers following continued rapid endemic expansion of a single meningococcal group W clone (sequence type 11 clonal complex), United Kingdom 2015". In: *Eurosurveillance* 20.28 (2015), p. 21188.

[39] https://www.niid.go.jp/niid/en/iasr-vol39-e/865-iasr/7802-455te.html.

[40] Ana Belén Ibarz-Pavón et al. "Epidemiology, molecular characterization and antibiotic resistance of Neisseria meningitidis from patients 15 years in Manhiça, rural Mozambique". In: *PLoS One* 6.6 (2011), e19717.

[41] https://www.who.int/images/default-source/health-topics/meningitis/map-serogroup-distribution-2019.png?sfvrsn=af422ab2_2.

[42] Juliette Paireau et al. "Seasonal dynamics of bacterial meningitis: a time-series analysis". In: *The Lancet global health* 4.6 (2016), e370–e377.

[43] Dennis K. S. Law et al. "Invasive Meningococcal Disease in Québec, Canada, Due to an Emerging Clone of ST-269 Serogroup B Meningococci with Serotype Antigen 17 and Serosubtype Antigen P1.19 (B:17:P1.19)". In: *Journal of Clinical Microbiology* 44.8 (2006), pp. 2743–2749. ISSN: 0095-1137. DOI: 10.1128/JCM.00601-06. eprint: https://jcm.asm.org/content/44/8/2743.full.pdf. URL: https://jcm.asm.org/content/44/8/2743.

[44] Philippe De Wals et al. "Impact of an immunization campaign to control an increased incidence of serogroup B meningococcal disease in one region of Quebec, Canada". In: *Clinical Infectious Diseases* 64.9 (2017), pp. 1263–1267.

[45] Anne Abio, Keith R. Neal, and Charles R. Beck. "An epidemiological review of changes in meningococcal biology during the last 100 years". In: *Pathogens and Global Health* 107 (2013), pp. 373 –380.

[46] Ray Borrow et al. "Effectiveness of meningococcal serogroup C vaccine programmes". In: *Vaccine* 31.41 (2013), pp. 4477–4486.

[47] Ashleigh R Tuite et al. "Respiratory virus infection and risk of invasive meningococcal disease in central Ontario, Canada". In: *PLoS One* 5.11 (2010), e15493.

[48] Judith E Mueller et al. "The association between respiratory tract infection incidence and localised meningitis epidemics: an analysis of high-resolution surveillance data from Burkina Faso". In: *Scientific reports* 7.1 (2017), pp. 1–9.

[49] KAV Cartwright et al. "Influenza A and meningococcal disease". In: *The Lancet* 338.8766 (1991), pp. 554–557.

[50] Bruno Hubert et al. "Meningococcal disease and influenza-like syndrome: a new approach to an old question". In: *Journal of Infectious Diseases* 166.3 (1992), pp. 542–545.

[51] Elise Snitker Jensen et al. "A 20-year ecological study of the temporal association between influenza and meningococcal disease". In: *European journal of epidemiology* 19.2 (2004), pp. 181–187.

[52] Lowell S Young et al. "A simultaneous outbreak of meningococcal and influenza infections". In: *New England Journal of Medicine* 287.1 (1972), pp. 5–9.

[53] Angela Domínguez et al. "Time-series analysis of meningococcal disease in Catalonia". In: *Annals of epidemiology* 17.9 (2007), pp. 654–662.

[54] Jean-Michel Alonso et al. "A model of meningococcal bacteremia after respiratory superinfection in influenza A virus-infected mice". In: *FEMS microbiology letters* 222.1 (2003), pp. 99–106.

[55] John F Brundage. "Interactions between influenza and bacterial respiratory pathogens: implications for pandemic preparedness". In: *The Lancet infectious diseases* 6.5 (2006), pp. 303–312.

[56] Marie-Anne Rameix-Welti et al. "Influenza A virus neuraminidase enhances meningococcal adhesion to epithelial cells through interaction with sialic acid-containing meningococcal capsules". In: *Infection and immunity* 77.9 (2009), pp. 3588–3595.

[57] Ray Borrow et al. "Meningococcal disease in the Middle East and Africa: Findings and updates from the Global Meningococcal Initiative". In: *Journal of Infection* 75.1 (2017), pp. 1–11.

[58] Matthew D Snape and Andrew J Pollard. "Meningococcal polysaccharide–protein conjugate vaccines". In: *The Lancet infectious diseases* 5.1 (2005), pp. 21–30.

[59] E Miller, D Salisbury, and M Ramsay. "Planning, registration, and implementation of an immunisation campaign against meningococcal serogroup C disease in the UK: a success story". In: *Vaccine* 20 (2001), S58–S67.

[60] Martin CJ Maiden et al. "Impact of meningococcal serogroup C conjugate vaccines on carriage and herd immunity". In: *The Journal of infectious diseases* 197.5 (2008), pp. 737–743.

[61] Paul A Kristiansen et al. "Impact of the serogroup A meningococcal conjugate vaccine, MenAfriVac, on carriage and herd immunity". In: *Clinical infectious diseases* 56.3 (2013), pp. 354–363.

[62] DM Daugla et al. "Effect of a serogroup A meningococcal conjugate vaccine (PsA–TT) on serogroup A meningococcal meningitis and carriage in Chad: a community study". In: *The Lancet* 383.9911 (2014), pp. 40–47.

[63] Paul A Kristiansen et al. "Persistent low carriage of serogroup A Neisseria meningitidistwo years after mass vaccination with the meningococcal conjugate vaccine, MenAfriVac". In: *BMC infectious diseases* 14.1 (2014), p. 663.

[64] Jukka Finne, Maija Leinonen, and P Helena Mäkelä. "Antigenic similarities between brain components and bacteria causing meningitis: implications for vaccine development and pathogenesis". In: *The Lancet* 322.8346 (1983), pp. 355–357.

[65] FoA Wyle et al. "Immunologic response of man to group B meningococcal polysaccharide vaccines". In: *Journal of infectious diseases* 126.5 (1972), pp. 514–522.

[66] Rodolfo Villena et al. "Global epidemiology of serogroup B meningococcal disease and opportunities for prevention with novel recombinant protein vaccines". In: *Human vaccines & immunotherapeutics* 14.5 (2018), pp. 1042–1057.

[67] DR Martin et al. "The VR2 epitope on the PorA P1. 7-2, 4 protein is the major target for the immune response elicited by the strain-specific group B meningococcal vaccine MeNZB". In: *Clinical and Vaccine Immunology* 13.4 (2006), pp. 486–491.

[68] Meta J Kuehn and Nicole C Kesty. "Bacterial outer membrane vesicles and the host–pathogen interaction". In: *Genes & development* 19.22 (2005), pp. 2645–2655.

[69] Johan Holst et al. "Vaccines against meningococcal serogroup B disease containing outer membrane vesicles (OMV): lessons from past programs and implications for the future". In: *Human vaccines & immunotherapeutics* 9.6 (2013), pp. 1241–1253.

[70] Rino Rappuoli. "Reverse vaccinology, a genome-based approach to vaccine development". In: *Vaccine* 19.17-19 (2001), pp. 2688–2691.

[71] Mariagrazia Pizza et al. "Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing". In: *Science* 287.5459 (2000), pp. 1816–1820.

[72] Herve Tettelin et al. "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58". In: *Science* 287.5459 (2000), pp. 1809–1815.

[73] https://www.fda.gov/vaccines-blood-biologics/vaccines/bexsero.

[74] https://www.fda.gov/vaccines-blood-biologics/vaccines/trumenba.

[75] https://www.ema.europa.eu/en/documents/overview/bexsero-epar-summary-public_en.pdf.

[76] https://www.tga.gov.au/auspar/auspar-multi-component-meningococcal-b-vaccine.

[77] https://www.gov.uk/government/publications/menb-vaccination-introduction-from-1-september-2015.

[78] Camile de Moraes et al. "Evaluation of the impact of serogroup C meningococcal disease vaccination program in Brazil and its regions: a population-based study, 2001-2013". In: *Memórias do Instituto Oswaldo Cruz* 112.4 (2017), pp. 237–246.

[79] Ana Lucia Andrade et al. "Impact of meningococcal C conjugate vaccination four years after introduction of routine childhood immunization in Brazil". In: *Vaccine* 35.16 (2017), pp. 2025–2033.

[80] Luciano Cesar Pontes Azevedo, Cristiana M Toscano, and Ana Luiza Bierrenbach. "Bacterial meningitis in Brazil: baseline epidemiologic assessment of the decade prior to the introduction of pneumococcal and meningococcal vaccines". In: *PloS one* 8.6 (2013), e64524.

[81] Ana Paula Silva de Lemos et al. "Clonal distribution of invasive Neisseria meningitidis serogroup C strains circulating from 1976 to 2005 in greater Sao Paulo, Brazil". In: *Journal of Clinical Microbiology* 45.4 (2007), pp. 1266–1273.

[82] Maria Cecília O Gorla et al. "Phenotypic and molecular characterization of serogroup C Neisseria meningitidis associated with an outbreak in Bahia, Brazil". In: *Enfermedades infecciosas y microbiologia clinica* 30.2 (2012), pp. 56–59.

[83] BPM Iser et al. "Outbreak of Neisseria meningitidis C in workers at a large food-processing plant in Brazil: challenges of controlling disease spread to the larger community". In: *Epidemiology & Infection* 140.5 (2012), pp. 906–915.

[84] http://www.sgc.goias.gov.br/upload/links/arq_626_menig.pdf.

[85] Shamez N. Ladhani et al. "Vaccination of Infants with Meningococcal Group B Vaccine (4CMenB) in England". In: *New England Journal of Medicine* 382.4 (2020), pp. 309–317. DOI: 10.1056/NEJMoa1901229. eprint: https://doi.org/10.1056/NEJMoa1901229. URL: https://doi.org/10.1056/NEJMoa1901229.

[86] Shamez N Ladhani et al. "Enter B and W: two new meningococcal vaccine programmes launched". In: *Archives of disease in childhood* 101.1 (2016), pp. 91–95.

[87] AW Dretler, NG Rouphael, and DS Stephens. "Progress toward the global control of Neisseria meningitidis: 21st century vaccines, current guidelines, and challenges for future vaccine development". In: *Human Vaccines & Immunotherapeutics* 14.5 (2018), pp. 1146–1160.

[88] John Donnelly et al. "Qualitative and quantitative assessment of meningococcal antigens to evaluate the potential strain coverage of protein-based vaccines". In: *Proceedings of the National Academy of Sciences* 107.45 (2010), pp. 19490–19495.

[89] Ulrich Vogel et al. "Predicted strain coverage of a meningococcal multicomponent vaccine (4CMenB) in Europe: a qualitative and quantitative assessment". In: *The Lancet infectious diseases* 13.5 (2013), pp. 416–425.

[90] Duccio Medini, Maria Stella, and James Wassil. "MATS: Global coverage estimates for 4CMenB, a novel multicomponent meningococcal B vaccine". In: *Vaccine* 33.23 (2015), pp. 2629–2636.

[91] Sydel R Parikh et al. "Effectiveness and impact of a reduced infant schedule of 4CMenB vaccine against group B meningococcal disease in England: a national observational cohort study". In: *The Lancet* 388.10061 (2016), pp. 2775–2782.

[92] Matthew D Snape et al. "Persistence of bactericidal antibodies following early infant vaccination with a serogroup B meningococcal vaccine and immunogenicity of a preschool booster dose". In: *Cmaj* 185.15 (2013), E715–E724.

[93] Helen S Marshall et al. "Meningococcal B vaccine and meningococcal carriage in adolescents in Australia". In: *New England Journal of Medicine* 382.4 (2020), pp. 318–327.

[94] Anne Schuchat and Beth P Bell. "Monitoring the impact of vaccines postlicensure: new challenges, new opportunities". In: *Expert review of vaccines* 7.4 (2008), pp. 437–456.

[95] Maarten van Wijhe. "The public health impact of vaccination programmes in the Netherlands". In: ().

[96] Marc Lipsitch. *Challenges of vaccine effectiveness and waning studies*. 2019.

[97] Lorenzo Argante, Michele Tizzoni, and Duccio Medini. "Fast and accurate dynamic estimation of field effectiveness of meningococcal vaccines". In: *BMC medicine* 14.1 (2016), p. 98.

[98] M Elizabeth Halloran and Claudio J Struchiner. "Study designs for dependent happenings". In: *Epidemiology* (1991), pp. 331–338.

[99] France Veyrier-du Lac et al. "Current challenges and new methodological approaches to assess vaccine effectiveness and vaccination impact". In: (2009).

[100] Lucia Helena De Oliveira et al. "Impact and effectiveness of meningococcal vaccines: a review". In: *Revista Panamericana de Salud Pública* 41 (2018), e158.

[101] William R Shadish, Thomas D Cook, Donald Thomas Campbell, et al. *Experimental and quasi-experimental designs for generalized causal inference/William R. Shedish, Thomas D. Cook, Donald T. Campbell.* Boston: Houghton Mifflin, 2002.

[102] James Lopez Bernal, Steven Cummins, and Antonio Gasparrini. "The use of controls in interrupted time series studies of public health interventions". In: *International journal of epidemiology* 47.6 (2018), pp. 2082–2093.

[103] Christian AW Bruhn et al. "Improving assessments of population-level vaccine impact". In: *Epidemiology (Cambridge, Mass.)* 28.2 (2017), p. 233.

[104] Marc Lipsitch, Ayan Jha, and Lone Simonsen. "Observational studies and the difficult quest for causality: lessons from vaccine effectiveness and impact studies". In: *International journal of epidemiology* 45.6 (2016), pp. 2060–2074.

[105] Kay H Brodersen et al. "Inferring causal impact using Bayesian structural time-series models". In: *The Annals of Applied Statistics* 9.1 (2015), pp. 247–274.

[106] Pejman Rohani and Aaron A King. "Never mind the length, feel the quality: the impact of long-term epidemiological data sets on theory, application and policy". In: *Trends in ecology & evolution* 25.10 (2010), pp. 611–618.

[107] Judith E Mueller and Bradford D Gessner. "A hypothetical explanatory model for meningococcal meningitis in the African meningitis belt". In: *International Journal of Infectious Diseases* 14.7 (2010), e553–e559.

[108] Lauren Ancel Meyers et al. "Epidemiology, hypermutation, within–host evolution and the virulence of Neisseria meningitidis". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270.1525 (2003), pp. 1667–1677.

[109] N Stollenwerk and M Maiden. "CJ, Jansen V AA. Diversity can cause outbreaks of meningococcal disease". In: *PNAS* 101.27 (2004), pp. 10229–10234.

[110] Caroline O Buckee et al. "Role of selection in the emergence of lineages and the evolution of virulence in Neisseria meningitidis". In: *Proceedings of the National Academy of Sciences* 105.39 (2008), pp. 15082–15087.

[111] Caroline L Trotter, Nigel J Gay, and W John Edmunds. "Dynamic models of meningococcal carriage, disease, and the impact of serogroup C conjugate vaccination". In: *American journal of epidemiology* 162.1 (2005), pp. 89–100.

[112] TJ Irving et al. "Modelling meningococcal meningitis in the African meningitis belt". In: *Epidemiology & Infection* 140.5 (2012), pp. 897–905.

[113] Andromachi Karachaliou et al. "Modeling long-term vaccination strategies with MenAfriVac in the African meningitis belt". In: *Clinical Infectious Diseases* 61.suppl_5 (2015), S594–S600.

[114] Domenico Martinelli et al. "Estimation of the impact of meningococcal serogroup C universal vaccination in Italy and suggestions for the multicomponent serogroup B vaccine introduction". In: *Journal of immunology research* 2015 (2015).

[115] AE Wormsbecker et al. "Epidemiology of serogroup C and Y invasive meningococcal disease (IMD) in Ontario, 2000–2013: Vaccine program impact assessment". In: *Vaccine* 33.42 (2015), pp. 5678–5683.

[116] Pantelis Samartsidis et al. "Review of methods for assessing the causal effect of binary interventions from aggregate time-series observational data". In: *arXiv preprint arXiv:1804.07683* (2018).

[117] Donald B Rubin. "Using propensity scores to help design observational studies: application to the tobacco litigation". In: *Health Services and Outcomes Research Methodology* 2.3-4 (2001), pp. 169–188.

[118] Robert H Shumway and David S Stoffer. "Time series analysis and its applications". In: *Studies In Informatics And Control* 9.4 (2000), pp. 375–376.

[119] Steven L Scott and Hal R Varian. "Predicting the present with bayesian structural time series". In: *International Journal of Mathematical Modelling and Numerical Optimisation* 5.1-2 (2014), pp. 4–23.

[120] James Durbin and Siem Jan Koopman. *Time series analysis by state space methods*. Oxford university press, 2012.

[121] Andrew Gelman et al. *Bayesian data analysis*. CRC press, 2013.

[122] Edward I George and Robert E McCulloch. "Variable selection via Gibbs sampling". In: *Journal of the American Statistical Association* 88.423 (1993), pp. 881–889.

[123] Michaela Dvorzak and Helga Wagner. "Sparse Bayesian modelling of underreported count data". In: *Statistical Modelling* 16.1 (2016), pp. 24–46.

[124] Arnold Zellner. "On assessing prior distributions and Bayesian regression analysis with g-prior distributions". In: *Bayesian inference and decision techniques* (1986).

[125] Hugh Chipman et al. "The practical implementation of Bayesian model selection". In: *Lecture Notes-Monograph Series* (2001), pp. 65–134.

[126] https://www.universiteitleiden.nl/binaries/content/assets/science/mi/scripties/statscience/2017-2018/2018_07_05_masterthesis_liu.pdf.

[127] Cynthia Schuck-Paim et al. "Effect of pneumococcal conjugate vaccine introduction on childhood pneumonia mortality in Brazil: a retrospective observational study". In: *The Lancet Global Health* 7.2 (2019), e249–e256.

[128] Carla Vizzotti et al. "Impact of a maternal immunization program against pertussis in a developing country". In: *Vaccine* 34.50 (2016), pp. 6223–6228.

[129] Joshua L Warren et al. "Impact of pneumococcal conjugate vaccines on pneumonia hospitalizations in high-and low-income subpopulations in Brazil". In: *Clinical Infectious Diseases* 65.11 (2017), pp. 1813–1818.

[130] https://github.com/weinbergerlab/InterventionEvaluatR.

[131] Kayoko Shioda et al. "Challenges in estimating the impact of vaccination with sparse data". In: *Epidemiology (Cambridge, Mass.)* 30.1 (2019), p. 61.

[132] Lucia H de Oliveira et al. "Declines in pneumonia mortality following the introduction of pneumococcal conjugate vaccines in Latin American and Caribbean countries". In: *Clinical Infectious Diseases* (2020).

[133] Robert B Cleveland et al. "STL: A seasonal-trend decomposition procedure based on loess. 1990". In: *DOI: citeulike-article-id* 1435502 ().

[134] http://portalsinan.saude.gov.br/meningite.

[135] Cristiane W Cardoso et al. "Impact of vaccination during an epidemic of serogroup C meningococcal disease in Salvador, Brazil". In: *Vaccine* 30.37 (2012), pp. 5541–5546.

[136] https://www.gov.uk/government/collections/meningococcal-disease-guidance-data-and-analysis.

[137] https://webarchive.nationalarchives.gov.uk/20110927180040/http://www.hpa.nhs.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1234510036546.

[138] Nita Bharti et al. "Spatial dynamics of meningococcal meningitis in Niger: observed patterns in comparison with measles". In: *Epidemiology & Infection* 140.8 (2012), pp. 1356–1365.

[139] Michael J Mina et al. "Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens". In: *Science* 366.6465 (2019), pp. 599–606.

[140] André Ricardo Ribas Freitas and Maria Rita Donalisio. "Respiratory syncytial virus seasonality in Brazil: implications for the immunisation policy for at-risk populations". In: *Memórias do Instituto Oswaldo Cruz* 111.5 (2016), pp. 294–301.

[141] Susanne Gschlößl and Claudia Czado. "Modelling count data with overdispersion and spatial effects". In: *Statistical papers* 49.3 (2008), p. 531.

[142] Danise Senna Oliveira et al. "Spatial analysis of pneumococcal meningitis in São Paulo in the pre-and post-immunization era". In: *Revista de saude publica* 53 (2019), p. 59.

[143] Rudolph Emil Kalman. "A new approach to linear filtering and prediction problems". In: (1960).

[144] James Durbin and Siem Jan Koopman. "A simple and efficient simulation smoother for state space time series analysis". In: *Biometrika* 89.3 (2002), pp. 603–616.

[145] Steven L Scott and Hal R Varian. *Bayesian variable selection for nowcasting economic time series*. Tech. rep. National Bureau of Economic Research, 2013.

[146] Hervé Abdi and Lynne J Williams. "Principal component analysis". In: *Wiley interdisciplinary reviews: computational statistics* 2.4 (2010), pp. 433–459.

[147] https://www.gov.uk/topic/health-protection/infectious-diseases.

[148] https://www.gov.uk/government/publications/respiratory-infections-laboratory-reports-2019.

[149] https://www.gov.uk/government/publications/pertussis-laboratory-confirmed-cases-reported-in-england-2017.

[150] https://www.gov.uk/government/publications/measles-mumps-and-rubella-lab-confirmed-cases-in-england-2019.

[151] https://www.gov.uk/government/collections/hepatitis-a-guidance-data-and-analysis.

[152] https://www.gov.uk/government/collections/hepatitis-c-guidance-data-and-analysis.

[153] https://www.gov.uk/government/statistics/hiv-annual-data-tables.

[154] https://www.gov.uk/government/collections/notifications-of-infectious-diseases-noids.

[155] https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsummarytablesenglandandwales/2019.

[156] https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2017.