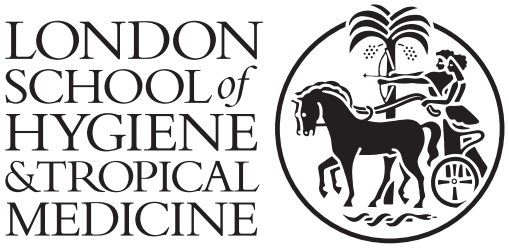


Using registry data to estimate the effects of long-term treatment use in cystic fibrosis

Author:
Simon NEWSOME

Supervisors:
Dr Ruth KEOGH
Dr Rhian DANIEL

Thesis submitted in accordance with the requirements for the degree Doctor of Philosophy

in the

Department of Medical Statistics
Faculty of Epidemiology and Population Health
London School of Hygiene & Tropical Medicine

funded by

Cystic Fibrosis Trust

29th January, 2019

Declaration of Authorship

I have read and understood the School's definition of plagiarism and cheating given in the Research Degrees Handbook. I declare that this thesis is my own work, and that I have acknowledged all results and quotations from the published or unpublished work of other people.

I have read and understood the School's definition and policy on the use of third parties (either paid or unpaid) who have contributed to the preparation of this thesis by providing copy editing and, or, proof reading services. I declare that no changes to the intellectual content or substance of this thesis were made as a result of this advice, and, that I have fully acknowledged all such contributions.

I have exercised reasonable care to ensure that the work is original and does not to the best of my knowledge break any UK law or infringe any third party's copyright or other intellectual property right.

Signed: 

Date: 29th January, 2019

Abstract

Using registry data to estimate the effects of long-term treatment use in cystic fibrosis

Cystic fibrosis (CF) is a disease affecting over 10,000 people in the UK. It has no cure, but there are many treatments to help improve health. Randomised controlled trials are the gold standard for establishing treatment efficacy, but most trials for CF treatments have no more than one year of follow-up. In practice treatments are commonly used for many years, and it is therefore important to evaluate their long-term effectiveness.

The UK CF Registry collects annual data on almost all people with CF in the UK. The overall aim of this work is to investigate how data from such registries can be harnessed to provide insights into the effects of long-term treatment use. My research illustrates the potential of registry data by investigating two CF treatments: DNase and ivacaftor.

DNase is a common CF treatment and generally, once started, it continues to be used indefinitely. Despite this, no studies have investigated its long-term effects. Estimating these effects using registry data is difficult due to time-dependent confounding. I investigate five methods that can account for this: sequential conditional mean models, inverse probability weighting of marginal structural models (MSM), history-adjusted MSM, g-computation formula and g-estimation of structural nested models. The performance of these methods is assessed through simulation studies, where it is shown that all methods perform similarly under correct model specification, suggesting that more than one method could be applied to assess consistency of results. My analysis of the UK CF Registry data suggests that DNase provides a step-change improvement in lung function only in individuals with ppFEV$_1$ $< 70\%$ (e.g. for a person starting DNase with ppFEV$_1$ of 20%, the one-year treatment effect was a 1.6% absolute difference in ppFEV$_1$, 95% CI 0.4, 2.8). However, the slope of lung function decline over five years remained unchanged.

Ivacaftor was introduced in the UK in 2012, but it is only available to people with a gating mutation. In this subgroup, it appears to be so beneficial that almost all eligible people are now receiving it. In this situation, it is difficult to estimate the treatment effect, because there are no eligible people not receiving treatment. Two possible comparator groups were identified: 1) those currently receiving ivacaftor, but using their data from years prior to its introduction, 2) those ineligible to receive ivacaftor due to their genotype. This work shows how analyses using negative controls can be used to assess the comparability of the different groups, and how differences between groups not due to treatment can be mitigated. Our analysis suggests that these two groups are comparable to people who are currently receiving ivacaftor, and the results of the analysis show that ivacaftor not only provides an initial step-change improvement in lung function (5.9% absolute difference in ppFEV$_1$, 95% CI 4.7, 7.1), but also decrease the rate of lung function decline (0.5% absolute decrease in ppFEV$_1$ decline per year, 95% CI 0.02, 1.0).

Acknowledgements

Firstly, the biggest thanks should go to my two supervisors: Ruth Keogh and Rhian Daniel. Their advice and support over the past three years have always been invaluable and I do not think I could have had a better pair of supervisors.

I am also grateful for all the help I have received from everyone in the CF-EpiNet group. A special thanks to Siobhán Carr and Diana Bilton for always being on hand to answer all my CF related questions. Thanks also to Olga Archangelidi, Susan Charman, Rebecca Cosgriff, Paul Cullinan, Elaine Gunn, Debbie Jarvis, Amy MacDougall, Bishal Mohindru, Daniela Schlüter, Sanja Stanojevic and David Taylor-Robinson.

Thanks should also go to the examiners of my upgrading, Nicholas Simmonds and Sarah Walker, for their helpful suggestions during the early stages of my work.

I would also like to thank the people with CF and their families for consenting to their data being held in the UK CF Registry, and NHS teams in CF centres and clinics for the input of data into the Registry.

I would also like to extend my gratitude to the UK Cystic Fibrosis Trust for providing access to the anonymised UK CF Registry data and for their generous financial support throughout my PhD.

Finally, thanks to Emma and Lizzie for their friendship during my time in London, and to Roel for his constant love and support even from afar.

Contents

List of Figures

List of Tables

List of Abbreviations

BMI	Body Mass Index
CF	Cystic Fibrosis
CF-EpiNet	Cystic Fibrosis Epidemiological Network
CFRD	Cystic Fibrosis Related Diabetes
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CI	Confidence Interval
DAG	Directed Acyclic Graph
DNase	Dornase Alfa
FEF_{25-75}	Forced Expiratory Flow 25-75%
FEV_1	Forced Expiratory Volume in One Second
FVC	Forced Vital Capacity
GLI	Global Lung Function Initiative
HA	History Adjusted
IPW	Inverse Probability Weighting
IRR	Incidence Rate Ratio
IV	Intravenous Antibiotic
LCI	Lung Clearance Index
MRSA	Meticillin-resistant *Staphylococcus aureus*
MSE	Mean Squared Error
MSM	Marginal Structural Models
NCCTE	Negative-Control-Corrected Treatment Effect
NCE	Negative-Control Effect
NTE	Naïve Treatment Effect
OR	Odds Ratio
PICOS	Participants, Interventions, Comparisons, Outcomes, Study designs
pp	Percent Predicted
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	Randomised Controlled Trial
RR	Relative Risk
SCMM	Sequential Conditional Mean Model
SE	Standard Error
SNM	Structural Nested Models
UK	United Kingdom
USA	United States of America

Part I

Introduction

Chapter 1

Introduction

1.1 Introduction

The work in this thesis was undertaken as part of the Cystic Fibrosis Epidemiological Network (CF-EpiNet), a Strategic Research Centre funded by the Cystic Fibrosis Trust. The network is carrying out several projects with the overall aim of *harnessing registry data to improve the lives of people with cystic fibrosis (CF)*. This project specifically aims to investigate how national registries can be used to estimate the effects of long-term treatment use.

Randomised controlled trials (RCTs) are known to be the gold-standard for estimating treament effects, but there are many situations where it is not possible to run a RCT.[1] For example, when we wish to estimate the effects of long-term treatment use, it may not be feasible to run a trial for such a long period of time and furthermore it may not be ethical to continue witholding a treatment from patients if it has already been proven to be beneficial in the short-term. In such situations, we can try to use observational data, such as registries, to obtain estimates of the treatment effects. However, estimation of treatment effects using observational data is not straightforward, as the people receiving treatment can be systematically different from those not receiving treatment, making comparisons difficult. These difficulties can be exacerbated when the data are longitudinal, due to there not only being differences between people at baseline, but also during follow-up, and also because people may start and stop treatments over time. The field of causal inference provides a conceptual framework to formally define what is meant by a 'treatment effect', even in longitudinal settings, and based on these clear definitions, together with clearly articulated assumptions, a wide variety of statistical methods have been developed that can be used to estimate causal effects of treatments in these challenging settings.

The overall aim of this thesis is to investigate the use of causal inference methods with national registry data to be able to accurately estimate the effects of long-term treatment

use in CF. This chapter contains a brief overview of CF and the basis of the causal inference methods that will be considered in this thesis. The chapter ends with an outline of the whole thesis.

1.2 Cystic Fibrosis

CF is one of the most common life-threatening genetic diseases in the world and in the United Kingdom (UK) there are approximately 10,500 people living with the disease.[2] CF is caused by a mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) gene which affects the way that salt and water move between cells. There are a number of different CF-causing mutations, and depending on the severity of the mutation, the synthesis, processing or regulation of the CFTR protein can be impaired. Overall, these issues result in a thick mucus which clogs up internal organs.

The most seriously affected organ is the lung, where the build-up of mucus causes breathing difficulties and leads to an increase in respiratory infections. Long term this results in a general trend of deteriorating lung function, but there can also be more sudden drops in lung function as well as periods when lung function improves. There are a number of bacterial and viral infections that commonly affect people with CF. Often these can at first be successfully eradicated, but long term many people end up with chronic infections. In order to avoid cross contamination, it is generally recommended that people with CF avoid contact with other people with CF.

As lung function deteriorates to very low levels it is often desirable for people to receive lung transplants. Sometimes other organs are also transplanted, but the most common transplant is a double lung transplant. Obviously, this depends on the availability of suitable donor organs, and in 2016 in the UK there were 96 people added onto the transplant wait list with 51 transplants performed (46 of which were double lung transplants).[3] After a lung transplant, people still have CF, but they no longer have CF lungs and as such they are generally very different from people who have not had a transplant, especially in terms of lung function. For this reason, for all the work in this thesis we will censor people after the date of their transplant.

The digestive system is also commonly affected by mucus which prevents the proper absorption of food. This can lead to difficulties in obtaining adequate nutrition and people will often receive enzyme supplements to aid digestion.

There is currently no cure for CF, but nowadays there are many treatments available to help people with CF live longer, healthier lives. In 1959 the median survival age of people with CF in England and Wales was estimated to be between 0 and 4 years.[4] However this has increased greatly and in 2017 the median predicted survival age was estimated to be 47 in the UK.[5] With people with CF living longer, treatments are also being used for

longer durations, as many treatments are expected to be used continuously once initiated. However, most treatments have only been shown to be efficacious in clinical trials with follow-up times of around one or two years.

The majority of these treatments target specific symptoms of CF, meaning that people often have to take many treatments for every area of health that is affected by the disease. This has led to treatment burden being a common complaint among people with CF; for example, a 2009 study from the United States of America (USA) showed that the median number of daily therapies per person was 7, requiring a mean time of 108 minutes per day.[6] One of the most common treatments available to people with CF is dornase alfa (DNase), which is administered by inhalation using a jet nebuliser. The treatment helps break down mucus in the airways with the aim of improving lung function. It has been licensed in the UK since 1994, but despite being available for over 20 years, no studies have looked at the effects of taking the treatment for more than four years. This will be the first treatment investigated as part of this thesis.

More recently there has been increased interest in disease-modifying treatments. A disease-modifying treatment would be a treatment that directly improves CFTR function, and therefore ideally it would improve all areas of health affected by CF.[7, 8] The first of these treatments to become available is a CFTR potentiator called ivacaftor, which has been available in the UK since 2012. It is only effective in a subset of people with CF, those with a gating-mutation, and this is approximately 5% of the UK CF population. However, there are a number of combination therapies in the pipeline and it is hoped that when combined with ivacaftor these treatments will be effective in a much larger proportion of the CF population. In fact, one such combination, ivacaftor and lumacaftor, is now available in some countries, but it is not yet generally available in the UK. For this reason, ivacaftor will be the second treatment that we investigate in this thesis, but restricted to its effect in people with a gating mutation.

1.3 Causal Inference of Treatment Effects

In this thesis, our main target of inference is a comparison of the mean outcome if hypothetically everybody received treatment for x years versus the mean outcome if hypothetically nobody received treatment for the same x years. In general, by 'everybody' we are referring to the UK CF population, but we may also be interested in specific subsets of this population in some analyses.

Ignoring all issues of feasibility, cost and ethics, the best way to estimate the causal effect of a treatment would be to perform a RCT. Half of the population would be randomised to receive treatment for x years and the other half would be randomised to not receive treatment over the same time period. There would be a number of additional steps,

such as using a placebo to ensure that people are blinded as to which group they in, but provided that the trial was well-designed and there were no issues during follow-up, we would then just compare the average level of outcome between the two groups at the end of follow-up and this would be our estimate of the causal effect of x years of treatment.

In reality, it is often not possible to perform a RCT and we therefore have to find other ways to obtain estimates of the causal effect of treatment. In observational settings, those receiving treatment will generally not be directly comparable to those not receiving treatment, for example people typically start treatment due to a deterioration in their health. Furthermore, people might not always receive treatment continuously, instead stopping and starting during follow-up, and the reasons for this may depend on their underlying health status, including prior measures of the outcome of interest. This is generally referred to as confounding by indication, and in longitudinal settings, there is often also the issue of time-dependent confounding, which will be introduced in Chapter 4. Depending on the specific situation, there are a number of methods that have been developed that can estimate causal effects from the analysis of observational data even in the presence of confounding by indication or time-dependent confounding. These methods all work in different ways, but overall they aim to remove any differences not due to treatment between those receiving treatment and those not receiving treatment, in order to provide an estimate of the causal effect of treatment.[9] In order to obtain these estimates, it is generally necessary to make a number of assumptions regarding how the data were generated. The treatment effect estimates will only be reliable if these assumptions are valid and it is therefore important to be explicit about any assumptions, so that it is possible to assess whether it is plausible that the results obtained are trustworthy.

1.4 Outline

This thesis is divided into four parts. In the remaining chapter of Part I, we will introduce the UK CF Registry which is the dataset that will be used throughout this thesis. The chapter gives details on how data are collected and contains a description of the key variables that will be important in our analyses. We end the chapter by discussing some of the strengths and weaknesses of the registry.

Part II will focus on the long-term effects of dornase alfa (DNase). Chapter 3 is a systematic review of studies that have investigated the effects of this treatment. This can be used as a starting point for our own investigations, finding out what is already known about the effects of the treatment and where there are gaps in knowledge.

Chapter 4 introduces the different statistical methods that we will consider for our analysis of the UK CF Registry to estimate the long-term effects of DNase. The main challenge of such an analysis is an issue known as time-dependent confounding. We introduce why this is a problem and then consider five methods that have been developed to provide unbiased estimates of long-term treatment effects even in the presence of time-dependent confounding: inverse probability weighting of marginal structural models, history-adjusted marginal structural models, g-computation formula, g-estimation of structural nested models, and sequential conditional mean models. In our analysis of the UK CF Registry we will focus on two outcomes: lung function and the annual number of days receiving intravenous antibiotics. The first of these can be measured as a continuous outcome, but the latter is measured as a count outcome, which brings some specific challenges for analysis and interpretation. We will therefore give details of how all the methods can handle outcomes that are most naturally modelled using non-linear models.

In Chapter 5 we take the five methods introduced in the previous chapter and perform simulation studies to assess their performance in a number of scenarios motivated by challenges faced in the analysis of routinely collected longitudinal data such as that found in the UK CF Registry. All five methods are known to be able to provide unbiased estimates of the treatment effect of interest in ideal settings. However with real data there will generally be some peculiarities that could cause issues in the analysis. It is not clear if some of the methods we consider will perform better than others when analysing the real UK CF Registry, and we therefore perform simulation studies to assess the performance of the methods to help guide an analysis strategy for the real data.

Based on the findings from the simulation studies, in Chapter 6 we perform an analysis of the UK CF Registry to estimate the effects of up to five-years of DNase use on two outcomes. The simulation studies suggested that any of the methods could be suitable for this analysis, depending on what assumptions are considered appropriate, and we therefore present the results from using all five available methods. We end the chapter with a discussion of our findings, comparing these to findings of previous studies identified in the systematic review.

In Part III, we focus on a second treatment, ivacaftor, and as with Part II, we start with a systematic review of studies that have investigated the effects of this treatment.

Chapter 8 will then look at the methods that could be used to analyse the long-term effects of ivacaftor using the UK CF Registry. Unlike with DNase, time-dependent confounding is not an issue for ivacaftor, because all people who are eligible for the treatment are currently receiving it. However, this leads to another challenge: namely that there are no directly comparable people who are not receiving treatment to be able to estimate what would happen to people if they did not receive ivacaftor. In this chapter, we consider different groups that could be used as comparator groups to those currently receiving

treatment and introduce the use of negative controls to assess whether the assumption that these groups are comparable is valid.

In Chapter 9 we then present the analysis of the UK CF Registry to investigate the effects of using ivacaftor for up to four years. As with the analysis of the effects of DNase, we consider two outcomes: lung function and annual number of days receiving intravenous antibiotics. In the analysis, we consider three different statistical models: marginal models, fixed-effects models and mixed-effects models. We compare the results from these three methods and also assess the usefulness of negative controls. Finally, we compare our findings to the findings of the studies identified in the systematic review.

Finally, Part IV of this thesis contains a summary of all of our findings and a discussion of some of the strengths and limitations of our analyses. We also consider some areas of potential future work, before concluding.

The thesis also contains a number of appendices. Appendix A contains a copy of the data resource profile of the UK CF Registry that was published in the International Journal of Epidemiology in 2018.[10] I was a co-author on this paper and it forms the basis of Chapter 2 of this thesis. Appendix B contains a copy of a paper published in Statistics in Medicine in 2018, which is based on the simulation studies found in Chapter 5 of this thesis, and on which I am the first author.[11] Appendix C contains some supplementary tables and figures of the simulation studies of Chapter 5. Appendix D contains a copy of a paper published in the Journal of Cystic Fibrosis in 2018 investigating the long-term effects of DNase on lung function.[12] It is based on the work presented in Chapter 6 of this thesis and I am the first author. Finally Appendices E and F contain supplementary tables and figures of the analyses of DNase (Chapter 7) and ivacaftor (Chapter 9) respectively.

Chapter 2

The UK Cystic Fibrosis Registry

2.1 Introduction

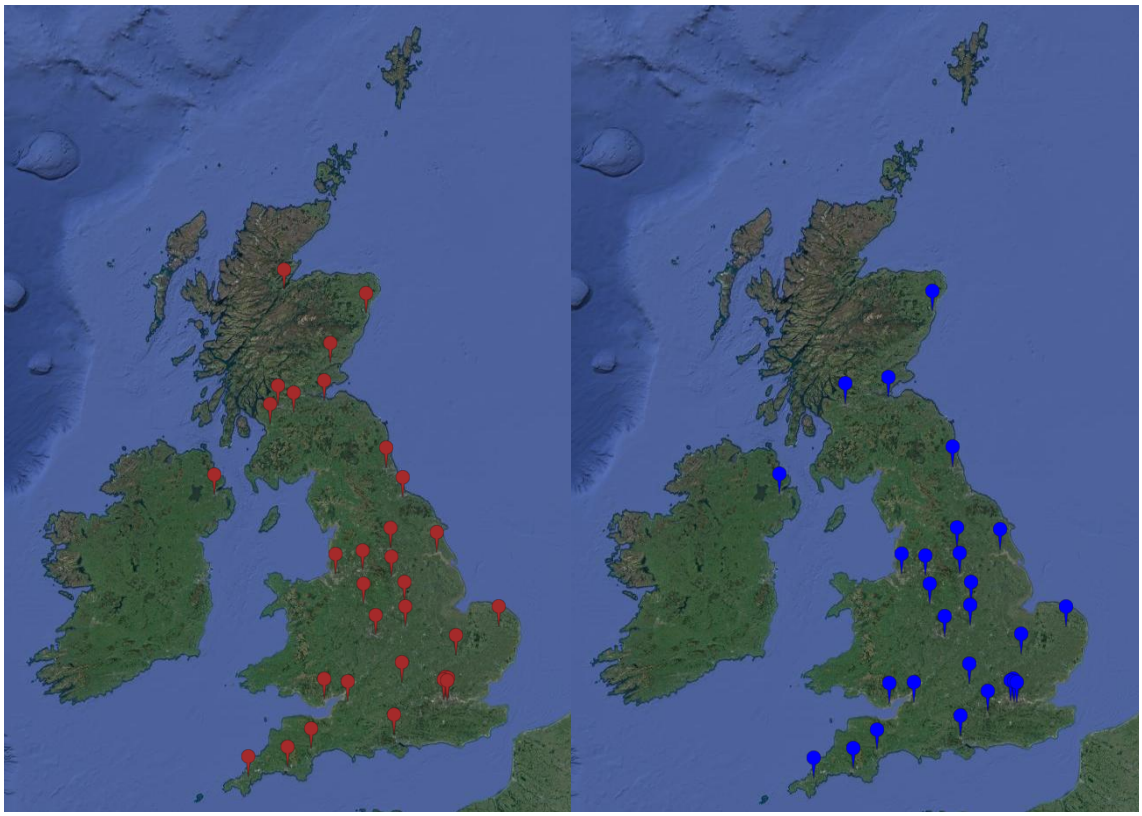
The UK Cystic Fibrosis Registry is a national database sponsored and managed by the Cystic Fibrosis Trust. First established in 1996, it aims to record annual health data on all people with CF in England, Wales, Scotland and Northern Ireland, and to date has captured data on over 12,000 individuals.

This chapter introduces the UK CF Registry, giving details of how the data are collected and some of the key variables that will be used in our analyses. The CF-EpiNet group published a data resource profile of the registry in the International Journal of Epidemiology in 2018. I was a co-author on that paper, and that work forms the basis of this chapter. A copy of the paper can be found in Appendix A.

2.2 Data Collection

In the UK, people with CF are treated in specialist centres. In 2016 data were submitted to the registry from 32 paediatric centres. At between 16 years and 18 years of age, children transfer to an adult specialist centre, and in 2016 data were submitted to the registry from 28 adult centres. The geographical distribution of these centres in the UK can be seen in Figure 2.1. When people with CF first attend a CF centre in the UK, they or their parents consent to information on their health and treatment being collected and stored in the registry. When transitioning to adult services or when changing primary centre of care, there is the opportunity to confirm or withdraw consent.

People with CF attend their specialist centre once a year specifically for an annual review. Data collected during this annual review are then submitted to the UK CF Registry through an online portal. Provided that consent is given, data are therefore usually first collected within one year from diagnosis. Future reviews are then carried out annually, with the aim that the reviews will be approximately twelve months apart. However,



(A) Paediatrics (B) Adults

FIGURE 2.1: Map of CF centres providing data to the UK CF Registry in 2016 (Note that there are four paediatric and four adult centres in London)

according to the Registry's protocol, annual reviews are also supposed to be conducted during a period of disease stability, and therefore, there is some variability in the spacing between annual reviews. Figure 2.2 shows a histogram of the number of months between two consecutive annual reviews since 2007. Overall, over 50% of annual reviews take place between eleven and thirteen months since the last annual review and over 80% take place between nine and fifteen months since the last annual review.

Figure 2.3 gives an overview of the number of individuals and annual reviews per year. The total number of individuals with data in the registry has been steadily increasing since the start of the registry, and the number of annual reviews carried out has also been increasing, except for a sharp dip in 2006, which was the year prior to the nationwide roll-out of the registry. In the most recent three years for which data were available (2014-2016), the number of annual reviews has been quite stable with just under 10,000 reviews carried out annually. Since 2007, the mean number of new people in the registry each year was 469 people (range 318 to 696). The number of deaths and losses to follow-up show more fluctuation, but with an overall upward trend consistent with the increasing total number of people in the registry. A person is defined as lost to follow-up if they miss two consecutive annual reviews.

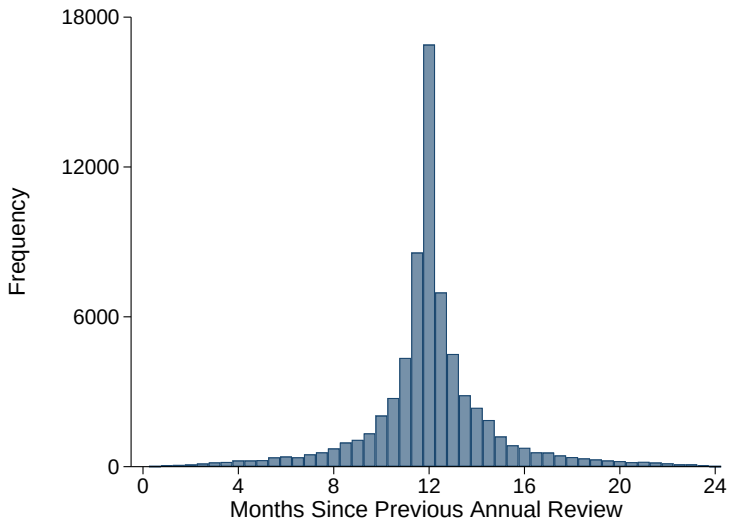
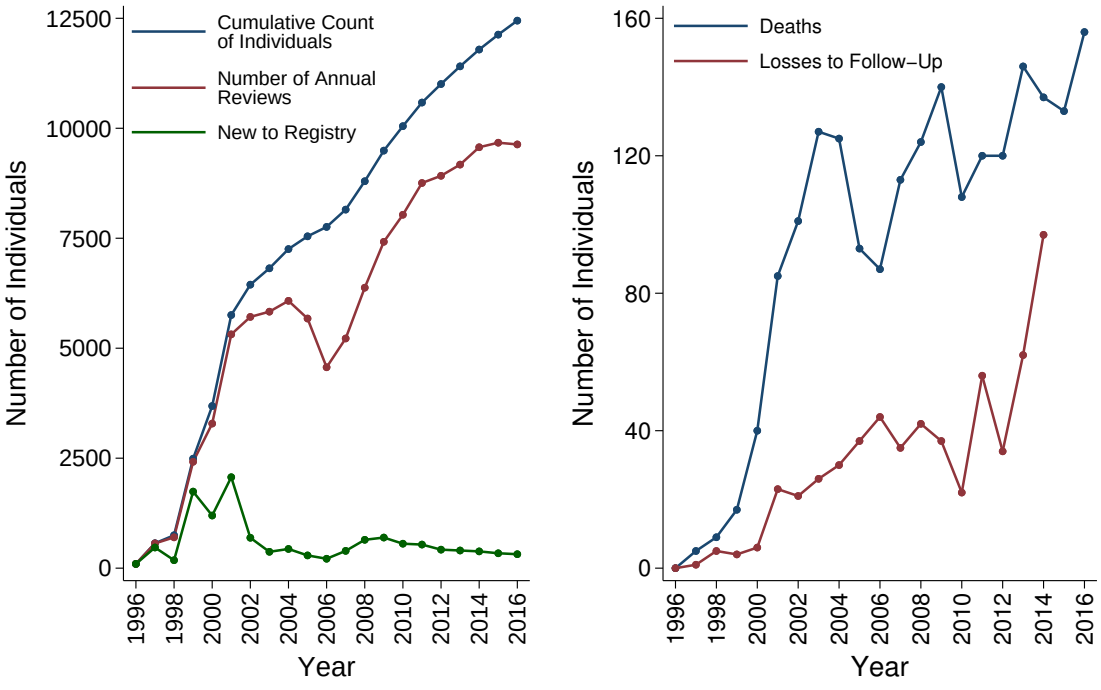


FIGURE 2.2: Histogram of number of months between consecutive annual reviews carried out between 2007 and 2016



(A) (B)

FIGURE 2.3: Summary of number of annual reviews, individuals, deaths and losses to follow-up in the UK CF Registry

At the first annual review, a number of demographic variables are collected, such as sex, ethnicity, genotype class and diagnosis date. Generally these variables will only need to be entered for the first annual review and will then remain fixed for all future annual reviews, but they can be updated to correct any mistakes or if previously missing data becomes available.

Most data collection for the annual review refers to events that have happened since the last annual review (i.e. the past twelve months for most people). For example, the number of hospitalisations since the last review, what treatments individuals have been receiving and any health complications that have happened in the past year. However, a number of variables are specific to the annual review date, such as height, weight and lung function, which are all measured at the annual review.

For Part II of this thesis, only data collected since 2007 is used, i.e. since the UK CF Registry was rolled out nationally, as a number of important variables were not routinely collected prior to this. For Part III, as ivacaftor is a very recently introduced treatment, the analyses are further restricted to data collected since 2008, which corresponds to the four years before the introduction of ivacaftor and four years of data since its introduction. In the following section, we will give an overview of the key variables that we will use for our analyses, restricting these summaries to data collected since 2007.

2.3 Key Variables

2.3.1 Demographic Data

Since 2007, there have been 11,373 unique individuals with at least one annual review in the UK CF Registry. Of these, 6,009 (52.8%) are male and 5,364 (47.2%) are female. In terms of ethnicity, the vast majority are white (10,830 people, 95.2%) and most of the remaining people report other or mixed ethnicity (501 people, 4.4%). Only 39 people are black (0.3%) and 3 people are East Asian (0.03%).

Figure 2.4 shows a histogram of the year of birth of all the people with at least one annual review between 2007 and 2016. Over a third of the individuals were born since the year 2000 (4,097 people, 36.0%), 61.2% were born since 1990 (6,961 people) and 82.4% since 1980 (9,377 people). However, there are 19 people who were born before 1940 (0.2%), and 78 people born before 1950 (0.7%). It is also noticeable that there are very few people with a date of birth in 2016, and this shows that most people born with CF do not have their first annual review until the year after their birth.

Nowadays, most people with CF will be genotyped to record which CF-causing mutation they have; of those with an annual review between 2007 and 2016, only 174 (1.5%) do not have a genotype recorded. The most common CF-causing mutation is f508del, and almost

half of people have two f508del mutations (5,581 people, 50.1%). A further 4,447 people (40.0%) have one f508del mutation and one other mutation, leaving 1,102 people (9.9%) with two non-f508del mutations.

The date of death will clearly not be captured at an annual review, but it is retrospectively added to the previous annual review through linkage with the Office of National Statistics. In total, there were 1,287 deaths recorded before 2018 (11.3%) for those with at least one annual review between 2007 and 2016. Figure 2.5 shows a plot of age at death against date of death; it can be seen that there is a slight upward trend in the age at death between 2007 and 2018, but note that this does not correspond to a formal survival analysis.

2.3.2 Treatment Data

The registry contains data on 30 different types of treatment within which there are 87 specific treatments. In this work we will focus on estimating the long-term effects of two treatments: dornase alfa (DNase) and ivacaftor, and more details about these treatments will be given in Parts II and III respectively.

In general, for each treatment, the Registry contains a marker to show if an individual has been receiving the treatment in the previous year. Therefore, this variable is either marked as "Yes" or is missing, and for this reason all missing values are assumed to be "No", which could affect the quality of the data. For example, Table 2.1 lists the ten most commonly used treatments in 2016, and it can be seen that overall 76.1% of patients were marked as using either Creon 10000 or Creon 25000. The proportion of CF patients who are pancreatic insufficient, and therefore, who would likely be receiving these enzyme replacement treatments is actually over 90%. The reason for this could be that the enzyme replacement treatments are so common that people do not always enter the same data every year, and it is hoped that this issue does not affect other data such as that on DNase use.

There is space to collect more detailed information about the treatment, such as exact start and stop dates, dosage and frequency, but these are not reliably collected and are missing for most people.

The data collected in the registry corresponds to the fact that the individual has been prescribed the specific treatment, but there is no data collected on whether they actually adhered to the treatment. Taking all annual reviews carried out in 2016 as an example, there were only 336 people (3.5%) who were not recorded to receive any treatments that year. The median number of treatments per person was 9 (IQR 7 to 12).

Table 2.1 lists the ten most commonly used treatments in 2016. The most common treatment was DNase, recorded for 5,830 people (60.6%)

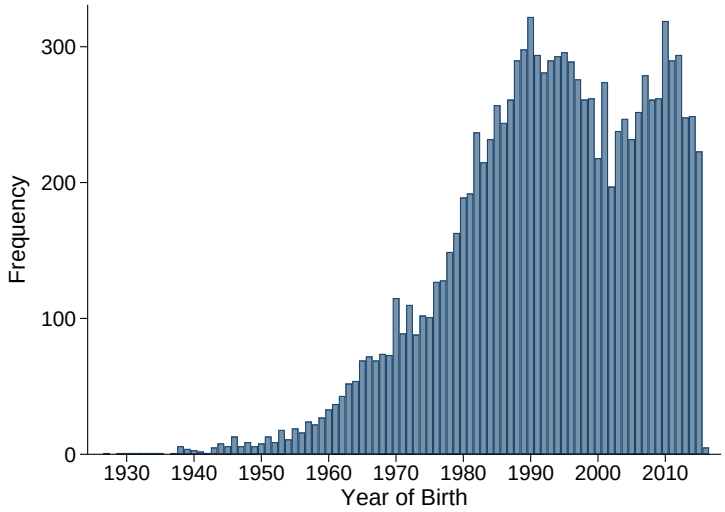


FIGURE 2.4: Histogram of year of birth for people with at least one annual review between 2007 and 2016

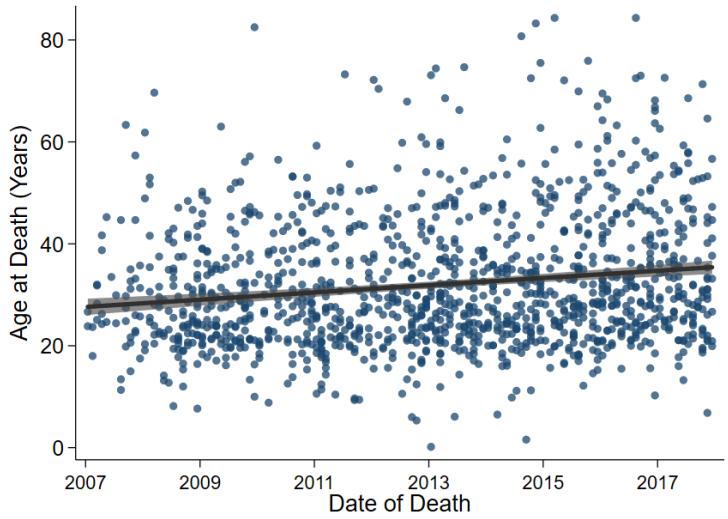


FIGURE 2.5: Scatter plot of age at death against date of death for people with at least one annual review between 2007 and 2016 and date of death before 2018

Treatment Name	Treatment Type	Frequency (%)
Dornase Alfa	Mucolytic	5830 (60.6)
Salbutamol	Bronchodilator	4623 (48.0)
Creon 10000	Pancreatic Enzyme	4438 (46.1)
Proton Pump	Acid Blocker	4261 (44.3)
Vitamin E	Vitamin	4216 (43.8)
Oral Azithromycin	Oral Antibiotic	3501 (36.4)
Hypertonic Saline	Osmotic Therapy	2889 (30.0)
Creon 25000	Pancreatic Enzyme	2853 (29.6)
Vitamin D	Vitamin	2539 (26.4)
Promixin	Inhaled Antibiotic	1996 (20.7)

TABLE 2.1: List of the ten most commonly used treatments in 2016

2.3.3 Lung Function

Measures of lung function are the most common outcomes used in CF studies, and this is because long-term lung function decline is very closely linked to survival, and it is therefore recognised by regulatory authorities as one of the key measures of treatment efficacy.[13] Unlike survival studies, which would generally require very long follow-up, changes in lung function can often be detected over much shorter periods of time, meaning that studies can often be shorter in duration when using lung function as the primary outcome. However, to really prolong the life of someone with CF it would be necessary to not only provide a one-off step-change increase in lung function, but also to provide a slope-change, decreasing the rate of lung function decline. This is often harder to investigate with short-term studies and is one of the key reasons why it is important to continue to investigate the long-term effects of treatments after the clinical trial phase.

Of the different lung function measures available by far the most commonly used is the forced expiratory volume in one second (FEV_1), which is the amount of air a person can blow out in one second.[14, 15] The absolute value of this measure is known to depend on age, sex, height and ethnicity, and therefore it is common to calculate percent-predicted values for these measures. The percent-predicted FEV_1 ($ppFEV_1$) is the ratio of a person's measured FEV_1 and their expected FEV_1 given their age, sex, height and ethnicity, based on a prediction equation. The normal range of $ppFEV_1$ for a healthy lung function would be expected to be between 80% and 100%. There are a number of different formulae that can be used to calculate the percent-predicted values, but for the work in this thesis we will always use those recently proposed by the Global Lung Function Initiative (GLI).[16] These GLI values are thought to deal with changes between childhood and adult lung function better than previously used prediction equations and are recognised as an international gold standard by major international respiratory societies.[17] Because the percent-predicted values are corrected for age, in a healthy population, we would expect the value to stay around 100%, but in people with CF we observe a long-term decline of on average approximately -1.5% per year. However, this trajectory is not linear, and each individual's lung function trajectory can be very variable, with sudden

drops, but also recoveries over time. This is shown in Figure 2.6 where a random selection of people's lung function trajectories are shown as well as the lowess-smoothed line of the population-average lung function trend. The population average slope will not be a good representation of an individual's long-term expected lung function as it is affected by survivor bias, whereby only those with healthy levels of lung function survive to older ages.

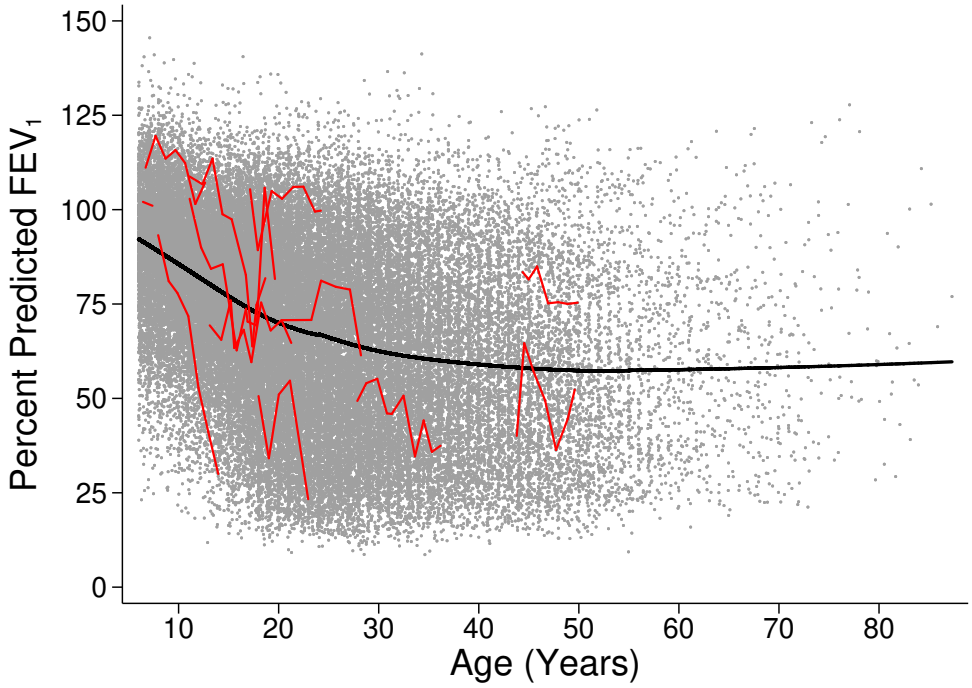


FIGURE 2.6: Graph showing ppFEV$_1$ measures against age between 2007 and 2016 (Highlighted in red are a random selection of individual trajectories and the black line shows the population average)

Two other common measures of lung function are forced vital capacity (FVC) and forced expiratory flow 25–75% (FEF$_{25-75}$). FVC is the total amount of air a person can blow out in one complete breath after taking a deep breath in, and FEF$_{25-75}$ is the flow (or speed) of air coming out of the lung during the middle portion of the FVC.[15] Similarly to FEV$_1$, these two measures can also be adjusted to percent-predicted values (ppFVC and ppFEF$_{25-75}$ respectively). FVC is thought to be less sensitive to acute changes in lung function than FEV$_1$, which is why FEV$_1$ is generally preferred in clinical trials.[18] FEF$_{25-75}$, on the other hand, is believed to be able to pick up the early stages of lung disease better than FEV$_1$, but unfortunately the variability of the measure is much greater than FEV$_1$ meaning that these changes can be lost in the noise of random variability. Figures 2.7 and 2.8 show the FVC and FEF$_{25-75}$ measures from the UK CF registry with a random selection of specific people's trajectories highlighted as well as the lowess-smoothed population-average slope. A recent study showed little difference between the

choice of these three lung function measures when used as an outcome measure, which is not surprising as all three measures would come from the same pulmonary function test and are therefore closely related to each other.[19] One drawback of using either of these measures as an outcome in this thesis is that they are less reliably collected than FEV_1 in the UK CF Registry and therefore the available sample size for these analyses will generally be lower.

Due to the non-linearity of lung function decline, it can be difficult to accurately model long-term lung function trends. Approaches such as the use of splines or stationary Gaussian process models have been proposed to more accurately model individuals' lung function trajectories.[20, 21] For the work in this thesis, we will only be considering treatment effects out to five years, and we will therefore either assume a linear trend over this shorter time-period or assess lung function discretely at yearly intervals.

One of the main drawbacks of these lung function measures is that they cannot be performed reliably in children under the age of six. This means that other outcome measures are generally needed when aiming to investigate treatment efficacy in young children. For the purposes of this thesis we will restrict our analyses to people over the age of six.

There are also other measures of lung function, such as the lung clearance index (LCI), which have been claimed to be better than the more traditional lung function measures at detecting the early signs of lung function deterioration.[22] However, this is not currently captured in the UK CF Registry due to the complexity of the equipment needed to take this measurement.[23]

2.3.4 Exacerbations & Number of Days Receiving Intravenous Antibiotics

As well as the overall long-term decline in lung function, people with CF often suffer exacerbations. These are periods where there is a sudden decrease in lung function often due to infections. There is no universal definition of what constitutes an exacerbation in CF, but it is an outcome that is often used in clinical trials where each trial will have a strict definition of what constitutes an exacerbation in their study. It is important to try to reduce the rate of exacerbations in the CF population as research has shown that in approximately one third of cases lung function measures do not recover to their pre-exacerbation levels.[24, 25]

The UK CF Registry does not specifically collect data on whether a person has suffered an exacerbation or not during a year, but a proxy is available, which is whether the person has received any intravenous antibiotics (IVs) during the year. Generally, if someone suffers an exacerbation they will receive IVs, therefore using this as a proxy should capture almost all exacerbations. However, some people receive IVs even during periods of

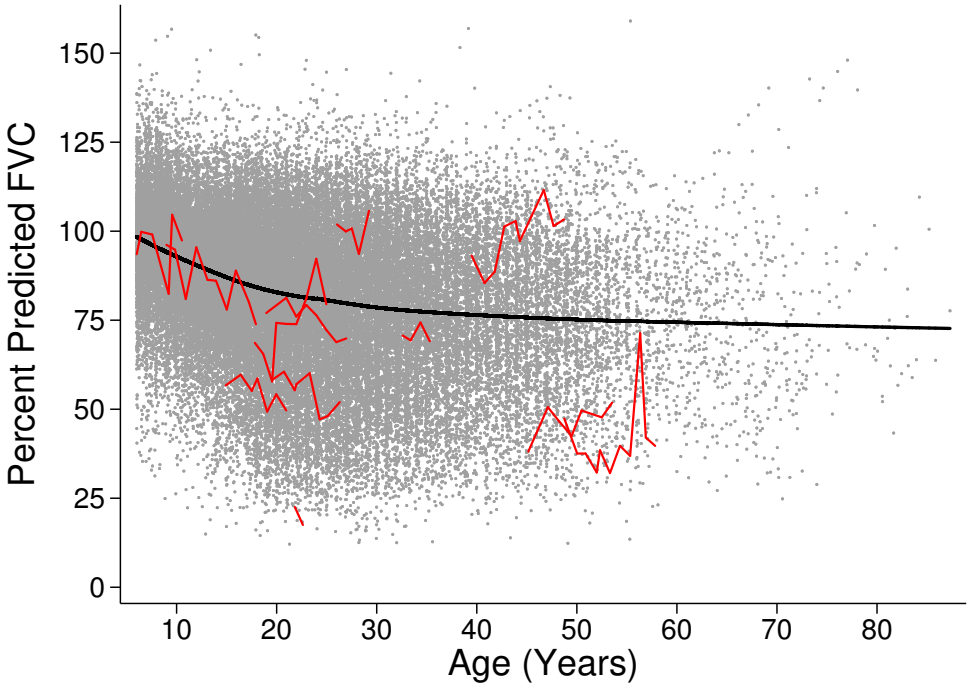


FIGURE 2.7: Graph showing ppFVC measures against age between 2007 and 2016 (Highlighted in red are a random selection of individual trajectories and the black line shows the population average)

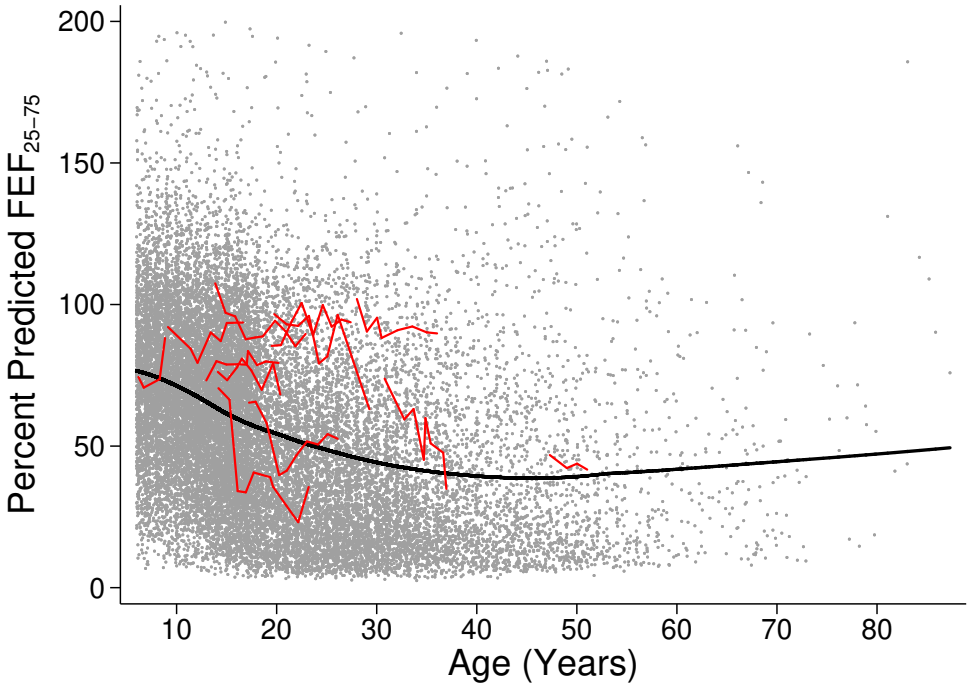


FIGURE 2.8: Graph showing ppFEF$_{25-75}$ measures against age between 2007 and 2016 (Highlighted in red are a random selection of individual trajectories and the black line shows the population average)

disease stability, meaning that using this variable would overestimate the rate of exacerbations. IVs can either be administered at home or in hospital and this data is collected separately in the Registry. However, for the purposes of this thesis, we use the combined total annual number of IV days.

In a given year, approximately half of the people in the UK CF Registry do not have any IV days. For people who do receive IV days the most common practice would be to receive them for two weeks. The number of people with more IV days per year quickly tails off, but there are a handful of patients who are receiving IVs for almost the whole year. This is shown in Table 2.2, which shows a breakdown of the number of annual IV days each person had in 2007.

Annual IV Days	Frequency	%	Cumulative %
0	2,995	56.4	56.4
1-14	810	15.3	71.6
15-28	483	9.1	80.7
29-42	389	7.3	88.1
43-56	307	5.8	93.8
57-70	155	2.9	96.8
70-98	98	1.8	98.6
98-365	74	1.4	100.0

TABLE 2.2: Breakdown of total number of annual IV days observed in 2007

2.3.5 Other Variables

There are a number of other variables, which will not be considered as outcomes of interest in the work presented in this thesis, but which could be important confounders in the analyses. This includes the demographic variables, which have previously been discussed as well as the variables introduced in this subsection.

Cystic fibrosis related diabetes (CFRD) is one of the main long-term complications of CF.[18] The prevalence of CFRD has been steadily increasing in line with the increasing survival age of the population. In 2016, the prevalence of CFRD in those under the age of 10 was 1.2%, in those between 10 and 20 was 15.4%, in those between 20 and 30 was 30.3% and in those over 30 was 38.1%.

People with CF are prone to a number of different respiratory infections. The two most common infections are *Staphylococcus aureus* and *Pseudomonas aeruginosa*, but there are a number of other important infections as well, such as *Aspergillus fumigatus*, meticillin-resistant *Staphylococcus aureus* (MRSA), influenza, *Stenotrophomonas maltophilia* and *Burkholderia cepacia* complex. Each year the annual assessment records whether a specific

infection was present in the lungs in the previous year. Figure 2.9 shows the prevalence of all of these infections by age group in 2016.

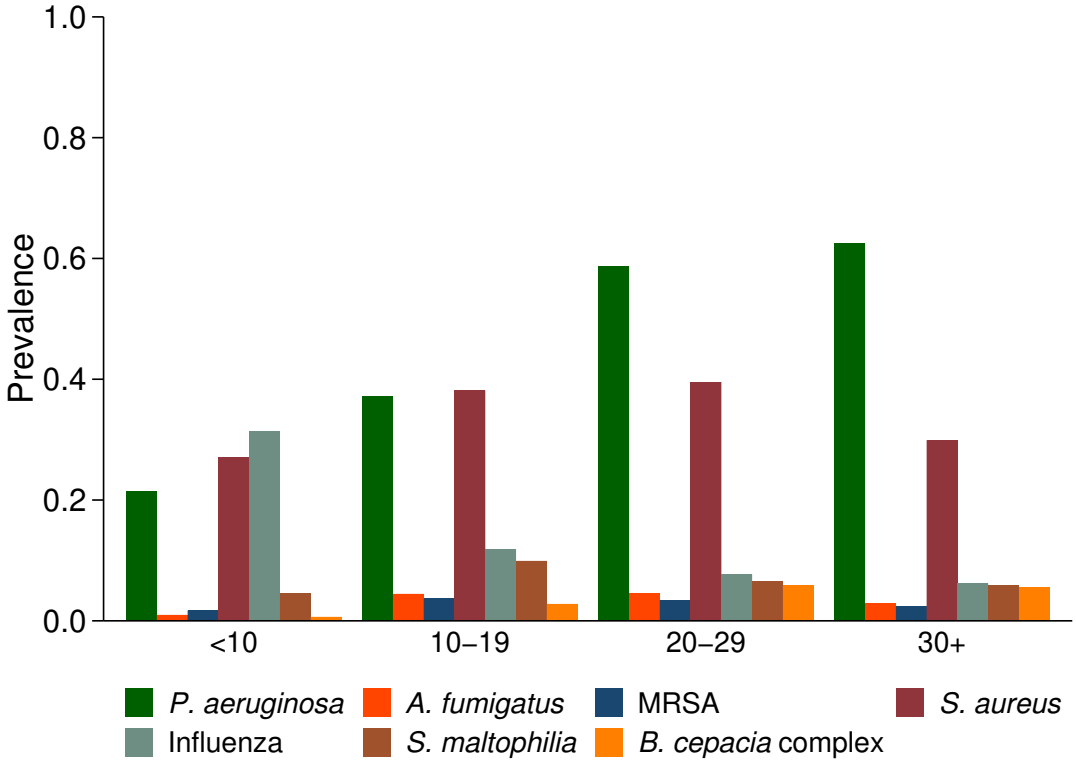


FIGURE 2.9: Bar chart showing prevalence of infections by age group in 2016

Data on whether a person smokes or not is collected at every annual assessment. The number of smokers appears to be very low, but we have no way of knowing how reliably people answer this question. For example, in 2007, only 97 of the 5,311 annual assessments report smoking (1.8%). Exposure to second-hand smoke is also thought to be an important predictor of long-term health in CF, but unfortunately this data is not currently captured in the UK CF Registry.[26]

2.4 Strengths & Weaknesses

One of the main strengths of the UK CF Registry is that it is estimated to contain data on over 99% of people with CF in the UK. Furthermore, new born screening for CF was implemented nationwide in 2007 (since 2003 in Scotland), meaning that since this date almost all people born with CF will be diagnosed soon after birth, allowing continual longitudinal data collection from a very early age. The numbers lost to follow-up are also very small and are generally due to people emigrating. These efforts combined with

the overall population size of the UK means that it is one of the largest CF datasets in the world.[27]

Data are supposed to be collected annually and Figure 2.2 shows good adherence to this protocol. As well as a wide range of demographic data, at each visit a wide range of data are collected, allowing both population level analyses as well as longitudinal analyses of individual health trajectories. Furthermore, the number of variables collected allow for careful adjustment for appropriate covariates in any statistical analyses, reducing the risk of unobserved confounders.

One drawback of annual review data is that any acute changes within a specific year may not be picked up. For example, if a person's lung function suddenly dropped, but then recovered before the next annual review, this would not be picked up in the data collected in the Registry. Some national registries collect data at more regular intervals, such as quarterly or monthly, but there is no evidence that this leads to improved analyses of long-term trends. Other registries also collect data on an encounter basis, which has the benefit of obtaining data on all clinic visits, but has the drawback that sicker patients will tend to end up with more data, which could easily bias analyses if not appropriately accounted for in the analysis. Related to this issue is that in the UK, the annual review is supposed to take place during a time of disease stability. This is what leads to visits not always being equally spaced apart, but even so, a recent study of a single centre in Sheffield found that approximately 20% of visits could have occurred during a period of instability according to their definition.[28]

Finally, although the number of variables that can be collected during an annual assessment is very extensive, and while many important variables including ppFEV_1, IV days and demographics are recorded with little missing data, not all variables are reliably collected. Just taking lung function as an example, Figure 2.6 of ppFEV_1 measures is based on 61,668 observations, whereas Figure 2.8 of ppFEF_{25-75} measures is based on only 23,238 observations. A number of binary variables, such as treatments or infections are collected via a yes tick box, leading to there artificially being no missing data, as all those without a tick are assumed to be no. This has been improved recently and the input system was updated for the 2016 data collection, but it remains a limitation for any data collected prior to this.

Part II

Dornase Alfa

Chapter 3

Dornase Alfa Use in Cystic Fibrosis: A Systematic Review

3.1 Overview

One of the most common symptoms of CF is a build up of mucus in the lungs, which leads to an increased prevalence of bacterial growth in the airways and a decline in lung function.[29] Therefore, it is common for people with CF to use aerosolized mucoactive agents, which help to break down the layer of mucus in the lungs making clearance easier.[18] Recombinant human deoxyribonuclease, commonly known as dornase alfa (DNase), and marketed under the name Pulmozyme®, is one such treatment which was authorised for use in January 1994.[30] Since then, it has become the most commonly used mucolytic treatment in CF in the UK, used by almost 60% of people in 2016.[3]

This chapter contains a systematic review of the studies that have investigated the effects of DNase use in CF, in order to summarise what is already known about the treatment and to identify the gaps in knowledge. This is followed in the next three chapters by work that will look at how the UK CF Registry data can be analysed to answer some of these questions.

3.2 Objectives

The main objective of this systematic review is to identify all studies which investigate the association between DNase use and outcomes for patients with CF.

The aim is then to use these studies, summarising their results, to obtain a broad understanding of what is already known about the overall effect of DNase use over different lengths of time. It was found that many of the studies investigated different outcomes over different lengths of follow-up, and for this reason a qualitative summary of results was deemed more appropriate then a quantitative meta-analysis.

3.3 Methods

The review was carried out with reference to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement which provides guidelines for how to conduct and report a systematic review.[31]

Studies were eligible for inclusion if the following PICOS criteria were met:

- ***P*articipants:** People with CF
- ***I*nterventions:** DNase given at any dose and frequency
- ***C*omparisons:** Placebo, no treatment and/or another mucolytic treatment, e.g. hypertonic saline
- ***O*utcomes:** All outcomes were considered
- ***S*tudy Designs:** All study designs were considered

All studies published prior to July 2009 were considered for inclusion in the Cochrane Review of Dornase Alfa for Cystic Fibrosis, and this review therefore contains references to all available studies.[32] However, the criteria for the review meant that only RCT were actually included in the review. Therefore for the purposes of this review both included and excluded studies were assessed for eligibility.

In order to identify studies carried out since July 2009 two search strategies were used:

- Hand search of two journals: Pediatric Pulmonology and Journal of Cystic Fibrosis
- Electronic search of two databases: MEDLINE and EMBASE.

Pediatric Pulmonology and the Journal of Cystic Fibrosis were hand searched, because in addition to publishing many relevant papers, they also contain the annual proceedings of the North American and European Cystic Fibrosis Conferences respectively.

The search strategy for the electronic databases was for any of the following terms to be included in the title of the article: DNase, d-nase, d nase, rhDNase, dornase, pulmozyme, or deoxyribonuclease. In addition to this, the term 'cystic fibrosis' needed to be included in the title, abstract or key words. Finally the searches were limited to papers published since July 2009, as papers published prior to this have already been identified through the Cochrane Review. The search criteria for the Cochrane Review were very similar to those employed here, including an electronic search of MEDLINE and EMBASE and a hand search of Pediatric Pulmonology and Journal of Cystic Fibrosis. Both the electronic and hand searches were originally performed in January 2016, and again in January 2018.

The abstracts of all papers identified from the Cochrane Review and the database searches were reviewed to ascertain whether any of the PICOS criteria listed above were violated and to exclude these papers from further review. All remaining papers were

then included in the full-text review which involved three elements: ensuring that all the PICOS criteria were fulfilled, extracting key data, and assessing the quality of the study.

For each study the following data were collected: date of publication, type of study, length of DNase administration, sample size, dose and frequency of treatment, baseline age and lung function data, the outcome measures, the results and the statistical methods.

In order to assess the quality of studies in systematic reviews it is common to use the 'risk of bias' tool developed by the Cochrane collaboration.[33] However, this was designed for RCTs, so it was not considered suitable for this review. A systematic review comparing all available tools for evaluating non-randomised studies was carried out in 2003 and out of 194 tools identified they recommended The Newcastle-Ottawa Scale or the Downs & Black scale.[34] Upon review, the Newcastle-Ottawa Scale is only suitable for cohort or case-control studies, whereas the Downs & Black scaled can assess the quality of any randomised or non-randomised study.[35, 36] Therefore each study included in the systematic was assessed against the Downs & Black checklist which contains 27 questions within five sub-scales of quality: reporting, external validity, bias, confounding and power. The scale assigns scores ranging from 0 to 32, where a higher score indicates a higher quality of study.

A qualitative review was carried out to summarise the findings of the effects of DNase on all available outcomes. However no formal meta-analyses were performed due to that fact that there was little overlap between papers containing the same outcomes and lengths of follow-up, making it difficult to synthesize the findings quantitatively.

The electronic and hand searches, the abstract and full-paper reviews, the quality assessment with the Downs & Black scale, and the qualitative review of included studies were carried out by me.

3.4 Results

In total there were 31 studies fulfilling all five PICOS criteria: 22 RCTs and 9 observational studies. Figure 3.1 gives a detailed flowchart of the number of studies identified at each stage of the search process. 43 studies were initially taken forward for the full-text review, but upon closer examination, twelve of them did not fulfil the inclusion criteria. This was mainly due to the studies not containing a suitable control group for comparison. Details of these studies with the reason for exclusion can be found in table 3.1.

Table 3.2 contains a summary of the 31 studies included in the systematic review. The median length of follow-up time was 12 weeks (range 1 week to 4 years) and the median number of patients in each study was 70 (range 9-8200). There were three studies only involving children under six years old, ten more studies only included patients under the

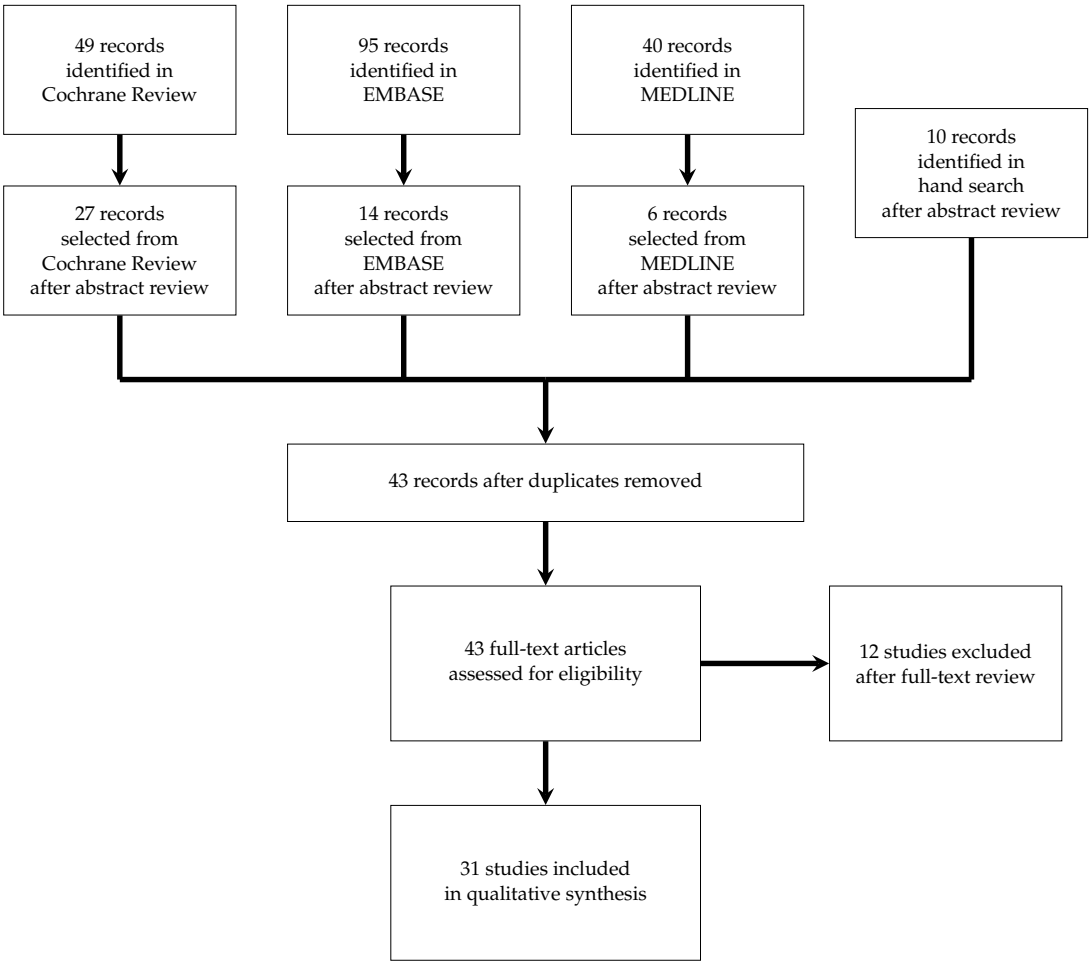


FIGURE 3.1: Flowchart of the number of studies at each stage of the DNase systematic review

Publication Date	Authors	Title	Reason for Exclusion	Full Reference
Jun-95	Heijerman et al.	Effect of rhDNase on lung function and quality of life in adult cystic fibrosis patients	No control group	[37]
Aug-95	Shah et al.	Two years experience with recombinant human DNase I in the treatment of pulmonary disease in cystic fibrosis	No control group	[38]
Feb-01	Furuya et al.	Efficacy of human recombinant DNase in pediatric patients with cystic fibrosis	No control group	[39]
Jul-04	Barker et al.	Effect of DNase on exercise capacity in cystic fibrosis	No control group	[40]
Jan-06	Riethmueller et al.	Recombinant human deoxyribonuclease shortens ventilation time in young, mechanically ventilated children	Not cystic fibrosis patients	[41]
Dec-08	McPhail et al.	Improvements in lung function outcomes in children with cystic fibrosis are associated with better nutrition, fewer chronic pseudomonas aeruginosa infections, and dornase alfa use	No Control Group	[42]
Sep-10	McPhail et al.	Initiation of dornase alfa before age six years is associated with improvements in the rate of decline in lung function in children who are "rapid early decliners"	No control group	[43]
Sep-10	Rozov et al.	Dornase alfa improves the health-related quality of life among Brazilian patients with cystic fibrosis - A one-year prospective study	No control group	[44]
Oct-13	Dasenbrook et al.	Combination inhaled 7% hypertonic saline and rhDNase: Retrospective cohort study of the US CFF national patient registry	No outcome	[45]
Dec-13	Rozov et al.	A first-year dornase alfa treatment impact on clinical parameters of patients with cystic fibrosis: the Brazilian cystic fibrosis multicenter study	No control group	[46]
Oct-15	VanDevanter et al.	Chronic high dose dornase alfa use in CF 1994-2005	No outcome	[47]
Nov-16	Shenoy et al.	The effects of early initiation with dornase alfa on childhood lung function in cystic fibrosis	No control group	[48]

TABLE 3.1: Studies excluded from DNase systematic review after full-text review

age of nineteen, and four studies only considered adults. The remaining fourteen studies included both children and adults. In terms of baseline lung function, two studies only included patients with an FEV_1 less than 40% predicted, for twelve studies the mean FEV_1 at baseline was between 40% and 80% predicted, and for six studies it was greater than 80%. Eleven studies did not report baseline FEV_1.

It should be noted that whilst the standard dose of DNase is now 2.5mg once per day, some earlier studies (before 2000) prescribed 2.5mg twice per day. However, studies have shown similar effects irrespective of dose frequency and for this reason all doses were considered equally in this review.[49, 50]

When assessing the quality of studies, the mean score from the Downs & Black questionnaire was 18.7 (SD 5.9) out of a possible 32. Looking at RCTs and observational studies separately shows that in general RCTs performed better (mean score 20.3 (SD 5.2) compared to 14.9 (SD 5.8). However, this can partly be explained by the weightings given to different sections: observational studies tended to perform well in terms of external bias and power, where the total possible scores were 3 and 5 respectively, whereas RCTs often performed poorly in these two areas, but performed better in internal bias and confounding, which respectively could contribute 6 and 7 points to the total score. Figure 3.2 presents a histogram showing the distribution of the Downs & Black scores.

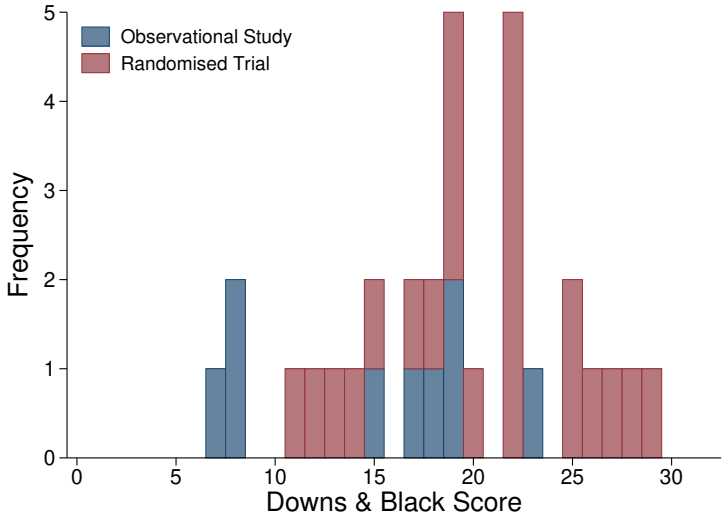


FIGURE 3.2: Histogram of Downs & Black scores for studies included in the DNase systematic review

Date	Authors	Title	Type of Study[1]	Quality[2]	Follow-Up (Weeks)	Sample Size	Age (Years)[3]	Baseline FEV_1 (%)[3]	Outcomes[4]	Full Reference
Mar-92	Hubbard et al.	A preliminary study of aerosolized recombinant human deoxyribonuclease I in the treatment of cystic fibrosis	RCT	20	1	16	27.0±3.3	45.0±20.0	L	[51]
Jul-93	Ranasinha et al.	Efficacy and safety of short-term administration of aerosolized recombinant human DNase I in adults with stable stage cystic fibrosis	RCT	25	1.5	71	26.2±7.9 (16 – 55)	46.8±19.4 (17 – 88)	L Q	[52]
Jul-93	Ramsey et al.	Efficacy and safety of short-term administration of aerosolized recombinant human deoxyribonuclease in patients with cystic fibrosis	RCT	25	1.5	181	19.2±9.5 (8 – 65)	62.5±25.7	L Q O	[53]
Sep-94	Fuchs et al.	Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis	RCT	28	24	968	19.9±9.0 (5 –54)	60.7±26.1 (17 – 142)	L E Q	[49]
Jun-95	Shah et al.	Recombinant human DNase I in cystic fibrosis patients with severe pulmonary disease: A short-term, double-blind study followed by six months open-label treatment	RCT	22	2	70	25.5±9.4 (5 – 48)	21.5±5.9 (12 – 39)	L Q	[54]
Feb-96	Laube et al.	Effect of rhDNase on airflow obstruction and mucociliary clearance in cystic fibrosis	RCT	19	1	20	26.0 (18 – 44)	54.9	L O	[55]
Jun-96	Wilmott et al.	Aerosolized recombinant human DNase in hospitalized cystic fibrosis patients with acute pulmonary exacerbations	RCT	22	2	80	19.9±9.2 (5 –)	40.0±18.9	L Q	[56]
Oct-96	McCoy et al.	Effects of 12-week administration of dornase alfa in patients with advanced cystic fibrosis lung disease	RCT	27	12	320	26.0±9.4 (7 – 57)	21.7±5.4 (9.2 – 39.2)	L E Q	[57]
Sep-99	Weck et al.	Efficacy of DNase in individual children using the N-of-1 study design	RCT	14	4	16	10.2 (5 – 17)	(– 40)	L Q	[58]
Jul-00	Robinson et al.	Effect of a short course of rhDNase on cough and mucociliary clearance in patients with cystic fibrosis	RCT	19	1	13	25.1±5.3 (18 – 38)	62.8±25.2 (27 – 103)	L O	[59]

[1] Randomised Controlled Trial or Observational Study
[2] Calculated using Downs & Black scale, maximum score possible is 32
[3] Mean±SD (Range)
[4] The abbreviations refer to the following outcomes: **L**ung function, **E**xacerbations, **Q**uality of life, **B**acterial infection, and **O**ther

Date	Authors	Title	Type of Study[1]	Quality[2]	Follow-Up (Weeks)	Sample Size	Age (Years)[3]	Baseline $FEV_1(\%)$[3]	Outcomes[4]	Full Reference
Apr-01	Shah et al.	A case-controlled study with dornase alfa to evaluate impact on disease progression over a 4-year period	Obs	15	208	60	25 (17 – 38)	42.5±13.1	L E B	[60]
Oct-01	Suri et al.	Comparison of hypertonic saline and alternate-day or daily recombinant human deoxyribonuclease in children with cystic fibrosis: a randomised trial	RCT	22	12	44	12.6±2.8 (7 – 17)	48.0±15.0 (14 – 77)	L E	[50]
Dec-01	Quan et al.	A two-year randomized, placebo-controlled trial of dornase alfa in young patients with cystic fibrosis with mild lung function abnormalities	RCT	29	96	474	8.3±1.5 (6 – 10)	85.0±30.0	L E O	[61]
Mar-02	Ballman & von der Hardt	Hypertonic saline and recombinant human DNase: A randomised cross-over pilot study in patients with cystic fibrosis	RCT	15	3	14	13.3±2.9		L	[62]
Sep-02	Robinson	Dornase alfa in early cystic fibrosis lung disease	RCT	26	96	474	8.3±1.5 (6 – 10)	85.0±30.0	L	[63]
Nov-03	Hodson et al.	Dornase alfa in the treatment of cystic fibrosis in Europe: A report from the epidemiologic registry of cystic fibrosis	Obs	19	104	4299	15.0±8.3 (6 –)	75.7±24.3	L E	[64]
Dec-03	ten Berge et al.	DNase in stable cystic fibrosis infants: A pilot study	RCT	18	2	9	1.4±0.4 (0 – 2)		L	[65]
Mar-04	Paul et al.	Effect of treatment with dornase alpha on airway inflammation in patients with cystic fibrosis	RCT	19	156	105	11.9±6.2 (5 – 37)	96.9±14.0	O	[66]
Jun-04	Adde et al.	Hypertonic saline X recombinant human DNase: A randomised cross-over study in 18 cystic fibrosis patients	RCT	13	4	18	(8 – 25)	48.0±16.6	L B	[67]
Sep-06	Frederiksen et al.	Effect of aerosolized rhDNase (Pulmozyme®) on pulmonary colonization in patients with cystic fibrosis	RCT	17	52	72	8.4 (0 – 25)		L B	[68]
Sep-08	Konstan	Dornase alfa and progression of lung disease in cystic fibrosis	Obs	18	104	6697	(8 – 38)	84.2	L	[69]

[1] Randomised Controlled Trial or Observational Study
[2] Calculated using Downs & Black scale, maximum score possible is 32
[3] Mean±SD (Range)
[4] The abbreviations refer to the following outcomes: **L**ung function, **E**xacerbations, **Q**uality of life, **B**acterial infection, and **O**ther

Date	Authors	Title	Type of Study[1]	Quality[2]	Follow-Up (Weeks)	Sample Size	Age (Years)[3]	Baseline $FEV_1(\%)$[3]	Outcomes[4]	Full Reference
Sep-09	Kraemer et al.	Long-term effect of dornase alfa on lung function in patients with CF evaluated over an age period of 5 to 18 years	Obs	8		170	(5 – 18)		L B	[70]
Sep-09	Castile et al.	Effects of nebulized recombinant human deoxyribonuclease (dornase alfa) in infants with CF evaluated using infant pulmonary function testing and high resolution computerized tomographic imaging of the chest	RCT	12	26	24	0.8±0.6	$FEV_{0.5}$ 98.1±13.4	L	[71]
Nov-09	Minasian et al.	Comparison of inhaled mannitol, daily rhDNase and a combination of both in children with cystic fibrosis: a randomised trial	RCT	22	12	20	13.2±2.4	64.4±10.4	L	[72]
Mar-10	Bonestroo et al.	No positive effect of rhdnase on the pulmonary colonization in children with cystic fibrosis	Obs	19	78	70	9.5±4.1 (1 – 18)	87.9±20.1 (47 – 143)	L E B	[73]
Apr-11	Amin et al.	The effect of dornase alfa on ventilation inhomogeneity in patients with cystic fibrosis	RCT	22	4	17	10.3±3.4 (6 –18)	90.7±9.1	L	[74]
Jun-11	Konstan et al.	Clinical use of dornase alfa is associated with a slower rate of FEV_1 decline in cystic fibrosis	Obs	23	104	8200	14.5±6.5 (8 – 38)	83.7±23.0	L	[75]
Nov-11	Pasta et al.	A mixed-effects piecewise linear model of the rate of lung function decline before and after the clinical use of dornase alfa in an observational study of cystic fibrosis	Obs	17	104	8200	(8 – 38)		L	[76]
Sep-14	Singh et al.	Clinical effectiveness of dornase alfa in children six years of age and younger with cystic fibrosis	Obs	7	130		(– 6)		L	[77]
Sep-14	Singer et al.	Effects of dornase alfa and hypertonic saline on gas mixing in lung periphery in children with cystic fibrosis	RCT	11	4	17			L	[78]
Apr-15	Stuckey et al.	The effect of dornase alfa (rhDNase) on recurrent gram negative infections in adult cystic fibrosis lung transplant recipients	Obs	8	78	52			B	[79]

TABLE 3.2: Key details of studies included in systematic review of DNase

[1] Randomised Controlled Trial or Observational Study
[2] Calculated using Downs & Black scale, maximum score possible is 32
[3] Mean±SD (Range)
[4] The abbreviations refer to the following outcomes: **L**ung function, **E**xacerbations, **Q**uality of life, **B**acterial infection, and **O**ther

There were many different outcomes measured across the studies, but these were categorised into five groups: lung function, exacerbations, quality of life, bacterial colonisation, and other outcomes. The following sections detail the findings in each of these areas.

3.4.1 Lung Function

All but two studies contained results relating to the effect of DNase on lung function. The most common lung function measurements were FEV_1 and FVC, but there was a lot of variability between papers on which measurements were used.

Children

There were thirteen studies whose participants were under the age of 18. Of these, two studies only investigated infants under 2 years old. Due to their age, this group of patients is unable to perform the same spirometry tests as used in older populations and different measurements are used to assess lung function. The RCT by ten Berge et al. used maximal flow at functional residual capacity as the primary outcome, and they found that after four weeks of DNase treatment it increased by an average of 72ml/s (95% confidence interval [CI] 38 to 107, $p = 0.002$), whereas the change in those treated with placebo was not statistically significant.[65] However, they did not formally test whether the DNase and placebo groups differed. The study by Castile et al. used several different infant pulmonary function tests as well as controlled ventilation high resolution computerised tomographic chest imaging, but reported that there was no detectable difference in any of these measurements between DNase and placebo patients over the six month follow-up period.[71]

In older children, three studies reported no statistically significant effects of DNase on FEV_1. Over a four week period Amin et al. found that in the DNase group the absolute change in predicted FEV_1 was 0.076% higher than in the placebo group ($p = 0.97$), and Bonestroo et al. found the difference to be 4.3% ($p = 0.22$) over a one year period.[73, 74] Kraemer et al. also noted no statistically significant effects of DNase on any lung function measures, but did not specify any of the estimates or p-values.[70] In contrast to these results, the papers by Quan et al. and Robinson et al., which were actually based on the same RCT, found that after 96 weeks those receiving DNase had maintained very similar FEV_1 levels, whereas those receiving placebo had worsened, resulting in a difference in absolute FEV_1 change of 3.2% ($p = 0.006$).[61, 63] A forest plot of these results can be seen in Figure 3.3.

Two of the above studies included results of other lung function tests. Quan et al. reported that FEF_{25-75} was 7.9% higher in the DNase group ($p = 0.0008$) and V_E50 was 8.2% higher ($p = 0.0002$), but FVC was only 0.7% better in the DNase group, which was not

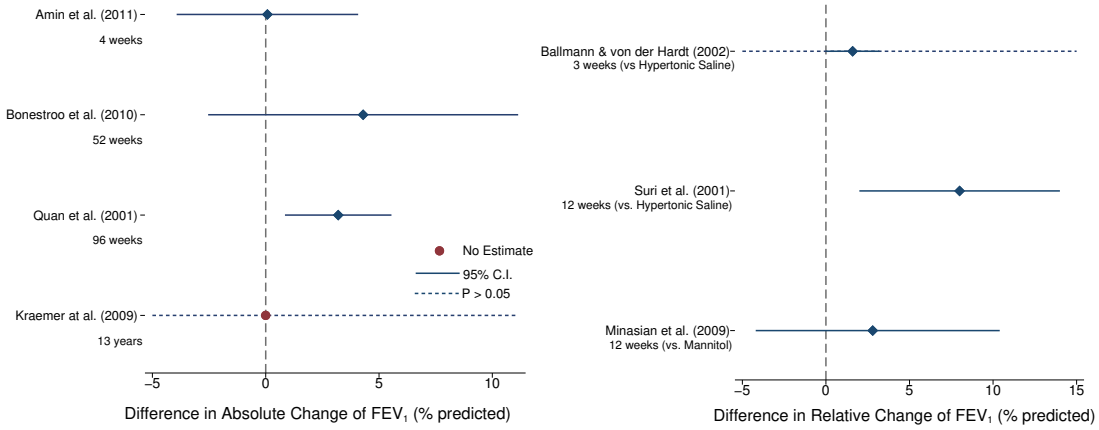


FIGURE 3.3: Forest plot of studies estimating effect of DNase on FEV_1 in children (Some CIs are approximations based on reported p-values)

statistically significant ($p = 0.51$).[61] Amin et al. also reported on FEF_{25-75} and FVC, again finding a positive treatment effect on FEF_{25-75} of 6.09% ($p = 0.03$), but FVC was actually estimated to be 3.61% worse in the DNase group, though this result was not statistically significant ($p = 0.14$).[74] These results can be seen in the forest plot in Figure 3.4.

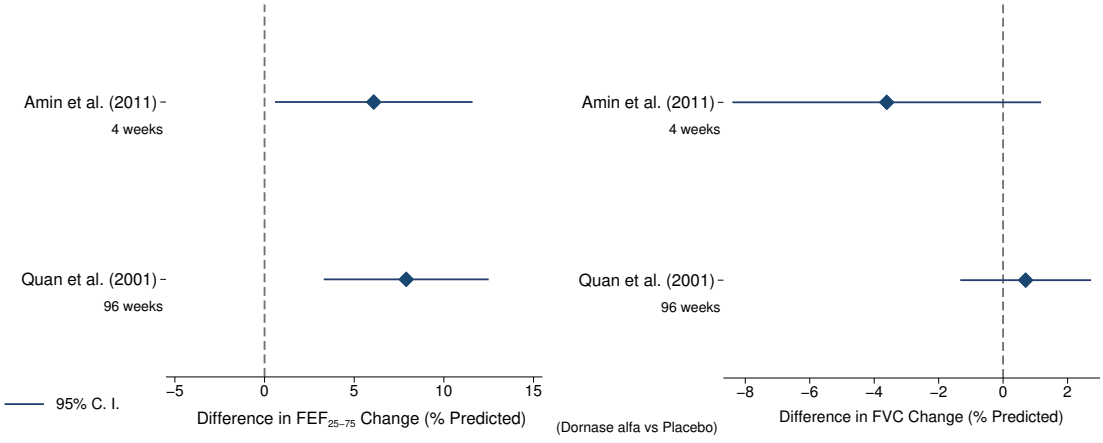


FIGURE 3.4: Forest plot of studies estimating effect of DNase on $ppFEF_{25-75}$ and ppFVC in children (Some CIs are approximations based on reported p-values)

The study by Singh et al. produced equivocal results, as they used two methods to assess the effect of DNase on lung function: a propensity score adjustment method showed a statistically non-significant difference ($p > 0.05$), but using the centre-specific prescription rate of DNase as an instrumental variable showed that $ppFEV_1$ was statistically significantly higher in the DNase takers ($p < 0.01$).[77]

Compared to the population average effects in all of the above studies, Weck et al. used an N-of-1 study design to individually assess for each patient whether they would benefit

from DNase. Of the sixteen children included in their study the results suggested that only four (25%) benefited from DNase.[58]

Most studies compared DNase to placebo or no treatment. However, two studies, whose results are shown in Figure 3.3, compared DNase to hypertonic saline and one study compared it to mannitol, both of which are other available mucolytic treatments. The study by Ballmann & von der Hardt found that over three weeks DNase increased FEV_1 by an average of 9.3% when compared to baseline ($p < 0.05$), but the difference was not statistically significant when compared to hypertonic saline.[62] Conversely, Suri et al. found that over twelve weeks the relative change of FEV_1 was 8% higher in the DNase group compared to the hypertonic saline group (95% CI 2 to 14, $p = 0.01$).[50] In the study comparing DNase to mannitol, DNase was shown to increase FEV_1 by 7.2% ($p = 0.029$) over 12 weeks, but this increase was slightly worse than the increase seen when receiving mannitol; the increase being 2.8% higher in the mannitol period than the DNase period (95% CI -4.2% to 10.4%, $p = 0.42$).[72]

Ventilation inhomogeneity measured via multiple-breath washout tests was an outcome for three studies. Two of the studies used LCI and reported a statistically significant effect of DNase on ventilation inhomogeneity: Amin et al. showed that four weeks of DNase treatment statistically significantly improved LCI compared to placebo by an average of 0.9 ($p = 0.02$), and Kraemer et al. also reported a statistically significant improvement in LCI for DNase users ($p = 0.024$).[70, 74] Singer et al. measured the average alveolar phase III slopes during the wash-out tests and found that the slope decreased by an average of 63.1% (95% CI 20.0 to 121.6) when using DNase.[78]

General Population

When considering adults and children three studies only administered DNase for one week, but there were some apparent treatment effects even over this short time period. Laube et al. noted that at the end of six days the relative change in $ppFEV_1$ and FVC were 11.2% and 12.3% higher respectively in the DNase group (both $p < 0.05$).[55] Robinson et al. found that in DNase users compared to placebo the mean percent change in predicted values from baseline was 4.1% for FEV_1, 7.6% for FVC and 3.8% for FEF_{25-75}, but only the change in FVC was statistically significant ($p < 0.05$).[59] Hubbard et al. reported similarly that taking DNase resulted in greater improvement in both FEV_1 and FVC (both $p < 0.01$).[51]

Two more studies had 1.5 weeks follow-up and found similar results as above. Ranasinha et al. reported that for the DNase group compared to placebo the relative change from baseline in FEV_1 was 13.5% (95% CI 7.6 to 19.4, $p < 0.001$) and FVC was 4.9% ($p > 0.05$).[52] In the same measurements, Ramsey et al. found a 15.4% better improvement in FEV_1 and a 11.3% better improvement in FVC (both $p < 0.001$).[53]

The study by Fuchs et al. had six months follow-up, but again the results are similar to those above. They found that the relative change in ppFEV$_1$ was 5.8% better ($p < 0.01$) in the DNase group, and FVC was 3.8% better ($p = 0.01$).[49]

Two studies specifically investigated patients with severe pulmonary disease, only enrolling patients with an FEV$_1$ less than 40% predicted. Over two weeks the study by Shah et al. showed a 2.8% better relative change in ppFEV$_1$ and a 4.9% better relative change in ppFVC, but neither of these were statistically significant.[54] Over twelve weeks, the study by McCoy et al. showed statistically significant results in both of these measures: a 7.3% better relative change in ppFEV$_1$ ($p < 0.001$) and a 5.1% better relative change in ppFVC ($p < 0.01$).[57]

The study by Wilmott et al. only recruited patients currently suffering from an exacerbation, but over fourteen days of treatment there were no statistically significant differences in either FEV$_1$ or FVC.[56] This and all the above results can be seen summarised in Figure 3.5.

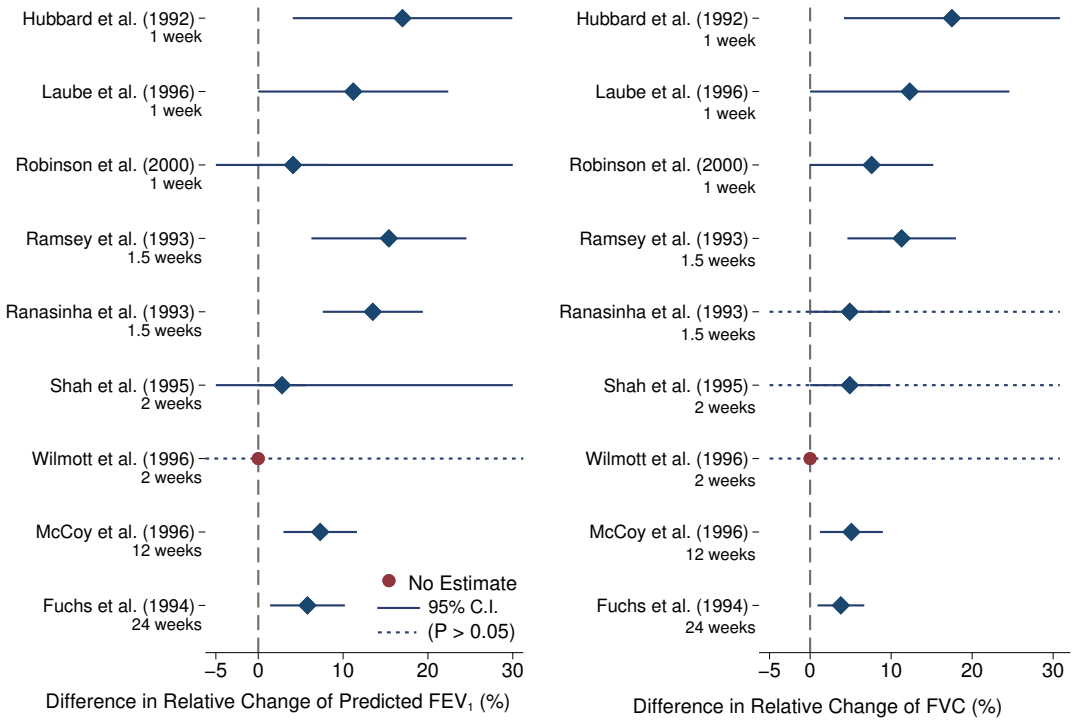


FIGURE 3.5: Forest plot of studies estimating relative effect of DNase on ppFEV$_1$ and ppFVC (Some CIs are approximations based on reported p-values)

Adde et al. compared the effects of DNase and hypertonic saline, but the results showed no statistically significant effect of DNase compared either to baseline or to hypertonic saline.[67]

Looking at longer follow-up (52 weeks) Frederiksen et al. reported the median absolute increase in ppFEV$_1$ was 6.4% better in DNase users than the placebo group after one year ($p < 0.05$).[68] Using the same measurement Hodson et al. found that after one year the difference between treated and untreated was 3.6% (95% CI 1.8 to 5.3), but by two years this difference had become slightly attenuated to 2.5% (95% CI 0.7 to 4.4).[64] A forest plot of these results can be seen in Figure 3.6.

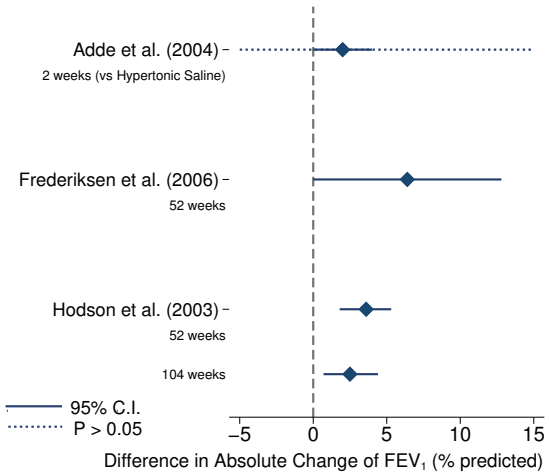


FIGURE 3.6: Forest plot of studies estimating absolute effect of DNase on FEV$_1$ (Some CIs are approximations based on reported p-values)

Unlike all of the above studies, which investigated change in lung function at specific time points, four studies (Shah et al. [2001], Konstan [2008], Pasta et al. [2011] and Konstan et al. [2011]) instead attempted to investigate the effects of DNase on the slope of lung function decline. All of these studies noted similarly to the findings above that upon initiation of DNase there was an acute improvement in FEV$_1$. Furthermore the mean rate of FEV$_1$ decline also appeared to be attenuated in those taking DNase.[76] In the first study by Konstan, DNase was estimated to result in a 46% relative reduction ($p < 0.001$) in the annual rate of FEV$_1$ decline.[69] However the later study by Konstan et al. reported smaller reductions: a 32.1% relative improvement in children ($p < 0.001$), and a 28.8% relative improvement in adults ($p = 0.068$).[75] However, none of these three studies formally compared these improvements with the group of non-users. Shah et al. used a matched study design to compare the change in slope to a group of non-users over a four year period and found that in DNase users the annual rate of ppFEV$_1$ decline improved by 0.93% compared to a worsening rate of decline in non-users of 1.43% ($p = 0.002$).[60]

3.4.2 Exacerbations

Four RCTs and three observational studies have been published which included exacerbations as an outcome. Unfortunately there is no standardised definition of what constitutes an exacerbation and the criteria did vary between studies. Figure 3.7 presents a

forest plot summarising the findings from these seven studies.

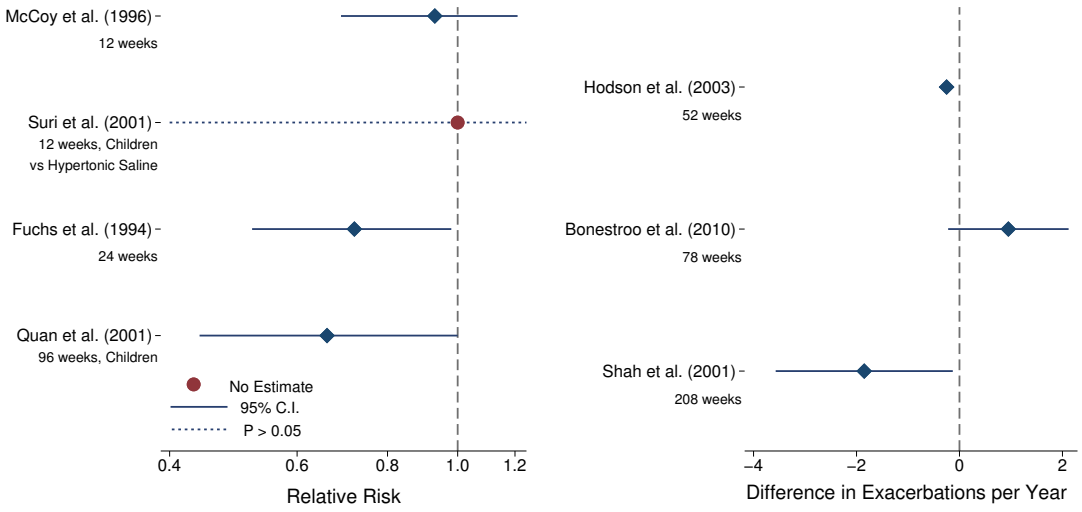


FIGURE 3.7: Forest plot of studies estimating effect of DNase on pulmonary exacerbations (Some CIs are approximations based on reported p-values)

Two of the RCTs only involved children. The study by Suri et al. reported no difference in the effect of DNase or hypertonic saline on the number of exacerbations during the twelve week follow up.[50] The papers by Quan et al. and Robinson et al. reported that over a 96 week period the relative risk (RR) of a respiratory tract infection in the DNase group was 0.66 (95% CI 0.44 to 1.00, $p = 0.048$) compared to placebo.[61, 63]

The other two trials contained a larger age range of patients, but the findings were similar. Over a twelve week period McCoy et al. found no evidence of a difference in risk of exacerbations when using DNase (RR 0.93, 95% CI 0.69 to 1.21, $p = 0.52$), whereas with a follow-up time of 24 weeks Fuchs et al. reported a RR of 0.72 (95% CI 0.52 to 0.98, $p = 0.04$).[49, 57]

The three observational studies by Hodson et al., Bonestroo et al. and Shah et al. had longer follow-up times of one year, one and a half years and four years respectively. Shah et al. reported that DNase users suffered 1.85 fewer exacerbations per year than non-users ($p = 0.035$), whereas the study by Bonestroo et al. reported that children using DNase suffered 0.95 more exacerbations per year than non-users ($p = 0.11$).[60, 73] However, neither of these studies adjusted for patients' average number of exacerbations prior to starting treatment. The study by Hodson et al. did adjust for baseline exacerbation rates and found that those taking DNase had 0.25 (95% CI 0.12 to 0.39) fewer exacerbations per year compared to those not on treatment.[64]

3.4.3 Quality of Life

Well-being and dyspnoea (breathlessness) were measured in seven studies. Five of these studies (Ranasinha et al., Shah et al., Wilmott et al., McCoy et al. and Weck et al.) reported no statistically significant difference in either outcome during follow-up, but they did not give any further details.[52, 54, 56–58]

However, despite the trial by Ramsey et al. also only containing 1.5 weeks follow-up, they reported a statistically significant change in dyspnoea (-45.8% DNase compared to -7.6% placebo, ($p < 0.05$). They also reported that several measures of well-being improved in the DNase group during the trial but none of them reached statistical significance.[53]

The trial by Fuchs et al. contained 24 weeks' follow-up and a larger number of patients than the other studies. It reported statistically significant improvements ($p < 0.05$) in both dyspnoea (on a 100mm scale a 2.1mm decrease in the DNase group compared to a 0.4mm increase in the placebo group) and well-being (on a 5-point scale a 0.019 improvement over baseline for the DNase group compared to a deterioration of 0.058 in the placebo group).[49]

3.4.4 Bacterial Colonisation

There were two RCTs and four observational studies which investigated the effects of DNase on bacterial colonisation of the lungs.

Only the trial by Frederiksen et al. reported evidence of an impact of DNase on bacteria cultures. They reported that after one year of treatment bacteria cultures were found in 72% of patients randomised to DNase compared to 82% of those randomised to placebo ($p < 0.05$).[68] However, they do not report the baseline figures for comparison.

The other RCT by Adde et al. only administered DNase for four weeks and compared it to hypertonic saline. Both *P. aeruginosa* and *S. aureus* growth were investigated, but over this short time-frame there was no statistically significant change in either group.[67]

The study by Bonestroo et al. observed patients for one and a half years but found no statistically significant difference in the percentage of positive cultures at the end of follow-up after adjusting for baseline values (2% increase in users compared to a 4% increase in non-users).[73] Similarly Shah et al. followed patients for four years, but observed no systematic change in bacterial colonisation.[60] Kraemer et al. investigated the onset of chronic *P. aeruginosa* infection during childhood and found that among users of DNase the median age of this was 7.3 years, compared to 13.3 years in non-users. However, this difference was not statistically significant.[70]

Finally the study by Stuckey et al. specifically investigated lung transplant recipients. The incidence of gram-negative infections over one year was found to be similar (20%

DNase vs 23% non-DNase) and the median time to infection was also not found to differ (32 days [IQR 14-46] DNase vs 38 days [IQR 13-219] non-DNase).[79]

3.4.5 Other Outcomes

Two studies investigated whether there was any evidence of DNase use improving mucociliary clearance, which was measured as the rate of retention of a dose of radioactivity. Both studies were RCTs with one-week follow-up time. Robinson et al. found an absolute increase of 2.9% in total lung clearance over 150 minutes after one week's use of DNase, compared to a decrease of 1.6% in the placebo group, but this difference was not statistically significant ($p = 0.2$).[59] The study by Laube et al. also reported that the mean percent retention remained unchanged between the two study groups. This study also investigated changes in aerosol distribution homogeneity, but again found no statistically significant differences between the treatment groups.[55]

The study by Paul et al. was the only study to investigate airway inflammation. When looking at the percentage of neutrophils in a bronchoalveolar lavage fluid sample, they noted that there was no change between the baseline and 3-year follow-up percentages in those randomised to DNase, whereas there was a statistically significant increase ($p < 0.02$) in the control group. However a test to formally compare the two groups was not performed.

None of the studies in the systematic review were powered to study mortality as an outcome and in most studies there were no deaths. In studies where deaths occurred they were reported as adverse events and no statistical analyses were performed to compare mortality between groups.

3.5 Discussion

3.5.1 Findings

Even though the results were not always statistically significant, all studies investigating the effect of DNase on lung function have found that FEV_1 is better among DNase users than among non-users. However, looking at time-trends does suggest that the benefits of DNase are most apparent immediately after initiating DNase and become more attenuated over time, although at two years there still appears to be a small, but statistically significant, benefit. The findings seem to be similar among both adults and children with one study suggesting it may even be beneficial in infants under two years old.

The few studies that investigated change in slope of FEV_1 decline showed promising results. However, only one of them formally compared DNase users and non-users.

The results for other lung function measures were equivocal and were measured in fewer studies, making it hard to draw conclusions.

Four of the seven studies analysing exacerbation rates reported beneficial results. However, all the improvements were modest with only borderline statistical significance. It is difficult to draw conclusions from a small number of studies, however the studies with the longest follow-up time showed the largest improvements, which could suggest that prolonged DNase use is more beneficial.

The clinical trial by Fuchs et al. was the only study to find that DNase improved well-being in patients. All other studies that investigated well-being found no evidence of an effect of DNase on well-being. However, each study used different tools to measure well-being making comparison between studies difficult.

Almost all the studies investigating bacterial colonisation found very similar rates in DNase users and non-users. The only study which found a statistically significant difference did not adjust for baseline prevalence of bacterial colonisation in the participants. However, there have been relatively few studies with bacterial colonisation as an outcome and in most studies the prevalence was already quite high at baseline making it difficult to notice any differences.

3.5.2 Limitations

One of the key limitations of this systematic review is the difficulty of synthesising the findings. It was not considered appropriate to perform any meta-analyses due to the small number of studies with overlapping follow-up times and outcomes.

Many studies in fact investigated many outcomes, for example many studies contained several lung function outcomes, but then only reported the results that were statistically significant, whereas all the other measured outcomes would often just be listed as not statistically significant with no further details.

The inclusion and exclusion criteria were very specific in many studies, especially in terms of baseline lung function, which could reduce the generalisability of their findings. However, few studies have investigated whether the effects of DNase differ between groups of patients, such as children or patients with severe pulmonary disease, which means questions remain as to whether DNase is equally effective in all patients.

Many of the RCTs only administered DNase for relatively short periods of time, whereas most of the observational studies had longer follow-up time and also larger sample sizes, but unfortunately many of them did not use appropriate statistical methods to control for confounding in a longitudinal setting. Often results at the end of follow up were compared between DNase users and non-users without accounting for baseline differences

between the two groups, or baseline values would be taken into account, but then no formal comparison would be made between the DNase users and non-users.

3.5.3 Conclusions

All of the results suggest that DNase does not have harmful effects on any of the outcomes, but it remains difficult to objectively assess the degree of benefit that DNase provides. It seems that the benefits may be modest (although still clinically important) and measuring these improvements accurately would require large datasets.

Furthermore, although a therapy which increases lung function is beneficial, if the rate of lung function decline remains unchanged, the initial improvement would soon be lost and the lifespan of the patient would not be considerably lengthened. Conversely, if the treatment is able to slow the rate of decline, this can have a much bigger impact on the prognosis of the disease.[69] For this reason it is important to find out if the observed improvement in lung function upon initiating DNase is sustained long term.

Similarly, so far only one study (Weck et al. [58]) has aimed to quantify the proportion of patients for whom DNase would be effective. Using a much larger dataset, it could be possible to investigate this further and see which patients respond well to DNase and if there are groups of patients who may benefit more from other available treatments.

A final point to mention is that most studies did not consider the rates of adherence to DNase. The administration of DNase can be time consuming and as the length of follow-up time increases patients may begin to feel that the treatment burden is too much. This could cause the adherence rates to drop, which could be a reason why observational studies with longer follow-up show more attenuated benefits.

Implications

- Although studies have shown that DNase treatment improves lung function, it is unclear if these improvements are sustained long term.
- The effect of DNase appears to be quite heterogeneous: some patients show a large benefit whereas others show no or minimal response.

Chapter 4

Statistical Methods to Estimate Effects of Dornase Alfa

4.1 Introduction

We wish to estimate the effect of long-term DNase use on health outcomes using the UK CF Registry data, and in this chapter the methods identified as possible options for enabling estimation of the effects of interest are introduced. Five separate methods are considered: inverse probability weighting (IPW) of marginal structural models (MSM), history-adjusted MSM (HA-MSM), g-computation formula, g-estimation of structural nested models (SNM) and sequential conditional mean models (SCMM).

We start by introducing the notation that will be used throughout the chapter and then explain the complexities faced in obtaining valid estimates of the long-term treatment effects using observational data, such as the UK CF Registry. For our analysis of the UK CF Registry data we focus on two outcomes: lung function (a continuous outcome) and IV days (a count outcome). We introduce each method in turn with reference to a continuous outcome, and afterwards we discuss how the methods can be altered to handle a count outcome. We end the chapter with an overview and comparison of the five methods and also discuss the assumptions that the methods rely on.

Work from this chapter and the following chapter formed the basis of a paper entitled "Estimating long-term treatment effects in observational data: a comparison of the performance of different methods under real-world uncertainty", which was published in Statistic in Medicine in May 2018.[11] The accepted version of this paper can be found in Appendix B.

4.1.1 Notation

We present the methods below in generic notation, but with reference to the way that data are collected in the UK CF Registry. In this set up, every individual will attend a

health assessment annually, and the outcome of interest, Y, is measured at every visit. If a person has T annual assessments, they would therefore have T outcome measures, Y_t $(t = 1, ..., T)$. The annual assessment also records whether the person has been exposed to the treatment of interest since the last annual assessment, X_t, and other health related data, which could have affected the decision to start treatment or not and which could also affect the outcome of interest, $\mathbf{V}_t$. The variables contained in $\mathbf{V}_t$ are taken to be time-varying, whereas we also separately refer to variables $\mathbf{B}$, which are non-time-varying variables measured only once at baseline.

$\overline{X}_t$ is a vector denoting the treatment history for an individual from visit 0 up to and including visit t, and similarly for other variables. $\overline{X}_{s,t}$ would refer to the treatment history between visits t and s $(t < s)$. We use the counterfactual notation $Y_t^{\overline{x}_t=\overline{1}}$ to refer to the outcome that would have been observed at visit t if an individual had received treatment at all visits up to and including visit t. Furthermore, $Y_s^{\overline{X}_{s,t}=\overline{1},\overline{X}_t=\overline{0}}$ would refer to the outcome that would have been observed at visit s if an individual had received treatment between visits s and t, but had not received treatment prior to visit t.

We assume that if an individual attends a visit at time t, all variables are measured and there are no missing data. However, a person can become censored at any time, $C_t = 1$, for example due to death or loss to follow-up, and no measurements would be recorded for this individual at time t. We assume that once an individual is censored that they remain censored, i.e.

$$C_t = 1 \implies C_s = 1 \quad \forall s > t.$$

4.1.2 Time-Dependent Confounding

Figure 4.1 presents a directed acyclic graph (DAG) of the assumed causal pathways between treatment, outcome and other time-varying covariates. The non-time-varying baseline variables, $\mathbf{B}$, have not been included in Figure 4.1 for clarity, but we assume these baseline variables can affect all other variables. The DAG visualises the outcome Y_{t-1} affecting treatment X_t, time-varying covariates $\mathbf{V}_t$ affecting treatment X_t, treatment X_t affecting the outcome Y_t and other time-varying covariates $\mathbf{V}_{t+1}$, and direct longitudinal effects of treatment, outcome and other time-varying covariates on later instances of the same variable.

This DAG highlights the main challenge of estimating long-term treatment effects: time-dependent confounding. If we consider just the effect of X_1 on Y_1, we see that $\mathbf{V}_1$ confounds this effect. Conventional statistical methods could easily be used to adjust for this confounding to give a valid estimate of the effect of X_1 on Y_1. However, if we are interested in the effect of $\overline{X}_2$, i.e. the joint effect of X_1 and X_2, on Y_2, we see that $\mathbf{V}_2$ is a confounder of the effect of X_2 on Y_2, but $\mathbf{V}_2$ also mediates some of the effect of X_1 on Y_2. As we are interested in both the effect of X_1 and X_2 on Y_2, it is not clear how to deal

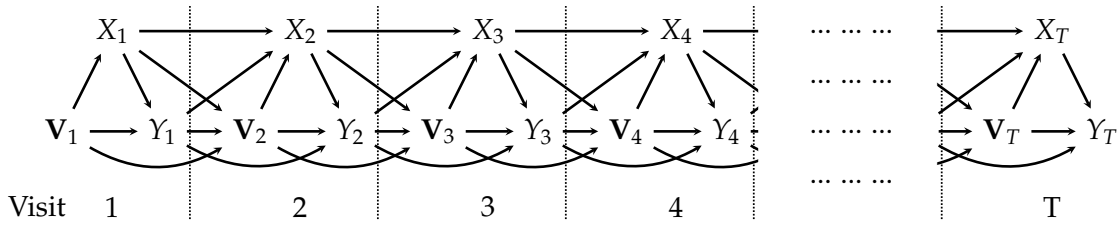


FIGURE 4.1: DAG of assumed causal pathways between treatment (X), outcome (Y) and time-varying confounders ($\mathbf{V}$)

with $\mathbf{V}_2$ using standard approaches. Variables such as $\mathbf{V}_2$ in this case are referred to as time-dependent confounders. A particular variable V_t is a time-dependent confounder if it fulfils the following three criteria[80]:

- V_t is affected by previous treatment X_{t-1}
- V_t affects the probability of starting, stopping or continuing treatment X_t
- V_t affects the outcome of interest Y_t

4.1.3 Treatment Effects

In this work, we wish to estimate the cumulative effect of treatment over time compared to never receiving treatment. We define the one-year effect of treatment as the effect of X_t on Y_t. This is represented by the expected difference in the counterfactual outcomes $Y_t^{\overline{X}_{t,t-1}=1,\overline{X}_{t-1}=\overline{0}}$ and $Y_t^{\overline{X}_t=\overline{0}}$, and we label this expectation ϕ_1.

In general, we define ϕ_{s-t+1} as the effect of X_t on Y_s ($1 \leq t \leq s \leq T$) not mediated through future treatment. The cumulative effect of receiving treatment for $s-t+1$ years compared to not receiving treatment over that period is then defined as:

$$\mathrm{E}\left(Y_s^{\overline{X}_{s,t}=\overline{1},\overline{X}_t=\overline{0}} - Y_s^{\overline{X}_s=\overline{0}}\right) = \sum_{i=1}^{s-t+1} \phi_i. \tag{4.1}$$

4.1.4 Assumptions

The following five sections give the details of the methods that we will consider for the analysis of the UK CF Registry: IPW of MSM, HA-MSM, g-computation formula, g-estimation of SNM and SCMM. All of these methods have been developed to estimate treatment effects in the presence of time-dependent confounding

For all five methods we make the following four assumptions: no interference, positivity, consistency, and no unmeasured confounding. No interference means that for a given individual their counterfactual outcome $Y_t^{\overline{x}}$ is not affected by the treatment that another

individual receives.[81] Positivity means that all individuals have a conditional probability strictly greater than 0 and strictly less than 1 of receiving treatment at all visits given their history, $0 < P(X_t = 1|\overline{X}_{t-1}, \overline{Y}_{t-1}, \overline{\mathbf{V}}_t, \mathbf{B}) < 1$.[82] Consistency means that for each individual, the counterfactual outcome under the observed treatment is equal to the observed outcome, $Y_i = Y_i^{x_i}$ when $x_i = X_i$.[83] Finally, no unmeasured confounding means that conditional on the past observed variables the treatment received at visit t is independent of the counterfactual outcome, $Y_t^{\overline{x}_t} \perp\!\!\!\perp X_t|\overline{X}_{t-1}, \overline{Y}_{t-1}, \overline{\mathbf{V}}_t, \mathbf{B}$.[80]

Thinking specifically about analysing the UK CF Registry, interference should not be an issue, because CF is a non-infectious condition. Furthermore, people with CF are generally kept out of direct contact with one another to avoid cross-infection of respiratory microorganisms.[18] The assumption of positivity is also likely to be valid, because although guidelines do exist to help advise when patients might benefit from DNase, it is not uncommon for patients to receive or not receive treatment despite the guidelines. Once DNase treatment has been initiated, it is usual to continue to receive the treatment indefinitely, but a number of people do also stop taking treatment for various reasons. Consistency concerns the definition of the intervention. The standard dosage and frequency of DNase is 2.5mg once a day, but a small number of patients receive a different dosage or frequency. Unfortunately, dosage data are not routinely collected in the Registry. However, consistency is considered to hold under an intervention defined as "receives DNase as prescribed by doctor". Finally, thanks to the large number of variables available in the UK CF Registry, we do not believe that there are likely to be any important unmeasured confounders.

4.2 Inverse Probability Weighting of Marginal Structural Models

IPW of MSM[84] has become an increasingly popular method to deal with time-dependent confounding. We consider MSM of the following form:

$$\mathrm{E}\left(Y_s^{\overline{x}_s}\right) = \beta_0 + \sum_{i=1}^{s} \phi_{s-i+1} x_i. \tag{4.2}$$

However, due to confounding, the conditional expectation $\mathrm{E}(Y_s|\overline{X}_s = \overline{x}_s)$ is not equal to the expected counterfactual $E(Y_s^{\overline{x}_s})$.

Under the assumptions listed in Section 4.1.4, IPW of the observations enables consistent estimation of the parameters of MSM by reweighting observations so that, at each point in time, the levels of past confounding and exposure variables become equally balanced between treated and untreated individuals at that time. This is achieved by assigning large weights to individuals who were estimated to be unlikely to receive the treatment

they actually received and downweighting observations for which there are lots of observations receiving the same treatment history.

To calculate the weights, one first estimates the propensity score, which is the probability of receiving treatment at each visit. Considering a binary treatment, this could therefore be estimated by fitting the following logistic regression model:

$$\mathrm{P}\Big(X_t = 1 \Big| \overline{X}_{t-1}, \overline{Y}_{t-1}, \overline{\mathbf{V}}_t, \mathbf{B}\Big) = \mathrm{expit}\Big(\beta_0 + \beta_X X_{t-1} + \beta_Y Y_{t-1} + \boldsymbol{\beta}_{\mathbf{V}} \mathbf{V}_t + \boldsymbol{\beta}_{\mathbf{B}} \mathbf{B}\Big). \tag{4.3}$$

This model, with its parameters estimated from the data, is then used to calculate the estimated probability that each person received the treatment they actually received, i.e. for those who did receive treatment we use the estimated probability from the above model and for those who did not receive treatment we use one minus the estimated probability.

$$\mathrm{P}\Big(X_t \Big| \overline{X}_{t-1}, \overline{Y}_{t-1}, \overline{\mathbf{V}}_t, \mathbf{B}\Big) = \begin{cases} \mathrm{P}\Big(X_t = 1 \Big| \overline{X}_{t-1}, \overline{Y}_{t-1}, \overline{\mathbf{V}}_t, \mathbf{B}\Big), & \text{if } X_t = 1 \\ 1 - \mathrm{P}\Big(X_t = 1 \Big| \overline{X}_{t-1}, \overline{Y}_{t-1}, \overline{\mathbf{V}}_t, \mathbf{B}\Big), & \text{if } X_t = 0 \end{cases} \tag{4.4}$$

The most simple method to create weights is to take the inverse of product of these estimated probabilities from visit 1 up to visit of the outcome, s, so that people who, based on their observed history, are estimated to be unlikely to have received the treatment they did receive, are assigned the highest weights.

$$W_s = \prod_{i=1}^{s} \frac{1}{\mathrm{P}\Big(X_i \Big| \overline{X}_{i-1}, \overline{Y}_{i-1}, \overline{\mathbf{V}}_i, \mathbf{B}\Big)} \tag{4.5}$$

It can be shown that such IPW leads to consistent estimation of parameters of MSM under the assumptions of Section 4.1.4 if, additionally, the statistical models used to estimate the weights are of the correct functional form.[84]

One problem that arises in this approach is that when the predictors of treatment are very strong the estimated probabilities can become very small for some individuals, resulting in extremely large weights. This has the negative consequence that the estimates for the final MSM will have large variances.[84] This problem can be mitigated to a certain extent by readjusting the weights using any covariates which do not cause time-dependent confounding, e.g. baseline covariates. These stabilised weights are created by changing the numerator of Equation 4.5 to the estimated probability that an individual received

their treatment history given only variables which are not time-dependent confounders:

$$\mathrm{SW}_s = \frac{\mathrm{P}\left(\overline{X}_s \middle| \overline{X}_{s-1}, \mathbf{B}\right)}{\mathrm{P}\left(\overline{X}_s \middle| \overline{X}_{s-1}, \overline{Y}_{s-1}, \overline{\mathbf{V}}_s, \mathbf{B}\right)} = \prod_{i=1}^{s} \frac{\mathrm{P}\left(X_i \middle| \overline{X}_{i-1}, \mathbf{B}\right)}{\mathrm{P}\left(X_i \middle| \overline{X}_{i-1}, \overline{Y}_{i-1}, \overline{\mathbf{V}}_i, \mathbf{B}\right)}. \tag{4.6}$$

If the numerator of the stabilised weights only includes previous treatment history, then the final MSM can then be fit following Equation 4.2 where the observations are weighted using the estimated weights. However, any baseline confounders included in the numerator of Equation 4.6 must then also be included in the MSM:

$$\mathrm{E}\left(Y_s^{\bar{x}_s} \middle| \mathbf{B}\right) = \beta_0 + \boldsymbol{\beta}_{\mathbf{B}}\mathbf{B} + \sum_{i=1}^{s} \phi_{s-i+1} X_i. \tag{4.7}$$

This would result in a conditional interpretation for the MSM parameters, meaning that if a marginal interpretation is desired then no baseline variables should be included in the numerator, so that they can also be excluded from the MSM. For the particular example of Equation 4.7, the conditional and marginal estimands of the treatment effect are equal, because there are no interactions between $\mathbf{B}$ and X, and because it is a linear MSM, the parameters of which have the property of collapsibility. However, in general this would not be the case, for example when there are interactions or other non-linearities, e.g. with a logistic MSM, which would give rise to non-collapsible parameters.

Due to the fact that time-varying covariates are not included in the MSM, this method does not allow for the estimation of effect modification by time-varying covariates. However, the method also does not make an assumption that there is no effect modification and will estimate consistent effects, marginal with respect to the time-dependent covariates, even if the effect of treatment is different at different levels of these time-dependent covariates.

Using stabilised weights helps to reduce the variability in the weights, but in cases where there are strong time-varying predictors of treatment the weights can remain highly variable which can lead to instability. Therefore, it can sometimes be preferable to truncate the most extreme weights, even though this may introduce some bias.[85] In general, the increase in bias due to the truncation of weights is greater than the decrease in precision, and therefore it has been suggested to only use minimal truncation, such as the 1st and 99th percentile.[86]

In the presence of censoring, it is also possible to incorporate censoring weights into the analysis. Similarly to the previously described weights, we weight individuals with stabilised inverse weights of their estimated probability of being censored, C, before visit s. An individual's observations will only be included in the analysis up until they are censored. We are, therefore, interested in the cumulative probability through time that

individuals remain uncensored:

$$\text{CW}_s = \frac{\text{P}\left(\overline{C}_s = 0 \middle| \mathbf{B}\right)}{\text{P}\left(\overline{C}_s = 0 \middle| \overline{X}_{s-1}, \overline{Y}_{s-1}, \overline{\mathbf{V}}_{s-1}, \mathbf{B}\right)} = \prod_{i=1}^{s} \frac{\text{P}\left(C_i \middle| \mathbf{B}\right)}{\text{P}\left(C_i \middle| X_{i-1}, Y_{i-1}, \mathbf{V}_{i-1}, \mathbf{B}\right)}. \tag{4.8}$$

Using this method we assume that future visits are missing at random, i.e. censoring is affected by previously measured variables. The estimated censoring weights can then be multiplied by the estimated stabilised weights at each time point to give the weights to be used to account for bias due to both confounding and censoring.

4.3 History-Adjusted Marginal Structural Models

As stated in the previous section, one limitation of IPW of MSM is that effect modification of the treatment effect by time-varying covariates cannot be estimated. Therefore, in cases where the estimation of an interaction term is desired, HA-MSM are an extension to IPW of MSM which do allow for this.[87] However, this is a different method to standard IPW of MSM and can be used even if there is no effect modification by time-varying covariates.

In the standard MSM described in the Section 4.2, observations are reweighted based on all covariates measured after baseline, $t = 1$ until the visit of the outcome of interest. In a HA-MSM, the reweighting is only done for variables measured between the time of treatment t and the time of outcome, s ($t \leq s$). Because this does not reweight for any covariates measured prior to the treatment at time t, these must be included in the final HA-MSM. Therefore, the MSM is only marginal with respect to any variables measured between t and s, but is conditional on the history prior to time t.

Formally, the stabilised weights at time s for an individual exposed at time t are given by

$$\text{HA-SW}_{ts} = \prod_{i=t+1}^{s} \frac{\text{P}\left(X_i \middle| \overline{X}_{i-1}, \mathbf{B}\right)}{\text{P}\left(X_i \middle| \overline{X}_{i-1}, \overline{Y}_{i-1}, \overline{\mathbf{V}}_i, \mathbf{B}\right)}. \tag{4.9}$$

Note that in the case where $t = s$, the weights are equal to 1, as there is no time-varying confounding for the weights to correct in the one-year treatment effect estimation.

An example of a HA-MSM is:

$$\begin{aligned} &\text{E}\left(Y_s^{\bar{x}_s} \middle| \overline{x}_{t-1}, \overline{y}_{t-1}, \overline{\mathbf{v}}_t, \mathbf{b}\right) = \\ &\qquad \beta_0 + \boldsymbol{\beta}_\mathbf{b}\mathbf{b} + \boldsymbol{\beta}_\mathbf{v}\mathbf{v}_t + \beta_x x_{t-1} + \beta_y y_{t-1} + \sum_{i=t}^{s} \phi_{s-i+1} x_i + \sum_{i=t}^{s} \phi_{\text{int}_{s-i+1}} x_i y_{t-1}. \end{aligned} \tag{4.10}$$

In equation 4.10, we have included an interaction term between previous measures of the outcome (itself a time-varying confounder) and treatment. This would allow us to estimate how the effect of treatment changes depending on previous levels of the outcome. Note that what is being estimated here is the extent to which the treatment effects $(X_t, ..., X_s)$ are modified by Y_{t-1}, with the model being agnostic as to the presence of any effect modification by later outcomes. This is in contrast to g-estimation (see Section 4.5), where the modification of the treatment effects of $X_t, ..., X_s$ by the most recent outcome measure $(Y_{t-1}, ..., Y_{s-1})$, respectively, is estimated.

As with IPW of MSM, extreme weights can be truncated to reduce the variance, but this would result in the same issue as explained in Section 4.2 that even small amounts of truncation could potentially lead to large increases in bias. It is also possible to estimate censoring weights, in this case estimating an individual's probability of being censored between visits t and s, and multiplying these weights with the stabilised weights.

$$\mathrm{CW}_{ts} = \prod_{i=t+1}^{s} \frac{\mathrm{P}\left(C_i\right)}{\mathrm{P}\left(C_i \middle| X_{i-1}, Y_{i-1}, \mathbf{V}_{i-1}, \mathbf{B}\right)}. \tag{4.11}$$

Again, it can be noted that in the case where $t = s$ the censoring weights are equal to 1, as data on X_t and Y_t are collected at the same time, so there is no possibility of censoring when estimating this effect.

4.4 G-Computation Formula

The g-computation formula (hereafter referred to as g-formula) first described by Robins[88] is another method that can deal with the issue of time-dependent confounding to give consistent estimates of long-term treatment effects under the assumptions given in Section 4.1.4, with a different set of parametric assumptions to those involved in IPW of MSM. In this method short-term models, i.e. models for one-year effects, for the outcome and all time-varying covariates (in our example, Y and $\mathbf{V}$) are specified. Although analytic methods can in some instances then be used in order to estimate the parameters of MSM using these fitted "one-year ahead" models, more commonly the integration involved is not tractable, and thus a more practical alternative is to approximate such integrals using Monte Carlo simulation. This has the advantage of being intuitively quite appealing: counterfactual outcomes under different treatment trajectories are simulated sequentially through time.

For example, the time-varying continuous outcome Y could be modelled as follows:

$$\mathrm{E}\left(Y_t \middle| \overline{X}_t, \overline{Y}_{t-1}, \overline{\mathbf{V}}_t, \mathbf{B}\right) = \beta_0 + \beta_{X_1} X_t + \beta_{X_2} X_{t-1} + \beta_Y Y_{t-1} + \boldsymbol{\beta}_{\mathbf{V}} \mathbf{V}_t + \boldsymbol{\beta}_{\mathbf{B}} \mathbf{B}, \tag{4.12}$$

and counterfactuals for Y_1 could then be simulated setting everyone either receiving or not receiving treatment at visit 1 (and assuming nobody received treatment prior to the start of follow-up):

$$\tilde{Y}_1^{x_1=1} = \hat{\beta}_0 + \hat{\beta}_Y Y_0 + \hat{\boldsymbol{\beta}}_{\mathbf{V}} \mathbf{V}_1 + \hat{\boldsymbol{\beta}}_{\mathbf{B}} \mathbf{B} + \hat{\beta}_{X_1} + \tilde{\varepsilon}, \tag{4.13}$$

$$\tilde{Y}_1^{x_1=0} = \hat{\beta}_0 + \hat{\beta}_Y Y_0 + \hat{\boldsymbol{\beta}}_{\mathbf{V}} \mathbf{V}_1 + \hat{\boldsymbol{\beta}}_{\mathbf{B}} \mathbf{B} + \tilde{\varepsilon}, \tag{4.14}$$

where $\tilde{\varepsilon}$ is a different, independent, random draw for each individual in the dataset from a normal distribution whose standard deviation is the model-estimated root-mean-square error, resulting in simulated counterfactual measures that have the correct distribution, if the assumptions in Section 4.1.4 hold, together with the parametric model 4.12, as well as an assumption of normally-distributed errors for model 4.12.

Similar short-term models would need to be specified for all time-varying covariates $\mathbf{V}$ to allow the counterfactuals for all covariates to be simulated.

The process can then be repeated sequentially for all visits up to T. For example, at visit 2 there would be four counterfactuals simulated for each individual, corresponding to: 1) receiving treatment at both visits, 2) at the first visit only, 3) at the second visit only or 4) never receiving treatment. These counterfactuals could be simulated respectively as follows:

$$\tilde{Y}_2^{x_1=1,x_2=1} = \hat{\beta}_0 + \hat{\beta}_Y \tilde{Y}_1^{x_1=1} + \hat{\boldsymbol{\beta}}_{\mathbf{V}} \tilde{\mathbf{V}}_2^{x_1=1} + \hat{\boldsymbol{\beta}}_{\mathbf{B}} \mathbf{B} + \hat{\beta}_{X_1} + \hat{\beta}_{X_2} + \tilde{\varepsilon}, \tag{4.15}$$

$$\tilde{Y}_2^{x_1=1,x_2=0} = \hat{\beta}_0 + \hat{\beta}_Y \tilde{Y}_1^{x_1=1} + \hat{\boldsymbol{\beta}}_{\mathbf{V}} \tilde{\mathbf{V}}_2^{x_1=1} + \hat{\boldsymbol{\beta}}_{\mathbf{B}} \mathbf{B} + \hat{\beta}_{X_1} + \tilde{\varepsilon}, \tag{4.16}$$

$$\tilde{Y}_2^{x_1=0,x_2=1} = \hat{\beta}_0 + \hat{\beta}_Y \tilde{Y}_1^{x_1=0} + \hat{\boldsymbol{\beta}}_{\mathbf{V}} \tilde{\mathbf{V}}_2^{x_1=0} + \hat{\boldsymbol{\beta}}_{\mathbf{B}} \mathbf{B} + \hat{\beta}_{X_2} + \tilde{\varepsilon}, \tag{4.17}$$

$$\tilde{Y}_2^{x_1=0,x_2=0} = \hat{\beta}_0 + \hat{\beta}_Y \tilde{Y}_1^{x_1=0} + \hat{\boldsymbol{\beta}}_{\mathbf{V}} \tilde{\mathbf{V}}_2^{x_1=0} + \hat{\boldsymbol{\beta}}_{\mathbf{B}} \mathbf{B} + \tilde{\varepsilon}. \tag{4.18}$$

Note that in Equations 4.15 to 4.17, we do not use the observed values of Y_1 and $\mathbf{V}_2$, but rather the counterfactuals for these that were simulated in the previous step.

Depending on the total number of visits, the number of possible treatment trajectories can become infeasible large to simulate all trajectories. There will be 2^T possible trajectories, and for each visit of each trajectory one would have to not only simulate the outcome for that visit, but also all the time-varying covariates. However, depending on the question of interest, it is possible to only simulate a subset of the trajectories. For example, if we are interested in estimating the effect of receiving treatment for T years compared to never receiving treatment, we would only need to simulate trajectories corresponding to those two regimes.

Having simulated the desired trajectories, the simulated counterfactual outcomes can then be compared with a MSM, e.g.:

$$\mathrm{E}\left(Y_s^{\bar{x}_s}\right) = \beta_0 + \sum_{i=1}^{s} \phi_{s-i+1} x_i. \tag{4.19}$$

As with IPW of MSM, because the final model is marginal, effect modification cannot be estimated with this method. Another well-known drawback of the use of this method with non-linear models is the g-null paradox.[88, 89] This is an issue whereby given a large enough sample size the causal null hypothesis will always be rejected even if there is in fact no treatment effect. This is due to the fact that the combination of different parametric models will be inconsistent with the null hypothesis.

4.5 G-Estimation of Structural Nested Models

The fourth method we will consider is g-estimation of SNM.[90] This method has been utilised less than the previously described methods and this may partly be due to the perceived difficulty of applying the method with standard statistical software.[91] However, a recent paper by Vansteelandt and Sjolander revisits g-estimation, showing how it can be implemented with standard software.[92]

Similar to HA-MSM, this method estimates the effect of receiving treatment continuously from visit t on outcomes at visits s where $t \leq s$.

The procedure for g-estimation as outlined by Vansteelandt and Sjolander consists of five steps[92]:

1. Estimate the propensity score of receiving treatment at each visit. This can be estimated in the same way as in IPW of MSM as given in Equation 4.3

2. Obtain an estimate for the one-year effect of treatment X_s on Y_s, represented by ϕ_1 in the following equation, where p_s is the propensity score for treatment at time s:

$$\begin{aligned} &\mathrm{E}\left(Y_s \middle| \overline{X}_s, \overline{Y}_{s-1}, \overline{\mathbf{V}}_s, \mathbf{B}, p_s\right) = \\ &\qquad \beta_s + \beta_{X_{s-1}} X_{s-1} + \beta_{Y_{s-1}} Y_{s-1} + \boldsymbol{\beta}_{\mathbf{V}_s} \mathbf{V}_s + \boldsymbol{\beta}_{\mathbf{B}} \mathbf{B} + \beta_{p_s} p_s + \phi_1 X_s, \end{aligned} \tag{4.20}$$

3. Use the estimates ϕ_j to construct counterfactuals of what the outcome would have been if people had not received treatment at between times j and s. (In the first iteration, we only have an estimate for ϕ_1, and $j = s - 1$):

$$H_{sj} = Y_s - \sum_{u=j+1}^{s} \phi_{s-u+1} X_u. \tag{4.21}$$

4. Fit a model to estimate the effect of treatment X_j on H_{sj} ($j \leq s$):

$$\mathrm{E}\left(H_{sj}\middle|\overline{X}_j, \overline{Y}_{j-1}, \overline{\mathbf{V}}_j, \mathbf{B}\right) = \beta_s + \beta_{X_s} X_{j-1} + \beta_{Y_s} Y_{j-1} + \boldsymbol{\beta}_{\mathbf{V}_s} \mathbf{V}_j + \boldsymbol{\beta}_{\mathbf{B}_s} \mathbf{B} + \beta_{p_s} p_j + \sum_{i=t}^{s} z \phi_{s-i+1} X_j, \tag{4.22}$$

   where $z = 1$ if $i = j$ and 0 otherwise.

5. Iterate steps 3 and 4 until all desired estimates ϕ_{s-j+1} have been obtained where $t \leq j \leq s$.

Iteration of steps 3 and 4 progressively allows for the estimation of longer-term treatment effects. For example, on the first run of the above procedure, we would obtain the one-year effect estimate at step 2, ϕ_1. Then, in step 3, the first time, we can only estimate H_{sj} for $j + 1 \geq s$, which means that step 4 will only provide estimates of ϕ_1 and ϕ_2 the first time around. However, now that we have an estimate for ϕ_2, repeating step 3, allows us to additional estimate H_{sj} for $j + 2 \geq s$, and with these estimates step 4 will now additionally estimate ϕ_3.

The procedure as described in the simple example given above assumes that the effect of covariates on Y_s is the same regardless of j, i.e. it would assume that the effect of Y_1 on Y_5 is the same as the effect of Y_4 on Y_5. We believe it is more realistic that these effects would change depending on the length of time between j and s. Therefore we fit models which do not assume these pooled effect estimates. This can be done by changing Equation 4.22 to the following equation:

$$\mathrm{E}\left(H_{sj}\middle|\overline{X}_j, \overline{Y}_{j-1}, \overline{\mathbf{V}}_j, \mathbf{B}, p_j\right) = \sum_{i=t}^{s} z\left(\beta_i + \beta_{X_i} X_{j-1} + \beta_{Y_i} Y_{j-1} + \boldsymbol{\beta}_{\mathbf{V}_i} \mathbf{V}_j + \boldsymbol{\beta}_{\mathbf{B}_i} \mathbf{B} + \beta_{p_i} p_j + \phi_{s-i+1} X_j\right), \tag{4.23}$$

where $z = 1$ if $i = j$ and 0 otherwise.

Censoring weights can also be incorporated into g-estimation and are calculated in the same way as for HA-MSM as given in Equation 4.11, weighting individuals by their estimated probability of being censored between visits t and s.

4.5.1 Effect Modification in G-Estimation

Modification of the treatment effect by time-varying covariates can also be estimated with g-estimation. To do this, the relevant product terms would be included in Equations 4.20,

4.21, and 4.22 or 4.23. For example, Equation 4.23 could be replaced with:

$$\begin{aligned}\mathrm{E}\left(H_{sj}\middle|\overline{X}_j,\overline{Y}_{j-1},\overline{\mathbf{C}}_j,\mathbf{B},p_j\right) = \\ \sum_{i=t}^{s} z\Big(\beta_i+\beta_{X_i}X_{j-1}+\beta_{Y_i}Y_{j-1}+\boldsymbol{\beta}_{\mathbf{V}_i}\mathbf{V}_j+\boldsymbol{\beta}_{\mathbf{B}_i}\mathbf{B}+\beta_{p_i}p_j+\beta_{p_{\mathrm{int}_i}}p_jY_{j-1} \\ +\psi_{s-i+1}X_j+\psi_{\mathrm{int}_{s-i+1}}X_jY_{j-1}\Big),\end{aligned} \tag{4.24}$$

where $z = 1$ if $i = j$ and 0 otherwise.

This allows the effect of treatment X_j on outcome Y_s to differ depending on Y_{j-1}. We could then calculate the total effect of treatment for given values of Y_{j-1} as:

$$\mathrm{E}\left(Y_s^{\overline{X}_{s,t}=\overline{1},\overline{X}_t=\overline{0}}-Y_s^{\overline{X}_s=\overline{0}}\middle|\overline{Y}_{s-1}\right)=\sum_{j=t}^{s}\psi_{s-j+1}+\psi_{\mathrm{int}_{s-j+1}}Y_{j-1}. \tag{4.25}$$

The intercept and interaction treatment effect terms ψ_i and ψ_{int_i} from g-estimation differ from the intercept and interaction treatment effect terms ϕ_i and ϕ_{int_i} estimated by HA-MSM. This is because the interaction effect in g-estimation is based on Y_{j-1}, where $t \leq j \leq s$, whereas in HA-MSM the interaction effect is always based on Y_{t-1}. It may be more relevant to model the effect of treatment being modified by the most recent level of the time-varying covariate, rather than the level of the covariate at the start of the treatment period of interest. However, when wishing to estimate a cumulative treatment effect, it would be more appropriate to just specify Y_{t-1}, rather than having to specify each value of Y_{j-1}, which is necessary if using Equation 4.25.

In the case of a continuous outcome, the estimates ψ_i and ψ_{int_i} from Equation 4.25 can be used to estimate ϕ_i and ϕ_{int_i}. For this, we wish to specify a value for y_{t-1} and then predict y_j, where $t < j \leq s$, using the estimates from Equation 4.24:

$$y_j=\beta_j+\beta_{Y_j}y_{t-1}+\boldsymbol{\beta}_{\mathbf{V}_j}\mathbf{v}_t+\boldsymbol{\beta}_{\mathbf{B}_j}\mathbf{b}+\beta_{p_j}p_t+\beta_{p_{\mathrm{int}_j}}p_ty_{t-1}+\sum_{i=1}^{j-t}\psi_i+\psi_{\mathrm{int}_i}y_{i-1}. \tag{4.26}$$

The above equation would need to be iterated $s - t$ times, the first time we would obtain a prediction for y_t, then y_{t+1}, up to y_{s-1}. It is also necessary to specify values of $\mathbf{v}_t$, $\mathbf{b}$ and p_t, for example their mean values could be used. The predicted values y_j can then be put into Equation 4.25 to estimate the treatment effect for a given value of y_{t-1}. This directly compares to ϕ_i and ϕ_{int_i} from HA-MSM. To show this, let $F_{t-1} = 0$ and F_j where $t < j \leq s$ be the parts of Equation 4.26 that are not multiplied by y_{t-1}, and let $G_{t-1} = 1$

and G_j where $t < j \leq s$ be the parts of Equation 4.26 that are multiplied by y_{t-1}, i.e.:

$$F_j = \beta_j + \boldsymbol{\beta}_{\mathbf{V}_j}\mathbf{V}_t + \boldsymbol{\beta}_{\mathbf{B}_j}\mathbf{B} + \beta_{p_j}p_t + \sum_{i=1}^{j-t}\psi_i + \psi_{\text{int}_i}F_{i-1} \tag{4.27}$$

$$G_j = \beta_{Y_j} + \beta_{p_{\text{int}_j}}p_t + \sum_{i=1}^{j-t}\psi_{\text{int}_i}G_{i-1}. \tag{4.28}$$

Then we can rearrange Equation 4.25:

$$\begin{aligned}
\mathrm{E}\left(Y_s^{\overline{X}_{s,t}=\overline{1},\overline{X}_t=\overline{0}} - Y_s^{\overline{X}_s=\overline{0}}\middle|\overline{Y}_{s-1}\right) &= \sum_{j=t}^{s}\psi_{s-j+1} + \psi_{\text{int}_{s-j+1}}Y_{j-1} && (4.29)\\
&= \sum_{j=t}^{s}\psi_{s-j+1} + \psi_{\text{int}_{s-j+1}}F_{j-1} + \psi_{\text{int}_{s-j+1}}G_{j-1}Y_{t-1} && (4.30)\\
&= \sum_{j=t}^{s}\phi_{s-j+1} + \phi_{\text{int}_{s-j+1}}Y_{t-1}. && (4.31)
\end{aligned}$$

The cumulative intercept term is therefore:

$$\sum_{j=t}^{s}\phi_{s-j+1} = \sum_{j=t}^{s}\psi_{s-j+1} + \psi_{\text{int}_{s-j+1}}F_{j-1}, \tag{4.32}$$

and the cumulative interaction term with Y_{t-1} is:

$$\sum_{j=t}^{s}\phi_{\text{int}_{s-j+1}} = \sum_{j=t}^{s}\psi_{\text{int}_{s-j+1}}G_{j-1} \tag{4.33}$$

The above decompositions are only possible because Y is continuous with the identity link function, allowing the y to be broken down into intercept and interaction terms f and g. In general with other link functions, the cumulative effect of treatment for a given value y_{t-1} could still be calculated, but it would not be possible to simplify it into one intercept and one interaction term.

4.6 Sequential Conditional Mean Models

Even in the presence of time-dependent confounding it is still possible to use standard regression methods, but these methods can only estimate total effects.[93] The total effect of a treatment, X_t, on an outcome Y_s $(s \geq t)$ would include not only the direct effect of X_t on Y_s and the indirect effects of X_t on Y_s mediated through time-varying covariates, but also the indirect effect of X_t on Y_s mediated through future exposures. An example of this two-year total effect is shown in the left panel of 4.2, and can be compared to the

two-year effect as estimated by the other methods introduced in this chapter, which is shown in the right panel of the same figure.

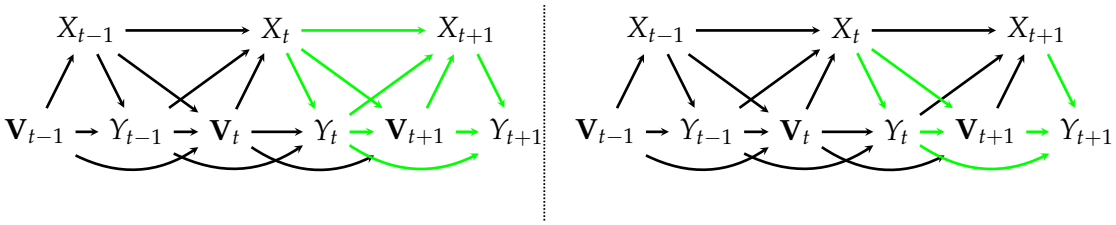


FIGURE 4.2: DAGs highlighting in green the different two-year treatment effects estimated by SCMM (left panel) compared to other methods (right panel)

Both of these long-term estimates are valid estimates of a treatment effect, but they answer different questions. The total effect as shown in the left panel of Figure 4.2 would answer the question: "What is the effect of starting treatment now on the outcome of interest in two-years' time compared to not starting treatment?". In this question, people who start treatment at time t may in fact discontinue treatment at time $t+1$, and similarly people untreated at time t could start at time $t+1$, with the effect of being treated at time $t+1$ likely to be affect by whether or not treatment is taken at time t. We would thus be estimating the total effect of taking treatment at time t, including its likely knock-on effect via the treatment taken at time $t+1$. This can be compared to the effect shown in the right panel of Figure 4.2, which would answer the question: "What is the effect of receiving treatment combination x_t and x_{t-1} versus some other treatment combination x'_t and x'_{t-1}?", e.g. being treated for both of these two years versus being treated for none of these two years. The fact that X_t likely affects X_{t+1} is not relevant in this estimand: instead, we imagine manipulating both of them. For this work, we are more interested in answering questions of the latter type, and therefore, this would suggest a preference towards the other methods introduced in this chapter over SCMM. Nevertheless, in our example, people are assumed to only start or stop treatment at each visit t, and therefore, in the case of the effect of X_t on Y_t (the one-year effect), the SCMM estimate coincides with the estimates of the other methods.

These SCMM will give a consistent estimate of the total effect of treatment as long as we appropriately control for all confounding effects of the effect of interest and correctly specify the parametric model involved. For example, the following model would suffice for the one-year effect if the most recent measures of all covariates were sufficient to remove confounding:

$$\mathrm{E}\left(Y_t \middle| \overline{X}_t, \overline{Y}_{t-1}, \overline{\mathbf{V}}_t, \mathbf{B}\right) = \beta_0 + \beta_{X_1} X_t + \beta_{X_2} X_{t-1} + \beta_Y Y_{t-1} + \boldsymbol{\beta}_{\mathbf{V}} \mathbf{V}_t + \boldsymbol{\beta}_{\mathbf{B}} \mathbf{B}. \tag{4.34}$$

It is also possible to incorporate propensity scores into the SCMM to provide a doubly robust estimator. For a binary treatment, the propensity score can be calculated as it was

in the IPW method by using Equation 4.3, and this is then incorporated into the SCMM as follows:

$$\mathrm{E}\left(Y_t \middle| \overline{X}_t, \overline{Y}_{t-1}, \overline{\mathbf{V}}_t, \mathbf{B}, p_t\right) = \beta_0 + \beta_{X_1} X_t + \beta_{X_2} X_{t-1} + \beta_Y Y_{t-1} + \boldsymbol{\beta}_{\mathbf{V}} \mathbf{V}_t + \boldsymbol{\beta}_{\mathbf{B}} \mathbf{B} + \beta_p p_t. \tag{4.35}$$

Equations 4.35 and 4.20 are the same and this means that for the one-year effect, g-estimation and SCMM are in fact identical. This also means that similarly to g-estimation, SCMM can also estimate modification of the treatment effect by including the relevant interaction term in the model. Note that for any interaction term involving the treatment variable, X_t, the corresponding interaction term with the propensity score p_t should also be included.

The paper by Keogh et al. also gives details of an extension to this method that tests whether there is evidence of a long-term effect of treatment, i.e. after having estimated the effect of X_s on Y_s, we can test whether there is any evidence of any additional effect of X_t on Y_s $(t < s)$ not mediated through X_s.[93] The test is performed in three stages:

1. Fit the model as given in Equation 4.35 to estimate the effect of X_t on Y_t.
2. Use this model to estimate the counterfactual $\hat{Y}_t^{x_t=0}$, i.e. what the observed outcomes would have been if everybody had not received treatment at time t (by setting X_t to 0):

$$\hat{Y}_t^{x_t=0} = \beta_0 + \beta_{X_2} X_{t-1} + \beta_Y Y_{t-1} + \boldsymbol{\beta}_{\mathbf{V}} \mathbf{V}_t + \boldsymbol{\beta}_{\mathbf{B}} \mathbf{B} + \beta_p p_t. \tag{4.36}$$

3. Estimate whether conditional on other covariates there is any residual association between $\hat{Y}_t^{x_t=0}$ and previous treatment X_{t-1}:

$$\begin{aligned} &\mathrm{E}\left(X_{t-1} \middle| X_{t-2}, Y_{t-1}, \mathbf{V}_{t-1}, \mathbf{B}, \hat{Y}_t^{x_t=0}\right) = \\ &\qquad \operatorname{expit}\left(\gamma_0 + \gamma_X X_{t-2} + \gamma_Y Y_{t-2} + \boldsymbol{\gamma}_{\mathbf{V}} \mathbf{V}_{t-1} + \boldsymbol{\gamma}_{\mathbf{B}} \mathbf{B} + \gamma_{\hat{Y}} \hat{Y}_t^{x_t=0}\right). \end{aligned} \tag{4.37}$$

A bootstrap method incorporating these three steps should be used to obtain valid standard errors (SEs) and CIs for parameter $\gamma_{\hat{Y}}$ in order to assess where there is any evidence of any effect of X_t on Y_s $(t < s)$.

Although SCMM cannot provide estimates for the effects of varying lengths of treatment duration, the simplicity of the method is appealing, and the one-year treatment effect estimates do coincide with the one-year treatment effect estimates from the other methods. Furthermore, if the test of no long-term treatment effects was not statistically significant, this could suggest that it is sufficient to only estimate the one-year treatment effect with a SCMM without resorting to the use of a more complex method.

4.7 Extension to Count Outcomes

For our motivating example we have two outcomes of interest, lung function and annual IV days. The first of these is a continuous outcome and all five of the methods can easily handle this outcome as outlined in the sections above. More care is needed when considering a count outcome.

Upon investigation, IV days can be considered approximately distributed as a zero-inflated negative binomial distribution. Modelling outcomes of this type requires two separate estimation procedures: 1) logistic regression to estimate the odds of zero count, and 2) negative binomial regression to estimate the rate of the count. Therefore, there are two separate parts to the treatment effect: the effect of treatment on having a zero count and the effect of treatment on the overall count.

4.7.1 IPW of MSM and g-formula

IPW of MSM and g-formula can both handle different types of outcome by just changing the final MSM, e.g.:

$$\mathrm{E}\Big(Y_s^{\bar{x}_s}\Big) = \mathrm{expit}\Big(\beta_0 + \sum_{i=1}^{s} \beta_{s-i+1} x_i\Big) \exp\Big(\gamma_0 + \sum_{i=1}^{s} \gamma_{s-i+1} x_i + u_i\Big), \tag{4.38}$$

where the expit term is a model for the odds of a zero count and the exponential term is a model for the rate of the count, which includes u_i to account for overdispersion:

$$e^{u_i} \sim \mathrm{Gamma}\big(1/\alpha, \alpha\big). \tag{4.39}$$

We then summarise the cumulative treatment effect as two separate processes. Firstly, the odds of having a zero count after receiving treatment for $s - t + 1$ years compared to not receiving treatment for $s - t + 1$ years:

$$\frac{\mathrm{P}\Big(Y_s^{\overline{X}_{s,t}=\overline{1},\overline{X}_t=\overline{0}} = 0\Big)}{\mathrm{P}\Big(Y_s^{\overline{X}_s=\overline{0}} = 0\Big)} = \mathrm{expit}\Big(\sum_{i=1}^{s-t+1} \beta_i\Big), \tag{4.40}$$

Secondly, the rate of the count after receiving treatment for $s - t + 1$ years compared to not receiving treatment for $s - t + 1$ years:

$$\frac{\mathrm{E}\Big(Y_s^{\overline{X}_{s,t}=\overline{1},\overline{X}_t=\overline{0}}\Big)}{\mathrm{E}\Big(Y_{s+1}^{\overline{X}_s=\overline{0}}\Big)} = \exp\Big(\sum_{i=1}^{s-t+1} \gamma_i\Big). \tag{4.41}$$

4.7.2 HA-MSM and SCMM

Similarly, HA-MSM can also easily handle this outcome by changing the final conditional model, and as with continuous outcomes, both can also include interaction terms between treatment and time-varying covariates. An example of a HA-MSM with an interaction term is:

$$\begin{aligned}\mathrm{E}\left(Y_s^{\bar{x}_s}\middle|\overline{x}_{t-1},\overline{y}_{t-1},\overline{\mathbf{v}}_t,\mathbf{b}\right) = \\ \operatorname{expit}\left(\beta_0+\boldsymbol{\beta}_{\mathbf{b}}\mathbf{b}+\boldsymbol{\beta}_{\mathbf{v}}\mathbf{v}_t+\beta_x x_{t-1}+\beta_y y_{t-1}+\sum_{i=t}^{s}\beta_{s-i+1}x_i+\beta_{\mathrm{int}_{s-i+1}}x_i y_{t-1}\right) \\ \exp\left(\gamma_0+\boldsymbol{\gamma}_{\mathbf{b}}\mathbf{b}+\boldsymbol{\gamma}_{\mathbf{v}}\mathbf{v}_t+\gamma_x x_{t-1}+\gamma_y y_{t-1}+\sum_{i=t}^{s}\gamma_{s-i+1}x_i+\gamma_{\mathrm{int}_{s-i+1}}x_i y_{t-1}+u_i\right),\end{aligned} \tag{4.42}$$

where

$$e^{u_i}\sim\mathrm{Gamma}\left(1/\alpha,\alpha\right). \tag{4.43}$$

The two cumulative treatment effects would then be given by:

$$\frac{\mathrm{P}\left(Y_s^{\overline{X}_{s,t}=\overline{1},\overline{X}_t=\overline{0}}=0\middle|\overline{x}_{t-1},\overline{y}_{t-1},\overline{\mathbf{v}}_t,\mathbf{b}\right)}{\mathrm{P}\left(Y_s^{\overline{X}_s=\overline{0}}=0\middle|\overline{x}_{t-1},\overline{y}_{t-1},\overline{\mathbf{v}}_t,\mathbf{b}\right)}=\operatorname{expit}\left(\sum_{i=1}^{s-t+1}\beta_i+\beta_{\mathrm{int}_i}Y_{t-1}\right) \tag{4.44}$$

$$\frac{\mathrm{E}\left(Y_s^{\overline{X}_{s,t}=\overline{1},\overline{X}_t=\overline{0}}\middle|\overline{x}_{t-1},\overline{y}_{t-1},\overline{\mathbf{v}}_t,\mathbf{b}\right)}{\mathrm{E}\left(Y_{s+1}^{\overline{X}_s=\overline{0}}\middle|\overline{x}_{t-1},\overline{y}_{t-1},\overline{\mathbf{v}}_t,\mathbf{b}\right)}=\exp\left(\sum_{i=1}^{s-t+1}\gamma_i+\gamma_{\mathrm{int}_i}Y_{t-1}\right). \tag{4.45}$$

For a SCMM, we could fit an unweighted model similar to Equation 4.42, including the propensity score and only the terms corresponding to the one-year effect, $\beta_1 x_1$ and $\gamma_1 x_1$.

4.7.3 G-Estimation

Unlike the other four methods, which can easily handle different types of outcome, the method of g-estimation described in Section 4.5 has until recently only been described for continuous outcomes. However, a recent paper has shown how this method can be adapted to allow for a count outcome using a gamma distribution.[94] This allows for the estimation of the effect of treatment on the rate of the count, but would not allow for the zero inflation as the other methods do.

The procedure for estimating an incidence rate ratio (IRR) in g-estimation follows the same 5-step procedure as for continuous outcomes given in Section 4.5. In step 2, we

would fit a gamma generalized linear model with a log link function and a scale parameter of one:

$$\begin{aligned}&\mathrm{E}\Big(Y_s\Big|\overline{X}_s,\overline{Y}_{s-1},\overline{\mathbf{V}}_s,\mathbf{B},p_s\Big)=\\&\quad\exp\Big(\beta_s+\beta_{X_{s-1}}X_{s-1}+\beta_{Y_{s-1}}Y_{s-1}+\boldsymbol{\beta}_{\mathbf{V}_s}\mathbf{V}_s+\boldsymbol{\beta}_{\mathbf{B}}\mathbf{B}+\beta_{p_s}p_s+\phi_1X_s\Big).\end{aligned}\tag{4.46}$$

In step 3 the counterfactuals are now estimated as:

$$H_{sj}=Y_s\prod_{u=j+1}^{s}e^{-\phi_{s-u+1}X_u}.\tag{4.47}$$

And, in step 4, we would again fit a gamma generalized linear model with a log link function and a scale parameter of one:

$$\begin{aligned}&\mathrm{E}\Big(H_{sj}\Big|\overline{X}_j,\overline{Y}_{j-1},\overline{\mathbf{V}}_j,\mathbf{B},p_j\Big)=\\&\quad\exp\Big(z\big(\beta_i+\beta_{X_i}X_{j-1}+\beta_{Y_i}Y_{j-1}+\boldsymbol{\beta}_{\mathbf{V}_i}\mathbf{V}_j+\boldsymbol{\beta}_{\mathbf{B}_i}\mathbf{B}+\beta_{p_i}p_j+\sum_{i=t}^{s}\phi_{s-i+1}X_j\big)\Big),\end{aligned}\tag{4.48}$$

where $z=1$ if $i=j$ and 0 otherwise.

The cumulative effect of receiving treatment for $s-t+1$ years compared to not receiving treatment over that time is then given by the IRR:

$$\frac{\mathrm{E}\Big(Y_s^{\overline{X}_{s,t}=\overline{1},\overline{X}_t=\overline{0}}\Big|\overline{x}_{t-1},\overline{y}_{t-1},\overline{\mathbf{v}}_t,\mathbf{b},p_t\Big)}{\mathrm{E}\Big(Y_{s+1}^{\overline{X}_s=\overline{0}}\Big|\overline{x}_{t-1},\overline{y}_{t-1},\overline{\mathbf{v}}_t,\mathbf{b},p_t\Big)}=\exp\Big(\sum_{i=1}^{s-t+1}\phi_i\Big).\tag{4.49}$$

Effect modification by time-varying covariates can again be incorporated into this g-estimation procedure, by including the desired product term in Equations 4.46, 4.47 and 4.48. The cumulative effect of receiving treatment for $s-t+1$ years compared to not receiving treatment over that time given Y_{j-1} would then be:

$$\frac{\mathrm{E}\Big(Y_s^{\overline{X}_{s,t}=\overline{1},\overline{X}_t=\overline{0}}\Big|\overline{x}_{t-1},\overline{y}_{t-1},\overline{\mathbf{v}}_t,\mathbf{b},p_t\Big)}{\mathrm{E}\Big(Y_{s+1}^{\overline{X}_s=\overline{0}}\Big|\overline{x}_{t-1},\overline{y}_{t-1},\overline{\mathbf{v}}_t,\mathbf{b},p_t\Big)}=\exp\Big(\sum_{i=1}^{s-t+1}\phi_i+\phi_{\mathrm{int}_i}Y_{j-1}\Big).\tag{4.50}$$

As with the continuous case, we may wish to only specify y_{t-1}, rather than every y_{j-1} ($t\leq j\leq s$). We can use the model given in Equation 4.48 with the interaction term included to sequentially predict y_j starting with a specific y_{t-1}.

$$y_j=\exp\Big(\beta_j+\beta_{Y_j}y_{t-1}+\boldsymbol{\beta}_{\mathbf{V}_j}\mathbf{v}_t+\boldsymbol{\beta}_{\mathbf{B}_j}\mathbf{b}+\beta_{p_j}p_t+\beta_{p_{\mathrm{int}_j}}p_ty_{t-1}+\sum_{i=1}^{j-t}\psi_i+\psi_{\mathrm{int}_i}y_{i-1}\Big).\tag{4.51}$$

We can then calculate the cumulative treatment effect for a given y_{t-1} by inputting all the predicted y_j into Equation 4.50. Due to the exponential term in Equation 4.51, it is not possible to decompose the cumulative effect into an intercept and interaction term as it was with a continuous outcome. This would make it difficult to succinctly summarise the treatment effect, which is a problem not suffered by HA-MSM. However, the g-estimation procedure may be more relevant as it estimates how the treatment effect is modified by the most recent measure of the covariate rather than only providing an estimate of how treatment is modified by the level of the covariate at exposure time t.

4.8 Overview & Comparison of Methods

We have introduced five methods that we will consider for the analysis of the UK CF Registry when investigating the effects of long-term DNase use on two health outcomes. We are interested in estimating the effect of using DNase continuously for a number of years compared to never receiving DNase. Four of the methods we have introduced are suitable for this, but one method, SCMM, will only be able to be used to estimate the effect of one-year of DNase use. SCMM can estimate some long-term effects: these are total effects including the indirect effects mediated through future treatment use. However, SCMM is the simplest of the methods to implement and along with its test of whether there are any long-term effects, it could be used as a first step before deciding if the more complex methods may be necessary.

In addition to being the simplest method, one further benefit of SCMM is that the model-based SE will be approximately correct when the propensity score is well estimated, meaning that the bootstrap does not need to be used and results can be obtained much faster than using the other methods presented in this paper. The asymptotic SEs have been derived for IPW of MSM, but only in a time-fixed setting,[95] and the difficulty of deriving these in a longitudinal setting necessitate the use of the bootstrap for all the methods other than SCMM. This is because all the methods contain a number of steps of estimation and just using the final model-based SE would fail to account for the uncertainty from the earlier steps.[86, 87, 96, 97] The bootstrap provides valid results as all of these methods produce regular estimators.[98]

All of the methods except for g-formula make use of the propensity score. However, in the case of IPW of MSM and HA-MSM, the propensity scores are used to calculate the estimated probability of the observed treatment for each individual, and because of this, the methods become much more complicated if we wish to use a non-binary treatment. G-estimation and SCMM do not have this issue and can easily handle treatments that are not binary, such as a continuous measure of dose of treatment received. Nevertheless, in our example, we only have reliable data for a binary marker indicating receiving or not receiving treatment, so this will not be discussed further here.

For continuous outcomes, conditional and marginal estimates are the same due to collapsibility. However, we also wish to use the methods with a count outcome, where the issue of non-collapsibility will be present. Both IPW of MSM and g-formula provide marginal estimates and would therefore be expected to give the same estimates on average if there are no issues with the analysis. Although, note that IPW of MSM will only provide a marginal estimate if no baseline covariates are included in the numerator of the weights. HA-MSM, g-estimation and SCMM provide conditional estimates, and due to non-collapsibility these estimates will not coincide with the marginal estimates from IPW of MSM and g-formula, even if all models are correct. Furthermore, SCMM condition on the propensity score, whereas HA-MSM do not, and therefore, these would be two different conditional estimates and would not be expected to be the same.

For all methods other than g-estimation we will model the count outcome using a zero-inflated negative binomial distribution, but for g-estimation, we instead use a gamma distribution. The results from this method cannot, therefore, be compared directly with the results from the other methods.

Linked to the issue of conditionality is the estimation of treatment effect modification by time-varying covariates. As IPW of MSM and g-formula give marginal estimates, it is not possible for them to estimate effect modification by time-varying covariates, but they do provide valid population average estimates even in the presence of effect modification, as neither method assumes that effect modification is not present. Conversely, the other three methods can estimate treatment effect modification by time-varying covariates, and in fact would not give valid estimates if there is effect modification but the interaction terms are not included in the models.

When considering modification of long-term treatment effects, HA-MSM and g-estimation differ in their approach. We use the example where the treatment effect is modified by previous levels of the outcome. HA-MSM estimates how the total effect of receiving treatment from X_t to X_s on Y_s is modified by Y_{t-1}, whereas g-estimation estimates how the effect of X_t on Y_s is modified by Y_{t-1}, and how the effect of X_{t+1} is modified by Y_t, etc, up to the effect of X_s on Y_s being modified by $Ys-1$. It may be more relevant to estimate how the effect of treatment at each time-point is modified by the most recent level of the outcome, rather than the level of the outcome at the start of the treatment period of interest, suggesting that g-estimation may provide more useful estimates of any effect modification.

The five methods introduced in this chapter are not the only methods that can be used. Two other methods that are sometimes used to analyse Registry data are methods involving patient matching or methods based on instrumental variables. The most common method involving matching patients is to match based on the propensity score.[99] This results in a method which is very similar to IPW by the propensity score, but when using matching methods, matched patients are given a weight of one and everyone else would

be given a weight of zero (i.e. excluded from the analysis). However, the use of matching methods can be challenging in longitudinal settings, as it is unlikely that there would be many patients who could be matched consistently over many years of follow-up.[100] For this reason, we chose to focus on IPW by the propensity score.

An instrumental variable is a variable that satisfies three conditions: 1) the variable affects the probability of receiving treatment, 2) the variable's only effect on the outcome is mediated through the treatment, and 3) there are no confounders between the variable and the outcome.[101] In our setting, it is difficult to find a variable which is predictive of treatment, but which does not itself affect either of our outcomes of interest. This is because the predictors of treatment are generally based on their health status. One potential instrumental variable could be CF centre, because there are noticeable differences between the proportion of people receiving DNase between CF centres. However, lung function and the use of IVs also varies between centres, plausibly for reasons other than the different treatment use, thus even this would probably not satisfy criteria two. It is for this reason that we did not consider using an instrumental variable approach in our analysis of the long-term effects of DNase.

Chapter 5

Simulation Studies to Assess the Performance of Statistical Methods

5.1 Introduction

The previous chapter introduced five methods that could potentially be used to analyse the UK CF Registry data in order to estimate the effects of long-term DNase use. However, there are no clear guidelines on when one method might be preferred over another. In an ideal setting, where we could be sure that all assumptions are met and that all models are correctly specified, any one of the available methods could be used to obtain consistent treatment effect estimates, albeit with some methods more efficient than others. However, with real data, it can be difficult to know if assumptions hold and all models will be misspecified to some degree.

In this chapter, we perform simulation studies under a number of different scenarios in order to help assess which of the methods might be the most appropriate for the analysis of the UK CF Registry data. The primary aims of these simulation studies are twofold:

1. Compare the ability and appropriateness of the different analysis methods for addressing the questions of interest, including to estimate effect modification by time-varying covariates and to handle different outcome types
2. Investigate the robustness of the different methods to handling practical challenges arising in longitudinal observational data, such as uncertainty about the relative temporality of measures and loss to follow-up.

We will consider all five methods introduced in the previous chapter: IPW of MSM, HA-MSM, g-formula, g-estimation of SNM, and SCMM. For IPW of MSM and HA-MSM we will consider the methods both with and without the use of truncated weights.

The results from these simulation studies will be used to help guide which methods could be used to analyse the UK CF Registry and specifically whether any of the methods should not be used.

As with the previous chapter, work from this chapter was also incorporated into the paper published in Statistics in Medicine which can be found in Appendix B.

5.2 Features & Challenges in the Analysis of the UK CF Registry

As is usual with most observational data, there are a number of issues which need to be considered when approaching the analysis of the UK CF Registry data.

One challenge is the use of the available methods with different types of outcomes. One of our outcomes of interest, lung function measured with ppFEV$_1$, is continuous and can be reasonably approximated by a conditionally normal distribution. All of the methods described in the previous chapter can easily accommodate such an outcome. However the other outcome, IV days, is a count outcome ranging from 0 to 365, with approximately half of the population having 0 days in a given year. We will model this outcome with a zero-inflated negative binomial distribution. As was discussed in the previous chapter, four of the methods can be used with zero-inflated negative binomial data, whereas, although g-estimation can handle count data, it models this with a gamma distribution, so it might not be the most suitable method if the real data are best approximated by a zero-inflated negative binomial distribution. Nevertheless, we will consider all five methods with the count outcome.

In addition to considering two different types of outcome, we have identified five key features of the analysis of the UK CF registry that post different challenges. The first three of these concern the ability and appropriateness of the different analysis methods to estimate the treatment effect:

1. whether there exists any treatment effect at all,
2. whether there are only short-term or also long-term effects, and
3. whether there is effect modification of the treatment effect by time-varying covariates.

The second category of challenges are those that may arise because of the nature of the data available to investigate the above questions:

4. uncertainty of the direction of causal pathways between variables, and
5. the presence of censoring

The following subsections give further details on each of these five issues.

5.2.1 Causal Null Hypothesis

Figure 5.1 shows the assumed causal pathways between DNase (D) and the two outcomes lung function (L) and IV days (V) in the UK CF Registry. This DAG allows for a causal effect of treatment on both outcomes of interest. To date, RCTs have demonstrated the efficacy of DNase treatment in improving lung function, but no studies have yet shown a statistically significant effect of the treatment on reducing the rate of IV days. Furthermore, in a non-trial setting, where, for example, adherence levels may not be as high as in clinical trials, the findings of a causal effect of treatment may not be replicated. For this reason, methods that benefit from a degree of robustness to model misspecification at the causal null, i.e. when there is in truth no treatment effect, would be attractive.

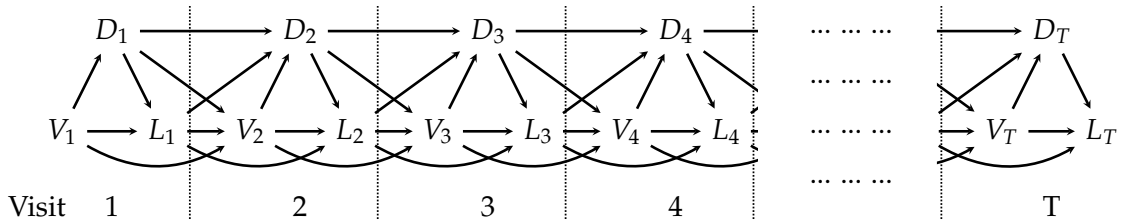


FIGURE 5.1: DAG of assumed causal pathways between DNase (D) and the two outcomes of interest: lung function (L) and IV days (V) (Other time-varying and baseline confounders are not shown for clarity)

5.2.2 Long-Term Treatment Effects

We define a long-term treatment effect as an effect of D_t on either L_s or V_{s+1} ($s > t$) not mediated via intermediate treatments. No studies have previously looked at the effects of DNase beyond two years, and it therefore remains unknown how the effect of treatment might change with length of use. Taking the example of lung function, two possible ways in which the treatment may affect the outcome are:

1. The lung function trajectories of those receiving and not receiving treatment continue to grow apart indefinitely through time, or
2. after the initial increase in lung function that has been observed at the start of taking treatment, the effectiveness of treatment may decrease with the two counterfactual trajectories no longer diverging.

These two hypothetical lung function trajectories compared to the trajectory when not receiving treatment are shown in Figure 5.2. As it is unknown how the effect of treatment might change through time, it is important that the methods are flexible enough to identify the true long-term effects.

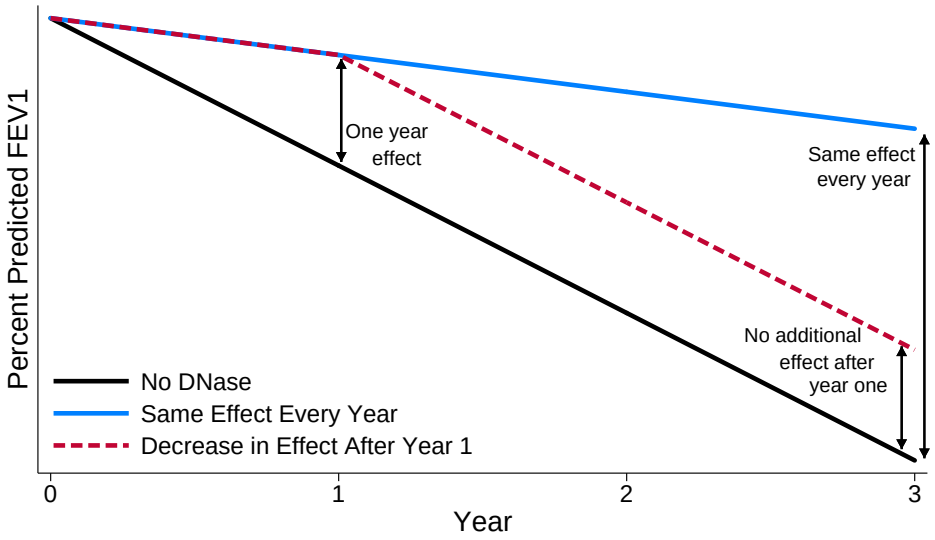


FIGURE 5.2: Two possible trajectories of lung function with long-term DNase use

5.2.3 Effect Modification by Time-Varying Covariates

We hypothesise that the effect of treatment may depend on previous lung function and number of IV days. This is because if a person starts treatment when they already have a high lung function, it is unlikely that treatment could further improve their lung function, whereas it is realistic that the treatment could be much more effective in an individual with very low lung function. For informing practice, rather than just identifying the population average effect of treatment, it is important to gain understanding of how the effect of treatment might change depending on other covariates, and for this reason, it would be preferable to use a method which can test for the presence and estimate the strength of any effect modification.

5.2.4 Misspecification of the Direction of Causal Pathways

The DAG in Figure 5.1 includes assumptions about the direction of the causal pathways. For some variables the appropriate direction of the causal pathway is clear due to the timing of when they are measured, i.e. causal pathways cannot go backwards in time. For example, the pathway from total IV days in one year to lung function measured at the end of the year could not possibly be in the other direction. However, a number of variables are summaries of the past year, and the causal pathways between these variables could conceivably go in either direction.

The direction of the causal pathway between treatment and number of IV days is particularly uncertain. Both variables are summaries of the previous year, and some individuals

may have had lots of IV days at the start of the year which prompted them to start treatment, whereas others may have started treatment earlier, but then had IV days later in the year. Figure 5.1 shows that we assumed that V_t affects D_t with D_t in turn affecting V_{t+1}. However, the way the data are collected could also be compatible with the DAG shown in Figure 5.3, where V_{t-1} affects D_t and D_t affects V_t.

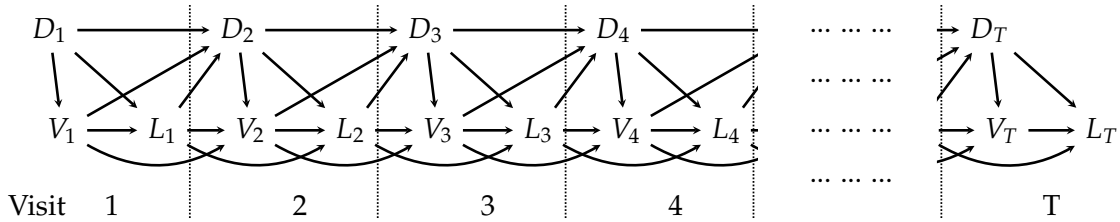


FIGURE 5.3: DAG showing possible reversed causal pathways between DNase (D) and IV days (V)

In reality, the causal pathway between D_t and V_t is likely to go both ways, or one way in some people and another way in other people, but in the methods investigated in this chapter it will be necessary to specify just one direction for this pathway. When considering IV days as an outcome, we have decided to focus on investigating the effect of D_t on V_{t+1}, as, due to temporality this pathway can only be directed this way, and to treat V_t as a potential confounder of this effect. In the real UK CF Registry data we cannot know whether the time ordering of certain variables in the DAG is misspecified or not, and therefore it is important to understand the potential extent of the bias in treatment effect estimates under different methods when the direction of this pathway has been misspecified.

5.2.5 Censoring

We are fortunate that there are very few people lost to follow-up in the UK CF Registry and each year there are relatively few deaths compared to the total number of people in the registry. Nevertheless, it is possible that the fact that some individuals are censored due to loss-to-follow-up or death may bias the results. Therefore, we also wish to investigate how the different methods handle censoring. Although in reality there would likely be different processes affecting the probability that an individual dies or is lost to follow-up, in this chapter we only consider one missing at random scenario where an individual's probability of being censored at a given time depends on previously measured variables.

5.3 Simulation Study Scenarios

In order to assess the five challenges and features, we simulated datasets under six different scenarios as shown in Figure 5.4. In each DAG, the arrows highlighted in red show the specific differences compared to the other scenarios.

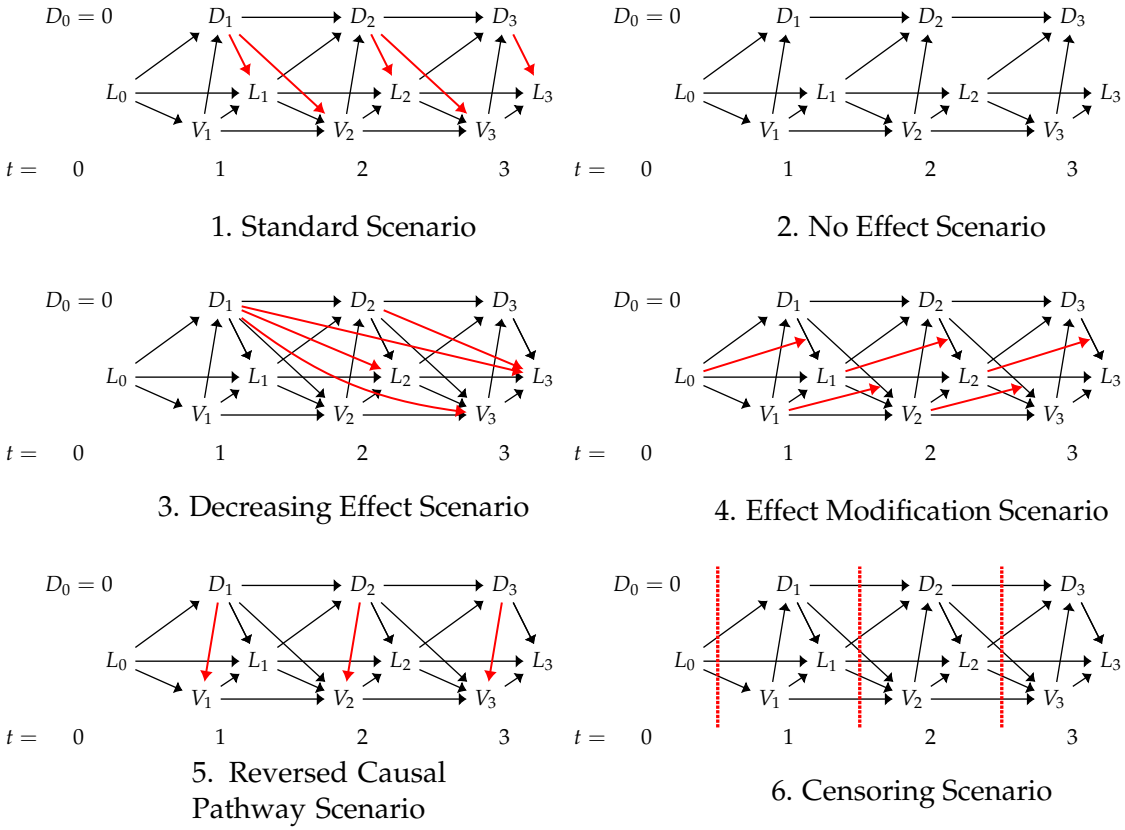


FIGURE 5.4: Simplified DAGs showing data generation process for the scenarios investigated (The actual simulation data were generated up to 5 visits, with an additional baseline confounder, age, affecting all variables)

The first scenario is the standard scenario, which will be the baseline to assess the performance of all methods and to compare this to their performance in the other scenarios. In this scenario there is a one-year treatment effect, but no direct long-term effects. There are, however, long-term effects mediated through other time-varying covariates. This is the scenario for which all the methods will be correctly specified and as such we would expect all methods to provide consistent estimates for the treatment effects in this scenario.

In the second scenario we simulate data where there is no treatment effect. This may affect the results obtained from g-computation formula due to an issue known as the g-null paradox, which can lead to the method providing incorrect treatment effect estimates in situations when there is no treatment effect.[88, 89]

The third scenario adds long-term direct effects of treatment. In this case, the long-term direct effects are actually negative effects, slightly counteracting the beneficial one-year effects, resulting in a decrease in the treatment effect through time. All of the methods should be able to deal with this scenario and provide consistent estimates of the treatment effect.

The fourth scenario simulates modification of the treatment effect by time-varying covariates. Although effect modification is generally not shown in DAGs, we have included arrows in Figure 5.4 to help illustrate this feature. The effect of D_t on L_t is modified by L_{t-1} and the effect of D_t on V_{t+1} is modified by V_t. Only three of the methods can provide estimates of treatment effect modification: HA-MSM, g-estimation and SCMM, but the other methods should still provide consistent estimates of the population-average treatment effect, as the models used do not assume no treatment effect modification.

The fifth scenario reverses the direction of the causal pathway between treatment and IV days in the same year. In our analyses, we will always assume that the direction of the causal pathway is from V_t to D_t, even when the data have actually been simulated the other way around, i.e. D_t affects V_t. This is a key issue in the UK CF Registry due to the fact that all data are collected at the end of each year, but many variables actually summarise the previous year (e.g. the total number of IV days since the previous annual assessment). Misspecifying the models means that all methods will probably perform poorly in this setting, but it is not clear to what extent they will show bias and whether some methods will be more affected than others.

In the final scenario some individuals are censored and so have fewer than five visits. The data were generate such that an individual's probability of being censored is affected by variables measured at the previous visit. This corresponds to a missing at random scenario, whereby the probability of being censored only depends on observed variables. Three of the methods (IPW of MSM, HA-MSM and g-estimation) can be extended to account for censoring by using censoring weights, and provided the censoring weights are correctly specified, these methods should therefore not be affected by censoring. SCMM should not be affected by censoring, because they only estimate short-term effects, and g-formula uses these same short-term models to then simulate data without censoring and so should similarly be unaffected by censoring.

5.4 Design of Simulation Studies

Simulation studies were performed following the guidelines given by Burton et al.[102] in order to investigate how the five features and challenges identified in Section 5.2 affect the performance of the methods described in Chapter 4 in each of the scenarios given above. The aims of the simulation studies are to understand how the performance of the

analysis methods might be affected by these challenges and to help provide a framework for the best analysis strategy for the real UK CF Registry data.

All simulations and analyses were carried out using Stata (Version 15). For each scenario, we simulated 1000 datasets, and we applied all the analysis methods to each dataset. This results in moderately independent simulations where for each scenario a new set of datasets was generated, but within each scenario the same set of simulated independent datasets was used to compare the statistical methods. The data were generated so as to imitate the observed data in the registry. In the real data, there are many treatments that individuals could be receiving and also many covariates which might be confounders. For the simulation studies we kept just one binary treatment, D, and the two outcome variables, lung function (L) and annual IV days (V), which also act as time-dependent confounders. Lung function was simulated as a continuous variable with a normal distribution, conditional on observed history, and IV days as a count outcome following a zero-inflated negative binomial distribution, conditional on observed history. In addition to these three variables, we also generated age from a beta distribution corresponding to what was observed in the real data, to act as a baseline confounder.

5.4.1 Simulation of Datasets

For each scenario the starting seeds used to generate each simulated dataset were pseudo-random numbers generated from a uniform distribution using the following formula:

$$\text{Seed} = \text{floor}(\text{unif}(0,1) \times 1000000 + 1000000 \times (n-1)) \qquad (1 \leq n \leq 1000). \qquad (5.1)$$

For each scenario datasets were simulated with 7500 individuals with 6 visits ($t = 0, ..., 5$). The data were simulated sequentially starting with visit 0. Table 5.1 lists the variables that were simulated.

Name	Variable	Type	Distribution	Range
Visit	T	Integer	-	0-5
Censored	C	Binary	Bernoulli	0, 1
Baseline Age	A	Continuous	Beta	6-90
Dornase Alfa	D	Binary	Bernoulli	0, 1
Lung Function	L	Continuous	Normal	10-
IV Days	V	Integer	Zero Inflated Negative Binomial	0-365
Exacerbation (IV Days > 0)	E	Binary	Bernoulli	0, 1

TABLE 5.1: List of simulated variables

The data for visit 0 were simulated using the following formulae:

$$a = 6 + 84 * \text{beta}(1.1, 5) \tag{5.2}$$

$$\begin{aligned} v_0 = \min\Big(365, \Big(\mathbb{1}\big(\text{unif}(0,1) < \text{expit}(0.004a - 0.6)\big)\Big) \\ \times \text{poisson}\Big(\text{gamma}\big(2, 0.5\text{exp}(2.4 - 0.002a)\big)\Big)\Big) \end{aligned} \tag{5.3}$$

$$e_0 = \mathbb{1}(v_0 > 0) \tag{5.4}$$

$$d_0 = 0 \tag{5.5}$$

$$l_0 = \max\big(10, \text{norm}(95 - 0.65a - 5.8e_0 - 0.3v_0, 20)\big) \tag{5.6}$$

$$c_0 = 0 \tag{5.7}$$

Data for visits 1 to 5 were then simulated sequentially using the following formulae, where α, β and γ were varied to create the six different scenarios (see Table 5.2):

$$\begin{aligned} v_t = \min\Big(365, \Big(\mathbb{1}\big(\text{unif}(0,1) < \text{expit}(1 + (\alpha_1 + \alpha_2 e_{t-1} + \alpha_3 v_{t-1}) d_{t-1} \\ + (\alpha_4 + \alpha_5 e_{t-1} + \alpha_6 v_{t-1}) \sum_{i=0}^{t-1} x_i \\ + e_{t-1} + 0.05 v_{t-1} - 0.025 l_{t-1} - 0.02a)\big)\Big) \\ * \text{poisson}\Big(\text{gamma}\big(\frac{1}{0.3}, 0.3\text{exp}(3.6 + (\alpha_7 + \alpha_8 e_{t-1} + \alpha_9 v_{t-1}) d_{t-1} \\ + (\alpha_1 0 + \alpha_1 1 e_{t-1} + \alpha_1 2 v_{t-1}) \sum_{i=0}^{t-1} d_i \\ + 0.1 e_{t-1} + 0.01 v_{t-1} - 0.0075 l_{t-1} - 0.003a)\big)\Big)\Big) \end{aligned} \tag{5.8}$$

$$e_t = \mathbb{1}(v_t > 0) \tag{5.9}$$

$$d_t = \mathbb{1}\big(\text{unif}(0,1) < \text{expit}(4d_{t-1} + 0.7e_t + 0.001v_t - 0.01l_{t-1} - 0.02a - 0.4)\big) \tag{5.10}$$

$$\begin{aligned} l_t = \max\big(10, \text{norm}(10 + (\beta_1 + \beta_2 l_{t-1}) x_t + (\beta_3 + \beta_4 l_{t-1}) \sum_{i=0}^{t} d_i \\ + 0.9 l_{t-1} - 0.7 e_t - 0.06 v_t - 0.08a, 10)\big) \end{aligned} \tag{5.11}$$

$$c_t = \gamma * \mathbb{1}\big(\text{unif}(0,1) < \text{expit}(0.02a - 0.03l_t - 0.9e_t + 0.02v_t - 0.1d_t - 2)\big) \tag{5.12}$$

The only scenario for which the data were simulated differently was the reversed causal pathways scenario, where treatment (d_t) was simulated prior to IV days (v_t), i.e. Equation 5.10 was simulated before Equation 5.8, with v_t and e_t replaced by v_{t-1} and e_{t-1} respectively in Equation 5.10, and d_{t-1} replaced by d_t in Equation 5.8.

The simulated datasets were checked to ensure that the previous formulae resulted in data which is distributed similarly to that observed in the real UK CF registry.

Scenario	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
Standard	-2	0	0	0	0	0	-0.8	0	0
No Effect	0	0	0	0	0	0	0	0	0
Decreasing Effect	-2	0	0	0.25	0	0	-0.8	0	0
Effect Modification	-0.8	-0.4	-0.02	0.1	0.05	0.0025	-0.2	-0.2	-0.02
Reversed Causal Pathway	-2	0	0	0	0	0	-0.8	0	0
Censoring	-2	0	0	0	0	0	-0.8	0	0

Scenario	α_{10}	α_{11}	α_{12}	β_1	β_2	β_3	β_4	γ
Standard	0	0	0	4	0	0	0	0
No Effect	0	0	0	0	0	0	0	0
Decreasing Effect	0.1	0	0	4	0	-0.5	0	0
Effect Modification	0.025	0.025	0.0025	8	-0.08	-1	0.01	0
Reversed Causal Pathway	0	0	0	4	0	0	0	0
Censoring	0	0	0	4	0	0	0	1

TABLE 5.2: Parameter values used in Equations 5.2 to 5.12 to simulate the six different scenarios

5.4.2 Analysis of Simulated Datasets

We analysed the data generated under each scenario using seven methods: IPW of MSM, IPW of MSM with truncated weights, HA-MSM, HA-MSM with truncated weights, SCMM, g-formula and g-estimation. The methods were all implemented as described in Chapter 4. For the methods involving truncation, the weights are truncated to the 1 and 99 percentile. We fit all seven methods without any interaction terms to estimate the population-average effect, even in the scenario where there was in fact some effect modification. Additionally, for the four methods that can accommodate estimation of interactions between treatment and time-varying covariates (HA-MSM, HA-MSM with truncation, SCMM and g-estimation), we analysed simulated data from three of the scenarios (no effect scenario, standard scenario and effect modification scenario) with interaction terms included.

For each scenario we perform analyses for two outcome variables: the first considering lung function (continuous) as the outcome and the second annual IV days (count) as the outcome. For each simulation, the coefficients corresponding to the cumulative treatment effects will be stored. For the continuous outcome, this is:

$$\mathrm{E}\left(L_s^{\overline{D}_s=\overline{1},\overline{D}_t=\overline{0}} - L_s^{\overline{D}_s=\overline{0}}\right) = \sum_{i=1}^{s-t+1} \phi_i, \tag{5.13}$$

and for the count outcome, there are two sets of coefficients. The first corresponding to the odds ratio (OR) of having a zero count:

$$\frac{\mathrm{P}\left(V_{s+1}^{\overline{D}_s=\overline{1},\overline{D}_t=\overline{0}}=0\right)}{\mathrm{P}\left(V_{s+1}^{\overline{D}_s=\overline{0}}=0\right)} = \exp\left(\sum_{i=1}^{s-t+1}\beta_i\right), \tag{5.14}$$

and the second corresponding to the IRR of the count:

$$\frac{\mathrm{E}\left(V_{s+1}^{\overline{D}_s=\overline{1},\overline{D}_t=\overline{0}}\right)}{\mathrm{E}\left(V_{s+1}^{\overline{D}_s=\overline{0}}\right)} = \exp\Big(\sum_{i=1}^{s-t+1}\gamma_i\Big). \tag{5.15}$$

For SCMM, as discussed in Chapter 4, only the coefficient corresponding to one-year of treatment will be stored. For all other methods we estimate the effects of one-, two-, three-, four-, and five-years' treatment use on the continuous outcome, i.e. in Equation 5.13 for $s-t+1=1,...,5$. For the count outcome, we only estimate the one-, two-, three- and four-year treatment effects ($s-t+1=1,...,4$ in Equations 5.14 and 5.15). The reason for this difference is that in the UK CF Registry, we define the one-year effect of treatment on lung function (continuous) as $D_t \rightarrow L_t$, whereas the one-year effect of treatment on IV days is defined as $D_t \rightarrow V_{t+1}$. As such, there is always one extra year of data available for the lung function outcome in the UK CF Registry, and we keep this consistent in the simulation study.

As detailed in Section 4.5.1, it is more complicated to calculate the cumulative effect of treatment when there is an interaction term with a time-varying variable included in the models. However, it is possible to do this for all methods when the outcome is continuous, and therefore in this case, we do present the results of the cumulative interaction terms per 10 change in L_{t-1}:

$$\mathrm{E}\left(L_s^{\overline{D}_s=\overline{1},\overline{D}_t=\overline{0}}\middle|L_{t-1}=10\right)-\mathrm{E}\left(L_s^{\overline{D}_s=\overline{1},\overline{D}_t=\overline{0}}\middle|L_{t-1}=0\right)=10\sum_{i=1}^{s-t+1}\phi_{\mathrm{int}_i}, \tag{5.16}$$

where $s-t+1=1,...,5$.

For the count outcome, in g-estimation, it is not possible to give one cumulative interaction effect, and therefore for the count outcome, rather than the cumulative effects, we present each separate interaction effect estimate per 14 day change in V_t (equivalent to a 2-week course of IVs): $14\psi_{\mathrm{int}_i}$ ($i=1,...,4$). Similarly, although it would be possible to give the cumulative interaction effect estimates in HA-MSM, for comparison purposes, we will present the separate interaction effects $14\beta_{\mathrm{int}_i}$ (for the zero count process) and $14\gamma_{\mathrm{int}_i}$ (for the rate of the count).

We compare the methods in terms of the bias, empirical SE and mean squared error. Although it is known that the model-based SEs are biased for most of these methods, we will also store the estimated robust SEs so as to compare them to the empirical SEs. We also count the number of simulated datasets for which each method failed to converge.

5.5 Results of Simulation Studies

In the following two subsections we present the results of the simulation studies for the continuous and count outcomes respectively.

5.5.1 Continuous Outcome

Figure 5.5 shows kernel density plots showing the results of the simulation studies for the conditionally normally-distributed continuous outcome. Table 5.3 presents the same results, showing the number of simulated datasets for which an estimate was obtained, the mean effect estimate, the bias, the empirical SE, the model-based SE and the mean squared error (MSE). We only present results for the one-year effect and the five-year effect to show the two extremes of short- to long-term effects. In all cases, the results for the two- to four-year effects followed the trend between the one-year and five-year effects. The results can be seen in full in Figure C.1 and Tables C.1 and C.2 of Appendix C.

Almost all methods appear to provide consistent estimators in the 'standard' scenario where all the models are correctly specified. The only method which performs poorly here is IPW of MSM with truncation of extreme weights, but this is to be expected as it is known that due to truncation of the weights this method no longer fully accounts for confounding. The five-year treatment effect estimates are slightly biased, but when using a much larger sample size all methods were unbiased, therefore we believe this residual bias is due to the sample size, which we have kept at 7500 individuals as it is unlikely that we would ever obtain a larger sample from the UK CF Registry.

These findings are repeated for the scenarios where there is no treatment effect, where the treatment effect decreases over time and where there is censoring (provided that censoring weights are used for IPW, HA-MSM and g-estimation).

For the scenario where the causal pathway between a confounder and treatment is reversed and the assumed direction is incorrect in the analysis, we find that the situation is the opposite: all methods are biased, but IPW and HA-MSM perform comparatively well when the weights are truncated. However, untruncated they perform very poorly with very large variability and even fail to converge in many simulated datasets.

When considering effect modification by time-varying covariates, all the methods can still be used to provide an estimate for the population-average effect. For the one-year effects all the methods appeared to provide consistent estimators, however at five-years there was some noticeable bias for g-estimation and HA-MSM. These are the two methods which can incorporate the estimation of effect modification by time-varying covariates, and not including these interactions terms when they are in fact present has

introduced bias. Conversely, although IPW and g-formula cannot estimate interaction terms, they do not assume that there is no effect modification and can therefore provide consistent estimators for the population-average effect.

If the aim is to estimate the extent of any effect modification by time-varying covariates, then it would be necessary to use HA-MSM, SCMM, or g-estimation and these results for the one- and five- year effects are presented in Figure 5.6 and Table 5.4. (The two-, three- and four-year effect results are given in Figure C.2 and Tables C.3 and C.3 of Appendix C). We see that in the case where is no treatment effect all three methods correctly estimate no interaction effect on average. However, in the scenario where there is a treatment effect, but no effect modification, incorrectly including the interaction terms in these models has introduced bias, and the methods estimate that there is some effect modification. When there is effect modification and hence including the interaction term is correct, then all three methods perform similarly well in estimating the strength of the effect modification.

When considering the SE, only in SCMM and HA-MSM did the model-estimated SEs approximate the empirical SEs. This is theoretically known in the case of SCMM with the propensity score known, and therefore the model-based SEs will be approximately correct when the propensity score is well estimated. In the case of HA-MSM, we believe this to be a peculiarity of our simulation setting and it is unlikely to be true generally. For this reason, for all methods other than SCMM, a bootstrap procedure should be used to obtain reliable estimates of the SEs. Comparing the methods, g-formula consistently shows the smallest empirical SEs, followed by SCMM, g-estimation and HA-MSM with similar SEs, and finally IPW with the largest SEs. In the scenario of reversed causal pathways, IPW and HA-MSM had especially large SEs when untruncated weights were used.

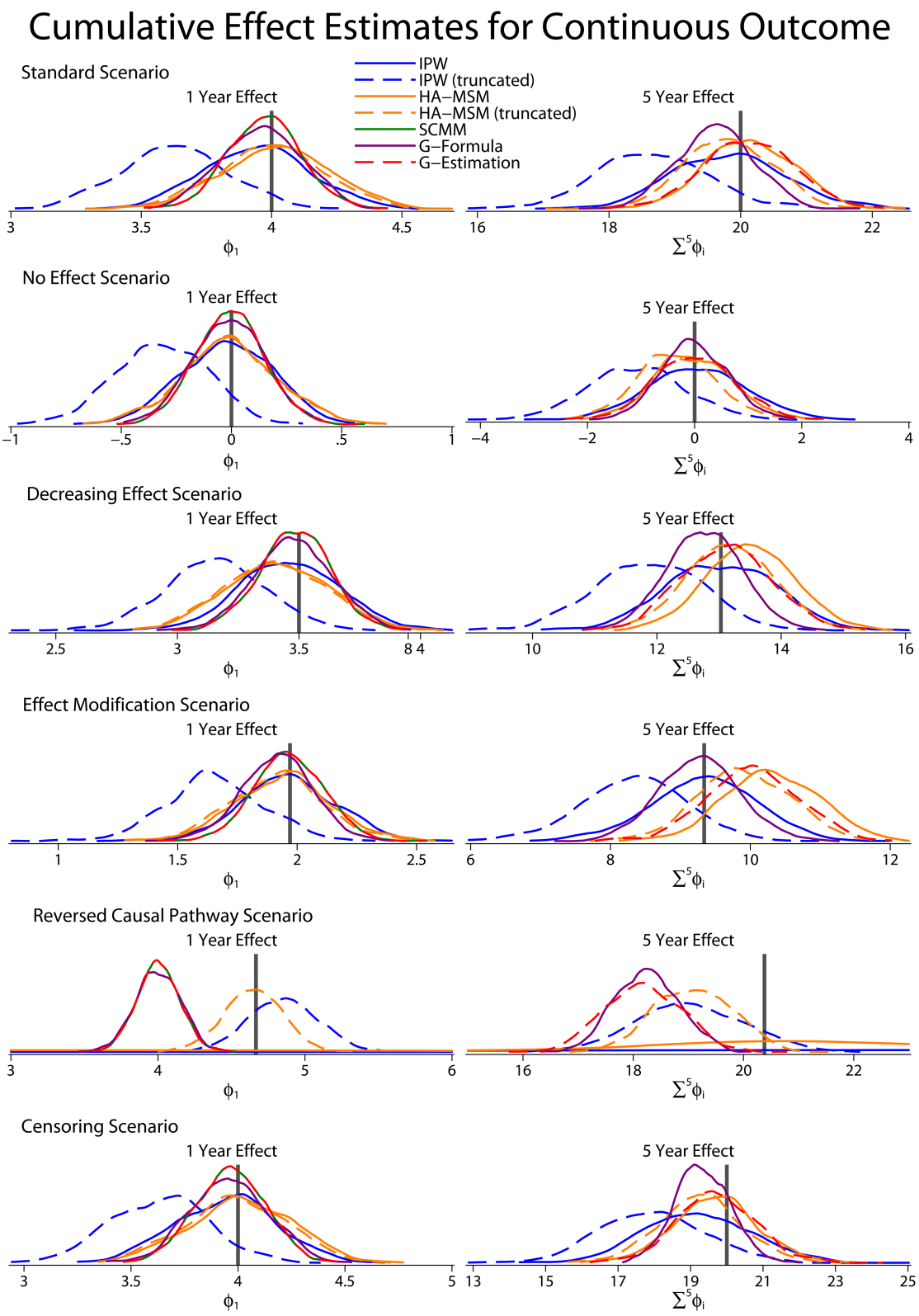


FIGURE 5.5: Kernel density plots of population-average effect estimates for a continuous outcome (The vertical line shows the correct effect)

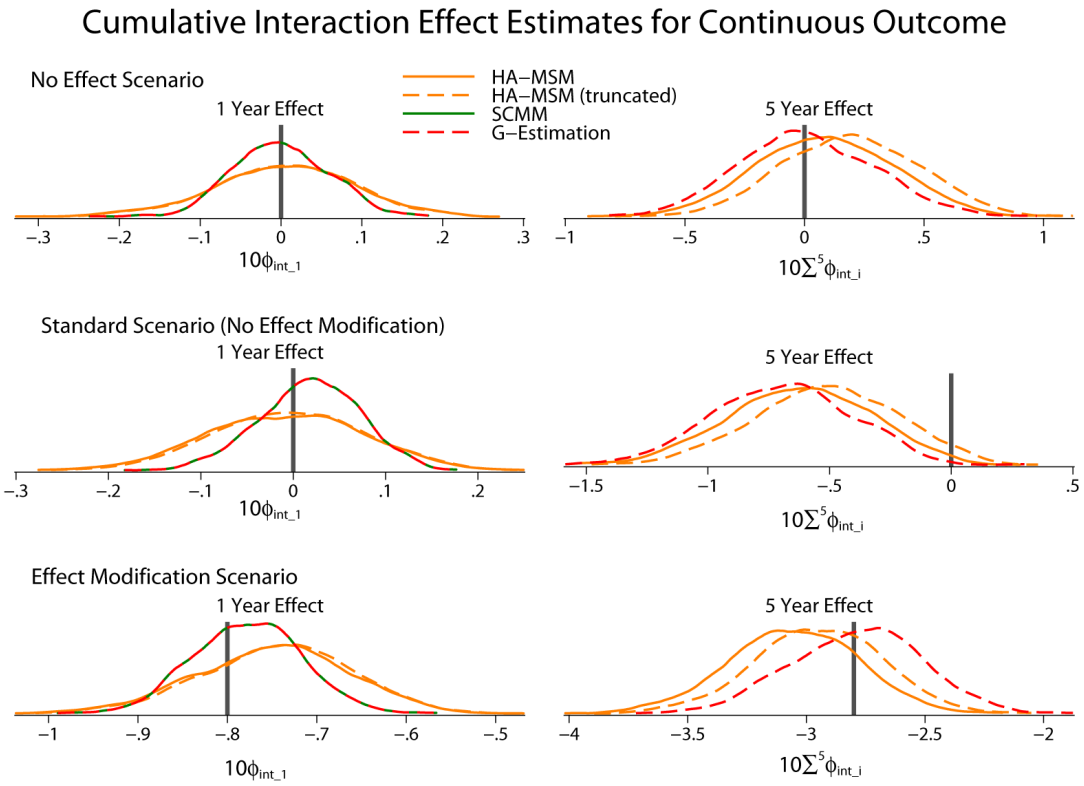


FIGURE 5.6: Kernel density plots of the interaction effect estimates for a continuous outcome (The vertical line shows the correct effect)

Scenario	Method	n	1 Year Treatment Effect (ϕ_1)					Cumulative 5 Year Treatment Effect ($\sum_{i=1}^{5} \phi_i$)				
			Mean	Bias	Empirical SE	Model SE	MSE	Mean	Bias	Empirical SE	Model SE	MSE
Standard $\phi_1 = 4.00$ $\sum_1^5 \phi_i = 20.00$	IPW of MSM	1000	3.94	-0.057	0.20	0.39	0.044	19.82	-0.18	0.87	1.05	0.79
	IPW of MSM (truncated)	1000	3.63	-0.37	0.20	0.38	0.18	18.66	-1.34	0.86	1.01	2.53
	HA-MSM	1000	4.02	-0.020	0.21	0.21	0.044	20.13	0.13	0.70	0.73	0.50
	HA-MSM (truncated)	1000	3.99	-0.007	0.21	0.21	0.042	19.81	-0.19	0.68	0.72	0.50
	SCMM	1000	3.97	-0.027	0.14	0.14	0.020	NA	NA	NA	NA	NA
	G-Formula	1000	3.96	-0.038	0.15	0.15	0.025	19.62	-0.38	0.58	0.33	0.49
	G-Estimation	1000	3.97	-0.027	0.14	0.14	0.020	20.12	0.12	0.69	1.19	0.49
No Effect $\phi_1 = 0.00$ $\sum_1^5 \phi_i = 0.00$	IPW of MSM	1000	-0.020	-0.020	0.21	0.39	0.044	0.050	0.005	0.92	1.08	0.85
	IPW of MSM (truncated)	1000	-0.32	-0.32	0.21	0.38	0.14	-1.09	-1.09	0.90	1.04	2.01
	HA-MSM	1000	-0.006	-0.006	0.22	0.22	0.047	-0.029	-0.029	0.72	0.73	0.53
	HA-MSM (truncated)	1000	-0.011	-0.011	0.21	0.22	0.046	-0.35	-0.35	0.72	0.72	0.64
	SCMM	1000	-0.007	-0.007	0.15	0.14	0.024	NA	NA	NA	NA	NA
	G-Formula	1000	-0.014	-0.014	0.16	0.15	0.026	-0.008	-0.008	0.61	0.33	0.37
	G-Estimation	1000	-0.007	-0.007	0.15	0.14	0.024	-0.006	-0.006	0.75	1.22	0.56
Decreasing Effect $\phi_1 = 3.50$ $\sum_1^5 \phi_i = 13.03$	IPW of MSM	1000	3.46	-0.039	0.21	0.39	0.046	13.02	-0.007	0.89	1.05	0.79
	IPW of MSM (truncated)	1000	3.15	-0.35	0.21	0.38	0.16	11.87	-1.16	0.88	1.01	2.12
	HA-MSM	1000	3.42	-0.082	0.21	0.22	0.052	13.51	0.48	0.71	0.73	0.73
	HA-MSM (truncated)	1000	3.40	-0.098	0.21	0.21	0.053	13.18	0.15	0.70	0.72	0.51
	SCMM	1000	3.48	-0.016	0.14	0.14	0.021	NA	NA	NA	NA	NA
	G-Formula	1000	3.48	-0.023	0.16	0.15	0.062	12.76	-0.27	0.61	0.33	0.44
	G-Estimation	1000	3.48	-0.016	0.14	0.14	0.021	13.12	0.088	0.72	1.19	0.53
Effect Modification $\phi_1 = 1.97$ $\sum_1^5 \phi_i = 9.34$	IPW of MSM	1000	1.96	-0.008	0.21	0.38	0.043	9.34	-0.0002	0.77	0.90	0.59
	IPW of MSM (truncated)	1000	1.65	-0.32	0.21	0.37	0.14	8.34	-1.00	0.76	0.86	1.58
	HA-MSM	1000	1.94	-0.029	0.20	0.22	0.041	10.31	0.97	0.67	0.67	1.39
	HA-MSM (truncated)	1000	1.91	-0.057	0.20	0.21	0.043	9.90	0.56	0.66	0.66	0.76
	SCMM	1000	1.97	-0.004	0.15	0.14	0.022	NA	NA	NA	NA	NA
	G-Formula	1000	1.92	-0.054	0.15	0.14	0.026	9.27	-0.067	0.57	0.30	0.33
	G-Estimation	1000	1.97	-0.004	0.15	0.14	0.022	10.03	0.69	0.67	1.12	0.92
Reversed Causal Pathway $\phi_1 = 4.67$ $\sum_1^5 \phi_i = 20.38$	IPW of MSM	942	-31.45	-36.12	19.36	3.37	1679.12	32.14	11.76	20.64	5.01	563.79
	IPW of MSM (truncated)	1000	4.85	0.19	0.24	0.43	0.090	19.03	-1.35	0.99	1.15	2.80
	HA-MSM	926	-4.62	-9.29	10.18	2.54	189.74	23.44	3.06	8.19	2.97	76.34
	HA-MSM (truncated)	1000	4.65	-0.022	0.21	0.22	0.043	19.09	-1.29	0.71	0.73	2.16
	SCMM	1000	4.00	-0.67	0.15	0.14	0.46	NA	NA	NA	NA	NA
	G-Formula	1000	4.00	-0.67	0.16	0.15	0.47	18.25	-2.13	0.58	0.32	4.89
	G-Estimation	1000	4.00	-0.67	0.15	0.14	0.46	18.18	-2.20	0.71	1.23	5.36

Scenario	Method	n	1 Year Treatment Effect (ϕ_1)					Cumulative 5 Year Treatment Effect ($\sum_{i=1}^{5} \phi_i$)				
			Mean	Bias	Empirical SE	Model SE	MSE	Mean	Bias	Empirical SE	Model SE	MSE
Censoring $\phi_1 = 4.00$ $\sum_1^5 \phi_i = 20.00$	IPW of MSM	1000	3.94	-0.058	0.23	0.45	0.058	19.20	-0.80	1.49	1.59	2.86
	IPW of MSM (truncated)	1000	3.63	-0.37	0.23	0.43	0.19	17.94	-2.06	1.46	6.38	1.54
	HA-MSM	1000	4.02	0.018	0.24	0.25	0.058	19.69	-0.31	1.17	1.46	1.16
	HA-MSM (truncated)	1000	3.99	-0.011	0.24	0.056	0.25	19.27	-0.73	1.15	1.14	1.87
	SCMM	1000	3.97	-0.035	0.16	0.17	0.027	NA	NA	NA	NA	NA
	G-Formula	1000	3.96	-0.041	0.18	0.17	0.035	19.33	-0.67	0.76	0.55	1.02
	G-Estimation	1000	3.97	-0.035	0.16	0.17	0.027	19.70	-0.30	1.10	1.68	1.31

TABLE 5.3: Simulation study results of population-average effects for continuous outcome ($\overline{D_t} \rightarrow L_t$) (NA signifies that the method does not estimate that effect)

Scenario	Method	n	1 Year Interaction Effect ($10\phi_{\text{int}_1}$)					Cumulative 5 Year Interaction Effect ($10\sum_{i=1}^{5} \phi_{\text{int}_i}$)				
			Mean	Bias	Empirical SE	Model SE	MSE	Mean	Bias	Empirical SE	Model SE	MSE
No Effect $10\phi_{\text{int}_1} = 0.00$ $10\sum_1^5 \phi_{\text{int}_i} = 0.00$	HA-MSM	1000	0.003	0.003	0.088	0.087	0.008	0.088	0.088	0.29	0.30	0.092
	HA-MSM (truncated)	1000	0.004	0.004	0.086	0.085	0.007	0.19	0.19	0.29	0.30	0.12
	SCMM	1000	-0.0002	-0.0002	0.060	0.060	0.004	NA	NA	NA	NA	NA
	G-Estimation	1000	-0.0002	-0.0002	0.060	0.060	0.004	0.009	0.009	0.28	0.46	0.080
Standard (no effect modification) $10\phi_{\text{int}_1} = 0.00$ $10\sum_1^5 \phi_{\text{int}_i} = 0.00$	HA-MSM	1000	-0.009	-0.009	0.084	0.085	0.007	-0.63	-0.63	0.30	0.30	0.48
	HA-MSM (truncated)	1000	-0.004	-0.004	0.081	0.083	0.007	-0.52	-0.52	0.29	0.30	0.35
	SCMM	1000	0.018	0.018	0.054	0.058	0.003	NA	NA	NA	NA	NA
	G-Estimation	1000	0.018	0.018	0.054	0.058	0.003	-0.69	-0.69	0.29	0.45	0.56
Effect Modification $10\phi_{\text{int}_1} = -0.80$ $10\sum_1^5 \phi_{\text{int}_i} = -2.80$	HA-MSM	1000	-0.75	0.050	0.085	0.087	0.010	-3.06	-0.26	0.27	0.27	0.14
	HA-MSM (truncated)	1000	-0.74	0.056	0.084	0.085	0.010	-2.95	-0.15	0.27	0.27	0.096
	SCMM	1000	-0.78	0.022	0.061	0.060	0.004	NA	NA	NA	NA	NA
	G-Estimation	1000	-0.78	0.022	0.061	0.060	0.004	-2.79	0.012	0.28	0.39	0.077

TABLE 5.4: Simulation study results of interaction effects for continuous outcome. Results show change in effect of $\overline{D_t}$ on L_t per 10 change in L_{t-1} (NA signifies that the method does not estimate that effect)

5.5.2 Count Outcome

Unlike with the continuous outcome, due to the issue of non-collapsibility, we do not compare the effect estimates for the count outcome to a 'correct' value. However, Figures 5.7 and 5.8, and Tables 5.5 and 5.6 present the one- and four-year effect estimates and SEs for both the odds of a zero count and the rate of the count. As with the continuous outcome, results for the two- and three- year effects can be found in Appendix C in Figures C.3 and C.4, and Tables C.5 and C.6.

Both IPW and g-formula provide marginal effect estimates, and in almost all cases provide very similar estimates. The only scenario in which they do not provide similar estimates is the case of reversed causal pathways where IPW performs very poorly with very large variability, as was also seen for the continuous outcome.

Considering the three methods which provide conditional effect estimates, (HA-MSM, g-estimation and SCMM), we note that the methods are not in general in agreement, and this is due to the fact that the final models condition on different subsets of variables. In the case of g-estimation, due to the fact that the method can only estimate a rate, (rather than also accounting for the separate process of excess zeroes), the estimates from this method are generally very different from all other methods.

The only scenario in which all five methods are in agreement is when there is no treatment effect. Here both the marginal and conditional effect estimates are zero. This suggests that any method could be used to perform a test of the null hypothesis of no treatment effect, but the strength of any effect estimates cannot directly be compared between methods.

The results for the one- and four-year interaction terms are presented in Figures 5.9 and 5.10, and Tables 5.7 and 5.8. (Appendix C contains the results for the two- and three-year interaction effects in Figures C.5 and C.6, and Tables C.7 and C.8). The findings are similar to those found in the case of interaction terms with continuous outcomes, except for the results from g-estimation: even when there is no effect modification in the data generation process, this method did not on average indicate no effect modification. This is again due to non-collapsibility, where although there is no effect modification present when assuming the outcome follows a zero-inflated negative binomial distribution, there may be under different outcome models.

As we found for the continuous outcome the model estimated SEs from HA-MSM and SCMM approximated the empirical SEs well, but again we would recommend a bootstrap procedure to be used for all methods other than SCMM.

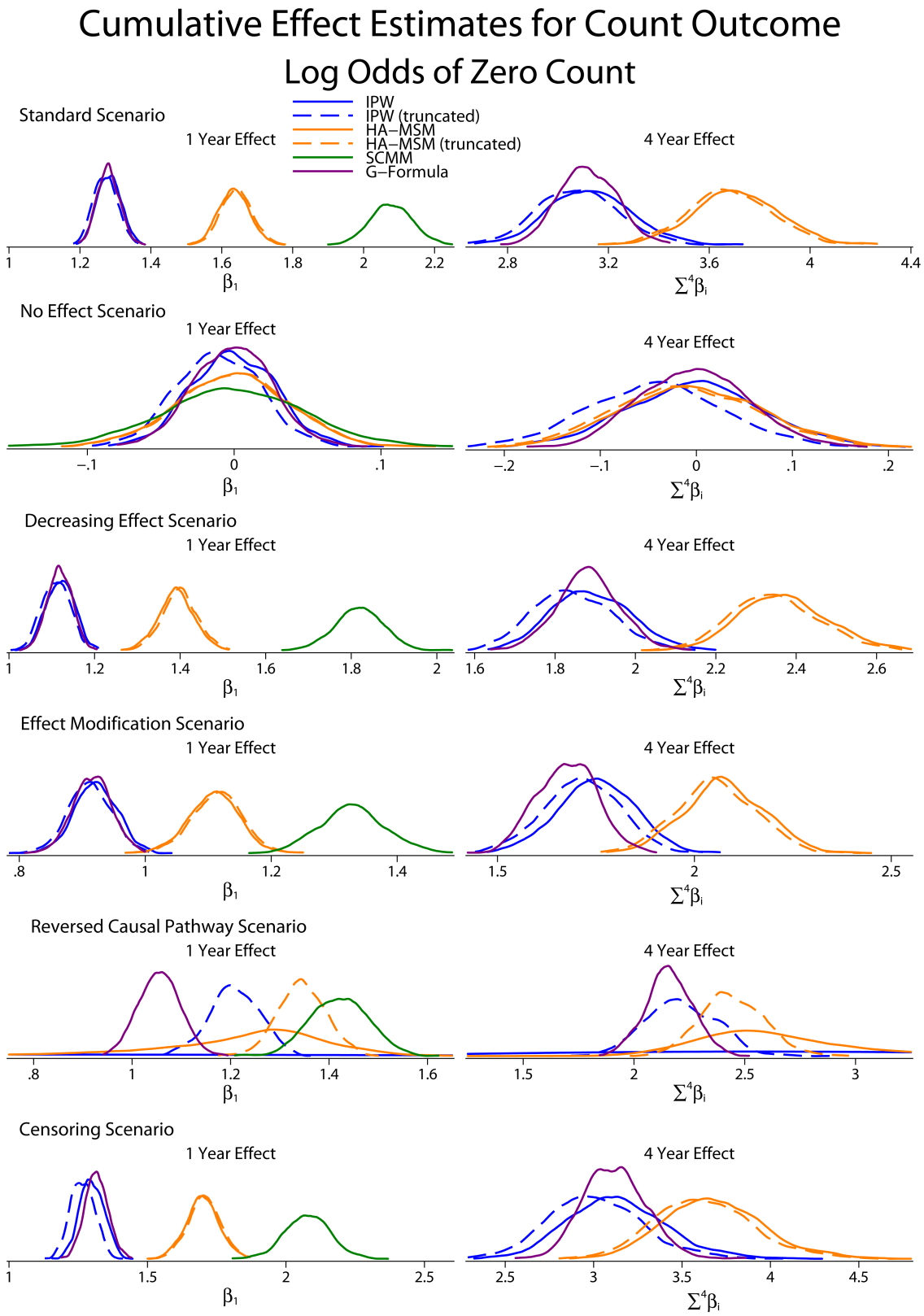


FIGURE 5.7: Kernel density plots of the population-average effect estimates for the odds of a zero count

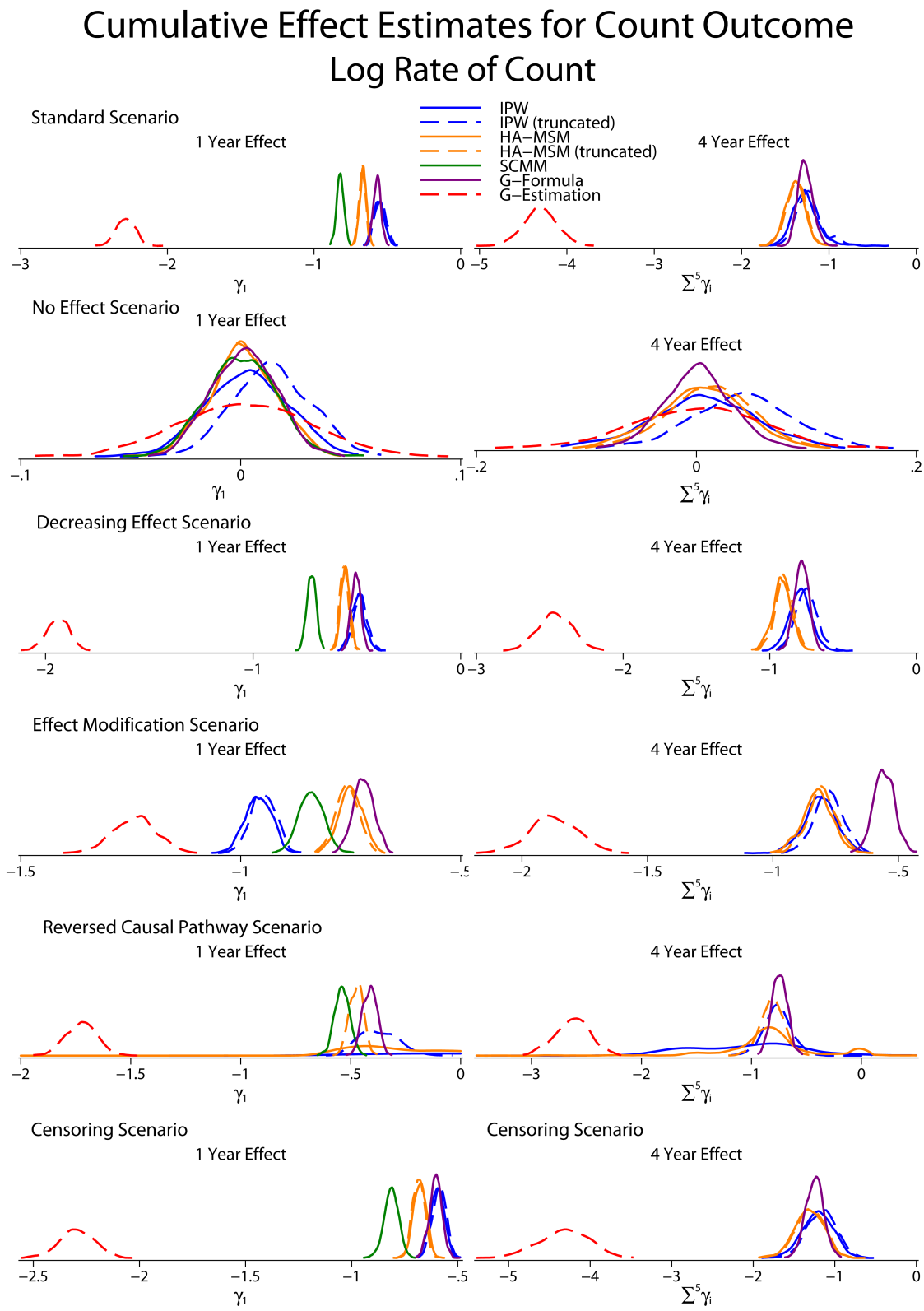


FIGURE 5.8: Kernel density plots of the population-average effect estimates for the rate of a count

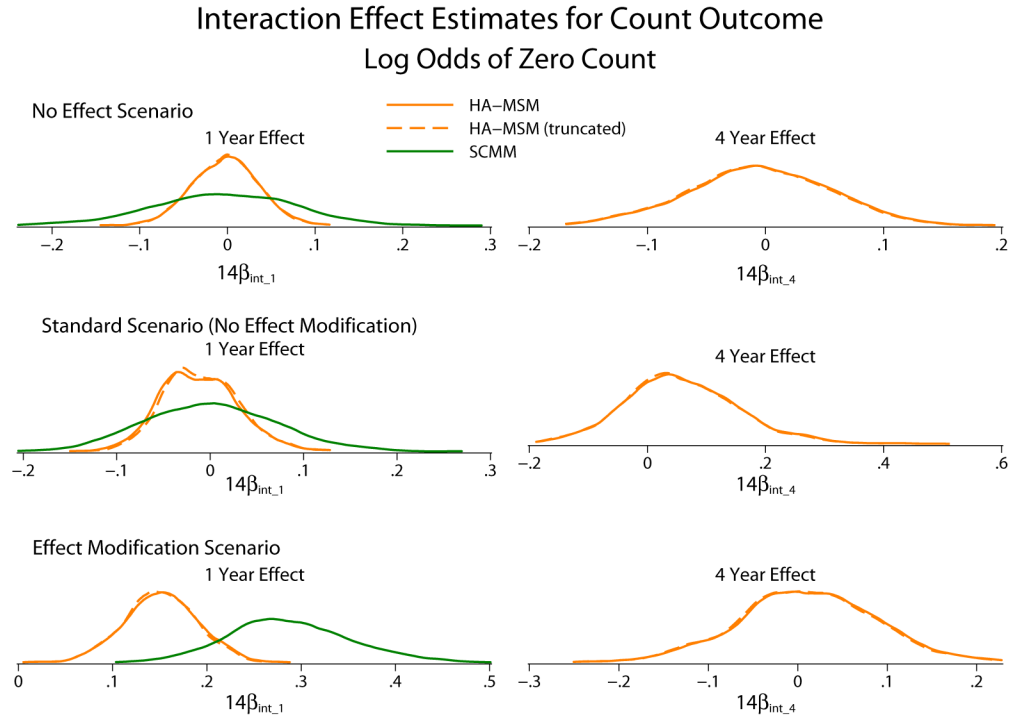


FIGURE 5.9: Kernel density plots of the interaction effect estimates for the odds of a zero count

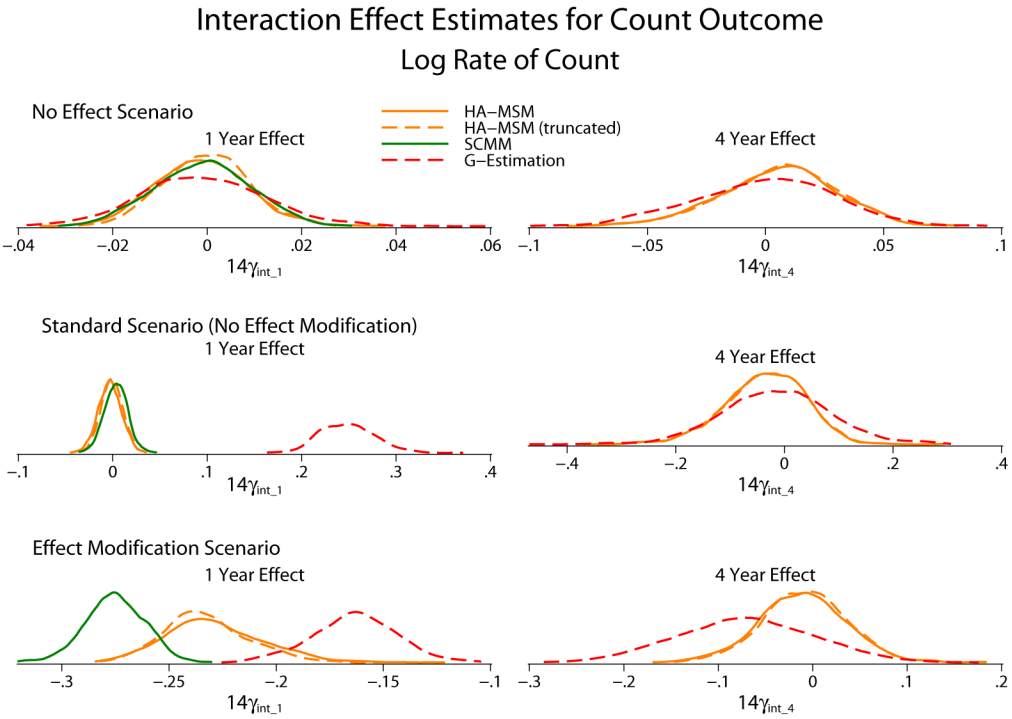


FIGURE 5.10: Kernel density plots of the interaction effect estimates for the rate of a count

Scenario	Estimate	Method	n	1 Year Treatment Effect					
				Log OR of Zero Count (β_1)			Log IRR of Count (γ_1)		
				Mean	Empirical SE	Model SE	Mean	Empirical SE	Model SE
Standard	Marginal	IPW of MSM	1000	1.28	0.033	0.040	-0.56	0.037	0.038
		IPW of MSM (truncated)	1000	1.27	0.033	0.039	-0.55	0.037	0.038
		G-Formula	1000	1.28	0.031	0.016	-0.57	0.026	0.017
	Conditional	HA-MSM	1000	1.63	0.043	0.045	-0.67	0.023	0.024
		HA-MSM (truncated)	1000	1.64	0.043	0.045	-0.67	0.023	0.023
		SCMM	1000	2.08	0.057	0.060	-0.82	0.025	0.024
		G-Estimation	1000	NA	NA	NA	-2.28	0.063	0.016
No Effect	Marginal	IPW of MSM	1000	-0.0007	0.028	0.035	0.003	0.019	0.022
		IPW of MSM (truncated)	1000	-0.011	0.028	0.035	0.014	0.018	0.020
		G-Formula	1000	-0.001	0.026	0.012	0.002	0.015	0.008
	Conditional	HA-MSM	1000	-0.0008	0.037	0.039	0.001	0.015	0.016
		HA-MSM (truncated)	1000	0.0005	0.037	0.039	0.001	0.015	0.0169
		SCMM	1000	-0.0007	0.047	0.048	0.0004	0.016	0.017
		G-Estimation	1000	NA	NA	NA	0.002	0.030	0.016
Decreasing Effect	Marginal	IPW of MSM	1000	1.12	0.032	0.038	-0.49	0.033	0.034
		IPW of MSM (truncated)	1000	1.11	0.033	0.038	-0.48	0.033	0.033
		G-Formula	1000	1.12	0.029	0.015	-0.50	0.022	0.015
	Conditional	HA-MSM	1000	1.39	0.041	0.042	-0.55	0.021	0.022
		HA-MSM (truncated)	1000	1.40	0.041	0.042	-0.56	0.021	0.022
		SCMM	1000	1.82	0.057	0.055	-0.72	0.022	0.022
		G-Estimation	1000	NA	NA	NA	-1.94	0.055	0.016
Effect Modification	Marginal	IPW of MSM	1000	0.92	0.034	0.039	-0.96	0.030	0.031
		IPW of MSM (truncated)	1000	0.91	0.034	0.039	-0.95	0.028	0.029
		G-Formula	1000	0.92	0.030	0.014	-0.72	0.023	0.010
	Conditional	HA-MSM	1000	1.11	0.041	0.042	-0.75	0.026	0.027
		HA-MSM (truncated)	1000	1.11	0.041	0.042	-0.076	0.026	0.026
		SCMM	1000	1.33	0.052	0.052	-0.84	0.028	0.029
		G-Estimation	1000	NA	NA	NA	-1.24	0.049	0.016

Scenario	Estimate	Method	1 Year Treatment Effect					
			n	Log OR of Zero Count (β_1)			Log IRR of Count (γ_1)	
				Mean	Empirical SE	Model SE	Mean	Empirical SE	Model SE
Reversed Causal Pathway	Marginal	IPW of MSM	758	-2.09	2.71	0.48	0.96	0.86	0.10
		IPW of MSM (truncated)	1000	1.21	0.051	0.058	-0.37	0.088	0.089
		G-Formula	1000	1.06	0.042	0.017	-0.41	0.033	0.014
	Conditional	HA-MSM	727	1.19	0.29	0.14	-0.11	0.50	0.37
		HA-MSM (truncated)	1000	1.35	0.049	0.049	-0.47	0.032	0.032
		SCMM	1000	1.42	0.060	0.061	-0.54	0.034	0.036
		G-Estimation	1000	NA	NA	NA	-1.72	0.071	0.016
Censoring	Marginal	IPW of MSM	1000	1.30	0.044	0.050	-0.60	0.033	0.034
		IPW of MSM (truncated)	1000	1.27	0.043	0.049	-0.59	0.032	0.033
		G-Formula	1000	1.32	0.042	0.020	-0.60	0.029	0.017
	Conditional	HA-MSM	1000	1.70	0.059	0.059	-0.68	0.031	0.030
		HA-MSM (truncated)	1000	1.71	0.059	0.058	-0.69	0.029	0.028
		SCMM	1000	2.08	0.082	0.081	-0.81	0.034	0.034
		G-Estimation	1000	NA	NA	NA	-2.29	0.083	0.082

TABLE 5.5: Simulation study results of one-year population-average effects for count outcome ($D_t \rightarrow V_{t+1}$) (NA signifies that the method does not estimate that effect)

Scenario	Estimate	Method	Cumulative 4 Year Treatment Effect					
			n	Log OR of Zero Count ($\sum_{i=1}^{4} \beta_i$)			Log IRR of Count ($\sum_{i=1}^{4} \gamma_i$)	
				Mean	Empirical SE	Model SE	Mean	Empirical SE	Model SE
Standard	Marginal	IPW of MSM	1000	3.13	0.16	0.16	-1.27	0.15	0.14
		IPW of MSM (truncated)	1000	3.08	0.16	0.16	-1.22	0.17	0.15
		G-Formula	1000	3.12	0.11	0.067	-1.27	0.085	0.067
	Conditional	HA-MSM	1000	3.71	0.17	0.17	-1.39	0.11	0.11
		HA-MSM (truncated)	1000	3.68	0.17	0.17	-1.37	0.11	0.10
		G-Estimation	1000	NA	NA	NA	-4.30	0.19	0.042
No Effect	Marginal	IPW of MSM	1000	-0.0007	0.071	0.075	0.006	0.047	0.053
		IPW of MSM (truncated)	1000	-0.042	0.071	0.074	0.042	0.045	0.048
		G-Formula	1000	-0.002	0.056	0.027	0.001	0.030	0.018
	Conditional	HA-MSM	1000	-0.002	0.073	0.074	0.004	0.039	0.041
		HA-MSM (truncated)	1000	-0.012	0.073	0.074	0.013	0.038	0.039
		G-Estimation	1000	NA	NA	NA	0.0006	0.058	0.042

Scenario	Estimate	Method	n	Cumulative 4 Year Treatment Effect					
				Log OR of Zero Count ($\sum_{i=1}^{4} \beta_i$)			Log IRR of Count ($\sum_{i=1}^{4} \gamma_i$)		
				Mean	Empirical SE	Model SE	Mean	Empirical SE	Model SE
Decreasing Effect	Marginal	IPW of MSM	1000	1.89	0.098	0.10	-0.79	0.074	0.076
		IPW of MSM (truncated)	1000	1.84	0.097	0.099	-0.75	0.074	0.074
		G-Formula	1000	1.88	0.077	0.040	-0.78	0.048	0.031
	Conditional	HA-MSM	1000	2.37	0.11	0.11	-0.92	0.065	0.063
		HA-MSM (truncated)	1000	2.35	0.11	0.11	-0.90	0.062	0.062
		G-Estimation	1000	NA	NA	NA	-2.46	0.12	0.042
Effect Modification	Marginal	IPW of MSM	1000	1.75	0.097	0.098	-0.81	0.065	0.066
		IPW of MSM (truncated)	1000	1.71	0.096	0.097	-0.78	0.058	0.060
		G-Formula	1000	1.68	0.076	0.039	-0.56	0.042	0.025
	Conditional	HA-MSM	1000	2.08	0.10	0.10	-0.82	0.060	0.058
		HA-MSM (truncated)	1000	2.06	0.10	0.10	-0.81	0.058	0.057
		G-Estimation	1000	NA	NA	NA	-1.87	0.096	0.042
Reversed Causal Pathway	Marginal	IPW of MSM	758	10.69	60.95	1.76	-0.93	12.37	0.18
		IPW of MSM (truncated)	1000	2.21	0.16	0.16	-0.76	0.13	0.12
		G-Formula	1000	2.16	0.11	0.057	-0.74	0.069	0.042
	Conditional	HA-MSM	727	3.13	2.19	0.65	-0.83	0.68	0.69
		HA-MSM (truncated)	1000	2.44	0.15	0.15	-0.81	0.11	0.11
		G-Estimation	1000	NA	NA	NA	-2.63	0.16	0.043
Censoring	Marginal	IPW of MSM	1000	3.14	0.28	0.28	-1.21	0.19	0.18
		IPW of MSM (truncated)	1000	3.01	0.27	0.27	-1.17	0.19	0.17
		G-Formula	1000	3.12	0.18	0.11	-1.25	0.11	0.080
	Conditional	HA-MSM	1000	3.68	0.28	0.29	-1.32	0.18	0.17
		HA-MSM (truncated)	1000	0.28	0.28	0.28	-1.30	0.18	0.17
		G-Estimation	1000	NA	NA	NA	-4.32	0.31	0.32

TABLE 5.6: Simulation study results of four-year population-average effects for count outcome ($\overline{D}_t \rightarrow V_{t+1}$) (NA signifies that the method does not estimate that effect)

Scenario	Estimate	Method	n	1 Year Interaction Effect					
				Log OR of Zero Count ($14\beta_{\text{int}_1}$)			Log IRR of Count ($14\gamma_{\text{int}_1}$)		
				Mean	Empirical SE	Model SE	Mean	Empirical SE	Model SE
No Effect	Conditional	HA-MSM	1000	-0.002	0.038	0.039	-0.001	0.010	0.009
		HA-MSM (truncated)	1000	-0.002	0.038	0.039	-0.0002	0.009	0.009
		SCMM	1000	-0.005	0.080	0.079	-0.0009	0.010	0.010
		G-Estimation	1000	NA	NA	NA	-0.0009	0.013	0.015
Standard (no effect modification)	Conditional	HA-MSM	1000	-0.012	0.041	0.040	-0.003	0.012	0.011
		HA-MSM (truncated)	1000	-0.009	0.040	0.039	-0.001	0.012	0.011
		SCMM	1000	0.001	0.070	0.069	0.003	0.012	0.012
		G-Estimation	1000	NA	NA	NA	0.25	0.028	0.016
Effect Modification	Conditional	HA-MSM	1000	0.15	0.039	0.039	-0.23	0.023	0.019
		HA-MSM (truncated)	1000	0.15	0.038	0.038	-0.23	0.020	0.017
		SCMM	1000	0.29	0.064	0.064	-0.28	0.014	0.013
		G-Estimation	1000	NA	NA	NA	-0.16	0.020	0.015

TABLE 5.7: Simulation study results of one-year interaction effects for count outcome. Results show change in effect of D_t on V_{t+1} per 14 day change in V_t (NA signifies that the method does not estimate that effect)

Scenario	Estimate	Method	n	4 Year Interaction Effect					
				Log OR of Zero Count ($14\beta_{\text{int}_4}$)			Log IRR of Count ($14\gamma_{\text{int}_4}$)		
				Mean	Empirical SE	Model SE	Mean	Empirical SE	Model SE
No Effect	Conditional	HA-MSM	1000	-0.004	0.060	0.059	0.005	0.025	0.023
		HA-MSM (truncated)	1000	-0.006	0.060	0.058	0.006	0.024	0.023
		G-Estimation	1000	NA	NA	NA	-0.0005	0.030	0.024
Standard (no effect modification)	Conditional	HA-MSM	1000	0.060	0.098	0.092	-0.035	0.075	0.051
		HA-MSM (truncated)	1000	0.058	0.097	0.092	-0.035	0.076	0.051
		G-Estimation	1000	NA	NA	NA	-0.015	0.10	0.026
Effect Modification	Conditional	HA-MSM	1000	0.012	0.077	0.074	-0.011	0.047	0.043
		HA-MSM (truncated)	1000	0.009	0.076	0.074	-0.008	0.045	0.042
		G-Estimation	1000	NA	NA	NA	-0.070	0.072	0.025

TABLE 5.8: Simulation study results of four-year interaction effects for count outcome. Results show change in effect of $\overline{D}_t$ on V_{t+1} per 14 day change in V_t (NA signifies that the method does not estimate that effect)

5.6 Discussion

In this chapter we have investigated the suitability of seven methods for estimating short- and long-term treatment effects in longitudinal observational data using simulation studies. The focus was on five scenarios that could be encountered in the UK CF Registry.

Our simulation studies showed that all the methods could be suitable for analysing the UK CF Registry data to investigate long-term treatment effects. Specifically in the standard scenario, where all models are correctly specified, all methods performed very similarly. However, the main issue is that when analysing the real UK CF Registry data we are not able to tell if the data are similar to the data simulated under the standard scenario, or whether one of the other scenarios is more realistic.

In the case of IPW, for example, there were noticeable differences between the results from applying the method with truncated and untruncated weights. In most scenarios, the untruncated weights performed best, but in the situation of the direction of causal pathways being misspecified the truncated weights showed much better performance. In a real data analysis, we would not know which scenario we are in; it would therefore be difficult to know when weights should be truncated or not. It may be sensible to only truncate when there are 'extreme' weights, but there is no clear definition of how large a weight must be before it is 'extreme'. This would suggest, therefore, in situations where there is uncertainty in the correct direction of causal pathways, that IPW not be used.

HA-MSM performed similarly to IPW of MSM in cases where there is no effect modification, but as it is a more complex method, it would be preferable to use standard IPW of MSM over HA-MSM in most cases.

For estimating the one-year effect of treatment, SCMM would probably be the preferred method due to its good performance in the simulation studies and its simplicity to implement. The obvious drawback is that the method cannot be used to estimate long-term effects, but we recommend that this method be used alongside other methods to check whether the more complex methods are in agreement with the one-year effect estimate of the SCMM. In cases where the one-year effect estimate is markedly different between SCMM and another method, this could act as a flag of potential issues with the analysis.

Another benefit of SCMM is that the model-based SEs will be approximately correct when the propensity score is well estimated, meaning that the bootstrap does not need to be used and results can be obtained much faster than using the other methods presented in this paper. The asymptotic SEs have been derived for IPW of MSM, but only in a time-fixed setting,[95] and the difficulty of deriving these in a longitudinal setting necessitates the use of the bootstrap for all the methods other than SCMM.

G-formula tended to perform as well as other methods, only performing poorly where other methods also performed poorly. The SEs were consistently smaller than for other methods, which would always be preferable in cases where we are confident in the specified models for the time-varying covariates. However, in cases where there is misspecification of these models the SEs remain small, and with real data it is unlikely that all the assumptions necessary for g-formula would be completely correct, which could result in a tight CI around an incorrect effect estimate. In our scenarios, we did not encounter any issue with the g-null paradox. This is because, for the g-null paradox to arise it is necessary for treatment to affect a time-dependent confounder without having any direct or indirect effect on the outcome.[103] In our 'no effect scenario' treatment had no effect on either lung function or IV days which are acting as both the outcome and the time-dependent confounders.

For continuous outcomes, g-estimation performed well with the SEs generally lying between those of g-formula and IPW, with the advantage that the method can also estimate effect modification by time-varying covariates, without the drawbacks of unstable weights which were sometimes observed in HA-MSM. However with the count outcome g-estimation used a gamma model rather than the zero-inflated negative binomial model like the other methods presented in this paper. This made comparison with other methods difficult. In situations where the count outcome is not as skewed as the annual IV days in the UK CF registry data, g-estimation may be a suitable method, but in our setting the other methods were generally preferable. A further complexity of the count outcome is the issue of non-collapsibility. In the simulations we found that when there is truly no treatment effect, both marginal and conditional estimates were correctly consistent with there being no treatment effect. However, in cases where there is a treatment effect, comparison between marginal and conditional estimates is not appropriate.

In most settings, more than one of the available methods would be suitable for the types of investigation considered in this paper. However, the results of the simulation studies suggest that IPW of MSM and HA-MSM will perform particularly poorly if the direction of the causal pathways is misspecified. This is a potential issue in the UK CF Registry due to that way that some data are collected. Therefore, even though we will still consider using these methods in the next chapter, it would not be unexpected if the results from these two methods are inconsistent with the results from other methods.

In conclusion, in many cases it may be beneficial to consider using more than one available method, to see if the results are consistent. Of course, in cases where two methods give the same effect estimate, this does not mean it is correct, but does add some reliability to the results. In cases where the methods gave very different effect estimates, this would act as a flag to re-examine the data, the assumptions of the methods and the suitability of the analyses performed.

Chapter 6

Estimating the Effects of Long-Term Dornase Alfa Use

6.1 Introduction

In Chapter 3 we performed a systematic review of studies that have investigated the effects of DNase. In a phase III clinical trial, using DNase over 24 weeks was shown to improve ppFEV$_1$ and reduce the risk of exacerbations.[49] Subsequently, a number of studies have examined the effect of longer-term use of DNase. For example, after 96 weeks of follow-up, DNase was shown to statistically significantly improve lung function and reduce the risk of respiratory tract infections in children aged 6 to 10.[61, 63]

Most studies have focused on the absolute effect of DNase on lung function over a specified time period, i.e. on a step-change effect. Its impact on the rate of lung function decline is also important and this has been investigated in two studies. These studies showed that during DNase use the rate of lung function decline was less than the rate of decline in the same patients prior to starting treatment and also less than the rate of decline in a comparator group of patients who never received treatment.[64, 75]

Only one study has attempted to evaluate the impact of using DNase for more than two years. This was a study with 76 patients and it was found that those receiving treatment over four years had a more gradual slope of decline in lung function and also suffered fewer exacerbations per year.[60]

In this chapter, we use the UK CF Registry to investigate the effects of one-, two-, three-, four- and five-years of DNase use on lung function and annual number of IV days. Long-term effects of treatment are particularly important for lung function as the hope is that treatments can modify the overall lung-function trajectory rather than just providing a one-off increase.[69, 104] Figure 5.2 in Chapter 5 shows two hypothetical lung-function trajectories for an individual treated with DNase compared to the lung-function trajectory of someone not receiving treatment. These trajectories show how treatment could

either be disease modifying, where the lung function of those receiving treatment continues to grow wider apart from the lung function of those not receiving treatment, or alternatively how the overall lung-function trajectory could remain unchanged after an initial increase.

Furthermore, we aimed to investigate whether there is evidence that treatment is more effective in younger people, as has previously been reported.[64] We also hypothesised that the effect of DNase on lung function may differ depending on lung function, and that its effect on IV days may differ depending on previous number of IV days.

For the main analysis, we have the choice of the five methods that were introduced in Chapter 4: IPW of MSM, HA-MSM, g-formula, g-estimation and SCMM. The simulation studies from Chapter 5 suggest that any of the methods could be used with little difference between the methods if all assumptions and models are correct. SCMM can only be used in our setting to estimate the effect of using DNase for one-year. However, it is the simplest method and can be supplemented by a test for evidence of any longer-term effects. All of the other methods can be used to estimate long-term effects, but in the case of IPW of MSM and g-formula they can only be used to estimate population-average effects, and therefore these methods cannot be used when we wish to investigate effect modification by a time-varying covariate. In this chapter, we will present the results from all available methods, as any major differences between the results from different methods would suggest potential violation of the assumptions made and inform subsequent investigations.

The work from this chapter formed the basis of a paper published in the Journal of Cystic Fibrosis in August 2018. A copy of this paper can be found in Appendix D.

6.2 Methods

6.2.1 UK Cystic Fibrosis Registry

The UK CF Registry was introduced in Chapter 2. For this analysis, we use annual review data from 2007 to 2015. At each annual assessment, the treatment section records whether a person has been receiving DNase since the last annual assessment. Although there is space to record a start date, this is not reliably collected, and for this analysis we assume that they have been receiving DNase since their last assessment.

People were eligible for inclusion in the study if they had at least two consecutive years of data in the registry between 2007 and 2015, had not received DNase prior to 2007, were not receiving DNase at their first visit, were at least six years old, had lung function data for at least two consecutive visits and had not received a lung transplant. The first

visit for everyone was therefore at a time when they were not and had never before been recorded to have received DNase treatment. Follow-up data were collected up to 2015 if available, or the most recent annual assessment otherwise. Figure 6.1 shows a flow chart of the number of people included and excluded from the study population.

The primary outcome is change in lung function expressed as absolute change over time in ppFEV$_1$ calculated using the GLI equations.[16] The secondary outcome was change in rate of IV Days. Sensitivity analyses for lung function were also conducted with ppFVC and ppFEF$_{25-75}$. The results of the sensitivity analyses are included in Appendix E.

As well as previous measures of the outcomes and DNase, we also adjusted for the following time-varying covariates: other muco-active treatments, smoking status, CFRD, and pseudomonas infection; and the following non-time-varying variables: baseline age, gender, ethnicity and genotype class. These were identified from the data collected in the registry as variables that could affect both outcomes and the decision to initiate DNase.

6.2.2 Statistical Methods

In a preliminary exploration of the data, baseline covariates, i.e. those measured at the first visit, were summarised overall and separately for those who never received DNase during subsequent follow-up and those who received DNase for at least one year. Baseline covariates were also summarised in the subgroups of those who received DNase for at least five years and people who stopped DNase during follow-up. Time-varying covariates were summarised at the first visit, and separately at visits when people were not receiving DNase, the first visit when people received treatment, subsequent visits among people receiving treatment and visits after people stopped treatment.

We start the main analysis by estimating the effect of one year of treatment using SCMM. As well as estimating the population average effect, we also allow modification of the treatment effect by baseline age or by previous lung function with both linear interaction terms and restricted cubic splines. Subsequent to this, we perform a test to assess the existence of longer-term effects. We then use IPW of MSM, HA-MSM, g-formula and g-estimation to estimate the effects of long-term DNase use. All four methods are used to estimate the population-average effects, and HA-MSM and g-estimation are also used to assess whether there was modification of the treatment effect.

The potential confounding variables listed above were included in all possible models, i.e. the same variables were included in the propensity score models, the censoring models and the final models for all methods. The only exception to this was in the case of g-formula, where due to the complexity of the method when there are many time-varying covariates, each of which would need a parametric model, the only time-varying confounders adjusted for were previous levels of lung function and IV days.

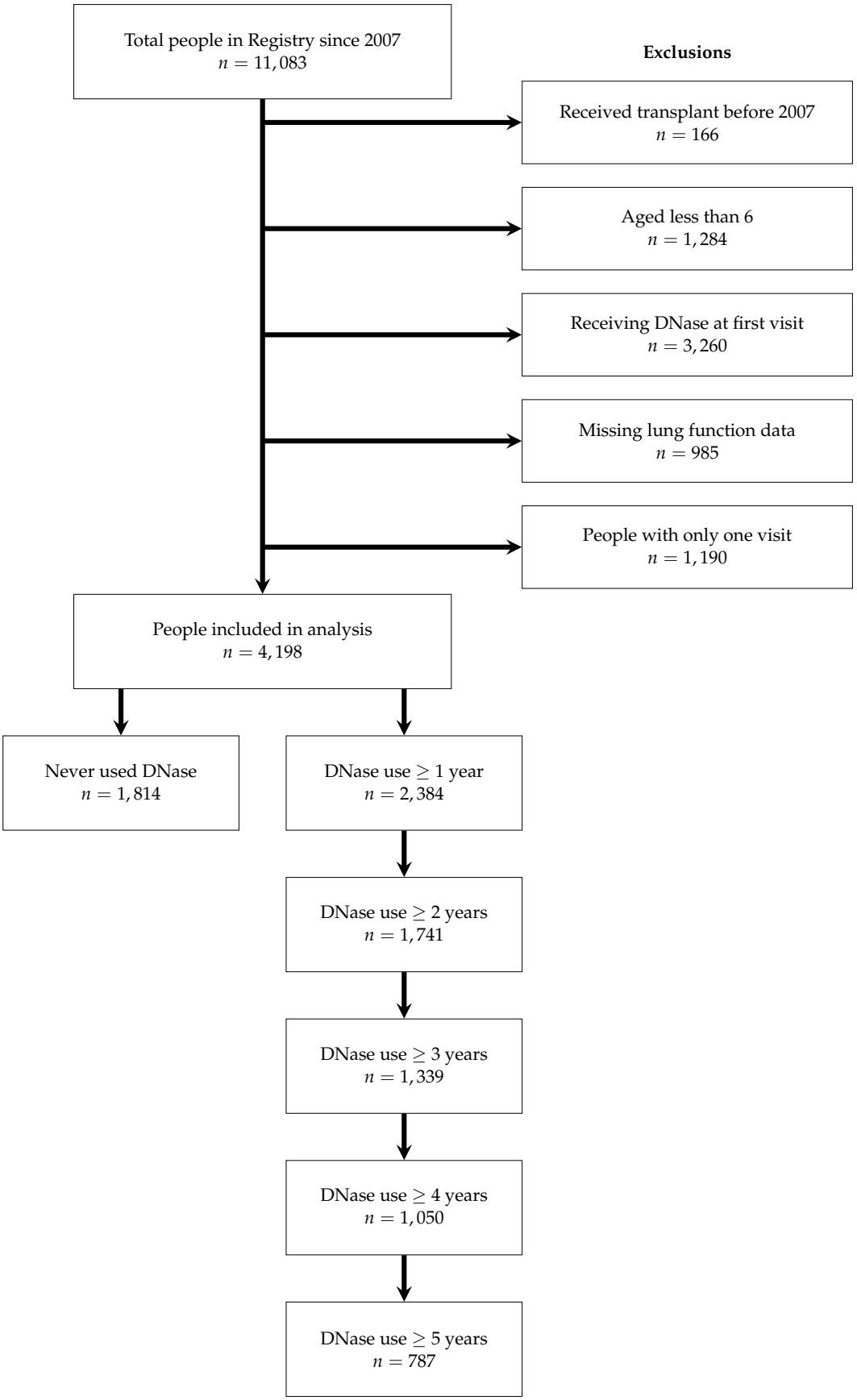


FIGURE 6.1: Flowchart of people included in DNase analysis

The majority of the additional covariates included in the models were binary: the use of other muco-active treatments, CFRD, pseudomonas infection, gender, and ethnicity. The last of which marking whether an individual was white or not. Genotype class is a variable with three levels: high, low, and not assigned. Smoking has four levels: none smoker, occasional smoker, smokes less than one packet per day, and smokes at least one packet per day. Some of these variables have missing data, and so as not to drop variables a missing category was also included for those variables. Previous DNase use was also adjusted for as a categorical variable equal to the sum of the total numbers of years that an individual has used treatment for prior to the current year.

The three continuous variables, baseline age, annual number of IV days and ppFEV$_1$, were included in the models using restricted cubic splines with four knots. However, for the models which considered interaction terms with these variables, we present models using linear terms to ease interpretability.

All analyses were performed using Stata (Version 15). SEs were estimated using a non-parametric bootstrap with 5000 resamples for all methods except for SCMM, where the model-based SEs were used.[105]

The following five subsections give further details of how each method was implemented for the continuous outcome, ppFEV$_1$. For IV days, the same methods were applied with the alterations for count outcomes as detailed in Chapter 4. We refer to DNase treatment as a binary variable, D, lung function as a continuous variable, L, IV days as a time-varying confounder, V, and age as a baseline confounder, A. We do not include other time-varying and baseline confounders in the equations below for conciseness, but they are all incorporated into the following models in the same way as V and A respectively. The rest of the notation is the same as detailed in Section 4.1.1.

Sequential Conditional Mean Models

First, we estimate the propensity score:

$$p_t = \mathrm{P}\Big(D_t = 1 \Big| D_{t-1}, L_{t-1}, V_t, A\Big) = \mathrm{expit}\Big(\gamma_0 + \gamma_d D_{t-1} + \gamma_l L_{t-1} + \gamma_v V_t + \gamma_a A\Big). \quad (6.1)$$

Then we fit the SCMM:

$$\begin{aligned} \mathrm{E}\Big(L_t \Big| \overline{D}_t, \overline{L}_{t-1}, \overline{V}_t, A, p_t\Big) = \\ \beta_0 + \sum_{j=1}^{t-1} \beta_{d_j} D_j + \beta_l L_{t-1} + \beta_v V_t + \beta_a A + \beta_p p_t + \phi_1 D_t. \end{aligned} \quad (6.2)$$

We also fit additional models to investigate effect modification. In these cases, the desired interaction terms are included with treatment and also with the propensity score. The first model investigates effect modification by baseline age:

$$\begin{aligned}&\mathrm{E}\Big(L_t\Big|\overline{D}_t,\overline{L}_{t-1},\overline{V}_t,A,p_t\Big)=\\&\beta_0+\sum_{j=1}^{t-1}\beta_{d_j}D_j+\beta_l L_{t-1}+\beta_v V_t+\beta_a A+\beta_p p_t+\beta_{i_2}p_t A+\phi_1 D_t+\phi_{int_a}D_t A.\end{aligned}\tag{6.3}$$

The second model investigates effect modification by previous lung function:

$$\begin{aligned}&\mathrm{E}\Big(L_t\Big|\overline{D}_t,\overline{L}_{t-1},\overline{V}_t,A,p_t\Big)=\\&\beta_0+\sum_{j=1}^{t-1}\beta_{d_j}D_j+\beta_l L_{t-1}+\beta_v V_t+\beta_a A+\beta_p p_t+\beta_{i_2}p_t L_{t-1}+\phi_1 D_t+\phi_{\mathrm{int}_1}D_t L_{t-1}.\end{aligned}\tag{6.4}$$

Finally, we test for evidence of longer-term effects by obtaining the predicted counterfactual, $\hat{L}_t^{d_t=0}$, as the fitted value from Equation 6.2 with D_t set to 0. We then fit the following model to see if there is any residual associated between these counterfactuals and previous treatment:

$$\begin{aligned}&\mathrm{E}\Big(D_{t-1}\Big|D_{t-2},L_{t-1},V_{t-1},A,\hat{L}_t^{d_t=0}\Big)=\\&\qquad\mathrm{expit}\Big(\gamma_0+\gamma_{\hat{l}}\hat{L}_t^{d_t=0}+\gamma_A A+\sum_{j=0}^{t-2}\gamma_d D_j+\gamma_l L_j+\gamma_V V_{j+1}\Big).\end{aligned}\tag{6.5}$$

Inverse Probability Weighting of Marginal Structural Models

We calculate the stabilised weights:

$$\mathrm{SW}_t=\frac{\mathrm{P}\Big(\overline{D}_t\Big|\overline{D}_{t-1},A\Big)}{\mathrm{P}\Big(\overline{D}_t\Big|\overline{D}_{t-1},\overline{L}_{t-1},\overline{V}_t,A\Big)}=\prod_{i=1}^{t}\frac{\mathrm{P}\Big(D_i\Big|\overline{D}_{i-1},A\Big)}{\mathrm{P}\Big(D_i\Big|\overline{D}_{i-1},\overline{L}_{i-1},\overline{V}_i,A\Big)},\tag{6.6}$$

where the denominator is estimated as given in Equation 6.1 and the numerator similarly estimated by removing Y_{t-1} and V_t from the model.

We also incorporate censoring weights, which are calculated as follows, and then multiplied by the stabilised weights to create the weights to be used for the MSM.

$$\mathrm{CW}_t=\frac{\mathrm{P}\Big(\overline{C}_t=0\Big|A\Big)}{\mathrm{P}\Big(\overline{C}_t=0\Big|\overline{D}_{t-1},\overline{L}_{t-1},\overline{V}_{t-1},A\Big)}=\prod_{i=1}^{t}\frac{\mathrm{P}\Big(C_i\Big|A\Big)}{\mathrm{P}\Big(C_i\Big|D_{i-1},L_{i-1},V_{i-1},A\Big)}.\tag{6.7}$$

Population average effects are then estimated by fitting the following MSM:

$$\mathrm{E}\left(L_t^{\bar{d}_t}\middle|A\right) = \beta_0 + \beta_a A + \sum_{j=1}^{t} \phi_j D_j. \tag{6.8}$$

This method is used with both non-truncated and truncated weights, where in the latter case the weights are truncated to the 1 and 99 percentile.

History-Adjusted Marginal Structural Models

The stabilised weights for HA-MSM are calculated similarly to those given above, except that the multiplication of the weights is not done from visit 1, but rather for time from the visit directly after exposure, $t+1$, up to and including the time of the outcome, s.

$$SW_{ts} = \prod_{i=t+1}^{s} \frac{\mathrm{P}\left(D_i\middle|\overline{D}_{i-1}, A\right)}{\mathrm{P}\left(D_i\middle|\overline{D}_{i-1}, \overline{L}_{i-1}, \overline{V}_i, A\right)}, \tag{6.9}$$

where $s - t + 1 \leq 5$.

The censoring weights are also calculated equivalently:

$$\mathrm{CW}_{ts} = \prod_{i=t+1}^{s} \frac{\mathrm{P}\left(C_i\middle|A\right)}{\mathrm{P}\left(C_i\middle|D_{i-1}, L_{i-1}, V_{i-1}, A\right)}. \tag{6.10}$$

The HA-MSM, given below, is then weighted by the product of the stabilised and censoring weights:

$$\mathrm{E}\left(L_s^{\bar{d}_s}\middle|\overline{d}_{t-1}, \overline{l}_{t-1}, \overline{v}_t, a\right) = \beta_0 + \beta_a a + \beta_v v_t + \sum_{j=1}^{t-1} \beta_{d_j} D_j + \beta_l l_{t-1} + \sum_{j=t}^{s} \phi_j d_j. \tag{6.11}$$

The interaction effect of interest can also be incorporated into the above equation, e.g. by adding

$$\sum_{j=t}^{s} \phi_{\mathrm{int}_j} d_j l_{t-1}$$

to the model given above.

Similarly to IPW of MSM, this method is used with both non-truncated and truncated weights, where in the latter case the weights are truncated to the 1 and 99 percentile.

G-Computation Formula

For g-formula, a parametric model needs to be specified for all time-varying covariates. Because of this complication, unlike with the other methods, where we included a number of time-varying covariates, in g-formula, we only consider the two most important time-varying covariates, lung function and annual IV days. These two variables are modelled as follows:

$$\begin{aligned}\mathrm{E}\Big(V_t\Big|\overline{D}_{t-1},\overline{L}_{t-1},\overline{V}_{t-1},A\Big) = \\ \mathrm{expit}\Big(\gamma_0+\gamma_{d_1}D_{t-1}+\sum_{j=1}^{t-2}\gamma_{d_j}D_j+\gamma_l L_{t-1}+\gamma_V V_{t-1}+\gamma_a A\Big) \\ \exp\Big(\delta_0+\delta_{d_1}D_{t-1}+\sum_{j=1}^{t-2}\delta_{d_j}D_j+\delta_l L_{t-1}+\delta_V V_{t-1}+\delta_a A\Big),\end{aligned} \tag{6.12}$$

$$\mathrm{E}\Big(L_t\Big|\overline{D}_t,\overline{L}_{t-1},\overline{V}_t,A\Big) = \beta_0+\beta_{d_1}D_t+\sum_{j=1}^{t-1}\beta_{d_j}D_j+\beta_l L_{t-1}+\beta_V V_t+\beta_a A. \tag{6.13}$$

For every individual, we simulate all the monotonic trajectories, whereby people can only start treatment and, once initiated, treatment is continued until the end of follow-up. The simulated counterfactual outcomes from the different treatment trajectories are then compared with a MSM:

$$\mathrm{E}\Big(L_t^{\bar{d}_t}\Big) = \beta_0+\sum_{i=1}^{t}\phi_i d_i. \tag{6.14}$$

G-Estimation of Structural Nested Models

We estimate the propensity score as in Equation 6.1 and the censoring weights as in Equation 6.10. Then we estimate the one-year effect:

$$\begin{aligned}\mathrm{E}\Big(L_s\Big|\overline{D}_s,\overline{L}_{s-1},\overline{V}_s,A,p_s\Big) = \\ \beta_s+\beta_{d_{s-1}}D_{s-1}+\beta_{l_{s-1}}L_{s-1}+\boldsymbol{\beta}_{V_s}V_s+\beta_a a+\beta_{p_s}p_s+\phi_1 D_s.\end{aligned} \tag{6.15}$$

The following two steps are then iterated four times to sequentially estimate the counterfactuals of what the outcome would have been if people had not received treatment between visits t and s, where $s-t+1\leq 5$:

$$H_{sj} = L_s - \sum_{u=j+1}^{s}\phi_{s-u+1}D_u, \tag{6.16}$$

$$\begin{aligned} &\mathrm{E}\left(H_{sj}\middle|\overline{D}_j, \overline{L}_{j-1}, \overline{V}_j, A, p_j\right) = \\ &\qquad \sum_{i=t}^{s} z\left(\beta_i + \beta_{d_i} D_{j-1} + \beta_{l_i} L_{j-1} + \beta_{v_i} V_j + \beta_{a_i} A + \beta_{p_i} p_j + \phi_{s-i+1} D_j\right), \end{aligned} \tag{6.17}$$

where $z = 1$ if $i = j$ and 0 otherwise.

Interaction terms can also be incorporated into Equations 6.15, 6.16 and 6.17. When a time-varying interaction term is included, then the cumulative treatment effect can be estimated by following the procedure given in Section 4.5.1.

6.3 Results

Overall, 4,198 people were included in the analysis, with a combined total of 26,363 annual assessments. The median number of follow-up visits per person was 6 (IQR 4-8). During follow-up, 2,384 (56.8%) people received DNase for at least one year, and most people who started using DNase continued to receive it indefinitely, with only 441 (18.5%) people stopping during follow-up. In total, 821 (34.5%) people had five or more consecutive years of DNase use.

Table 6.1 summarises the variables that were considered as confounders for the main analysis. When looking at this summary of the data, we found that those taking DNase tend to have worse disease. This is expected, because those needing treatment are expected to be in worse health. This finding is part of the phenomenon of confounding by indication. Particularly, the mean ppFEV$_1$ measured prior to visits when people were not receiving DNase was 80% compared to 72% prior to the visit when people started to receive DNase. Furthermore, the group who ever received DNase had a higher proportion of people with a high genotype class compared with those who never received DNase (75% vs 57%), a higher proportion with CFRD (23% vs 18%) and had more annual IV days (mean 19 vs 9). Table 6.1 also summarises the group of people who stopped taking DNase during follow-up, but there were no noticeable differences between this group of people and those who continued to take DNase throughout follow-up.

6.3.1 Lung Function

One-Year Treatment Effect

The SCMM estimated that the effect of using DNase for one year was a -0.07% absolute change in ppFEV$_1$ (95% CI -0.47, 0.33, $p = 0.73$). There was no evidence that the treatment effect was modified by baseline age: linear interaction $p = 0.93$, spline interaction

Demographics (4,198 people)	**All (n=4,198)**	**Never used DNase (n=1,814)**	**DNase use $\geq$ 1 year (n=2,384)**	**DNase use $\geq$ 5 year (n=821)**	**Stopped DNase during follow-up (n=441)**
Baseline Age (years)	19.8 (12.7)	21.4 (13.5)	18.7 (11.9)	19.2 (11.1)	21.0 (11.1)
Female	1,923 (45.8)	806 (44.4)	1,117 (46.9)	380 (46.3)	210 (47.6)
Caucasian	4,061 (96.7)	1,761 (97.1)	2,300 (96.5)	792 (96.5)	432 (98.0)
Genotype Class					
High	2,828 (67.4)	1,031 (56.8)	1,797 (75.4)	659 (80.3)	333 (75.5)
Low	563 (13.4)	353 (19.5)	210 (8.8)	52 (6.3)	35 (7.9)
None Assigned	153 (3.6)	77 (4.2)	76 (3.2)	28 (3.4)	17 (3.9)
Missing	654 (15.6)	353 (19.5)	301 (12.6)	82 (10.0)	56 (12.7)
Longitudinal Data (26,363 observations)	**First Visit [no DNase use] (n=4,198)**	**Other visits where never used DNase (n=12,194)**	**First year using DNase (n=2,384)**	**Subsequent years using DNase (n=6,633)**	**Years after stopping DNase (n=954)**
Previous ppFEV$_1$	***	80.1 (20.9)	72.4 (21.3)	68.4 (21.5)	67.8 (21.5)
Annual Change in ppFEV$_1$	***	-1.0 (9.9)	-0.4 (12.5)	-1.6 (9.5)	-1.4 (9.1)
Previous ppFVC*	***	90.1 (17.1)	84.8 (18.2)	82.4 (18.1)	82.4 (18.1)
Annual Change in ppFVC*	***	-0.7 (10.1)	-0.3 (12.1)	-1.2 (10.1)	-0.7 (9.8)
Previous ppFEF$_{25-75}$**	***	71.4 (31.2)	63.7 (28.9)	59.2 (26.7)	62.6 (34.3)
Annual Change in ppFEF$_{25-75}$**	***	-1.3 (21.7)	0.1 (26.7)	-1.1 (19.4)	-3.5 (17.9)
Annual IV Days	8.2 (17.1)	9.2 (19.4)	19.4 (27.5)	23.1 (31.6)	21.2 (32.9)
Smoker					
No	3,058 (72.8)	10,567 (86.7)	2,138 (89.7)	5,949 (89.7)	845 (88.6)
Occasionally	33 (0.8)	126 (1.0)	19 (0.8)	53 (0.8)	12 (1.3)
<1 packet per day	49 (1.2)	256 (2.1)	24 (1.0)	59 (0.9)	21 (2.2)
$\geq$1 packet per day	20 (0.5)	69 (0.6)	6 (0.3)	18 (0.3)	7 (0.7)
Missing	1,038 (24.7)	1,176 (9.6)	197 (8.3)	554 (8.4)	69 (7.2)
Pseudomonas infection	2,453 (58.4)	6,460 (53.0)	947 (39.7)	2,336 (35.2)	321 (33.6)
CFRD					
No	2,226 (53.0)	7,455 (61.1)	1,360 (57.0)	3,418 (51.5)	483 (50.6)
Yes	504 (12.0)	2,150 (17.6)	550 (23.1)	2,088 (31.5)	297 (31.1)
Missing	1,468 (35.0)	2,589 (21.2)	474 (19.9)	1,127 (17.0)	174 (18.2)
Other Muco-Active Treatments	200 (4.8)	1,578 (12.9)	502 (21.1)	2,229 (33.6)	356 (37.3)
Acetylcysteine	15 (0.4)	133 (1.1)	41 (1.7)	112 (1.7)	11 (1.2)
Hypertonic Saline	184 (4.4)	1,440 (11.8)	469 (19.7)	2,095 (31.6)	340 (35.6)
Mannitol	2 (0.05)	39 (0.3)	13 (0.5)	145 (2.2)	14 (1.5)

TABLE 6.1: Descriptive statistics of key variables: Data presented as mean (SD) for continuous variables and n (%) for binary and categorical variables (*FVC data based on 26,086 observations, **FEF$_{25-75}$ data based on 4,452 observations, ***First visit, so no previous measures available)

$p = 0.77$. This is shown in Figure 6.2a where it can be seen that including an interaction term does not result in a statistically significantly different treatment effect than not including an interaction term. Conversely, there was strong evidence of modification of the treatment effect by previous ppFEV$_1$: linear interaction $p = 0.003$, spline interaction $p = 0.007$. These results are shown in Figure 6.2b. The model with an interaction including restricted cubic splines shows that the relationship is not entirely linear, but for ease of interpretability, we use the linear interaction term from now on. The results from the population average model and the two models with linear interaction terms are also shown in Table 6.2.

The interaction effect with previous ppFEV$_1$ suggests that treatment is most beneficial in people with lower lung function, with the effect estimated to decrease by 0.32% (95% CI 0.11, 0.53) for every 10% increase in ppFEV$_1$. This model estimates that treatment would improve lung function at one year if starting treatment with a ppFEV$_1$ less than 69.7%.

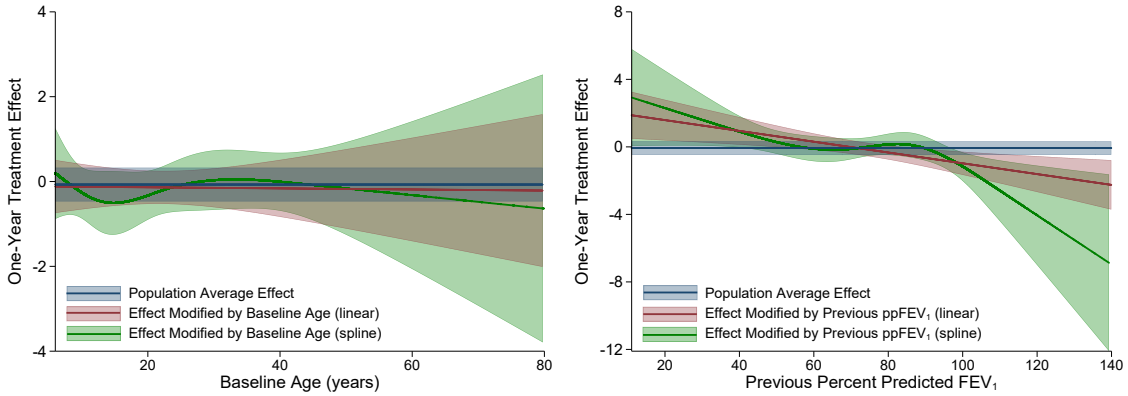
The test for longer-term effects was highly statistically significant ($p < 0.0001$) suggesting that the effect of DNase on lung function changes with increasing years of DNase use.

Long-Term Treatment Effect

Results from the methods that can estimate long-term effects are shown in Figure 6.3 and Table 6.3. G-formula and g-estimation gave similar one-year effect estimates as SCMM, but IPW of MSM and HA-MSM both estimated a statistically significant decrease in lung function after one year of treatment: IPW of MSM -2.24% (95% CI -3.31, -1.16, $p < 0.0001$), and HA-MSM -0.93 (95% CI -1.48, -0.39, $p < 0.001$). These two results were very similar even when the most extreme weights were truncated.

Looking at longer-term effects all four methods showed a trend to the decrease in lung function becoming larger over time. In the cases of HA-MSM and g-estimation this decrease was modest: at five years the estimated treatment effect for HA-MSM was -1.93% (95% CI -3.43, -0.42, $p = 0.012$) and g-estimation -3.18% (95% CI -4.48, -1.88, $p < 0.0001$). However, IPW of MSM and g-formula estimated a stronger negative effect: five-year effect using IPW of MSM -5.65% (95% CI -7.66, -3.63, $p < 0.0001$), using g-formula -5.54% (95% CI -6.72, -4.36, $p < 0.0001$).

HA-MSM and g-estimation gave similar results when considering effect modification by previous ppFEV$_1$. These results are shown in Figure 6.4 and Table 6.4. For example, for an individual starting DNase with a ppFEV$_1$ of 20%, after one year their lung function was estimated to increase by 1.59% (95% CI 0.38, 2.79, $p = 0.010$) using g-estimation and to increase by 1.54% (95% CI -0.09, 3.16, $p = 0.064$) using HA-MSM. These estimates remained more or less stable out to five years, although with much larger CIs: at five years, g-estimation estimate 2.98% (95% CI -1.46, 7.42, $p = 0.19$), HA-MSM estimate 2.79% (95%



(A) (B)

FIGURE 6.2: Estimated one-year effect of DNase on ppFEV$_1$ depending on baseline age or previous ppFEV$_1$

Method		1 Year Treatment Effect		
		Est.	95% CI	p
Population Average		-0.071	-0.47, 0.33	0.73
Modified by Age	Intercept	-0.11	-0.89, 0.67	0.78
	Interaction	-0.013	-0.32, 0.29	0.93
Modified by Previous ppFEV$_1$	Intercept	2.23	0.61, 3.84	0.007
	Interaction	-0.32	-0.53, -0.11	0.003

TABLE 6.2: Estimated one-year effects of DNase on ppFEV$_1$ (The intercept term is the estimated effect if the age or previous ppFEV$_1$ is 0, and the interaction term is the change in effect per 10 increase in the modifier)

CI -1.64, 7.21, $p = 0.22$). For individuals starting DNase with a ppFEV$_1$ of 100% the results were again similar between the two methods, with a negative effect at one year of -0.97% (95% CI -1.65, -0.29, $p = 0.005$) for g-estimation and -2.10% (95% CI -2.87, -1.33, $p < 0.0001$) for HA-MSM. This negative effect became slightly more pronounced at five years: g-estimation estimate -5.75% (95% CI -7.82, -3.69, $p < 0.0001$), HA-MSM estimate -4.36% (95% CI -6.65, -2.08, $p < 0.001$). Overall, the results for HA-MSM with truncated weights were very similar to the results without truncation of weights shown above.

Including the interaction with previous ppFEV$_1$, it was estimated that DNase would improve lung function over one-year for anyone starting treatment with a ppFEV$_1$ less than 69.6% (g-estimation) or less than 53.8% (HA-MSM). At five years, treatment was estimated to be improve lung function in anyone who had started treatment with a ppFEV$_1$ less than 47.3% (g-estimation) or 51.2% (HA-MSM).

We also repeated the same analyses using ppFVC and ppFEF$_{25-75}$ as outcomes and obtained broadly similar results. The results from these analyses can be seen in Appendix E in Figures E.1 to E.6 and Tables E.1 to E.6.

6.3.2 Annual IV Days

One-Year Treatment Effect

Using SCMM, one year of DNase use was estimated to decrease an individual's odds of having zero IV days by 20% (95% CI 0.72, 0.88, $p < 0.0001$) and increase the overall rate of IV days by 8% (95% CI 1.04, 1.12, $p < 0.0001$). However, there was estimated to be a statistically significant change in the effect of DNase by both baseline age ($p = 0.009$) and previous number of IV days ($p = 0.003$), and both these interaction terms also remained statistically significant when included in the model together. This model estimated that for an individual aged 6 with zero IV days in the previous year, the odds of having zero IV days in the following year decreased by 40% if using DNase (95% CI 0.50, 0.70, $p < 0.0001$), and the overall rate of IV days increased by 12% (95% CI 1.05, 1.19, $p < 0.001$). For every 10 year increase in age, the odds of having zero IV days were estimated to increase by 14% (95% CI 1.04, 1.24, $p = 0.005$) and the rate of IV Days was estimated to decrease by 1% (95% CI 0.96, 1.02, $p = 0.46$). For every 14 day increase in the number of IV days in the previous year, the odds of having zero IV days were estimated to increase by 18% (95% CI 1.06, 1.30, $p = 0.001$) and the rate of IV days was estimated to decrease by 1% (95% CI 0.97, 1.01, $p = 0.21$). These results can be seen in Figures 6.5 and 6.6, and Table 6.5. In the figures, we again also include the results showing the interaction using restricted cubic splines as well as the linear interaction terms.

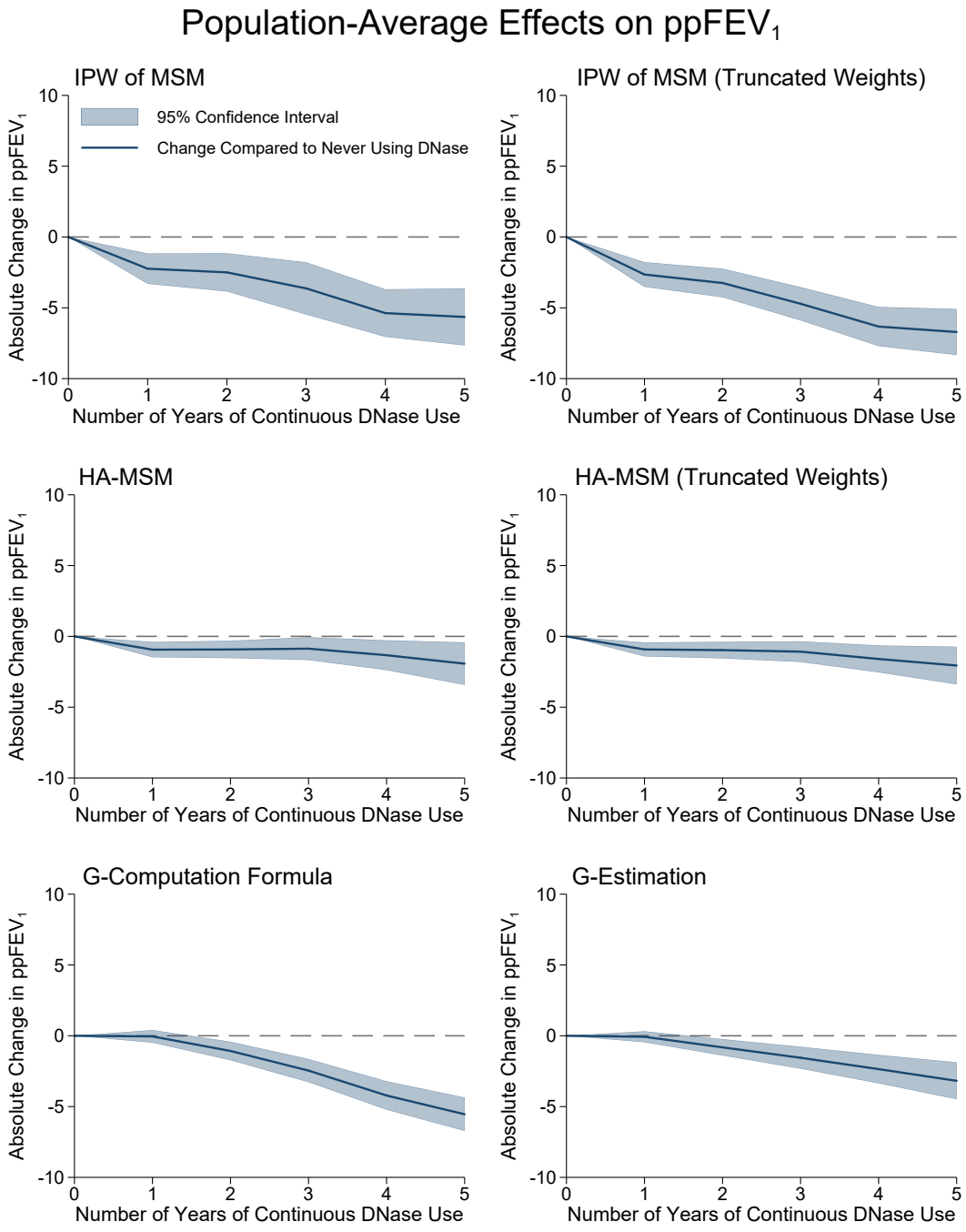


FIGURE 6.3: Estimated population-average effects of DNase on $ppFEV_1$

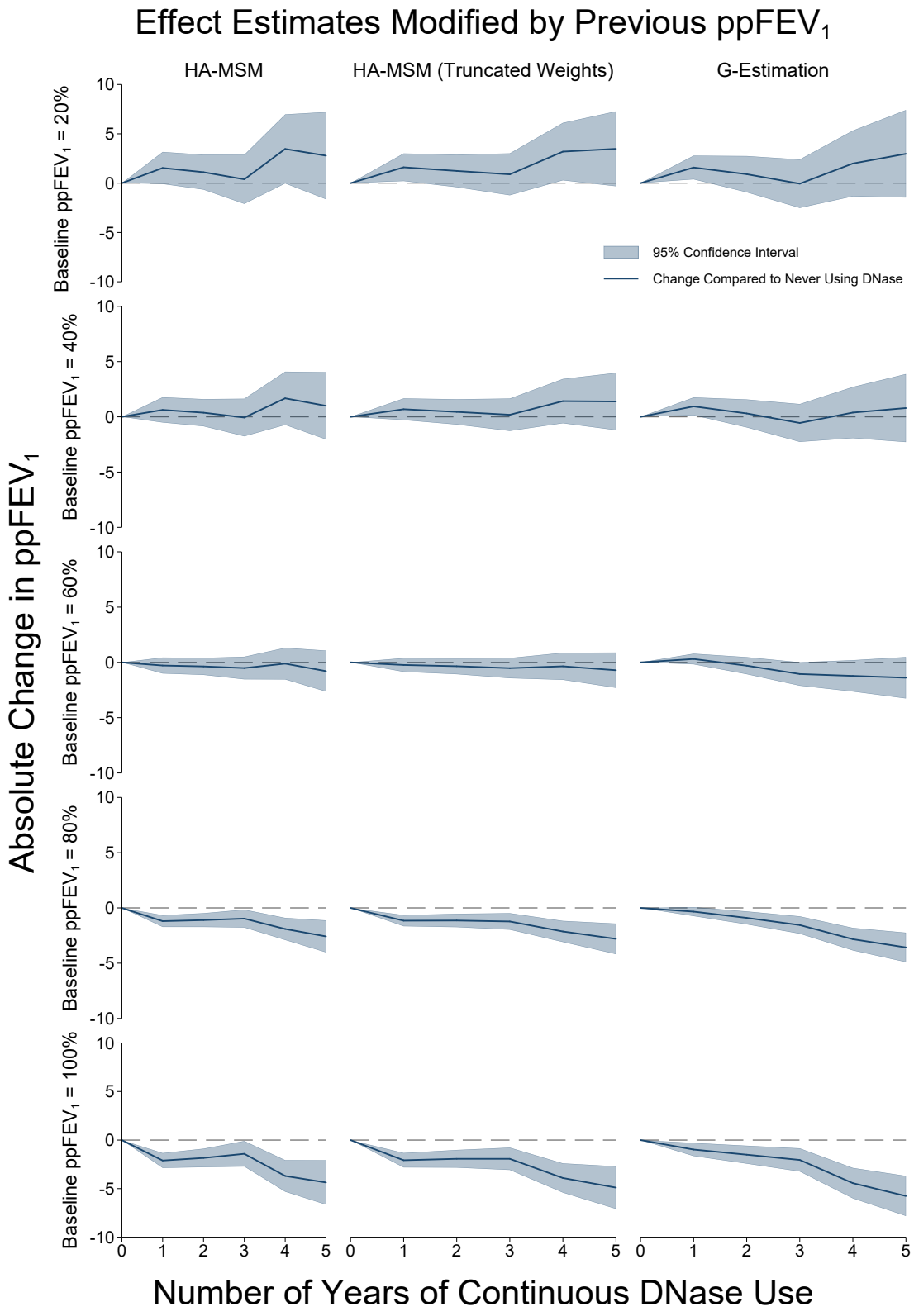


FIGURE 6.4: Estimated effects of DNase on $ppFEV_1$ modified by previous $ppFEV_1$

Method	1 Year Treatment Effect			Cumulative 2 Year Treatment Effect			Cumulative 3 Year Treatment Effect		
	Est.	95% CI	p	Est.	95% CI	p	Est.	95% CI	p
IPW of MSM	-2.24	-3.31, -1.16	<0.0001	-2.50	-3.85, -1.15	<0.001	-3.63	-5.74, -1.78	<0.001
IPW of MSM (truncated)	-2.65	-3.53, -1.77	<0.0001	-3.25	-4.26, -2.23	<0.0001	-4.71	-5.89, -3.54	<0.0001
HA-MSM	-0.93	-1.48, -0.39	<0.001	-0.92	-1.54, -0.31	0.003	-0.87	-1.68, -0.06	0.036
HA-MSM (truncated)	-0.92	-1.41, -0.43	<0.001	-0.97	-1.56, -0.38	0.001	-1.08	-1.81, -0.34	0.004
G-Formula	-0.045	-0.49, 0.40	0.85	-1.07	-1.73, -0.42	0.001	-2.46	-3.28, -1.64	<0.0001
G-Estimation	-0.071	-0.46, 0.32	0.72	-0.82	-1.40, -0.24	0.006	-1.55	-2.33, -0.78	<0.0001
Method	Cumulative 4 Year Treatment Effect			Cumulative 5 Year Treatment Effect					
	Est.	95% CI	p	Est.	95% CI	p			
IPW of MSM	-5.37	-7.05, -3.69	<0.0001	-5.65	-7.66, -3.63	<0.0001			
IPW of MSM (truncated)	-6.32	-7.71, -4.94	<0.0001	-6.71	-8.34, -5.08	<0.0001			
HA-MSM	-1.33	-2.39, -0.27	0.014	-1.93	-3.43, -0.42	0.012			
HA-MSM (truncated)	-1.59	-2.55, -0.64	0.001	-2.06	-3.39, -0.73	0.002			
G-Formula	-4.22	-5.22, -3.21	<0.0001	-5.54	-6.72, -4.36	<0.0001			
G-Estimation	-2.35	-3.36, -1.35	<0.0001	-3.18	-4.48, -1.88	<0.0001			

TABLE 6.3: Estimated cumulative population average effects of DNase on ppFEV$_1$

Method		1 Year Treatment Effect			Cumulative 2 Year Treatment Effect			Cumulative 3 Year Treatment Effect		
		Est.	95% CI	p	Est.	95% CI	p	Est.	95% CI	p
HA-MSM	Intercept	2.45	0.32, 4.58	0.024	1.85	-0.50, 4.20	0.12	0.84	-2.48, 4.15	0.62
	Interaction	-0.46	-0.72, -0.19	<0.001	-0.37	-0.67, -0.069	0.016	-0.23	-0.65, 0.20	0.30
HA-MSM (truncated)	Intercept	2.53	0.68, 4.37	0.007	2.03	-0.16, 4.21	0.069	1.60	-1.22, 4.41	0.27
	Interaction	-0.46	-0.69, -0.23	<0.001	-0.40	-0.68, -0.12	0.006	-0.35	-0.71, 0.008	0.056
G-estimation	Intercept	2.23	0.62, 3.84	0.007	1.51	-0.92, 3.95	0.22	0.44	-2.80, 3.68	0.79
	Interaction	-0.32	-0.53, -0.11	0.003	-0.30	-0.61, 0.007	0.056	-0.25	-0.65, 0.16	0.23
Method		Cumulative 4 Year Treatment Effect			Cumulative 5 Year Treatment Effect					
		Est.	95% CI	p	Est.	95% CI	p			
HA-MSM	Intercept	5.26	0.63, 9.89	0.026	4.57	-1.28, 10.43	0.13			
	Interaction	-0.89	-1.48, -0.32	0.003	-0.89	-1.64, -0.15	0.019			
HA-MSM (truncated)	Intercept	4.97	1.11, 8.84	0.012	5.57	0.51, 10.64	0.031			
	Interaction	-0.89	-1.38, -0.39	<0.001	-1.05	-1.71, -0.38	0.002			
G-estimation	Intercept	3.60	-0.80, 7.99	0.11	5.16	-0.68, 11.00	0.083			
	Interaction	-0.80	-1.35, -0.26	0.004	-1.09	-1.82, -0.36	0.003			

TABLE 6.4: Estimated cumulative effects of DNase on ppFEV$_1$ modified by previous ppFEV$_1$. The intercept term is the estimated effect if previous ppFEV$_1$ was 0%, and the interaction term is the change in effect per 10% increase in previous ppFEV$_1$.

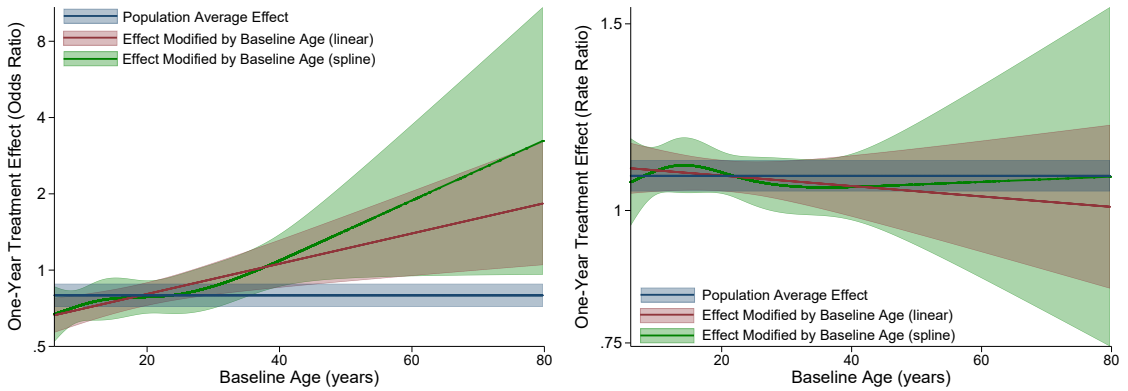
These results suggest that DNase would be least useful in young people with no previous IV days. For example, for a six year old, treatment would not be estimated to increase the odds of having zero IV days unless they had previously had at least 45 IV days, and would not decrease the overall rate of IV days unless they had had at least 148 previous IV days. However, by age 18, the number of previous IV days necessary for treatment to appear beneficial would only be 31 days for the odds of zero IV days and 130 days for the rate of IV days.

The test for long-term effects was not statistically significant, $p = 0.17$, suggesting that the estimated results as given above will remain the same no matter how long DNase is used for.

Long-Term Treatment Effect

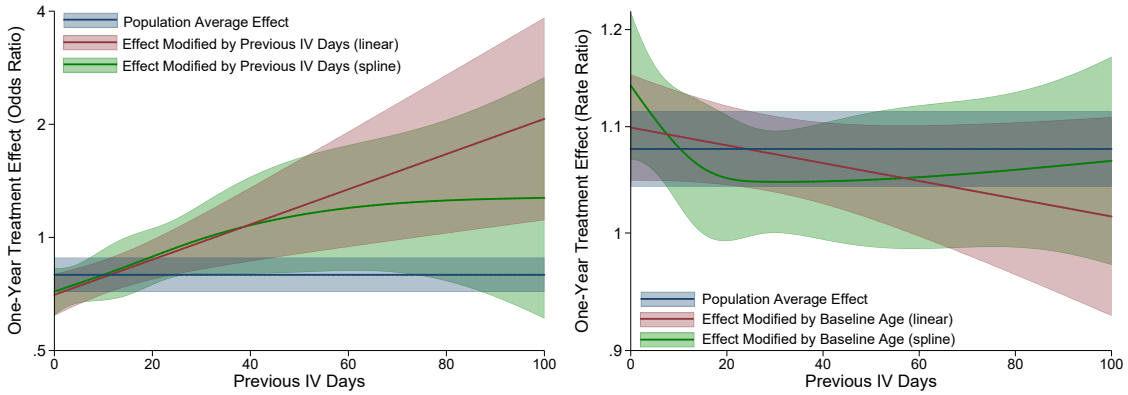
The results from IPW of MSM and HA-MSM concord with the test for long-term effects from the SCMM. Both methods show a decrease in the odds of zero IV days after one year: IPW of MSM OR 0.67 (95% CI 0.60, 0.75, $p < 0.0001$), HA-MSM OR 0.75 (95% CI 0.67, 0.84, $p < 0.0001$), and an increase in the rate of IV days: IPW of MSM IRR 1.12 (95% CI 1.03, 1.22, $p = 0.009$), HA-MSM IRR 1.12 (95% CI 1.07, 1.18, $p < 0.0001$). These estimates then remained stable with increased length of DNase use, for example after four years the OR of zero IV days compared to never using DNase was 0.66 (95% CI 0.55, 0.80, $p < 0.0001$) for IPW of MSM and 0.88 (95% CI 0.72, 1.07, $p = 0.20$) for HA-MSM), and the IRR of IV days was 1.04 (95% CI 0.90, 1.20, $p = 0.58$) for IPW of MSM and 1.16 (95% CI 1.06, 1.28, $p = 0.002$) for HA-MSM. As with the lung function outcome, truncation of the most extreme weights did not result in statistically significantly different results. These results can be seen in Figures 6.7 and 6.8 and Table 6.6. The results from g-formula and g-estimation can also been seen in the same figures and tables, but both of these methods estimated the effects of DNase on both the odds of zero IV days and the rate of IV days to become stronger through time. For g-formula, the OR of zero IV days was estimated to be 0.84 (95% CI 0.77, 0.91, $p < 0.0001$) at one year and changed to 0.66 (95% CI 0.57, 0.76, $p < 0.0001$) after four years. The IRR similarly changed from 1.04 (95% CI 0.99, 1.09, $p = 0.11$) at one year to 1.20 (95% CI 1.10, 1.32, $p < 0.0001$) after four years. For g-estimation, which only estimates an overall rate of IV days, the IRR was estimated to be 1.23 (95% CI 1.15, 1.33, $p < 0.0001$) at one year and 1.56 (95% CI 1.56, 1.80, $p < 0.0001$) after four years.

The results from methods which can include interaction terms are shown in Figures 6.9 to 6.13 and Table 6.7. As with the SCMM, there was strong evidence of an interaction between treatment and the previous number of IV days. In the figures, the number of previous IV days changes from 0 in the left column to 42 in the right column, and it can



(A) Effect on Odds of Zero IV Days (B) Effect on Rate of IV Days

FIGURE 6.5: Estimated one-year effect of DNase on IV Days depending on baseline age



(A) Effect on Odds of Zero IV Days (B) Effect on Rate of IV Days

FIGURE 6.6: Estimated one-year effect of DNase on IV Days depending on previous IV days

Method		1 Year Treatment Effect					
		OR	95% CI	p	IRR	95% CI	p
Population Average		0.80	0.72, 0.88	<0.0001	1.08	1.04, 1.12	<0.0001
Modified by Age	Intercept	0.61	0.50, 0.75	<0.0001	1.10	1.03, 1.18	0.006
	Interaction	1.14	1.05, 1.26	0.003	0.99	0.96, 1.02	0.45
Modified by Previous IV days	Intercept	0.70	0.62, 0.80	<0.0001	1.10	1.05, 1.15	<0.0001
	Interaction	1.16	1.06, 1.28	0.002	0.99	0.97, 1.00	0.18
Modified by Age & Previous IV days	Intercept	0.55	0.45, 0.68	<0.0001	1.12	1.04, 1.21	0.002
	Interaction by Age	1.14	1.04, 1.24	0.005	0.99	0.96, 1.02	0.46
	Interaction by Previous IV days	1.18	1.06, 1.30	0.001	0.99	0.97, 1.01	0.21

TABLE 6.5: Estimated one-year effects of DNase on odds of zero IV days (OR) and rate of total IV days (IRR) (The intercept term is the estimated effect if the age or previous IV days is 0, and the interaction term is the change in effect per 10 increase age or per 14 increase in IV days)

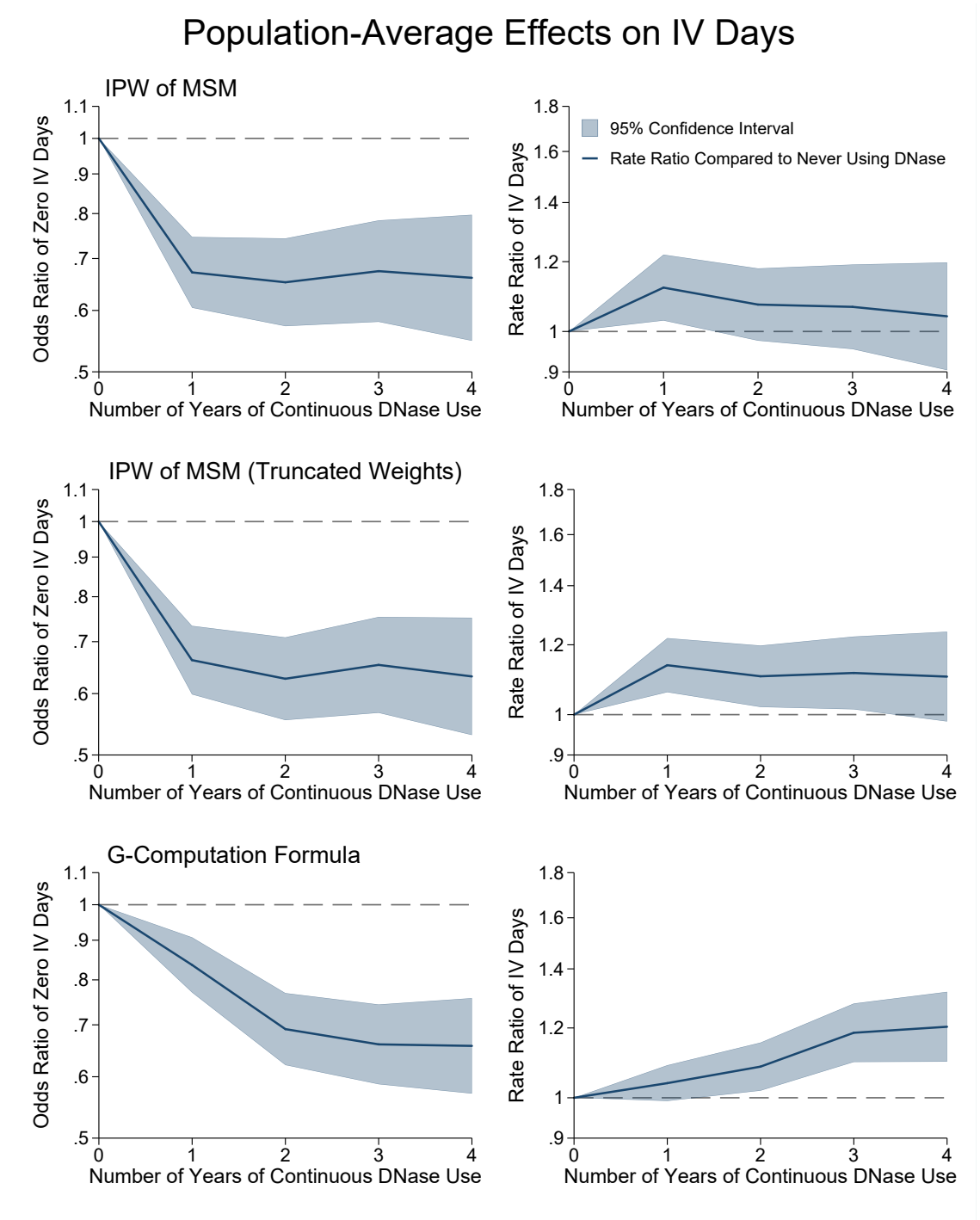


FIGURE 6.7: Estimated population-average effects of DNase on IV Days (1)

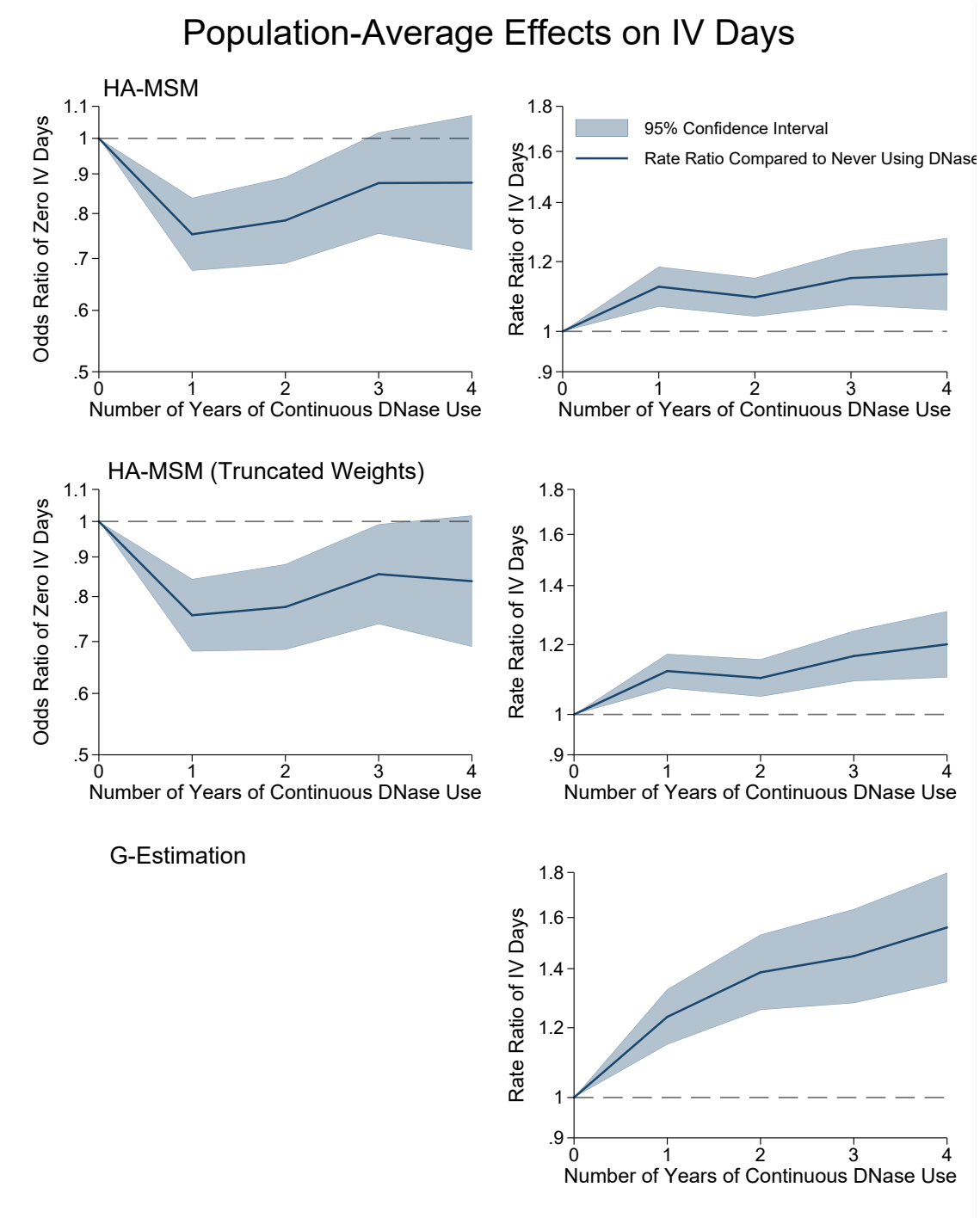


FIGURE 6.8: Estimated population-average effects of DNase on IV Days (2)

Method	1 Year Treatment Effect					
	OR	95% CI	p	IRR	95% CI	p
IPW of MSM	0.67	0.60, 0.75	<0.0001	1.12	1.03, 1.22	0.009
IPW of MSM (truncated)	0.66	0.60, 0.73	<0.0001	1.14	1.06, 1.22	<0.001
HA-MSM	0.75	0.67, 0.84	<0.0001	1.12	1.07, 1.18	<0.0001
HA-MSM (truncated)	0.76	0.68, 0.84	<0.0001	1.12	1.07, 1.17	<0.0001
G-formula	0.84	0.77, 0.91	<0.0001	1.04	0.99, 1.09	0.11
G-estimation	-	-	-	1.23	1.15, 1.33	<0.0001
Method	Cumulative 2 Year Treatment Effect					
	OR	95% CI	p	IRR	95% CI	p
IPW of MSM	0.65	0.57, 0.74	<0.0001	1.07	0.98, 1.18	0.14
IPW of MSM (truncated)	0.63	0.55, 0.71	<0.0001	1.11	1.02, 1.20	0.015
HA-MSM	0.78	0.69, 0.89	<0.001	1.09	1.04, 1.15	<0.001
HA-MSM (truncated)	0.78	0.68, 0.88	<0.0001	1.10	1.05, 1.15	<0.001
G-formula	0.69	0.62, 0.77	<0.0001	1.08	1.02, 1.16	0.011
G-estimation	-	-	-	1.39	1.26, 1.53	<0.0001
Method	Cumulative 3 Year Treatment Effect					
	OR	95% CI	p	IRR	95% CI	p
IPW of MSM	0.67	0.58, 0.78	<0.0001	1.07	0.95, 1.19	0.25
IPW of MSM (truncated)	0.65	0.57, 0.75	<0.0001	1.11	1.01, 1.23	0.025
HA-MSM	0.88	0.75, 1.02	0.083	1.15	1.07, 1.23	<0.001
HA-MSM (truncated)	0.86	0.74, 0.99	0.038	1.16	1.09, 1.24	<0.0001
G-formula	0.66	0.59, 0.74	<0.0001	1.18	1.10, 1.28	<0.0001
G-estimation	-	-	-	1.45	1.28, 1.64	<0.0001
Method	Cumulative 4 Year Treatment Effect					
	OR	95% CI	p	IRR	95% CI	p
IPW of MSM	0.66	0.55, 0.80	<0.0001	1.04	0.90, 1.20	0.58
IPW of MSM (truncated)	0.63	0.53, 0.75	<0.0001	1.10	0.98, 1.24	0.097
HA-MSM	0.88	0.72, 1.07	0.20	1.16	1.06, 1.28	0.002
HA-MSM (truncated)	0.84	0.69, 1.02	0.075	1.20	1.10, 1.31	<0.0001
G-formula	0.66	0.57, 0.76	<0.0001	1.20	1.10, 1.32	<0.0001
G-estimation	-	-	-	1.56	1.35, 1.80	<0.0001

TABLE 6.6: Estimated long-term population-average effects of DNase on odds of zero IV days (OR) and rate of total IV days (IRR)

be seen that the estimated effects all become weaker as the number of previous IV days increases. The interaction with baseline age was estimated to be less strong than had been estimated with SCMM, but there were still statistically significant changes in the estimates as baseline age increased. This can be seen in the rows of Figures 6.9 to 6.13, where increasing age appears to result in less negative treatment effect estimates. However, even in the case of an individual aged 30 with 42 previous IV days at baseline (the bottom-right graph of each figure), the effect estimates are still generally only suggesting no effect of treatment on IV days, and treatment is never estimated to be statistically significantly beneficial on either the odds of zero IV days or the overall rate of IV days.

As in the case of the population-average results, using truncated weights did not noticeably change the results from HA-MSM. The results from g-estimation once again estimated a much strong treatment effect on the rate of IV days than HA-MSM, and this is because g-estimation does not give a separate estimate for the odds of zero IV days.

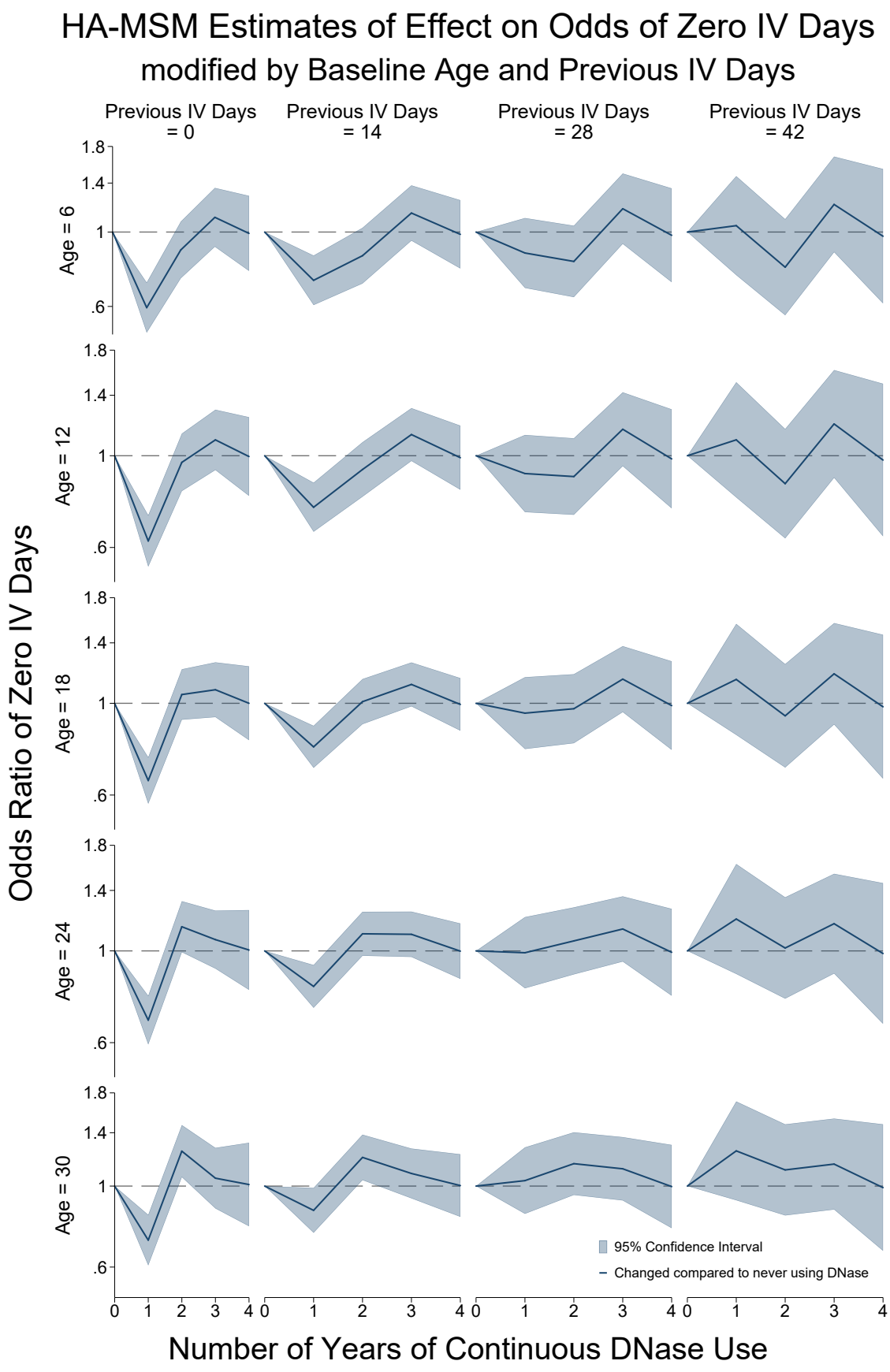


FIGURE 6.9: Estimated effects from HA-MSM of DNase on odds of zero IV days modified by baseline age and previous IV Days

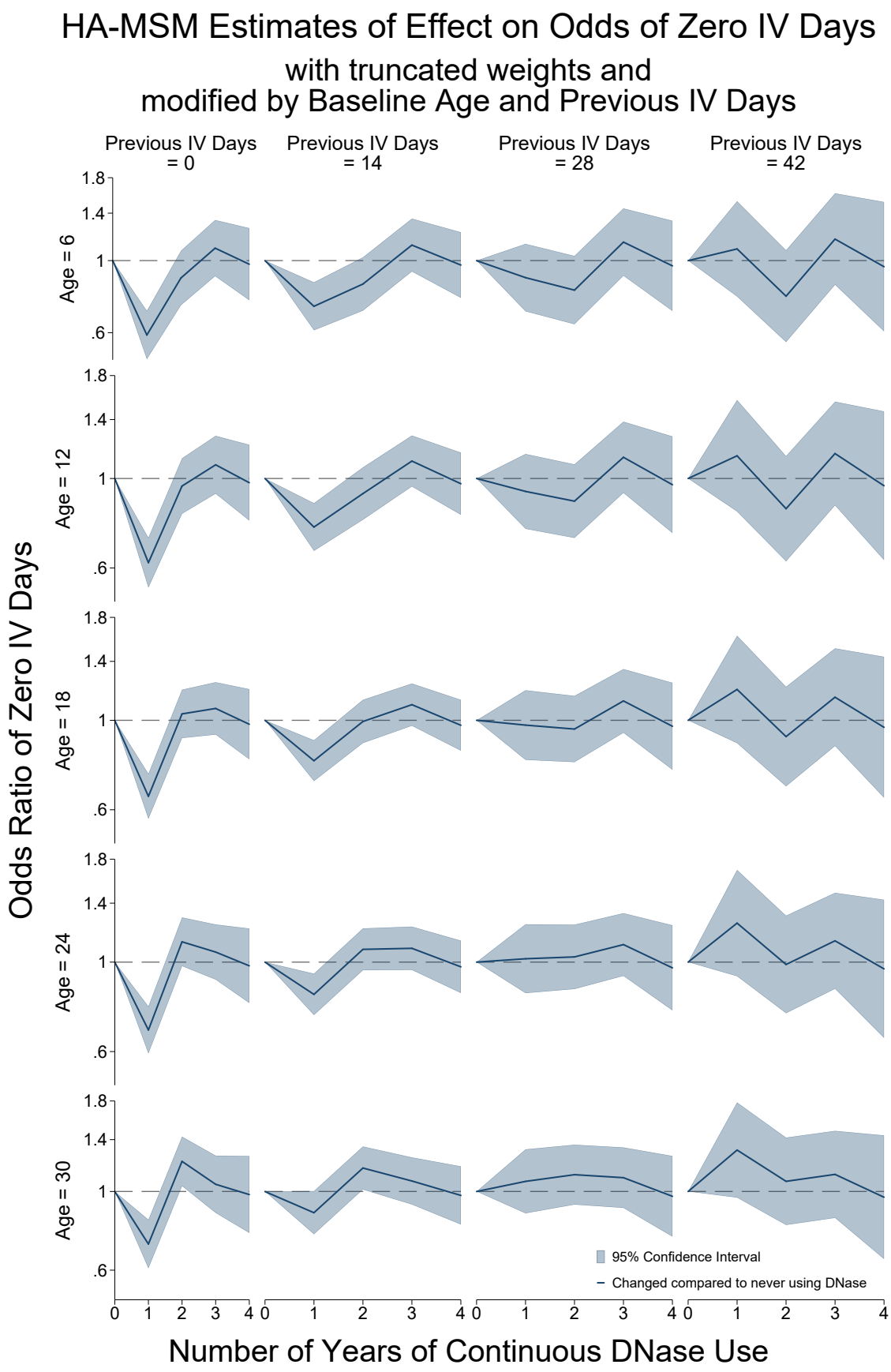


FIGURE 6.10: Estimated effects from HA-MSM with truncated weights of DNase on odds of zero IV days modified by baseline age and previous IV Days

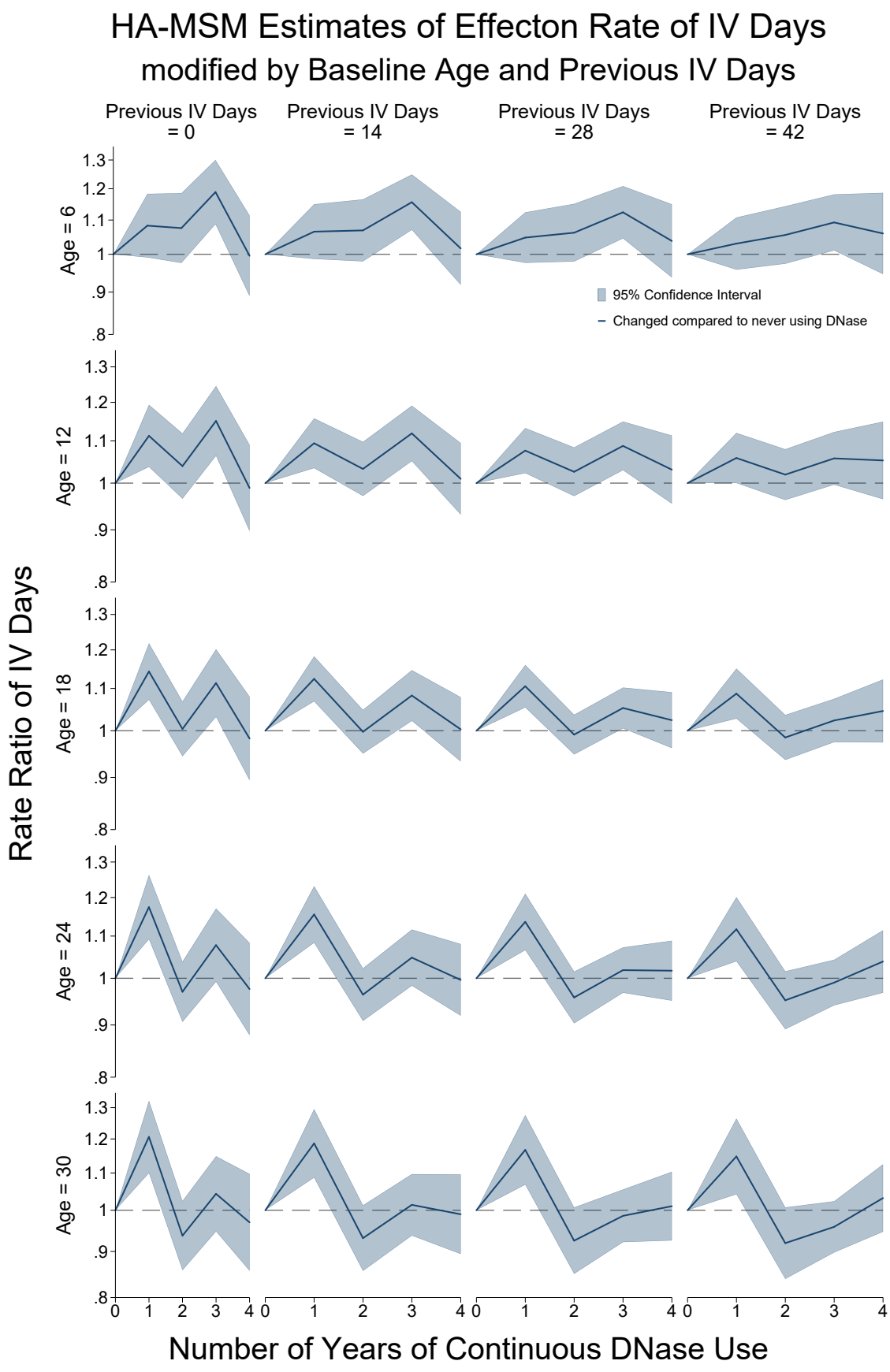


FIGURE 6.11: Estimated effects from HA-MSM of DNase on rate of IV days modified by baseline age and previous IV Days

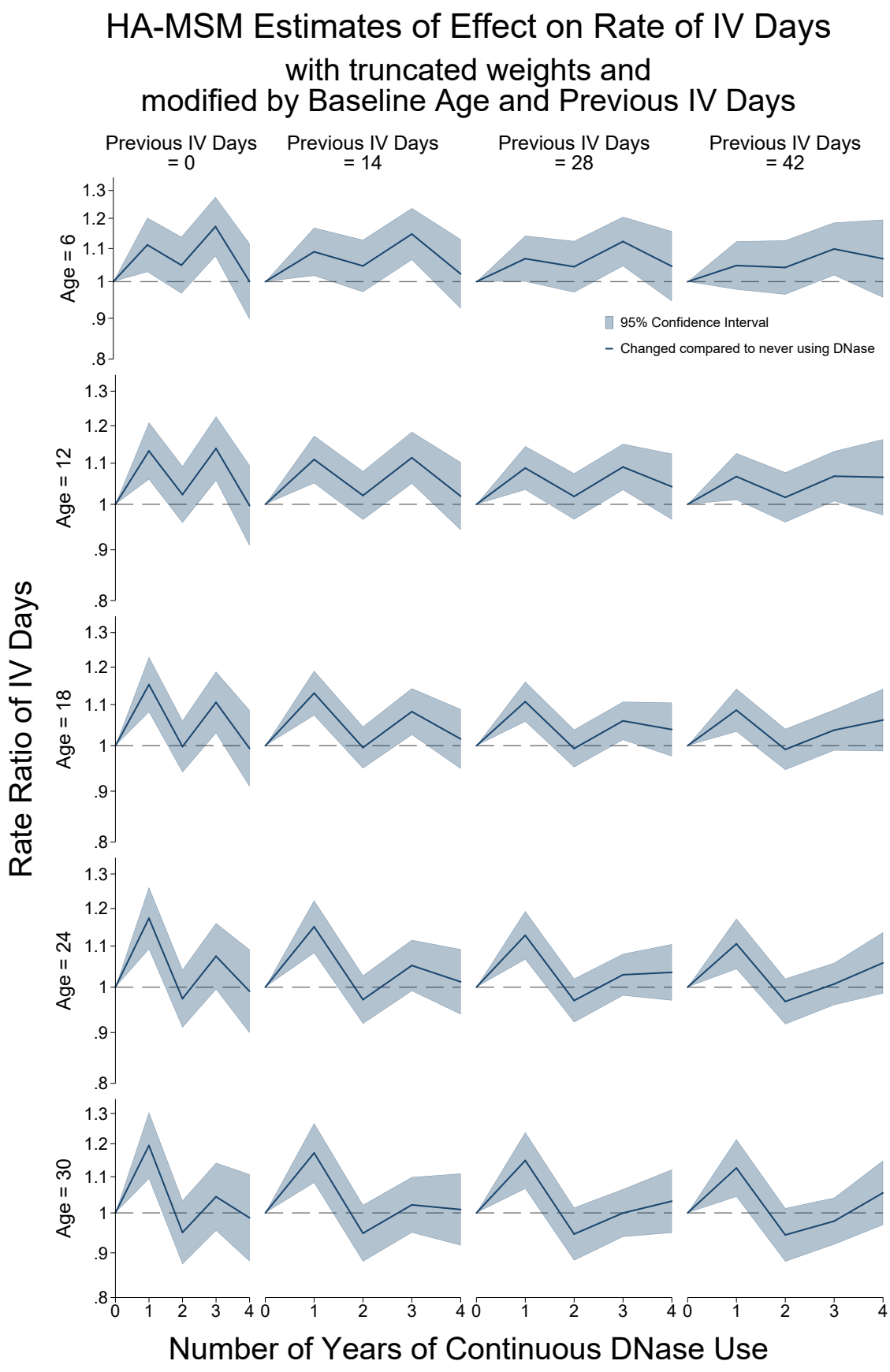


FIGURE 6.12: Estimated effects from HA-MSM with truncated weights of DNase on rate of zero IV days modified by baseline age and previous IV Days

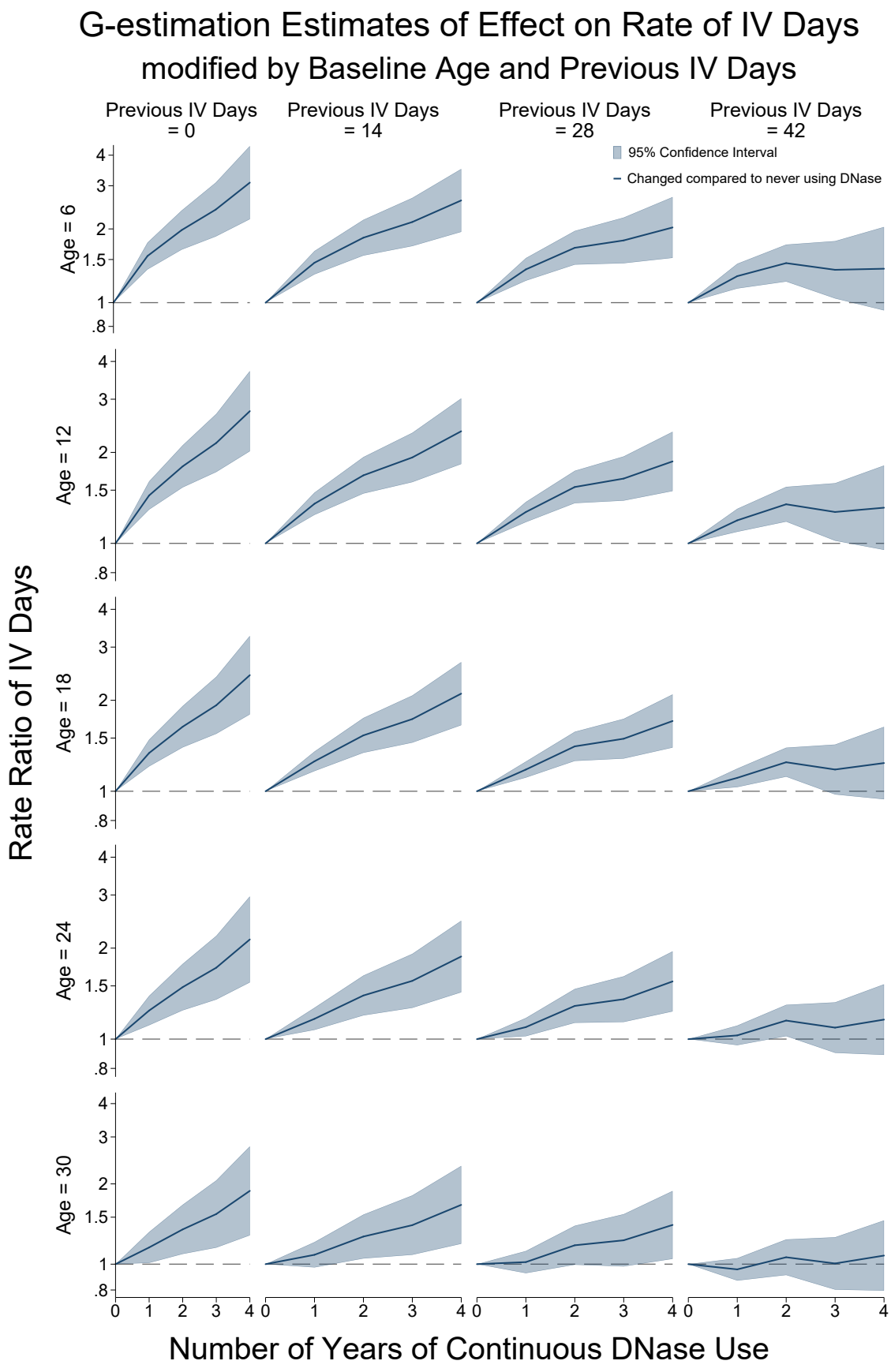


FIGURE 6.13: Estimated effects from g-estimation of DNase on rate of IV days modified by baseline age and previous IV Days

Method		1 Year Treatment Effect					
		OR	95% CI	*p*	IRR	95% CI	*p*
HA-MSM	Intercept	0.57	0.46, 0.70	<0.0001	1.05	0.94, 1.18	0.36
	Interaction by Age	1.08	0.99,1.17	0.093	1.05	0.99, 1.10	0.10
	Interaction by Previous IV days	1.21	1.07, 1.36	0.002	0.98	0.96, 1.01	0.18
HA-MSM (truncated)	Intercept	0.56	0.46, 0.69	<0.0001	1.09	0.99, 1.20	0.072
	Interaction by Age	1.08	0.99, 1.18	0.077	1.03	0.99, 1.08	0.18
	Interaction by Previous IV days	1.23	1.09, 1.38	<0.001	0.98	0.96, 1.00	0.062
G-estimation	Intercept	-	-	-	1.67	1.43, 1.95	<0.0001
	Interaction by Age	-	-	-	0.88	0.83, 0.95	<0.001
	Interaction by Previous IV days	-	-	-	0.94	0.90, 0.98	0.002
Method		Additional 2nd Year Treatment Effect					
		OR	95% CI	*p*	IRR	95% CI	*p*
HA-MSM	Intercept	0.81	0.64, 1.03	0.086	1.11	0.98, 1.26	0.095
	Interaction by Age	1.15	1.05, 1.27	0.003	0.94	0.89, 1.00	0.052
	Interaction by Previous IV days	0.96	0.86, 1.07	0.49	0.99	0.97, 1.02	0.57
HA-MSM (truncated)	Intercept	0.82	0.65, 1.03	0.093	1.07	0.97, 1.19	0.17
	Interaction by Age	1.14	1.04, 1.25	0.005	0.96	0.92, 1.01	0.090
	Interaction by Previous IV days	0.96	0.86, 1.07	0.44	1.00	0.98, 1.02	0.83
G-estimation	Intercept	-	-	-	1.33	1.15, 1.55	<0.001
	Interaction by Age	-	-	-	0.96	0.90, 1.03	0.23
	Interaction by Previous IV days	-	-	-	0.95	0.92, 0.99	0.007
Method		Additional 3rd Year Treatment Effect					
		OR	95% CI	*p*	IRR	95% CI	*p*
HA-MSM	Intercept	1.12	0.87, 1.44	0.37	1.23	1.11, 1.37	<0.001
	Interaction by Age	0.98	0.88, 1.09	0.69	0.95	0.91, 0.99	0.015
	Interaction by Previous IV days	1.03	0.92, 1.16	0.62	0.97	0.95, 1.00	0.038
HA-MSM (truncated)	Intercept	1.10	0.87, 1.41	0.43	1.21	1.09, 1.33	<0.001
	Interaction by Age	0.98	0.89, 1.09	0.74	0.95	0.91, 0.99	0.024
	Interaction by Previous IV days	1.02	0.91, 1.15	0.71	0.98	0.95, 1.00	0.090
G-estimation	Intercept	-	-	-	1.27	1.07, 1.51	0.008
	Interaction by Age	-	-	-	0.97	0.89, 1.06	0.49
	Interaction by Previous IV days	-	-	-	0.94	0.90, 0.98	0.008
Method		Additional 4th Year Treatment Effect					
		OR	95% CI	*p*	IRR	95% CI	*p*
HA-MSM	Intercept	0.99	0.72, 1.35	0.93	1.00	0.88, 1.15	0.97
	Interaction by Age	1.01	0.88, 1.15	0.91	0.99	0.93, 1.05	0.71
	Interaction by Previous IV days	0.99	0.84, 1.18	0.94	1.02	0.99, 1.06	0.24
HA-MSM (truncated)	Intercept	0.97	0.71, 1.33	0.86	1.00	0.88, 1.15	0.96
	Interaction by Age	1.00	0.88, 1.14	0.97	0.99	0.94, 1.05	0.85
	Interaction by Previous IV days	0.99	0.84, 1.18	0.95	1.02	0.99, 1.06	0.19
G-estimation	Intercept	-	-	-	1.39	1.10, 1.76	0.007
	Interaction by Age	-	-	-	0.97	0.87, 1.07	0.50
	Interaction by Previous IV days	-	-	-	0.94	0.88, 0.99	0.027

TABLE 6.7: Estimated long-term effects of DNase on odds of zero IV days (OR) and rate of total IV days (IRR) (The intercept term is the estimated effect if the age or previous IV days is 0, and the interaction term is the change in effect per 10 increase age or per 14 increase in IV days)

6.4 Comparison of Methods

In Chapter 5, we assessed the performance of five different methods that could be used to analyse the UK CF Registry data to estimate the effects of long-term DNase use. The results from these simulation studies suggested that any of the methods could be used to provide consistent effect estimates in an ideal scenario where all models are correctly specified and all assumptions are met. In this chapter, we analysed the UK CF Registry data with all available methods as any discrepancies between the methods could highlight potential problems.

We started by analysing the data using SCMM, as although this method can only estimate the effect of one-year of treatment in our setting, the method is much simpler than the other available methods, and a subseqeunt test can be used to test for the presence of longer-term effects. In our setting, the results of this test suggested that there is an effect of long-term DNase use on lung function, but that in the case of IV days the estimated effects of long-term treatment use would be similar to the one-year effect estimates. For lung function, this finding was borne out in the results from the methods which allow estimation of long-term effects, as all four other methods did show that the estimated effect of treatment on lung function changed as the cumulative number of years using DNase increased. For IV days the results on long-term effects were more ambiguous, as two of the methods (IPW of MSM and HA-MSM) showed that the estimated effect stayed the same through years one to five, but g-formula and g-estimation estimated that the effect did change with continuous use of DNase. In the case of g-estimation, however, the CIs were quite wide and could be consistent with no long-term effect.

The long-term treatment effects obtained using the four methods did not differ substantially, and the overall conclusions would be the same no matter which results were used. However, there did appear to be important modification of the treatment effect by age or previous levels of the outcome, and as such we show a preference towards HA-MSM and g-estimation which can estimate the desired interaction terms. The results from these two methods were remarkably similar. Another issue that was highlighted in the simulation studies in Chapter 5 was that the results would be biased if a causal pathway between two variables was specified in the wrong direction. In the UK CF Registry data, number of IV days and treatment status are collected over the same time period and it is therefore not clear which direction this causal pathway should be specified. In the simulation studies, misspecifying the pathway often led to very extreme weights for HA-MSM and the method performed very poorly unless truncated weights were used. We saw in our analysis of the UK CF Registry data that there were no very extreme weights and truncation had hardly any effect on the effect estimates. This suggests that misspecification of the causal pathway was not an important issue in this analysis.

6.5 Discussion

We used UK CF Registry data to estimate long-term effects of DNase use, controlling for confounding by indication by using state-of-the-art statistical methodology not previously applied to CF registry data. For individuals with a reduced lung function and not using DNase, we have shown that initiating DNase treatment and using it for one year brings a benefit such that ppFEV$_1$ is higher after one year than it would have been had those individuals not initiated DNase treatment. This beneficial effect appeared to remain with continued use of DNase out to five years, but with no overall modification of the lung function trajectory, as the estimated effect remained stable between years one and five.

Crude comparisons between those who received and did not receive DNase clearly indicate that there is confounding by indication, such that individuals taking DNase tend to have worse health status than those not taking DNase. If this is not appropriately handled in the analysis, any estimates of the treatment effect would not have a causal interpretation. In this study, we made use of appropriate statistical methods to address the confounding by indication, accounting for the longitudinal setting, thereby showing how registries can be used to evaluate the long-term effects of treatment. RCTs are the gold standard for establishing treatment efficacy, but as previously discussed, it is preferable for a CF treatment to alter lung function trajectory rather than to provide a one-off improvement, and assessing change in trajectory requires longer follow-up than would typically be feasible in trials.[20] The analyses we have used take advantage of clear heterogeneity in treatment practices as the proportion of patients receiving DNase at individual CF centres ranges from less than 20% to more than 80%.[3] As the groups who receive and do not receive DNase include individuals with wide-ranging clinical characteristics, the statistical methods used in this paper can correct for confounding by indication as long as data on all confounders have been collected. We were able to adjust for a large number of variables using the data available in the UK CF Registry, but it is not possible to verify whether confounding by indication has ever been completely dealt with and the causal interpretation of our results therefore depends on the assumption of no unmeasured confounding.[106]

The results from our analysis appear to show a much weaker treatment effect than has been reported in previous studies. For example, a previous registry study by Hodson et al. of the European Epidemiologic Registry of Cystic Fibrosis estimated a one-year treatment effect of DNase on ppFEV1 of 3.6% (95% CI 1.8% to 5.3%) and a two-year treatment effect of 2.5% (0.7% to 4.4%).[64] These are larger population-average treatment effect estimates than obtained in our study, but the population for those studies had lower average baseline ppFEV$_1$, who were the patients we found benefited most from treatment.

Only two previous studies have investigated the effects of DNase in people with ppFEV$_1$ $>$ 80% and only one of these included lung function as an outcome.[66, 74] That study only administered DNase for four weeks and found no effect of treatment on ppFEV$_1$.[74] In our study, for individuals with higher lung function, those on DNase treatment had steeper trajectories of lung function decline than comparable individuals not receiving treatment. This may suggest that it would be more beneficial, in terms of lung function outcomes, to wait until lung function starts to decline before initiating DNase. However, as with all observational studies, it is possible that unobserved confounders affect these results. With the rich registry data, we believe we have accounted for the covariates that could affect both lung function and the probability of receiving DNase treatment to account for confounding by indication, but it is possible that there are some unmeasured health-related variables that affected the decisions to initiate treatment in these individuals. For example, although an individual may have had a high lung function measurement at the previous annual assessment, we do not have an assessment of their overall lung function decline during the previous year, and it would be plausible that the patients who did start to receive DNase with a high lung function measure in the Registry might have been starting to show signs of lung function deterioration that were not picked up by having only one lung function measure per year.

The results from the analysis with IV days as the outcome do not appear to agree with the findings from previous studies that investigated the effects of DNase on exacerbations. Although most of these studies did not have statistically significant results, all but one previous study has estimated DNase to result in a reduction in the risk of exacerbations. Although there did appear to be statistically significant effect modification of the treatment effect by age and previous number of IV days, this only resulted in DNase being estimated to be beneficial in older people with many previous IV days, which is a very small proportion of the general CF population. However, one of the limitations of the way the data on IV days is collected in the UK CF Registry is that sometimes people are prescribed IV days as a protective measure rather than to actually treat an exacerbation. It seems plausible that people who are more likely to received DNase may also be more likely to receive these additional non-exacerbation-related IV days. This practice is more common in children than adults, which would also explain why DNase appeared more harmful in younger people. Overall, this seems to suggest that the results from this analysis are unreliable and that the way IV days are collected in the UK CF registry could be enhanced by recording the reason they are being given.

Upon initiation of DNase, most people continue to receive the treatment indefinitely, with only very small numbers of people stopping treatment. Due to this, with the sample size available, it was not possible to estimate the effect of stopping treatment. However, as we observed that treatment only appeared to be beneficial in individuals with reduced lung function, future studies could investigate whether treatment needs to be continued

in people who recover to higher levels of lung function.

A major strength of this study is the use of the UK CF Registry data, which are collected at regular intervals according to a standardised protocol. The data include a large number of variables that we could account for as potential confounders. One of the main limitations of this study is that there are no data available on levels of adherence to treatment. It is known to be particularly hard to measure adherence levels, but previous studies have estimated that average adherence levels to nebulised therapies, such as DNase, may range between 60% and 70%.[107, 108] Specifically, for a longitudinal study, we may expect adherence levels to be higher at treatment initiation and decrease through time, which may partly explain why the estimated effects are not as pronounced as those estimated from RCTs, where adherence would typically be higher.[109]

It is also acknowledged that spirometry measures, such as FEV_1, may not be sufficiently sensitive to detect the early stages of lung function decline, and it has been suggested that other measures such as LCI may give a better indication of early lung function deterioration.[22, 110] Unfortunately, LCI is not collected in the UK CF Registry, so it was not possible to investigate this. However, FEF_{25-75} is collected in the registry, albeit less reliably than FEV_1 (3,320 FEF_{25-75} measurements compared to 20,923 FEV_1 measurements in this analysis), and the results showed similar findings to the findings with FEV_1, but with much larger CIs, reflecting the smaller sample size and higher variability of this measure (see Appendix E).

In conclusion, we have shown a beneficial long-term effect of DNase in people with reduced lung function, but with no overall change in lung-function trajectory. There is a differential effect of treatment based on lung function at treatment initiation with no improvements in lung function seen in individuals initiating treatment with $ppFEV_1$ higher than 70%, suggesting that, in terms of lung function, it could be more effective to initiate treatment only when lung function starts to decline. Finally, these analyses highlight the potential of registries in investigating the effects of long-term treatment use and that issues of confounding by indication can be addressed with appropriate statistical methods.

Part III

Ivacaftor

Chapter 7

Systematic Review of Ivacaftor Use in Cystic Fibrosis

7.1 Overview

Part II of this thesis focussed on DNase, which works by reducing the viscosity of mucus in the airways helping to combat the effects of CF on the lungs. Unfortunately, it does not tackle the underlying cause of the disease and until recently all other available treatments had similarly only reduced the symptoms of the disease. Recently there has been a push to discover treatments that can target and correct the underlying problems with the synthesis and function of the CFTR protein. The first of these treatments to become available was ivacaftor (trade name Kalydeco®).

Ivacaftor works by increasing the open probability of the ion channel.[111] Figure 7.1 shows the different CFTR mutations grouped by functional class and their effect on CFTR protein synthesis and function.[112] In this figure, it can be seen how in a class III mutation, the CFTR protein is synthesised almost as normal and reaches the cell wall, but the gating of the protein is defective and the ion channel remains closed. Ivacaftor was developed for this class of mutation, allowing the gating regulation to be corrected. More recently, studies have investigated the potential effectiveness of ivacaftor in people with other classes of mutation, but for this thesis we only consider the effects of ivacaftor on people with a class III mutation.

Ivacaftor has been available to people with an eligible CF mutation through the NHS in the UK since 2012, meaning that there are four years' of data available in the UK CF Registry up to 2016. During this period, almost all people with a gating mutation have started to receive ivacaftor, with most people starting to receive it as soon as it became available. However, due to its recent availability there remain a lot of unanswered questions about its effects, especially its long-term effects. However, even in this relatively short time there have been many studies published about ivacaftor and therefore this chapter contains a systematic review of the studies investigating the effects of ivacaftor on people

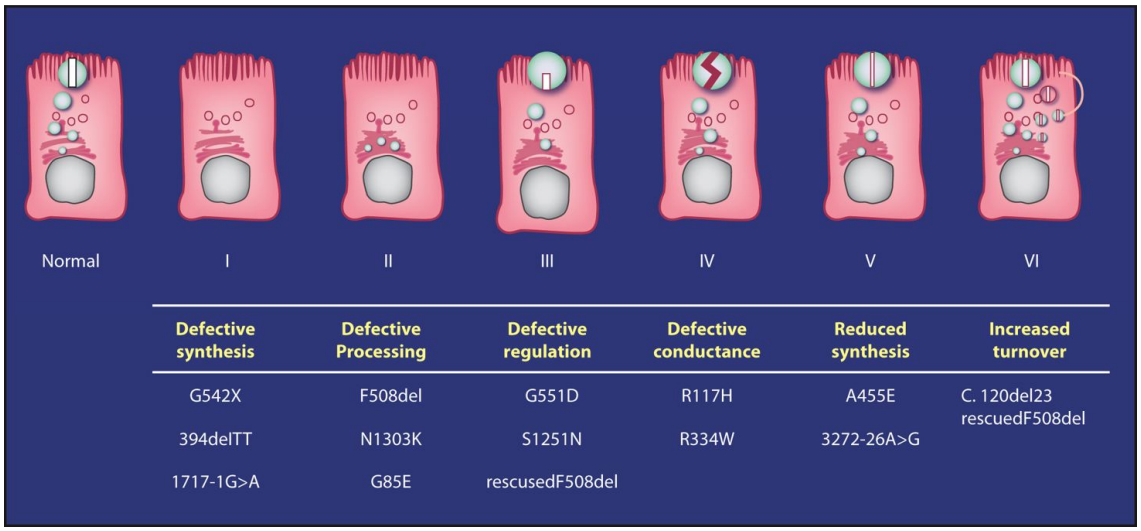


FIGURE 7.1: Diagrams of how different CFTR mutations interfere with CFTR protein synthesis and function (Taken from Hodson and Geddes' Cystic Fibrosis. Taylor & Francis; 2015)

with a gating mutation. This will be followed in the next two chapters by analyses of the UK CF Registry data investigating the four-year effects of ivacaftor.

There have been some more recent studies which suggest that ivacaftor may be effective in people with a residual function mutation, but so far in the UK only people with gating mutations have been eligible to receive ivacaftor.[113, 114] Any analysis we can currently perform with the UK CF Registry data would thus only be applicable to people with a gating mutation, and therefore, we will only consider the treatment effect of ivacaftor in those with a gating mutation in this thesis.

7.2 Objectives

This systematic review aims to identify and summarise relevant studies which have investigated the effects of ivacaftor in people with a CF-causing gating mutation.

The results of the studies identified in the systematic review will be used to compare the results of the analysis of the UK CF registry data presented in this PhD. Therefore, we will only include studies in the systematic review if they contain an outcome that is collected in the UK CF Registry.

7.3 Methods

As with the systematic review of DNase in Chapter 3, this review was carried out with reference to the PRISMA Statement.[31]

Studies were eligible for inclusion if they fulfilled the following PICOS criteria:

- ***P*articipants:** People with a CF-causing gating mutation (the comparator group not receiving ivacaftor may have a different CF-causing mutation).
- ***I*nterventions:** Ivacaftor given at any dose and frequency.
- ***C*omparisons:** Placebo or no treatment
- ***O*utcomes:** We consider two outcomes: lung function measured by FEV_1, and exacerbations and/or number of IV days. We restrict the review to these outcomes, as these are the ones we consider for the our analysis in Chapter 9.
- ***S*tudy Designs:** Both RCTs and observational studies, but treatment must be compared to a group of people not receiving ivacaftor, i.e. studies only comparing change since baseline are not included.

Two search strategies were employed to identify potential studies for the systematic review:

- Hand search of two journals: Pediatric Pulmonology and Journal of Cystic Fibrosis
- Electronic search of two databases: MEDLINE and EMBASE.

The hand search was carried out on all issues, including supplements, of both journals since 2008. The search strategy for the electronic databases was for any of the following terms to be included in the title of the article: ivacaftor, VX-770, CFTR potentiator, or Kalydeco. These searches were performed in January 2018.

After removing any papers which did not fulfil all of the PICOS criteria listed above, the following data were collected from all of the remaining studies: publication date, type of study, follow-up time, sample size, baseline distribution of age and lung function, the outcome measures, the comparator group used and the results. The quality of the studies was also assessed using the Downs & Black checklist.[36]

The findings of the systematic review are summarised in the following section. During the systematic review many of the identified papers and abstracts were found to come from the same studies, but at different time points. For this reason, the results reported in these papers would generally include the same participants, and therefore no meta-analyses have been performed on the results as the observations are not all independent.

The electronic and hand searches, the abstract and full-paper reviews, the quality assessment with the Downs & Black scale, and the qualitative review of included studies were carried out by me.

7.4 Results

After combining the results of the searches from EMBASE, MEDLINE, the Journal of Cystic Fibrosis and Pediatric Pulmonology there were 141 potential papers for inclusion in the systematic review (see Figure 7.2). However, after the full-text review of these studies, 95 of them were not in fact eligible to be included in the systematic review. This was due to two reasons: the studies either did not contain one of the prespecified outcomes of interest, or they did not contain any comparator group. The most common outcomes in these excluded studies were sweat chloride levels or growth variables such as weight or BMI. Tables 7.1 and 7.2 list the studies that were excluded due to having no outcome of interest and no comparator group respectively.

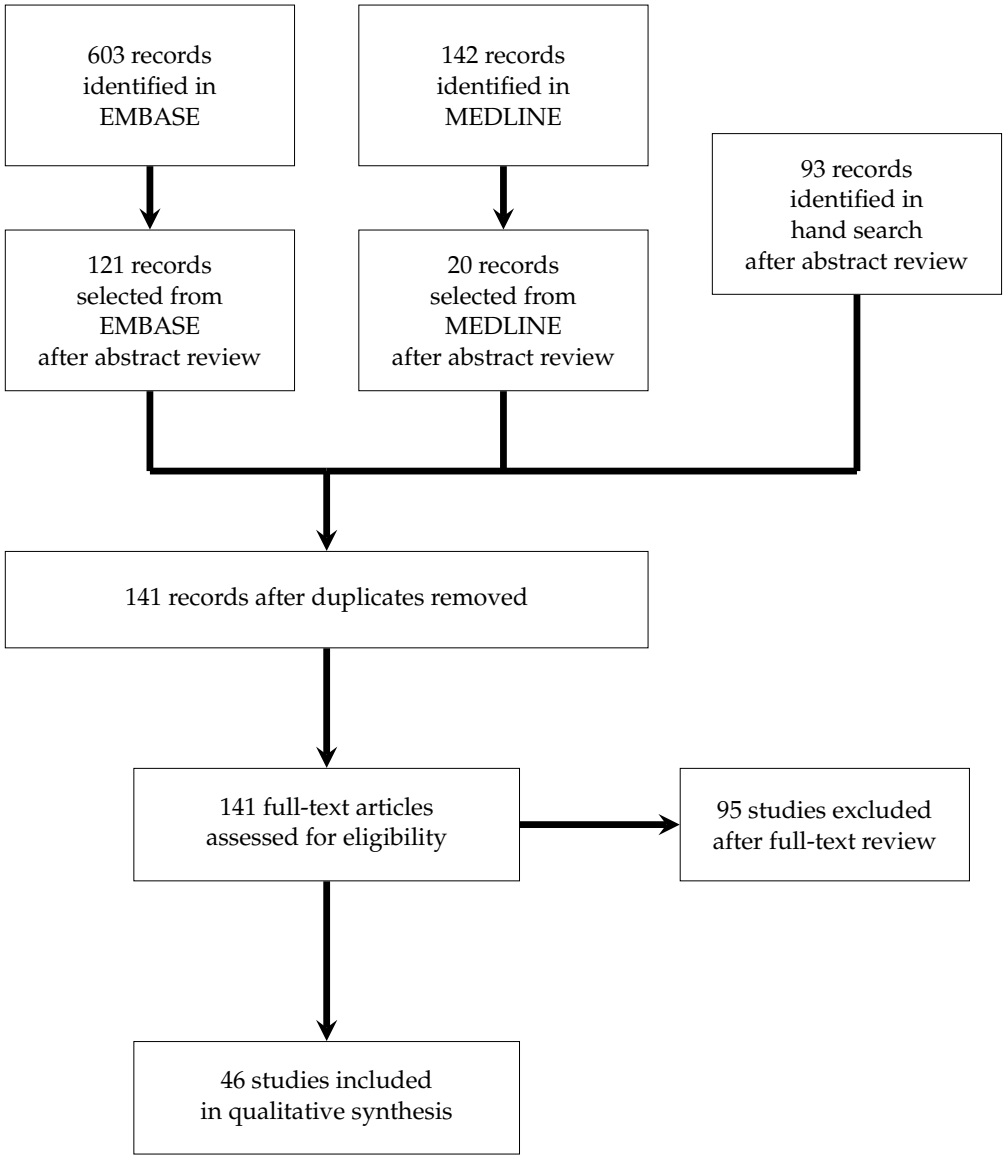


FIGURE 7.2: Flowchart of the number of studies at each stage of the ivacaftor systematic review

Publication Date	Authors	Title	Full Reference
Oct-09	Boyle et al.	Effect of VX-770, a CFTR potentiator, on spirometry and QOL assessment in subjects with CF and the G551D-CFTR mutation	[115]
Jun-10	Rowe et al.	Improvement in ion transport biomarkers and spirometry with the investigational CFTR potentiator VX-770 in subjects with cystic fibrosis and the G551D-CFTR mutation	[116]
Jun-12	Borowitz et al.	Measures of nutritional status in two Phase 3 trials of Ivacaftor in subjects with cystic fibrosis who have the G551D-CFTR mutation	[117]
Jun-13	Plant et al.	Lung function, weight and sweat chloride responses in patients with cystic fibrosis and the G551D-CFTR mutation treated with ivacaftor: A secondary analysis	[118]
Sep-13	Elborn et al.	Lung function, weight and sweat chloride responses in patients with cystic fibrosis and the G551D-CFTR mutation treated with ivacaftor: A secondary analysis	[119]
Oct-13	De Boeck et al.	Ivacaftor, a CFTR potentiator, in cystic fibrosis patients who have a non-G551D-CFTR gating mutation: Phase 3, Part 1 results	[120]
Apr-14	Wainwright et al.	The effect of ivacaftor in individuals with cystic fibrosis and severe lung disease: analysis of data from the Australian named patient programme	[121]
Jun-14	De Boeck et al.	The effect of Ivacaftor, a CFTR potentiator, in patients with cystic fibrosis and a non-G551D-CFTR gating mutation, the KONNECTION study	[122]
Jun-14	Sawicki et al.	The effect of ivacaftor on weight over three years in patients with CF and a G551D-CFTR mutation	[123]
Sep-14	Heltshe et al.	Ivacaftor is associated with Pseudomonas Aeruginosa reduction in cystic fibrosis patients with G551D-CFTR	[124]
Sep-14	Pace et al.	Ivacaftor improves linear growth in G551D pre-pubertal children	[125]
Sep-14	Wainwright et al.	The effect of ivacaftor in individuals with CF and severe lung disease	[126]
Nov-14	Varghese et al.	Outcomes after 1 year of Ivacaftor treatment in CF patients with at least one G551D-CFTR mutation: a single centre experience	[127]
May-15	Konstan et al.	Efficacy response in CF patients treated with ivacaftor: post-hoc analysis	[128]
Jun-15	McKay et al.	The effect of ivacaftor on exocrine pancreatic function in patients with cystic fibrosis and the G551D CFTR mutation who are naïve for ivacaftor	[129]
Jun-15	Pope et al.	Short-term and long-term effects of ivacaftor treatment on sputum microbiota in people with the G551D CFTR mutation	[130]
Jun-15	Prosser et al.	Transformational care at the All Wales Adult CF Centre (AWACFC) - is ivacaftor making us fat? The impact of ivacaftor (Kalydeco®) on body composition	[131]
Jun-15	Rosenfeld et al.	An open-label study of the safety, pharmacokinetics and pharmacodynamics of ivacaftor in patients aged 2 to 5 years with cystic fibrosis and a CFTR gating mutation: The KIWI study	[132]
Jun-15	Tierney et al.	Ivacaftor and its effects on body composition in adults with G551D realted cystic fibrosis	[133]
Oct-15	Barry et al.	Ivacaftor decreases mortality in G551D patients with severe lung disease	[134]
Oct-15	Rosenfeld et al.	Extended evaluation of ivacaftor treatment in pediatric patients with cystic fibrosis and a CFTR gating mutation	[135]
Oct-15	Stalvey et al.	Ivacaftor improves linear growth in G551D cystic fibrosis children: Results of a multicenter, placebo-controlled study	[136]
Dec-15	Davies et al.	Ivacaftor treatment in preschool children with cystic fibrosis and a CFTR gating mutation: extended evaluation	[137]
Jan-16	Borowitz et al.	Nutritional status improved in cystic fibrosis patients with the G551D mutation after treatment with ivacaftor	[138]

Publication Date	Authors	Title	Full Reference
Jun-16	Stalvey et al.	Ivacaftor improves linear growth in children with cystic fibrosis (CF) and a G551D-CTFR mutation: data from the ENVISION study	[139]
Oct-16	Adam et al.	Quantitative CT scan assessment of lung structure and function after one year of Ivacaftor therapy	[140]
Oct-16	Looi et al.	Ivacaftor therapy increases BMI but does not affect serum cholesterol in patients with gating mutations	[141]
Oct-16	Rosenfeld et al.	Long-term safety and efficacy of ivacaftor in pediatric patients aged 2-5 years with CF and a CFTR gating mutation	[142]
Feb-17	Stalvey et al.	Growth in prepubertal children with cystic fibrosis treated with ivacaftor	[143]
Sep-17	McCullagh et al.	Long term microbiological outcomes of Ivacaftor use. A single-centre retrospective study	[144]

TABLE 7.1: Studies excluded from the ivacaftor systematic review after full-text review due to no relevant outcome

Publication Date	Authors	Title	Full Reference
Oct-08	Accurso et al.	Interim results of phase 2a study of VX-770 to evaluate safety, pharmacokinetics, and biomarkers of CFTR activity in cystic fibrosis subjects with G551D	[145]
Jun-09	Accurso et al.	Final results of a 14- and 28-day study of VX-770 in subjects with CF	[146]
Oct-11	McKone et al.	Long-term safety and efficacy of investigational CFTR potentiator, VX-770, in subjects with CF	[147]
Jun-12	McKone et al.	Long-term safety and efficacy of ivacaftor in subjects with cystic fibrosis who have the G551D-CFTR mutation	[148]
Sep-12	McKone et al.	Long-term safety and efficacy of ivacaftor in persons with cystic fibrosis who have the G551D-CFTR mutation	[149]
Sep-12	McKone et al.	Long-term safety and efficacy of ivacaftor in subjects with CF who have the G551D-CFTR mutation	[150]
Apr-13	Wood et al.	Observational study of the clinical effects of ivacaftor in patients with severe cystic fibrosis (CF) lung disease	[151]
May-13	Mondal et al.	Postmarketing experience with ivacaftor therapy in cystic fibrosis patients having G551D mutation	[152]
Jun-13	Barry et al.	UK and Ireland review of Ivacaftor in severe CF: Impact on lung function and weight	[153]
Jun-13	Hubert et al.	Ivacaftor in French patients with cystic fibrosis and a G551D mutation in the real world setting	[154]
Oct-13	Hubert et al.	Ivacaftor in French patients with cystic fibrosis and a G551D mutation	[155]
Oct-13	McGarry et al.	Normalization of sweat chloride concentration and clinical improvement with ivacaftor in a patient with cystic fibrosis with mutation S549N	[156]
Oct-13	McKone et al.	Long-term safety and efficacy of ivacaftor in patients with cystic fibrosis who have the G551D-CFTR mutation: response through 144 weeks of treatment (96 weeks of PERSIST)	[157]
Oct-13	Rowe et al.	Results of the G551D observational study: The effect of Ivacaftor in G551D patients following FDA approval	[158]
Oct-13	Spencer-Clegg et al.	Ivacaftor in the real world - early experience in a large adult CF centre	[159]
Oct-13	Trinh et al.	Ivacaftor in adults with cystic fibrosis: one-year experience in the real world setting	[160]
Dec-13	Barry et al.	Sweat chloride is not a useful marker of clinical response to Ivacaftor	[161]
Dec-13	Ewence et al.	Does Ivacaftor improve objective measurements of health in patients with the G551D cystic fibrosis transmembrane conductance regulator (CFTR) protein mutation? The experience of a UK cystic fibrosis centre	[162]
Dec-13	Hebestreit et al.	Effects of ivacaftor on severely ill patients with cystic fibrosis carrying a G551D mutation	[163]
Jun-14	Barry et al.	Sweat chloride is not a useful marker of clinical response to ivacaftor	[164]

Publication Date	Authors	Title	Full Reference
Jun-14	Green et al.	The effect of ivacaftor therapy on clinical and PCR-identified microbial diversity of cystic fibrosis lung infection	[165]
Jun-14	Jenkins et al.	The use of LCI as an effective tool for monitoring clinical response to Ivacaftor therapy in CF patients with at least one G551D-allele	[166]
Jul-14	Rowe et al.	Clinical mechanism of the cystic fibrosis transmembrane conductance regulator potentiator ivacaftor in G551D-mediated cystic fibrosis	[167]
Sep-14	Adam et al.	Ivacaftor rapidly improves airway distensibility and vascular tone in people with G551D-CFTR suggesting a CF-related smooth muscle abnormality	[168]
Sep-14	Carnovale et al.	Effects of Ivacaftor in cystic fibrosis patients carrying the G1244E mutation with severe lung disease	[169]
Sep-14	Harman et al.	Exploring indices derived from multibreath washout (MBW) following treatment with Ivacaftor	[170]
Sep-14	Reilly et al.	Sustained effects of ivacaftor on muscle strength, body composition, anxiety and depression scores	[171]
Sep-14	Sawicki et al.	The effect of ivacaftor treatment on the rate of lung function decline in CF patients with a G551D-CFTR mutation	[172]
Nov-14	McKone et al.	Long-term safety and efficacy of ivacaftor in patients with cystic fibrosis who have the Gly551Asp-CFTR mutation: a phase 3, open-label extension study (PERSIST)	[173]
Dec-14	Green et al.	The effect of Ivacaftor therapy on the microbial diversity of cystic fibrosis lung infection	[174]
Dec-14	Harman et al.	Changes in indices derived from multibreath washout (MBW) following treatment with Ivacaftor in patients with cystic fibrosis	[175]
Jan-15	Sheikh et al.	Computed tomography correlates with improvement with ivacaftor in cystic fibrosis patients with G551D mutation	[176]
Feb-15	Sheikh et al.	Ivacaftor improves appearance of sinus disease on computerised tomography in cystic fibrosis patients with G551D mutation	[177]
Mar-15	Heltshe et al.	Pseudomonas aeruginosa in cystic fibrosis patients with G551D-CFTR treated with ivacaftor	[178]
Jun-15	Grasemann et al.	Effect of ivacaftor therapy on exhaled nitric oxide in patients with cystic fibrosis	[179]
Jun-15	Ronan et al.	Clinical outcomes of real-world Klaydeco (CORK) study - a prospective 12 month analysis addressing the impact of CFTR modulation on the cystic fibrosis lung	[180]
Jun-15	Zwolsman et al.	Partial recovery of exocrine pancreatic function and intestinal fat absorption after ivacaftor treatment	[181]
Sep-15	Grasemann et al.	Airway nitric oxide production in patients with cystic fibrosis increases with Ivacaftor therapy	[182]
Oct-15	Einarsson et al.	The effect of Ivacaftor treatment on airway microbial community dynamics in patients with G551D (The "Celtic" mutation)	[183]
Oct-15	Fink et al.	Treatment response to Ivacaftor in clinical practice: analysis of the US CF foundation patient registry	[184]
Oct-15	Kane et al.	Lung clearance index response in CF patients with class III CFTR mutations	[185]
Oct-15	Ronan et al.	The impact of CFTR modulation with ivacaftor on circulating inflammatory mediators and their correlation with clinical parameters in patients with the G551D mutation	[186]
Jan-16	Taylor-Cousar et al.	Effect of ivacaftor in patients with advance cystic fibrosis and a G551D-CFTR mutation: Safety and efficacy in an expanded access program in the United States	[187]
Feb-16	Davies et al.	Safety, pharmacokinetics, and pharmacodynamics of ivacaftor in patients aged 2-years with cystic fibrosis and a CFTR gating mutation (KIWI): An open-label single-arm study	[188]
Jun-16	Bertin et al.	Clinical effects of ivacaftor on chronic rhinosinusitis	[189]

Publication Date	Authors	Title	Full Reference
Jun-16	Button et al.	Increased maximal exercise capacity is associated with increased wellness across the spectrum of age and lung function in adults with cystic fibrosis (CF) after treatment with ivacaftor	[190]
Jun-16	Iacotucci et al.	Effects of ivacaftor in cystic fibrosis patients carrying a non-G551D gating mutation	[191]
Jun-16	Stallings et al.	3-month ivacaftor (Kalydeco) treatment resulted in weight gain, decreased resting energy expenditure and gut inflammation, and increased pulmonary function, muscle mass and strength, and growth status	[192]
Jun-16	Zeybel et al.	Ivacaftor treatment is associated with a decline in symptoms of extra-oesophageal reflux in patients with cystic fibrosis and the G551D mutation	[193]
Sep-16	Aziz et al.	Ivacaftor - as effective in clinical practice?	[194]
Oct-16	Button et al.	Increased total exercise time and maximal exercise capacity are associated with increased wellness in adults with CF across spectrum of age and lung function after treatment with Ivacaftor	[195]
Oct-16	Sainath et al.	Improvement in weight, pulmonary function and other outcomes with 3-month ivacaftor treatment differed by exocrine pancreatic status in people with cystic fibrosis gating mutations	[196]
Oct-16	Sainath et al.	Exocrine pancreatic status effects outcomes following 3-month ivacaftor therapy in subjects with cystic fibrosis gating mutations	[197]
Oct-16	Schall et al.	3-month ivacaftor treatment resulted in weight gain, improved pulmonary function and growth status, and reduced resting energy expenditure and gut inflammation in people with cystic fibrosis gating mutations	[198]
Oct-16	Schall et al.	Improved pulmonary function and weight gain, reduced resting energy expenditure, and improved gut inflammaton after 3-month ivacaftor treatment in cystic fibrosis	[199]
Jan-17	Mesbahi et al.	Changes of CFTR functional measurements and clinical improvements in cystic fibrosis patients with non p.Gly551Asp gating mutations treated with ivacaftor	[200]
Feb-17	Stallings et al.	Ivacaftor treatment in cystic fibrosis and improvement in resting energy expenditure gut inflammation and fat absorption	[201]
Jun-17	Luna-Paredes et al.	Impact of treatment with Ivacaftor on Spanish cystic fibrosis patients with gating mutations	[202]
Jul-17	Strang et al.	Pseudomonas eradication and clinical effectiveness of ivacaftor in four hispanic patients with S549N	[203]
Sep-17	Dagan et al.	Ivacaftor for the p.ser549arg gating mutation - the Israeli experience	[204]
Sep-17	Mouzaki et al.	Weight increase in CF patients on Kalydeco is due to decrease in resting energy expenditure and associated with increase in adipose tissue	[205]
Sep-17	Sainath et al.	Effect of ivacaftor treatment on dietary intake in Italian and North American subjects with cystic fibrosis	[206]
Nov-17	Mouzaki et al.	Increase in body weight of patients with cystic fibrosis receiving Kalydeco treatment is due to increase in adipose tissue	[207]
Nov-17	Sainath et al.	Changes in dietary intake with ivacaftor treatment in Italian and North American subjects with cystic fibrosis	[208]
Jan-18	Dryden et al.	The impact of 12 months treatment with ivacaftor on Scottish paediatric patients with cystic fibrosis with the G551D mutation: a review	[209]

TABLE 7.2: Studies excluded from the ivacaftor systematic review after full-text review due to no comparator group

Details of the 46 studies which fulfilled all the criteria can be found in Table 7.3. In total there were 21 RCTs and 25 observational studies. The median length of follow-up time was 48 weeks (range 2 weeks to 4 years), and the median number of patients in each study was 99 (range 8 to 9056). The sample sizes varied greatly depending on the comparator group used: in RCTs only people with a gating mutation would be eligible for inclusion, limiting the available population for these studies, whereas in some of the observational studies people receiving ivacaftor were compared to controls with other CF-causing mutations, hence the total sample size could be much bigger in these studies. Compared to the findings of the DNase systematic review in Chapter 3, the median follow-up time is much longer (48 weeks compared to 12 weeks for the DNase studies) and the median sample size is also larger (99 compared to 70) despite the number of people who would be eligible for ivacaftor being much smaller than the number of people eligible for DNase.

Assessing the quality of the studies with the Downs & Black questionnaire showed a mean score of 19.7 (SD 4.8) out of a total of 32. This is only slightly higher than the average of the studies included in the DNase systematic review (mean 18.7, SD 5.9), and as in that review, we again found here that RCTs tended to score more highly than observational studies: the mean score for RCTs was 22.8 (SD 3.4) compared to 17.2 (SD 4.2) for observational studies. A histogram of these scores is shown in figure 7.3

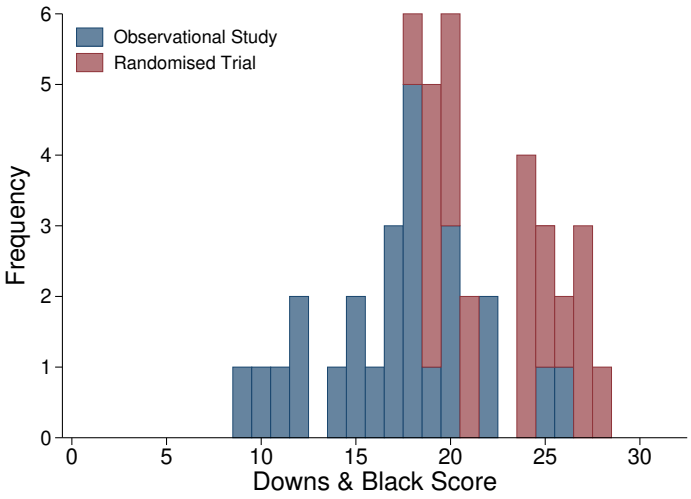


FIGURE 7.3: Histogram of Downs & Black scores for studies included in the ivacaftor systematic review

Most of the studies included in the systematic review contained both children and adults, but there were four studies that only contained children between 6 and 12 years old, and five studies only looking at adults. None of the studies selected included participants under 6 years old, as although there have been some studies investigating the effects of ivacaftor in younger patients, the outcome measures are generally different for young children and these are not routinely collected in the UK CF Registry.

Date	Authors	Title	Type of Study[1]	Quality[2]	Follow-Up (Weeks)	Sample Size	Age (Years)[3]	Baseline $FEV_1(\%)$[3]	Outcomes[4]	Full Reference
Nov-10	Accurso et al.	Effect of VX-770 in persons with cystic fibrosis and the G551D+CFTR mutation	RCT	24	2	8	(19 – 48)	(42 – 97)	L	[210]
Sep-11	Plant et al.	VX-770, an investigational CFTR potentiator, in subjects with CF and the G551D mutation	RCT	19	48	161	(12 –)		L E	[211]
Oct-11	Aherns et al.	VX-770 in subjects 6 to 11 years with Cystic Fibrosis and the G551D-CFTR mutation	RCT	28	24	52	8.9±2.0 (6-11)	84.2 (40 – 105)	L	[212]
Oct-11	Ramsey et al.	Efficacy and safety of VX-770 in subjects with cystic fibrosis and the G551D-CFTR mutation	RCT	20	48	161	(12 –)		L E	[213]
Nov-11	Ramsey et al.	A CFTR potentiator in patients with cystic fibrosis and the G551D mutation	RCT	28	48	161	25.5 (12 –)	63.6 (32 – 98)	L E	[214]
May-12	Elborn et al.	Pulmonary effects of the investigational CFTR potentiator, Ivacaftor, in two phase 3 trials in subjects with CF who have the G551D-CFTR mutation	RCT	25	48	161	(5 –)	(40 – 105)	L E	[215]
Jun-12	Davies et al.	Ivacaftor in subjects 6 to 11 years of age with cystic fibrosis and the G551D-CFTR mutation	RCT	21	48	52	(6 – 11)	84.2±18.1 (40 – 105)	L	[216]
Jun-12	Davies et al.	Effect of Ivacaftor on lung function in subjects with CF who have the G551D-CFTR mutation and mild lung disease: a comparison of lung clearance index (LCI) vs. spirometry	RCT	20	4	14	14.0±8.6 (6 –)	(90 –)	L	[217]
Jun-12	Greiese et al.	Pulmonary exacerbations in a Phase 3 trial of Ivacaftor in subjects with cystic fibrosis who have the G551D-CFTR mutation	RCT	24	48	161	(12 –)	(40 – 90)	E	[218]
Sep-12	Davies et al.	Lung clearance index to evaluate the effect of Ivacaftor on lung function in subjects with CF who have the G551D-CFTR mutation and mild lung disease	RCT	19	4	34	16.6±10.9 (6 –)	97.2±10.6 (90 –)	L	[219]
Sep-12	Elborn et al.	Effects of the CFTR potentiator, Ivacaftor, in two phase 3 trials in subjects with CF who have the G551D-CFTR mutation	RCT	25	48	161	(6 –)	(40 – 105)	L E	[220]

[1] Randomised Controlled Trial or Observational Study
[2] Calculated using Downs & Black scale, maximum score possible is 32
[3] Mean±SD (Range)
[4] The abbreviations refer to the following outcomes: **L**ung function, **E**xacerbations, and **S**urvival

Date	Authors	Title	Type of Study[1]	Quality[2]	Follow-Up (Weeks)	Sample Size	Age (Years)[3]	Baseline $FEV_1(\%)$[3]	Outcomes[4]	Full Reference
Sep-12	Ratjen et al.	Effect of ivacaftor on lung clearance index and FEV1 in subjects with CF who have the G551D-CFTR mutation and mild lung disease	RCT	24	4	40	16.6±10.9 (6 –)	97.2±10.6 (90 –)	L	[221]
Dec-12	Davies et al.	Lung clearance index to evaluate the effect of Ivacaftor on lung function in subjects with CF who have the G551D-CFTR mutation and mild lung disease	RCT	196	4	34	16.6±10.9 (6 –)	97.2±10.6 (90 –)	L	[222]
Jun-13	Barry et al.	UK and Ireland review of Ivacaftor in severe CF: Impact on hospitalisations and antibiotic use	Obs	12	26	32		(– 40)	E	[223]
Jun-13	Davies et al.	Efficacy and safety of ivacaftor in patients aged 6 to 11 years with cystic fibrosis with a G551D mutation	RCT	26	48	52	8.9 (6 – 12)	84.2 (44 – 133.8)	L E	[224]
Jun-13	Flume et al.	Pulmonary exacerbations in CF patients with the G551D-CFTR mutation treated with ivacaftor	RCT	20	48	161	(6 –)		L E	[225]
Jun-13	Davies et al.	Assessment of clinical response to ivacaftor with lung clearance index in cystic fibrosis patients with a G551D-CFTR mutation and preserved spirometry: a randomised controlled trial	RCT	27	4	34	16.6±11.2 (8 – 43)	97.2±10.8 (90 –)	L	[226]
Oct-13	Elborn et al.	Effect of Ivacaftor in patients with cystic fibrosis at the G551D-CFTR mutation who have baseline FEV1>90% of predicted	RCT	19	7	28	(6 –)	99.2 (90 –)	L	[227]
May-14	Ronan et al.	Real world sustained efficacy, tolerability and satisfaction with ivacaftor use in a single adult cystic fibrosis centre cohort	Obs	16	26	30	(16 –)		E	[228]
Jun-14	Ronan et al.	Clinical outcomes of real world Kalydeco (CORK) study	Obs	11	26	58	(6 –)		E	[229]
Jul-14	Barry et al.	Effects of ivacaftor in patients with cystic fibrosis who carry the G551D mutation and have severe lung disease	Obs	20	34	56	22.6 (20 – 31)	28.9±7.6 (– 40)	L E	[230]

[1] Randomised Controlled Trial or Observational Study
[2] Calculated using Downs & Black scale, maximum score possible is 32
[3] Mean±SD (Range)
[4] The abbreviations refer to the following outcomes: **L**ung function, **E**xacerbations, and **S**urvival

Date	Authors	Title	Type of Study[1]	Quality[2]	Follow-Up (Weeks)	Sample Size	Age (Years)[3]	Baseline $FEV_1(\%)$[3]	Outcomes[4]	Full Reference
Oct-14	Nedevska et al.	Ivacaftor: a new treatment for cystic fibrosis which results in reduced sweat chloride and improved clinical outcomes	Obs	10	52	22	23.5 (6 – 36)		E	[231]
Dec-14	De Boeck et al.	Efficacy and safety of ivacaftor in patients with cystic fibrosis and a non-G551D gating mutation	RCT	27	8	78	22.8 (6 – 57)	78.4 (43 - -119)	L	[232]
Jun-15	Guha-niyogi et al.	Transformational care at the All Wales Adult CF Centre (AWACFC) - the impact of Ivacaftor (Kalydeco®) one year on	Obs	15	52	22	28±13	63.5±26.2	E	[233]
Sep-15	Edge-worth et al.	Exercise improvements in ivacaftor treated G551D cystic fibrosis patients are not solely related to FEV1 and sweat changes	RCT	18	17	40			L	[234]
Oct-15	Bai et al.	Ivacaftor long-term safety study: Analysis of 2013 US CF Foundation patient registry data	Obs	18	73	5931			E S	[235]
Oct-15	Sawicki et al.	Sustained benefit from ivacaftor demonstrated by combining clinical trial and cystic fibrosis patient registry data	Obs	26	156	1075	21.5±10.7	67.2±20.3	L	[236]
Jun-16	Bai et al.	Real-world outcomes in patients (pts) with cystic fibrosis (CF) treated with ivacaftor (IVA): analysis of 2014 US and UK CF registries	Obs	22	52	7456			E S	[237]
Jun-16	Bai et al.	Real-world outcomes in young (6- to 12-year-old) patients (pts) with cystic fibrosis (CF) treated with ivacaftor (IVA): analysis of 2014 US and UK CF registries data	Obs	22	52	1324	(6 – 12)		E	[238]
Jun-16	Fila et al.	Ivacaftor in cystic fibrosis adults: Czech experience with six years of follow-up	Obs	12	52	10	28.6 (21 – 36)	45 (16 – 85)	L	[239]
Jun-16	Hassan et al.	Reduction in pulmonary exacerbations (Pex) after initiation of Ivacaftor: a retrospective cohort study among patients with cystic fibrosis (CF) treated in real-world settings	Obs	15	52	168	22.1±12.7 (6 –)		E	[240]

[1] Randomised Controlled Trial or Observational Study
[2] Calculated using Downs & Black scale, maximum score possible is 32
[3] Mean±SD (Range)
[4] The abbreviations refer to the following outcomes: **L**ung function, **E**xacerbations, and **S**urvival

Date	Authors	Title	Type of Study[1]	Quality[2]	Follow-Up (Weeks)	Sample Size	Age (Years)[3]	Baseline $FEV_1(\%)$[3]	Outcomes[4]	Full Reference
Jun-16	Hubert et al.	Retrospective observational study of French patients with cystic fibrosis and a G551D mutation after 1 and 2 years of treatment with ivacaftor	Obs	14	104	110	17.7 (6 – 52)	72.3±26.5	E	[241]
Jun-16	Iacotucci et al.	Effects of ivacaftor in cystic fibrosis patients carrying a non-G551D gating mutation with severe lung disease	Obs	9	52	26		35.1±14.3 (40 –)	E	[242]
Jun-16	Volkova et al.	Disease progression in patients (pts) with cystic fibrosis (CF) treated with ivacaftor (IVA): analysis of real-world data from the UK CF registry	Obs	18	104	1642		71.3±23.8	L E	[243]
Oct-16	Bessonova et al.	Analysis of real-world outcomes in patients with CF treated with Ivacaftor from the 2014 US and UK CF registries	Obs	20	52	7456			E S	[244]
Oct-16	Hassan et al.	One-year evaluation of pulmonary exacerbation outcomes among patients with cystic fibrosis initiated on Ivacaftor in a multistate medicaid population	Obs	17	52	88	13.1±6.7 (6 –)		E	[245]
Oct-16	Volkova et al.	Analysis of disease progression in patients with CF treated with ivacaftor in the real world using data from the UK CF registry	Obs	18	104	1642		71.3±23.8	L E	[246]
Jun-17	Bessonova et al.	Disease progression in patients (pts) with cystic fibrosis (CF) treated with ivacaftor (IVA): analysis of real-world data from the UK CF Registry	Obs	19	156	1549			L	[247]
Jun-17	Bessonova et al.	Real-world outcomes in patients (pts) with cystic fibrosis (CF) treated with ivacaftor (IVA): analysis of 2015 US and UK CF registries	Obs	20	52	9056			E S	[248]
Jun-17	Kirwan et al.	Temporal trends in key outcome measures in cystic fibrosis patients treated with Ivacaftor: real-world data from the Irish CF registry	Obs	17	130	228			E	[249]

[1] Randomised Controlled Trial or Observational Study
[2] Calculated using Downs & Black scale, maximum score possible is 32
[3] Mean±SD (Range)
[4] The abbreviations refer to the following outcomes: **L**ung function, **E**xacerbations, and **S**urvival

Date	Authors	Title	Type of Study[1]	Quality[2]	Follow-Up (Weeks)	Sample Size	Age (Years)[3]	Baseline $FEV_1(\%)$[3]	Outcomes[4]	Full Reference
Jul-17	Edgeworth et al.	Improvement in exercise duration, lung function and well-being in G551D-cystic fibrosis patients: a double-blind, placebo-controlled, randomized, cross-over study with ivacaftor treatment	RCT	24	4	40	32 (18 – 65)	54 (23 –110)	L	[250]
Sep-17	Volkova et al.	Disease progression in patients with CF treated with ivacaftor: Analyses of real world data from the US and UK CF Registries	Obs	18	208	3024		82.1±23.2	L E	[251]
Sep-17	Volkova et al.	Real-world outcomes in patients with CF treated with Ivacaftor: 2015 US and UK CF Registry analyses	Obs	17	52	9056			E S	[252]
Oct-17	Dagan et al.	Ivacaftor for the p.Ser549Arg (S549R) gating mutation - The Israeli experience	Obs	18	52	16	21±10 (12 – 40)	74	E	[253]
Jan-18	Flume et al.	Recovery of lung function following a pulmonary exacerbation in patients with cystic fibrosis and the G551D-CFTR mutation treated with ivacaftor	RCT	21	48	161	(12 –)	63.6±16.4 (40 – 90)	E	[254]
Jan-18	Hubert et al.	Retrospective observational study of French patients with cystic fibrosis and a Gly551Asp-GFTR mutation after 1 and 2 years of treatment with ivacaftor in a real-world setting	Obs	25	104	113	21.5±13.1 (6 – 52)	72.3±26.4	E	[255]

TABLE 7.3: Key details of studies included in the systematic review of ivacaftor

[1] Randomised Controlled Trial or Observational Study
[2] Calculated using Downs & Black scale, maximum score possible is 32
[3] Mean±SD (Range)
[4] The abbreviations refer to the following outcomes: **L**ung function, **E**xacerbations, and **S**urvival

In terms of baseline lung function most studies either had no exclusion criteria or had a wide range of eligibility. There were, however, three studies which only included people with severe lung disease (defined in the studies as FEV_1<40% predicted) and six studies which only included people with FEV_1>90% predicted.

Of the two outcome measures we considered, exacerbations and/or IV days were the most commonly reported outcomes, with 31 studies analysing these. FEV_1 was an outcome measure in 26 studies. The following two subsections look at the findings in each of these outcomes in turn.

7.4.1 Exacerbations & IV Days

The most common way to estimate the effect of ivacaftor on exacerbations was to look at the RR of having an exacerbation during the follow-up period between those receiving ivacaftor and those not. However, the way this was measured differed markedly between studies. Firstly, there is no universal definition of an exacerbation, and therefore, each study used its own definition. Secondly, some studies obtained a RR by comparing the proportions of people who had at least one exacerbation in the study period in the ivacaftor and comparison groups, whereas other studies obtained a RR based on the total number of exacerbations during the study period in the two groups. These two issues, combined with the fact that RR are non-collapsible, means that the estimates from different studies will not technically be comparable. However, Figure 7.4 shows a forest plot of all the estimated RR of exacerbations regardless of the specific definitions to help show the overall trend of change in exacerbations when taking ivacaftor.

Three studies had follow-up of approximately half a year (24 weeks or 26 weeks) and all three showed very large reductions in the risk of exacerbations in people taking ivacaftor. The study by Ramsey et al. was a RCT in patients aged over 12 and the estimated RR of an exacerbation in the ivacaftor group relative to the control group was 0.38 (95% CI 0.22, 0.64).[214] The other two studies were observational studies carried out in Ireland by Ronan et al. and estimated the RR of requiring IVs. In adults the estimated RR was 0.15 (95% CI 0.02, 1.0), and in a group containing both adults and children it was 0.069 (95% CI 0.01, 0.49).[228, 229] These were the two most extreme RR estimates, but they also have much larger CIs due to the smaller sample sizes of these two studies.

The majority of studies looked at the risk of exacerbations over a one-year period (either 48 weeks or 52 weeks). Of these studies, three only included children aged between 6 and 12 years. Two of these studies were RCTs and actually reported a higher risk of exacerbations in the group receiving ivacaftor: Davies et al. reported RR 1.33 (95% CI 0.33, 5.38), Flume et al. reported RR 1.75 (95% CI 0.58, 5.27).[224, 225] These were the only two studies to report a negative effect of ivacaftor, but in both studies the overall rate of exacerbations was very low and the CIs were wide. The third study to only report

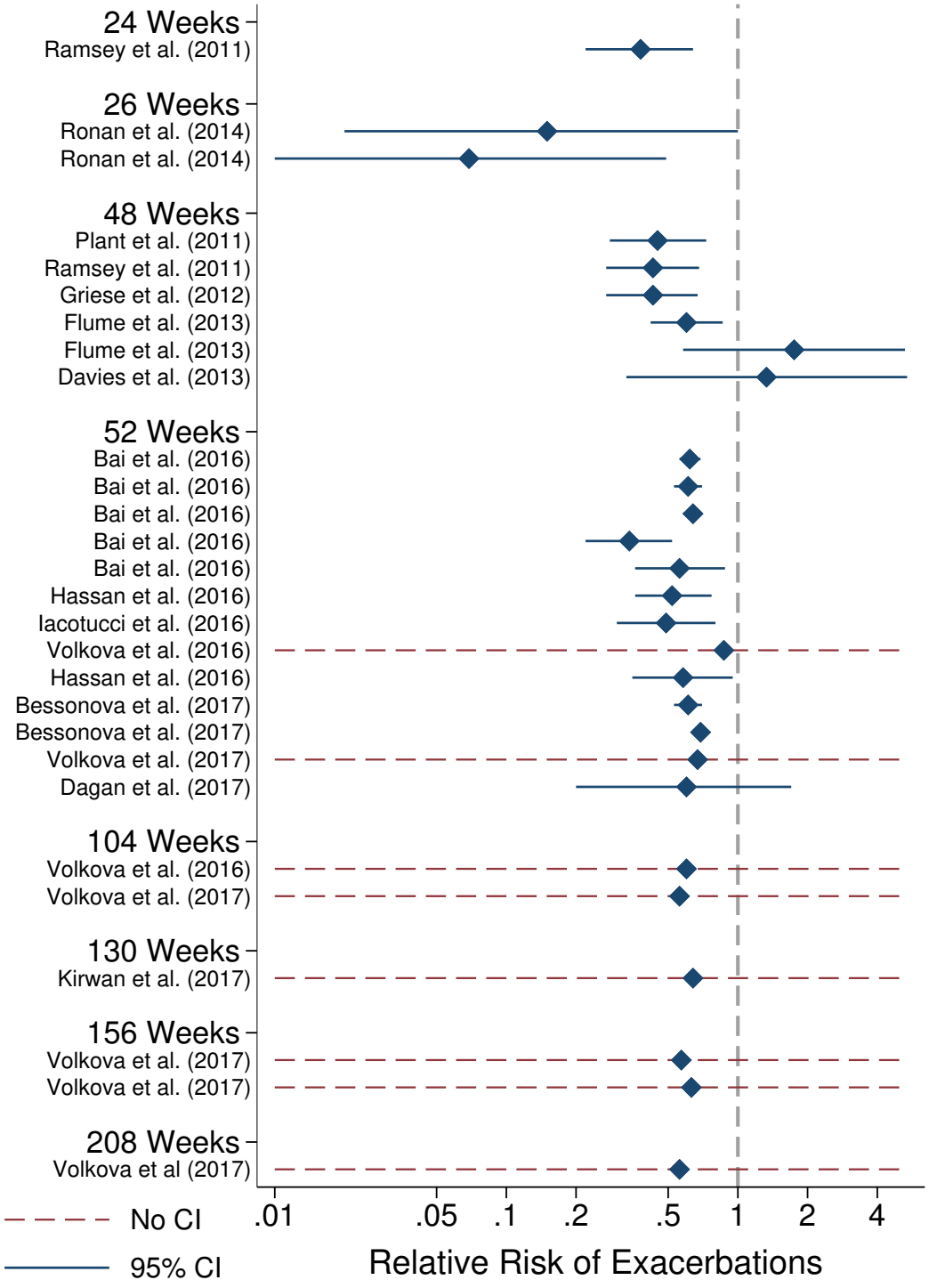


FIGURE 7.4: Forest plot of studies comparing the risk of exacerbations in those taking ivacaftor compared to controls (Some CIs are approximations based on reported *p*-values and papers reporting the same results are only plotted once)

the effect in children was by Bai et al. and reported separate RR for the US and UK: US RR 0.34 (95% CI 0.22, 0.52), UK RR 0.56 (95% CI 0.36, 0.88).[238]

Of the studies reporting one-year exacerbation rates, there were no studies reporting the effects of ivacaftor on exacerbations only in adults, but nine studies did only include people over the age of 12. Four of these papers reported on the same RCT which showed a RR of 0.45 (95% CI 0.28, ,0.73) of an exacerbation in the ivacaftor group.[211, 213, 215, 220] In another two papers, the reported RR was 0.60 (95% CI 0.42, 0.86)[225, 254], another RCT reported a RR of 0.43 (95% CI 0.27, 0.67)[218], and the final RCT reported almost the exact same result of 0.43 (95% CI 0.27, 0.68)[214]. The one observational study to look at the one-year exacerbation risk in patients over 12 compared the number of exacerbations in patients receiving ivacaftor to the number of exacerbations they had suffered in the year prior to starting ivacaftor. The reported RR here was 0.6 (95% CI 0.2, 1.7).[253]

The other studies to report on one-year exacerbation risks were all observational studies with a minimum age of six, but no upper age limit. The majority of these studies looked at the rate of exacerbations in the UK and US CF registries separately, comparing those receiving ivacaftor to matched controls. These papers reported the RR of exacerbations in the years 2012, 2013, 2014 and 2015, but did not account for the length of time using ivacaftor. In the US in 2012 the estimated RR was 0.67 (no CI reported)[251]. In 2013, the estimated RR in the US was 0.62 (95% CI 0.56, 0.69) and in the UK in 2013 it was 0.87 (CI not reported).[235, 243, 246] In 2014, the estimated RR in the US was 0.64 (95% CI 0.58, 0.70), and in the UK was 0.61 (95% CI 0.53, 0.70).[237, 244]. In 2015, the estimated RR in the US was 0.69 (95% CI 0.63, 0.76), and in the UK was 0.61 (95% CI 0.53, 0.70).[248, 252] Two other studies were also carried out in the USA and reported one-year exacerbation RR of 0.52 (95% CI 0.36, 0.77) and 0.58 (95% CI 0.35, 0.95).[240, 245] The last study estimating one-year exacerbation rates was carried out in Italy and only included patients with severe lung disease; here the RR of exacerbations in patients receiving ivacaftor was 0.49 (95% CI 0.30, 0.80) compared to the same patients in the year prior to starting ivacaftor.[242]

There were only four studies which investigated the effect of receiving ivacaftor for longer than a year on exacerbations. Three studies analysed the US and UK CF registries separately, and only included people who had started ivacaftor during its first year of availability. In those patients, in the second year of treatment, the RR of exacerbations in the USA was estimated to be 0.56 (no CI reported) and in the UK was 0.60 (no CI reported).[243, 246, 251] In the third year in the USA the RR was 0.57 (no CI reported) and in the UK was 0.63 (no CI reported), and in the fourth year in the US it was 0.56 (no CI reported).[251] There is no four year estimate for the UK as ivacaftor became available one year later in the UK than the USA.

The final study to look at the risk of exacerbations had a follow-up period of 30 months. In this study by Kirwan et al. the RR of exacerbations in the 30 months after starting

ivacaftor was 0.64 (no CI reported) compared to the 30 months prior to starting treatment.[249]

Seven studies estimated the effect of ivacaftor on the total number of IV days. One study was a RCT and it estimated the mean number of annual IV days to be 4.4 lower in those receiving ivacaftor (95% CI 0.7, 8.0).[254] The observational study by Hubert et al. estimated the number of IV days to decrease by 7.1 (95% 1.5, 12.7) in the first year of treatment compared to the year prior to starting treatment.[241, 255] Two other studies also estimated this difference: Guhaniyogi et al. -32.5 (95% CI -51.8, -13.2) and Nedevska et al. -12.6 (no CI reported).[231, 233] Two studies only included patients with severe lung disease and showed a more extreme decrease in the number of annual IV days of 41 (95% CI 16,67) and 46 (95% CI 21, 71).[223, 230] The study by Hubert et al. also estimated the difference between number of IV days in the second year of ivacaftor treatment compared to the year prior to treatment and showed a mean decrease of 7.3 days (95% CI 1.4, 13.2).[241, 255]

Five of the seven studies with total IV days as an outcome also investigated the effect of ivacaftor on just hospital IV days. The two studies of people with severe lung disease showed mean decreases in the annual number of hospital IV days of 24 (95% CI 7, 41) and 15 (95% CI 6, 24).[223, 230] However, none of the other studies found statistically significant effects of ivacaftor on this outcome: estimated decrease in annual hospital IV days in first year of treatment 1.2 (no CI reported, $p > 0.5$) and 3.7 (no CI reported, $p > 0.5$).[231, 241, 255] Similarly, the estimated effect in the second year of treatment was also not statistically significant, with an estimated decrease in the mean number of annual IV days of only 0.2 (no CI reported, $p > 0.5$).[241, 255] All of these results can be seen in Figure 7.5.

7.4.2 Lung Function

There were 26 studies measuring absolute change in ppFEV$_1$. The follow-up time of these studies ranged from just two weeks to four years, with most of the shorter studies being RCTs (range 2 weeks to 48 weeks) and the longer studies being observational (range 34 weeks to 4 years). All of the results from these studies can be seen in Figure 7.6.

The shortest studies were just two weeks long, but even over this short time period the observed treatment effect was still quite large: Accurso et al. showed a 9.8% increase in ppFEV$_1$ (95% CI -1.9, 21.5) and De Boeck et al. an 8.3% increase (95% CI 4.5, 12.1).[210, 232] The estimated effect was slightly larger at four weeks: De Boeck et al. 10.0% (95% CI 6.2, 13.8) and Edgeworth et al. 13.7% (95% CI 7.0, 20.3).[232, 250], but at eight weeks remained similar to four weeks: 10.7% (95% CI 7.3, 14.1).[232]

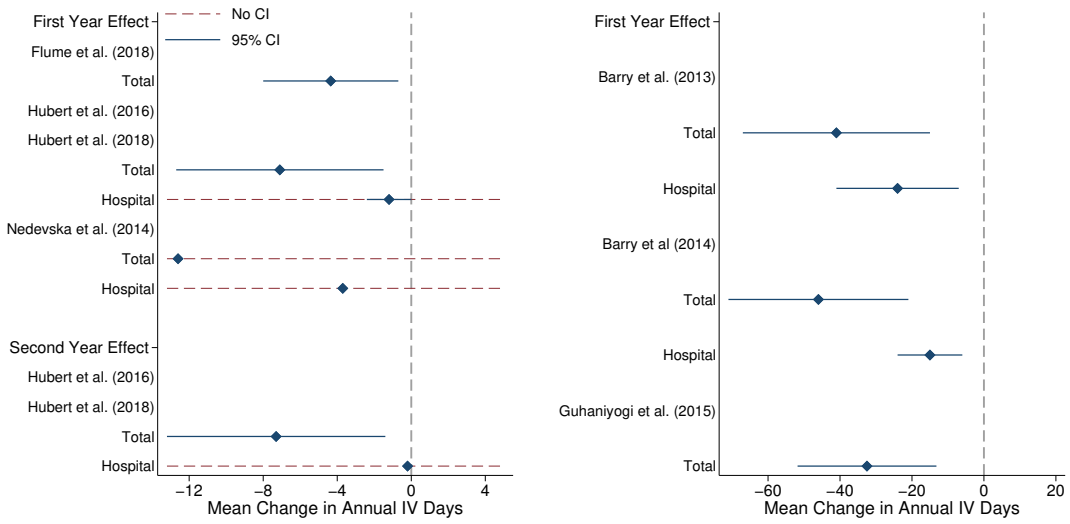


FIGURE 7.5: Forest plot of studies investigating the effect of ivacaftor on annual total and annual hospital IV Days (Some CIs are approximations based on reported *p*-values and larger estimates are plotted separately on right figure)

In studies with similarly short follow-up times of between 2 and 7 weeks, but restricted to people with baseline ppFEV$_1$ above 90%, the observed treatment effects were similar. At two weeks Davies et al. observed a 11.0% increase (95% CI 2.3, 19.7)[226], and slightly longer at 3.5 weeks a similar 12.8% (no CI reported) increase was seen in children aged between 6 and 11, but only a 5.9% (no CI reported) increase in people aged over 12, and by 7 weeks these had changed to 16.8% (no CI reported) and 5.9% (no CI reported) respectively.[227] At four weeks, in studies not stratified by age the estimated treatment effects were an increase of 7.2% (95% CI -2.0, 16.4), 7.0% (95% CI 1.6, 12.4), 5.5% (no CI reported), and 7.01% (95% CI 1.8, 12.2).[217, 219, 221, 222, 226, 227].

Three RCTs had 8 weeks, 12 weeks and 17 weeks of follow-up and respectively the estimates of the treatment effect from these studies were 10.7% (95% CI 7.3, 14.1), 9.3% (no CI reported) and 11.7% (95% CI 5.3, 18.1).[214, 232, 234]

Although many papers reported on the treatment effect observed at 24 weeks, they were all based on just two RCTs: one in people aged 6 to 11 and one in people aged over 12. In the former, the estimated treatment effect was 12.5% (95% CI 6.3, 18.7), and in the latter it was 10.6% (95% CI 5.3, 15.9).[211–216, 220, 224]

The shortest observational study estimated the treatment effect at 34 weeks and was restricted to people with ppFEV$_1$ less than 40%. The treatment effect estimated here was more modest with an increase of only 3.2% (95% CI 0.8, 5.6).[230]

The one-year treatment effect was estimated in two RCTs, again one involving children aged 6 to 11 and the other anyone aged 12 or over. The estimated treatment effect was

very similar in both groups, however, with the 6 to 11 year old estimated treatment effect being 10.0% (95% CI 4.3, 15.7) and those aged 12 or over having an estimated treatment effect of 10.5% (95% CI 5.3, 15.7).[211, 213, 214, 216, 224]

An observational study of patients in the Czech Republic estimated the one year treatment effect to be 3.6% (95% CI 0.1, 7.1).[239] In the USA, Sawicki et al. estimated the one-, two- and three-year treatment effects to be 8.3% (95% CI 3.4, 13.2), 9.9% (95% CI 4.0, 15.8), and 10.7% (95% CI 4.4, 17.0) respectively.[236] A separate study in the US estimated the one-, two-, three- and four-year effects to be 3.2%, 5.6%, 6.2% and 7.2% (no CI reported),[251] and in the UK slightly larger estimates were obtained for the one-, two- and three-year effects: 6.0%, 7.8% and 8.9%.[243, 246, 247, 251]

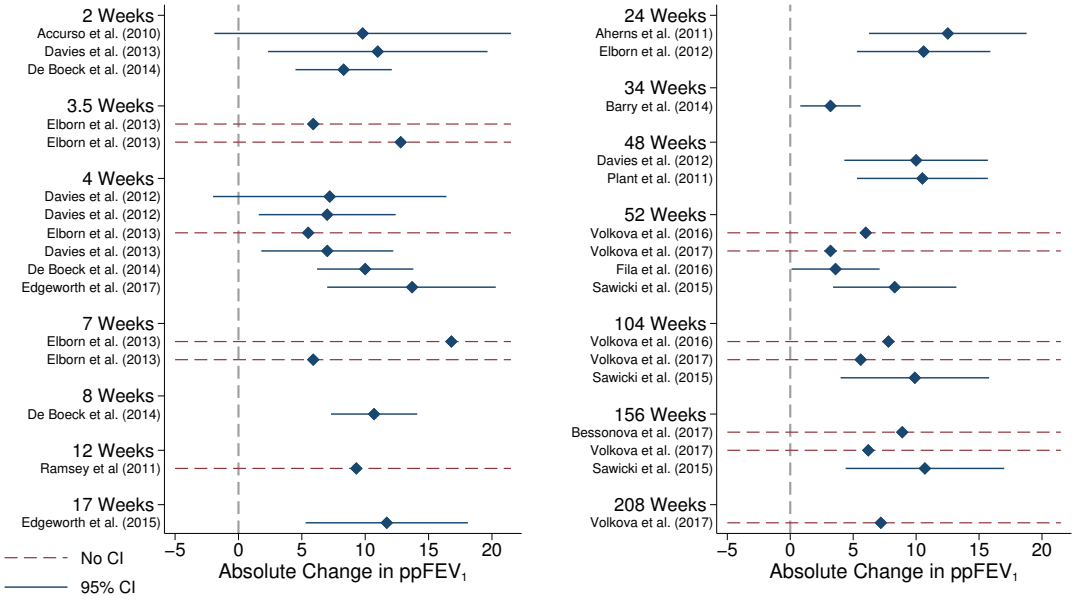


FIGURE 7.6: Forest plot of studies estimating the effect of ivacaftor on absolute change in $ppFEV_1$ (Some CIs are approximations based on reported p-values and papers reporting the same results are only plotted once)

Two RCTs reported the relative change in lung function as well as the absolute change.in lung function and these results are shown in Figure 7.7. The study by Aherns et al. was carried out in patients aged between 6 and 11 and showed a 17.4% (95% CI 8.7, 26.1) mean relative change in $ppFEV_1$ compared to placebo at 24 weeks.[212] In patients aged over 12, the study by Ramsey et al. showed a similar result of 16.9% (95% CI 13.6, 20.2) over 24 weeks, and this remained similar at 48 weeks: 16.8% (95% CI 13.5, 20.1).[214]

Only one study has so far estimated the effect of ivacaftor on the annual rate of lung function decline. This observation study by Sawicki et al. used data from the US CF Registry, and showed that the rate of decline in $ppFEV_1$ was 0.81 less per year (95% CI 0.08, 1.54) in those using ivacaftor.[236]

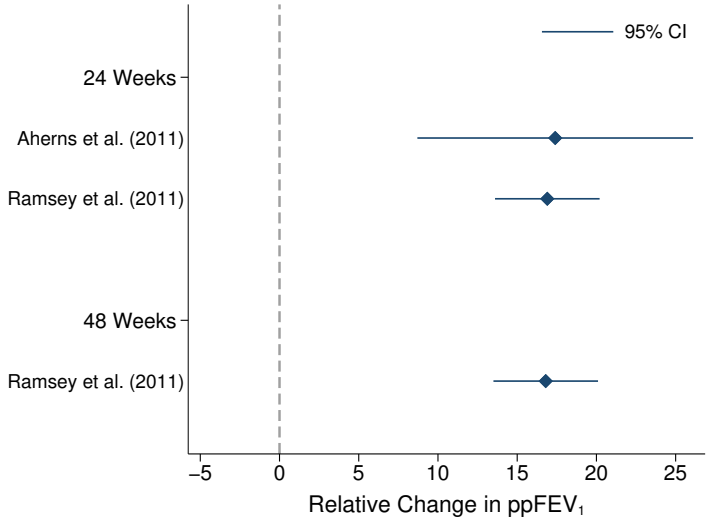


FIGURE 7.7: Forest plot of studies estimating the effect of ivacaftor on relative change in $ppFEV_1$ (Some CIs are approximations based on reported p-values)

7.5 Discussion

7.5.1 Findings

This systematic review shows that there is consistent evidence across a number of studies that ivacaftor is very beneficial in terms of all three outcomes studied. Although there are variable definitions of what constitutes an exacerbation, all but four of the studies estimated relatively similar RR of exacerbations, suggesting that ivacaftor reduces the risk of exacerbations by approximately half. This did not appear to change with time, although the only studies which reported the effects of more than one year of treatment did not provide estimates of the SE or 95% CI. The four studies which reported very different results were much smaller studies and the estimates had very large CIs. All four of these studies estimated more extreme RR, two of a drastic reduction in the risk of exacerbations when using ivacaftor and two estimating a small increase in risk of exacerbations, but in all four studies the large CIs overlapped with the results from the other studies.

All the studies that reported the effects of ivacaftor on lung function showed a positive effect. However, the results were more heterogeneous than the results on exacerbations, with some studies showing more modest benefits of approximately a 5% improvement in $ppFEV_1$ and others showing improvements larger than 15%. There did not appear to be any relationship between the length of follow-up and the estimated effect of treatment. However, the one study which did investigate the effect of ivacaftor on rate of lung function decline did suggest that the rate of decline was almost 1% less per year

once treated with ivacaftor, suggesting that the treatment effect at specific time points should grow over time.

Most studies included people with a wide range of characteristics, but there were some studies that only included children, adults, people with severe lung disease, or people with healthy lung function levels. Within all of these subgroups the effects of ivacaftor were positive, with no obvious differences in treatment effect between the groups.

7.5.2 Limitations

In RCTs there is always a group with whom to compare those treated, as by design approximately half of the people eligible for the RCT will not receive treatment. However, when considering observational studies, if we wish to be able to say something about the treatment effect, we want to be able to estimate what would have happened to people had they not received treatment. A lot of studies were excluded from this systematic review because they only compared people to their baseline values, but without any group with whom to compare them, we are unable to say anything about what might have happened to them had they not received treatment.

In the observational studies that were included in this review there were two main comparator groups. Some studies matched people receiving treatment to controls who are ineligible to receive treatment due to their genotype. Although matching will ensure that patients are comparable at baseline, it cannot ensure that the matched control will have a similar longitudinal change in variables that the person receiving treatment would have had, had they not received treatment. There is an implicit assumption that people with different CF-causing mutations matched on other baseline variables would have similar disease progression. This assumption is not acknowledged in any of the papers and if the assumption is not valid it would mean that the estimated treatment effects are not accurate.

Other studies used people as their own controls, comparing the a period of time prior to treatment to the time since treatment. This was more common for the outcome of exacerbations, where the risk of exacerbations in the year prior to treatment was compared to the risk of exacerbations since treatment was initiated. Again, however, there is an implicit assumption here that is not mentioned in the papers, that there have been no other changes in health care in this time period. Over a short time period, it may be reasonable that such an assumption would not be violated, but when looking longer term, it becomes more likely that there could have been other changes that could affect the health of individuals not related to the treatment.

Although we have only considered studies that provided ivacaftor to people with gating mutations, we have not taken into account that some studies only included one specific

gating mutation. We have therefore assumed that the effect of ivacaftor is comparable no matter the specific gating mutation.

7.5.3 Conclusions

Although ivacaftor is a relatively new treatment there have already been many studies published to investigate its effects and all of them have shown very strong positive results. However, given the newness of the treatment, there is still no conclusive evidence of its long-term effectiveness. Furthermore, many of the observational studies initially found through the search of this systematic review were excluded as they provided no estimate of the causal effect of treatment due to having no control group.

Compared to the number of studies from the systematic review of DNase, there has clearly been an increase in interest in using observational data to estimate the effects of long-term treatment use, but it is unfortunate that in many cases the methods used are inadequate to estimate a causal effect of treatment in an unbiased way.

Implications

- All studies show very promising results, but only a few studies have estimated the effects of more than one year of treatment
- As almost all people who are eligible for ivacaftor are now receiving it, so it is important to consider who we can compare them with to be able to accurately estimate treatment effects

Chapter 8

Statistical Methods to Estimate Effects of Ivacaftor

8.1 Introduction

The systematic review in the previous chapter demonstrated that ivacaftor has shown very promising results so far. However most of the studies only followed patients for a year after initiating ivacaftor, with just a few studies with longer follow-up out to four years. In real world practice, people would generally be expected to continue receiving ivacaftor indefinitely for the rest of their lives, so it is important to be able to continue to estimate the effectiveness of the treatment long term.

Clearly, it is neither feasible nor ethical to continue to withhold treatment from some of the eligible population which is what would be necessary to estimate the long-term treatment effect using a RCT. However, in order to estimate a treatment effect, we do need to have an estimate of what would happen to people long term if they were not receiving ivacaftor, so that we can compare this to what has happened to people since they have been receiving the treatment. Because everyone who is eligible to receive ivacaftor is now receiving it, we no longer have a control group with whom to compare the treatment group.

In this chapter, we will introduce two possible comparator groups that have been identified and that could be used to compare outcomes with long-term follow-up on ivacaftor to those with long-term follow-up without ivacaftor. We also discuss the assumptions underlying the use of these comparator groups. We introduce the use of negative controls, which can be used to test the validity of the assumptions we must make in order for the analyses to be correct. Finally, we introduce possible statistical analysis methods that could be used when comparing these groups.

8.2 Comparator Groups

As seen in the systematic review, in the absence of direct control patients, most observational studies have compared people now receiving ivacaftor either to people not receiving ivacaftor because they do not have an eligible genotype, or to the people eligible for ivacaftor in the time period prior to when ivacaftor was available. These two comparator groups can be seen summarised in Figure 8.1. Ivacaftor is only available for people in group B of Figure 8.1, this is the group of patients with an eligible genotype for ivacaftor ($G = 1$) in the time period since ivacaftor has been available ($P = 1$) . The data from the patients in group B allow us to measure what has happened to them since they started to receive ivacaftor. However, to estimate the treatment effect, we also need to know what would have happened to them had they not received treatment. Because everybody in group B is receiving ivacaftor, we cannot directly measure this.

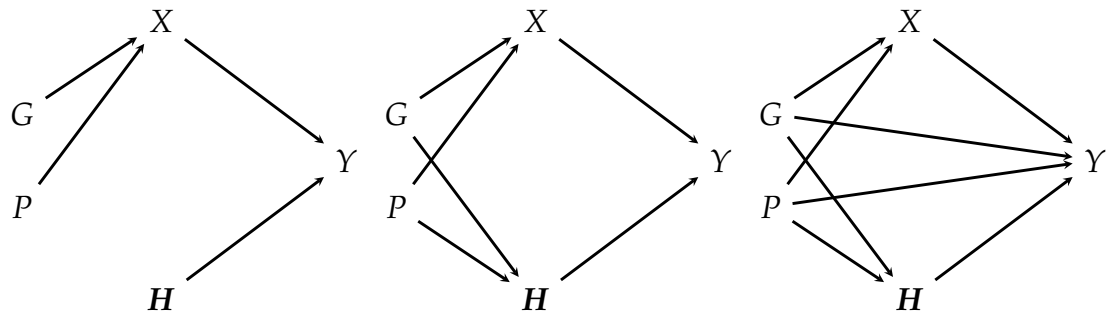
			Ivacaftor	
			Pre-	Post-
			$P = 0$	$P = 1$
Genotype	Gating	$G = 1$	**A**	**B**
	Other	$G = 0$	**C**	**D**

FIGURE 8.1: Division of CF population into four groups based on genotype (rows) and time period (columns)

The other three groups of Figure 8.1 are not receiving ivacaftor, either because it was not yet available (group A), or because they do not have an eligible genotype (group D), or due to both of these reasons (group C). If we are willing to assume that all four of these groups are comparable in terms of their long-term health trajectories, then we can simply observe what happened in these groups (A, B and C) of patients over the same length of follow-up as those in group B. We would then assume that what we observe in these groups is what would have happened to those in group B had they not received ivacaftor, and any difference observed between group B and the other groups would then be attributed as the treatment effect of ivacaftor. We define this estimate of the treatment effect, the 'Naïve Treatment Effect' (NTE); it is an estimate of the treatment effect assuming that these different groups of patients are directly comparable. A more precise definition of the NTE and the other treatment effects we consider is given in Section 8.3.1.

For example, for patients in group B who started receiving ivacaftor in 2012 and were followed-up until 2016, we wish to estimate the effect of receiving ivacaftor for four years compared to not receiving ivacaftor for four years. We could therefore compare them to the four-year follow-up observed in group D, again from 2012 to 2016. Simply comparing these two groups would assume that people with different genotypes have on average

the same disease severity, i.e. that the genotype does not affect the outcome of interest. This scenario is reflected in the DAG shown in Figure 8.2a. Here genotype is assumed to have no effect except on the eligibility to receive ivacaftor. Thus in this setting, the final outcome of interest could simply be compared between the two genotype groups.



(A) (B) (C)

FIGURE 8.2: DAGs showing potential causal pathways between ivacaftor (X), Genotype (G), time Period (P), covariates measuring Health at baseline ($\boldsymbol{H}$) and outcome (Y)

A number of studies have looked at comparing the long-term health between people with different genotypes and overall the findings suggest that people with a gating mutation eligible for ivacaftor have similar lung function decline and survival as people with a class II mutation who form the majority of people with CF.[256, 257] This would suggest that the causal pathways shown in Figure 8.2a may be valid. However, we also present two other possible DAGs. In Figure 8.2b, genotype is now assumed to affect health at baseline (which we take to be measured through observed covariates), but after accounting for these baseline differences, genotype is assumed to not directly affect future health. In this setting, genotype would be a confounder of the treatment effect that we wish to estimate, but rather than adjusting for genotype (which is not possible due to perfect collinearity with ivacaftor), adjustment for the baseline covariates ($\boldsymbol{H}$) would be sufficient to control for this confounding, because adjusting for $\boldsymbol{H}$ would block the path from G to Y not through X. (See Section 8.3.2 for details on the situation when there are additionally some unmeasured baseline determinants of health).

The final scenario presented in Figure 8.2c shows genotype being directly associated with the outcome of interest. In this situation, the NTE obtained from comparing group B to group D would no longer be an unbiased estimate of the treatment effect, because it would not be possible to correct for the differences between these groups due to genotype. For the NTE to be an unbiased estimate of the treatment effect, we must therefore assume that after adjustment for baseline differences, genotype has no effect on the outcome of interest, i.e. the scenarios shown in Figure 8.2a or Figure 8.2b. Furthermore, in the scenario shown in Figure 8.2b, we must adjust for $\boldsymbol{H}$, whereas this would not be necessary in the scenario shown in Figure 8.2a.

As well as comparing those currently receiving ivacaftor (group B) to those not eligible for ivacaftor (group D), we could also compare outcomes in group B to outcomes in group A. This would compare the four years since people have been receiving ivacaftor (2012 to 2016) to the four years directly preceeding the introduction of the ivacaftor (2008 to 2012) in the same group of patients. In this situation, we face the same predicament as in the comparison of those eligible and not eligible for ivacaftor. The same three scenarios are presented again in Figure 8.2, where in this analysis the time period (P) is a perfect predictor of ivacaftor. We can adjust for any diffences between people at baseline, i.e. whether the average health of people differed in 2008 compared to 2012, but must then assume that the outcome trajectories have not changed between 2008-2012 and 2012-2016. Over a longer period of time, we know that there have been improvements in the survival and average health of the CF population thanks to the introduction of other treatments and improvements in health care practice.[258] However, whether there would be an observable effect over the eight year time period of our analysis is uncertain. For this reason, we would ideally like to be able to test these assumptions of the comparability of group B to group A, and also the comparability of group B to group D.

8.3 Negative Controls

For either of the analyses introduced in the previous section to provide unbiased treatment estimates, we must assume that either genotype or time period has no direct effect on our outcome of interest. This assumption could be valid for some outcomes and not for others and therefore it would be good to be able to test the validity of these assumptions.

The use of a negative control analysis is a technique often utilised in biological laboratory studies, where all analyses are repeated under identical conditions except that the active substance of interest is not used. The results of the negative control analysis are therefore excepted to be null, and if this is the case, it suggests that the results observed in the actual analysis are due to the active substance of interest and not due to any other potential sources of bias in the experiment.

Negative controls have not often been used in epidemiological studies, but it has been proposed that they could be a useful tool in detecting bias in analyses, especially bias due to unobserved confounding, which is often a key issue in observational studies.[259] In our setting, we do not technically have any unobserved confounders of the association between treatment and outcome, as we have observed for everyone the reason why they are or are not receiving ivacaftor (i.e. their genotype and time period). However, because these two variables cannot be accounted for in the analysis, we do wish to learn whether there may be any bias from ignoring them in the analyses.

In the literature on the use of negative controls in epidemiological studies, two types of negative controls have been identified: negative exposure controls and negative outcome controls.[259–261] A negative outcome control would be an outcome that is believed to not be affected in any way by the exposure of interest.[262] As ivacaftor is generally believed to be a 'disease-modifying' treatment, it is possible that the treatment affects many different outcomes, and it is unlikely to be plausible to assume that ivacaftor would have no effect on any of the health data collected in the registry.

We instead propose the use of a negative control exposure. A negative control exposure is an exposure which is believed not to affect the outcome of interest, but except for this is as similar as possible to the exposure of interest. In our setting, we have had to assume that neither genotype nor time period is a cause of the outcome of interest except via its effect on ivacaftor use. If we can identify groups that differ by genotype and/or time period, but where neither group receives ivacaftor, we can test the assumptions that these have no effect on the outcome. In the previous section, we identified two analyses to estimate the NTE of ivacaftor: compare group B to group A, or compare group B to group D. In neither case did we utilise group C, but this group can be used to assess whether these four different groups of patients are comparable except for ivacaftor.

In the comparison of group B to group A, we assess how the outcome changed at the end of the four year period prior to ivacaftor compared to the four year period since its introduction. To test the assumption that time period has no effect on the outcome after adjustment for baseline covariates, we can perform the same analysis, but compare group D to group C, i.e. compare the same time periods, but in a group of patients where ivacaftor was not introduced. We define the difference between these groups (or similarly the difference between group A and group C when looking at whether there is an effect of genotype) as the negative control effect (NCE). For the NTE estimate to be an unbiased estimate of the true treatment effect, we expect the NCE estimate to be null.

Similarly, in the analysis comparing group B to group D, we assume that genotype has no effect on outcomes after adjustment for baseline covariates. This can be tested, under the assumption that the effect of genotype on outcome is the same in both time periods, by performing the same analysis, but comparing group A to group C, i.e. comparing whether genotype had an effect on the outcome before the introduction of ivacaftor. If this result were non-null, it would suggest that genotype has a direct effect on the outcome of interest (as shown in Figure 8.2c) and that the NTE estimate from the main analysis is biased.

8.3.1 Negative-Control-Corrected Treatment Effect

The ideal result would be for the NCE to be zero, as this would suggest that the NTE is an unbiased estimator of the true treatment effect. However, in reality, the estimate

of the NCE will never be exactly zero, and even if it were, there would still be some uncertainty around the estimate. To obtain a more reliable estimate of the true treatment effect, we should therefore account for both the NCE estimate and its SE. The simplest way to do this is to subtract the NCE estimator from the NTE estimator. We define this as the negative-control-corrected treatment effect (NCCTE) estimator:

$$\text{NCCTE} = \text{NTE} - \text{NCE} \tag{8.1}$$

This 'difference in differences' approach will be valid provided that the magnitude of the effect not due to ivacaftor is the same in both populations.[263] For example, for the comparison of genotype groups, we would assume that any difference observed between the groups does not change over time. Similarly, for the comparison of the same individuals before and after the introduction of ivacaftor, we would assume that any change over time observed in those ineligible for ivacaftor is the same as any change in those eligible for ivacaftor apart from the influence of ivacaftor. Both of these assumptions are weaker than the assumptions necessary for the NTE estimate to be an unbiased estimate of the true treatment effect, as there we had to assume that there was no effect of time period or genotype except the effect mediated through ivacaftor use. With the NCCTE estimator, we allow for differences in the outcome by time period and genotype, but assume that the differences are the same in the negative control comparison as in the comparison between the groups receiving and not receiving ivacaftor.

The variance of the NCCTE estimator can be calculated as follows:

$$\text{Var(NCCTE)} = \text{Var(NTE} - \text{NCE)} = \text{Var(NTE)} + \text{Var(NCE)} - 2\text{Cov(NTE,NCE)}. \tag{8.2}$$

This means that in general we would need to know the covariance between the NTE and the NCE. As estimates of the NTE and NCE are obtained from two separate models, we will not generally be able to estimate this quantity analytically and thus a bootstrap method will usually be necessary to obtain estimates of the variance of the NCCTE. However, in the analysis comparing the time period since ivacaftor to the time period before ivacaftor, the NCE groups are completely independent of the NTE groups (i.e. the same person cannot be in both analyses, because people cannot change genotype). As such, the covariance of the two estimates is expected to be zero, meaning the variance of the NCCTE will simply be the sum of the variances of the NTE and the NCE. This is not true in the case of the analysis comparing those eligible for ivacaftor to those ineligible for ivacaftor as the same individual can be in both the NTE analysis and the NCE analysis.

In summary, we have proposed two different analyses, each with a different negative control:

- Analysis 1:

  - Naïve Treatment Effect:
    - Compare outcomes post-ivacaftor to outcomes pre-ivacaftor in those eligible for ivacaftor, i.e. group B vs group A
    - $\text{NTE}_P = \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 1, G = 1) - \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 0, G = 1)$
  - Negative-Control Effect:
    - Compare outcomes post-ivacaftor to outcomes pre-ivacaftor in those ineligible for ivacaftor, i.e. group D vs group C
    - $\text{NCE}_P = \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 1, G = 0) - \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 0, G = 0)$
  - Negative-Control-Corrected Treatment Effect:
    - $\text{NCCTE}_P = \text{NTE}_P - \text{NCE}_P = \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 1, G = 1) - \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 0, G = 1) - \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 1, G = 0) + \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 0, G = 0)$
  - Assumptions:
    - NTE_P assumes that P does not affect Y after adjustment for $\boldsymbol{H}$
    - NCCTE_P assumes that the effect of P on Y is not modified by G

- Analysis 2:
  - Naïve Treatment Effect:
    - Compare outcomes in those eligible for ivacaftor to outcomes in those ineligible for ivacaftor in the four years since ivacaftor was introduced, i.e. group B vs group D
    - $\text{NTE}_G = \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 1, G = 1) - \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 1, G = 0)$
  - Negative-Control Effect:
    - Compare outcomes in those eligible for ivacaftor to outcomes in those ineligible for ivacaftor in the four years prior to the introduction of ivacaftor, i.e. group A vs group C
    - $\text{NCE}_G = \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 0, G = 1) - \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 0, G = 0)$
  - Negative-Control-Corrected Treatment Effect:
    - $\text{NCCTE}_G = \text{NTE}_G - \text{NCE}_G = \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 1, G = 1) - \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 1, G = 0) - \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 0, G = 1) + \text{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 0, G = 0)$
  - Assumptions:
    - NTE_G assumes that G does not affect Y after adjustment for $\boldsymbol{H}$
    - NCCTE_G assumes that the effect of G on Y is not modified by P

It can be seen that the NCCTE estimator is the same in both analyses, but this will only be an unbiased estimator of the true treatment effect provided that the assumption of no interaction between genotype and time period on the outcome is valid. Conversely, the NTE estimates are different between the analyses, and will only both be unbiased if there is no direct effect of either time period or genotype, i.e. $\mathrm{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 0, G = 1) = \mathrm{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 1, G = 0) = \mathrm{E}(Y|\boldsymbol{H} = \boldsymbol{h}, P = 0, G = 0)$.

8.3.2 Unmeasured Covariates

As well as correcting for any direct effect of genotype or time period on the outcome, the use of the negative controls can also correct for any effects of unmeasured covariates on the outcome. Figure 8.3 shows three possible assumed causal pathways where now some of the baseline covariates are unmeasured, $\boldsymbol{U}$. In Figure 8.3a, neither genotype nor time period affect any of the baseline covariates, measured or unmeasured, nor do they directly affect the outcome. In this situation, the use of negative controls would not be required to obtain an unbiased estimate of the treatment effect. However, in Figure 8.3b, genotype and time period are now assumed to affect the baseline covariates. Previously, when we had no unmeasured covariates, as shown in Figure 8.2b, we still did not require the use of a negative control, as conditioning on the baseline covariates, $\boldsymbol{H}$ would have been sufficient to control for the confounding by genotype and time period, and the NCE estimate would be expected to be null. In the situation where we no longer assume that all the baseline covariates have been measured, we would need to use negative controls to obtain an unbiased estimate of the treatment effect. In this setting, the NCE estimate would be the indirect effects of genotype and time period through the unmeasured baseline covariates. In the setting shown in Figure 8.3c, the NCE would now be both the direct effect of genotype and time period on the outcome, as well as the indirect effects through the unmeasured baseline covariates.

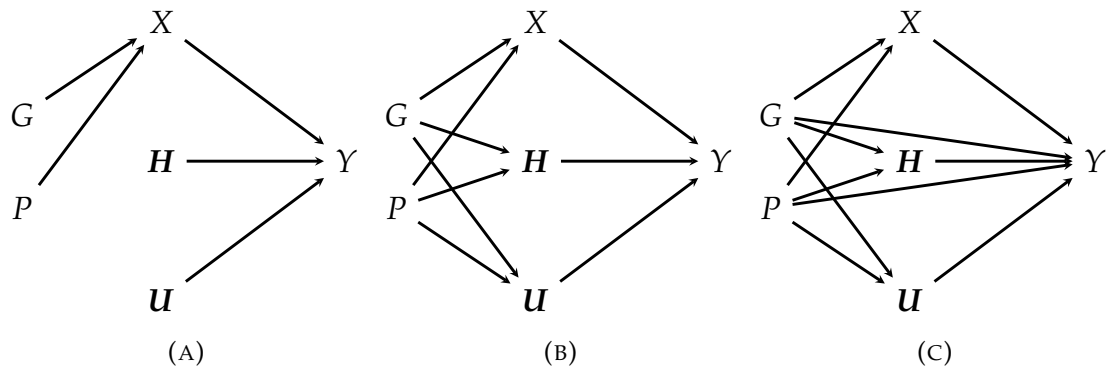


FIGURE 8.3: DAGs showing potential causal pathways between ivacaftor (X), Genotype (G), time Period (P), measured covariates of Health at baseline ($\boldsymbol{H}$), Unmeasured covariates of health at baseline ($\boldsymbol{U}$) and outcome (Y)

This suggests that even if we are willing to believe that genotype and time period have no direct effect on the outcome after account for baseline differences between groups, the use of a negative control is still beneficial as we no longer need to assume that we have measured all of the important baseline variables.

8.3.3 Combined Model

Analyses 1 and 2 described above involve fitting four models. Rather than perform each of these analyses separately, it is also possible to combine them into one analysis. One of the main benefits of comparing all groups simultaneously is that ivacaftor is no longer perfectly collinear with either time period or genotype, meaning that the negative controls can be included in the main model and the NCCTE estimate will then be directly estimated from the model.[264] For example, using all four groups, we could fit the following model:

$$\mathrm{E}(Y|H=h,P=p,G=g,X=x)=\beta_0+\boldsymbol{\beta_h}\boldsymbol{H}+\beta_p P+\beta_g G+\beta_x X, \tag{8.3}$$

where $X = P \times G$.

Here β_x corresponds to the NCCTE, as both genotype and time period have been accounted for through β_p and β_g. However, we are still making the same assumption, as before, that there is no interaction between P and G and it would not be possible to include an interaction term between P and G, as $P = 1, G = 1 \iff X = 1$. This analysis also makes the further assumption that the effect of $\boldsymbol{H}$ is common among all four groups, whereas in the previous analyses this effect could be different between analyses 1 and 2.

By combining all the groups into one model we increase our sample size which may improve the efficiency. However, the NCCTE must now account for the uncertainty in both the effect of genotype and the effect of time period, and therefore any improvements in the variance of the estimates due to increases in sample size may be offset this.

8.4 Analysis Methods

Due to the longitudinal nature of the registry data, it would be common to use mixed-effects models to account for the variability between individuals. However, as we are only interested in the average effect of treatment, it would actually be possible to use simpler methods and still obtain unbiased estimators of the treatment effect. In this section, we will first introduce two simpler methods: marginal models and fixed-effects models, before exploring the potential benefits and drawbacks of using mixed-effects models. We introduce the methods with a continuous outcome, and will end the section with a

discussion of any additional complexities when using count outcomes. All of the examples below are presented for the example of estimating the NTE, but by definition, the negative-control analyses should be identical to the main analysis, except that treatment, X, would be replaced by the negative control exposure, G or P.

8.4.1 Marginal Models

The simplest form of a marginal model is a generalised estimating equation assuming an independent working correlation matrix, which is simply the ordinary least squares estimator obtained through standard regression procedures. For example,

$$Y_{ij} = \beta_0 + \beta_x X_{ij} + \beta_{xt} X_{ij} T_{ij} + \beta_t T_{ij} + \boldsymbol{\beta_h} \boldsymbol{H}_{ij} + \epsilon_{ij}, \tag{8.4}$$

where i refers to an individual, j refers to the visit number, T refers to time elapsed since baseline (defined as the start of the pre-ivacaftor period or the start of the post-ivacaftor period) and ϵ is the residual error.

First we define the two treatment effects that we are interested in estimating. At the first visit after treatment was initiated ($j = 1$), we hope that the treatment will have resulted in an acute improvement in the outcome. This is the estimate given by β_x, which we refer to as the step-change treatment effect. Furthermore, if there is a decline in the outcome over time, as estimated by β_t, we hope that treatment can slow this decline, and this would be estimated by β_{xt}, which we refer to as the slope-change treatment effect.

Overall, the model specified in Equation 8.4 should provide an unbiased estimate of the NTE, because j is not a confounder, i.e. receipt of the treatment is determined only by genotype and time period, and not by additional individual characteristics. Furthermore, it is expected that the baseline covariates contained in $\boldsymbol{H}$ can explain a lot of the variance in outcome between individuals. However, if i remains a strong predictor of Y after adjustment for $\boldsymbol{H}$, then although this method will on average give unbiased results, it will be very inefficient, as the error term ϵ would be much larger than if i were somehow accounted for.

Using standard regression methods assumes an independent working correlation matrix, i.e. that observations from the same individual are not correlated. Violations of this assumption only impact the efficiency of the estimation of the average treatment effect, but do not result in bias.[265] A simple alternative is to assume exchangeability, which is that all observations are equally correlated to one another. The benefit of assuming independent or exchangeable working correlation matrices is that the observations do not have to happen at specific times, so the exact time since baseline can be used for every observation. Other types of working correlation matrix can be used, such as designs which allow observations that are closer together in time to be more strongly correlated to one

another than observations further apart. However, these matrices now require visits to happen at discrete intervals. In the UK CF Registry, visits are supposed to happen every twelve months, so this may be a possibility, but any efficiency gained from better specification of the working correlation matrix could be offset by ignoring the exact amount of time elapsed since baseline.

Finally, it should be noted that the standard model-based estimates for the SE would be incorrect as they would not account for the non-independence of observations from the same individual. For this reason, it would be necessary to use a robust estimator of the SE.

8.4.2 Fixed-Effects Regression

If between individuals there are large differences in the outcome even after accounting for baseline covariates, then it is possible to use fixed-effects regression to further correct for baseline differences between individuals. For this, we consider models of the form:

$$Y_{ij} = \boldsymbol{\beta}_i + \beta_x X_{ij} + \beta_{xt} X_{ij} T_{ij} + \beta_t T_{ij} + \epsilon_{ij}. \tag{8.5}$$

The main difference between Equation 8.4.2 and Equation 8.4 is that the single intercept term, β_0, has now been replaced by a separate intercept term for each person, $\boldsymbol{\beta}_i$. However, another difference is that we no longer include the baseline covariates, $\boldsymbol{H}$, in this model. This is because every person now has a unique estimate for their 'average' outcome, $\boldsymbol{\beta}_i$, and the baseline variables $\boldsymbol{H}$ do not vary within an individual. Thus, the effects of each individual baseline variable can no longer be estimated, but if we are only interested in estimating the treatment effects then this is not an issue.

As with marginal models a robust estimator should be used to calculate SEs.

8.4.3 Mixed-Effects Regression

Rather than estimating a separate intercept term for every individual, mixed models only estimate one intercept term and then also estimate how each individual varies around this average. It is necessary to make an assumption about the distribution of these random effects, and the most common assumption would be that they are normally distributed with mean zero and a common variance. Then, the mixed model only estimates one intercept term and one variance term, rather than having to estimate a separate intercept term for each individual.

It is common to not only include a random intercept, but also a random slope, for example:

$$Y_{ij} = \beta_0 + \beta_x X_{ij} + \beta_{xt} X_{ij} T_{ij} + \beta_t T_{ij} + \boldsymbol{\beta_h} \boldsymbol{H}_{ij} + u_{i0} + u_{it} T_{ij} + \epsilon_{ij}, \tag{8.6}$$

where u_{i0} is the random-intercept term and u_{it} is the random-slope term allowing for how different each individual is from average, β_0 and β_t respectively.

In a fixed-effects model, the intercepts are essentially estimated non-parametrically, i.e. there is no distribution assumption behind the estimates. However, in the case of a mixed-effects model, we must now specify the distribution of the effects, and a misspecification of this could lead to bias in the estimated treatment effect, but would lead to a gain in efficiency if it is correctly specified. One example would be that rather than a common variance among all individuals as stated above, there could actually be less variability among people receiving ivacaftor. For example a very effective treatment could result in everybody having a healthy lung function and thus little variability between people compared to those not receiving treatment. This type of heteroscedasticity can easily be incorporated into mixed models, by choosing to estimate two separate variances for those receiving and not receiving ivacaftor, but it highlights how there are more modelling decisions that need to be made when using mixed-effects models compared to fixed-effects models or marginal models.

Finally, unlike with marginal and fixed-effects models, the standard model-based estimates of the SE are consistent estimates of the true SE.

8.4.4 Models for Count Outcomes

All three methods introduced above can easily handle count outcomes, but there are a number of technicalities which must be addressed.

Firstly, as previously discussed in Section 4.8, it is important to remember that due to non-collapsibility, there is no unique treatment effect estimate with non-linear models. Fixed- and mixed-effects models are conditional on the individual and estimate the within-individual treatment effect, i.e. an estimate of what would be observed on average to one person if they did or did not receive treatment. Conversely, marginal models do not condition on the individual and thus will estimate the population average treatment effect. If no other covariates are included in the model, this would be the average effect seen over the whole population if everyone did or did not receive treatment, but if other baseline covariates are included in the model, the estimate would be conditional on these covariates. This is also an issue when comparing fixed- and mixed- effects models, as when using a fixed-effects model we cannot condition on baseline covariates, whereas this can be done in a mixed-effects model. Thus, even these two within-individual

treatment effect estimates will not be expected to be the same. All of these treatment effect estimators can be different whilst all still being correct and this makes comparisons between methods more difficult. It is therefore important to carefully consider what each method is actually estimating and what is the treatment effect we are interested in estimating.[266]

In Part II, we modelled the annual number of IV days from a zero-inflated negative binomial distribution, and in the case of a marginal model, this is still a possibility. However, fixed-effects models should not be used with logistic regression (which is used for the zero-inflated part of the distribution). This is because a lot of individuals can easily end up with no variability in outcome between observations leading to estimation problems.[267] Fixed-effects models can, however, be used for other types of non-linear models, such as a negative binomial model.[268] This means that this method could be used if we were willing to ignore the excess of zeroes. It is also computationally challenging to fit a mixed-effects model under a zero-inflated negative binomial distribution, and there are no packages currently available in Stata which can fit this type of model.

The models for count outcomes will use a log link function. The NTE and NCE estimates can then be expressed in terms of an IRR by taking the exponential of the coefficients estimated from the models. In this case, we are again hoping that the coefficient for the NCE estimate is null, i.e. an IRR of 1. The NCCTE estimate can then be obtained either on the log scale:

$$\log\left(\text{NCCTE}\right) = \log\left(\text{NTE}\right) - \log\left(\text{NCE}\right), \tag{8.7}$$

or directly as an IRR:

$$\text{NCCTE} = \frac{\text{NTE}}{\text{NCE}}. \tag{8.8}$$

8.4.5 Comparison of Methods

Under the assumptions listed, all three methods should provide consistent estimators of the treatment effect, and therefore one of the key criteria when assessing the methods is their efficiency. As discussed in Section 8.4.1, using standard regression methods will ignore the variability between individuals, which could lead to inefficiencies. However, if a lot of the variability between individuals can be explained by other baseline covariates, there may not be a noticeable decrease in the efficiency of this method compared to the other two methods. In general, mixed-effects models require a large number of people to be able to accurately estimate the random-effects, but this should not be an issue with the UK CF Registry. However, this could cause issues with fixed-effects regression where

a separate intercept will need to be estimated for each of the thousands of individuals. Furthermore, each individual only has a small number of observations, making it difficult to accurately estimate the fixed intercepts, which may affect the efficiency for estimating the treatment effect.

Another key consideration is computation time. In general, the computation time required for marginal or fixed-effects models will be noticeably quicker than that required for mixed-models. If only fitting one mixed-model, this might not be much of an issue. However, in order to estimate the NCCTE estimate, we generally need to use a bootstrap procedure, meaning that any differences in computation time will be exacerbated. For non-linear models, this is even more of an issue, because they are slower to fit, and there can often be difficulties with achieving convergence for these models.

In the next chapter, when analysing the UK CF Registry to estimate the effects of ivacaftor, we will use all three available methods to compare results. In the case of a continuous outcome, such as lung function, we would expect all three methods to give similar results, which could highlight issues if there are any stark differences between methods. We can then assess whether the additional computation time for mixed-models has resulted in noticeable gains in reducing the SE of the treatment effect estimates. For count outcomes, such as IV days, the estimates cannot be compared as the marginal and conditional estimates are expected to be different.

Chapter 9

Estimating the Effects of Long-Term Ivacaftor Use

9.1 Introduction

The systematic review in Chapter 7 showed that there is a lot of interest in ivacaftor and specifically in its long-term effectiveness to assess to what extent it is 'disease-modifying'. The RCTs for ivacaftor showed very impressive results with a maximum follow-up time of 48 weeks, and since then ivacaftor has generally been initiated in all eligible patients.

As previously discussed, to be eligible for ivacaftor CF patients must generally have a gating mutation. There have also been studies that have investigated the effectiveness of ivacaftor in people with a residual function mutation[269], but it is not generally available for these people in the UK. This means the UK CF Registry could not be used to investigate the effectiveness of treatment in this group of patients.

For people with a gating mutation, ivacaftor was initially introduced only in those patients over six years old. Since then, the minimum age has been lowered to two, but there is still only a small amount of data for these younger patients in the UK CF Registry, especially as the lung function measures are not typically used in children under six. For this reason, we restrict the analyses in this chapter to people over six years old.

Almost all people with CF in the UK who are aged over six and have a gating mutation have been receiving ivacaftor since 2012.[270] This means we currently have four years of follow-up data (up to the end of 2016). One of the key benefits of such long-term data is in terms of lung function outcomes, where it is hoped that a disease-modifying treatment can not only provide a step-change increase in lung function, but can also slow the rate of lung function decline. This latter treatment effect is hard to measure with short-term data. Figure 9.1 shows the two treatment effects we wish to estimate.

Since the introduction of ivacaftor, there have been a number of studies using observational data aiming to estimate it long-term effects. Due to the fact that almost all eligible

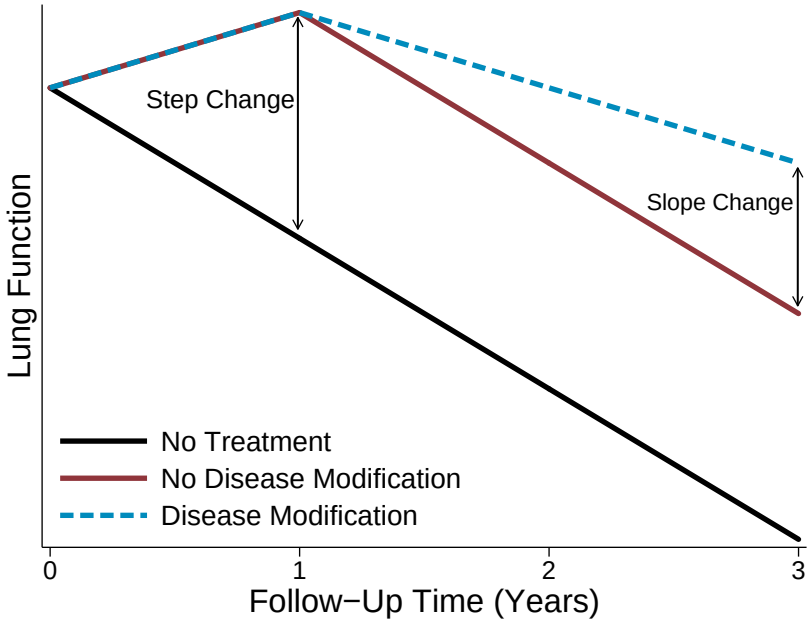


FIGURE 9.1: Possible effects of ivacaftor on lung function in terms of a one-off step-change or a long-term slope-change

people are now receiving treatment, most of these studies have compared people currently receiving treatment either to people without a gating mutation or to people with a gating mutation but in the time period prior to the availability of ivacaftor. Steps to ensure that these groups are comparable at baseline are often taken, such as adjusting for or matching on baseline characteristics, but after this there is the assumption that these two separate groups of patients would have had similar long-term health trajectories were it not for ivacaftor. Studies have suggested that people with and without gating mutations do have similar long-term health trajectories, suggesting that these comparisons may be valid.[256, 257] However, it is known that the health of patients with CF has been improving with time, suggesting that pre- & post- comparisons may not be valid depending on the length of follow-up being considered.[258]

In this chapter, we aim to estimate the effect of ivacaftor on lung function and the rate of annual IV days. In the case of lung function, we wish to estimate the step-change effect as the effect of ivacaftor on lung function after receiving treatment for one year, and also the slope-change effect as the change in annual decline over four years. For annual IV days, we estimate the IRR of IV days after using ivacaftor for one, two and three years. We perform two analyses: 1) compare those currently receiving ivacaftor to people not currently receiving ivacaftor due to not having a gating mutation, and 2) compare those currently receiving ivacaftor to those with a gating mutation in the period prior to the availability of ivacaftor. In both cases these analyses estimate the NTE. Further to this, we use the negative control exposures as defined in Chapter 8 to estimate the NCE, which

can be used to test the assumptions of the comparability of the group currently receiving ivacaftor to the other groups. Finally, we use the NTE and NCE to estimate the NCCTE, correcting for any differences between the groups not due to ivacaftor.

9.2 Methods

We divide the UK CF Registry data into four groups: A) People with a gating mutation in years 2008 to 2012, B) People with a gating mutation in years 2012 to 2016, C) People without a gating mutation in years 2008 to 2012, D) People without a gating mutation in years 2012 to 2016. This is shown in Figure 9.2.

			Ivacaftor	
			Pre-	Post-
			$P = 0$	$P = 1$
Genotype	Gating	$G = 1$	**A**	**B**
			I=437	I=397
			n=1,763	n=1,765
	Other	$G = 0$	**C**	**D**
			I=6,382	I=7,378
			n=25,449	n=31,759

FIGURE 9.2: Number of people (I) and total number of observations (n) per group

The following exclusion criteria were applied to observations, which resulted in some people being entirely excluded if all their observations were excluded: observations post-transplant (58 people excluded), observations where people were aged under six (1,479 people excluded), observations missing lung function measures (165 people excluded), people without two consecutive visits [baseline visit and year 1 visit] (836 people excluded). There were also sixteen people who did not have a gating mutation, but who were recorded to be receiving ivacaftor; these people were excluded from the analysis. These exclusions and the final numbers in each group are shown in Figure 9.3.

9.2.1 Notation

The following notation will be used to explain the methods.

Let i denote the individual and j denote the visit. Individuals are observed at up to five visits ($j = 0, ..., 4$) per group and baseline is defined as the visit where $j = 0$. If the same individual is in more than one group, e.g. both in group A and group B, they have a visit

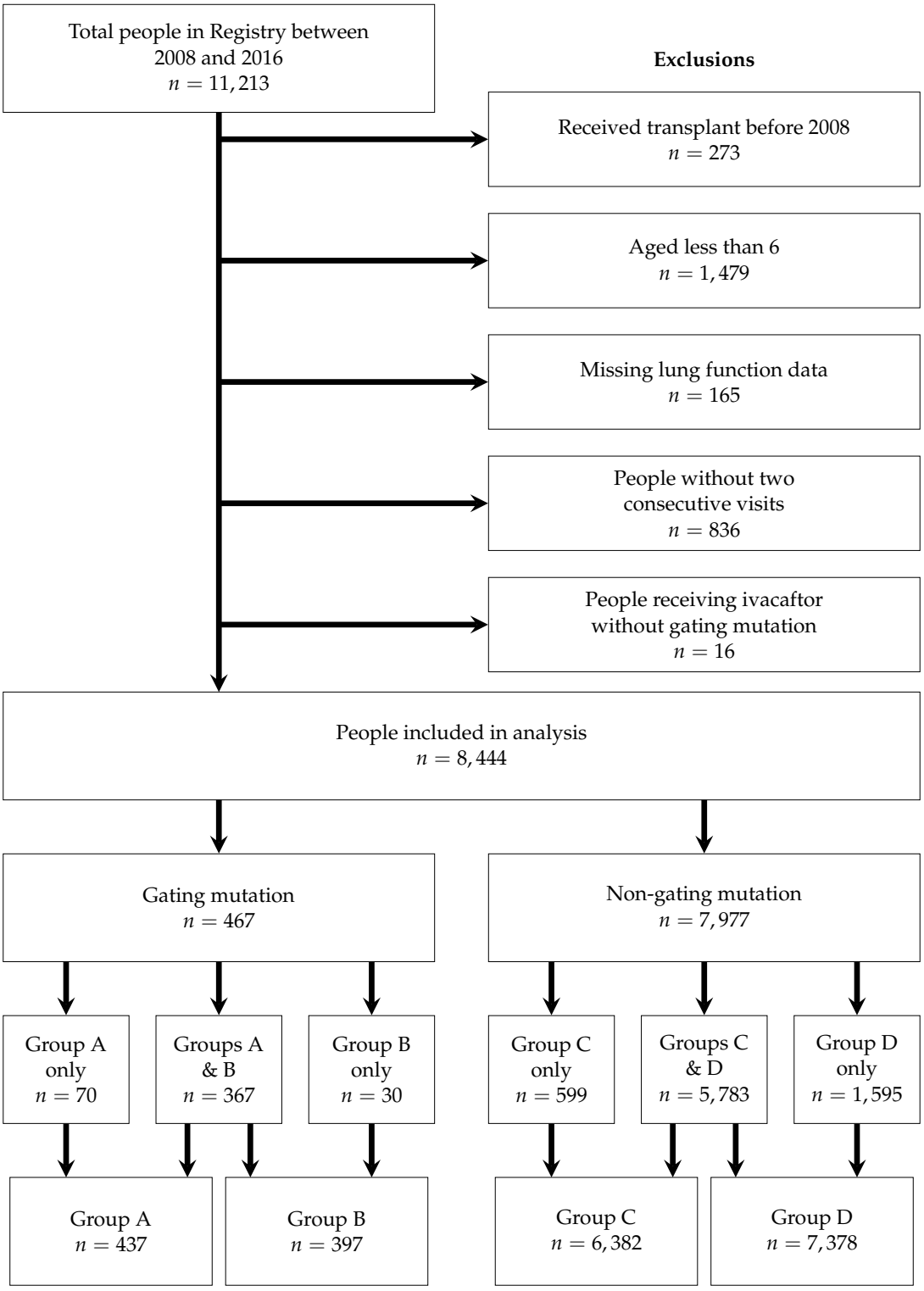


FIGURE 9.3: Flowchart of people included in ivacaftor analysis

$j = 0$ for Group A, which is separate from visit $j = 0$ in Group B. The total population is split into four groups based on genotype, G, and time period P. Anybody with a gating mutation in the time period 2012-2016 ($G = 1 \cap P = 1$) is in the treatment group, $X = 1$, and individuals without a gating mutation or in the pre-ivacaftor era ($G = 0 \cup P = 0$) do not receive treatment, $X = 0$. If an individual is included in both time periods, these are treated separately, and this individual will have two separate baseline visits and two separate follow-up periods.

We are interested in the effect of treatment, X, on outcomes Y. Each outcome is measured four times, at visits $j = 1, 2, 3$ and 4, and we also record the exact time since baseline that the visit took place, T. In the analyses, we will allow an interaction either between treatment, X, and time T, or between treatment, X, and visit, j, to be able to estimate the long-term effects of treatment use.

We also measured a number of variables related to health at baseline, $\boldsymbol{H}$, but allow for that fact that not all baseline measures of health related to the outcome may have been measured. These unobserved baseline variables are referred to as $\boldsymbol{U}$. We only use measures taken at baseline, as any variable measured after baseline would be on the causal pathway between treatment and outcome.

In Figure 9.4 we present the assumed causal pathways between our variables of interest. In Chapter 8, we introduced six possible DAGs for our investigation. In this analysis, we assume the least restrictive of these DAGs that allow for causal pathways between all variables forward in time (Figure 8.3c). Importantly, we allow both genotype and time period to affect both measured and unmeasured baseline variables of health and also to have a direct effect on the outcomes. The methods we introduce will allow us to estimate an unbiased treatment effect even in this 'worst-case' scenario, and would therefore also still provide unbiased estimates of the treatment effect even if some of the assumed causal pathways are null.

9.2.2 Effects of Interest

We wish to estimate the effect of ivacaftor on two outcomes of interest: lung function and annual number of IV days. The primary measure of lung function will be ppFEV$_1$, but we also perform supplementary analyses with ppFVC and ppFEF$_{25-75}$. All three measures are calculated using the GLI equations.[16] All three of these outcome measures are continuous and will be modelled with linear models. In each case, we wish to estimate the 'step-change', defined as the difference in absolute change in lung function between baseline and one year in those receiving treatment to those not receiving treatment. We also aim to estimate the 'slope-change', defined as the difference in absolute change in the annual decline of lung function between year one and the end of follow-up. We assume a linear slope of decline, as with only four visits there is not enough data to accurately

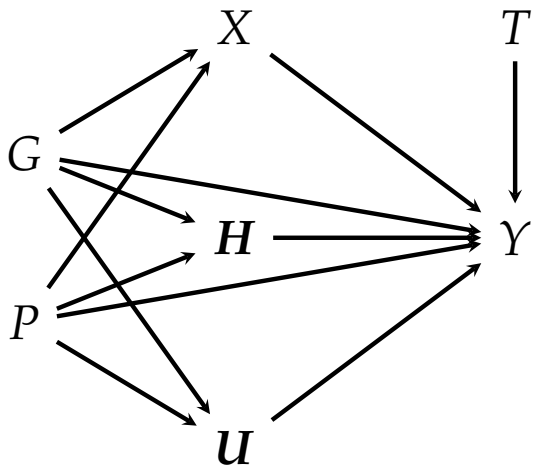


FIGURE 9.4: DAG of assumed causal pathways for ivacaftor analysis between ivacaftor (X), Genotype (G), time Period (P), measured covariates of Health at baseline ($\boldsymbol{H}$), Unmeasured covariates of health at baseline ($\boldsymbol{U}$), outcome (Y), and follow-up Time (T)

model non-linear slopes of decline. These two treatment effects are highlighted in Figure 9.1.

The annual number of IV days is a count outcome and in this chapter will be modelled with a negative binomial distribution. This will ignore any excess of zeroes, which had in previous chapters been accounted for using a zero-inflated negative binomial model. Unlike with lung function, we do not expect there to be an average slope trajectory for the rate of IV days over time, as although the number of IV days will vary year to year, there is no overall trend. We therefore estimate the effect of ivacaftor on the rate of IV days at three discrete time points: after one year of treatment, after two years of treatment and after three years of treatment.

As explained in Chapter 8, we will perform two main comparisons: 1) Group B vs Group A, and 2) Group B vs Group D. For each comparison we obtain an estimate of the treatment effect of interest; this treatment effect estimate is termed the NTE. We also perform the same analyses using negative controls: 1) Group D vs Group C, and 2) Group A vs Group C. From this we obtain an estimate of the NCE. This is the estimated difference between the groups not due to ivacaftor. We then obtain the NCCTE estimate by subtracting the respective NCE from each NTE, (for IV days, this is done on the log scale). This is summarised below:

- Analysis 1
  - Naïve Treatment Effect (NTE_P): Group B vs Group A
  - Negative-Control Effect (NCE_P): Group D vs Group C
  - Negative-Control-Corrected Treatment Effect: $\text{NCCTE}_P = \text{NTE}_P - \text{NCE}_P$

- Analysis 2
  - Naïve Treatment Effect (NTE_G): Group B vs Group D
  - Negative-Control Effect (NCE_G): Group A vs Group C
  - Negative-Control-Corrected Treatment Effect: $\text{NCCTE}_G = \text{NTE}_G - \text{NCE}_G$

We also perform a third analysis in which all four groups are compared in a combined model. This directly estimates the NCCTE estimate, and corrects for any differences in outcome due to genotype or time period.

All three analyses make assumptions about the comparability of the groups. NTE_P in analysis 1 assumes that time period has no effect on the outcome (either direct or mediated through $\boldsymbol{U}$), i.e. that all of the observed difference is due to ivacaftor. Similarly, NTE_G in analysis 2 assumes that genotype has no effect on the outcome (either direct or mediated through $\boldsymbol{U}$). In both cases, these assumptions can be relaxed through estimation of the NCCTE. In the case of analysis 1, we now assume that any difference over time not due to ivacaftor is the same as the difference observed over time in those without a gating mutation. In the case of analysis 2, the assumption is that any difference between the genotype groups not due to ivacaftor is the same as it was in the time period prior to the availability of ivacaftor. Analysis 3 combines both of these assumptions, and in this case can be phrased in terms of interaction effects. Here, we assume that there is no interaction between genotype and time period, i.e. that any changes in time between 2008-2012 and 2012-2016 are the same irrespective of genotype, and that any differences between the two genotype groups have not changed over time, except for the introduction of ivacaftor. Anaysis 3 also makes the additional assumption that the effect of other baseline covariates included in the model is the same in all four groups.

9.2.3 Statistical Methods

We estimate the treatment effects using three types of model: marginal, fixed-effects and mixed-effects models. We address each of these in turn in the following subsections.

Marginal Models

For lung function, we consider four different marginal models:

$$Y_{ij} = \beta_0 + \beta_x X_{ij} + \beta_{xt} X_{ij} T_{ij} + \beta_t T_{ij} + \epsilon_{ij}, \tag{9.1}$$

$$Y_{ij} = \beta_0 + \beta_x X_{ij} + \beta_{xt} X_{ij} j + \beta_t j + \epsilon_{ij}, \tag{9.2}$$

$$Y_{ij} = \beta_0 + \beta_x X_{ij} + \beta_{xt} X_{ij} T_{ij} + \beta_t T_{ij} + \boldsymbol{\beta_h} \boldsymbol{H}_{ij} + \epsilon_{ij}, \tag{9.3}$$

$$Y_{ij} = \beta_0 + \beta_x X_{ij} + \beta_{xt} X_{ij} j + \beta_t j + \boldsymbol{\beta_h} \boldsymbol{H}_{ij} + \epsilon_{ij}. \tag{9.4}$$

In each case, β_x is the step-change effect of treatment and β_{xt} is the slope-change effect of treatment.

In the models shown in Equations 9.1 and 9.2, we do not include any baseline health variables and we refer to these models as unadjusted. Conversely, Equations 9.3 and 9.4 are adjusted models. If the DAG shown in Figure 9.4 is correct, then adjusting for $\boldsymbol{H}$ will correct for some of the confounding due to genotype or time period. For this reason, we would not expect the NTE to be the same between the unadjusted and adjusted models, but the NCCTE estimate should be the same on average, as in the case of the adjusted analysis, the NCE would correct for any direct effect of genotype or time period on the outcome and any indirect effect mediated through unobserved baseline covariates, while the unadjusted analysis would additionally correct for any indirect effects mediated through the observed baseline covariates.

The models shown in Equations 9.1 and 9.3 differ from the models of Equations 9.2 and 9.4 in that the former models use the actual measured time since baseline (continuous time), whereas the latter models assume that visits are all equally spaced at yearly intervals (discrete time). We would expect to gain some precision by using the more accurate continuous measure of follow-up time in the models.

For each model, we must also specify the working correlation matrix. We compare three different choices: independent, exchangeable and unstructured working correlation matrices. An independent working correlation matrix assumes that for each person the outcomes observed at different visits are conditionally uncorrelated. An exchangeable correlation matrix assumes that for each person all outcome measures have the same conditional correlation. Finally, an unstructured correlation matrix allows estimates of a separate correlation for every pair of outcomes. In the case of the unstructured correlation matrix, visits must happen at evenly spaced intervals and for this reason, this type of working correlation matrix can only be used when time is modelled as a discrete variable.

For the annual number of IV days, we are interested in the treatment effect at three discrete time-points and therefore we only use models using visit number, but again we can either adjust for $\boldsymbol{H}$ or not, which should not matter for the final NCCTE estimates:

$$Y_{ij} = \exp\big(\gamma_0 + \gamma_j \mathbb{1}\left[j = k\right] + \gamma_{xj} X_{ij} \mathbb{1}\left[j = k\right] + v_{ij} + \epsilon_{ij}\big), \tag{9.5}$$

$$Y_{ij} = \exp\big(\gamma_0 + \gamma_j \mathbb{1}\left[j = k\right] + \gamma_{xj} X_{ij} \mathbb{1}\left[j = k\right] + \boldsymbol{\gamma_h} \boldsymbol{H}_{ij} + v_{ij} + \epsilon_{ij}\big), \tag{9.6}$$

$$\text{where} \quad e^{v_{ij}} \sim \text{Gamma}\big(1/\alpha, \alpha\big), \text{ and } k = 0, ..., 3. \tag{9.7}$$

Here, the treatment effect estimates of interest are $e^{\gamma_{x1}}$, $e^{\gamma_{x2}}$ and $e^{\gamma_{x3}}$, which correspond to IRR at the end of years one, two and three respectively. As with lung function, we consider independent, exchangeable and unstructured working correlation matrices. Note

that the α is the estimate of overdispersion of the counts, but does not affect the interpretation of the treatment effect estimates.

Fixed-Effects Models

For fixed-effect models, we only consider unadjusted models, as any differences due to baseline covariates are incorporated into the person-specific intercept term. For lung function, this results in the following two models:

$$Y_{ij} = \boldsymbol{\beta}_j + \beta_x X_{ij} + \beta_{xt} X_{ij} T_{ij} + \beta_t T_{ij} + \epsilon_{ij}, \tag{9.8}$$

$$Y_{ij} = \boldsymbol{\beta}_j + \beta_x X_{ij} + \beta_{xt} X_{ij} j + \beta_t j + \epsilon_{ij}. \tag{9.9}$$

And for IV days,

$$Y_{ij} = \exp\big(\gamma_i + \gamma_j \mathbb{1}\left[j = k\right] + \gamma_{xj} X_{ij} \mathbb{1}\left[j = k\right] + v_{ij} + \epsilon_{ij}\big), \tag{9.10}$$

$$Y_{ij} = \exp\big(\gamma_i + \gamma_j \mathbb{1}\left[j = k\right] + \boldsymbol{\gamma_{xj}} X_{ij} \mathbb{1}\left[j = k\right] + \boldsymbol{\gamma_h H}_{ij} + v_{ij} + \epsilon_{ij}\big), \tag{9.11}$$

$$\text{where} \quad e^{v_{ij}} \sim \text{Gamma}\big(1/\alpha, \alpha\big), \text{ and } k = 0, ..., 3. \tag{9.12}$$

Mixed-Effects Models

As with marginal models, we consider four main forms of mixed-effects models corresponding to whether time is measured continuously or discretely and whether measured baseline health variables are included in the model:

$$Y_{ij} = \beta_0 + \beta_x X_{ij} + \beta_{xt} X_{ij} T_{ij} + \beta_t T_{ij} + u_{i0} + u_{it} T_{ij} + \epsilon_{ij}, \tag{9.13}$$

$$Y_{ij} = \beta_0 + \beta_x X_{ij} + \beta_{xt} X_{ij} j + \beta_t j + u_{i0} + u_{it} j + \epsilon_{ij}, \tag{9.14}$$

$$Y_{ij} = \beta_0 + \beta_x X_{ij} + \beta_{xt} X_{ij} T_{ij} + \beta_t T_{ij} + \boldsymbol{\beta_h H}_{ij} + u_{i0} + u_{it} T_{ij} + \epsilon_{ij}, \tag{9.15}$$

$$Y_{ij} = \beta_0 + \beta_x X_{ij} + \beta_{xt} X_{ij} j + \beta_t j + \boldsymbol{\beta_h H}_{ij} + u_{i0} + u_{it} j + \epsilon_{ij}, \tag{9.16}$$

$$\text{where} \quad u_{i0} \sim \text{N}\big(0, \sigma_0^2\big), \quad u_{it} \sim \text{N}\big(0, \sigma_t^2\big), \quad \text{and} \quad \text{Cov}\big(u_{i0}, u_{it}\big) = \sigma_{0t}^2. \tag{9.17}$$

Here, we allow for a random intercept and slope term, but assume that in each case the random effects come from the same distribution for all people. We refer to this as a combined covariance matrix structure. Alternatively, we can model the random effects separately by group, if we believe that ivacaftor could result in people becoming more or less variable. This results in a separate covariance matrix structure for each group. For example, in analysis 1, comparing group B to group A, we would assume the following

structure for the random effects:

$$u_{i0a} \sim \mathrm{N}\left(0, \sigma^2_{0a}\right), \quad u_{ita} \sim \mathrm{N}\left(0, \sigma^2_{ta}\right), \quad \text{and} \quad \mathrm{Cov}\left(u_{i0a}, u_{ita}\right) = \sigma^2_{0ta}, \tag{9.18}$$

$$u_{i0b} \sim \mathrm{N}\left(0, \sigma^2_{0b}\right), \quad u_{itb} \sim \mathrm{N}\left(0, \sigma^2_{tb}\right), \quad \text{and} \quad \mathrm{Cov}\left(u_{i0b}, u_{itb}\right) = \sigma^2_{0tb}. \tag{9.19}$$

For estimating the treatment effect on the rate of annual IV days, we only include a random intercept term:

$$Y_{ij} = \exp\left(\gamma_0 + \gamma_j \mathbb{1}\left[j = k\right] + \gamma_{xj} X_{ij} \mathbb{1}\left[j = k\right] + u_{i0} + v_{ij} + \epsilon_{ij}\right), \tag{9.20}$$

$$Y_{ij} = \exp\left(\gamma_0 + \gamma_j \mathbb{1}\left[j = k\right] + \gamma_{xj} X_{ij} \mathbb{1}\left[j = k\right] + \boldsymbol{\gamma_h} \boldsymbol{H}_{ij} + u_{i0} + v_{ij} + \epsilon_{ij}\right), \tag{9.21}$$

$$\text{where} \quad e^{v_{ij}} \sim \mathrm{Gamma}\left(1/\alpha, \alpha\right), \text{ and } u_{i0} \sim \mathrm{N}\left(0, \sigma^2_0\right), \text{ and } k = 0, ..., 3. \tag{9.22}$$

Again we perform the analyses with two different covariance matrix structures: combined through all groups, or separate by group.

Estimation of Standard Errors

To analytically estimate the SE of the NCCTE it is necessary to be able to estimate the covariance between the NTE and NCE. For analysis 1, where the NCE analysis is independent of the NTE analysis (because the same person cannot be in both analyses), the covariance will be zero and the SE of the NCCTE could be calculated as the sum of the variances of the NCE and NTE. However, in the case of analysis 2, the same people will be in both the NCE analysis and the NTE analysis. For this reason, in this case it will be necessary to use a bootstrap procedure to obtain estimates of the SE in this case.

In order to ensure comparability between the methods, for these analyses we perform a non-parametric boostrap with 1000 resamples and for each resample we calculate all effect estimates, in order to obtain bootstrap estimates for the SEs, even those that could have been obtained analytically.

9.3 Results

There were 397 people who received ivacaftor and fulfilled the inclusion criteria to be included in the analysis. Overall, in this group, there were 1,765 observations, with a mean follow up time of 3.5 years (SD 0.9). Figure 9.2 shows the breakdown for all four groups. As expected, the available population without gating mutations is much larger than that with gating mutations: 7,977 people (57,208 total observations) in those without gating mutations compared to 467 people (3,528 observations) in those with gating mutations. In those with a gating mutation, the number of inviduals in the pre- & post-ivacaftor

phase are similar, but in those without a gating mutation we see a much larger number in the post-phase than the pre-phase. This is due to the fact that for those receiving ivacaftor, they needed to have a baseline measure directly prior to the year they started to receiving ivacaftor, whereas in those without a gating mutation, there is no ivacaftor start date, so people just needed two consecutive visits at any point in the period.

Table 9.1 shows a summary of baseline covariates by group. Overall, all four groups are very similar, but there are some small systematic differences that can be noticed between the groups in the pre-ivacaftor phase (groups A and C) compared to the groups in the post-ivacaftor phase (groups B and D). The mean age is slightly higher in groups B & D, due to these groups being from a later time period. These two groups also have longer follow-up on average, suggesting less loss to follow-up in the post-ivacaftor phase. Furthermore, there are higher numbers of people in the pre-ivacaftor phase (groups A and C) who have CFRD and who receive mucolytic treatments. However, when comparing those with a gating mutation (groups A and B) those without a gating mutation (groups C and D) there are no obvious differences between these groups.

Variable	Group			
	A I=437	B I=397	C I=6,382	D I=7,378
Ivacaftor	0 (0.0)	397 (100.0)	0 (0.0)	0 (0.0)
Total Follow-Up Time (Years)	3.12 (1.18)	3.49 (0.90)	3.07 (1.16)	3.35 (1.05)
Total Number of Visits	3.07 (1.15)	3.50 (0.88)	3.05 (1.11)	3.36 (1.01)
Baseline Age (Years)	20.4 (10.8)	22.4 (11.2)	20.9 (11.6)	21.9 (12.6)
Female	205 (46.9)	186 (46.9)	2,971 (46.6)	3,465 (47.0)
White Ethnicity	428 (97.9)	390 (98.2)	6,150 (96.4)	7,043 (95.5)
Baseline ppFEV$_1$	71.0 (23.2)	69.7 (23.2)	71.6 (23.3)	72.0 (23.4)
Baseline ppFVC*	84.8 (19.4)	84.1 (18.9)	84.0 (19.5)	84.4 (19.6)
Baseline ppFEF$_{25-75}$**	56.3 (31.3)	55.9 (32.4)	60.9 (32.8)	58.4 (31.0)
Baseline IV Days	18.4 (28.1)	20.2 (30.5)	17.6 (27.7)	18.6 (28.3)
Baseline Infection	358 (81.9)	350 (88.2)	4,847 (75.9)	5,948 (80.6)
Baseline CFRD	69 (15.8)	90 (22.7)	1,198 (18.8)	1,751 (23.7)
Baseline Smoker	9 (2.1)	9 (2.3)	154 (2.4)	201 (2.7)
Baseline Mucolytic Treatment	223 (51.0)	264 (66.5)	2,992 (46.9)	4,822 (65.4)

TABLE 9.1: Summary Statistics by Group (*ppFVC based on 14,556 observations. **ppFEF$_{25-75}$ based on 5,711 observations)

By design, all people had a baseline visit ($j = 0$) and one visit in the year directly after baseline ($j = 1$). The total number of possible visits post-baseline was 4, which 7,933 (54.4%) people had. However, a number of people missed a visit during follow-up rather than missing the final visit, meaning that in total there were 8,620 (59.1%) people with data out to four years post-baseline. The number of people with only 1 year of follow-up was 1,667 (11.4%), only 2 years was 2,035 (13.9%) and only 3 years was 2,272 (15.6%).

Table 9.2 compares visit number to continuous follow-up time measured in years. On average the two variables coincide as expected with visits generally happening every year. The IQR shows that approximately 50% of visits generally happen between 0.9

and 1.1 years since the previous visit. However, the range does show that there are a number of people who do not always have regularly spaced visits, for example some people's second visit post-baseline actually happens only one year after baseline, whereas for others it's almost three years after baseline.

Continuous Follow-Up Time	Visit Number			
	1 n=14,594	2 n=12,365	3 n=10,563	4 n=8,620
Mean (SD)	1,00 (0.23)	2.01 (0.26)	3.00 (0.29)	4.01 (0.30)
Median (IQR)	1.00 [0.93, 1.07]	2.00 [1.92, 2.11]	3.00 [2.87, 3.12]	4.00 [3.87, 4.14]
Range	0.03, 1.97	1.00, 2.96	2.00, 3.97	3.00, 4.97

TABLE 9.2: Comparison of continuous follow-up time to discrete visit number

9.3.1 Estimated Effect of Ivacaftor on Lung Function

The full results of analysis 1 comparing those currently taking ivacaftor to those with a gating mutation but in the time period before ivacaftor was available (group B vs A) are presented in Figures 9.5 and 9.6, and Tables 9.3 and 9.4. In terms of the step-change at one year, all methods gave similar estimates of the NTE. Across methods the estimate NTE was around 6%. The smallest estimate was from an unadjusted marginal model with an independent working correlation matrix using time measured continuously (estimate 5.17%, 95% CI 3.31% to 7.03%), and the largest estimate was from an adjusted marginal model with an independent working correlation matrix using time measured continuously (estimate 8.12%, 95% CI 6.38% to 9.44%). The NCE estimates (groups D vs C) were all small and positive, suggesting a slight improvement in ppFEV$_1$ over time not due to ivacaftor. The smallest estimate was from an unadjusted marginal model with an unstructured working correlation matrix and time measured at discrete intervals (estimate 0.22%, 95% CI -0.04% to 0.48%), and the largest estimate was from an adjusted marginal model with an independent working correlation matrix and time measured continuously (estimate 1.59%, 95% CI 1.21% to 1.98%). This resulted in the NCCTE being slightly lower than the NTE, but all methods still showed a statistically significant improvement in lung function due to ivacaftor. The smallest NCCTE estimate was from an unadjusted marginal model with an independent working correlation matrix using time measured continuously (estimate 4.77%, 95% CI 2.87% to 6.67%), and the largest estimate was from an adjusted marginal model with an independent working correlation matrix using time measured continuously (estimate 6.53%, 95% CI 4.76% to 8.30%). Overall, all methods gave very similar NCCTE estimates with largely overlapping CIs.

The results from different modelling approaches were also similar when considering the effect of ivacaftor on the slope of lung function decline. The NTE estimates were all relatively similar, indicating approximately a 0.75% improvement in the rate of decline in

the post- versus pre-ivacaftor phase. All but two of the estimates were statistically significant. The smallest estimate was from an unadjusted marginal model with an independent working correlation matrix using time measured continuously (estimate 0.48%, 95% CI -0.42% to 1.39%), and the largest estimate was from an adjusted marginal model with an independent working correlation matrix using time measured at discrete intervals (estimate 1.04%, 95% CI 0.30% to 1.79%). Once again, the NCE estimates were small and positive suggesting a small improvement in the rate of lung function decline over time, but many of these estimates were not statistically significant. The smallest estimate was from an unadjusted mixed-effects model with separate random-effects by group using time measured continuously (estimate 0.05%, 95% CI -0.09% to 0.19%), and the largest estimate was from an adjusted marginal model with an independent working correlation matrix using time measured at discrete intervals (estimate 0.60%, 95% CI 0.43% to 0.77%). This resulted in the NCCTE being lower than the NTE, estimating approximately a 0.6% improvement in the rate of lung function decline. Overall, fifteen out of twenty of the NCCTE estimates were statistically significantly different from zero. The smallest estimate was from an adjusted marginal model with an independent working correlation matrix using time measured continuously (estimate 0.35%, 95% CI -0.49% to 1.18%), and the largest estimate was from an adjusted mixed-effects model using time measured continuously (estimate 0.63%, 95% CI 0.12% to 1.15%).

The full results of analysis 2 comparing those eligible for ivacaftor to those ineligible for ivacaftor due to genotype (group B vs D)are presented in Figures 9.7 and 9.8, and Tables 9.5 and 9.6. Overall, the results are very similar to those from analysis 1. When considering the step-change effect, the NTE estimates indicated an improvement in ppFEV$_1$ of between 5.5% and 6%, except for the unadjusted marginal models with an independent working correlation matrix which gave a smaller estimate of 3.5%. The largest estimate was from an adjusted marginal model with an unstructured working correlation matrix and time measured at discrete intervals (estimate 6.03%, 95% CI 4.81% to 7.26%). The NCE esimates were all estimated to be negative, suggesting that people with a gating mutation actually had a slightly greater decline in lung function at one year of follow-up when ivacaftor was not available. However, none of the NCE estimates were statistically significantly different from zero. Again, the estimates from the marginal models with an independent working correlation matrix were the most extreme, with the estimated difference between groups being between -1.57% and -1.06%. All other methods estimated a much smaller difference of approximately -0.3%. The least extreme result was from an unadjusted marginal model with an unstructured working correlation matrix and time measured at discrete intervals (estimate -0.09%, 95% CI -1.23% to 1.05%). All methods gave very similar estimates for the NCCTE. This is because the methods which had more extreme estimates for the NTE, also had more extreme estimates for the NCE, which cancelled each other out when calculating the NCCTE.

The slope-effect estimates from analysis 2 were also similar to the slope effect estimates from analysis 1, except that the SEs were generally slightly larger. This did not materially impact the NTE estimates as all but one estimate was statistically significant. The estimate that was not statistically significant was from an adjusted marginal model with an independent working correlation matrix and time measured continuously (estimate 0.47%, 95% CI -0.14% to 1.09%). The largest estimate was from an adjusted mixed-effects model with time measured continuously (estimate 0.75%, 95% CI 0.38% to 1.13%). The NCE estimates were all positive (but not statistically significant) suggesting a slower rate of decline in those with a gating mutation even without ivacaftor. The smallest slope-change estimate was from an unadjusted mixed-effects model with separate random-effects by group using time measured at discrete intervals (estimate 0.15%, 95% CI -0.29% to 0.59%), and the largest estimate was from an unadjusted marginal model with an independent working correlation matrix using time measured continuously (estimate 0.43%, 95% CI -0.45% to 1.32%). This small positive estimate for the NCE resulted in the NCCTE being lower than the NTE and although all methods estimated an improvement in the rate of lung function decline, none of them were statistically significant. The method with the smallest p-value for the NCCTE slope-effect was an adjusted mixed-effects model with time measured at discrete intervals (estimate 0.56%, 95% CI -0.02% to 1.13%).

Finally, the results from the combined analysis directly obtain the NCCTE estimate and are presented in Figures 9.9 and 9.10, and in Tables 9.7 and 9.8. Again, the step-change effect estimates indicated around a 6% improvement in lung function. The smallest estimate was from an unadjusted marginal model with an independent working correlation matrix using time measured continuously (estimate 4.88%, 95% CI 3.04% to 6.73%), and the largest estimate was from an adjusted marginal model with an independent working correlation matrix using time measured continuously (estimate 6.07%, 95% CI 4.33% to 7.82%). As with the other analyses, the estimates of the slope-change effect were all positive, but only thirteen out of twenty were statistically significant. The smallest estimate was from an unadjusted marginal model with an independent working correlation matrix using time measured at discrete intervals (estimate 0.32%, 95% CI -0.53% to 1.17%), and the largest estimate was from an adjusted mixed-effects model using time measured at discrete intervals (estimate 0.59%, 95% CI 0.07% to 1.10%).

We also present the results for two other lung function outcome measures (ppFVC and ppFEF_{25-75}) in Appendix F. The results for ppFVC were very similar to the results presented in this chapter, with strong evidence of an step-change increase in ppFVC at one year and an estimated decrease in the slope of decline that was not always statistically significant. For ppFEF_{25-75} the direction of the results was the same, but there was only weak evidence of a step-change effect and none of the slope-change effect estimates were statistically significant.

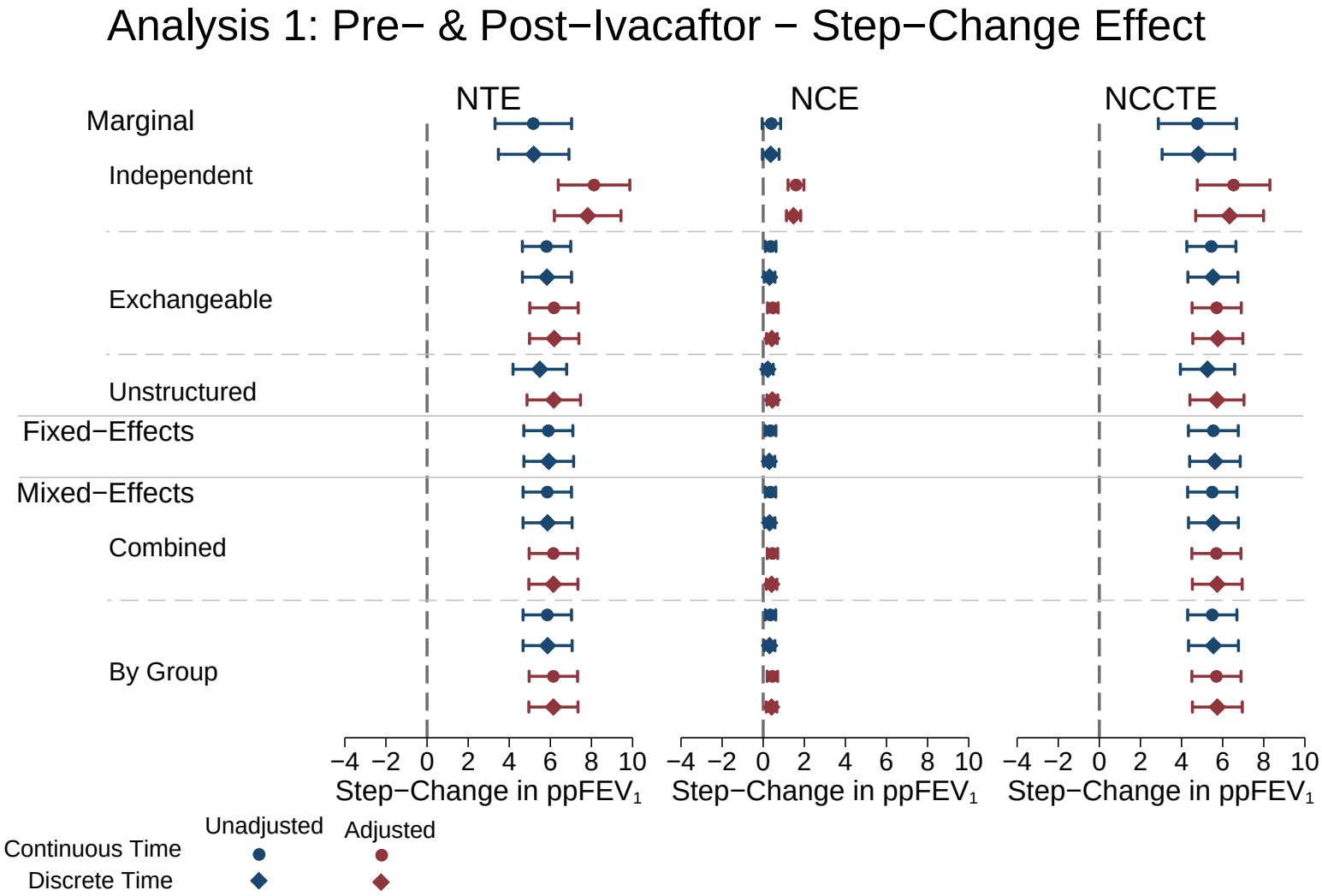


FIGURE 9.5: Estimated step-change effect on $ppFEV_1$ when comparing post-ivacaftor period to pre-ivacaftor period

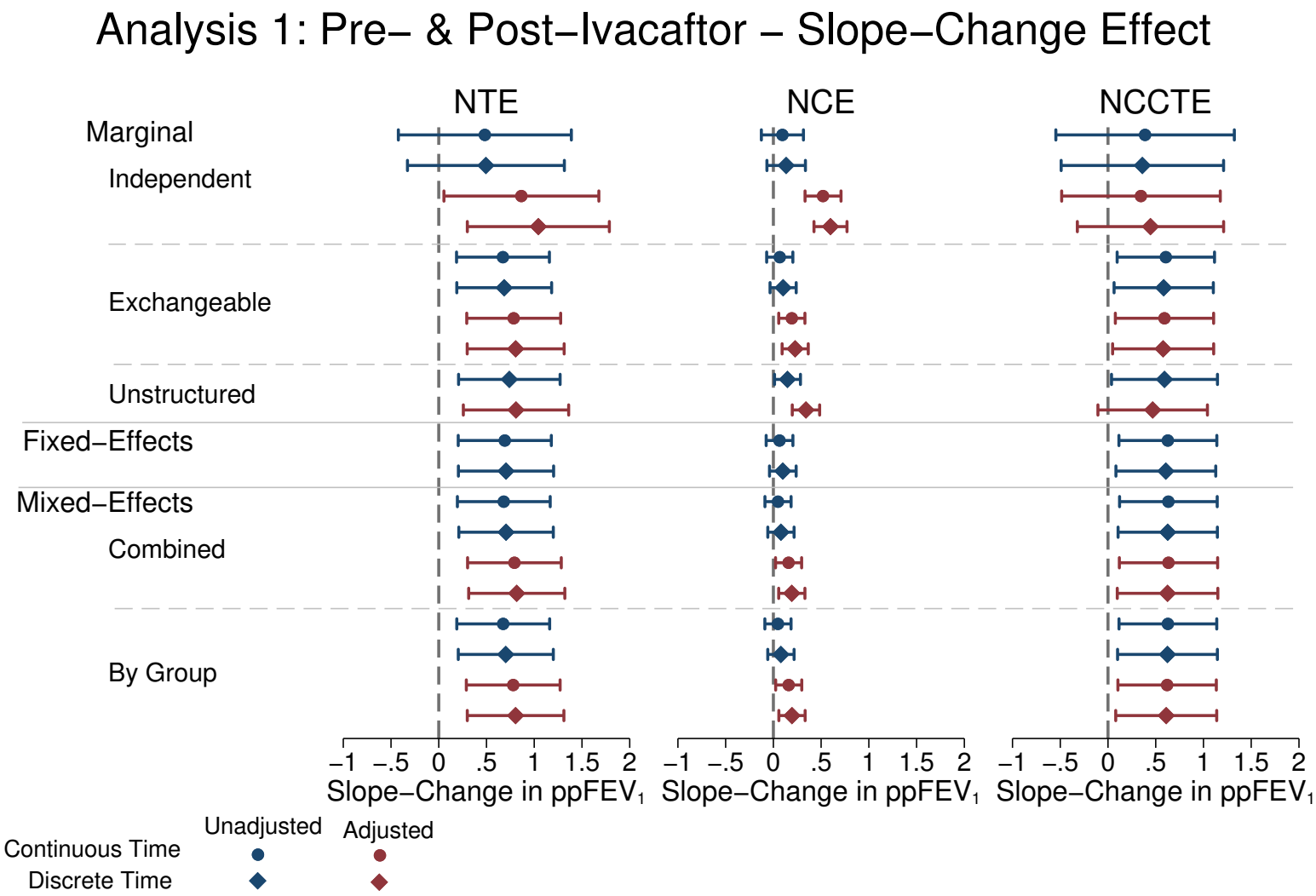


FIGURE 9.6: Estimated slope-change effect on $ppFEV_1$ when comparing post-ivacaftor period to pre-ivacaftor period

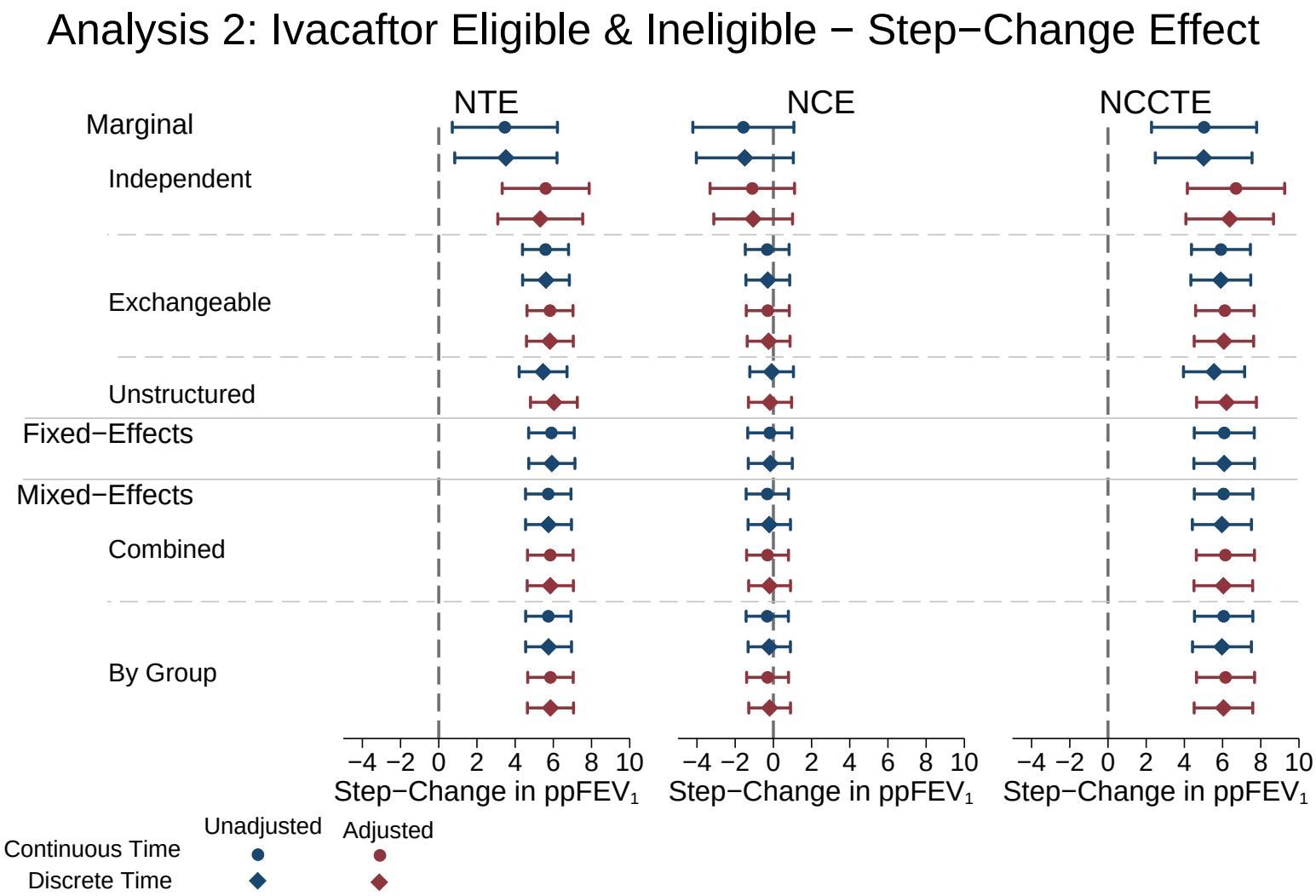


FIGURE 9.7: Estimated step-change effect on $ppFEV_1$ when comparing those eligible for ivacaftor to those ineligible for ivacaftor

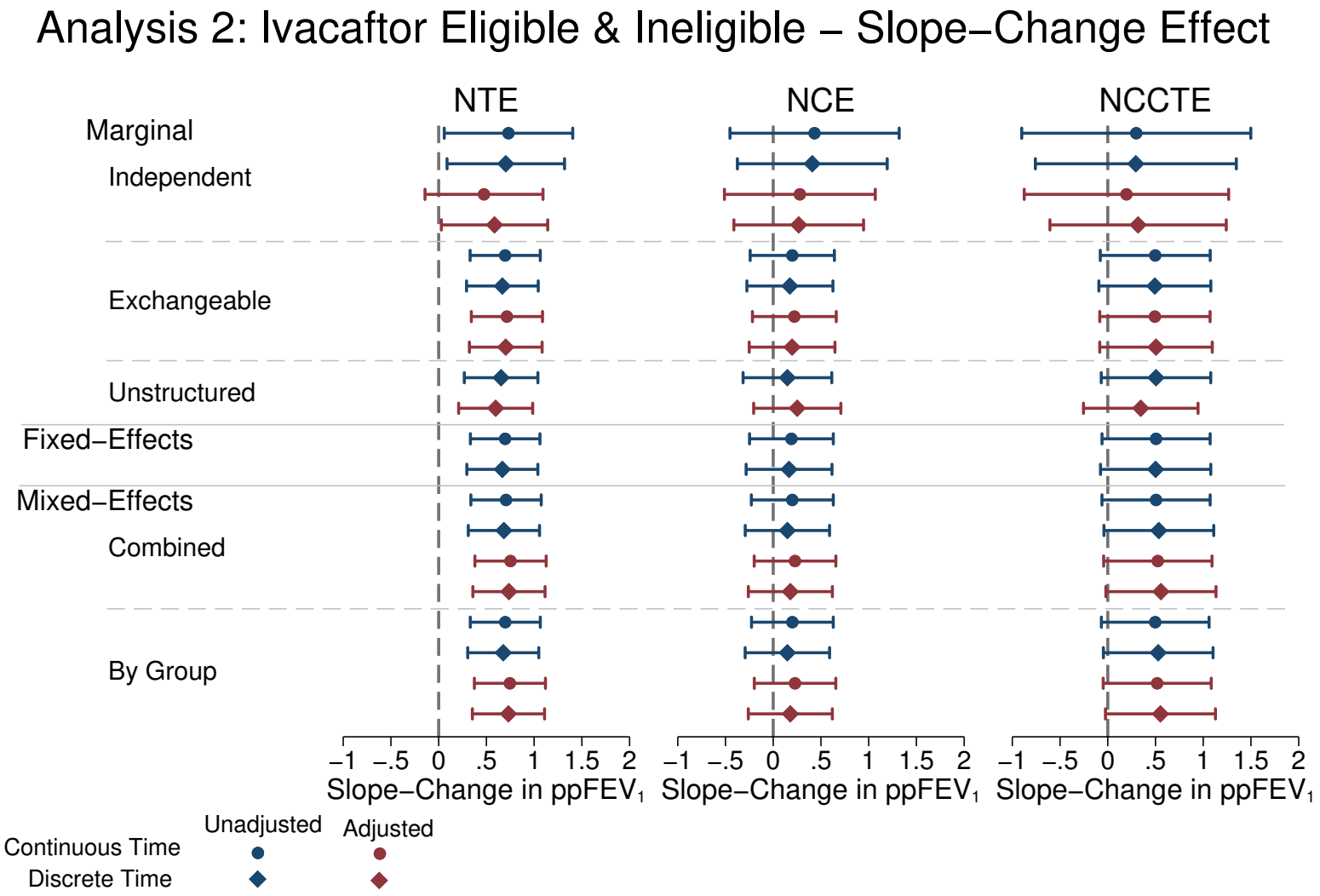


FIGURE 9.8: Estimated slope-change effect on $ppFEV_1$ when comparing those eligible for ivacaftor to those ineligible for ivacaftor

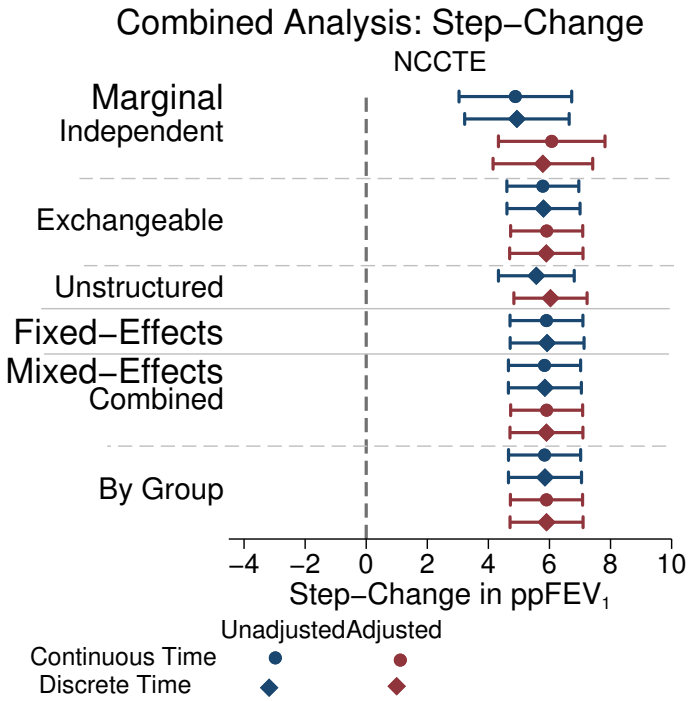


FIGURE 9.9: Estimated step-change effect on $ppFEV_1$ from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time period prior to the availability of ivacaftor

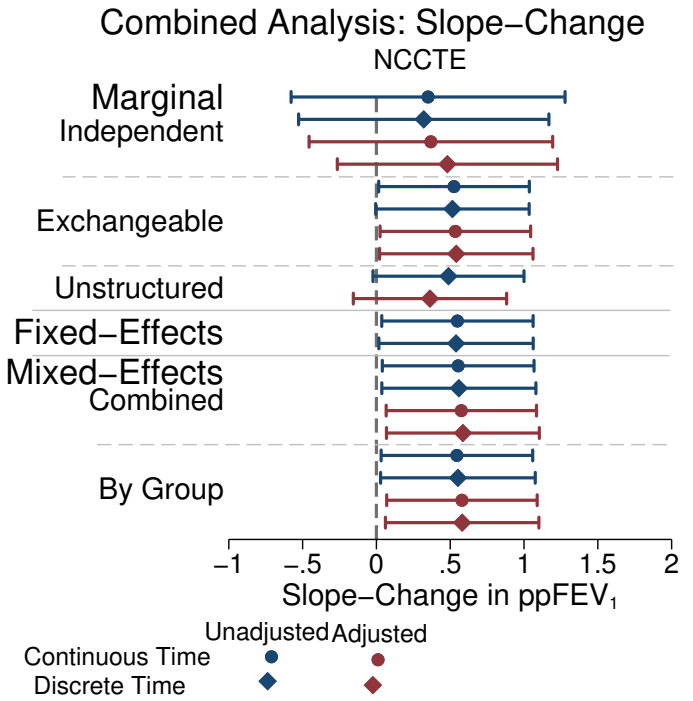


FIGURE 9.10: Estimated slope-change effect on $ppFEV_1$ from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time period prior to the availability of ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 1: Pre- & Post Comparison - Step-Change Effect											
				NTE (B vs A)				NCE (D vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	5.17	0.95	3.31, 7.03	<0.001	0.40	0.23	-0.05, 0.85	0.080	4.77	0.97	2.87, 6.67	<0.001
			Discrete	5.19	0.88	3.47, 6.91	<0.001	0.37	0.21	-0.04, 0.77	0.077	4.82	0.90	3.05, 6.59	<0.001
		Adjusted	Continuous	8.12	0.89	6.38, 9.87	<0.001	1.59	0.20	1.21, 1.98	<0.001	6.53	0.90	4.76, 8.30	<0.001
			Discrete	7.82	0.83	6.20, 9.44	<0.001	1.48	0.18	1.13, 1.82	<0.001	6.34	0.84	4.69, 7.99	<0.001
	Exchangeable	Unadjusted	Continuous	5.81	0.60	4.64, 6.99	<0.001	0.36	0.13	0.10, 0.62	0.006	5.45	0.61	4.25, 6.65	<0.001
			Discrete	5.84	0.61	4.64, 7.03	<0.001	0.31	0.13	0.05, 0.57	0.020	5.52	0.62	4.31, 6.74	<0.001
		Adjusted	Continuous	6.18	0.60	5.00, 7.36	<0.001	0.47	0.13	0.21, 0.73	<0.001	5.71	0.61	4.51, 6.91	<0.001
			Discrete	6.19	0.61	4.99, 7.39	<0.001	0.42	0.13	0.16, 0.68	0.001	5.76	0.62	4.55, 6.98	<0.001
	Unstructured	Unadjusted	Discrete	5.49	0.67	4.18, 6.79	<0.001	0.22	0.13	-0.04, 0.48	0.091	5.26	0.67	3.94, 6.58	<0.001
		Adjusted		6.17	0.66	4.87, 7.47	<0.001	0.45	0.13	0.18, 0.71	<0.001	5.72	0.67	4.41, 7.04	<0.001
Fixed-Effects		Unadjusted	Continuous	5.90	0.61	4.71, 7.10	<0.001	0.35	0.13	0.09, 0.62	0.008	5.55	0.62	4.33, 6.76	<0.001
			Discrete	5.92	0.62	4.71, 7.13	<0.001	0.30	0.13	0.03, 0.56	0.027	5.62	0.63	4.39, 6.85	<0.001
Mixed-Effects	Combined	Unadjusted	Continuous	5.85	0.60	4.67, 7.03	<0.001	0.36	0.13	0.10, 0.61	0.006	5.49	0.61	4.30, 6.69	<0.001
			Discrete	5.86	0.61	4.67, 7.06	<0.001	0.31	0.13	0.05, 0.57	0.018	5.55	0.62	4.34, 6.76	<0.001
		Adjusted	Continuous	6.15	0.60	4.97, 7.32	<0.001	0.45	0.13	0.20, 0.70	<0.001	5.70	0.61	4.50, 6.89	<0.001
			Discrete	6.15	0.61	4.95, 7.34	<0.001	0.41	0.13	0.15, 0.67	0.002	5.74	0.62	4.52, 6.95	<0.001
	By Group	Unadjusted	Continuous	5.85	0.60	4.67, 7.03	<0.002	0.36	0.13	0.10, 0.61	0.006	5.50	0.61	4.30, 6.69	<0.002
			Discrete	5.87	0.61	4.67, 7.06	<0.001	0.31	0.13	0.05, 0.57	0.018	5.55	0.62	4.34, 6.77	<0.001
		Adjusted	Continuous	6.15	0.60	4.97, 7.32	<0.001	0.45	0.13	0.20, 0.70	<0.001	5.70	0.61	4.50, 6.89	<0.001
			Discrete	6.15	0.61	4.95, 7.34	<0.001	0.41	0.13	0.15, 0.66	0.002	5.74	0.62	4.53, 6.95	<0.001

TABLE 9.3: Estimated step-change effect on ppFEV$_1$ when comparing post-ivacaftor period to pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 1: Pre- & Post Comparison - Slope-Change Effect											
				NTE (B vs A)				NCE (D vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	0.48	0.46	-0.42, 1.39	0.30	0.09	0.11	-0.13, 0.31	0.40	0.39	0.48	-0.55, 1.32	0.42
			Discrete	0.49	0.42	-0.33, 1.32	0.24	0.13	0.10	-0.07, 0.34	0.19	0.36	0.43	-0.49, 1.21	0.41
		Adjusted	Continuous	0.87	0.41	0.05, 1.68	0.037	0.52	0.10	0.33, 0.71	<0.001	0.35	0.42	-0.49, 1.18	0.42
			Discrete	1.04	0.38	0.30, 1.79	0.006	0.60	0.09	0.43, 0.77	<0.001	0.45	0.39	-0.32, 1.21	0.25
	Exchangeable	Unadjusted	Continuous	0.67	0.25	0.19, 1.16	0.007	0.07	0.07	-0.07. 0.20	0.34	0.61	0.26	0.10, 1.12	0.020
			Discrete	0.69	0.25	0.19, 1.18	0.007	0.10	0.07	-0.03, 0.24	0.14	0.58	0.27	0.06, 1.10	0.028
		Adjusted	Continuous	0.78	0.25	0.29, 1.28	0.002	0.19	0.07	0.06, 0.33	0.006	0.59	0.26	0.08, 1.11	0.024
			Discrete	0.81	0.26	0.30, 1.31	0.002	0.23	0.07	0.09, 0.37	0.001	0.58	0.27	0.05, 1.11	0.032
	Unstructured	Unadjusted	Discrete	0.74	0.27	0.21, 1.27	0.006	0.15	0.07	0.01, 0.28	0.033	0.59	0.28	0.04, 1.15	0.037
		Adjusted		0.81	0.28	0.26, 1.36	0.004	0.34	0.07	0.20, 0.48	<0.001	0.47	0.29	-0.11, 1.04	0.11
Fixed-Effects		Unadjusted	Continuous	0.69	0.25	0.20, 1.18	0.005	0.06	0.07	-0.08, 0.20	0.37	0.63	0.26	0.11, 1.14	0.017
			Discrete	0.70	0.25	0.21, 1.20	0.006	0.10	0.07	-0.04, 0.24	0.17	0.61	0.27	0.08, 1.13	0.023
Mixed-Effects	Combined	Unadjusted	Continuous	0.68	0.25	0.20, 1.17	0.006	0.05	0.07	-0.09, 0.19	0.49	0.63	0.26	0.12, 1.14	0.015
			Discrete	0.70	0.25	0.21, 1.20	0.005	0.08	0.07	-0.06, 0.22	0.26	0.63	0.27	0.11, 1.15	0.018
		Adjusted	Continuous	0.79	0.25	0.30, 1.28	0.002	0.16	0.07	0.02, 0.30	0.023	0.63	0.26	0.12, 1.15	0.016
			Discrete	0.82	0.26	0.31, 1.32	0.001	0.19	0.07	0.06, 0.33	0.006	0.62	0.27	0.10, 1.15	0.020
	By Group	Unadjusted	Continuous	0.68	0.25	0.19, 1.16	0.007	0.05	0.07	-0.09, 0.19	0.49	0.63	0.26	0.12, 1.14	0.016
			Discrete	0.70	0.25	0.20, 1.20	0.006	0.08	0.07	-0.06, 0.22	0.26	0.62	0.27	0.10, 1.15	0.019
		Adjusted	Continuous	0.78	0.25	0.29, 1.27	0.002	0.16	0.07	0.02, 0.30	0.022	0.62	0.26	0.10, 1.13	0.019
			Discrete	0.80	0.26	0.30, 1.31	0.002	0.19	0.07	0.06, 0.33	0.005	0.61	0.27	0.08, 1.14	0.024

TABLE 9.4: Estimated slope-change effect on ppFEV$_1$ when comparing post-ivacaftor period to pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 2: Eligible & Ineligible Comparison - Step-Change Effect											
				NTE (B vs D)				NCE (A vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	3.46	1.40	0.71, 6.21	0.014	-1.57	1.35	-4.22, 1.08	0.24	5.03	1.40	2.28, 7.79	<0.001
			Discrete	3.52	1.37	0.84, 6.19	0.010	-1.50	1.30	-4.03, 1.04	0.25	5.01	1.29	2.48, 7.54	<0.001
		Adjusted	Continuous	5.60	1.16	3.32, 7.88	<0.001	-1.10	1.13	-3.32, 1.11	0.33	6.70	1.30	4.16, 9.25	<0.001
			Discrete	5.32	1.13	3.09, 7.54	<0.001	-1.06	1.05	-3.12, 1.00	0.32	6.37	1.17	4.08, 8.67	<0.001
	Exchangeable	Unadjusted	Continuous	5.59	0.62	4.39, 6.80	<0.001	-0.32	0.59	-1.47, 0.82	0.58	5.92	0.79	4.37, 7.46	<0.001
			Discrete	5.62	0.62	4.39, 6.84	<0.001	-0.29	0.58	-1.44, 0.85	0.62	5.91	0.80	4.34, 7.47	<0.001
		Adjusted	Continuous	5.83	0.61	4.62,7.03	<0.001	-0.29	0.58	-1.42, 0.84	0.61	6.12	0.78	4.59, 7.65	<0.001
			Discrete	5.82	0.63	4.60, 7.05	<0.001	-0.25	0.57	-1.37, 0.87	0.66	6.07	0.79	4.51, 7.63	<0.001
	Unstructured	Unadjusted	Discrete	5.46	0.64	4.21, 6.72	<0.001	-0.09	0.58	-1.23, 1.05	0.88	5.55	0.82	3.95, 7.16	<0.001
		Adjusted		6.03	0.63	4.81, 7.26	<0.001	-0.18	0.58	-1.31, 0.95	0.76	6.21	0.80	4.64, 7.77	<0.001
Fixed-Effects		Unadjusted	Continuous	5.90	0.61	4.71, 7.10	<0.001	-0.19	0.59	-1.34, 0.97	0.75	6.09	0.80	4.52, 7.66	<0.001
			Discrete	5.92	0.62	4.71, 7.13	<0.001	-0.16	0.59	-1.31, 0.99	0.78	6.08	0.81	4.50, 7.67	<0.001
Mixed-Effects	Combined	Unadjusted	Continuous	5.73	0.61	4.54, 6.92	<0.001	-0.32	0.57	-1.43, 0.79	0.57	6.05	0.78	4.52, 7.59	<0.001
			Discrete	5.75	0.61	4.55, 6.95	<0.001	-0.22	0.57	-1.33, 0.89	0.70	5.96	0.79	4.42, 7.51	<0.001
		Adjusted	Continuous	5.84	0.61	4.65, 7.04	<0.001	-0.31	0.56	-1.40, 0.79	0.58	6.15	0.78	4.63, 7.67	<0.001
			Discrete	5.84	0.62	4.63, 7.05	<0.001	-0.20	0.56	-1.2, 0.90	0.72	6.04	0.78	4.50, 7.58	<0.001
	By Group	Unadjusted	Continuous	5.74	0.61	4.55, 6.92	<0.001	-0.32	0.57	-1.43, 0.79	0.57	6.06	0.78	4.53, 7.59	<0.001
			Discrete	5.75	0.61	4.55, 6.95	<0.001	-0.22	0.56	-1.32, 0.89	0.70	5.97	0.79	4.42, 7.51	<0.001
		Adjusted	Continuous	5.85	0.61	4.66, 7.04	<0.001	-0.30	0.56	-1.40, 0.79	0.59	6.15	0.78	4.63, 7.67	<0.001
			Discrete	5.85	0.62	4.64, 7.06	<0.001	-0.20	0.56	-1.29, 0.90	0.73	6.05	0.78	4.51, 7.58	<0.001

TABLE 9.5: Estimated step-change effect on ppFEV$_1$ when comparing those eligible for ivacaftor to those ineligible for ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 2: Eligible & Ineligible Comparison - Slope-Change Effect											
				NTE (B vs D)				NCE (A vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	0.73	0.34	0.06, 1.40	0.033	0.43	0.45	-0.45, 1.32	0.34	0.30	0.61	-0.90, 1.50	0.63
			Discrete	0.70	0.31	0.09, 1.32	0.025	0.41	0.40	-0.38, 1.19	0.31	0.29	0.54	-0.76, 1.35	0.58
		Adjusted	Continuous	0.47	0.32	-0.14, 1.09	0.13	0.28	0.40	-0.51, 1.07	0.49	0.20	0.55	-0.88, 1.27	0.72
			Discrete	0.59	0.28	0.03, 1.14	0.040	0.27	0.35	-0.41, 0.95	0.44	0.32	0.47	-0.61, 1.24	0.50
	Exchangeable	Unadjusted	Continuous	0.70	0.19	0.33, 1.06	<0.001	0.20	0.23	-0.24, 0.64	0.37	0.50	0.29	-0.08, 1.07	0.091
			Discrete	0.67	0.19	0.29, 1.04	<0.001	0.17	0.23	-0.28, 0.63	0.45	0.49	0.30	-0.09, 1.08	0.099
		Adjusted	Continuous	0.72	0.19	0.34, 1.09	<0.001	0.22	0.22	-0.22, 0.66	0.32	0.49	0.29	-0.08, 1.07	0.094
			Discrete	0.70	0.19	0.32, 1.08	<0.001	0.20	0.23	-0.25, 0.65	0.39	0.51	0.30	-0.08, 1.09	0.093
	Unstructured	Unadjusted	Discrete	0.65	0.20	0.27, 1.04	<0.001	0.15	0.24	-0.32, 0.61	0.53	0.51	0.29	-0.07, 1.08	0.084
		Adjusted		0.60	0.20	0.21, 0.98	0.003	0.25	0.23	-0.21, 0.71	0.28	0.35	0.31	-0.25, 0.95	0.26
Fixed-Effects		Unadjusted	Continuous	0.70	0.19	0.33, 1.06	<0.001	0.19	0.22	-0.25, 0.63	0.39	0.51	0.29	-0.06, 1.07	0.080
			Discrete	0.67	0.19	0.29, 1.04	<0.001	0.17	0.23	-0.28, 0.62	0.47	0.50	0.29	-0.08, 1.08	0.089
Mixed-Effects	Combined	Unadjusted	Continuous	0.71	0.19	0.34, 1.07	<0.001	0.20	0.22	-0.23, 0.63	0.36	0.51	0.29	-0.06, 1.07	0.080
			Discrete	0.68	0.19	0.31, 1.05	<0.001	0.15	0.23	-0.29, 0.59	0.51	0.53	0.29	-0.04, 1.11	0.068
		Adjusted	Continuous	0.75	0.19	0.38, 1.13	<0.001	0.23	0.22	-0.20, 0.66	0.29	0.52	0.29	-0.04, 1.09	0.070
			Discrete	0.74	0.19	0.36, 1.11	<0.001	0.18	0.22	-0.26, 0.62	0.42	0.56	0.29	-0.02, 1.13	0.058
	By Group	Unadjusted	Continuous	0.70	0.19	0.33, 1.06	<0.001	0.20	0.22	-0.23, 0.63	0.36	0.50	0.29	-0.07, 1.06	0.084
			Discrete	0.68	0.19	0.30, 1.05	<0.001	0.15	0.23	-0.29, 0.59	0.51	0.53	0.29	-0.05, 1.10	0.071
		Adjusted	Continuous	0.75	0.19	0.37, 1.12	<0.001	0.23	0.22	-0.20, 0.66	0.29	0.52	0.29	-0.05, 1.08	0.073
			Discrete	0.73	0.19	0.35. 1.11	<0.001	0.18	0.22	-0.26, 0.62	0.43	0.55	0.29	-0.02, 1.13	0.060

TABLE 9.6: Estimated slope-change effect on ppFEV$_1$ when comparing those eligible for ivacaftor to those ineligible for ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Combined Analysis Step-Change Effect			
				est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	4.88	0.94	3.04, 6.73	<0.001
			Discrete	4.93	0.87	3.22, 6.64	<0.001
		Adjusted	Continuous	6.07	0.89	4.33, 7.82	<0.001
			Discrete	5.78	0.83	4.15, 7.42	<0.001
	Exchangeable	Unadjusted	Continuous	5.78	0.60	4.61, 6.96	<0.001
			Discrete	5.80	0.61	4.61, 7.00	<0.001
		Adjusted	Continuous	5.91	0.60	4.73, 7.09	<0.001
			Discrete	5.90	0.61	4.70, 7.10	<0.001
	Unstructured	Unadjusted	Discrete	5.57	0.63	4.33, 6.81	<0.001
		Adjusted		6.03	0.61	4.71, 7.10	<0.001
Fixed-Effects		Unadjusted	Continuous	5.90	0.61	4.71, 7.10	<0.001
			Discrete	5.92	0.62	4.71, 7.13	<0.001
Mixed-Effects	Combined	Unadjusted	Continuous	5.84	0.60	4.66, 7.02	<0.001
			Discrete	5.85	0.61	4.66, 7.04	<0.001
		Adjusted	Continuous	5.91	0.60	4.73, 7.09	<0.001
			Discrete	5.90	0.61	4.71, 7.09	<0.001
	By Group	Unadjusted	Continuous	5.84	0.60	4.66, 7.02	<0.001
			Discrete	5.85	0.61	4.66, 7.05	<0.001
		Adjusted	Continuous	5.90	0.60	4.72, 7.08	<0.001
			Discrete	5.90	0.61	4.71, 7.10	<0.001

TABLE 9.7: Estimated step-change effect on ppFEV_1 from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time period prior to the availability of ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Combined Analysis Slope-Change Effect			
				est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	0.35	0.47	-0.58, 1.28	0.46
			Discrete	0.32	0.43	-0.53, 1.17	0.46
		Adjusted	Continuous	0.37	0.42	-0.46, 1.19	0.38
			Discrete	0.48	0.38	-0.26, 1.23	0.21
	Exchangeable	Unadjusted	Continuous	0.53	0.26	0.02, 1.04	0.043
			Discrete	0.51	0.26	-0.00, 1.03	0.052
		Adjusted	Continuous	0.53	0.26	0.02, 1.04	0.040
			Discrete	0.54	0.27	0.02, 1.06	0.041
	Unstructured	Unadjusted	Discrete	0.49	0.26	-0.02, 1.00	0.061
		Adjusted		0.36	0.26	-0.16, 0.88	0.17
Fixed-Effects		Unadjusted	Continuous	0.55	0.26	0.04, 1.06	0.036
			Discrete	0.54	0.27	0.02, 1.06	0.043
Mixed-Effects	Combined	Unadjusted	Continuous	0.55	0.26	0.04, 1.07	0.035
			Discrete	0.56	0.27	0.04, 1.08	0.036
		Adjusted	Continuous	0.58	0.26	0.07, 1.08	0.027
			Discrete	0.59	0.26	0.07, 1.10	0.027
	By Group	Unadjusted	Continuous	0.55	0.26	0.03, 1.06	0.037
			Discrete	0.55	0.27	0.03, 1.08	0.039
		Adjusted	Continuous	0.58	0.26	0.07, 1.09	0.026
			Discrete	0.58	0.27	0.06, 1.10	0.028

TABLE 9.8: Estimated slope-change effect on ppFEV_1 from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time period prior to the availability of ivacaftor

9.3.2 Estimated Effect of Ivacaftor on Rate of Annual IV Days

As with the lung function results, we present the results of the effect of ivacaftor on the rate of annual IV days separately for each analysis. In all cases, as expected, the estimates are more heterogeneous than was the case for the lung function analyses, and this is due to the issue of non-collapsibility. In general, the estimates from the mixed-effects models were more extreme than the results from marginal or fixed-effects models, which were more similar to one another. Another issue encountered in these analyses is that a number of the mixed-effects models did not reach convergence. This is shown in the results tables and is the reason why some methods do not appear in the graphs of the results.

The results of analysis 1 are presented in Figures 9.11, 9.12 and 9.13, and in Tables 9.9, 9.10 and 9.11. For analysis 1 neither of the adjusted mixed-effects models reached convergence and hence only unadjusted mixed-effects models are presented. After one year, the NTE estimates all showed a sharp decrease in the rate of IV days in the post- versus pre-ivacaftor era: the smallest effect estimate was from an unadjusted marginal model with an unstructured working correlation matrix (IRR 0.61, 95% CI 0.50 to 0.74), and the largest effect estimate was from an unadjusted mixed-effects model with separate random effects for each group (IRR 0.30, 95% CI 0.21 to 0.44). However, the NCE estimates were also all smaller than 1 and statistically significant, suggesting a decline in the rate of IV days over time unrelated to ivacaftor. Here, the smallest effect estimate was from an unadjusted marginal model with an unstructured working correlation matrix (IRR 0.94, 95% CI 0.91 to 0.98), and the largest effect estimate was from an adjusted marginal model with an independent working correlation matrix (IRR 0.79, 95% CI 0.74 to 0.83). This again resulted in the NCCTE estimates all being slightly closer to 1, but all the results were still highly statistically significant, indicating a positive effect of ivacaftor. The smallest NCCTE effect estimate was from an unadjusted marginal model with an unstructured working correlation matrix (IRR 0.65, 95% CI 0.53 to 0.79), and the largest effect estimate was from an unadjusted mixed-effects model with separate random effects for each group (IRR 0.33, 95% CI 0.22 to 0.49). The results were remarkably stable over time with very similar effect estimates seen at two years and at three years.

The results of analysis 2 were very similar to those of analysis 1 and are shown in Figures 9.14, 9.15 and 9.16, and in Tables 9.12, 9.13 and 9.14. Unlike in analysis 1, in analysis 2 one of the adjusted mixed-effects model did achieve convergence and these results are presented, however the model which allowed the random effects to differ between groups still did not attain convergence. At the end of year one, the smallest NTE estimate was from an adjusted marginal model with an unstructured working correlation matrix (IRR 0.62, 95% CI 0.52 to 0.73), and the largest effect estimate was from the adjusted mixed-effects model (IRR 0.29, 95% CI 0.22 to 0.38). As with analysis 1, the NCE estimates were

also smaller than one, suggesting that on average people with gating mutations had fewer IV days per year than people with non-gating mutations even before ivacaftor was available. The smallest NCE estimate was from an unadjusted marginal model with an independent working correlation matrix (IRR 0.96, 95% CI 0.80 to 1.14), and the largest effect estimate was from an adjusted marginal model with an independent working correlation matrix (IRR 0.79, 95% CI 0.67 to 0.92). This resulted in the NCCTE estimates being slightly closer to 1 than the NTE estimates, but all were still highly statistically significant. The smallest effect estimate was from an unadjusted marginal model with an unstructured working correlation matrix (IRR 0.66, 95% CI 0.54 to 0.81), and the largest effect estimate was from an unadjusted mixed model with different random effects by group (IRR 0.34, 95% CI 0.23 to 0.50). As with analysis 1, the results in years two and three remained stable, suggesting the effect of ivacaftor on the rate of IV days does not change over time.

Finally, the results of the combined analysis are shown in Figure 9.17, and in Tables 9.15, 9.16 and 9.17. Again, these analyses directly estimate the NCCTE. In this case, only one of the four attempted mixed-effects models reached convergence. This was the unadjusted model with one set of random effects. At one year, the smallest effect estimate was from an unadjusted marginal model with an unstructured working correlation matrix (IRR 0.67, 95% CI 0.57 to 0.81), and the largest effect estimate was from the unadjusted mixed model (IRR 0.38, 95% CI 0.27 to 0.54). As with the other two analyses, these results stayed stable out to years two and three.

9.4 Discussion

All of the results that have been presented suggest quite a strong positive effect of ivacaftor on both the step-change in lung function at one year and the rate of IV days out to three years. The estimates of a change in the slope of lung function trajectory were also suggestive of a beneficial effect of ivacaftor, but these results were not always statistically significant. However, rather than just obtaining a general idea of the direction of the effect of ivacaftor on the outcomes, it would be nice to be able to give a more specific estimate of the treatment effect. Although the methods and analyses that we have considered were generally in agreement, there were a number of differences that sometimes could change the message of whether we believe there is enough evidence in the data to firmly conclude that the observed effects are real treatment effects. In the following subsections, we will compare the different analyses and methods and give suggestions as to whether some approaches may be more reliable than others for providing a summary result and for recommendation for future analyses.

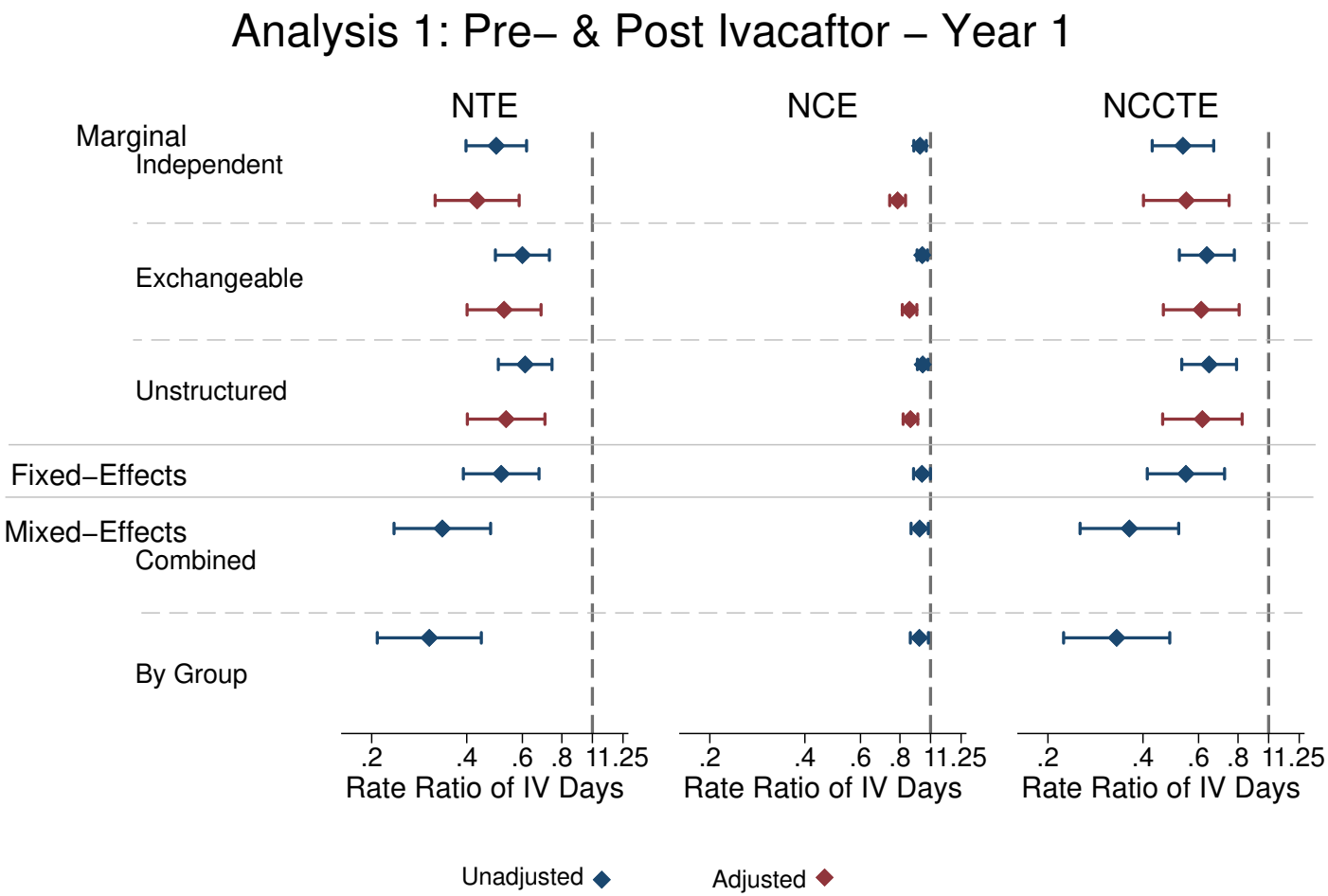


FIGURE 9.11: Estimated effect of ivacaftor on rate of annual IV days after one year of treatment when comparing post-ivacaftor period to pre-ivacaftor period

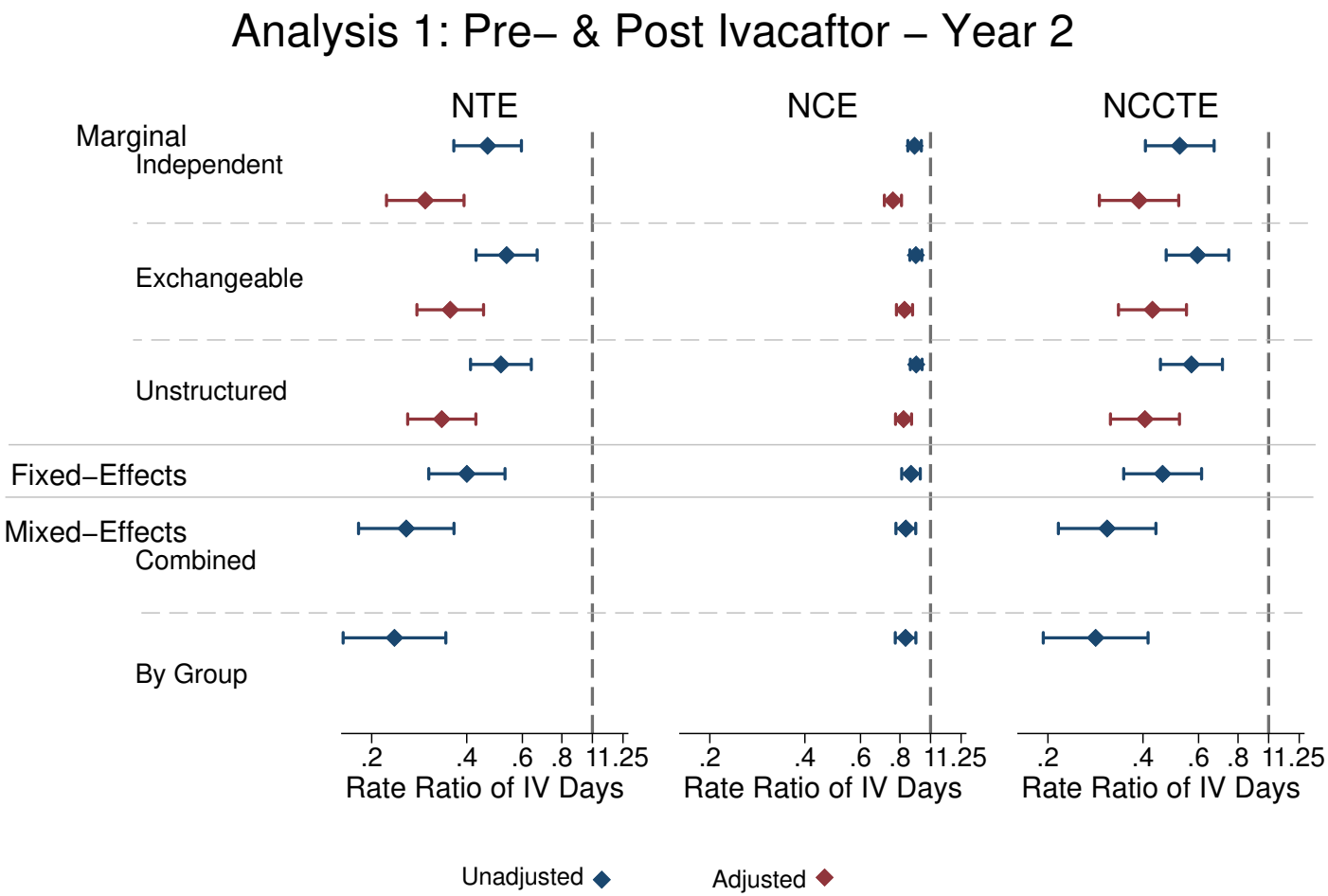


FIGURE 9.12: Estimated effect of ivacaftor on rate of annual IV days after two years of treatment when comparing post-ivacaftor period to pre-ivacaftor period

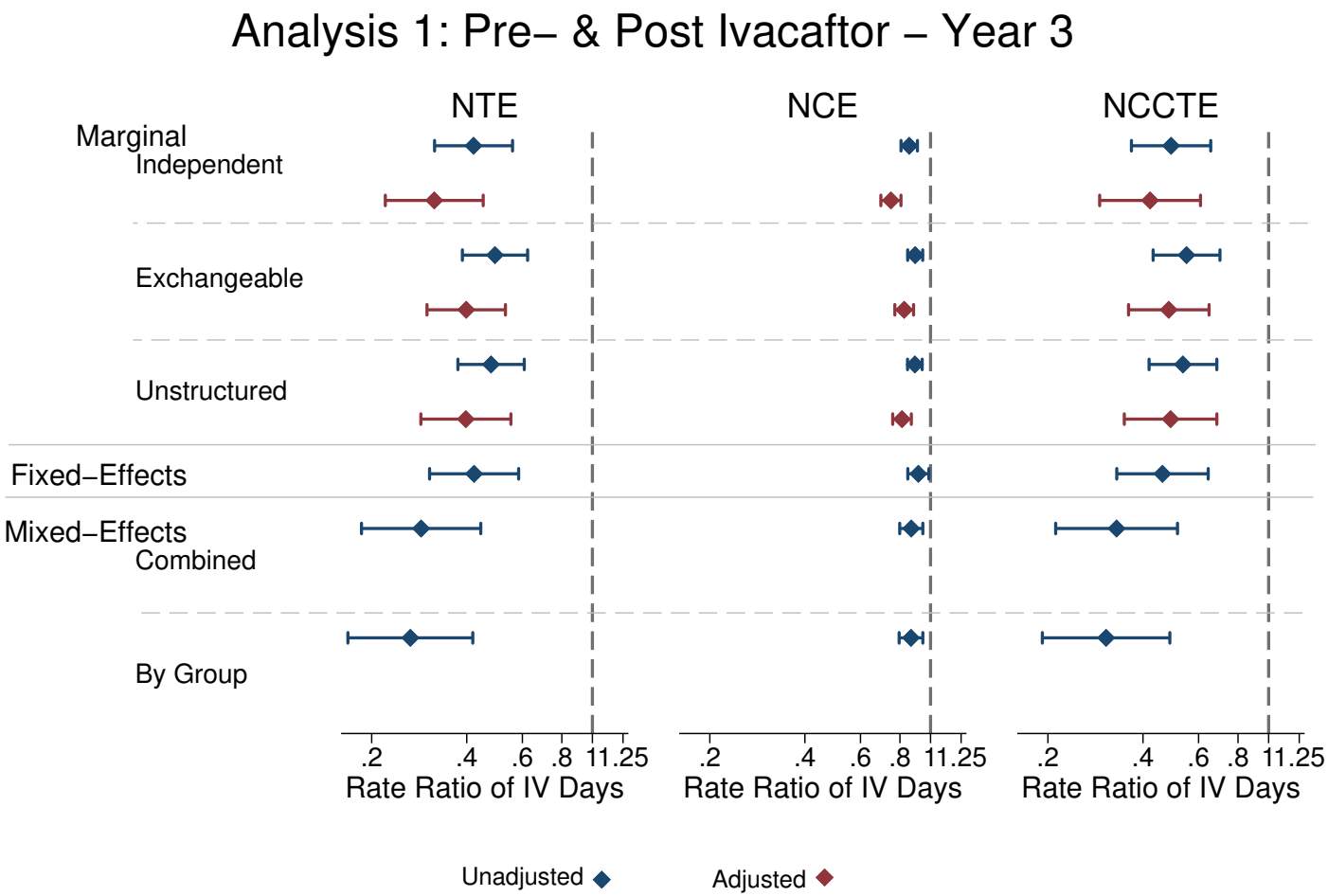


FIGURE 9.13: Estimated effect of ivacaftor on rate of annual IV days after three years of treatment when comparing post-ivacaftor period to pre-ivacaftor period

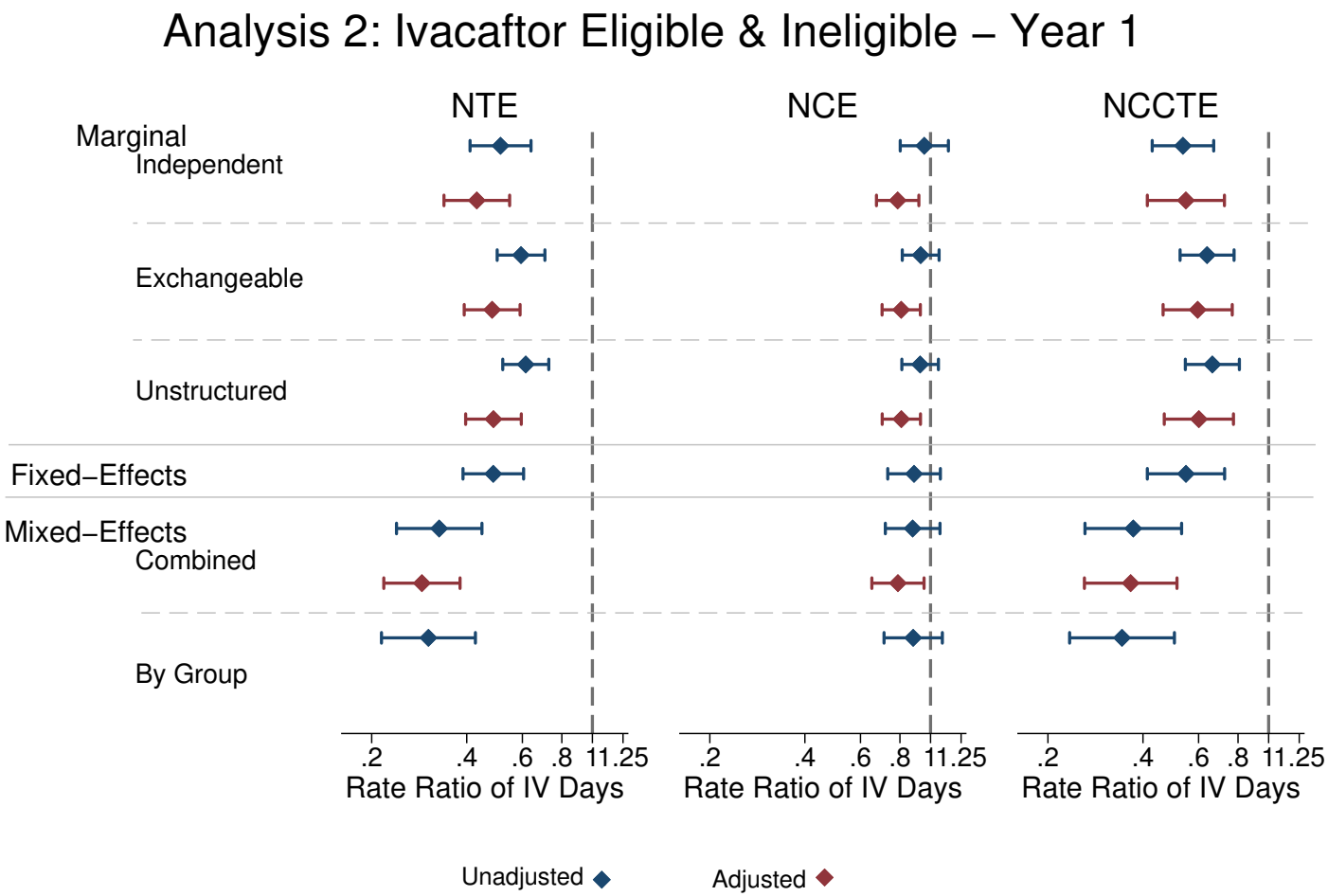


FIGURE 9.14: Estimated effect of ivacaftor on rate of annual IV days after one year of treatment when comparing those eligible for ivacaftor to those ineligible for ivacaftor

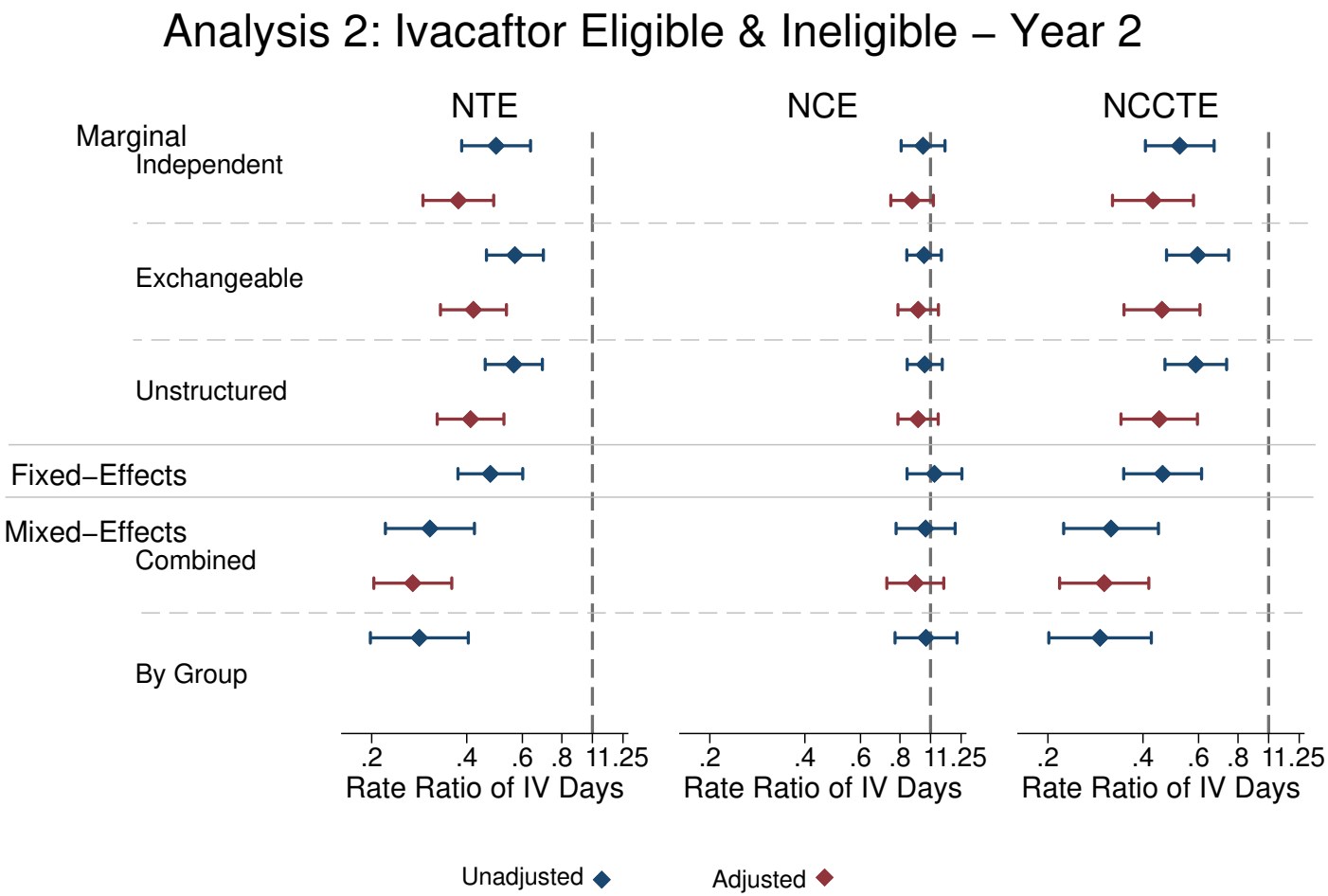


FIGURE 9.15: Estimated effect of ivacaftor on rate of annual IV days after two years of treatment when comparing those eligible for ivacaftor to those ineligible for ivacaftor

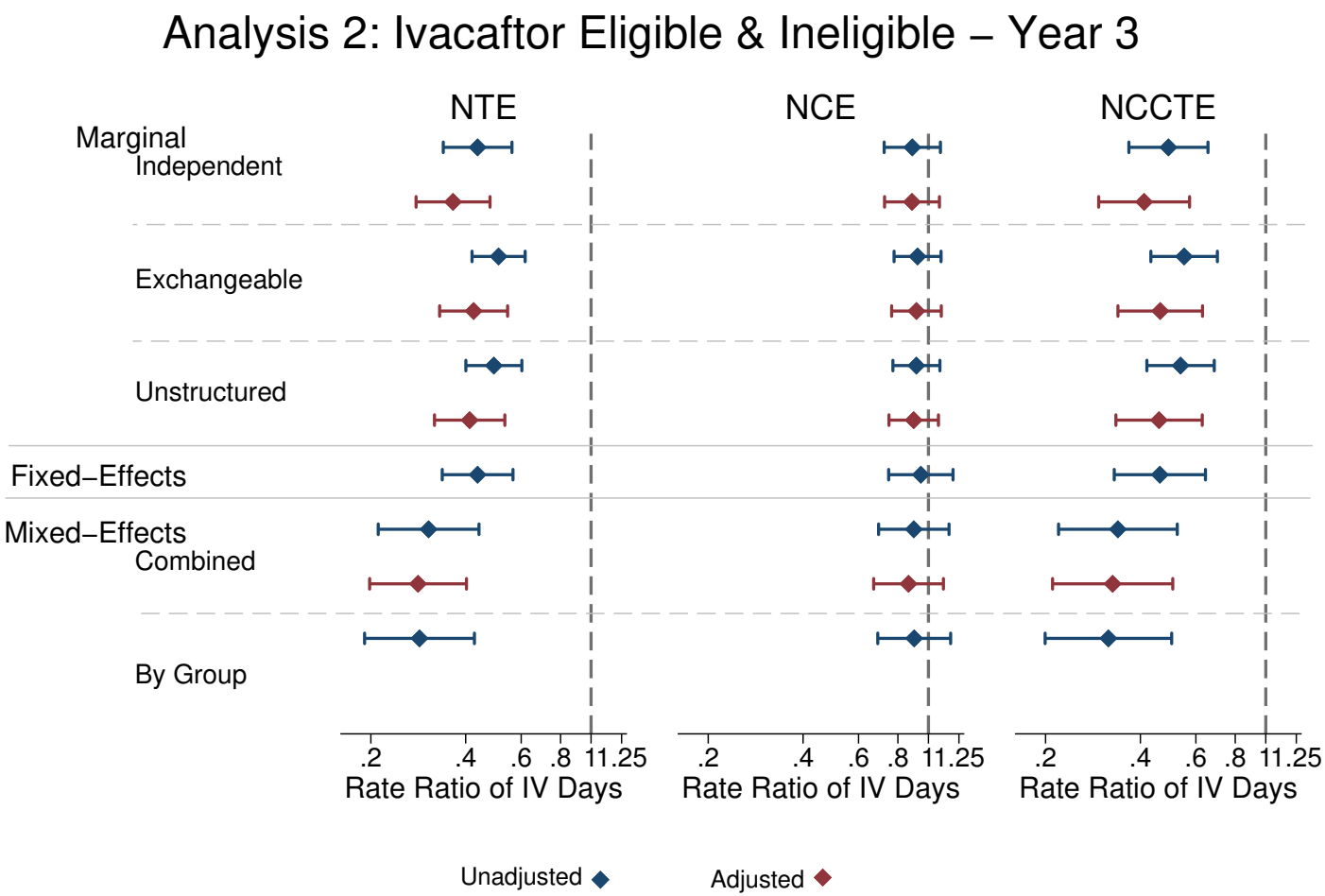


FIGURE 9.16: Estimated effect of ivacaftor on rate of annual IV days after three years of treatment when comparing those eligible for ivacaftor to those ineligible for ivacaftor

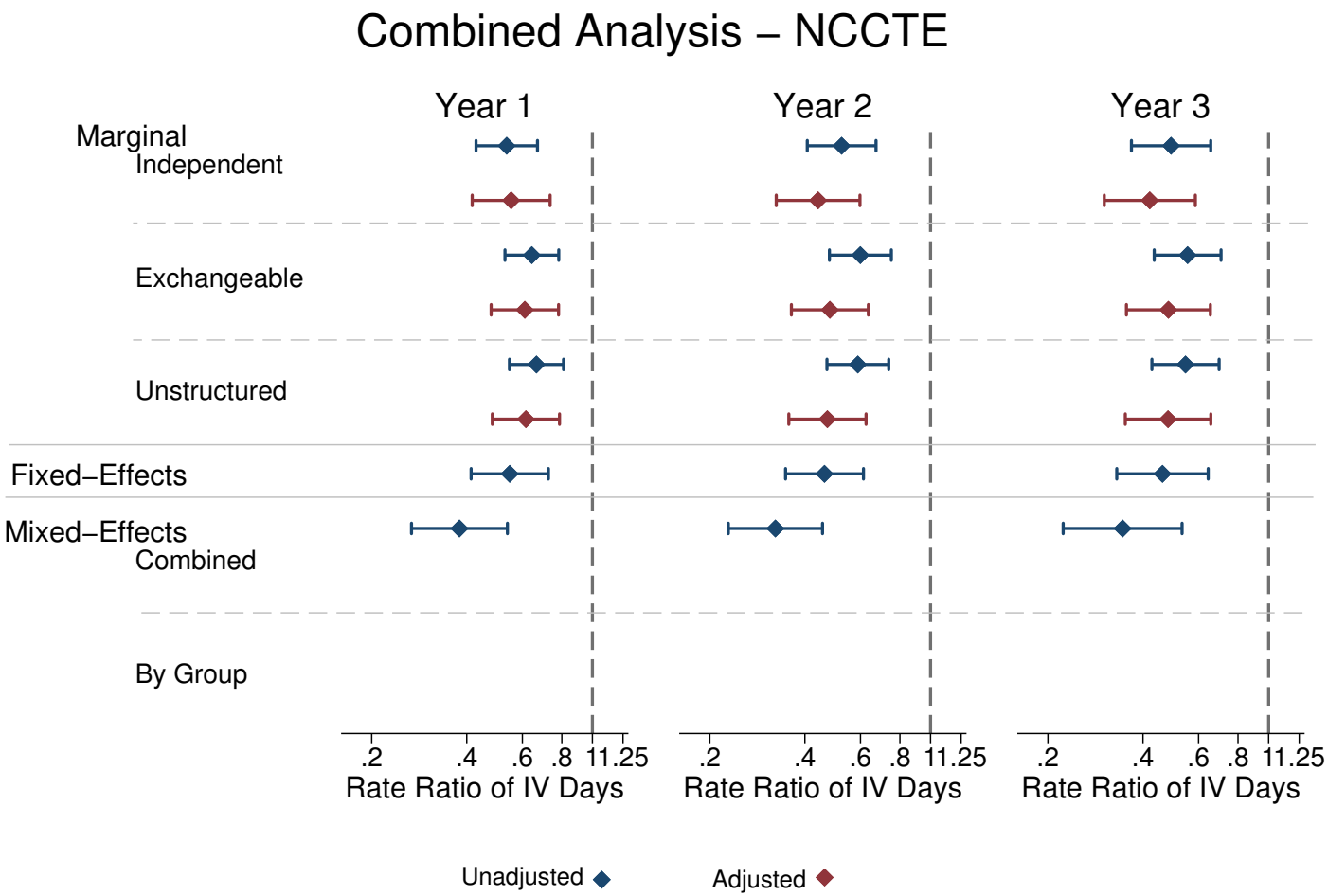


FIGURE 9.17: Estimated effect of ivacaftor on rate of annual IV days when comparing those currently receiving ivacaftor both to those not currently receiving ivacaftor and those in the pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Analysis 1: Pre- & Post Comparison - Year 1											
			NTE (B vs A)				NCE (D vs C)				NCCTE			
			IRR	SE	95% CI	p	IRR	SE	95% CI	p	IRR	SE	95% CI	p
Marginal	Independent	Unadjusted	0.50	0.06	0.40, 0.62	<0.001	0.93	0.02	0.89, 0.97	<0.001	0.54	0.06	0.43, 0.67	<0.001
		Adjusted	0.43	0.07	0.32, 0.59	<0.001	0.79	0.02	0.74, 0.83	<0.001	0.55	0.09	0.40, 0.75	<0.001
	Exchangeable	Unadjusted	0.60	0.06	0.49, 0.73	<0.001	0.94	0.02	0.91, 0.98	0.002	0.64	0.07	0.52, 0.78	<0.001
		Adjusted	0.53	0.76	0.40, 0.69	<0.001	0.86	0.02	0.81, 0.91	<0.001	0.61	0.08	0.46, 0.81	<0.001
	Unstructured	Unadjusted	0.61	0.06	0.50, 0.74	<0.001	0.94	0.02	0.91, 0.98	0.003	0.65	0.07	0.53, 0.79	<0.001
		Adjusted	0.52	0.07	0.40, 0.71	<0.001	0.86	0.02	0.82, 0.91	<0.001	0.62	0.09	0.46, 0.82	0.001
Fixed-Effects		Unadjusted	0.51	0.07	0.39, 0.68	<0.001	0.94	0.03	0.88, 1.00	0.047	0.55	0.08	0.41, 0.73	<0.001
Mixed-Effects	Combined	Unadjusted	0.33	0.06	0.24, 0.48	<0.001	0.92	0.03	0.87, 0.98	0.014	0.36	0.07	0.25, 0.52	<0.001
		Adjusted	Does not converge											
	By Group	Unadjusted	0.30	0.06	0.21, 0.44	<0.001	0.92	0.03	0.86, 0.98	0.015	0.33	0.07	0.22, 0.49	<0.001
		Adjusted	Does not converge											

TABLE 9.9: Estimated effect of ivacaftor on rate of annual IV days after one year of treatment when comparing post-ivacaftor period to pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Analysis 1: Pre- & Post Comparison - Year 2											
			NTE (B vs A)				NCE (D vs C)				NCCTE			
			IRR	SE	95% CI	p	IRR	SE	95% CI	p	IRR	SE	95% CI	p
Marginal	Independent	Unadjusted	0.47	0.06	0.36, 0.60	<0.001	0.89	0.02	0.85, 0.94	<0.001	0.52	0.07	0.41, 0.67	<0.001
		Adjusted	0.30	0.04	0.22, 0.39	<0.001	0.76	0.02	0.71, 0.81	<0.001	0.39	0.05	0.29, 0.52	<0.001
	Exchangeable	Unadjusted	0.53	0.06	0.43, 0.67	<0.001	0.90	0.02	0.86, 0.94	<0.001	0.60	0.07	0.47, 0.75	<0.001
		Adjusted	0.35	0.04	0.28, 0.45	<0.001	0.83	0.02	0.78, 0.88	<0.001	0.43	0.05	0.33, 0.55	<0.001
	Unstructured	Unadjusted	0.51	0.06	0.41, 0.64	<0.001	0.90	0.02	0.86, 0.94	<0.001	0.57	0.07	0.45, 0.71	<0.001
		Adjusted	0.33	0.04	0.26, 0.43	<0.001	0.82	0.02	0.77, 0.87	<0.001	0.41	0.05	0.32, 0.52	<0.001
Fixed-Effects		Unadjusted	0.40	0.06	0.30, 0.53	<0.001	0.87	0.03	0.81, 0.93	<0.001	0.46	0.07	0.35, 0.61	<0.001
Mixed-Effects	Combined	Unadjusted	0.26	0.05	0.18, 0.36	<0.001	0.84	0.03	0.78, 0.90	<0.001	0.31	0.06	0.22, 0.44	<0.001
		Adjusted	Does not converge											
	By Group	Unadjusted	0.24	0.05	0.16, 0.34	<0.001	0.83	0.03	0.77, 0.90	<0.001	0.28	0.06	0.19, 0.42	<0.001
		Adjusted	Does not converge											

TABLE 9.10: Estimated effect of ivacaftor on rate of annual IV days after two years of treatment when comparing post-ivacaftor period to pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Analysis 1: Pre- & Post Comparison - Year 3											
			NTE (B vs A)				NCE (D vs C)				NCCTE			
			IRR	SE	95% CI	p	IRR	SE	95% CI	p	IRR	SE	95% CI	p
Marginal	Independent	Unadjusted	0.42	0.06	0.32, 0.56	<0.001	0.86	0.03	0.81, 0.91	<0.001	0.49	0.07	0.37, 0.66	<0.001
		Adjusted	0.32	0.06	0.22, 0.45	<0.001	0.75	0.03	0.70, 0.81	<0.001	0.42	0.08	0.29, 0.61	<0.001
	Exchangeable	Unadjusted	0.49	0.06	0.39, 0.62	<0.001	0.89	0.02	0.85, 0.93	<0.001	0.55	0.07	0.43, 0.70	<0.001
		Adjusted	0.40	0.06	0.30, 0.53	<0.001	0.83	0.03	0.77, 0.88	<0.001	0.48	0.07	0.36, 0.65	<0.001
	Unstructured	Unadjusted	0.48	0.06	0.38, 0.61	<0.001	0.89	0.02	0.85, 0.94	<0.001	0.54	0.07	0.42, 0.69	<0.001
		Adjusted	0.40	0.06	0.29, 0.55	<0.001	0.81	0.03	0.76, 0.87	<0.001	0.49	0.08	0.35, 0.69	<0.001
Fixed-Effects		Unadjusted	0.42	0.07	0.30, 0.58	<0.001	0.92	0.04	0.85, 0.99	0.023	0.46	0.08	0.33, 0.64	<0.001
Mixed-Effects	Combined	Unadjusted	0.29	0.07	0.19, 0.44	<0.001	0.87	0.04	0.80, 0.95	0.001	0.33	0.08	0.21, 0.51	<0.001
		Adjusted	Does not converge											
	By Group	Unadjusted	0.27	0.06	0.17, 0.42	<0.001	0.87	0.04	0.80, 0.95	0.001	0.31	0.08	0.19, 0.49	<0.001
		Adjusted	Does not converge											

TABLE 9.11: Estimated effect of ivacaftor on rate of annual IV days after three years of treatment when comparing post-ivacaftor period to pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Analysis 2: Eligible & Ineligible - Year 1											
			NTE (B vs D)				NCE (A vs C)				NCCTE			
			IRR	SE	95% CI	p	IRR	SE	95% CI	p	IRR	SE	95% CI	p
Marginal	Independent	Unadjusted	0.51	0.06	0.41, 0.64	<0.001	0.96	0.09	0.80, 1.14	0.61	0.54	0.06	0.43, 0.67	<0.001
		Adjusted	0.43	0.05	0.34, 0.55	<0.001	0.79	0.06	0.67, 0.92	0.002	0.55	0.08	0.41, 0.72	<0.001
	Exchangeable	Unadjusted	0.59	0.05	0.50, 0.71	<0.001	0.93	0.06	0.81, 1.06	0.29	0.64	0.06	0.52, 0.78	<0.001
		Adjusted	0.48	0.05	0.39, 0.59	<0.001	0.81	0.06	0.70, 0.93	0.003	0.60	0.08	0.46, 0.77	<0.001
	Unstructured	Unadjusted	0.62	0.05	0.52, 0.73	<0.001	0.93	0.06	0.81, 1.06	0.27	0.66	0.07	0.54, 0.81	<0.001
		Adjusted	0.49	0.05	0.40, 0.60	<0.001	0.81	0.06	0.70, 0.93	0.003	0.60	0.08	0.47, 0.77	<0.001
Fixed-Effects		Unadjusted	0.49	0.06	0.39, 0.61	<0.001	0.89	0.09	0.73, 1.07	0.22	0.55	0.08	0.41, 0.73	<0.001
Mixed-Effects	Combined	Unadjusted	0.33	0.05	0.24, 0.45	<0.001	0.88	0.09	0.72, 1.07	0.20	0.37	0.07	0.26, 0.53	<0.001
		Adjusted	0.29	0.04	0.22, 0.38	<0.001	0.79	0.08	0.65, 0.95	0.014	0.37	0.06	0.26, 0.51	<0.001
	By Group	Unadjusted	0.30	0.05	0.21, 0.43	<0.001	0.88	0.10	0.71, 1.09	0.24	0.34	0.07	0.23, 0.50	<0.001
		Adjusted	Does not converge											

TABLE 9.12: Estimated effect of ivacaftor on rate of annual IV days after one year of treatment when comparing those eligible for ivacaftor to those ineligible for ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Analysis 2: Eligible & Ineligible - Year 2											
			NTE (B vs D)				NCE (A vs C)				NCCTE			
			IRR	SE	95% CI	p	IRR	SE	95% CI	p	IRR	SE	95% CI	p
Marginal	Independent	Unadjusted	0.50	0.06	0.39, 0.64	<0.001	0.95	0.08	0.81, 1.11	0.51	0.52	0.07	0.41, 0.67	<0.001
		Adjusted	0.38	0.05	0.29, 0.49	<0.001	0.87	0.07	0.75, 1.02	0.090	0.43	0.06	0.32, 0.58	<0.001
	Exchangeable	Unadjusted	0.57	0.06	0.46, 0.70	<0.001	0.95	0.06	0.84, 1.08	0.47	0.60	0.07	0.48, 0.75	<0.001
		Adjusted	0.42	0.05	0.33, 0.53	<0.001	0.91	0.07	0.79, 1.06	0.23	0.46	0.06	0.35, 0.61	<0.001
	Unstructured	Unadjusted	0.56	0.06	0.46, 0.69	<0.001	0.96	0.06	0.84, 1.09	0.52	0.59	0.07	0.47, 0.74	<0.001
		Adjusted	0.41	0.05	0.32, 0.52	<0.001	0.91	0.07	0.79, 1.06	0.22	0.45	0.06	0.34, 0.60	<0.001
Fixed-Effects		Unadjusted	0.48	0.06	0.38, 0.60	<0.001	1.03	0.11	0.84, 1.26	0.78	0.46	0.07	0.35, 0.61	<0.001
Mixed-Effects	Combined	Unadjusted	0.31	0.05	0.22, 0.42	<0.001	0.96	0.11	0.78, 1.20	0.74	0.32	0.06	0.22, 0.45	<0.001
		Adjusted	0.27	0.04	0.20, 0.36	<0.001	0.89	0.10	0.73, 1.10	0.30	0.30	0.05	0.22, 0.42	<0.001
	By Group	Unadjusted	0.28	0.05	0.20, 0.40	<0.001	0.97	0.11	0.77, 1.21	0.77	0.29	0.06	0.20, 0.43	<0.001
		Adjusted	Does not converge											

TABLE 9.13: Estimated effect of ivacaftor on rate of annual IV days after two years of treatment when comparing those eligible for ivacaftor to those ineligible for ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Analysis 2: Eligible & Ineligible Comparison - Year 3											
			NTE (B vs D)				NCE (A vs C)				NCCTE			
			IRR	SE	95% CI	p	IRR	SE	95% CI	p	IRR	SE	95% CI	p
Marginal	Independent	Unadjusted	0.44	0.06	0.34, 0.56	<0.001	0.89	0.09	0.72, 1.09	0.26	0.49	0.07	0.37, 0.66	<0.001
		Adjusted	0.36	0.05	0.28, 0.48	<0.001	0.89	0.09	0.73, 1.08	0.24	0.41	0.07	0.29, 0.57	<0.001
	Exchangeable	Unadjusted	0.51	0.05	0.42, 0.62	<0.001	0.92	0.08	0.78, 1.10	0.36	0.55	0.07	0.43, 0.70	<0.001
		Adjusted	0.42	0.05	0.33, 0.54	<0.001	0.92	0.08	0.76, 1.10	0.34	0.46	0.07	0.34, 0.63	<0.001
	Unstructured	Unadjusted	0.49	0.05	0.40, 0.60	<0.001	0.92	0.08	0.77, 1.09	0.32	0.54	0.07	0.42, 0.69	<0.001
		Adjusted	0.41	0.05	0.32, 0.53	<0.001	0.90	0.08	0.75, 1.08	0.24	0.46	0.08	0.33, 0.63	<0.001
Fixed-Effects		Unadjusted	0.44	0.06	0.34, 0.57	<0.001	0.95	0.11	0.75, 1.20	0.64	0.46	0.08	0.33, 0.64	<0.001
Mixed-Effects	Combined	Unadjusted	0.31	0.06	0.21, 0.44	<0.001	0.90	0.12	0.69, 1.16	0.42	0.34	0.08	0.22, 0.52	<0.001
		Adjusted	0.28	0.05	0.20, 0.40	<0.001	0.86	0.11	0.67, 1.12	0.26	0.33	0.08	0.21, 0.51	<0.001
	By Group	Unadjusted	0.29	0.06	0.19, 0.43	<0.001	0.90	0.12	0.69, 1.18	0.44	0.32	0.08	0.20, 0.50	<0.001
		Adjusted	Does not converge											

TABLE 9.14: Estimated effect of ivacaftor on rate of annual IV days after three years of treatment when comparing those eligible for ivacaftor to those ineligible for ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Combined Analysis Year 1			
			IRR	SE	95% CI	p
Marginal	Independent	Unadjusted	0.54	0.06	0.43, 0.67	<0.001
		Adjusted	0.55	0.08	0.43, 0.73	<0.001
	Exchangeable	Unadjusted	0.64	0.06	0.53, 0.78	<0.001
		Adjusted	0.61	0.08	0.48, 0.78	<0.001
	Unstructured	Unadjusted	0.67	0.07	0.55, 0.81	<0.001
		Adjusted	0.62	0.08	0.48, 0.79	<0.001
Fixed-Effects		Unadjusted	0.55	0.08	0.41, 0.73	<0.001
Mixed-Effects	Combined	Unadjusted	0.38	0.07	0.27, 0.54	<0.001
		Adjusted	Does not converge			
	By Group	Unadjusted	Does not converge			
		Adjusted	Does not converge			

TABLE 9.15: Estimated effect of ivacaftor on rate of annual IV days after one year of treatment when comparing those currently receiving ivacaftor both to those not currently receiving ivacaftor and those in the pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Combined Analysis Year 2			
			IRR	SE	95% CI	p
Marginal	Independent	Unadjusted	0.52	0.07	0.41, 0.67	<0.001
		Adjusted	0.44	0.07	0.32, 0.60	<0.001
	Exchangeable	Unadjusted	0.60	0.07	0.48, 0.75	<0.001
		Adjusted	0.48	0.07	0.36, 0.64	<0.001
	Unstructured	Unadjusted	0.59	0.07	0.47, 0.74	<0.001
		Adjusted	0.47	0.07	0.36, 0.62	<0.001
Fixed-Effects		Unadjusted	0.46	0.07	0.35, 0.61	<0.001
Mixed-Effects	Combined	Unadjusted	0.32	0.06	0.23, 0.45	<0.001
		Adjusted	Does not converge			
	By Group	Unadjusted	Does not converge			
		Adjusted	Does not converge			

TABLE 9.16: Estimated effect of ivacaftor on rate of annual IV days after two years of treatment when comparing those currently receiving ivacaftor both to those not currently receiving ivacaftor and those in the pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Combined Analysis Year 3			
			IRR	SE	95% CI	p
Marginal	Independent	Unadjusted	0.49	0.07	0.37, 0.66	<0.001
		Adjusted	0.42	0.07	0.30, 0.59	<0.001
	Exchangeable	Unadjusted	0.55	0.07	0.43, 0.71	<0.001
		Adjusted	0.48	0.08	0.35, 0.65	<0.001
	Unstructured	Unadjusted	0.55	0.07	0.43, 0.70	<0.001
		Adjusted	0.48	0.08	0.35, 0.66	<0.001
Fixed-Effects		Unadjusted	0.46	0.08	0.33, 0.64	<0.001
Mixed-Effects	Combined	Unadjusted	0.34	0.08	0.22, 0.53	<0.001
		Adjusted	Does not converge			
	By Group	Unadjusted	Does not converge			
		Adjusted	Does not converge			

TABLE 9.17: Estimated effect of ivacaftor on rate of annual IV days after three years of treatment when comparing those currently receiving ivacaftor both to those not currently receiving ivacaftor and those in the pre-ivacaftor period

9.4.1 Comparability of Groups

The key assumption of the NTE analyses is that the groups presented in Figure 9.2 are comparable except for the introduction of ivacaftor. We aimed to test this assumption through the use of negative controls and the results of these analyses suggested that on the whole the assumption of comparability of the groups was reasonable. The NCE estimates were all quite small and would likely not be deemed clinically significant. Some of the estimates were, however, statistically significant, and this was generally due to the very large sample size available in the non-gating mutation groups, which allowed very precise estimation of the NCE. Nevertheless, the fact that the estimates were not exactly zero and also tended to be in the same direction as the NTE estimates does suggest that there could be some small biases in the NTE estimates. This bias can be accounted for by calculating the NCCTE, which requires weaker assumptions about the comparability of the groups. Therefore, it should generally be preferable to use the NCCTE estimates of the NTE estimates.

The SEs of the NCCTE estimates were generally similar to the SEs of the NTE estimates, meaning that this was not a concern for these analyses. This was because in analysis 1 we were able to obtain such precise estimates of the NCE and in analysis 2 the same people were in both the NTE and the NCE resulting in a large covariance between the two estimates. In the case of analysis 2, this meant that sometimes the SE of the NCCTE was actually smaller than the SE of the NTE, as we gained precision through having the same individuals in both the main and the negative control analyses. A situation where there was only a relatively small negative control group could result in a much larger SE for the NCCTE. Depending on the analysis, this could make it difficult for the NCCTE to reach statistical significance, but this is not necessarily a bad thing, as the large SE would reflect the uncertainty regarding the comparability of the groups.

One of the benefits of analyses one and two is that the NCE are obtained separately. This allows us to directly see an estimate of how comparable the groups are. However, this benefit is probably outweighed by the benefit of the combined analysis, which gives a direct estimate of the NCCTE and which is generally more precise due to the larger sample size of combining both groups. Furthermore, although the combined analysis includes comparison to both people of a different genotype and people from a different time period, the overall assumption is actually the same as the assumption for the NCCTE estimates of the separate analyses: that there is no interaction between the effects of genotype and time period.

9.4.2 Comparison of Statistical Methods

We considered three main types of models: marginal models, fixed-effects models and mixed-effects models. In the case of the lung function analysis most of the methods gave very similar results. However, the point estimates from the marginal models using an independent working correlation matrix were more divergent from the other methods and also tended to have larger SEs. Checking the estimated correlations from exchangeable and unstructured correlation matrices suggests that there is strong correlation in repeat measures from the same individual and this would suggest that although marginal models can be used, the working correlation matrix should allow for correlations between the measures. Taking the unadjusted marginal model for the NTE of analysis 1: using an exchangeable correlation matrix estimated the correlation between all measures for the same individual to be 0.87, and using an unstructured correlation matrix, the correlations were all estimated to be between 0.82 and 0.91. This suggests that here the use of an exchangeable working correlation matrix is sufficient as the correlations between measures made at different time points do not vary much and furthermore the exchangeable working correlation matrix does not need the visits to be equally spaced allowing it to be used with the continuous measure of time.

For mixed-effects models, preliminary analyses suggested that the random effects structure should be separated by group, as it seemed that lung function measures generally became less variable when taking ivacaftor. This finding was repeated in the final analyses, where allowing separate random effects by group was shown to lead to a statistically significantly better fit to the data. However, this change in random effects structure had almost no effect on the treatment effect estimates, suggesting that as long as the random effects specification is approximately correct it will not lead to any observable bias in the estimation of fixed-effects terms.

For the lung function analyses, thanks to collapsibility, the estimands for all models are the same and the observed estimates found across methods were similar as we expected. This would suggest that when choosing an analysis approach, we should aim to choose the method that could give the smallest SE. However, our results show that all analyses also gave almost identical SEs. This means that in the case of lung function, it does not matter whether one uses a marginal model (as long as an independent working correlation matrix is not used), a fixed-effects model or a mixed-effects model.

When considering the analysis with the number of IV days as the outcome, the same problem of using marginal models with an independent working correlation matrix was observed. Although the correlation between observations was less than in the case of lung function, it was still relatively large: the exchangeable working correlation matrix estimated the correlation between observations from the same person to be 0.59, and the unstructured working correlation matrix gave estimates between 0.47 and 0.73.

With a non-linear model, we no longer expect the estimates from the different models to be the same due to non-collapsibility, i.e. the estimands are different. This means we have to consider what effect estimate we are interested in: the unadjusted marginal models give the population-average effect, whereas the mixed-effects and fixed-effects models estimate the subject-specific effect. The former is the effect we would see over the whole population if everyone received ivacaftor compared to if nobody received ivacaftor. The latter is the effect we would expect to see in one person when they are taking ivacaftor compare to if they were not taking ivacaftor. Surprisingly, the effect estimates from the fixed-effects models were actually closer to the estimates from the marginal model than the estimates from the mixed-effects model. It is known that fixed-effects regression should not be used with logistic models, but it has been suggested that this was not an issue with negative binomial models.[268] However, in our analysis more than half of the observed IV days outcomes are zeroes, which could be why these estimates do not coincide with those of the mixed-effects models. Therefore, in the case of non-continuous outcomes, we suggest either using marginal models or mixed-effects models and avoiding fixed-effects models.

One further issue of non-linear models is that a number of the mixed-effects models did not converge. This is generally an issue as the models become more complicated, i.e. by using more complex random effects structures or adjusting for more variables. As our results appear to suggest that adjusting for additional variables or using more complex random effects structures make little difference, this should not affect our results and we can safely just use the results from the simpler mixed-effects models that did converge.

9.4.3 Adjustment for Measured Baseline Health Variables

As shown in Figure 9.4, we expected the use of the negative control to correct the treatment effect estimates whether observed measures of baseline health were adjusted for or not. This appears to be correct, as generally the unadjusted and adjusted NCCTE estimates were more similar to each other than the unadjusted and adjusted NTE estimates, as the NCE estimates corrected for these differences. However, in reality, even the unadjusted and adjusted NTE estimates were remarkably similar. This suggests that the four groups were actually quite well balanced at baseline, which is what was shown in Table 9.1. If there had been imbalances in the groups at baseline, the use of the negative control should still have been able to account for this, but additionally adjusting for $\boldsymbol{H}$ should have decreased the SE of the treatment effect estimates. As this was not the case in our analyses, the SEs were very similar regardless of adjustments.

9.4.4 Accuracy of Time Measurement

For the lung function analysis, we performed all analyses twice, one using the precise estimate of time since baseline and the other just using visit number (1, 2, 3 or 4). As presented in Table 9.2, on average the visits do happen annually, which suggested that ignoring the exact timing of the visits may not have a big impact on the results. It was hoped, however, that using the more precise estimates of time could improve the efficiency of the methods. The results of the analysis show that this was not the case, as there was barely any difference in either the point estimates or the SEs of any of the results when using continuous or discrete measures of time. This is actually a positive result though as it suggests that there will be no bias in just using the discrete measure of time in situations where it is not appropriate to use the continuous measure of time, such as in the IV days analysis or in any of the methods presented in Part II of this thesis. However, even though it makes little difference, in situations where it is possible, it is probably best to use the more accurate continuous measure of time.

9.4.5 Computation Time

There were two procedures which heavily affected the computation time of the analyses: the bootstrap and the mixed-effects models. In isolation, neither of these procedures resulted in unmanageable computation times, but put together the analyses would take many days. Obviously, if the only correct method was a mixed-effects model and if it were necessary to use a bootstrap to obtain the SEs then the computation time would be unavoidable. However, the bootstrap can be avoided provided that the same person is not in the NTE analysis and the NCE analysis, as in that situation, the covariance between the two estimates is known to be zero. Similarly, the combined analysis directly obtains the NCCTE, and so the bootstrap is not necessary there either. In fact, in the analyses presented in this chapter, the bootstrap was only necessary to obtain the NCCTE estimates of analysis 2. As the results of analysis 2 were very similar to those of analysis 1 and the combined analysis, it would suggest that we could have avoided the bootstrap altogether.

9.4.6 Long-Term Effects of Ivacaftor

Finally, we present the results of the chosen method for each outcome. For lung function, we use the combined analysis and an adjusted marginal model with an exchangeable working correlation matrix and time measured continuously. Here, ivacaftor was estimated to result in an step-change increase in ppFEV$_1$ of 5.91% after one year of treatment (95% CI 4.73% to 7.09%, $p < 0.001$). It also resulted in a decrease in the annual rate of lung function decline of 0.53% (95% CI 0.02% to 1.04%, $p = 0.040$). For reference, if this

group of people were not currently receiving ivacaftor their annual rate of lung function decline was estimated to be -1.15% (95% CI -1.52% to -0.78%). This suggests that the rate of lung function decline has almost been halved.

For IV days, we present both an estimate of the population-average effect and an estimate of the subject-specific effect. Firstly, the population-average effect is taken from the combined analysis using an unadjusted marginal model with an exchangeable working correlation matrix. (We use the unadjusted rather then adjusted model here, so that the effect estimates are truly marginal, which was not an issue with the lung function results). Here, the IRR of IV days after one year of using ivacaftor was estimated to be 0.64 (95% CI 0.53, 0.78), after two years the effect became slightly stronger, 0.60 (95% CI 0.48, 0.75), and then slightly stronger again after three years, 0.55 (95% CI 0.43, 0.71).

For the subject-specific estimates we again use the combined analysis, but this time with the unadjusted mixed-effects analysis. Here the estimated treatment effect at one year was 0.38 (95% CI 0.27, 0.54), at two years was 0.32 (95% CI 0.23, 0.45), and at three years was 0.34 (95% CI 0.22, 0.53). These subject-specific estimates are all stronger than the population-average effect estimated, and this is what is generally expected to be observed in non-linear models.

9.4.7 Comparison of Results to Other Studies

The systematic review in Chapter 7 identified two RCTs with one year of follow-up. The estimated step-change effect on ppFEV$_1$ in these studies was 10.0% (95% CI 4.3% to 15.7%) and 10.5% (5.3% to 15.7%).[211, 213, 214, 216, 224] This is quite a bit larger than the estimates from our study, but previous observational studies showed more comparable results of between a 3.2% improvement and an 8.3% improvement.[236, 251] These observational studies did only estimate the NTE, but as our NCE estimates were very close to zero, it suggests that these estimates are also reliable.

Only one study had previously investigated the effect of ivacaftor on the rate of lung function decline. In this study, which had three years of follow-up, the rate of annual decline was estimated to decrease by 0.81% (95% CI 0.08% to 1.54%). Again this is a NTE, and it is slightly larger than our NCCTE estimate, but overall they are similar estimates.[236]

There have been a large number of studies that have looked at the RR of exacerbations due to ivacaftor, but relatively fewer studies that investigated the effect of ivacaftor on IV days. The studies that have included IV days as an outcome all looked at absolute differences in IV days between baseline and follow-up and therefore did not estimate an IRR as we did in our study. For this reason, we cannot directly compare our results to any previous studies for this outcome.

9.4.8 Limitations

One of the key strengths of our analysis is that we have assessed and hopefully corrected for any differences between groups not due to ivacaftor. However, the methods used still rely on the assumption that any differences between genotypes have not changed over time, and similarly that any differences over time are the same regardless of genotype. The fact that in our analyses the NCE were very close to zero suggests that the groups were comparable anyway and therefore it seems reasonable to assume that the groups remained comparable after the introduction of ivacaftor. However, if along with the introduction of ivacaftor, there was also a change in healthcare policy only affecting those receiving ivacaftor, e.g. if those receiving ivacaftor also started to attend the clinic for more regular check-ups, then this could affect the validity of our analyses. We do not know of any changes of any nationwide changes in healthcare practice that could have had such an effect, but there remains the possibility that some local centres may have started treated those on ivacaftor differently, which could affect the results.

One of the key aims of this work is to show how the effects of long-term treatment use can be estimated. In this analysis, we are limited by the fact that ivacaftor has only been available since 2012 and therefore we can only estimate the effects of receiving treatment for four years. Over this four year period the different groups appeared to be more or less comparable, which suggested that the use of the negative controls may not have actually been necessary, as the NTE would still have provided good estimates of the treatment effect. However, with longer follow-up time it seems likely that the comparability of the groups would diminish, necessitating the use of the negative controls. For example, if we wish to look at the effects of receiving treatment for ten years compared to never receiving treatment, we would need to compare the most recent ten years to the ten years prior to that. Expecting observations that are ten years apart to be directly comparable is much less reasonable than assuming that observations that are four years apart are comparable. Similarly, the long-term follow-up of people with different genotypes is more likely to diverge as the length of follow-up time increases. Theoretically, the negative control methods introduced here should still work even with much longer follow-up times, but it would be interesting to assess that with real data once it is available.

A further issue is that ivacaftor is the first disease-modifying treatment available to people with CF. This meant that we knew that everyone not receiving ivacaftor was not receiving any other kind of disease modifying treatment and these people could then be used as a control group. There are many treatments in the pipeline that will hopefully bring disease-modifying treatments to a much larger number of people with CF. As the number of people receiving these treatments increases, the ease of finding treatment-free control groups will rapidly diminish. This is not a limitation of our current study, but it does mean that the methods introduced here may not be applied so easily in the future without adaptation.

In our analyses, we modelled lung function trajectory over time linearly. As we only had four lung function measures per person, it was not really possible to model the trajectory non-linearly without risk of overfitting. Assuming a linear trajectory is the most common way to model lung function in these types of analyses and one of the benefits of this is that it is easily interpretable. However, long-term lung function decline is known to not be linear and other modelling strategies such as including a quadratic term or using splines could be considered.[20]

There are also known to be limitations to using IV days as a proxy for exacerbations, as sometimes people will receive IVs despite not currently suffering from an exacerbation. If these protective courses of IVs became less common once people started to receive ivacaftor, then the results of our analyses would be biased overestimating the effect of ivacaftor on improving exacerbation rates. However, given that the estimated effect was so strong on decreasing the rate of IV days, it does not seem plausible that all of this effect could come just from protective IV days, and therefore the results do suggest that ivacaftor also reduces the rate of preventative IV days as well.

Finally, the average long-term lung function trajectories can become biased due to deaths during follow-up. For example, if some people had very steep lung function declines in the first couple of years of follow-up, but then subsequently died, long term the average slope would be estimated to become more shallow even though this is not actually the case for specific individuals. One possible solution to this issue is to use an approach known as joint modelling. Joint models simultaneously model both the survival and the longitudinal trajectories, and include terms to estimate how the two functions are related.[271] This approach does have its own drawbacks as the results obtained refer to a hypothetical group of people where nobody dies.[272] Depending on the length of follow-up and the age of the people in the cohort this may or may not therefore provide sensible results. Furthermore, the survival models used in joint modelling cannot actually be used to estimate the treatment effect on survival. This is because, the survival trajectories are affected by the time-updating lung function measures, which are on the causal pathway between treatment and survival. Therefore, a separate model would need to be fit if interest was in the total effect of treatment on survival. It is for these reasons, along with the fact that with our relatively short follow-up time there were not many deaths, that we did not consider the use of joint models here.

9.4.9 Conclusion

The analyses presented in this chapter concord with previous studies that there is a very beneficial effect of ivacaftor in terms of both improving lung function and reducing the rate of annual IV days. Furthermore, there does appear to be evidence that the slope of

lung function decline is slowed when taking ivacaftor, although longer follow-up would be needed to confirm this finding.

Previous observational studies have relied on the assumption that the group of people currently receiving ivacaftor are comparable either to themselves in the time period prior to the availability of ivacaftor or to people who are not eligible to receive ivacaftor due to genotype. In this study, we have shown how the use of negative controls can be used to test these assumptions and if necessary to correct for any differences between the groups not due to treatment. Over the four years of follow-up of this study, it appeared that the assumption of comparability of groups was met, suggesting that the treatment effect estimates from other observational studies are unlikely to be biased. Nevertheless, even though the negative controls showed relatively little difference between the groups not due to ivacaftor, we still believe that it is beneficial to incorporate this uncertainty into the final treatment effect estimate.

Finally, there are a wide variety of methods that can be used with longitudinal data, and it not always necessary to resort straightaway to mixed-effects models. These are the most flexible of the models we considered, but if interest only lies in the average treatment effect, then this can be obtained with much simpler methods that will also be computationally much quicker. However, in terms of both lung function and IV days, it appears that the assumption of an independent working correlation matrix is unreasonable, and for this reason standard regression methods should be avoided. Using an exchangeable working correlation matrix appeared to be sufficient in our examples, but other types of working correlation matrix could also be considered. Fixed-effects methods are also a simple alternative, but our results suggest that they should be used with caution for non-linear models.

Part IV

Discussion

Chapter 10

Discussion

10.1 Summary of Findings

The main findings from this thesis can be summarised in four main sections: methods for dealing with time-dependent confounding, the long-term effects of DNase, the use of negative controls in observational studies, and the long-term effects of ivacaftor.

10.1.1 Time-dependent confounding

Two chapters of this thesis were devoted to issues surrounding time-dependent confounding. In Chapter 4, methods that can deal with time-dependent confounding were introduced and then in Chapter 5 the performance of these methods under different scenarios was investigated with simulation studies.

As a brief reminder, a time-dependent confounder is defined as a covariate measured during follow-up that: 1) is affected by previous treatment, 2) affects the probability of future treatment, and 3) affects the outcome of interest.[80] Time-dependent confounders cannot generally be handled like baseline confounders, where for example they could be adjusted for in a multivariable model, because of the fact that they are themselves affected by previous treatment, meaning that as well as being confounders they are also mediators of the total effect of treatment.

In this thesis, five methods were considered that can estimate treatment effects in the presence of time-dependent confounding: IPW of MSM[84], HA-MSM[87], g-formula[88], g-estimation of SNM[90] and SCMM[93]. Although all of these methods can be used to remove bias due to time-dependent confounding, there were no clear guidelines on whether there are situations where one method would be preferred over another. IPW of MSM is probably the most popular of the five methods, having recently become quite a common method used in epidemiological research. However, this is believed to be due to its perceived simplicity rather than due to any superiority of this method over other methods.[92]

The findings from the simulation studies in this thesis confirmed that in ideal settings there was relatively little difference between the five methods. All five methods provided unbiased estimates of the treatment effect and there was relatively little difference between the SEs of the methods. Ideal settings refers to a scenario where the following assumptions are valid: no interference, positivity, consistency, and no unmeasured confounding. Furthermore, the models are assumed to be correctly specified in terms of the parametric functional forms, short-term and long-term causal pathways, interaction effects, the direction of causal pathways, and any censoring is accounted for.

Both IPW of MSM and HA-MSM performed very similarly to one another in all scenarios, but HA-MSM can provide estimates of effect modification by time-varying covariates. Standard IPW methods are agnostic on whether there is effect modification meaning they do not provide biased results if there is effect modification, but they can only ever give estimates of the population-average effect. Ensuring that the weights are combined correctly through time is more difficult with HA-MSM than with the standard IPW method though, suggesting that unless there is particular interest in estimating any effect modification that standard IPW methods be used over HA-MSM.

G-formula requires parametric models to be specified for all time-varying covariates and the resulting estimates will be biased if any of these models are misspecified. In our scenarios, we only had two time-varying confounders and the amount of bias from g-formula was never noticeably more than from the other methods considered. However, in other settings, where there might be many time-varying covariates the method could quickly become very difficult to use. One benefit of g-formula over the other methods considered is that it is very easy to compare many different types of long-term treatment patterns. In this thesis, we have focussed on comparing taking treatment continuously to never taking treatment, but g-formula could easily estimate the effects of other trajectories, such as taking treatment for two years, stopping treatment for two years and then restarting for another two years.

G-estimation of SNM was one of the methods that allowed for the estimation of effect modification by time-varying covariates and in all scenarios studied in the simulations it performed similarly to the other methods.

Finally, SCMM could only be used to estimate the one-year effect of treatment in our situation. However, it was shown to be able to fully account for all confounding of this short-term effect, (provided that all confounders are observed), meaning that the more complex methods introduced in this thesis do not always have to be used when there is time-dependent confounding. As this method is a much simpler method than the other methods introduced, essentially being a multivariable regression model, we suggest this as a good starting point for analyses to first look at the short-term effects, before deciding whether to continue to investigate the long-term effects with the other available methods.

For our specific analysis, it was clear from the preliminary short-term analyses using SCMM that there was modification of the treatment effect by time-varying covariates. For this reason, it was most suitable to use either HA-MSM or g-estimation for the long-term effects analyses. Although the other two methods could still be used to provide unbiased estimates of the population-average effect of treatment.

Perhaps unsurprisingly, misspecification of the direction of causal pathways resulted in biased treatment effect estimates for all five methods. However, IPW of MSM and HA-MSM were shown to perform especially poorly in these situations. Truncation of extreme weights reduced the bias in these situations, but in a real-life setting, it would be more difficult to know when truncation of the weights would lead to less biased estimates. For this reason, in situations where it is difficult to know the direction of causal pathways with certainty, it is recommended not to use IPW or MSM or HA-MSM. This situation would be common whenever data are collected at regular intervals, but the data are related to the time period since the last data were collected.

10.1.2 Dornase Alfa

From the systematic review presented in Chapter 3, it is clear that there have not been many studies investigating the effect of long-term DNase use. The longest study had four years of follow-up and the results did suggest that the rate of decline of lung function was slowed by DNase, but the study only had 60 participants.[60] No other study had longer than two years of follow-up, but in general the findings were all positive that DNase improved lung function and reduced the rate of pulmonary exacerbations. Almost all studies were, however, restricted to patients who already showed declines in lung function, and therefore may not be applicable to people taking DNase with high lung function levels.

In the UK CF Registry, we saw very variable rates of DNase use depending on CF centre, and this is generally due to different practices, where some centres routinely initiate DNase when people reach a certain age (often 6 years old), whereas other centres wait until lung function drops below a certain level (often 80%). Overall, there was clear confounding by indication in the UK CF population, with those taking DNase having worse health on average, but there was a lot of overlap and positivity therefore did not appear to be an issue for the analysis.

Based on the findings of the simulation studies, it appeared that any of the five methods could be used to estimate the effects of DNase. However, there was uncertainty about the direction of the causal pathway between DNase and IV days in a given year and for this reason, IPW of MSM and HA-MSM, may perform particularly poorly if this is misspecified.

Initially, the data were analysed using SCMM to estimate the effect of one-year of DNase use on lung function. The population-average estimates showed almost no effect of DNase on lung function. However, there was evidence that the treatment effect was modified by previous levels of lung function, with those with lower lung function seeing improvements in lung function after one year of DNase use.

Long term, the effect modification by previous levels of lung function appeared to remain important suggesting that either g-estimation or HA-MSM be used. The results from both of these methods were very similar, estimating the treatment effect to remain more or less stable between years one and five. This suggests that DNase does not alter the trajectory of lung function decline, as at five-years the difference between untreated and treatment was similar to what it was at one-year. The confidence intervals did, however, become much larger at five years due to the smaller sample size of people who had continuously used treatment for this period of time.

One of the main limitations of our study (and indeed most observational studies) was that it is not possible to confirm whether confounding had been adequately accounted for or whether there remained important unobserved confounders. For the lung function analysis, the main confounder appeared to be previous lung function and once this was included in the models, there was actually little effect of including a range of other potential confounders. However, we also wished to investigate the effects of DNase on the rate of IV days, and in these analyses, the results also found that DNase was having a strong negative effect on the rate of IV days. It seems unlikely that this is a real effect of DNase, and suggests that there are some unmeasured confounders that affect both the probability of receiving DNase and the probability of receiving IVs.

One of the assumptions of all the methods we considered for this analysis is that visits are equally spaced. As shown in Figure 2.2, annual assessments are in most circumstances carried out very close to annually, but a small number of people do have much smaller or bigger gaps between assessments. It is not clear if this may affect the results of the analysis, but for the ivacaftor analyses carried out in Chapter 9, we were able to investigate whether using exact times of visits affected the analyses and there was almost no affect. This suggests that the repeated visits are not so irregular as to cause any issues in analyses of the registry data.

10.1.3 Negative Controls

The next treatment we wished to investigate was ivacaftor. This recent treatment is thought to be the first 'disease-modifying' treatment available to a sub-group of people with CF. Because it was shown to be so efficacious in trials, everybody in the UK who has an eligible genotype is now receiving ivacaftor. Therefore, if we wish to estimate the effect of long-term treatment use in these people, it has been suggested that they should be

compared to another group of people who cannot receive treatment but who otherwise are similar. Negative controls are a tool that can be used to formally test the assumption of the comparability of groups, i.e. to test whether there is any confounding, thus helping to ensure there is no systematic bias in the analysis.[259]

In Chapter 8 we presented six different DAGs in Figures 8.2 and 8.3 showing possible causal pathways in this situation where there is a perfect predictor of treatment, here genotype or time period. In three of these scenarios we would expect there to be no bias in analyses ignoring the perfect predictor of treatment. These are situations where the perfect predictor of treatment has either no direct effect on the outcome or only indirect effects on treatment through observed covariates, so that the indirect confounding effects can be accounted for through adjustment of the observed variables. If the perfect predictor has an effect on the outcome, either directly or indirectly through unobserved covariates, then the simple comparisons of the groups would be biased, and this would be reflected in non-null results of the negative control analysis.

We defined three different estimators: the NTE, the NCE, and the NCCTE. The NTE is obtained by comparing those receiving treatment to those not receiving treatment, ignoring any effect the perfect predictor of treatment may have on the outcome. The NCE is an estimator of the effect of the perfect predictor of treatment on the outcome not mediated through treatment. The NCCTE is then an estimator of the effect of treatment on outcome, correcting the NTE for any estimated effect of the perfect predictor of treatment in the NCE analysis.

The simplest way to calculate an estimate of the NCCTE is to subtract the estimated NCE from the estimated NTE. This assumes that the effect of the perfect predictor is the same in the negative control groups as in the main analysis groups. In settings where the negative control groups are independent of the main analysis groups, the variance of the NCCTE can be estimated as the sum of the variance of the NCE and the variance of the NTE. However, in situations where the same person can be in both the negative control analysis and the main analysis, there will likely be some covariance between the two analyses meaning that a bootstrap is necessary to obtain accurate estimates of the variance of the NCCTE. In such settings, the variance of the NCCTE can actually be less than the variance of the NTE, as the NCE analysis provides information on the comparability of the groups without treatment.

In our setting, depending on the comparator group, we actually had two perfect predictors of treatment: time period and genotype. We could therefore perform the analyses described in the previous two paragraphs twice with different negative controls. Alternatively, by including the whole CF population in one analysis, neither time period or genotype is a perfect predictor of treatment, and they can be included in a multivariable model. However, this analysis does still make the same assumption as the negative control analyses: namely that there is no interaction between the effects of genotype and time

period on the outcome.

10.1.4 Ivacaftor

Chapter 7 presented a systematic review of studies investigating the effects of ivacaftor on lung function and pulmonary exacerbations. Even though ivacaftor has only been available for a few years, there have already been a large number of studies investigating the effects of ivacaftor. The length of these studies has mainly been limited by the time ivacaftor has been available, but there are a number of recent studies with two to four years of follow-up.

The longest RCTs had one-year of follow-up and these all showed very beneficial effects of ivacaftor on lung function and the rate of exacerbations. However, due to the relatively short follow-up time, it was not possible to estimate whether the treatment had resulted in a change of the long-term lung function trajectory.

Due to that fact that almost all people eligible for ivacaftor are now receiving treatment, observational studies have generally had to find control groups among people who could not receive ivacaftor. This meant either doing a pre- and post- study comparing people in the years prior to the availability of ivacaftor to the years since they have been receiving it. This assumes that there have been no other changes in healthcare that could affect the outcome over this time period. We know that the average health of people with CF has been improving over time, but it is not clear if this would have an important effect on an analysis with two to four years of follow-up.[258] Alternatively, studies have compared people receiving ivacaftor with people who are not receiving ivacaftor due to their genotype. Often these types of studies employ matching methods to ensure that patients are matched on a number of baseline covariates. However, even with matching, these comparisons assume that these groups of people would have had similar long-term follow-up were it not for ivacaftor. Some studies have suggested that people with gating mutations have similar long-term follow-up as those with a class II mutation[256, 257]. However, to obtain accurate estimates of the treatment effect, even small non-statistically significant differences between groups should be accounted for.

Our analyses used negative controls to test the assumptions of these two comparison groups. The results showed that over four years the different groups, defined by genotype and time period, were generally comparable with only very small estimates of the NCE. However, although this did not result in large differences in the point estimates of the NTE and the NCCTE, the variance of the NCCTE was generally larger than that of the NTE, reflecting the uncertainty in the comparability of the groups. This meant that the final results were not always statistically significant.

We also compared the use of three different analysis methods: marginal models, fixed-effects models and mixed-effects models. Marginal models appeared to be appropriate as long as an independent working correlation matrix was not used, as generally there was strong correlation between measures from the same individual. Fixed-effects models appeared to perform well for linear outcomes, but there appeared to be some issues with the results when the outcome was non-linear. Mixed-effects models also performed well, but the computation time was much slower and sometimes models would not reach convergence. This was more of an issue when the outcome was non-linear and in analyses where it was necessary to use a bootstrap procedure.

Overall, the analyses suggested a strong beneficial effect of ivacaftor, resulting in a step-change increase in lung function over the first year of treatment. There then appeared to be a clinically important reduction in the annual rate of lung function decline, which was of marginal statistical significance. There also appeared to be a very strong effect of ivacaftor on reducing the rate of IV days that appeared to remain stable throughout follow-up.

The analyses suggested that the use of negative controls was not necessary as the groups did appear comparable. However, this study only had four years of follow-up, and it is plausible that any differences between groups may only become apparent over longer periods of follow-up. For this reason, we recommend continued use of negative controls in the future when estimating longer-term effects of ivacaftor.

10.2 Strengths and Limitations

This thesis has highlighted how registries can be used to estimate treatment effects specifically in settings where it would not be feasible to run RCTs, such as when estimating long-term effects. Although RCTs will remain the gold-standard for estimating treatment efficacy, there are a number of strengths to registries when compared to RCTs. Firstly, the UK CF Registry contains long-term health data on over 99% of people with CF in the UK. Although it will still generally be necessary to exclude a number of people from analyses to avoid bias, the sample size available for analysis will often be much larger than was available in RCTs. The results from the analyses in this thesis should therefore be more generalisable to the whole CF population, rather than only those who fulfil strict inclusion criteria of RCTs.

Another key strength of the analyse of registry data is that RCTs are generally only short-term, and the long-term effects of treatment can be quite different to the short-term effects. This is particularly highlighted by the analysis in this thesis of the long-term effects of DNase, where the short-term benefits of the treatment were not shown to translate into a long-term change in the slope of lung function decline. The analyses in this thesis

estimate effects of up to five years of treatment, but in theory the methods investigated could be used to estimate much longer-term effects.

In this thesis, we have illustrated that suitable statistical methods exist which can be used to estimate treatment effects using Registry data. The methods that are appropriate for estimating long-term treatment effects will not always be the same and this is highlighted by the very different statistical methods presented in Chapters 4 and 8, and the discussion of their features, assumptions, and what they can and cannot be used to estimate. None of the methods used in this thesis are new, but they are still not regularly used in practice. This thesis provides novel assessments of the methods and shows how they can be applied to Registry data. It is hoped that these analyses can be used as illustrations for future research using CF or other national disease registries.

For the methods introduced in Chapter 4, the work from Chapter 5 highlights some potential challenges that could be faced when attempting such analyses in practice and provides results concerning how the performance of the different methods can be affected in different scenarios provides recommendations for when some methods should be preferred over others. Data were simulated under six different scenarios, all of which were felt could reflect the real UK CF Registry data. However, there are undoubtedly many other scenarios which could also have been tested. One such example is the censoring scenario, where data were censored based on observed data. Other types of censoring, such as censoring based on unobserved features and interval censoring, were not considered. Due to the low levels of drop-out in the UK CF Registry, different censoring scenarios would probably not have had much of an impact on the findings. However, some people in the UK CF Registry are interval-censored, meaning that they just happen to miss one or two visits, before re-entering the Registry in later years. In the analyses in Chapter 6, these patients were censored as soon as they missed one visit, meaning that a lot of data was dropped from the analyses. The methods considered for this analysis require sequential visit data, so it was not possible to include their future visits in the analyses, but if these patients were systematically different from other patients, for example, if they were responding well to treatment and thus did not attend the clinic, this could affect the validity of the results from this chapter. This was not an issue in Chapter 9, as the methods considered in this chapter do not require sequential visits. The methods used in Chapter 9 can, however, result in bias of the long-term trajectories due to deaths during follow-up, and methods which aim to correct for this, such as joint modelling, were not used due to their own limitations of referring to a hypothetical group of people where nobody dies.[271, 272]

One key limitation of the analyses in Chapters 6 and 9 is the data quality of the Registry data. While data quality is generally believed to be very good, there are some variables which are thought to be less reliably recorded. For example, in Chapter 2, we highlighted an issue with the reported proportion of patients who are receiving enzyme replacement

therapies, and it is not possible to know if such an issue also affects the DNase data. More recently, as interest in long-term follow-up studies using registry data have become more common, there have been improvements in ensuring that treatment data are captured accurately, and therefore the ivacaftor data are believed to be more reliable. However, even here exact start and end dates are not always reliably entered. It is also known that levels of adherence to long-term treatments can be particularly poor, and thus, all of the analyses in this thesis are estimating the effect of being prescribed a specific treatment, rather than the actual effect of the treatment if it is taken as prescribed.

All of our analysis make use of the lung function measures taken specifically on the day of the annual assessment. One strength of this approach is that it removes any ambiguity around the direction of causal pathways. However, a single lung function measure does not give any indication of the trajectory of lung function, which is often of more interest. With only one measurement per year, it can be difficult to estimate the overall trajectory, but if the collection frequency of lung function measures were increased or with more years of data, it may have been possible to incorporate this into the analyses.

The data entry platform for the UK CF Registry is very comprehensive with a lot of data able to be captured. However, not all fields are reliably collected, and this can be very centre dependent. The key variables used for the analyses in this thesis, such as ppFEV$_1$ and IV days have very low levels of missingness, but some variables that were considered as potential confounders, such as smoking, have much higher levels of missingness. No attempt has been made in this thesis to consider whether this type of missing data could bias the results from the analyses, and this would be an important element of any future work to ensure that analyses of registry data are robust.

Another limitation of the work presented in this thesis is that the outcomes used are not always the same as those that would be used in a RCT, and hence findings are not always directly comparable. This is particularly an issue when considering exacerbations or IV days as an outcome. Most RCTs have a strict definition of what constitutes an exacerbation, and the analyses then look at the effect of treatment on time to next exacerbation or the rate of exacerbations in a given time period. In the UK CF Registry there is no specific marker for whether a patient has suffered an exacerbation and IV days are used as a proxy. For the analyses in this thesis, we chose to use the number of IV days, and to look at the effect of treatment on the rate of IV days. It may have been possible to try and recreate exacerbation data, by considering one course of IVs as an exacerbation, but this transformation would not have been perfect as people often receive IVs even though they are not suffering an exacerbation. Our analyses, therefore, answer a different question to the analyses commonly used in RCTs: whether treatment reduces the use of IVs, compared to whether treatment reduced the incidence of exacerbations. Although this limits the comparability between our analyses and RCTs, reducing the number of IV days is still seen as an important question, both for patients, as well as from a resource standpoint.

10.3 Future Work

The analyses presented in this thesis investigate two very different CF treatments. There are a number of other treatments already available that may be investigated in a way similar to DNase, one such example would be azithromycin, which is an antibiotic that is often prescribed to patients with CF. However, there has never been a RCT looking at azithromycin use in CF and its use in CF is currently off-label in the UK.[273] In terms of disease-modifying treatments, there are number of combination treatments in the pipeline that will combine ivacaftor with other treatments in order for people without gating mutations to receive these treatments. If these are licensed for use in the UK, then it is likely that similar to with ivacaftor, all people who are eligible for them will receive them. In such situations, it would therefore again be necessary to consider what comparator groups are available and if the groups' comparability can be tested through the use of negative controls.

As well as estimating the effect of a treatment in isolation, it would also be of interest to be able to estimate the effect of common combinations of treatment. It is thought that some common treatments could have an antagonistic effect on one another, such as azithromycin and tobramycin, which it would be important to investigate further.[274] Furthermore, treatment burden is a common complaint among people with CF, and simplifying the treatment burden of people with CF has been recognised as a top research priority.[275] This may therefore entail future research into not only the long-term effectiveness of treatments, but also whether there are any effects of stopping treatments.

There are a number of other national CF registries around the world, and other members of the CF-EpiNet group have used data from the US, Canadian and Danish CF registries.[276–278] The way data are collected and what data are collected can vary greatly between registries and it is, therefore, not always easy to compare results obtained from different registries. One key difference is the regularity of data collection, with some registries collecting data monthly or quarterly and some collecting data on an encounter basis. The methods introduced in this thesis should be relevant as long as data are collected at relatively regular intervals, but for data collected on an encounter basis, it is likely that people who are sicker will have more encounters. It would probably be necessary to account for this in analyses in order to ensure it does not result in any bias.

In this thesis, we have shown how the methods can be used with two different outcomes: lung function and number of IV days. There are two of the most important outcome measures in CF studies, but there are also a number of other outcomes that could be important. Firstly, all of our studies excluded anyone under six years of age, because lung function measures are not reliable before this age. In children under the age of six, body mass index (BMI) is often used as an outcome. This is a continuous outcome and therefore the analysis methods should be similar to those presented for lung function.

However, the long-term trajectory of BMI is obviously very different to lung function. In lung function, we approximated the trajectory linearly, which appeared a reasonable assumption to aid interpretability of the results, but it is unlikely that a linear trajectory of BMI would be realistic. Another key outcome of long-term studies would be survival. The methods required for survival analyses are quite different from those used for longitudinal measures. However, there are a number of methods that have been developed to investigate the effects of treatment on survival in the presence of time-dependent confounding, such as marginal structural Cox models[279], and structural nested cumulative failure time models[280].

Our analysis of the effects of ivacaftor suggested that visits are on average equally spaced apart and that ignoring the exact visit dates did not lead to any noticeable bias in the effect estimates. However, it is not clear if this would always be the case and extensions to the methods dealing with time-dependent confounding to account for non-equally spaced visits could be desirable.

Throughout this thesis we have ignored any issues related to missing data. Fortunately, the variables used for our analyses had very low levels of missingness meaning that very few people were excluded from the analyses due to missing data. However, missing data can lead to biased effect estimates unless restrictive assumptions are placed on the missing data mechanisms.[281] Multiple imputation is a common method to account for missing data, and previous work has shown how this method can be combined with IPW of MSM and g-formula.[282–284] SCMM only use standard regression methods, suggesting that there should not be any additional issues if using multiple imputation, but we are not aware of any work investigating the use of multiple imputation with HA-MSM or g-estimation.

Finally, it is known that lung function decline and survival are two very closely related processes in CF.[285] Therefore if deaths during follow-up are not accounted for in some way, the results associated with lung function can be biased, as the population-average rate of decline would naturally become shallower over time due to survivor bias. For the DNase analysis, g-formula and SCMM are not affected by this issue, as both methods only use short-term models. The other methods, IPW of MSM, HA-MSM, and g-estimation, utilised censoring weights to account for loss to follow-up. However, this issue was not addressed in the analysis of the long-term effects of ivacaftor. Over the four year period of this analysis, there were very few deaths, meaning that it is unlikely to have led to substantial bias. However, in future work with longer follow-up, it is more likely that this could become an issue. Joint modelling is an approach that can be used to simultaneously model both longitudinal outcomes and survival, and there has already been some work showing how joint models can be used in CF.[285–287] However, these methods have not yet been used in any studies to estimate treatment effects in CF.

10.4 Conclusions

In this thesis, we have shown that registry data can be used to estimate the long-term effects of CF treatments. However, there are generally a number of complexities that must be addressed when analysing longitudinal observational data. Two such issues are time-dependent confounding or lack of a comparator group, for which suitable statistical methods already exist to be able to obtain consistent estimators of the treatment effect. It is usually necessary to make a number of untestable assumptions when performing these analyses, for example no unmeasured confounding, but in the case of some assumptions it is possible to formally test whether they seem plausible, allowing for an assessment of how trustworthy any results from the analyses might be.

Bibliography

[1] LE Bothwell et al. "Assessing the Gold Standard - Lessons from the History of RCTs". In: *New Engl J Med* 374.22 (2016), pp. 2175–81. DOI: 10.1056/NEJMms1604593.

[2] *Cystic Fibrosis Trust*. URL: https://www.cysticfibrosis.org.uk.

[3] *UK Cystic Fibrosis Registry Annual Data Report 2016*. Tech. rep. 2017.

[4] HL Barr et al. "Association between socioeconomic status, sex, and age at death from cystic fibrosis in England and Wales (1959 to 2008): Cross sectional study". In: *BMJ* 343 (2011). DOI: 10.1136/bmj.d4662.

[5] *UK Cystic Fibrosis Registry Annual data report 2017*. Tech. rep. 2018.

[6] GS Sawicki, DE Sellers, and WM Robinson. "High treatment burden in adults with cystic fibrosis: Challenges to disease self-management". In: *J Cyst Fibros* 8.2 (2009), pp. 91–96. DOI: 10.1016/j.jcf.2008.09.007.

[7] I Fajac and K De Boeck. "New horizons for cystic fibrosis treatment". In: *Pharmacol Therapeut* 170 (2017), pp. 205–11. DOI: 10.1016/j.pharmthera.2016.11.009.

[8] K Harman, R Dobra, and JC Davies. "Disease-modifying drug therapy in cystic fibrosis". In: *Paediatr Respir Rev* 26 (2018), pp. 24–26. DOI: 10.1016/j.prrv.2017.03.008.

[9] MA Hernán and JM Robins. *Causal Inference*. Chapman & Hall/CRC, 2018.

[10] D Taylor-Robinson et al. "Data Resource Profile: The UK Cystic Fibrosis Registry". In: *Int J Epidemiol* 47.1 (2017), 9–10e. DOI: 10.1093/ije/dyx196.

[11] SJ Newsome, RH Keogh, and RM Daniel. "Estimating long-term treatment effects in observational data: a comparison of the performance of different methods under real-world uncertainty". In: *Stat Med* 37.15 (2018), pp. 2367–90. DOI: 10.1002/sim.7664.

[12] SJ Newsome et al. "Investigating the effects of long-term dornase alfa use on lung function using registry data". In: *J Cyst Fibros* (2018). DOI: 10.1016/j.jcf.2018.08.004.

[13] BW Ramsey and TF Boat. "Outcome measures for clinical trials in cystic fibrosis. Summary of a Cystic Fibrosis Foundation Consensus Conference". In: *J Pediatr* 124.2 (1994), pp. 177–92. DOI: 10.1016/S0022-3476(94)70301-9.

[14] DR VanDevanter and MW Konstan. "Outcome measures for clinical trials assessing treatment of cystic fibrosis lung disease". In: *Clin Investig* 2.2 (2012), pp. 163–75. DOI: 10.4155/cli.11.174.Outcome.

[15] *Spirometry and reversibility testing*. Tech. rep. 2017.

[16] PH Quanjer et al. "Multi-ethnic reference values for spirometry for the 3-95-yr age range: The global lung function 2012 equations". In: *Eur Respir J* 40.6 (2012), pp. 1324–43. DOI: 10.1183/09031936.00080312.

[17] S Stanojevic et al. "The impact of switching to the new global lung function initiative equations on spirometry results in the UK CF Registry". In: *J of Cyst Fibros* 13.3 (2014), pp. 319–27. DOI: 10.1016/j.jcf.2013.11.006.

[18] A Bush, D Bilton, and M Hodson. *Hodson and Geddes' Cystic Fibrosis*. Taylor & Francis, 2015. ISBN: 1444180002.

[19] JS Wagener et al. "Pulmonary function outcomes for assessing cystic fibrosis care". In: *J Cyst Fibros* 14.3 (2015), pp. 376–83. DOI: 10.1016/j.jcf.2014.11.008.

[20] R Szczesniak et al. "Use of FEV1 in cystic fibrosis epidemiologic studies and clinical trials: A statistical perspective for the clinical researcher". In: *J Cyst Fibros* 16.3 (2017), pp. 318–26. DOI: 10.1016/j.jcf.2017.01.002.

[21] D Taylor-Robinson et al. "Understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: A longitudinal study". In: *Thorax* 67.10 (2012), pp. 860–66. DOI: 10.1136/thoraxjnl-2011-200953.

[22] JC Davies et al. "Lung clearance index in CF: A sensitive marker of lung disease severity". In: *Thorax* 63.2 (2008), pp. 96–97. DOI: 10.1136/thx.2007.088575.

[23] AR Horsley et al. "Lung clearance index is a sensitive, repeatable and practical measure of airways disease in adults with cystic fibrosis". In: *Thorax* 63.2 (2008), pp. 135–40. DOI: 10.1136/thx.2007.082628.

[24] DB Sanders et al. "Return of FEV1 after pulmonary exacerbation in children with cystic fibrosis". In: *Pediatr Pulm* 45.2 (2010), pp. 127–34. DOI: 10.1002/ppul.21117.

[25] DB Sanders et al. "Failure to recover to baseline pulmonary function after cystic fibrosis pulmonary exacerbation". In: *Am J Resp Crit Care* 182.5 (2010), pp. 627–32. DOI: 10.1164/rccm.200909-1421OC.

[26] BT Kopp et al. "The impact of secondhand smoke exposure on children with cystic fibrosis: A review". In: *Int J Env Res Pub He* 13.10 (2016), pp. 1–10. DOI: 10.3390/ijerph13101003.

[27] AD Jackson and CH Goss. "Epidemiology of CF: How registries can be used to advance our understanding of the CF population". In: *J Cyst Fibros* 17.3 (2018), pp. 297–305. DOI: 10.1016/j.jcf.2017.11.013.

[28] ZH Hoo et al. "The importance of data issues when comparing cystic fibrosis registry outcomes between countries: Are annual review FEV1 in the UK only collected when subjects are well?" In: *J Eval Clin Pract* (2018), pp. 1–7. DOI: 10.1111/jep.12967.

[29] T Pressler. "Review of recombinant human deoxyribonuclease (rhDNase) in the management of patients with cystic fibrosis". In: *Biologics* 2.4 (2008), pp. 611–17.

[30] *EMC*. URL: https://www.medicines.org.uk/emc/medicine/1723.

[31] D Moher et al. "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement". In: *PLoS Med* 6.7 (2009), pp. 873–80. DOI: 10.1371/journal.pmed.1000097.

[32] AP Jones and C Wallis. "Dornase alfa for cystic fibrosis". In: *Cochrane Database Syst Rev* 3 (2010). DOI: 10.1002/14651858.CD001127.pub2.

[33] JPT Higgins et al. "The Cochrane Collaboration's tool for assessing risk of bias in randomised trials". In: *BMJ* 343 (2011). DOI: 10.1136/bmj.d5928.

[34] JJ Deeks et al. "Evaluating non-randomised intervention studies". In: *Health Technol Assess* 7.27 (2003). DOI: 10.3310/hta7270.

[35] G Wells et al. *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analysis*. Tech. rep. 2011.

[36] SH Downs and N Black. "The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions". In: *J Epidemiol Commun H* 52.6 (1998), pp. 377–84. DOI: 10.1136/jech.52.6.377.

[37] HGM Heijerman, RN van Rossem, and W Bakker. "Effect of rhDNase on lung function and quality of life in adult cystic fibrosis patients". In: *Neth J Med* 46.6 (1995), pp. 293–97. DOI: 10.1016/0300-2977(95)00022-F.

[38] PL Shah et al. "Two years experience with recombinant human DNase I in the treatment of pulmonary disease in cystic fibrosis". In: *Resp Med* 89.7 (1995), pp. 499–502. DOI: 10.1016/0954-6111(95)90126-4.

[39] ME Furuya et al. "Efficacy of human recombinant DNase in pediatric patients with cystic fibrosis". In: *Arch Med Res* 32.1 (2001), pp. 30–34. DOI: 10.1016/S0188-4409(00)00258-7.

[40] M Barker et al. "Effect of DNase on exercise capacity in cystic fibrosis". In: *Pediatr Pulm* 38.1 (2004), pp. 70–74. DOI: 10.1002/ppul.20048.

[41] J Riethmueller et al. "Recombinant human deoxyribonuclease shortens ventilation time in young, mechanically ventilated children". In: *Pediatr Pulm* 41.1 (2006), pp. 61–66. DOI: 10.1002/ppul.20298.

[42] GL McPhail et al. "Improvements in lung function outcomes in children with cystic fibrosis are associated with better nutrition, fewer chronic Pseudomonas aeruginosa infections, and dornase alfa use". In: *J Pediatr* 153.6 (2008), pp. 752–757. DOI: 10.1016/j.jpeds.2008.07.011.

[43] GL McPhail et al. "Initiation of dornase alfa before age six years is associated with improvements in the rate of decline in lung function in children who are "rapid early decliners"". In: *Pediatr Pulm* 45.S33 (2010), p. 363. DOI: 10.1002/ppul.21339.

[44] T Rozov et al. "Dornase alfa improves the health-related quality of life among Brazilian patients with cystic fibrosis — A one-year prospective study". In: *Pediatr Pulm* 45.9 (2010), pp. 874–82. DOI: 10.1002/ppul.21267.

[45] EC Dasenbrook et al. "Combination inhaled 7% hypertonic saline and rhDNase: Retrospective cohort study of the US CFF national patient registry". In: *Pediatr Pulm* 48.S36 (2013), p. 343. DOI: 10.1002/ppul.22898.

[46] T Rozov et al. "A first-year dornase alfa treatment impact on clinical parameters of patients with cystic fibrosis: the Brazilian cystic fibrosis multicenter study". In: *Rev Paul Pediatr* 31.4 (2013), pp. 420–30. DOI: 10.1590/S0103-05822013000400002.

[47] D VanDevanter et al. "Chronic High Dose Dornase Alfa Use in CF 1994-2005". In: *Pediatr Pulm* 50.S41 (2015), p. 350. DOI: 10.1002/ppul.23297.

[48] A Shenoy et al. "The effects of early initiation with dornase alfa on childhood lung function in cystic fibrosis". In: *Pediatr Pulm* 51.S45 (2016), pp. 364–5. DOI: 10.1002/ppul.23576.

[49] HJ Fuchs et al. "Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis". In: *New Engl J Med* 331.10 (1994), pp. 637–42. DOI: 10.1056/NEJM199409083311003.

[50] R Suri et al. "Comparison of hypertonic saline and alternate-day or daily recombinant human deoxyribonuclease in children with cystic fibrosis: A randomised trial". In: *Lancet* 358.9290 (2001), pp. 1316–21. DOI: 10.1016/S0140-6736(01)06412-1.

[51] RC Hubbard et al. "A preliminary study of aerosolized recombinant human deoxyribonuclease I in the treatment of cystic fibrosis". In: *New Engl J Med* 326.12 (1992), pp. 812–5.

[52] C Ranasinha et al. "Efficacy and safety of short-term administration of aerosolised recombinant human DNase I in adults with stable stage cystic fibrosis". In: *Lancet* 342.8865 (1993), pp. 199–202. DOI: 10.1016/0140-6736(93)92297-7.

[53] BW Ramsey et al. "Efficacy and safety of short-term administration of aerosolized recombinant human deoxyribonuclease in patients with cystic fibrosis." In: *Am Rev Respir Dis* 148.1 (1993), pp. 145–51. DOI: 10.1164/ajrccm/148.1.145.

[54] PI Shah et al. "Recombinant human DNase I in cystic fibrosis patients with severe pulmonary disease: a short-term, double-blind study followed by six months open-label treatment". In: *Eur Respir J* 8.6 (1995), pp. 954–8. DOI: 10.1183/09031936.95.08060954.

[55] BL Laube et al. "Effect of rhDNase on airflow obstruction and mucociliary clearance in cystic fibrosis". In: *Am J Resp Crit Care* 153.2 (1996), pp. 752–60. DOI: 10.1164/ajrccm.153.2.8564129.

[56] RW Wilmott et al. "Aerosolized recombinant human DNase in hospitalized cystic fibrosis patients with acute pulmonary exacerbations". In: *Am J Resp Crit Care* 153.6 (1996), pp. 1914–7. DOI: 10.1164/ajrccm.153.6.8665055.

[57] K McCoy, S Hamilton, and C Johnson. "Effects of 12-week administration of dornase alfa in patients with advanced cystic fibrosis lung disease". In: *Chest* 110.4 (1996), pp. 889–95. DOI: 10.1378/chest.110.4.889.

[58] MB Weck, GZ Retsch-Bogart, and CS Scott. "Efficacy of DNase in Individual Children Using the N-of-1 Study Design". In: *Pediatr Pulm* 28.19 (1999), pp. 56–82.

[59] M Robinson et al. "Effect of a short course of rhDNase on cough and mucociliary clearance in patients with cystic fibrosis". In: *Pediatr Pulm* 30.1 (2000), pp. 16–24.

[60] PL Shah et al. "A case-controlled study with dornase alfa to evaluate impact on disease progression over a 4-year period". In: *Respiration* 68.2 (2001), pp. 160–4. DOI: 10.1159/000050486.

[61] JM Quan et al. "A two-year randomized, placebo-controlled trial of dornase alfa in young patients with cystic fibrosis with mild lung function abnormalities". In: *J Pediatr* 139.6 (2001), pp. 813–20. DOI: 10.1067/mpd.2001.118570.

[62] M Ballmann and Horst von der Hardt. "Hypertonic saline and recombinant human DNase: A randomised cross-over pilot study in patients with cystic fibrosis". In: *J Cyst Fibros* 1.1 (2002), pp. 35–7. DOI: 10.1016/S1569-1993(01)00009-1.

[63] PJ Robinson. "Dornase alfa in early cystic fibrosis lung disease". In: *Pediatr Pulm* 34.3 (2002), pp. 237–41. DOI: 10.1002/ppul.10136.

[64] ME Hodson et al. "Dornase alfa in the treatment of cystic fibrosis in Europe: a report from the Epidemiologic Registry of Cystic Fibrosis". In: *Pediatr Pulm* 36.5 (2003), pp. 427–32. DOI: 10.1002/ppul.10348.

[65] M ten Berge et al. "DNase in stable cystic fibrosis infants: A pilot study". In: *J Cyst Fibros* 2.4 (2003), pp. 183–8. DOI: 10.1016/S1569-1993.

[66] K Paul et al. "Effect of treatment with dornase alpha on airway inflammation in patients with cystic fibrosis". In: *Am J Resp Crit Care* 169.6 (2004), pp. 719–25. DOI: 10.1164/rccm.200307-959OC.

[67] FV Adde et al. "Hypertonic Saline X Recombinant Human DNase: A Randomised Cross-Over Study in 18 Cystic Fibrosis Patients". In: *J Cyst Fibros* 3.S1 (2004), p. 66. DOI: 10.1016/j.jcf.2004.03.001.

[68] B Frederiksen et al. "Effect of aerosolized rhDNase (Pulmozyme) on pulmonary colonization in patients with cystic fibrosis". In: *Acta Paediatr* 95.9 (2006), pp. 1070–4. DOI: 10.1080/08035250600752466.

[69] MW Konstan. "Dornase alfa and progression of lung disease in cystic fibrosis". In: *Pediatr Pulm* 43.S9 (2008), pp. 24–28. DOI: 10.1002/ppul.20858.

[70] R Kraemer et al. "Long-term effect of dornase alfa on lung function in patients with CF evaluated over an age period of 5 to 18 years". In: *Pediatr Pulm* 44.S32 (2009). DOI: 10.1002/ppul.21133.

[71] R Castile et al. "Effects of nebulized recombinant human deoxyribonuclease (dornase alfa) in infants with CF evaluated using infant pulmonary function testing and high resolution computerized tomographic imaging of the chest". In: *Pediatr Pulm* 44.S32 (2009). DOI: 10.1002/ppul.21133.

[72] C Minasian et al. "Comparison of inhaled mannitol, daily rhDNase and a combination of both in children with cystic fibrosis: A randomised trial". In: *Thorax* 65.1 (2010), pp. 51–56. DOI: 10.1136/thx.2009.116970.

[73] HJC Bonestroo, MG Slieker, and HGM Arets. "No Positive Effect of rhDNase on the pulmonary colonization in children with cystic fibrosis". In: *Monaldi Arch Chest Dis* 73.1 (2010), pp. 12–17. DOI: 10.4081/monaldi.2010.308.

[74] R Amin et al. "The effect of dornase alfa on ventilation inhomogeneity in patients with cystic fibrosis". In: *Eur Respir J* 37.4 (2011), pp. 806–12. DOI: 10.1183/09031936.00072510.

[75] MW Konstan et al. "Clinical use of dornase alpha is associated with a slower rate of FEV1 decline in cystic fibrosis." In: *Pediatr Pulm* 46.6 (2011), pp. 545–53. DOI: 10.1002/ppul.21388.

[76] DJ Pasta, SJ Millar, and L Rasouliyan. "A mixed-effects piecewise linear model of the rate of lung function decline before and after the clinical use of dornase alfa in an observational study of cystic fibrosis". In: *Value Health* 14.7 (2011), A330. DOI: 10.1016/j.jval.2011.08.533.

[77] SB Singh, T Burns, and TD Starner. "Clinical effectiveness of dornase alfa in children six years of age and younger with cystic fibrosis". In: *Pediatr Pulm* 49.S38 (2014), p. 357. DOI: 10.1002/ppul.23108.

[78] F Singer et al. "Effects of dornase alfa and hypertonic saline on gas mixing in lung periphery in children with cystic fibrosis". In: *Eur Respir J* 44.S58 (2014), p. 1976.

[79] LJ Stuckey et al. "The effect of dornase alfa (rhDNase) on recurrent gram negative infections in adult cystic fibrosis lung transplant recipients". In: *J Heart Lung Transpl* 34.4 (2015), S240. DOI: 10.1016/j.healun.2015.01.666.

[80] RM Daniel et al. "Methods for dealing with time-dependent confounding". In: *Stat Med* 32.9 (2013), pp. 1584–618. DOI: 10.1002/sim.5686.

[81] DB Rubin. "Bayesian inference for causal effects: The role of randomization". In: *Ann Statist* 6.1 (1978), pp. 34–58.

[82] L Orellana, A Rotnitzky, and JM Robins. "Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part II: Proofs of results". In: *Int J Biostat* 6.2 (2010). DOI: 10.2202/1557-4679.1242.

[83] TJ VanderWeele. "Concerning the consistency assumption in causal inference". In: *Epidemiology* 20.6 (2009), pp. 880–3. DOI: 10.1097/EDE.0b013e3181bd5638.

[84] JM Robins, MA Hernán, and B Brumback. "Marginal structural models and causal inference in epidemiology". In: *Epidemiology* 11.5 (2000), pp. 550–60. DOI: 10.1097/00001648-200009000-00011.

[85] Y Wang et al. "Diagnosing bias in the inverse probability of treatment weighted estimator resulting from violation of experimental treatment assignment". In: *UC Berkeley Division of Biostatistics Working Paper Series* 211 (2006). URL: http://www.bepress.com/ucbbiostat/paper211/.

[86] SR Cole and MA Hernán. "Constructing inverse probability weights for marginal structural models". In: *Am J Epidemiol* 168.6 (2008), pp. 656–64. DOI: 10.1093/aje/kwn164.

[87] ML Petersen et al. "History-adjusted marginal structural models for estimating time-varying effect modification". In: *Am J Epidemiol* 166.9 (2007), pp. 985–93. DOI: 10.1093/aje/kwm232.

[88] J Robins. "A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect". In: *Math Mod* 7.9–12 (1986), pp. 1393–1512. DOI: 10.1016/0270-0255(86)90088-6.

[89] JM Robins and L Wasserman. "Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs". In: *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*. 1997. DOI: 10.1088/0963-0252/22/1/015008.

[90] JM Robins et al. "G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients". In: *Epidemiology* 3.4 (1992), pp. 319–36. DOI: 10.1097/00001648-199207000-00007.

[91] S Vansteelandt and M Joffe. "Structural nested models and g-estimation: The partially realized promise". In: *Stat Sci* 29.4 (2014), pp. 707–31. DOI: 10.1214/14-STS493.

[92] S Vansteelandt and A Sjolander. "Revisiting g-estimation of the effect of a time-varying exposure subject to time-varying confounding". In: *Epidemiologic Methods* 5.1 (2016), pp. 37–56. DOI: 10.1515/em-2015-0005.

[93] RH Keogh et al. "Analysis of longitudinal studies with repeated outcome measures: adjusting for time-dependent confounding using conventional methods". In: *Am J Epidemiol* 187.5 (2018), pp. 1085–92. DOI: 10.1093/aje/kwx311.

[94] O Dukes and S Vansteelandt. "A note on g-estimation of causal risk ratios". In: *Am J Epidemiol* 187.5 (2018), pp. 1079–84. DOI: 10.1093/aje/kwx347.

[95] EJ Williamson, A Forbes, and IR White. "Variance reduction in randomised trials by inverse probability weighting using the propensity score". In: *Stat Med* 33.5 (2014), pp. 721–37. DOI: 10.1002/sim.5991.

[96] JM Snowden, S Rose, and KM Mortimer. "Implementation of g-computation on a simulated data set: Demonstration of a causal inference technique". In: *American Journal of Epidemiology* 173.7 (2011), pp. 731–738. DOI: 10.1093/aje/kwq472.

[97] AI Naimi et al. "Estimating the effect of cumulative occupational asbestos exposure on time to lung cancer mortality: Using structural nested failure-time models to account for healthy-worker survivor bias". In: *Epidemiology* 25.2 (2014), pp. 246–54. DOI: 10.1097/EDE.0000000000000045.

[98] EB Laber et al. "Dynamic treatment regimes: Technical challenges and applications". In: *Electron J Stat* 8.1 (2014), pp. 1225–72.

[99] E Williamson et al. "Propensity scores: from naive enthusiasm to intuitive understanding". In: *Stat Methods Med Res* 21.3 (2012), pp. 273–93. DOI: 10.1177/0962280210394483.

[100] KA Kupzyk and SJ Beal. "Advanced Issues in Propensity Scores: Longitudinal and Missing Data". In: *J Early Adolescence* 37.1 (2016), pp. 59–84. DOI: 10.1177/0272431616636229.

[101] MA Hernán and JM Robins. "Instruments for Causal Inference: An Epidemiologist's Dream?" In: *Epidemiology* 17.4 (2006), pp. 360–72. DOI: 10.1097/01.ede.0000222409.00878.37.

[102] A Burton et al. "The design of simulation studies in medical statistics". In: *Stat Med* 25.24 (2006), pp. 4279–92. DOI: 10.1002/sim.2673.

[103] JG Young and EJ Tchetgen Tchetgen. "Simulation from a known Cox MSM using standard parametric models for the g-formula". In: *Stat Med* 33.6 (2014), pp. 1001–14. DOI: 10.1002/sim.5994.

[104] MW Konstan and F Ratjen. "Effect of dornase alfa on inflammation and lung function : Potential role in the early treatment of cystic fibrosis". In: *J Cyst Fibros* 11.2 (2012), pp. 78–83. DOI: 10.1016/j.jcf.2011.10.003.

[105] B Efron and R Tibshirani. "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy". In: *Statistical Science* 1.1 (1986), pp. 54–77.

[106] MA Hernán. "The C-word: Scientific euphemisms do not improve causal inference from observational data". In: *Am J Public Health* 108.5 (2018), pp. 616–19. DOI: 10.2105/AJPH.2018.304337.

[107] A Shakkottai et al. "A five-year retrospective analysis of adherence in cystic fibrosis". In: *Pediatr Pulm* 50.12 (2015), pp. 1224–29. DOI: 10.1002/ppul.23307.

[108] AL Quittner et al. "Pulmonary medication adherence and health-care use in cystic fibrosis". In: *Chest* 146.1 (2014), pp. 142–51. DOI: 10.1378/chest.13-1926.

[109] LJ Kettler et al. "Determinants of adherence in adults with cystic fibrosis". In: *Thorax* 57.5 (2002), pp. 459–64. DOI: 10.1136/thorax.57.5.459.

[110] J Pittman and M Rosenfeld. "Elementary, My Dear Watson! The accumulating evidence for the lung clearance index in monitoring early cystic fibrosis lung disease". In: *Am J Resp Crit Care* 195.9 (2017), pp. 1131–32. DOI: 10.1164/rccm.201701-0025ED.

[111] F Van Goor et al. "Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770". In: *Proc Natl Acad Sci* 106.44 (2009), pp. 18825–30. DOI: 10.1073/pnas.0904709106.

[112] KD MacDonald, KR McKenzie, and PL Zeitlin. "Cystic fibrosis transmembrance regulator protein mutations: 'class' opportunity for novel drug innovation". In: *Paediatr Drugs* 9.1 (2007), pp. 1–10. DOI: 10.2165/00148581-200709010-00001.

[113] *Clinical Commissioning Policy: Ivacaftor for Cystic Fibrosis (named mutations)*. Tech. rep. 2015.

[114] *Clinical Commissioning Policy: Ivacaftor for children aged 2-5 years with cystic fibrosis (named mutations)*. Tech. rep. 2016.

[115] M Boyle et al. "Effect of VX-770, a CFTR potentiator, on spirometry and QOL assessment in subjects with CF and the G551D-CFTR mutation". In: *Pediatr Pulm* 44.S32 (2009). DOI: 10.1002/ppul.21133.

[116] SM Rowe et al. "Improvement in ion transport biomarkers and spirometry with the investigational CFTR potentiator VX-770 in subjects with cystic fibrosis and the G551D-CFTR mutation". In: *Paediatr Respir Rev* 11.S1 (2010), S99. DOI: 10.1016/S1526-0542(10)70137-8.

[117] D Borowitz et al. "Measures of nutritional status in two Phase 3 trials of ivacaftor in subjects with cystic fibrosis who have the G551D-CFTR mutation". In: *J Cyst Fibros* 11.S1 (2012), S13. DOI: 10.1016/S1569-1993(12)60041-1.

[118] BJ Plant et al. "Lung function, weight, and sweat chloride responses in patients with cystic fibrosis and the G551D-CFTR mutation treated with ivacaftor: A secondary analysis". In: *J Cyst Fibros* 12.S1 (2013), S62. DOI: 10.1016/S1569-1993(13)60195-2.

[119] S Elborn et al. "Lung function, weight, and sweat chloride responses in patients with cystic fibrosis and the G551D-CFTR mutation treated with Ivacaftor: A secondary analysis". In: *Eur Respir J* 42.S57 (2013), p. 5059.

[120] K De Boeck et al. "Ivacaftor, a CFTR potentiator, in cystic fibrosis patients who have a non-G551D-CFTR gating mutation: phase 3, part 1 results". In: *Pediatr Pulm* 48.S36 (2013), p. 292. DOI: 10.1002/ppul.22898.

[121] C Wainwright et al. "The effect of Ivacaftor in individuals with cystic fibrosis and severe lung disease: analysis of data from the Australian named patient programme". In: *Respirology* 19.S2 (2014), p. 112. DOI: 10.1111/resp.12263_10.

[122] K De Boeck et al. "The effect of ivacaftor, a CFTR potentiator, in patients with cystic fibrosis and a non-G551D-CFTR gating mutation, the KONNECTION study". In: *J Cyst Fibros* 13.S2 (2014), S1.

[123] GS Sawicki et al. "The effect of ivacaftor on weight over three years in patients with CF and a G551D-CFTR mutation". In: *Pediatr Pulm* 49.S38 (2014), p. 289. DOI: 10.1002/ppul.23108.

[124] SL Heltshe et al. "Ivacaftor is associated with Pseudomonas aeruginosa reduction in cystic fibrosis patients with G551D-CFTR". In: *Pediatr Pulm* 49.S38 (2014), p. 284. DOI: 10.1002/ppul.23108.

[125] J Pace et al. "Ivacaftor improves linear growth in G551D pre-pubertal children". In: *Pediatr Pulm* 49.S38 (2014), pp. 305–6. DOI: 10.1002/ppul.23108.

[126] C Wainwright et al. "The effect of Ivacaftor in individuals with CF and severe lung disease". In: *Pediatr Pulm* 49.S38 (2014), pp. 376–7. DOI: 10.1002/ppul.23108.

[127] C Varghese et al. "Outcomes after 1 year of Ivacaftor treatment in CF Patients with at least one G551D-CFTR mutation: a single centre experience". In: *Irish J Med Sci* 183.S11 (2014), p. 540.

[128] MW Konstan et al. "Efficacy response in CF patients treated with ivacaftor: Post-hoc analysis". In: *Pediatr Pulm* 50.5 (2015), pp. 447–55. DOI: 10.1002/ppul.23173.

[129] K McKay et al. "The effect of ivacaftor on exocrine pancreatic function in patients with cystic fibrosis and the G551D CFTR mutation who are naïve for ivacaftor". In: *J Cyst Fibros* 14.S1 (2015), S117. DOI: 10.1016/S1569-1993(15)30406-9.

[130] CE Pope et al. "Short-term and long-term effects of Ivacaftor treatment on sputum microbiota in people with the G551D CFTR mutation". In: *J Cyst Fibros* 14.S1 (2015), S3. DOI: 10.1016/S1569-1993(15)30010-2.

[131] A Prosser et al. "Transformational care at the All Wales Adult CF Centre (AWACFC) - is Ivacaftor making us fat? The impact of ivacaftor (Kalydeco®) on body composition". In: *J Cyst Fibros* 14.S1 (2015), S129. DOI: 10.1016/S1569-1993(15)30312-X.

[132] M Rosenfeld et al. "An open-label study of the safety, pharmacokinetics, and pharmacodynamics of ivacaftor in patients aged 2 to 5 years with cystic fibrosis and a CFTR gating mutation: the KIWI study". In: *J Cyst Fibros* 14.S1 (2015), S2. DOI: 10.1016/S1569-1993(15)30005-9.

[133] AC Tierney et al. "Ivacaftor and its effects on body composition in adults with G551D related cystic fibrosis". In: *J Cyst Fibros* 14.S1 (2015), S50. DOI: 10.1016/S1569-1993(15)30162-4.

[134] PJ Barry et al. "Ivacaftor decreases mortality in G551D patients with severe lung disease". In: *Pediatr Pulm* 50.S41 (2015), pp. 275–6. DOI: 10.1002/ppul.23297.

[135] M Rosenfeld et al. "Extended evaluation of Ivacaftor treatment in pediatric patients with cystic fibrosis and a CFTR gating mutation". In: *Pediatr Pulm* 50.S41 (2015), pp. 284–5. DOI: 10.1002/ppul.23297.

[136] MS Stalvey et al. "Ivacaftor improves linear growth in G551D cystic fibrosis children: Results of a multicenter, placebo-controlled study". In: *Pediatr Pulm* 50.S41 (2015), p. 402. DOI: 10.1002/ppul.23297.

[137] JC Davies et al. "Ivacaftor treatment in preschool children with cystic fibrosis and a CFTR gating mutation: extended evaluation". In: *Thorax* 70.S3 (2015), A14–15. DOI: 10.1136/thoraxjnl-2015-207770.24.

[138] D Borowitz et al. "Nutritional status improved in cystic fibrosis patients with the G551D mutation after treatment with ivacaftor". In: *Digest Dis Sci* 61.1 (2016), pp. 198–207. DOI: 10.1007/s10620-015-3834-2.

[139] MS Stalvey et al. "Ivacaftor improves linear growth in children with cystic fibrosis (CF) and a G551D-CTFR mutation: Data from the ENVISION study". In: *J Cyst Fibros* 15.S1 (2016), S101. DOI: 10.1016/S1569-1993(16)30436-2.

[140] R Adam et al. "Quantitative CT scan assessment of lung structure and function after one year of Ivacafotr therapy". In: *Pediatr Pulm* 51.S45 (2016), p. 365. DOI: 10.1002/ppul.23576.

[141] E Looi, A Jones, and PJ Barry. "Ivacaftor therapy increases BMI but does not affect serum cholesterol in patients with gating mutations". In: *Pediatr Pulm* 51.S45 (2016), p. 431. DOI: 10.1002/ppul.23576.

[142] M Rosenfeld et al. "Long-term safety and efficacy of Ivacaftor in pediatric patients aged 2-5 years with CF and a CFTR gating mutation". In: *Pediatr Pulm* 51.S45 (2016), pp. 258–9. DOI: 10.1002/ppul.23576.

[143] MS Stalvey et al. "Growth in Prepubertal Children With Cystic Fibrosis Treated With Ivacaftor". In: *Pediatrics* 139.2 (2017). DOI: 10.1542/peds.2016-2522.

[144] M McCullagh et al. "Long term microbiological outcomes of Ivacaftor use. A single-centre resprospective study". In: *Pediatr Pulm* 52.S47 (2017), p. 346. DOI: 10.1002/ppul.23840.

[145] FJ Accurso et al. "Intermin results of phase 2A study of VX-770 to evaluate safety, pharmacokinetics, and biomarkers of CFTR activity in cystic fibrosis subjects with G551D". In: *Pediatr Pulm* 43.S31 (2008), p. 295.

[146] FJ Accurso et al. "Final results of a 14- and 28-day study of VX-770 in subjects with CF". In: *J Cyst Fibros* 8.S2 (2009), S25. DOI: 10.1016/S1569-1993(09)60101-6.

[147] EF McKone et al. "Long-term safety and efficacy of investigational CFTR potentiator, VX-770, in subjects with CF". In: *Pediatr Pulm* 46.S34 (2011), p. 284. DOI: 10.1002/ppul.21583.

[148] EF McKone et al. "Long-term safety and efficacy of Ivacaftor in subjects with CF who have the G551D-CFTR mutation". In: *J Cyst Fibros* 11.S1 (2012), S13. DOI: 10.1016/S1569-1993(12)60042-3.

[149] EF McKone et al. "Long-term safety and efficacy of Ivacaftor in persons with cystic fibrosis who have the G551D-CFTR mutation". In: *Pediatr Pulm* 47.S35 (2012), pp. 296–7. DOI: 10.1002/ppul.22682.

[150] EF McKone et al. "Long-term safety and efficacy of Ivacaftor in subjects with CF who have the G551D-CFTR mutation". In: *Eur Respir J* 40.S56 (2012), p. 1842.

[151] ME Wood et al. "Observational study of the clinical effects of Ivacaftor in patients with severe cystic fibrosis (CF) lung disease". In: *Respirology* 18.S2 (2013), p. 64. DOI: 10.1111/resp.12046.

[152] P Mondal et al. "Postmarketing experience with Ivacaftor therapy in cystic fibrosis patients having G551D mutation". In: *Am J Respir Crit Care Me* 187 (2013), A1179.

[153] PJ Barry et al. "UK and Ireland review of ivacaftor in severe CF: Impact on lung function and weight". In: *J Cyst Fibros* 12.S1 (2013), S62. DOI: 10.1016/S1569-1993(13)60196-4.

[154] D Hubert et al. "Ivacaftor in French patients with cystic fibrosis and a G551D mutation in the real world setting". In: *J Cyst Fibros* 12.S1 (2013), S63. DOI: 10.1016/S1569-1993(13)60196-4.

[155] D Hubert et al. "Ivacaftor in French patients with cystic fibrosis and a G551D mutation". In: *Pediatr Pulm* 48.S36 (2013), p. 286. DOI: 10.1002/ppul.22898.

[156] ME McGarry and DW Nielson. "Normalization of sweat chloride concentration and clinical improvement with ivacaftor in a patient with cystic fibrosis with mutation S549N". In: *Chest* 144.4 (2013), pp. 1376–8. DOI: 10.1378/chest.13-0239.

[157] EF McKone et al. "Long-term safety and efficacy of Ivacaftor in patients with cystic fibrosis who have the G551D-CFTR mutation: response through 144 weeks of treatment (96 weeks of persist)". In: *Pediatr Pulm* 48.S36 (2013), p. 287. DOI: 10.1002/ppul.22898.

[158] SM Rowe et al. "Results of the G551D observational study: The effect of ivacaftor in G551D patients following FDA approval". In: *Pediatr Pulm* 48.S36 (2013), p. 278. DOI: 10.1002/ppul.22898.

[159] E Spencer-Clegg et al. "Ivacaftor in the real world - early experience in a large adult CF centre". In: *Pediatr Pulm* 48.S36 (2013), p. 287. DOI: 10.1002/ppul.22898.

[160] I Trinh et al. "Ivacaftor in adults with cystic fibrosis: one-year experience in the real world setting". In: *Int J Clin Pharm* 35 (2013), p. 1331. DOI: 10.1007/s11096-013-9886-5.

[161] PJ Barry et al. "Sweat chloride is not a useful marker of clinical response to Ivacaftor". In: *Thorax* 68.S3 (2013), A120–1. DOI: 10.1136/thoraxjnl-2013-204457.252.

[162] AE Ewence et al. "Does ivacaftor improve objective measurements of health in patients with the G551D cystic fibrosis transmembrane conductance regulator (CFTR) protein mutation? The experience of a UK cystic fibrosis centre". In: *Thorax* 68.S3 (2013), A120. DOI: 10.1136/thoraxjnl-2013-204457.251.

[163] H Hebestreit et al. "Effects of ivacaftor on severely ill patients with cystic fibrosis carrying a G551D mutation". In: *J Cyst Fibros* 12.6 (2013), pp. 599–603. DOI: 10.1016/j.jcf.2013.05.006.

[164] PJ Barry et al. "Sweat chloride is not a useful marker of clinical response to ivacaftor". In: *Thorax* 69.6 (2014), p. 586. DOI: 10.1016/j.jcf.2013.03.004.

[165] HD Green et al. "The effect of ivacaftor therapy on clinical and PCR-identified microbial diversity of cystic fibrosis lung infection". In: *J Cyst Fibros* 13.S2 (2014), S56. DOI: 10.1016/S1569-1993(14)60173-9.

[166] LE Jenkins et al. "The use of LCI as an effective tool for monitoring clinical response to ivacaftor therapy in CF patients with at least one G551D-allele". In: *J Cyst Fibros* 13.S2 (2014), S40. DOI: 10.1016/S1569-1993(14)60125-9.

[167] SM Rowe et al. "Clinical mechanism of the cystic fibrosis transmembrane conductance regulator potentiator ivacaftor in G551D-mediated cystic fibrosis". In: *Am J Resp Crit Care* 190.2 (2014), pp. 175–84. DOI: 10.1164/rccm.201404-0703OC.

[168] RJ Adam et al. "Ivacaftor rapidly improves airway distensibility and vascular tone in people with G551D-CFTR suggesting a CF-related smooth muscle abnormality". In: *Pediatr Pulm* 49.S38 (2014), p. 354. DOI: 10.1002/ppul.23108.

[169] V Carnovale et al. "Effects of ivacaftor in cystic fibrosis patients carrying the G1244E mutation with severe lung disease". In: *Pediatr Pulm* 49.S38 (2014), p. 300. DOI: 10.1002/ppul.23108.

[170] KM Harman et al. "Exploring indices derived from multibreath washout (MBW) following treatment with ivacaftor". In: *Pediatr Pulm* 49.S38 (2014), pp. 267–8. DOI: 10.1002/ppul.23108.

[171] C Reilly et al. "Sustained effects of ivacaftor on muscle strength, body composition, anxiety and depression scores". In: *Pediatr Pulm* 49.S38 (2014), p. 298. DOI: 10.1002/ppul.23108.

[172] GS Sawicki et al. "The effect of ivacaftor on the rate of lung function decline in CF patients with a G551D-CFTR mutation". In: *J Cyst Fibros* 13.S2 (2014), S6. DOI: 10.1016/S1569-1993(14)60019-9.

[173] EF McKone et al. "Long-term safety and efficacy of ivacaftor in patients with cystic fibrosis who have the Gly551Asp-CFTR mutation: A phase 3, open-label extension study (PERSIST)". In: *Lancet Resp Med* 2.11 (2014), pp. 902–10. DOI: 10.1016/S2213-2600(14)70218-8.

[174] HD Green et al. "The effect of ivacaftor therapy on the microbial diversity of cystic fibrosis lung infection". In: *Thorax* 69.S2 (2014), A162. DOI: 10.1136/thoraxjnl-2014-206260.325.

[175] KML Harman et al. "Changes in indices derived from multibreath washout (MBW) following treatment with Ivacaftor in patients with cystic fibrosis". In: *Thorax* 69.S2 (2014), A167. DOI: 10.1136/thoraxjnl-2014-206260.335.

[176] SI Sheikh et al. "Computed tomography correlates with improvement with ivacaftor in cystic fibrosis patients with G551D mutation". In: *J Cyst Fibros* 14.1 (2015), pp. 84–9. DOI: 10.1016/j.jcf.2014.06.011.

[177] SI Sheikh et al. "Ivacaftor improves appearance of sinus disease on computerised tomography in cystic fibrosis patients with G551D mutation". In: *Clin Otolaryngol* 40.1 (2015), pp. 16–21. DOI: 10.1111/coa.12310.

[178] SL Heltshe et al. "Pseudomonas aeruginosa in cystic fibrosis patients with G551D-CFTR treated with ivacaftor". In: *Clin Infect Dis* 60.5 (2015), pp. 703–12. DOI: 10.1093/cid/ciu944.

[179] H Grasemann et al. "Effect of treatment with ivacaftor on exhaled nitric oxide". In: *J Cyst Fibros* 14.S1 (2015), S95. DOI: 10.1016/S1569-1993(15)30321-0.

[180] NJ Ronan et al. "Clinical outcomes of real-world Kalydeco (CORK) study - a prospective 12 month analysis addressing the impact of CFTR modulation on the cystic fibrosis lung". In: *J Cyst Fibros* 14.S1 (2015), S37. DOI: 10.1016/S1569-1993(15)30119-3.

[181] MG Zwolsman et al. "Partial recovery of exocrine pancreatic function and intestinal fat absorption after ivacaftor treatment". In: *J Cyst Fibros* 14.S1 (2015), S142. DOI: 10.1016/S1569-1993(15)30503-8.

[182] H Grasemann et al. "Effect of ivacaftor therapy on exhaled nitric oxide in patients with cystic fibrosis". In: *J Cyst Fibros* 14.6 (2015), pp. 727–32. DOI: 10.1016/j.jcf.2015.07.001.

[183] GG Einarsson et al. "The effect of ivacaftor treatment on airway microbial community dynamics in patients with G551D (the "Celtic" mutation)". In: *Pediatr Pulm* 50.S41 (2015), pp. 302–3. DOI: 10.1002/ppul.23297.

[184] A Fink et al. "Treatment response to ivacaftor in clinical practice: Analysis of the US CF Foundation Patient Registry". In: *Pediatr Pulm* 50.S41 (2015), p. 361. DOI: 10.1002/ppul.23297.

[185] M Kane et al. "Lung clearance index response in CF patients with class III CFTR mutations". In: *Pediatr Pulm* 50.S41 (2015), p. 278. DOI: 10.1002/ppul.23297.

[186] N Ronan et al. "The impact of CFTR modulation with ivacaftor on circulating inflammatory mediators and their correlation with clinical parameters in patients with the G551D mutation". In: *Pediatr Pulm* 50.S41 (2015), pp. 267–8. DOI: 10.1002/ppul.23297.

[187] J Taylor-Cousar et al. "Effect of ivacaftor in patients with advanced cystic fibrosis and a G551D-CFTR mutation: Safety and efficacy in an expanded access program in the United States". In: *J Cyst Fibros* 15.1 (2016), pp. 116–22. DOI: 10.1016/j.jcf.2015.01.008.

[188] JC Davies et al. "Safety, pharmacokinetics, and pharmacodynamics of ivacaftor in patients aged 2-years with cystic fibrosis and a CFTR gating mutation (KIWI): An open-label, single-arm study". In: *Lancet Resp Med* 4.2 (2016), pp. 107–15. DOI: 10.1016/S2213-2600(15)00545-7.

[189] S Bertin et al. "Clinical effects of ivacaftor on chronic rhinosinusitis". In: *J Cyst Fibros* 15.S1 (2016), S41–2. DOI: 10.1016/S1569-1993(16)30207-7.

[190] BM Button et al. "Increased maximal exercise capacity is associated with increased wellness across the spectrum of age and lung function in adults with cystic fibrosis (CF) after treatment with ivacaftor". In: *J Cyst Fibros* 15.S1 (2016), S15. DOI: 10.1016/S1569-1993(16)30110-2.

[191] P Iacotucci et al. "Effects of ivacaftor in cystic fibrosis patients carrying a non-G551D gating mutation". In: *J Cyst Fibros* 15.S1 (2016), S59. DOI: 10.1016/S1569-1993(16)30271-5.

[192] V Stallings et al. "3-month ivacaftor (Kalydeco®) treatment resulted in weight gain, decreased resting energy expenditure and gut inflammation, and increased pulmonary function, muscle mass and strength, and growth status". In: *J Cyst Fibros* 15.S1 (2016), S101–2. DOI: 10.1016/S1569-1993(16)30437-4.

[193] GL Zeybel et al. "Ivacaftor treatment is associated with a decline in symptoms of extra-oesophageal reflux in patients with cystic fibrosis and the G551D mutation". In: *J Cyst Fibros* 15.S1 (2016), S58–9. DOI: 10.1016/j.jcf.2016.07.004.

[194] A Aziz et al. "Ivacaftor-as effective in clinical practice?" In: *Eur Respir J* 48 (2016), p. 1257. DOI: 10.1183/13993003.congress-2016.PA1257.

[195] BM Button et al. "Increased total exercise time and maximal exercise capacity are associated with increased wellness in adults with CF across spectrum of age and lung function after treatment with Ivacaftor". In: *Pediatr Pulm* 51.S45 (2016), p. 368. DOI: 10.1002/ppul.23576.

[196] NN Sainath et al. "Improvement in weight, pulmonary function and other outcomes with 3-month Ivacaftor treatment differed by exocrine pancreatic status in people with cystic fibrosis gating mutations". In: *Pediatr Pulm* 51.S45 (2016), p. 418. DOI: 10.1002/ppul.23576.

[197] N Sainath et al. "Exocrine pancreatic status effects outcomes following 3-month Ivacaftor therapy in subjects with cystic fibrosis gating mutations". In: 63.S2 (2016), S107–8. DOI: 10.1097/MPG.0b013e3182849d2b.

[198] JI Schall et al. "3-month ivacaftor treatment resulted in weight gain, improved pulmonary function and growth status, and reduced resting energy expenditure and gut inflammation in people with cystic fibrosis gating mutations". In: *Pediatr Pulm* 51.S45 (2016), p. 422. DOI: 10.1002/ppul.23576.

[199] J Schall et al. "Improved pulmonary function and weight gain, reduced resting energy expenditure, and improved gut inflammation after 3-month Ivacaftor treatment in cystic fibrosis". In: *J Pediatr Gastr Nutr* 63.S2 (2016), pp. 406–7.

[200] M Mesbahi et al. "Changes of CFTR functional measurements and clinical improvements in cystic fibrosis patients with non p.Gly551Asp gating mutations treated with ivacaftor". In: *J Cyst Fibros* 16.1 (2017), pp. 45–8. DOI: 10.1016/j.jcf.2016.08.006.

[201] VA Stallings et al. "Ivacaftor treatment in cystic fibrosis and improvement in resting energy expenditure gut inflammation, and fat absorption". In: *Ital J Pediatr* 43.S1 (2017), p. 21. DOI: 10.1186/s13052-017-0327-2.

[202] C Luna-Paredes et al. "Impact of treatment with ivacaftor on Spanish cystic fibrosis patients with gating mutations". In: *J Cyst Fibros* 16.S2 (2017), S74. DOI: 10.1016/S1569-1993(17)30406-X.

[203] A Strang, AJ Fischer, and A Chidekel. "Pseudomonas eradication and clinical effectiveness of Ivacaftor in four Hispanic patients with S549N". In: *Pediatr Pulm* 52.7 (2017), E37–9. DOI: 10.1002/ppul.23665.

[204] A Dagan et al. "Ivacaftor for the p.ser549arg gating mutation - the Israeli experience". In: *Pediatr Pulm* 52.S47 (2017), S302. DOI: 10.1002/ppul.23840.

[205] M Mouzaki et al. "Weight increase in CF patients on Kalydeco is due to decrease in resting energy expenditure and associated with increase in adipose tissue". In: *Pediatr Pulm* 52.S47 (2017), p. 448. DOI: 10.1002/ppul.23840.

[206] NN Sainath et al. "Effect of ivacaftor treatment on dietary intake in Italian and North American subjects with cystic fibrosis". In: *Pediatr Pulm* 52.S47 (2017), pp. 462–3. DOI: 10.1002/ppul.23840.

[207] M Mouzaki et al. "Increase in body weight of patients with cystic fibrosis receiving Kalydeco treatment is due to increase in adipose tissue". In: *J Pediatr Gastr Nutr* 65.S2 (2017), S213.

[208] N Sainath et al. "Changes in dietary intake with Ivacaftor treatment in Italian and North American subjects with cystic fibrosis". In: *J Pediatr Gastr Nutr* 65.S2 (2017), S322.

[209] C Dryden et al. "The impact of 12 months treatment with ivacaftor on Scottish paediatric patients with cystic fibrosis with the G551D mutation: A review". In: *Arch of Dis Child* 103 (2018), pp. 68–70. DOI: 10.1136/archdischild-2015-310420.

[210] FJ Accurso et al. "Effect of VX-770 in Persons with Cystic Fibrosis and the G551D-CFTR Mutation". In: *New Engl J Med* 363.21 (2010), pp. 1991–2003. DOI: 10.1056/NEJMoa0909825.

[211] BJ Plant et al. "VX-770, an investigational CFTR potentiator, in subjects with CF and the G551D mutation". In: *Eur Respir J* 38 (2011), p. 4655.

[212] R Aherns et al. "VX-770 in subjects 6 to 11 years with cystic fibrosis and the G551D-CFTR mutation". In: *Pediatr Pulm* 46.S34 (2011), S283. DOI: 10.1002/ppul.21581.

[213] B Ramsey et al. "Efficacy and safety of VX-770 in subjects with cystic fibrosis and the G551D-CFTR mutation". In: *Pediatr Pulm* 46.S34 (2011), pp. 286–7. DOI: 10.1002/ppul.21581.

[214] BW Ramsey et al. "A CFTR potentiator in patients with cystic fibrosis and the G551D mutation". In: *New Engl J Med* 365.18 (2011), pp. 1663–72. DOI: 10.1056/NEJMoa1306220.

[215] S Elborn et al. "Pulmonary effects of the investigational CFTR potentiator, ivacaftor, in two phase 3 trials in subjects with CF who have the G551D-CFTR mutation". In: *Am J Resp Crit Care* 185 (2012), A2464. DOI: 10.1164/ajrccm-conference.2012.185.1_MeetingAbstracts.A2464.

[216] JC Davies et al. "Ivacaftor in subjects 6 to 11 years of age with cystic fibrosis and the G551D-CFTR mutation". In: *J Cyst Fibros* 11.S1 (2012), S13. DOI: 10.1016/s1569-1993%2812%2960043-5.

[217] JC Davies et al. "Effect of ivacaftor on lung function in subjects with CF who have the G551D-CFTR mutation and mild lung disease: A comparison of lung clearance index (LCI) vs. spirometry". In: *J Cyst Fibros* 11.S1 (2012), S15. DOI: 10.1016/S1569-1993(12)60049-6.

[218] M Griese et al. "Pulmonary exacerbations in a phase 3 trial of ivacaftor in subjects with cystic fibrosis who have the G551D-CFTR mutation". In: *J Cyst Fibros* 11.S1 (2012), S67. DOI: 10.1016/S1569-1993(12)60041-1.

[219] JC Davies et al. "Lung clearance index to evaluate the effect of Ivacaftor on lung function in subjects with CF who have the G551D-CFTR mutation and mild lung disease". In: *Pediatr Pulm* 47.S35 (2012), p. 311. DOI: 10.1002/ppul.22682.

[220] S Elborn et al. "Effects of the CFTR potentiator, Ivacaftor, in two phase 3 trials in subjects with CF who have the G551D-CFTR mutation". In: *Eur Respir J* 40.1840 (2012).

[221] F Ratjen et al. "Effect of Ivacaftor on lung clearance index and FEV1 in subjects with CF who have the G551D-CFTR mutation and mild lung disease". In: *Eur Respir J* 40.1840 (2012).

[222] JC Davies et al. "Lung clearance index to evaluate the effect of ivacaftor on lung function in subjects with CF who have the G551D-CFTR mutation and mild lung disease". In: *Thorax* 67.S2 (2012), A58–9. DOI: 10.1136/thoraxjnl-2012-202678.124.

[223] PJ Barry et al. "UK and Ireland review of ivacaftor in severe CF: Impact on hospitalisations and antibiotic use". In: *J Cyst Fibros* 12.S1 (2013), S62. DOI: 10.1016/S1569-1993(13)60196-4.

[224] JC Davies et al. "Efficacy and safety of ivacaftor in patients aged 6 to 11 years with cystic fibrosis with a G551D mutation". In: *Am J Resp Crit Care* 187.11 (2013), pp. 1219–25. DOI: 10.1164/rccm.201301-0153OC.

[225] P Flume et al. "Pulmonary exacerbations in CF patients with the G551D-CFTR mutation treated with ivacaftor". In: *J Cyst Fibros* 12.S1 (2013), S63. DOI: 10.1016/S1569-1993(13)60200-3.

[226] JC Davies et al. "Assessment of clinical response to ivacaftor with lung clearance index in cystic fibrosis patients with a G551D-CFTR mutation and preserved spirometry: A randomised controlled trial". In: *Lancet Resp Med* 1.8 (2013), pp. 630–8. DOI: 10.1016/S2213-2600(13)70182-6.

[227] JS Elborn et al. "Effect of ivacaftor in patients with cystic fibrosis and the G551D-CFTR mutation who have baseline FEV1>90% of predicted". In: *Pediatr Pulm* 48.S36 (2013), p. 298. DOI: 10.1002/ppul.22898.

[228] NJ Ronan et al. "Real world sustained efficacy, tolerability and satisfaction with ivacaftor use in a single adult cystic fibrosis centre cohort". In: *Am J Resp Crit Care* 189 (2014), A2850.

[229] NJ Ronan et al. "Clinical outcomes of real world Kalydeco (CORK) study". In: *J Cyst Fibros* 13.S2 (2014), S56. DOI: 10.1016/S1569-1993(14)60174-0.

[230] PJ Barry et al. "Effects of ivacaftor or in patients with cystic fibrosis who carry the G551D mutation and have severe lung disease". In: *Chest* 146.1 (2014), pp. 152–8. DOI: 10.1378/chest.13-2397.

[231] L Nedevska et al. "Ivacaftor: A new treatment for cystic fibrosis which results in reduced sweat chloride and improved clinical outcomes". In: *Clin Chem Lab Med* 52.11 (2014), A268–A9. DOI: 10.1515/cclm-2014-0890.

[232] K De Boeck et al. "Efficacy and safety of ivacaftor in patients with cystic fibrosis and a non-G551D gating mutation". In: *J Cyst Fibros* 13.6 (2014), pp. 674–80. DOI: 10.1016/j.jcf.2014.09.005.

[233] L Guhaniyogi et al. "Transformational care at the All Wales Adult CF Centre (AWACFC) - the impact of ivacaftor (Kalydeco®) one year on". In: *J Cyst Fibros* 14.S1 (2015), S92. DOI: 10.1016/S1569-1993(15)30312-X.

[234] D Edgeworth et al. "Exercise improvements in ivacaftor treated G551D cystic fibrosis patients are not solely related to FEV1 and sweat changes". In: 46 (2015), A2047. DOI: 10.1183/13993003.congress-2015.PA2047.

[235] Y Bai et al. "Ivacaftor long-term safety study: Analysis of 2013 US CF Foundation patient registry data". In: *Pediatr Pulm* 50.S41 (2015), p. 284. DOI: 10.1002/ppul.23297.

[236] GS Sawicki et al. "Sustained benefit from ivacaftor demonstrated by combining clinical trial and cystic fibrosis patient registry data". In: *Am J Resp Crit Care* 192.7 (2015), pp. 836–42. DOI: 10.1164/rccm.201503-0578OC.

[237] Y Bai et al. "Real-world outcomes in patients (pts) with cystic fibrosis (CF) treated with ivacaftor (IVA): analysis of 2014 US and UK CF registries". In: *J Cyst Fibros* 51.S1 (2016), S41. DOI: 10.1016/S1569-1993(16)30205-3.

[238] Y Bai et al. "Real-world outcomes in young (6- to 12-year-old) patients (pts) with cystic fibrosis (CF) treated with ivacaftor (IVA): analysis of 2014 US and UK CF registries data". In: *J Cyst Fibros* 15.S1 (2016), S57–8. DOI: 10.1016/S1569-1993(16)30206-5.

[239] L Fila et al. "Ivacaftor in cystic fibrosis adults: Czech experience with six years of follow-up". In: *Biomedical Papers* 160.2 (2016), pp. 276–9. DOI: 10.5507/bp.2016.029.

[240] M Hassan et al. "Reduction in pulmonary exacerbations (PEx) after initiation of ivacaftor: A retrospective cohort study among patients with cystic fibrosis (CF) treated in real-world settings". In: *J Cyst Fibros* 15.S1 (2016), S58. DOI: 10.1016/S1569-1993(16)30268-5.

[241] D Hubert et al. "Retrospective observational study of French patients with cystic fibrosis and a G551D-CFTR mutation after 1 and 2 years of treatment with ivacaftor". In: *J Cyst Fibros* 15.S1 (2016), S58. DOI: 10.1016/j.jcf.2017.07.001.

[242] P Iacotucci et al. "Effects of ivacaftor in cystic fibrosis patients carrying a non-G551D gating mutation with severe lung disease". In: *J Cyst Fibros* 15.S1 (2016), S59. DOI: 10.1016/S1569-1993(16)30270-3.

[243] N Volkova et al. "Disease progression in patients (pts) with cystic fibrosis (CF) treated with ivacaftor (IVA): Analysis of real-world data from the UK CF Registry". In: *J Cyst Fibros* 15.S1 (2016), S41. DOI: 10.1016/S1569-1993(16)30206-5.

[244] L Bessonova et al. "Analysis of real-world outcomes in patients with CF treated with Ivacaftor from the 2014 US and UK CF registries". In: *Pediatr Pulm* 51.S45 (2016), pp. 381–2. DOI: 10.1002/ppul.23576.

[245] M Hassan et al. "One-year evaluation of pulmonary exacerbation outcomes among patients with cystic fibrosis initiated on ivacaftor in a multistate Medicaid population". In: *Pediatr Pulm* 51.S45 (2016), p. 355. DOI: 10.1002/ppul.23576.

[246] N Volkova et al. "Analysis of disease progression in patients with CF treated with Ivacaftor in the real world using data from the UK CF registry". In: *Pediatr Pulm* 51.S45 (2016), p. 382. DOI: 10.1002/ppul.23576.

[247] L Bessonova et al. "Disease progression in patients (pts) with cystic fibrosis (CF) treated with ivacaftor (IVA): analysis of real-world data from the UK CF Registry". In: *J Cyst Fibros* 16.S1 (2017), S167. DOI: 10.1016/S1569-1993(17)30321-1.

[248] L Bessonova et al. "Real-world outcomes in patients (pts) with cystic fibrosis (CF) treated with ivacaftor (IVA): analysis of 2015 US and UK CF registry data". In: *J Cyst Fibros* 16.S1 (2017), S49–50. DOI: 10.1016/S1569-1993(17)30321-1.

[249] L Kirwan et al. "Temporal trends in key outcome measures in cystic fibrosis patients treated with ivacaftor: Real-world data from the Irish CF registry". In: *J Cyst Fibros* 16.S1 (2017), S48. DOI: 10.1016/S1569-1993(17)30317-X.

[250] D Edgeworth et al. "Improvement in exercise duration, lung function and well-being in G551D-Cystic Fibrosis patients: A double-blind, placebo-controlled, randomised, cross-over study with ivacaftor". In: *Clin Sci* 131.15 (2017), pp. 2037–45. DOI: 10.1042/CS20170995.

[251] N Volkova et al. "Disease progression in patients with CF treated with ivacaftor: Analyses of real-world data from the US and UK CF registries". In: *Pediatr Pulm* 52.S47 (2017), p. 407. DOI: 10.1002/ppul.23840.

[252] N Volkova et al. "Real-world outcomes in patients with CF treated with ivacaftor: 2015 US and UK CF registry analyses". In: *Pediatr Pulm* 52.S47 (2017), p. 407. DOI: 10.1002/ppul.23840.

[253] A Dagan et al. "Ivacaftor for the p.Ser549Arg (S549R) gating mutation – The Israeli experience". In: *Resp Med* 131 (2017), pp. 225–8. DOI: 10.1016/j.rmed.2017.08.026.

[254] PA Flume et al. "Recovery of lung function following a pulmonary exacerbation in patients with cystic fibrosis and the G551D-CFTR mutation treated with ivacaftor". In: *J Cyst Fibros* 17.S1 (2018), S83–8. DOI: 10.1016/j.jcf.2017.06.002.

[255] D Hubert et al. "Retrospective observational study of French patients with cystic fibrosis and a Gly551Asp-CFTR mutation after 1 and 2 years of treatment with ivacaftor in a real-world setting". In: *J Cyst Fibros* 17.S1 (2018), pp. 89–95. DOI: 10.1016/j.jcf.2017.07.001.

[256] EF McKone et al. "Effect of genotype on phenotype and mortality in cystic fibrosis: A retrospective cohort study". In: *Lancet* 361.9370 (2003), pp. 1671–6. DOI: 10.1016/S0140-6736(03)13368-5.

[257] GS Sawicki et al. "Patients with cystic fibrosis and a G551D or homozygous F508del mutation: Similar lung function decline". In: *Am J Respir Crit Care Med* 195.12 (2017), pp. 1673–6. DOI: 10.1164/rccm.201608-1678LE.

[258] DP Stevens and BC Marshall. "A decade of healthcare improvement in cystic fibrosis: Lessons for other chronic diseases". In: *BMJ Qual Saf* 23.S1 (2014), pp. 2–3. DOI: 10.1136/bmjqs-2014-002871.

[259] M Lipsitch, EJ Tchetgen Tchetgen, and T Cohen. "Negative controls: A tool for detecting confounding and bias in observational studies". In: *Epidemiology* 21.3 (2010), pp. 383–8. DOI: 10.1097/EDE.0b013e3181d61eeb.Negative.

[260] GD Smith. "Negative control exposures in epidemiologic studies". In: *Epidemiology* 23.2 (2012), pp. 351–2. DOI: 10.1097/EDE.0b013e3182460c23.

[261] MG Weisskopf, EJ Tchetgen Tchetgen, and R Raz. "On the use of imperfect negative control exposures in epidemiologic studies". In: *Epidemiology* 27.3 (2016), pp. 365–7. DOI: 10.1097/EDE.0000000000000454.

[262] EJ Tchetgen Tchetgen. "The control outcome calibration approach for causal inference with unobserved confounding". In: *Am J Epidemiol* 179.5 (2014), pp. 633–40. DOI: 10.1093/aje/kwt303.

[263] T Sofer et al. "On negative outcome control of unobserved confounding as a generalization of difference-in-differences". In: *Stat Sci* 31.3 (2016), pp. 348–61. DOI: 10.1007/s11897-014-0247-z.Pathophysiology.

[264] WD Flanders, MJ Strickland, and M Klein. "A new method for partial correction of residual confounding in time-series and other observational studies". In: *Am J Epidemiol* 185.10 (2017), pp. 941–949. DOI: 10.1093/aje/kwx013.

[265] YG Wang and VJ Carey. "Working correlation misspecification, estimation and covariate design: Implications for generalized estimating equation performance". In: *Biometrika* 90.1 (2003), pp. 29–41. DOI: 10.1093/biomet/90.1.29.

[266] G Molenberghs and G Verbeke. *Models for discrete longitudinal data*. Springer US, 2005.

[267] S Rabe-Hesketh and A Skrondal. *Multilevel and longitudinal modeling using Stata - Volume II: categorical responses, counts and survival*. College Station, Texas: Stata Press, 2012.

[268] PD Allison and R Waterman. "Fixed-effects negative binomial regression models". In: *Sociology Methodology* 32.1 (2002), pp. 247–65. DOI: 10.1111/1467-9531.00117.

[269] ME McGarry et al. "In vivo and in vitro ivacaftor response in cystic fibrosis patients with residual CFTR function: N-of-1 studies". In: *Pediatr Pulm* 52.4 (2017), pp. 472–9. DOI: 10.1002/ppul.23659.

[270] *Clinical Commissioning Policy: Ivacaftor for Cystic Fibrosis*. Tech. rep. 2013.

[271] D Rizopoulos. *Joint models for longitudinal and time-to-event data with applications in R*. Boca Raton, FL: CRC Press, 2012.

[272] L Wen, GM Terrera, and SR Seaman. "Methods for handling longitudinal outcome processes truncated by dropout and death". In: *Biostatistics* (2018), pp. 1–19. DOI: 10.1093/biostatistics/kxy001.

[273] *Cystic fibrosis: long-term azithromycin*. Tech. rep. 2014.

[274] JA Nick et al. "Azithromycin may antagonize inhaled tobramycin when targeting Pseudomonas aeruginosa in cystic fibrosis". In: *Ann Am Thorac Soc* 11.3 (2014), pp. 342–50. DOI: 10.1513/AnnalsATS.201310-352OC.

[275] NJ Rowbotham et al. "The top 10 research priorities in cystic fibrosis developed by a partnership between people with CF and healthcare providers". In: *Throax* 73.4 (2018), pp. 388–90. DOI: 10.1136/thoraxjnl-2017-210473.

[276] *Cystic Fibrosis Foundation Patient Registry 2016 Annual Data*. Tech. rep. 2017.

[277] *The Canadian Cystic Fibrosis Registry 2016 Annual Data Report*. Tech. rep. 2017.

[278] E Lynge, JL Sandegaard, and M Rebolj. "The Danish national patient register". In: *Scand J Public Healt* 39.7 (2011), pp. 30–3. DOI: 10.1177/1403494811401482.

[279] C Lusivika-Nzinga et al. "Performance of the marginal structural cox model for estimating individual and joined effects of treatments given in combination". In: *BMC Med Res Methodol* 17.1 (2017), pp. 1–11. DOI: 10.1186/s12874-017-0434-1.

[280] S Picciotto et al. "Structural nested cumulative failure time models to estimate the effects of interventions". In: *J Am Stat Assoc* 107.499 (2012), pp. 1–35. DOI: 10.1080/01621459.2012.682532.Structural.

[281] G Molenberghs et al. *Handbook of Missing Data Methodology*. Chapman & Hall/CRC, 2015.

[282] RM Daniel, BL De Stavola, and SN Cousens. "gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula". In: *Stata J* 11.4 (2011), pp. 479–517.

[283] SR Seaman et al. "Combining multiple imputation and inverse-probability weighting". In: *Biometrics* 68.1 (2012), pp. 129–37. DOI: 10.1111/j.1541-0420.2011.01666.x.

[284] P Han. "Combining inverse probability weighting and multiple imputation to improve robustness of estimation". In: *Scand J Stat: Theory Applications* 43.1 (2015), pp. 246–60. DOI: 10.1111/sjos.12177.

[285] MD Schluchter, MW Konstan, and PB Davis. "Jointly modelling the relationship between survival and pulmonary function in cystic fibrosis patients". In: *Stat Med* 21.9 (2002), pp. 1271–87. DOI: 10.1002/sim.1104.

[286] J Barrett et al. "Joint modelling of repeated measurements and time-to-event outcomes: Flexible model specification and exact likelihood inference". In: *J R Stat Soc B* 77.1 (2015), pp. 131–48. DOI: 10.1111/rssb.12060.

[287] E Waldmann et al. "Boosting joint models for longitudinal and time-to-event data". In: *Biom J* 59.6 (2017), pp. 1104–21. DOI: 10.1002/bimj.201600158.

Appendix A

Paper 1 - October 2017

This appendix contains the published version of the paper entitled "Data Resource Profile: The UK Cystic Fibrosis Registry". It was published in the International Journal of Epidemiology in October 2017.[10] Chapter 2 is based on work found in this paper.

For the preparation of this paper, I helped in cleaning the UK CF Registry data, drafting the "Data collected" section, producing Figure 1 and approving the final version of the paper.

Student's Signature 

Date: 29th January, 2019

Supervisor's Signature: 

Date: 29th January, 2019

International Journal of Epidemiology, 2018, 9–10e
doi: 10.1093/ije/dyx196
Advance Access Publication Date: 3 October 2017
Data Resource Profile

Data Resource Profile

Data Resource Profile: The UK Cystic Fibrosis Registry

David Taylor-Robinson,[1]* Olia Archangelidi,[2] Siobhán B Carr,[3] Rebecca Cosgriff,[4] Elaine Gunn,[4] Ruth H Keogh,[5] Amy MacDougall,[2] Simon Newsome,[5] Daniela K Schlüter,[6] Sanja Stanojevic[7] and Diana Bilton;[2] on behalf of the CF-EpinNet collaboration

[1]Department of Public Health and Policy, University of Liverpool, Liverpool, UK, [2]National Heart and Lung Institute, Imperial College London, London, UK, [3]Department of Paediatric Respiratory Medicine, Royal Brompton and Harefield NHS Foundation Trust, London, [4]Cystic Fibrosis Trust, Aldgate, London, UK, [5]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK, [6]Centre for Health Informatics, Computing and Statistics (CHICAS), Lancaster University, Lancaster, UK and [7]Translational Medicine, Hospital for Sick Children and Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, ON, Canada

*Corresponding author. Department of Public Health and Policy, Farr Institute, University of Liverpool, Liverpool L69 3GB, UK. E-mail: dctr@liv.ac.uk)

Editorial decision 9 August 2017; Accepted 22 August 2017

Data resource basics

The UK Cystic Fibrosis Registry is a national, secure, centralized database sponsored and managed by the Cystic Fibrosis Trust, with UK National Health Service (NHS) research ethics approval and consent from each person for whom data are collected. First established in 1995, it records longitudinal health data on all people with cystic fibrosis (CF) in England, Wales, Scotland and Northern Ireland, and to date has captured data on over 12 000 individuals.

Cystic fibrosis is an inherited, chronic, progressive condition occurring in around 1 in 2500 live births in the UK, with around 200–300 new diagnoses annually. Children are generally diagnosed in the first few months of life with universal newborn screening being implemented in 2007 in the UK, though some people are diagnosed into adulthood. For instance, 29 people aged over 16 years were diagnosed with CF in the UK in 2015.[1] Patients diagnosed with CF subsequently require intensive support from family and health care services. Most patients die prematurely from their disease through respiratory failure, and in the 1930 s and 40 s survival beyond childhood was rare.[2] There have been impressive improvements in survival over subsequent decades; for instance, the median life expectancy of children with cystic fibrosis born in 1990 was estimated to be 40 years, double that of estimates 20 years earlier.[3]

In the UK, children with CF are treated in one of 33 specialist centres (associated with over 100 smaller network clinics). At between 16 years and 18 years of age, children transfer to one of 27 adult specialist centres. All centres and network clinics routinely collect data in a standardized fashion. When patients with CF attend a new CF centre in the UK, they or their parents consent to information on their health and treatment being collected and stored in the CF Registry. The patient information and consent form also covers the issue of linking registry data to the UK Office for National Statistics. When transitioning to adult services, the young adult is given the opportunity to confirm or withdraw consent. People with CF will also re-consent if they change their primary centre of care. The Registry records information about the health and

treatment, and health care use (e.g. hospital days) and outcomes of patients from diagnosis onwards. The Registry contains personal identifiers, seen only by the local centre team and CF Registry data managers. Reports generated by the Registry are used as the evidence base for commissioning care and pharmacovigilance of new therapies. Harnessing the rich data in CF registries in the UK and beyond offers the opportunity to improve the lives of patients with CF now, and to establish an essential data resource for future research.

History of the registry

The UK CF Registry started as the UK CF Database, which was established at the University of Dundee, Scotland, in 1995. Initially data were collected from 56 paediatric and adult CF clinics, using standardized forms, and validated through a system of double data entry, range checking and error correction.[4] Between 2005 and 2007, the Cystic Fibrosis Trust rolled out a national UK-wide web-based system, following new ethics approval and re-consent of all patients, with migration of the data to a new system. The data collection system thus changed from a paper-based return system to using the online 'PortCF' software which mimicked the data collection and storage system used by the Cystic Fibrosis Foundation Patient Registry in the USA. During this transfer there was extensive retrospective data cleaning and checking. In 2012. the Registry commenced production of reports that are used by the NHS England to make payment by results (PbR) tariff payments to CF centres. Linking the Registry to NHS reimbursement processes significantly improved the completion of data. Recently, the PortCF system has been replaced by new UK CF Registry software developed by the UK-based web development specialists Net Solving Ltd. The new system has been developed with patient involvement and includes interactive elements that may, in due course, allow patients to access their own data. The UK CF Registry has thus far led to the production of 10 annual reports, with the latest 2015 data published in August 2016.[1] The UK CF Registry Steering Committee was established in 2007 to oversee development of the Registry, annual reports and research governance; it meets regularly and includes medical, sponsor (Cystic Fibrosis Trust), commissioner, statistician, patient and parent representation. It has recently set up a subcommittee, the UKCF Registry Research Committee, to allow more detailed oversight and governance of this area.

Data collected

The dataset contains: time-invariant variables, such as sex, genotype and date of birth; and longitudinal variables that change over time, such as weight and measures of lung function. CF patients are seen in the outpatient clinic for a comprehensive annual review, including evaluation of clinical status, pulmonary function, microbiology of respiratory tract secretions and use of major CF-related therapies. The minimum data collection requirement for the UK CF Registry is an annual dataset, usually taken from the annual review clinic visit. The data collected at the annual review are indicated in the dataset, and can be distinguished from 'encounter'-based data collected in the interval between clinic visits. In this data resource profile, we only describe the annual review data which are the basis of the CF-Epinet dataset described below. Some clinics use the Registry to collect encounter data, but these are not systematically collected. Thus the data in the registry mostly derive from the annual review, rather than being encounter-based, though the annual review data include certain summaries of information since the previous review–for example. the number of days a patient has been on intravenous antibiotics and the best % forced epiratory volume in 1 s (FEV_1) recorded in the previous year. Data are collected in key areasincluding: demographics (including genotyping and diagnosis data), hospital admissions and intravenous therapies, pulmonary function, chronic medications, culture and microbiology, health complications, nutritional assessment, physiotherapy, smoking, socioeconomic status and outcomes (death and transplants).[4] Templates of the data collection forms showing all variables collected are available to download from the CF Registry portal as well as from the Cystic Fibrosis Trust Registry web page. [5]

Coverage

Of the 22 countries providing data to the wider European Cystic Fibrosis Society Patient Registry,[6] the UK CF Registry is the largest national database and the most complete in terms of coverage. Currently data on 12 201 patients are captured in the UK Registry (alive, dead or lost to follow-up) with 9734 (79.8%) still in follow-up at the end of 2015. In total there are data on over 100 000 annual assessments. Figure 1 shows the cumulative count of patients captured in the Registry by year, including patients who have died. Figure 1 also shows annual counts of annual reviews, deaths, and losses to follow-up (defined as patients with no annual reviews for 2 years in succession). The number of people captured in the dataset has increased year on year, with increases coinciding with the move to the web-based system, followed by the incentivizing of data collection for NHS funding purposes in England and Scotland in 2012. The number of patients for whom a 'complete' dataset, defined as the data required to produce

International Journal of Epidemiology, 2018, Vol. 47, No. 1 **10a**

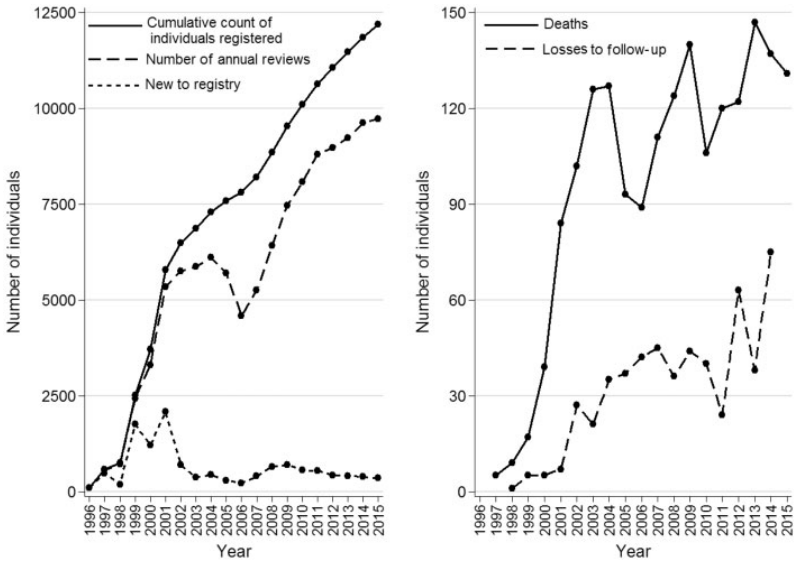


Figure 1. Cumulative count of individuals captured in the registry and number with annual review data in each year (left panel). Count of deaths and losses to follow-up (right panel).

the range of key clinical outcomes relating to growth, lung function and treatment presented in the annual reports, was recorded at 82% in 2009, and this has increased year on year, with the figure up to 89% for the latest (2015) annual report.[1]

For the purposes of illustrating the longitudinal structure of the data, we consider the patient's weight, since weight is one of the most commonly collected outcomes in the dataset, collected at 117 482 annual reviews on 12 201 patients between 1996 up to 2015 in the UK. A total 0f 78% of individuals had five or more weight measurements, with a mean number of nine measurements. Figure 2 shows all patient's weight data, presented as age-standardized z-scores[7] plotted against age, with randomly selected individual trajectories highlighted. Z-scores in adults were calculated assuming the weight-for-age distribution at age 19.

Quality

Data quality is assured through a number of mechanisms. For clinicians and others entering data into the Registry, up-to-date user guides are available on the UK CF Registry portal and contextual help text is available next to individual variables, to instruct users as to how to interpret the question and use the software. Training videos are also available to assist clinicians in entering and monitoring data from their own clinic. Software functionality encourages CF centres to monitor their own data on an ongoing basis. A quality dashboard gives users an on-demand, at a glance view of their data completeness as well as summaries of key clinical indicators such as lung function, with live benchmarks against national averages.

The Registry data collection software performs data validation checks at point of entry. These include: the enforcing of mandatory data; range checks for clinically valid values; ensuring that only valid characters are entered; and text and visual prompts to ensure that data (including dates) are not illogical or conflicting and to encourage completion of core data. The Cystic Fibrosis Trust also supports a Registry Annual Meeting in July each year. This event is free to attend for all CF centre employees who enter data onto the Registry, and is designed to showcase current registry research and offer training and best practice-sharing opportunities.

The Cystic Fibrosis Trust Registry team perform training and validation visits at CF centres, to ensure that users are aware of data entry guidelines and encouraged to improve the completeness and accuracy of their data. After the data entry deadline on the 31 January each year, Registry data managers and statisticians perform a variety of data cleaning checks, such as tracking of patient transition from paediatric to adult clinics. Where longitudinal analysis detects apparent inconsistencies that cannot be automatically cleaned, the Registry team liaise directly with CF centres to check and, where relevant, correct data.

Data resource use

Registry studies have been crucial in informing our understanding of the epidemiology, changing demographics,

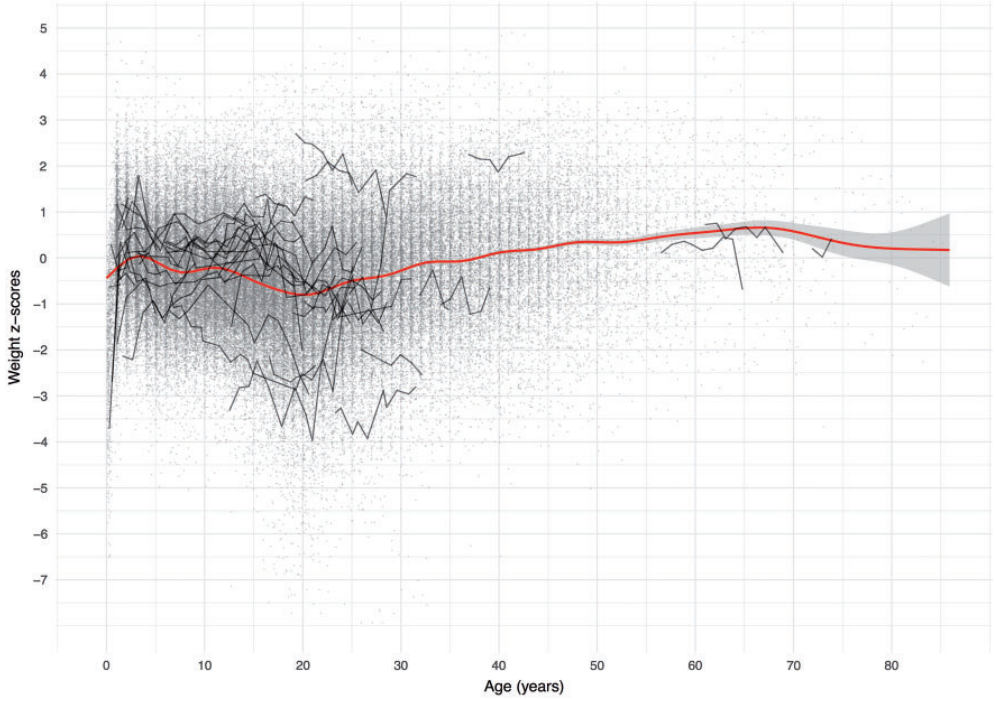


Figure 2. Spaghetti plot for weight z-score versus age, illustrating the longitudinal nature of the data collected in the UKCF Registry. Each dot ($n = 117\,482$) represents a weight z-score measure on a person in the dataset. The smoothed cross-sectional population average is shown in red (95% confidence intervals) and 50 randomly selected individual trajectories are in black.

outcomes and treatments in CF. Two examples of analyses of UK Registry data are provided in the text box.

In addition to epidemiological studies, the UK CF Registry team produce reports on long-term drug safety required by the European Medicines Agency (EMA). These reports contain anonymized, aggregated data and allow safety- and efficacy-monitoring of new therapies for cystic fibrosis. There is also a facility to support Registry-based clinical trials. For instance, the UK CF Registry is currently, via a specially designed study module, running a CF Registry-based clinical trial, the cystic fibrosis (CF) anti-staphylococcal antibiotic prophylaxis trial (CF START): a randomized registry trial to assess the safety and efficacy of flucloxacillin as a long-term prophylaxis agent for infants with CF.[8]

Making use of the Port-CF system, with similar variables collected in the UK and USA, Goss and colleagues compare CF outcomes and use of treatments between the two countries. Their cross-sectional analysis suggested that the USA does better in terms of lung function in children, and one hypothesis raised is that this may be due to more intensive treatment in the early years in the USA.[9]

Studies making use of data from the UK CF Registry are increasing,[5–11,15,16] but relatively few have made use of the longitudinal nature of the data. One longitudinal study has assessed the impact of socioeconomic status on outcomes and treatment use in the UK population.[7] More disadvantaged children with CF in the UK were found to have significantly worse growth and lung function, and were more likely to have chronic *P. aeruginosa* infection. There was evidence that in the NHS, clinicians in making decisions about treatments for children take deprivation as well as disease status into account, and this may mitigate some effects of social disadvantage. The study raises concerns about the provision of therapies such as DNase to people living in disadvantaged areas.

The Cystic Fibrosis Epidemiological Network (CF-EpiNet)

In recognition of the potential to better harness data from registries to improve patient outcomes, the Cystic Fibrosis Trust have funded a Strategic Research Centre (CF-EpiNet) focused on CF data and epidemiology. CF-EpiNet is focused on: Registry enhancement; application of state-of-the art

International Journal of Epidemiology, 2018, Vol. 47, No. 1 10c

statistical modelling techniques to longitudinal data; and economic modelling. An important part of this project will involve data cleaning, harmonization of variables over time and the generation of a research-ready 'CF-EpiNet' dataset. Algorithms and code for cleaning the data should allow researchers to a obtain more up-to-date version of the data, with extra years of longitudinal data added.

CF-EpiNet aims to develop a holistic view of how CF impacts on patients across the life course, and to identify modifiable targets for clinical and policy intervention (e.g. early life exposures, access to therapies, educational support) by the application of modern statistical methods for longitudinal analysis and by direct comparisons with international datasets. Key aims are to optimize the scientific value of the Registry by linking it to data from other administrative datasets, and by undertaking a quality of life survey. The quality of life component of the project has been developed within a portal, 'My CF Registry', allowing patient-reported outcome measures to be entered directly by patients. In the future, this portal could enable people with CF to view their historical clinical data, self-report data and opt into additional uses of Registry data that will enhance the value of the Registry to the CF community.

Strengths and weaknesses of the Registry

Registries can provide valuable insights into variations in clinical outcomes, quality of care and the safety and/or effectiveness of treatments. However, the usefulness and applicability of Registry data rely on the quality of several aspects: the measurements and information recorded; the accuracy of data input, data storage and export; and appropriate data analysis and interpretation, bearing in mind the inherent shortcomings of routinely collected data.[17,18]

A key strength of the UK CF Registry is the population-level coverage. In the UK, the Registry is estimated to capture almost all of the CF population; any consenting patients attending NHS clinics will have annual data routinely collected into the database. Furthermore, the dataset is of high quality, with robust systems for data cleaning and checking. The UK dataset represents one of the largest national CF datasets outside the USA,[17] and this provides the statistical power to precisely estimate parameters of interest. The UK CF Registry contains a wide range of clinical, health care and social information, allowing for robust adjustment for appropriate covariates in statistical analyses. A further consequence of the high level of population coverage in the UK, coupled with a universal health care system, is that analyses can cover individuals across the full range of the socioeconomic spectrum in the UK. The unique Registry identification number facilitates longitudinal research, an advantage over other registries which rely on yearly snapshot population data. In addition, the coverage for core variables summarized in the annual reports, such as weight and $\%FEV_1$, are high.

One key limitation of the UK CF Registry compared with some other CF registries is that it at present mandatorily requires only annual review data, rather than all clinical encounters, which limits some of the research questions that can be addressed. There are also some gaps in information collected, such as primary and secondary non-CF-related care episodes. Data linkage to other national datasets has been a long-standing aim of the UK CF Registry, and the CF-EpiNet project is exploring relevant linkages to include those with primary and secondary care databases, mortality data, census data, databases holding area-level information on environmental exposures and the National Pupil Database. For each of these, linkage poses different ethical and practical challenges and strict data protection guidelines are followed.

Common to many registries, in contrast to inception cohorts, survival times of individuals in the UK dataset are subject to left truncation because the Registry captures the living population at the inception of the Registry and incident cases subsequent to this. This leads to potential survivor bias, whereby the living population at the outset of the Registry represented healthier individuals from their respective birth cohorts who have survived to the point of being included in the dataset. This is a common issue in registry analyses. There are strong cohort effects in the data, which are likely to represent a mixture of survivorship effects, and the 'true' cohort effects representing improving treatment over time, as demonstrated in other studies.[7,19] With longer follow-up, as the UK CF Registry matures, separating age and cohort effects will become possible and eventually it will be possible to analyse incident individuals alone, ensuring that the longitudinal experience of all individuals from a particular birth cohort will be captured.[18] Making projections about outcomes for people with CF in the future will always be a challenge, in particular because there have been and continue to be substantial improvements in treatment and care over a relatively short time period.

Data resource access

All the information in the UK CF Registry is held confidentially and available only to two assigned members of the UK CF Registry data management team. Clinical teams can access only the data that relate to patients in their care, and are required to validate their identity using two-factor authentication before logging into the system. The system also provides an audit trail for any data access. The CF Registry is registered under the Data Protection Act (1998) which was designed to provide a legal framework upon

which to protect the privacy of personal data when used with information technology.

Data are stored within secure Microsoft Azure data centres in Ireland and Holland, which are certified as suitable for official government data, including health care data. The database servers are not directly connected to the internet; all administrative access is via a secure VPN connection which requires a dedicated security certificate to allow access. The application hosting environment has been security-tested by certified security consultants. All data are viewed and entered via a secure HTTPS web portal.

NHS Research Ethics approval (Huntingdon Research Ethics Committee 07/Q0104/2) has been granted for the collection of data into the UK CF Registry. Each patient or their parent provided written informed consent for collection of data in the registry as outlined above, and this includes for use of pseudonymized data in research. There is a formal process for requesting access to the UK CF registry, and an application form can be found using this link: [www.cysticfibrosis.org.uk/registry]. In 2016 there were 29 applications for access to anonymized Registry data; 26 were approved after review by the UK CF Registry Research Committee, with the remaining three requests being withdrawn or rejected.

Profile in a nutshell

- The UK Cystic Fibrosis Registry is a national, secure, centralized database sponsored and managed by the Cystic Fibrosis Trust. It was set up to record longitudinal health data on people with cystic fibrosis (CF) in England, Wales, Scotland and Northern Ireland. Containing data on almost all people with CF in the UK, it is one of the largest and most complete national CF databases available.
- First established in 1995, to date the UK CF Registry has captured data on over 12 000 individuals with a combined total of more than 100 000 annual assessments.
- Patient data are recorded in the Registry after a confirmed CF diagnosis and consent for data to be collected has been obtained. Subsequently, health data are added to the Registry at every annual assessment.
- Data are collected in several key areas including: demographics (including genotyping andand diagnosis data), hospital admissions and intravenous therapies, pulmonary function, chronic medications, culture andand microbiology, health complications, nutritional assessment, physiotherapy, lifestyle and outcomes (death and transplants).
- There is a formal process for requesting access to the UK CF Registry. An application form and more details can be found using the following link: [https://www.cysticfibrosis.org.uk/the-work-we-do/uk-cf-registry/apply-for-data-from-the-cf-registry].

Funding and competing interests

E.G. and R.C. work for the CF Trust; all other authors are part of the CF-EpiNet Strategic Research Centre supported by a grant from the Cystic Fibrosis Trust. S.C. is a principal investigator (PI) for a CF Trust Registry-based pharmacovigilance (PASS) study for Pharmaxis, and is a PI for a CF Trust Vertex-funded Registry-based observational study. D.B. is a principal investigator for two CF Trust Registry-based pharmacovigilance studies for Vertex and Teva pharmaceutical companies. DTR is funded by the MRC on a Clinician Scientist Fellowship (MR/P008577/1).

Acknowledgements

We thank people with cystic fibrosis and their families for consenting to their data being held in the UK CF Registry, and NHS teams in CF centres and clinics for the input of data into the Registry. We also thank the UK Cystic Fibrosis Trust and the Registry Steering Committee for access to anonymized UK CF Registry data.

References

1. Cystic Fibrosis Trust. *UK Cystic Fibrosis Registry 2015 Annual Data Report*. London: CFT, 2015.
2. Dodge JA, Lewis PA, Stanton M, Wilsher J. Cystic fibrosis mortality and survival in the UK: 1947–2003. *Eur Respir J* 2007;**29:** 522–26.
3. Elborn JS, Shale DJ, Britton JR. Cystic fibrosis: current survival and population estimates to the year 2000. *Thorax* 199;**46:** 881–85 .
4. Mehta G, Sims EJ, Culross F, McCormick JD, Mehta A. Potential benefits of the UK Cystic Fibrosis Database. *J R Soc Med* 2004;**97(Suppl 44):**60–71.
5. Cystic Fibrosis Trust. *Apply for Data From the UK CF Registry*. 2017. https://www.cysticfibrosis.org.uk/the-work-we-do/uk-cf-registry/apply-for-data-from-the-cf-registry (11 September 2017, date last accessed).
6. European Cystic Fibrosis Society. *ECFS Patient Registry Annual Reports*. 2017. https://www.ecfs.eu/projects/ecfs-patient-registry/annual-reports (11 September 2017, date last accessed).
7. Taylor-Robinson D, Smyth RL, Diggle P, Whitehead M. The effect of social deprivation on clinical outcomes and the use of treatments in the UK cystic fibrosis population: A longitudinal study. *Lancet Respir Med* 2013;**1:**121–28.
8. CF Start: A Randomised Registry Trial. 2017. http://www.cfstart.org.uk (11 September 2017, date last accessed).
9. Goss CH, MacNeill SJ, Quinton HB *et al.* Children and young adults with CF in the USA have better lung function compared with the UK. *Thorax* 2015;**70:**229–36.
10. Adler AI, Shine BSF, Chamnan P, Haworth CS, Bilton D. Genetic determinants and epidemiology of cystic fibrosis-related diabetes: results from a British cohort of children and adults. *Diabetes Care* 2008;**31:**1789–94.

International Journal of Epidemiology, 2018, Vol. 47, No. 1 **10e**

11. Sims EJ, Mugford M, Clark A *et al*. Economic implications of newborn screening for cystic fibrosis: a cost of illness retrospective cohort study. *Lancet* 2007;**369:**1187–95.
12. Sims EJ, Clark A, McCormick J *et al*. Cystic fibrosis diagnosed after 2 months of age leads to worse outcomes and requires more therapy. *Pediatrics* 2007;**119:**19–28.
13. Sims EJ, McCormick J, Mehta G, Mehta A; UK CF Database Steering Committee. Newborn screening for cystic fibrosis is associated with reduced treatment intensity. *J Pediatr* 2005;**147:** 306–11.
14. McCormick J, Sims EJ, Green MW, Mehta G, Culross F, Mehta A. Comparative analysis of Cystic Fibrosis Registry data from the UK with USA, France and Australasia. *J Cyst Fibros Off J Eur Cyst Fibros Soc*. 2005 May;**4**(2):115–22.
15. Stanojevic S, Stocks J, Bountziouka V, Aurora P, Kirkby J, Bourke S *et al*. The impact of switching to the new global lung function initiative equations on spirometry results in the UK CF registry. *J Cyst Fibros* 2014;**13:**319–27.
16. Stanojevic S, Bilton D, McDonald A *et al*. Global Lung Function Initiative equations improve interpretation of FEV1 decline among patients with cystic fibrosis. *Eur Respir J* 2015;**46:** 262–64.
17. Salvatore D, Buzzetti R, Baldo E *et al*. An overview of international literature from cystic fibrosis registries. Part 4: update 2011. *J Cyst Fibros* 2012;**11:**480–93.
18. Gliklich RE, Dreyer NA, Leavy MB (eds). *Registries for Evaluating Patient Outcomes: A User's Guide]*. 3rd edn. Rockville (MD): Agency for Healthcare Research and Quality (US), 2014.
19. Buzzetti R, Salvatore D, Baldo E *et al*. An overview of international literature from cystic fibrosis registries: 1. Mortality and survival studies in cystic fibrosis. *J Cyst Fibros* 2009;**8:**229–37.

Appendix B

Paper 2 - April 2018

This appendix contains the published version of the paper entitled "Estimating long-term treatment effects in observational data: a comparison of the performance of different methods under real-world uncertainty". It was published in Statistics in Medicine in April 2018.[11] This paper is based on the work found in Chapters 4 and 5.

As the first author of this paper, I carried out the simulation studies and analysis of the UK CF registry data, drafted all sections of the paper, produced all figures and tables, and prepared and approved the final version of the paper.

Student's Signature: 

Date: 29th January, 2019

Supervisor's Signature: 

Date: 29th January, 2019

Received: 20 September 2017 | Revised: 19 January 2018 | Accepted: 26 February 2018

DOI: 10.1002/sim.7664

RESEARCH ARTICLE

WILEY Statistics in Medicine

Estimating long-term treatment effects in observational data: A comparison of the performance of different methods under real-world uncertainty

Simon J. Newsome[1] | Ruth H. Keogh[1] | Rhian M. Daniel[2]

[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

[2]Division of Population Medicine, Cardiff University, Cardiff, UK

Correspondence
Simon Newsome, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.
Email: Simon.Newsome@lshtm.ac.uk

Funding information
Medical Research Council Methodology Fellowship, Grant/Award Number: MR/M014827/1; Wellcome Trust and the Royal Society, Grant/Award Number: 107617/Z/15/Z; Cystic Fibrosis Trust

In the presence of time-dependent confounding, there are several methods available to estimate treatment effects. With correctly specified models and appropriate structural assumptions, any of these methods could provide consistent effect estimates, but with real-world data, all models will be misspecified and it is difficult to know if assumptions are violated.

In this paper, we investigate five methods: inverse probability weighting of marginal structural models, history-adjusted marginal structural models, sequential conditional mean models, g-computation formula, and g-estimation of structural nested models. This work is motivated by an investigation of the effects of treatments in cystic fibrosis using the UK Cystic Fibrosis Registry data focussing on two outcomes: lung function (continuous outcome) and annual number of days receiving intravenous antibiotics (count outcome). We identified five features of this data that may affect the performance of the methods: misspecification of the causal null, long-term treatment effects, effect modification by time-varying covariates, misspecification of the direction of causal pathways, and censoring.

In simulation studies, under ideal settings, all five methods provide consistent estimates of the treatment effect with little difference between methods. However, all methods performed poorly under some settings, highlighting the importance of using appropriate methods based on the data available. Furthermore, with the count outcome, the issue of non-collapsibility makes comparison between methods delivering marginal and conditional effects difficult. In many situations, we would recommend using more than one of the available methods for analysis, as if the effect estimates are very different, this would indicate potential issues with the analyses.

KEYWORDS
causal inference, g-computation formula, g-estimation, inverse probability weighting, time-dependent confounding

1 | INTRODUCTION

Advanced methods for causal inference in longitudinal observational studies are an important tool for investigating treatment effects in nontrial settings where the presence of time-dependent confounders generally precludes the use of simpler conventional methods. Time-dependent confounding is an issue in longitudinal studies when a time-varying covariate is affected by treatment, but this covariate then also subsequently affects the probability of receiving future treatment as well as affecting the outcome of interest.[1]

In these situations, there are a number of methods available to researchers, and one of these methods, in particular, inverse probability weighting (IPW) of marginal structural models (MSM), has become increasingly popular in applied research. The increasing use of this method over other methods may in part be due to its relative simplicity, but it is not clear if other methods may be better suited to some analyses. The methods investigated in this paper are motivated by questions about the efficacy of long-term treatment use in cystic fibrosis (CF) and the challenges for addressing these using longitudinal observational data from a patient registry. In addition to IPW of MSM, we identified four other methods, which could be used in this setting: history-adjusted marginal structural models (HA-MSM), sequential conditional mean models (SCMM), g-computation formula, and g-estimation of structural nested models (SNM).

The primary aims of this paper are twofold: (1) to compare the ability and appropriateness of the different analysis methods for addressing the questions of interest, including to estimate treatment effect modification and to handle different outcome types and (2) to investigate the robustness of the different methods to handling practical challenges arising in longitudinal observational data, such as uncertainty about the relative temporality of measures and loss to follow-up. In an ideal setting, where we could be sure that all assumptions are met and that all models are correctly specified, any one of the available methods could be used to obtain consistent treatment effect estimates. However, in reality, it can be difficult to know if assumptions hold and all models will be misspecified to some degree.

Section 2 of this paper gives more details about the UK CF Registry, the questions we wish to address, and specific details of the data that present challenges. This is followed in Section 3 by an overview of the five different methods, which we considered for the analysis of the Registry data. In Section 4, we present simulation studies investigating the performance of these five methods with two different types of outcomes (normally distributed continuous data and zero-inflated negative binomial count data). An analysis of the UK CF registry data is presented in Section 5, and finally, we discuss the implications of the results of the simulation studies and data analysis in Section 6.

2 | MOTIVATING EXAMPLE

2.1 | CF and the UK Cystic Fibrosis Registry

Cystic fibrosis is the most common life-threatening inherited disease in white people, and in the UK, there are over 10 000 people living with the disease.[2,3] Cystic fibrosis most seriously affects the lungs, where a build-up in mucus causes breathing difficulties and leads to an increase in respiratory infections. There are now many treatments available that can help improve the health of people with CF, but many of these treatments are very time consuming and often treatments are not stopped once started. This leads to an accumulation of treatments, and treatment burden is a common complaint among people with CF.[4]

Almost all treatments currently used in CF care were approved following a successful clinical trial. However, a limitation of many trials is that they are short in duration, whereas in practice, treatments are used long term. In most cases, it would not be feasible to run trials for such long periods of times, and it could also be unethical to continue to withhold treatment from patients if a strong short-term benefit has been observed.

The UK CF Registry is a national database, which has collected annual data on almost all people with CF in the UK since 2007. At an annual assessment, detailed information is obtained on many different measures of health status as well as all the treatments received in the past year.[5]

This paper will focus on one common CF treatment, dornase alfa (DNase), which was licensed for use in the UK in 1994 after a randomised trial showed efficacy at improving lung function over a 6-month period.[6] Subsequent studies have investigated the effects of up to 2 years' use of DNase, but in practice, patients generally continue to receive DNase indefinitely once treatment has been started.[7]

To illustrate the potential of the statistical methods with different types of outcomes, this paper will consider the effects of treatment on two important clinical outcomes: percent predicted forced expiratory volume in 1 second ($ppFEV_1$) (a continuous outcome measuring an individual's lung function) and annual number of days of intravenous

antibiotic therapy (IV days) (a count outcome of the number of days an individual received intravenous antibiotics in a given year). The decision to start prescribing DNase to a patient depends on many factors, including pretreatment measures of $ppFEV_1$ and annual IV days, which are then in turn potentially affected by treatment use.

Figure 1 shows a directed acyclic graph (DAG) of the assumed causal pathways between key variables in the UK CF Registry. Data are obtained at annual visits. Treatment status at visit t is denoted X_t, F_t denotes $ppFEV_1$, and V_t denotes IV days. We focus on patients not using treatment at a baseline visit 0, $X_0 = 0$. At subsequent annual visits, their $ppFEV_1$ on that day is recorded, and data about the previous year are also collected, such as which treatments they received throughout the year and their total number of IV days.

The DAG visualises $ppFEV_1$ at visit $t-1$ affecting treatment at visit t, annual IV days at visit t affecting treatment at visit t, DNase use at visit t affecting all future measures of $ppFEV_1$ and annual IV days, and direct longitudinal associations between $ppFEV_1$ and annual IV days. There may also be other important baseline or time-varying confounders, which have not been included in the DAG for clarity. In this paper, we focus on investigating the effect of treatment on lung function and IV days and how this effect might change with continued use of treatment over several years.

2.2 | Features and challenges in the analysis of this data

As is common with most observational data, there are a number of issues that need to be considered when approaching the analysis of the UK CF Registry data.

One challenge is the use of the available methods with different types of outcomes. One of our outcomes of interest, $ppFEV_1$, is continuous and can be approximated by a conditionally normal distribution. All of the methods described in this paper can easily accommodate such an outcome. However, the other outcome, annual IV days, is a count outcome ranging from 0 to 365, which we model with a zero-inflated negative binomial distribution. This can be harder to incorporate into some of the methods, and we will discuss these issues in Section 3.7.

In addition to considering two different types of outcome, we have identified 5 key features of the analysis of the UK CF registry. The first three of these question the ability and appropriateness of the different analysis methods for estimating the treatment effect: whether there exists any treatment effect at all, whether there are only short-term or also long-term effects, and whether there is effect modification of the treatment effect by time-varying covariates. The second category are challenges that may arise because of the nature of the data available to investigate the above questions. Here, we consider the issues of censoring and uncertainty of the direction of causal pathways between variables.

The following subsections give further details on each of these five issues.

2.2.1 | Causal null hypothesis

The DAG shown in Figure 1 shows a causal effect of treatment on the outcomes of interest. To date, randomised trials have demonstrated the efficacy of DNase treatment in improving $ppFEV_1$, but no studies have yet shown a significant effect of the treatment on reducing the rate of IV days. Furthermore, in a nontrial setting, where, for example, adherence levels may not be as high as in clinical trials, the findings of a causal effect of treatment may not be replicated. For this reason, methods that benefit from a degree of robustness to model misspecification at the causal null would be attractive.

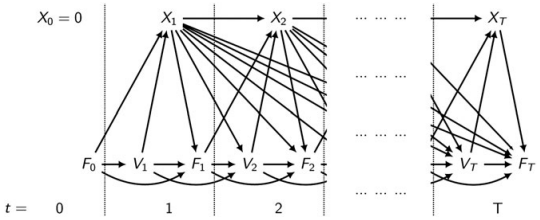


FIGURE 1 Directed acyclic graph of causal pathways between treatment (X_t), $ppFEV_1$ (F_t), and IV days (V_t). Baseline and other time-varying confounders are not shown for clarity

2.2.2 | Long-term treatment effects

We define a long-term treatment effect as an effect of X_t on Y_s ($s > t$) not mediated via intermediate treatments. No studies have previously looked at the effects of DNase beyond 2 years, and it therefore remains unknown how the effect of treatment might change with length of use. Taking the example of $ppFEV_1$, two possible ways in which the treatment may affect the outcome are (1) the $ppFEV_1$ trajectories of those receiving and not receiving treatment continue to grow apart indefinitely through time or (2) after the initial increase in $ppFEV_1$ that has been observed at the start of taking treatment, the effectiveness of treatment may decrease with the two counterfactual trajectories no longer diverging. These two hypothetical lung function trajectories compared with the trajectory when not receiving treatment are shown in Figure 2. As it is unknown how the effect of treatment might change through time, it is important that the methods are flexible enough to identify the true long-term effects.

2.2.3 | Effect modification by time-varying covariates

We hypothesise that the effect of treatment may depend on the previous levels of $ppFEV_1$ and number of IV days. This is because if a person starts treatment when they already have a healthy $ppFEV_1$ level, it is unlikely that treatment could further improve $ppFEV_1$, whereas it is realistic that the treatment could be much more effective in an individual with an lower $ppFEV_1$. For informing practice, rather than just identifying the population average effect of treatment, it is important to gain understanding of how the effect of treatment might change depending on other covariates, and for this reason, it would be preferable to use a method that can test for the presence and estimate the strength of any effect modification.

2.2.4 | Misspecification of the direction of causal pathways

The DAG in Figure 1 includes assumptions about the direction of the causal pathways. For some variables, the appropriate direction of the causal pathway is clear (eg, the pathway from total IV days in 1 year to $ppFEV_1$ measured at the end of the year). However, for other pathways, the appropriate direction for the arrow is less clear.

The direction of the causal pathway between treatment and number of IV days is particularly challenging. Both variables are summaries of the previous year, and some individuals may have had lots of IV days at the start of the year, which prompted them to start treatment, whereas others may have started treatment earlier, but then had IV days later in the year. In reality, therefore, the causal pathway between X_t and V_t is likely to go both ways, but in many methods, it will be necessary to specify just one direction for this pathway.

We have decided to focus on investigating the effect of X_t on V_{t+1}, as, due to temporality, this pathway can only be directed this way, and to treat V_t as a confounder of this effect. In the real Registry data, we cannot know whether this is misspecified or not, and therefore, it is important to understand the potential extent of the bias in treatment effect estimates under different methods when the direction of this pathway has been misspecified.

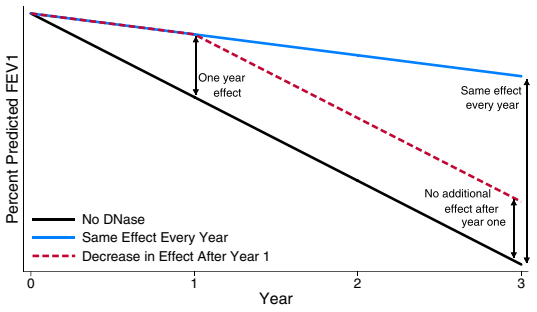


FIGURE 2 Two possible trajectories of lung function with long-term dornase alfa treatment [Colour figure can be viewed at wileyonlinelibrary.com]

2.2.5 | Censoring

We are fortunate that there are very few people lost to follow-up in the UK CF Registry, and each year, there are relatively few deaths compared with the total number of people in the Registry. Nevertheless, it is possible that the fact that some individuals are censored for either of these reasons may bias the results. Therefore, we also wish to investigate how the different methods handle censoring. Although in reality there would likely be different processes affecting the probability that an individual dies or is lost to follow-up, in this paper, we only consider one missing at random scenario where an individual's probability of being censored depends on previously measured variables.

3 | METHODS

3.1 | Notation and assumptions

We discuss the statistical methods with generic notation. Consider a cohort followed up annually from visit $t=0$ up to visit $t=T$. The treatment received at time t is denoted X_t, and each year, a person can receive ($X_t=1$) or not receive ($X_t=0$) treatment during the period since the last visit. The outcome of interest, Y_t, is also measured annually. We assume that at each visit, X_t precedes Y_t and define a 1-year treatment effect to be the effect of X_t on Y_t. We also have baseline confounders, $\mathbf{B}$, and time-varying confounders, $\mathbf{C}$.

$\bar{X}_t$ is a vector of the treatment history for an individual from visit 0 up to and including visit t, and we use the counterfactual notation $Y_t^{\bar{x}_t=\bar{1}}$ to refer to the outcome that would have been observed at visit t if an individual had received treatment up to and including visit t.

For all methods, we make the following four assumptions: no interference, positivity, consistency, and no unmeasured confounding. No interference means that for a given individual, their counterfactual outcome $Y_t^{\bar{x}}$ is not affected by the treatment that another individual receives.[8] Positivity means that all individuals had a conditional probability strictly greater than 0 and strictly less than 1 of receiving treatment at all visits given their history, $0 < P(X_t = 1|\bar{X}_{t-1}, \bar{Y}_{t-1}, \bar{\mathbf{C}}_t, \mathbf{B}) < 1$.[9] Consistency means that for each individual, the counterfactual outcome under the observed treatment is equal to the observed outcome, $Y_i = Y_i^{x_i}$ when $x_i = X_i$.[10] Finally, no unmeasured confounding means that conditional on the past observed variables the treatment received at visit t is independent of the counterfactual outcome, $Y_t^{\bar{x}_t} \perp\!\!\!\perp X_t|\bar{X}_{t-1}, \bar{Y}_{t-1}, \bar{\mathbf{C}}_t, \mathbf{B}$. [1]

The following subsections give an overview of the methods that are considered for the analysis of the UK CF Registry. We introduce the methods with a continuous outcome in mind. Referring back to our motivating example and the DAG in Figure 1, we can consider $\text{ppFEV}_1(F_t)$ to be the outcome of interest with IV days (V_t) acting as a time-dependent confounder. The count variable of IV days is also of interest as an outcome, and in Section 3.7, we outline how the methods can be extended for use with a count outcome.

3.2 | IPW of marginal structural models

Inverse probability weighting of MSM[11] has become an increasingly popular method to deal with time-dependent confounding. We consider MSM of the following form:

$$\mathrm{E}\left(Y_t^{\bar{x}_t}\right) = \beta_0 + \sum_{i=1}^{t} \beta_{x_i} x_i, \tag{1}$$

where the β_{x_1} to β_{x_t} represent separate effects for treatment at each visit, thereby allowing for long-term treatment effects. However, due to confounding, directly using the observed values and calculating $\mathrm{E}[Y_t|\bar{X}_t = \bar{x}_t]$ does not equate to the counterfactual $E[Y_t^{\bar{x}_t}]$.

Inverse probability weighting of the observations enables consistent estimation from an MSM by reweighting observations so that the levels of confounding variables become equally balanced between treated and untreated individuals. This is achieved by assigning large weights to individuals who were estimated to be unlikely to receive the treatment they actually received and downweighting observations for which there are lots of observations estimated to have similar propensities to receive the same treatment history.

To calculate the weights, one first estimates the propensity score, which is the probability of receiving treatment at each visit:

$$\mathrm{P}(X_t = 1|\bar{X}_{t-1}, \bar{Y}_{t-1}, \bar{\mathbf{C}}_t, \mathbf{B}) = \mathrm{expit}(\beta_0 + \beta_X X_{t-1} + \beta_Y Y_{t-1} + \boldsymbol{\beta}_{\mathbf{C}}\mathbf{C}_t + \boldsymbol{\beta}_{\mathbf{B}}\mathbf{B}). \quad (2)$$

Then this model is used to calculate the estimated probability that each person received the treatment they actually received, ie, for those who did receive treatment, we use the estimated probability from the above model and for those who did not receive treatment, 1 minus the estimated probability. The probability of their treatment history is then the product of these estimated probabilities from visit 1 up to visit t.

The inverse of the estimated probabilities can be used directly as the weights, but it is usually preferable to use so-called stabilised weights[1] where the numerator of the weights is the probability of receiving treatment based on previous treatment history and baseline covariates only,

$$SW_t = \frac{\mathrm{P}(\bar{X}_t|\bar{X}_{t-1}, \mathbf{B})}{\mathrm{P}(\bar{X}_t|\bar{X}_{t-1}, \bar{Y}_{t-1}, \bar{\mathbf{C}}_t, \mathbf{B})} = \prod_{i=1}^{t} \frac{\mathrm{P}(X_i|\bar{X}_{i-1}, \mathbf{B})}{\mathrm{P}(X_i|\bar{X}_{i-1}, \bar{Y}_{i-1}, \bar{\mathbf{C}}_i, \mathbf{B})}. \quad (3)$$

A final MSM, such as that given in Equation 1, can then be fit where the observations are weighted using the estimated weights. However, note that any baseline confounders included in the numerator of Equation 3 must also be included in the MSM. This would result in a conditional estimate, meaning if a marginal estimate is desired then no confounders should be included in the numerator.

Due to the fact that time-varying covariates are not included in the MSM, this method does not allow for the estimation of effect modification by time-varying covariates. However, the method also does not need the assumption that there is no effect modification and will estimate consistent population average effects even if the effect of treatment is not uniform for the whole population.

Using stabilised weights helps to reduce the variability in the weights, but in cases where there are strong time-varying predictors of treatment, the weights can remain highly variable that can lead to instability. Therefore, it can sometimes be preferable to truncate the most extreme weights, even though this may introduce some bias.[12,13] In this paper, we will present the results of IPW analyses with and without truncation of the stabilised weights to the 1st and 99th percentile.

In the presence of censoring, it is also possible to incorporate censoring weights into the analysis. Similarly to the previously described weights, we weight individuals with stabilised inverse weights of their estimated probability of being censored before visit t,

$$\mathrm{LTFU}W_t = \prod_{i=1}^{t} \frac{\mathrm{P}(\mathrm{LTFU}_i)}{\mathrm{P}(\mathrm{LTFU}_i|\bar{X}_{i-1}, \bar{Y}_{i-1}, \bar{\mathbf{C}}_{i-1}, \mathbf{B})}. \quad (4)$$

Using this method, we assume that future visits are missing at random, ie, censoring is affected by previously measured variables. The estimated censoring weights can then be multiplied by the estimated stabilised weights to give the weights to be used to account for bias due to both confounding and censoring.

3.3 | History-adjusted marginal structural models

As stated in the previous section, one limitation of IPW of MSM is that effect modification of the treatment effect by time-varying covariates cannot be estimated. Therefore, in cases where the estimation of an interaction term is desired, HA-MSM are an extension to IPW of MSM, which do allow for this.[14]

In the standard MSM described in the Section 3.2, observations are reweighted based on all covariates measured after baseline until the visit of the outcome of interest. In an HA-MSM, the reweighting is done separately from each time of treatment t up to the time of outcome, s ($t \leq s$). Covariates measured prior to the treatment at time t can be included in the final HA-MSM in the same way as baseline covariates were included in the standard MSM.

Formally, the stabilised weights for exposure at time t on outcome at time s are given by

$$SW_{ts} = \prod_{i=t}^{s} \frac{\mathrm{P}(X_i|\bar{X}_{i-1}, \mathbf{B})}{\mathrm{P}(X_i|\bar{X}_{i-1}, \bar{Y}_{i-1}, \bar{\mathbf{C}}_i, \mathbf{B})}, \quad (5)$$

and an example of the HA-MSM could be given by

$$\mathrm{E}\left(Y_s^{\bar{x}_s}|\bar{x}_{t-1},\bar{y}_{t-1},\bar{\mathbf{c}}_t,\mathbf{b}\right)=\beta_0+\boldsymbol{\beta}_{\mathbf{b}}\mathbf{b}+\boldsymbol{\beta}_{\mathbf{c}}\mathbf{c}_t+\beta_x x_{t-1}+\beta_y y_{t-1}+\sum_{i=t}^{s}\beta_{x_i}x_i+\sum_{i=t}^{s}\beta_{\mathrm{int}_i}x_i y_{t-1}. \tag{6}$$

In Equation 6, we have included an interaction term between previous measures of the outcome (itself a time-varying confounder) and treatment so as to allow the estimation of any effect modification.

As with IPW of MSM, it is also possible to estimate censoring weights, in this case estimating an individual's probability of being censored between visits t and s and multiplying these weights with the stabilised weights.

3.4 | Sequential conditional mean models

Even in the presence of time-dependent confounding, it is still possible to use standard regression methods, but these methods can only estimate total effects.[15] The total effect of a treatment, X_t, on an outcome Y_s ($s>t$) would include not only the direct effect of X_t on Y_s and the indirect effects of X_t on Y_s through time-varying covariates but also the indirect effect of X_t on Y_s mediated through future exposures. It cannot, therefore, be used to investigate the effect of receiving 2 years' treatment in our example, as some people discontinue treatment. For this reason, in the examples here, this method is only used to estimate "short-term" effects, which we define as the effect of 1-year treatment on the outcome measured at the end of the year.

These SCMM will give a consistent estimate of the 1-year effect of treatment as long as we appropriately control for all confounding effects of this short-term effect. For example, the following short-term model would suffice if the most recent measures of all covariates were sufficient to remove confounding:

$$\mathrm{E}(Y_t|\bar{X}_t,\bar{Y}_{t-1},\bar{\mathbf{C}}_t,\mathbf{B})=\beta_0+\beta_{X_1}X_t+\beta_{X_2}X_{t-1}+\beta_Y Y_{t-1}+\boldsymbol{\beta}_{\mathbf{C}}\mathbf{C}_t+\boldsymbol{\beta}_{\mathbf{B}}\mathbf{B}. \tag{7}$$

It is also possible to incorporate propensity scores into the SCMM to provide a doubly robust estimator. The propensity score can be calculated as it was in the IPW method by using Equation 2, and this is then incorporated into the SCMM as follows:

$$\mathrm{E}(Y_t|\bar{X}_t,\bar{Y}_{t-1},\bar{\mathbf{C}}_t,\mathbf{B},p_t)=\beta_0+\beta_{X_1}X_t+\beta_{X_2}X_{t-1}+\beta_Y Y_{t-1}+\boldsymbol{\beta}_{\mathbf{C}}\mathbf{C}_t+\boldsymbol{\beta}_{\mathbf{B}}\mathbf{B}+\beta_p p_t. \tag{8}$$

Although this method cannot provide estimates for the effects of varying lengths of treatment duration, the simplicity of the method is appealing, and these short-term effect estimates can also be compared with the 1-year treatment effect estimates from the other methods. SCMM also form the first step of the next 2 methods: g-computation formula and g-estimation of SNM.

3.5 | G-computation formula

The g-computation formula first described by Robins[16] is another method that can deal with the issue of time-dependent confounding to give consistent estimates of long-term treatment effects. In this method, short-term models, ie, models for 1-year time effects, for all time-varying covariates (in our example, Y and $\mathbf{C}$) are used to simulate counterfactual outcomes under different treatment trajectories sequentially through time.

For example, the time-varying continuous outcome Y could be modelled by Equation 7 and counterfactuals for Y_1 could then be simulated setting everyone either receiving or not receiving treatment at visit 1:

$$\tilde{Y}_1^{x_1=1}=\hat{\beta}_0+\hat{\beta}_Y Y_0+\hat{\boldsymbol{\beta}}_{\mathbf{C}}\mathbf{C}_1+\hat{\boldsymbol{\beta}}_{\mathbf{B}}\mathbf{B}+\hat{\beta}_{X_1}+\tilde{\varepsilon}, \tag{9}$$

$$\tilde{Y}_1^{x_1=0}=\hat{\beta}_0+\hat{\beta}_Y Y_0+\hat{\boldsymbol{\beta}}_{\mathbf{C}}\mathbf{C}_1+\hat{\boldsymbol{\beta}}_{\mathbf{B}}\mathbf{B}+\tilde{\varepsilon}, \tag{10}$$

where $\tilde{\varepsilon}$ is a random draw from a normal distribution whose standard deviation is the model-estimated root mean square error, resulting in simulated counterfactual measures.

Similar short-term models would need to be specified for all time-varying covariates $\mathbf{C}$ to allow the counterfactuals for all covariates to be simulated at visit 1. In our example, we have one time-varying confounder, which follows a zero-

inflated negative binomial distribution. Therefore, this was used to model the data and then to simulate random draws for the count for each individual.

The process can then be repeated sequentially for all visits. For example, at visit 2, there would be four counterfactuals simulated for each individual, corresponding to (1) receiving treatment at both visits, (2) at the first visit only, (3) at the second visit only, or (4) never receiving treatment. These counterfactuals could be simulated, respectively, as follows:

$$\tilde{Y}_2^{x_1=1,x_2=1} = \hat{\beta}_0 + \hat{\beta}_Y \tilde{Y}_1^{x_1=1} + \hat{\boldsymbol{\beta}}_{\mathbf{C}} \tilde{\mathbf{C}}_2^{x_1=1} + \hat{\boldsymbol{\beta}}_{\mathbf{B}} \mathbf{B} + \hat{\beta}_{X_1} + \hat{\beta}_{X_2} + \tilde{\varepsilon}, \tag{11}$$

$$\tilde{Y}_2^{x_1=1,x_2=0} = \hat{\beta}_0 + \hat{\beta}_Y \tilde{Y}_1^{x_1=1} + \hat{\boldsymbol{\beta}}_{\mathbf{C}} \tilde{\mathbf{C}}_2^{x_1=1} + \hat{\boldsymbol{\beta}}_{\mathbf{B}} \mathbf{B} + \hat{\beta}_{X_2} + \tilde{\varepsilon}, \tag{12}$$

$$\tilde{Y}_2^{x_1=0,x_2=1} = \hat{\beta}_0 + \hat{\beta}_Y \tilde{Y}_1^{x_1=0} + \hat{\boldsymbol{\beta}}_{\mathbf{C}} \tilde{\mathbf{C}}_2^{x_1=0} + \hat{\boldsymbol{\beta}}_{\mathbf{B}} \mathbf{B} + \hat{\beta}_{X_1} + \tilde{\varepsilon}, \tag{13}$$

$$\tilde{Y}_2^{x_1=0,x_2=0} = \hat{\beta}_0 + \hat{\beta}_Y \tilde{Y}_1^{x_1=0} + \hat{\boldsymbol{\beta}}_{\mathbf{C}} \tilde{\mathbf{C}}_2^{x_1=0} + \hat{\boldsymbol{\beta}}_{\mathbf{B}} \mathbf{B} + \tilde{\varepsilon}. \tag{14}$$

The counterfactual outcomes under different treatment trajectories can then be compared with a MSM, eg,

$$\mathrm{E}\left(Y_t^{\bar{x}_t}\right) = \beta_0 + \sum_{i=1}^{t} \beta_{x_i} x_i. \tag{15}$$

One well-known drawback of the use of this method with non-linear models is the g-null paradox.[16,17] This is an issue whereby given a large enough sample size, the causal null hypothesis will always be rejected even if there is in fact no treatment effect. This is due to the fact that the combination of different parametric models will be inconsistent with the null hypothesis.

3.6 | G-estimation of structural nested models

The final method we will consider is g-estimation of SNM.[18] This method has been used less than the previously described methods, and this may partly be due to the perceived difficulty of applying the method with standard statistical software.[19] However, a recent paper by Vansteelandt and Sjolander revisits g-estimation, showing how it can be implemented with standard software.[20]

Similar to HA-MSM, this method can estimate the effect of all treatments at visits t on outcomes at visits s where $t \leq s$. Starting from the short-term model as in the SCMM, we obtain an estimate for the 1-year effect of treatment, β_{X_1},

$$\mathrm{E}(Y_t | \bar{X}_t, \bar{Y}_{t-1}, \bar{\mathbf{C}}_t, \mathbf{B}, p_t) = \beta_0 + \beta_1 X_{t-1} + \beta_2 Y_{t-1} + \boldsymbol{\beta}_3 \mathbf{C}_t + \boldsymbol{\beta}_4 \mathbf{B} + \beta_5 p_t + \beta_{X_1} X_t, \tag{16}$$

where p_t is the estimated propensity score.

The estimate β_{X_1} can then be used to construct counterfactuals by subtracting the estimated 1-year effect to be able to see if there is any extra effect for additional years of treatment,

$$H_{st} = Y_s - \sum_{u=t+1}^{s} \beta_{X_{s-u+1}} X_u. \tag{17}$$

It can be seen that in the case where $t = s$, H_{st} is simply equal to Y_s as expected, whereas intermediate treatment effects are subtracted if $t < s$. In the first iteration, as we only have an estimate for β_{X_1}, we can only calculate H_{st} where $s \leq t+1$. However, this now allows us to estimate both the 1-year and 2-year effects with the following model:

$$\mathrm{E}\left(H_{sj} | \bar{X}_j, \bar{Y}_{j-1}, \bar{\mathbf{C}}_j, \mathbf{B}, p_j\right) = \beta_0 + \beta_1 X_{j-1} + \beta_2 Y_{j-1} + \boldsymbol{\beta}_3 \mathbf{C}_j + \boldsymbol{\beta}_4 \mathbf{B} + \beta_5 p_j + \beta_{X_{s-j+1}} X_j. \tag{18}$$

Iteration of Equations 17 and 18 allows the estimation of all $\beta_{X_{s-j+1}}$ where $1 \leq j \leq s$ and $1 \leq s \leq T$.

Similarly to HA-MSM, g-estimation is a method that allows the estimation of effect modification by time-varying covariates by including interaction terms in both Equations 17 and 18.

Censoring weights as described in Section 3.3 can also be incorporated into g-estimation, weighting individuals by their estimated probability of being censored between visits t and s.

3.7 | Use of methods with a count outcome

For our motivating example, we have 2 outcomes of interest, ppFEV_1 and annual IV days. The first of these is a continuous outcome, and all 5 of the methods can easily handle this outcome. More care is needed when considering annual IV days, which is a count outcome ranging from 0 to 365.

Upon investigation, IV days can be considered approximately distributed by a zero-inflated negative binomial distribution. Modelling this outcome therefore requires two separate estimation procedures: (1) logistic regression to estimate the odds of a count of zero IV days and (2) negative binomial regression to estimate the rate of IV days. Therefore, there are two separate parts to the treatment effect: the estimated effect of treatment on having zero IV days (an odds ratio) and the estimated effect of treatment on the number of IV days (a rate ratio).

For SCMM, this is not an issue, as one can simply fit a zero-inflated negative binomial model to estimate both effects. Similarly, IPW, HA-MSM, and g-computation formula can all handle different types of outcome by just changing the final MSM, eg,

$$\mathrm{E}\left(Y_t^{\bar{x}_t}\right) = \operatorname{expit}\left(\beta_0 + \sum_{i=1}^{t} \beta_{x_i} x_i\right) \exp\left(\gamma_0 + \sum_{i=1}^{t} \gamma_{x_i} x_i\right). \tag{19}$$

Unlike the other 4 methods, which can easily handle different types of outcome, the method of g-estimation described in Section 3.6 has until recently only been described for continuous outcomes. However, a recent paper has shown how this method can be adapted to allow for a count outcome by modelling with a gamma distribution.[21] This allows for the estimation of the effect of treatment on the rate of IV days, but would still not allow for the decomposition of the effect into the probability of a zero count and a rate, as the other methods do.

Another issue one needs to consider when modelling a count outcome is non-collapsibility. Unlike with the continuous outcome where, thanks to collapsibility, the marginal and conditional effects are the same, this no longer holds for models suitable for count outcomes. Thus, the treatment effect on IV days will differ between methods depending on whether the method delivers a marginal effect (IPW and g-computation formula) or a conditional effect (SCMM, HA-MSM and g-estimation).

3.8 | Overview of methods

Referring back to the five features of the UK CF Registry introduced in Sections 2.2.1 to 2.2.5, we would hope for any method to estimate no treatment effect on average when there is no treatment effect, but the g-null paradox may mean that the g-computation formula could perform poorly in this setting.

Except for SCMM, all methods can estimate long-term treatment effects, and in all our analyses, we will include separate terms for treatment at each visit making no assumptions about a continuous effect or a trend effect. SCMM will only be used to estimate the short-term treatment effect, but the method will consistently estimate this even if there are longer term effects.[15]

In terms of effect modification, three of the methods (SCMM, g-estimation, and HA-MSM) allow interaction terms, meaning that when this is of interest, only these methods can be used. IPW of MSM and g-computation formula can still be used to estimate population average effects even in the presence of effect modification by time-varying covariates, whereas the other three methods may show bias in estimating population average effects if there is in fact effect modification and it is not explicitly modelled.

When there is censoring, three of the methods (IPW of MSM, HA-MSM, and g-estimation) can use censoring weights to correct for the individuals who do not have full follow-up. Censoring should not affect the short-term models used in SCMM, and similarly, the g-computation formula uses the same short-term models and then simulates follow-up without censoring.

With the exception of SCMM, it is normally advised to use a bootstrap procedure to obtain standard errors (SE). This is because all the methods contain a number of steps of estimation and just using the final model-based SE would fail to

account for the uncertainty from the earlier steps.[13,14,22,23] The bootstrap provides valid results as all of these methods produce regular estimators.[24]

4 | SIMULATION STUDIES

The following section gives details of simulation studies that were performed to investigate how the features and challenges identified in Section 2.2 (robustness to misspecification of the causal null, long-term treatment effects, effect modification by time-varying covariates, misspecification of the direction of causal pathways, and censoring) affect the performance of the five methods given in Section 3 (SCMM, IPW of MSM, HA-MSM, g-computation formula, and g-estimation of SNM).

The aims of the simulation studies are to understand how the performance of the analysis methods might be affected by these challenges and to help provide a framework for the best analysis strategy for the real UK CF Registry data. The simulation studies were performed following the guidelines given by Burton et al[25] and full details of the design of the simulation studies can be found in Supporting Information, with a summary below.

4.1 | Design of simulation studies

Datasets were simulated for six different scenarios as shown in Figure 3. In each DAG, the arrows highlighted in red show the specific differences compared with the other scenarios.

The first scenario is the standard scenario, which will be the baseline with which to compare the other methods. In this scenario, there is a 1-year treatment effect, there are no direct long-term effects, although there are long-term effects mediated through other time-varying covariates. This is the scenario for which all the methods will be correctly specified and as such we would expect all methods to provide consistent estimates for the treatment effects in this scenario.

In the second scenario, we simulate without any treatment effect, and the third scenario adds long-term direct effects of treatment. In this case, the long-term direct effects are actually negative effects, slightly counteracting the beneficial 1-year effects, resulting in a decrease in the treatment effect through time.

The fourth scenario simulates effect modification of treatment by time-varying covariates. Although effect modification is generally not shown in DAGs, we have included arrows in Figure 3 to help illustrate how the presence of effect modification would change the treatment effect.

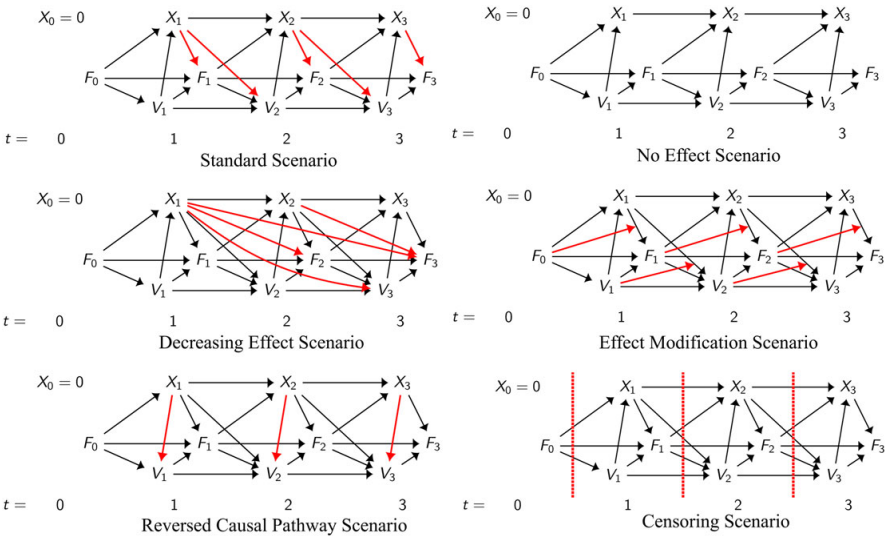


FIGURE 3 Simplified directed acyclic graphs showing data generation process for each of the 6 scenarios investigated. The real data were generated for up to 5 visits, and there is additionally a baseline confounder, age, affecting all variables

The fifth scenario concerns the direction of the causal pathway between treatment and IV days in the same year. In our analysis, we will always analyse the data as if the direction of the causal pathway is from V_t to X_t, even when the data have actually been simulated the other way around, ie, X_t affects V_t.

In the final scenario, individuals can either all be followed up for 5 visits, or there can be some censoring, whereby "unhealthy" individuals are more likely to be censored at an earlier visit. This corresponds to a missing at random scenario, whereby the probability of being censored depends on observed variables.

For each scenario, we simulated 1000 datasets, each with 7500 individuals. The data were generated so as to imitate the observed data in the Registry as closely as possible. In the real data, there are many treatments that individuals could be receiving and also many covariates that might be confounders. For the simulation studies, we kept just one binary treatment, X_t, and the 2 outcome variables, $\text{ppFEV}_1(F_t)$ and annual IV days (V_t), which also act as time-dependent confounders. Lung function was simulated as a continuous variable with a normal distribution and IV days as a count outcome following a zero-inflated negative binomial distribution. In addition to these 3 variables, we also generated age from a beta distribution corresponding to what was observed in the real data to act as a baseline confounder.

For each method and each scenario, we run two analyses: the first considering ppFEV_1 as the outcome and the second annual IV days as the outcome. For each simulation, the coefficients corresponding to the treatment effects will be stored. For SCMM, which can only measure short-term effects, only the coefficient corresponding to 1 year of treatment will be stored. For all other methods, we estimate the effects of up to 5 years' treatment use on ppFEV_1 and up to 4 years' treatment use on IV days. The reason for this difference is due to the 1-year effect of treatment on lung function being defined as $\bar{X}_t \rightarrow F_t$, whereas the 1-year effect of treatment on IV days is $\bar{X}_t \rightarrow V_{t+1}$. As such, there is always one extra year of data available for the lung function outcome.

We will compare the methods based on the bias, empirical SE, and mean squared error (MSE). Although it is known that the model-based SE are biased for most of these methods, we will also store the estimated robust SE so as to compare them to the empirical SE.

4.2 | Results of simulation studies

4.2.1 | Continuous outcome

In Figure 4, we present kernel density plots showing the results of the simulation studies for the normally distributed continuous outcome. We only present results for the 1-year effect and the 5-year effect to show the 2 extremes of short- to long-term effects. In all cases, the results for the 2- to 4-year effects followed the trend between the 1-year and 5-year effects. More details of the results can be found in Table S3.

As expected, all 5 methods to provide consistent estimates for the "standard" scenario where all the models are correctly specified. The only method that performs poorly here is using truncation with IPW, but this is also to be expected as it is known that due to truncation the weights would no longer fully account for confounding. The 5-year treatment effect estimates are slightly biased, but when using a much larger sample size, all methods were unbiased; therefore, we believe this residual bias is due to the sample size, which we have kept at 7500 individuals as it is unlikely that we would ever obtain a larger sample from the UK CF Registry.

These findings are repeated for the scenarios where there is no treatment effect, where the treatment effect decreases over time, and where there is censoring (provided that censoring weights are used for IPW, HA-MSM, and g-estimation).

For the scenario where the causal pathway between a confounder and treatment is specified the wrong way round, we find that the situation is the opposite: All methods are biased, but IPW and HA-MSM perform comparatively well when the weights are truncated. However, untruncated, they perform very poorly with very large variability and even fail to converge on an estimate many times.

When considering effect modification by time-varying covariates, all the methods can still be used to provide an estimate for the population-average effect. For the 1-year effects, all the methods provided consistent estimates; however, at 5 years, there was some noticeable bias for g-estimation and HA-MSM. These are the 2 methods that can incorporate the estimation of effect modification by time-varying covariates, and not including these interactions terms when they are in fact present has introduced bias. Conversely, although IPW and g-computation formula cannot estimate interaction terms, they do not assume that there is no effect modification and can provide consistent estimates for the population-average effect.

If the aim is to estimate the strength of any effect modification by time-varying covariates, then it would be necessary to use HA-MSM, SCMM, or g-estimation, and these results are presented in Figure 5 (and Table S4). We see that all

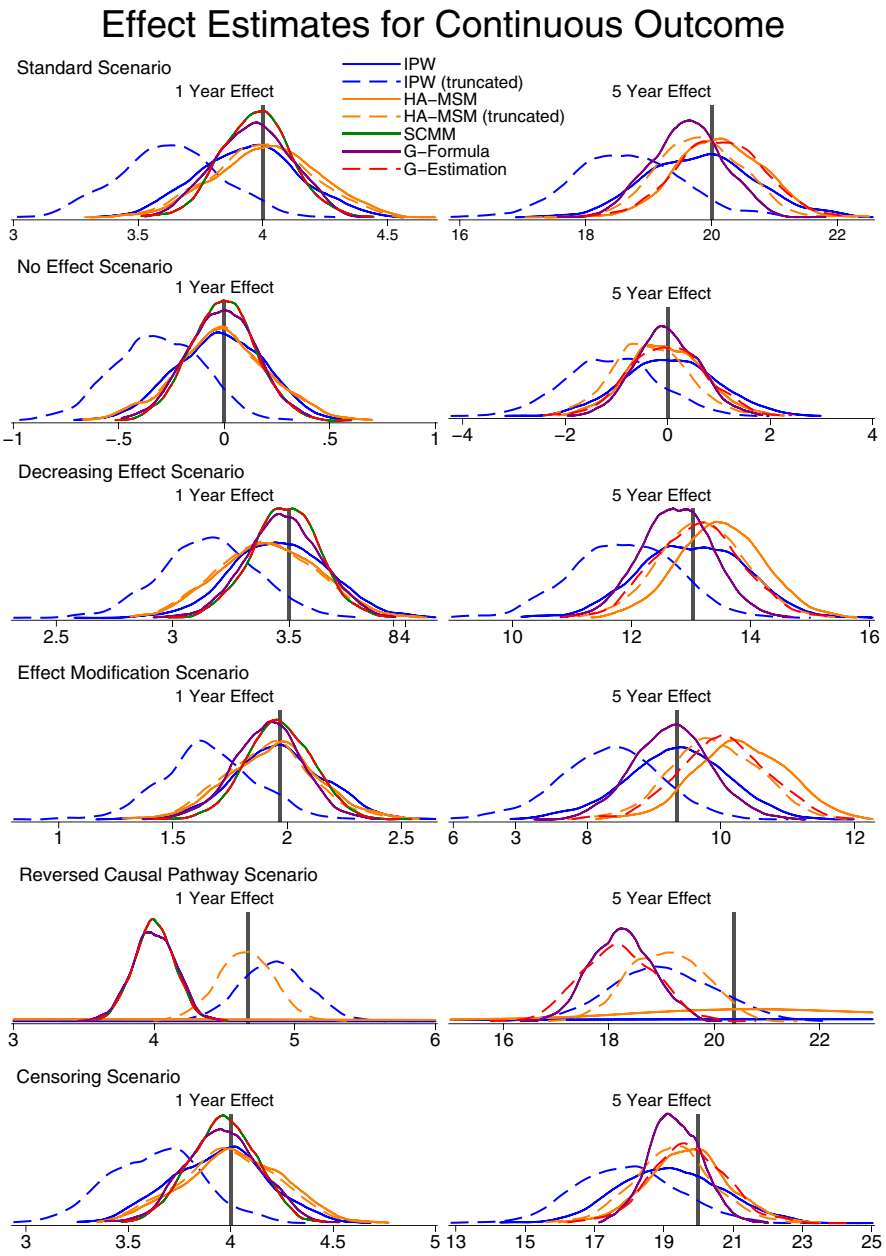


FIGURE 4 Kernel density plots showing the distribution of population-average effect estimates for a continuous outcome. The vertical line shows the correct effect. HA-MSM, history-adjusted marginal structural models; IPW, inverse probability weighting; SCMM, sequential conditional mean models

3 methods perform similarly well in estimating interaction terms, although there is still some finite sample size bias, and a much larger sample size would be needed to accurately estimate the interaction terms. Even in cases where there is no effect modification, including an interaction term in the models did not introduce bias, and the methods correctly estimate zero for the interaction term on average.

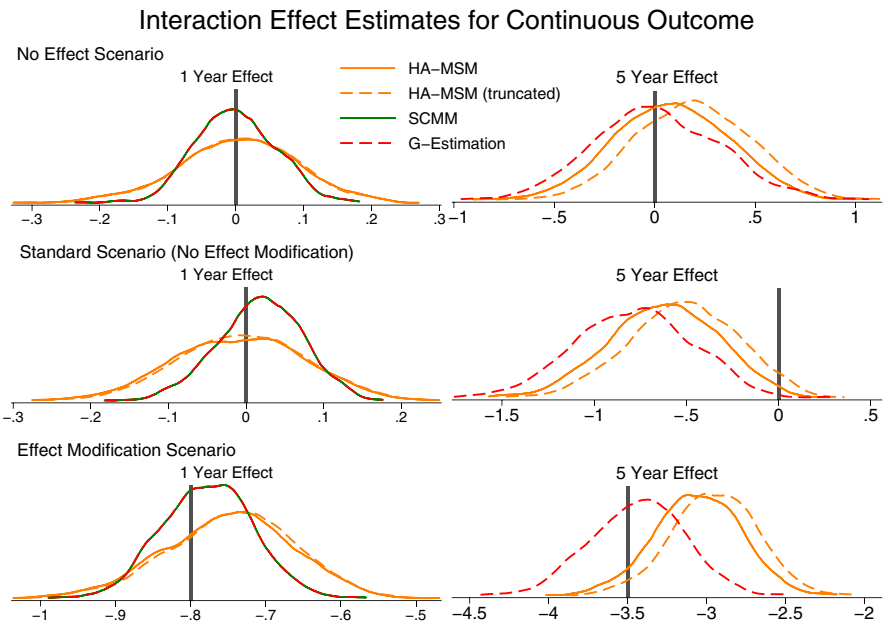


FIGURE 5 Kernel density plots showing the distribution of interaction effect estimates for a continuous outcome. The vertical line shows the correct effect. HA-MSM, history-adjusted marginal structural models; IPW, inverse probability weighting; SCMM, sequential conditional mean models

When considering the SE, only in SCMM and HA-MSM did the model-estimated SE approximate the empirical SE. This is theoretically known in the case of SCMM with the propensity score known and, therefore, will be approximately correct when the propensity score is well estimated. In the case of HA-MSM, we believe this to be a peculiarity of our simulation setting, and it is unlikely to be true generally. For this reason, for all methods other than SCMM, a bootstrap procedure should be used to obtain reliable SE estimates. Comparing the methods, g-computation formula consistently shows the smallest empirical SE, followed by SCMM, g-estimation, and HA-MSM with similar SE and, finally, IPW with the largest SE. In the scenario of reversed causal pathways, IPW and HA-MSM had especially large SE when untruncated weights were used.

4.2.2 | Count outcome

Unlike with the continuous outcome, due to the issue of non-collapsibility, we do not compare the effect estimates for the count outcome to a "correct" value. However, Figures 6 and 7 present the effect estimates and SE for both the odds of a zero count and the rate of the count. As with the continuous outcome, more detailed results can be found in Table S5.

Both IPW and g-computation formula provide marginal effect estimates and in almost all cases provide very similar estimates. The only setting where they do not provide similar estimates is the case of reversed causal pathways where IPW performs very poorly with very large variability, as was also seen for the continuous outcome.

Considering the 3 methods that provide conditional effect estimates, we note that the methods are not in general in agreement, and this is due to the fact that the final models condition on different subsets of variables. In the case of g-estimation, due to the fact that the method can only estimate a rate (rather than also accounting for the separate process of excess zeroes), the estimates from this method are generally very different from all other methods.

The only case where all 5 methods are in agreement is when there is no treatment effect. Here, both the marginal and conditional effect estimates are zero. This suggests that any method could be used to perform a test of the null hypothesis of no treatment effect, but the strength of any effect estimates cannot directly be compared between methods.

The results for estimating interaction terms are presented in Figures 8 and 9 (Table S6). The findings are similar to the case of interaction terms with continuous outcomes, except for the case of g-estimation where even in the case

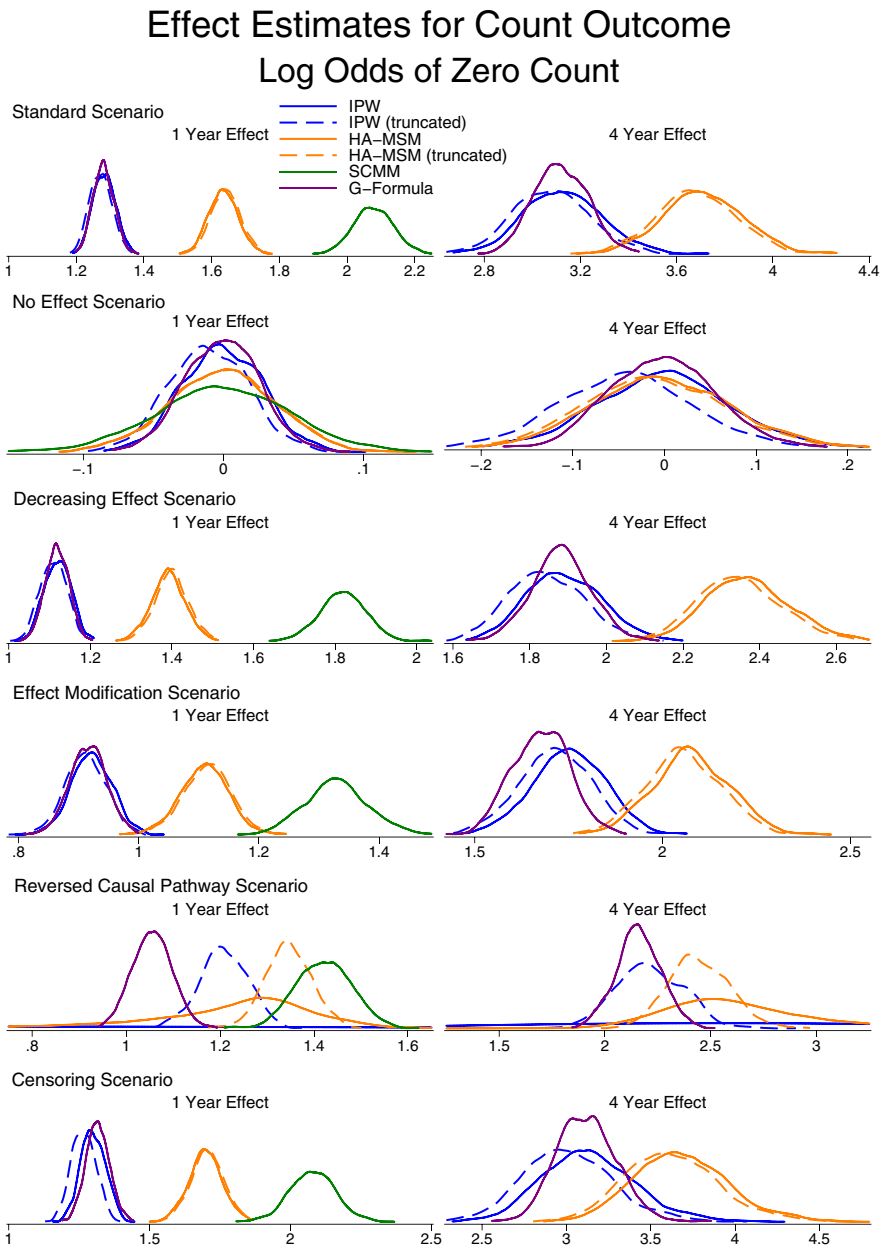


FIGURE 6 Kernel density plots showing the distribution of population-average effect estimates for the odds of a zero count. HA-MSM, history-adjusted marginal structural models; IPW, inverse probability weighting; SCMM, sequential conditional mean models

where there is no effect modification in the data generation process, the method did not average on no effect modification. This is again due to non-collapsibility, where although there is no effect modification present when assuming the data follow a zero-inflated negative binomial distribution, there may be under different distributional models.

Similar to continuous outcomes, the model estimated SE from HA-MSM and SCMM approximated the empirical SE well, but again, we would recommend a bootstrap procedure to be used for all methods other than SCMM.

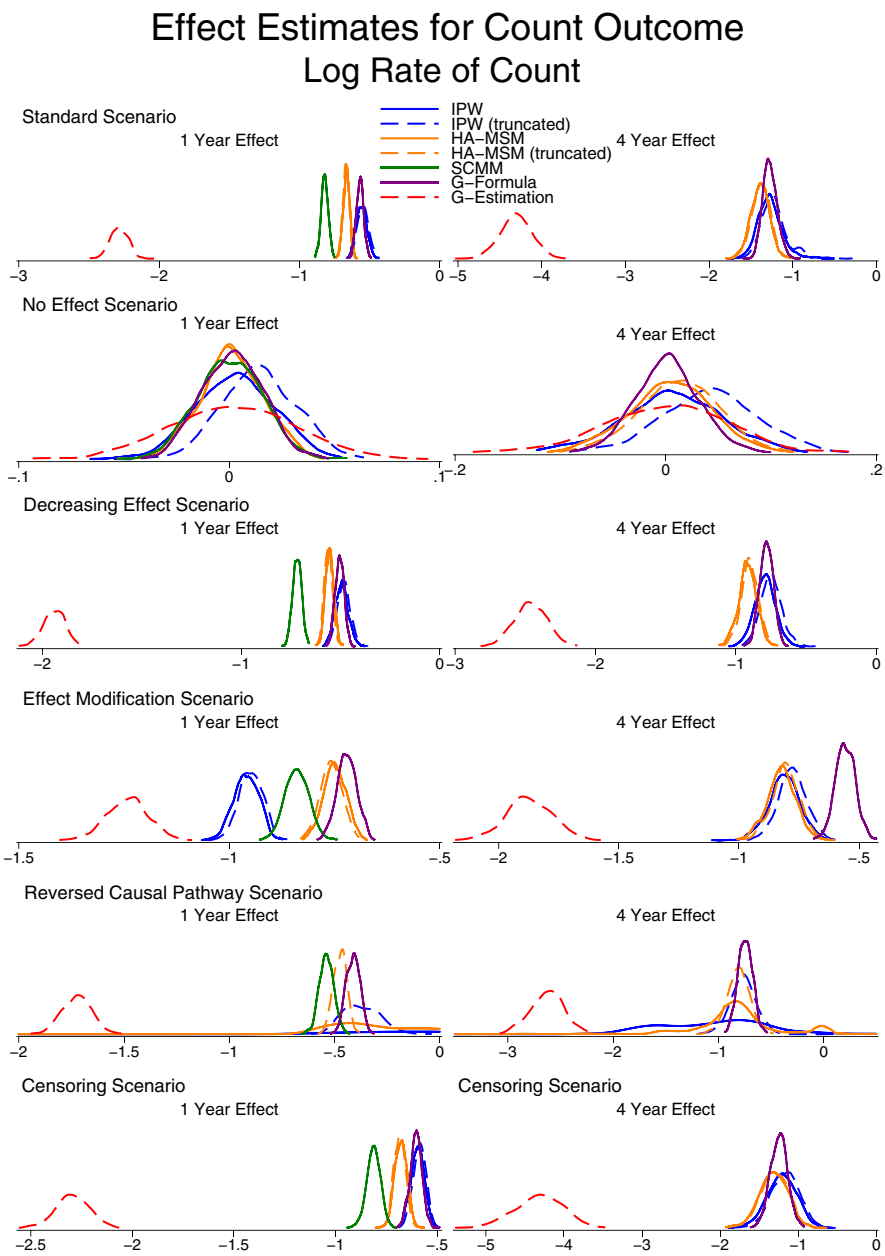


FIGURE 7 Kernel density plots showing the distribution of population-average effect estimates for the rate of a count outcome. HA-MSM, history-adjusted marginal structural models; IPW, inverse probability weighting; SCMM, sequential conditional mean models

5 | DATA ANALYSIS

Based on the findings from the simulation studies, if the real causal pathways in the Registry data are similar to those used in the simulation studies, all 5 available statistical methods would be suitable to investigate the effects of DNase on ppFEV$_1$ and annual number of IV days.

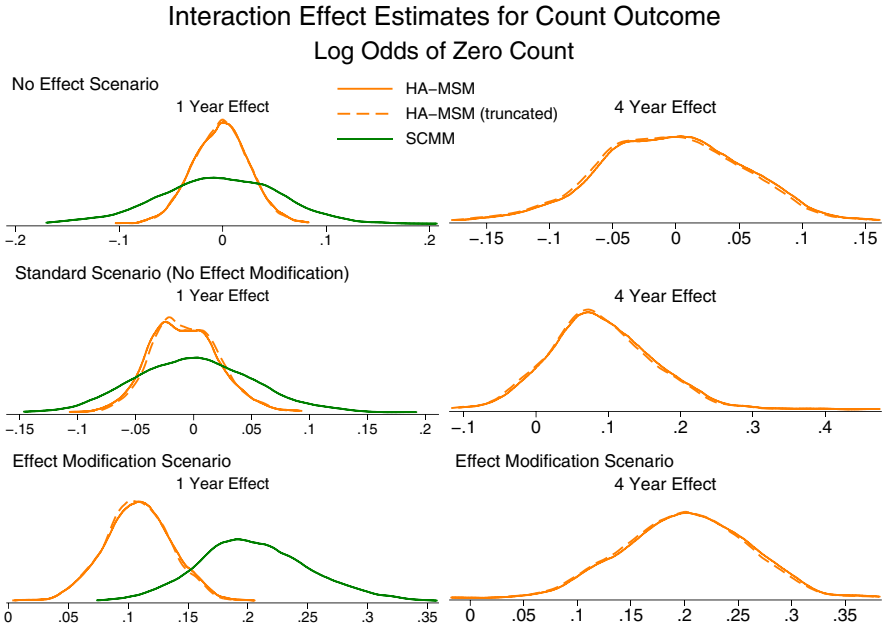


FIGURE 8 Kernel density plots showing the distribution of interaction effect estimates for the odds of a zero count. HA-MSM, history-adjusted marginal structural models; SCMM, sequential conditional mean models

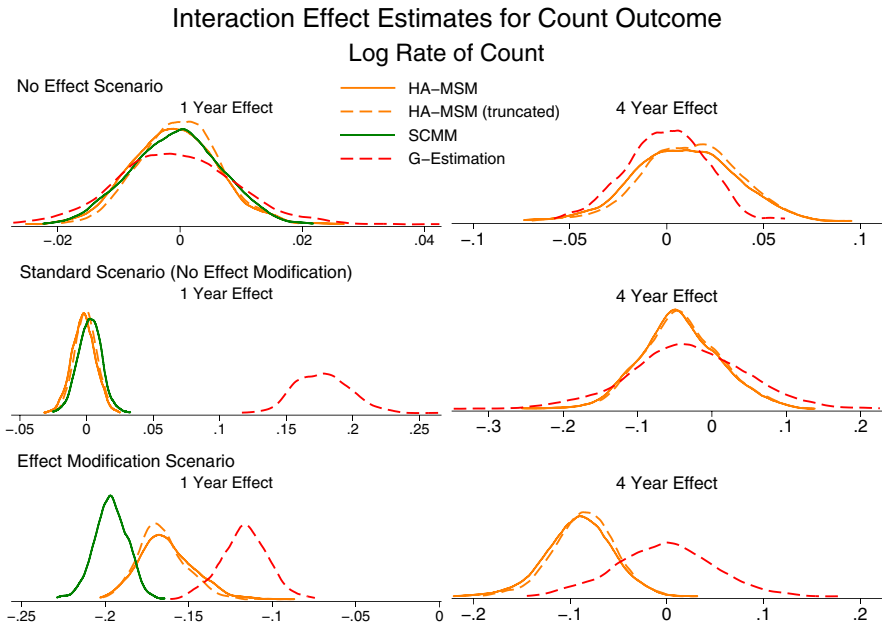


FIGURE 9 Kernel density plots showing the distribution of interaction effect estimates for the rate of a count outcome. HA-MSM, history-adjusted marginal structural models; SCMM, sequential conditional mean models

Unfortunately, the key challenge identified in the simulation studies was that misspecifying the direction of a causal pathway will introduce bias no matter which method is used. In the Registry data, it is likely that the real direction of the causal pathway between treatment (X_t) and annual IV days (V_t) is somewhere between the two extremes of the best-case scenario where X_t is only affected by V_t and the worse-case scenario where X_t only affects V_t. In this setting, the simulation studies showed that we might expect IPW and HA-MSM to perform particularly poorly if the extreme weights are not truncated. Nevertheless, we still perform the analysis here both with and without truncation to compare the effect of truncating weights.

For these analyses, we must also consider the four assumptions highlighted in Section 3.1: no interference, positivity, consistency, and no unmeasured confounding. Interference should not be an issue, because CF is a non-infectious condition. Furthermore, people with CF are generally kept out of direct contact with one another to avoid cross-infection of respiratory microorganisms.[4] The assumption of positivity was also considered to be valid for this investigation. Although guidelines do exist to help advise when patients might benefit from DNase, it is not uncommon for patients to receive or not receive treatment despite the guidelines. Once DNase treatment has been initiated, it is usual to continue to receive the treatment indefinitely, but a number of people do also stop taking treatment for various reasons. Furthermore, in the IPW analysis, there were no extreme weights, suggesting that the assumption of positivity held. Consistency concerns the definition of the intervention. The standard dosage and frequency of DNase is 2.5 mg once a day, but a small number of patients receive a different dosage or frequency. Unfortunately, dosage data are not routinely collected in the Registry. However, consistency is considered to hold under an intervention defined as "receives DNase as prescribed by doctor'.

All models included the time-varying covariates $ppFEV_1$ and IV days as both outcomes and confounders. The analyses also adjusted for baseline confounders: age, sex, ethnicity, and genotype class (a binary marker of the severity of the CF-causing mutation). It is possible that there is residual confounding of the treatment-outcome association and there were a number of other covariates measured in the UK CF Registry that could have been adjusted for, eg, smoking status or body mass index (BMI). However, there is a large amount of missing data in these variables, resulting in many observations being dropped from the analyses if they were included. In sensitivity analyses based on the subset without missing data adjusting for time-varying smoking status and BMI had only a very small impact on the effect estimates.

Our analysis included 22 357 annual assessments from 3847 people. The median number of visits per person was 8 (IQR, 5-9). DNase was used for at least 1 year by 2251 people (58.5%) and for at least 5 years by 823 people (21.4%). Table 1 gives an overview of the people included in the analysis at baseline.

5.1 | Results of lung function analysis

Figure 10 presents the results of the estimated population-average effect of DNase on $ppFEV_1$ depending on length of treatment use. At 1 year, all methods except g-computation formula estimate that treatment has a negative effect on $ppFEV_1$. The results for SCMM, g-computation formula, and g-estimation are, however, not significant ($P = .86$, .89, and .86, respectively), whereas IPW and HA-MSM estimate a stronger, significant, negative effect ($P < .001$ and .005), which does not change much upon truncation of the extreme weights. Looking at longer term effects, all methods showed a trend with the treatment effect becoming more negative through time, with truncated IPW estimating the largest difference in $ppFEV_1$ between those taking and not taking treatment of −8.81% (95% CI, −10.50 to −7.12, $P < .001$) and HA-MSM the smallest effect of −1.52% (95% CI, −3.30 to 0.27, $P = .097$). Full results from this analysis can be found in Table S7.

SCMM, and g-estimation were also used to investigate effect modification of the treatment effect by time-varying $ppFEV_1$. These results are presented in Figure 11 and show that treatment was estimated to be beneficial in people with lower baselined $ppFEV_1$. HA-MSM estimated an intercept term of 3.32, with treatment becoming less beneficial by 0.57 per 10% change in baseline $ppFEV_1$. This equates to a beneficial effect for people with a baseline $ppFEV_1$ below 58% and a negative effect for people with $ppFEV_1$ above 58%. SCMM and g-estimation estimated a more attenuated interaction effect where treatment became less effective by 0.37 per 10% change in baseline $ppFEV_1$. This means that for these methods, treatment was estimated to be beneficial for people with a baseline $ppFEV_1$ up to 73%.

Looking at the 5-year treatment effect, the interaction between treatment and $ppFEV_1$ was estimated to increase in strength leading to a bigger differentiation in effect between those with low and high baseline $ppFEV_1$. In the case of HA-MSM, the intercept was estimated to be 8.30 with a change in effect of −1.29 per 10% increase in $ppFEV_1$, leading to a boundary for a beneficial effect of 64%. G-estimation showed a stronger interaction effect at 5 years of −2.69%, but

TABLE 1 Descriptive baseline statistics of people included in data analysis. Mean (SD) are given for continuous variables and n (%) for categorical variables

	Received DNase During Follow-Up?	
Variable	**No (n = 1596)**	**Yes (n = 2251)**
Age, y	20.8 (13.9)	16.0 (11.7)
$ppFEV_1$	84.5 (19.9)	78.6 (20.5)
Annual IV days	5.8 (14.4)	10.6 (19.4)
Sex		
Female	716 (44.9)	1081 (48.0)
Male	880 (55.1)	1170 (52.0)
Ethnicity		
Caucasian	1547 (96.9)	2172 (96.5)
Other	49 (3.1)	79 (3.5)
Genotype class		
High	909 (57.0)	1708 (75.9)
Low	310 (19.4)	183 (8.1)
Unassigned	377 (23.6)	360 (16.0)

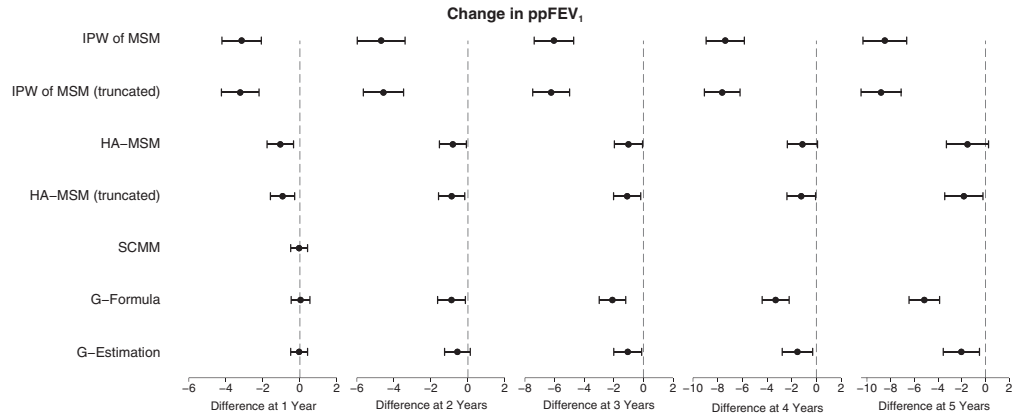


FIGURE 10 Plots showing the estimated population-average effect of DNase treatment on $ppFEV_1$. HA-MSM, history-adjusted marginal structural models; IPW, inverse probability weighting; MSM, marginal structural models; SCMM, sequential conditional mean models

due to the increased SE, this was not significant ($P = .16$). The full results from the analysis including the interaction term can be seen in Table S8.

5.2 | Results of IV days analysis

Similarly to the $ppFEV_1$ analysis, we generally estimated a negative effect when considering the population average effect of DNase on the annual number of IV days. These results are shown in Figure 12 and can also be seen in more detail in Table S9.

At 1 year, all methods estimated a strong, significant decrease in the odds of having zero IV days and an increase in the overall rate of the number of IV days for those receiving treatment. As we observed in the simulation studies, the estimates from g-estimation were larger due to the fact that it does not estimate the odds of a zero count separately to the overall rate.

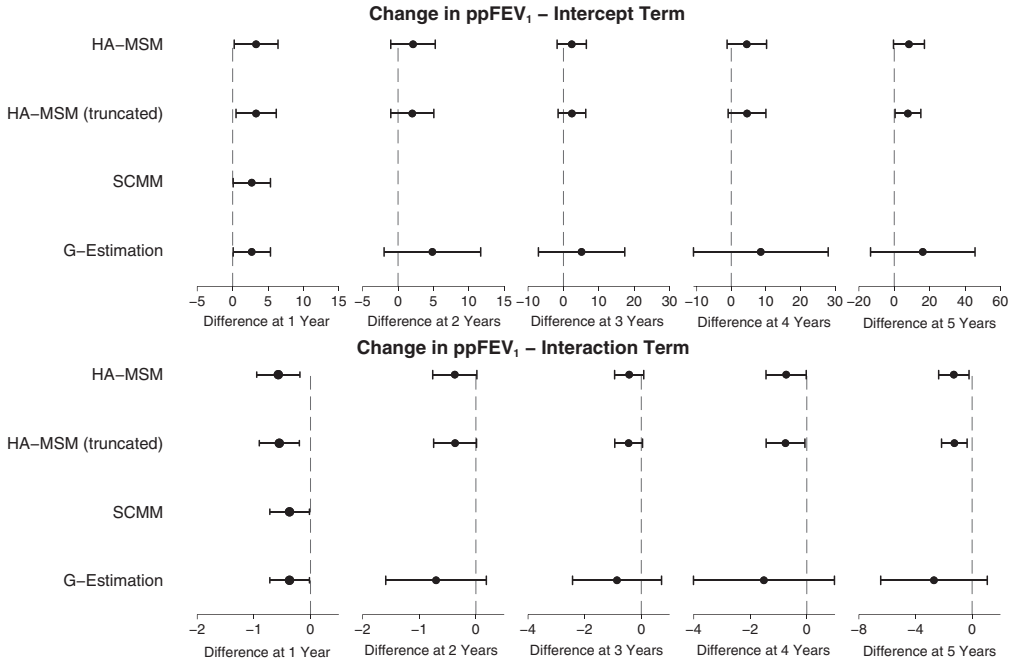


FIGURE 11 Plots showing the estimated effect of DNase treatment on $ppFEV_1$ with effect modification by previous measure of $ppFEV_1$. The intercept term is the estimated effect for an individual with $ppFEV_1$ equal to 0 in the previous year, and the interaction effect is the estimated change per 10 increase in the previous year's $ppFEV_1$. HA-MSM, history-adjusted marginal structural models; SCMM, sequential conditional mean models

The estimates for the 4-year treatment effects were very similar to the 1-year treatment effect estimates, so although treatment was still not estimated to be beneficial, we did not observe a trend of divergence between the treated and nontreated as was observed with $ppFEV_1$.

The results including an interaction term between previous number of IV days and treatment are shown in Figure 13. In people who had previously had zero IV days, treatment was estimated to decrease their odds of zero future IV days by between 0.62 (HA-MSM) and 0.73 (SCMM), but for every 10 additional previous IV days, the odds of zero future IV days increased by between 1.16 (HA-MSM) and 1.17 (truncated HA-MSM). This means that treatment would be estimated to become beneficial on the odds of zero future IV days in individuals who previously had more than between 21 IV days (SCMM) or 32 IV days (HA-MSM).

Considering the overall rate of IV days, the interaction effect was not significant for HA-MSM but was for SCMM and g-estimation, where for people with zero previous IV days, treatment was estimated to increase the rate of future IV days by between 1.12 (SCMM) and 1.36 (g-estimation), and this was estimated to decrease by a rate of 0.98 (SCMM) and 0.94 (g-estimation) per 10 IV days, resulting in a treatment estimated to be beneficial for people with more than 56 previous IV days for SCMM or more than 50 previous IV days (g-estimation).

By 4 years, the interaction present at 1 year modifying the effect of the odds of a zero count had attenuated from 1.16 to 1.08 and was no longer significant. However, there was moderate evidence of interaction when considering the overall rate of IV days with treatment estimated to be beneficial at 4 years in those who had previously had more than 43 days (HA-MSM) or 162 IV days (g-estimation). Table S10 contains the full results from this analysis.

6 | DISCUSSION

We have investigated the suitability of five methods for estimating treatment effects in longitudinal observational data using simulation studies and applied the methods to the UK CF Registry. The focus was on five features encountered in

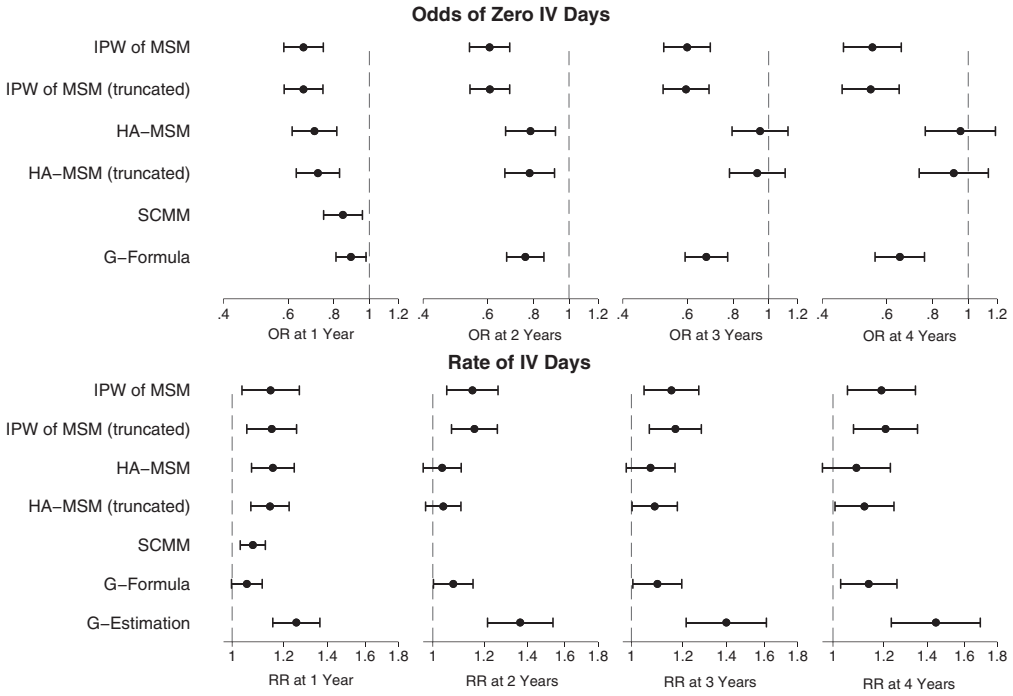


FIGURE 12 Plots showing the estimated population-average effect of DNase treatment on annual IV days. HA-MSM, history-adjusted marginal structural models; IPW, inverse probability weighting; MSM, marginal structural models; SCMM, sequential conditional mean models

these investigations (Section 2.2). The suitability and performance of the methods differs depending on the research question, the nature of the treatment effect of interest, and the features of the data. Here, we provide an overview and recommendations based on our findings.

Our simulation studies showed that all the methods we considered are suitable for analysing registry data to investigate treatment effects in many scenarios. Specifically in the standard scenario, where all models are correctly specified, all methods performed very similarly with little impact depending on the method chosen.

In the case of IPW, however, there were noticeable differences between the method with truncated and untruncated weights. In most situations, the untruncated weights performed best, but in the situation of causal pathways being misspecified, the truncated weights showed much better performance. In a real scenario, we would not know which scenario we are in; it would therefore be difficult to know when weights should be truncated or not. It may be sensible to only truncate when there are "extreme" weights, but there is no clear definition of how large a weight must be before it is "extreme." This would suggest, therefore, in situations where there is uncertainty in the correct direction of causal pathways, that IPW not be used.

HA-MSM performed similarly to IPW of MSM in cases where there is no effect modification, but as it is a more complex method, it would be preferable to use standard IPW of MSM over HA-MSM in most cases.

For measuring the 1-year effect of treatment, SCMM would probably be the preferred method due to its good performance in the simulation studies and its simplicity to implement. The obvious drawback is that the method cannot be used to estimate long-term effects like the other methods, but we recommend that this method be used alongside other methods to check whether the more complex methods are in agreement with the 1-year effect estimate of the SCMM. In cases where the 1-year effect estimate is markedly different between SCMM and another method, this could act as a flag of potential issues with the analysis.

Another benefit of SCMM is that the model-based standard errors will be approximately correct when the propensity score is well estimated, meaning that the bootstrap does not need to be used and results can be obtained much faster than using the other methods presented in this paper. The asymptotic SEs have been derived for IPW of MSM, but only

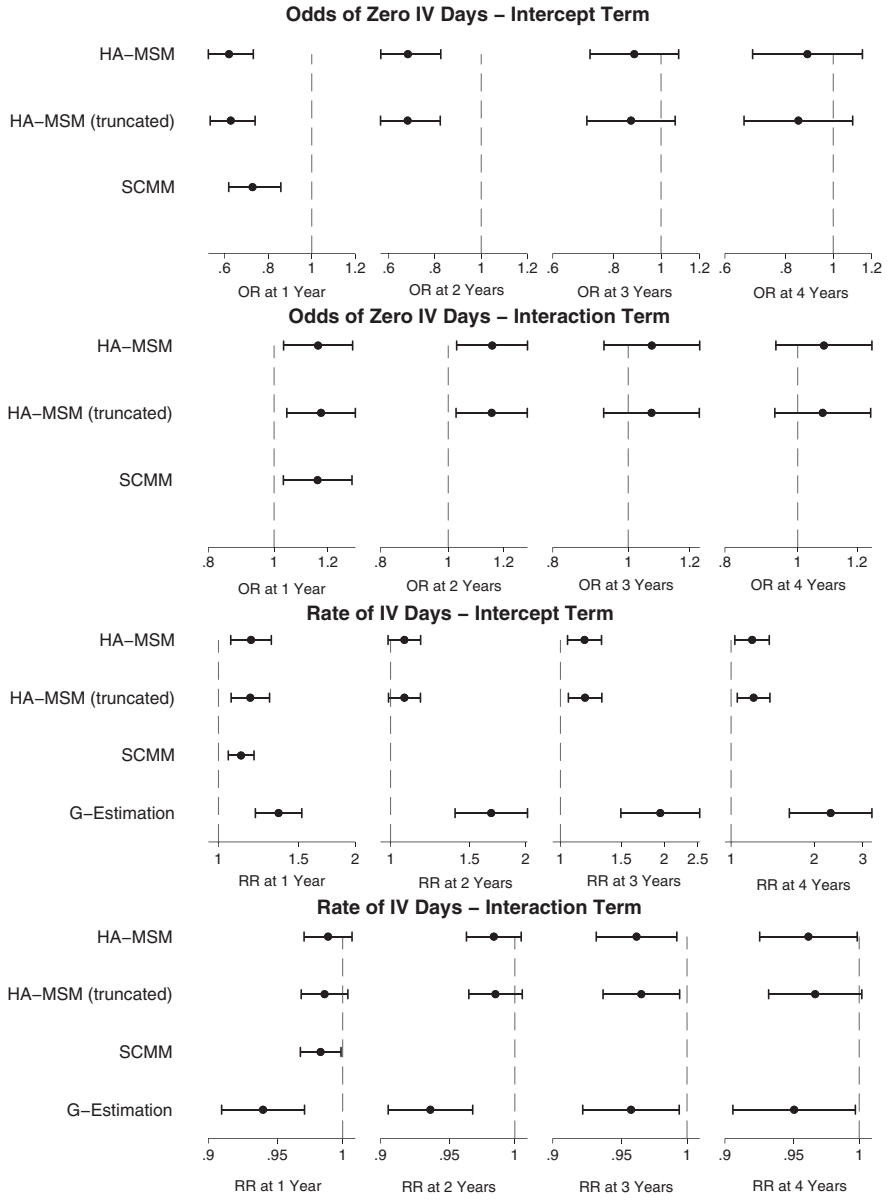


FIGURE 13 Plots showing the estimated effect of DNase treatment on IV days with effect modification by previous number of IV days. The intercept term is the estimated effect for an individual with 0 IV days in the previous year, and the interaction effect is the estimated change per 10 increase in the number of IV days in the previous year. HA-MSM, history-adjusted marginal structural models; SCMM, sequential conditional mean models

in a time-fixed setting,[26] and the difficulty of deriving these in a longitudinal setting necessitate the use of the bootstrap for all the methods other than SCMM.

G-computation formula tended to perform as well as other methods, only performing poorly where other methods also performed poorly. The SE were consistently smaller than for other methods, which would always be preferable in cases where we are confident in the specified models for the time-varying covariates. However, in cases where there is

misspecification, the SE remains small, and with real data, it is unlikely that all the assumptions necessary for g-computation formula would be completely correct, which could result in tight confidence intervals around an incorrect effect estimate. In our scenarios, we did not encounter any issue with the g-null paradox. This is because, for the g-null paradox to arise it is necessary for treatment to affect a time-dependent confounder without having any direct or indirect effect on the outcome.[27] In our "no effect scenario," treatment had no effect on either lung function or IV days, which are acting as both the outcome and the time-dependent confounders.

For continuous outcomes, g-estimation performed well with the SE generally lying between those of g-computation formula and IPW, with the advantage that the method can also estimate effect modification by time-varying covariates, without the drawbacks of unstable weights which were sometimes observed in HA-MSM. However, with the count outcome, g-estimation used a gamma model rather than the zero-inflated negative binomial model like the other methods presented in this paper. This resulted in only one rate ratio compared with the two distinct effect estimates of the other methods making comparison difficult. In situations where the count outcome is not as skewed as the annual IV days in the UK CF registry data, g-estimation may be a suitable method, but in our setting, the other methods were generally preferable.

We outlined how all methods can handle a count outcome, with the outcome model being restricted to a gamma model in g-estimation. A further complexity of the count outcome is the issue of non-collapsibility. In the simulations, we found that when there is truly no treatment effect, both marginal and conditional estimates were correctly consistent with there being no treatment effect. However, in cases where there is a treatment effect, comparison between marginal and conditional estimates from different methods is not as useful.

In addition to the five methods considered in this paper, there are other methods that could have been considered for estimation of treatment effects in the analysis of the Registry data. One such method is targeted maximum likelihood estimation, which is related to the g-computation formula.[28] This method has previously been compared with both IPW and the g-computation formula.[29-31]

Considering the analysis of the UK CF Registry data, as hypothesised, there did appear to be effect modification of the treatment effect by previous ppFEV$_1$ and previous annual IV days. This resulted in the population average estimates hiding the fact that treatment could be beneficial for a group of people. Therefore, in this situation, we would prefer to use SCMM, HA-MSM, or g-estimation, which can estimate effect modification of the treatment effect by time-varying covariates. Due to the fact that we are unsure of the correct specification of some of the causal pathways, HA-MSM may not be a suitable method as shown in the simulation studies. However, the results from all four methods (Figure 11) were very similar, suggesting that the direction of the causal pathways may not be misspecified and any of the methods may in fact be suitable.

There was also evidence of effect modification of the treatment effect on the annual number of IV days. However, depending on the method used, treatment was not estimated to become beneficial until individuals had had over at least 21 IV days in the previous year. In our data, almost 80% of people had fewer than 21 IV days, meaning treatment would only be beneficial in reducing IV days in a small subset of people if these results are reliable. However, a further issue with the annual IV days is that people are not only prescribed IVs as a result of an exacerbation of symptoms, but sometimes they are prescribed as a protective measure to avoid a future exacerbation. It is plausible that people who are more likely to be prescribed treatment are also more likely to be prescribed IVs and it may not be possible to account for this confounding with the available data in the Registry. The issue of unmeasured confounding has not been considered in this paper, because it is an assumption of all the considered methods that there is no unmeasured confounding, but it is important to remember this when considering if the data available are suitable for the desired analysis.

Previous work using more traditional statistical methods has only investigated the effects of up to 2 years of DNase treatment. We have shown in this paper how the data available in registries can be harnessed with appropriate statistical methods to investigate the effects of longer term use of treatments. Many treatments for CF would actually be used for more than 5 years, which was the maximum time-frame considered in this paper due to the limited sample size with follow-up longer than 5 years, but as more data are collected in the UK CF Registry, further analyses with longer follow-up could be performed.

Unfortunately, as with a lot of observational data, there are high levels of missingness in some of the variables collected in the UK CF Registry. As missing data were not the focus of this paper, we presented the results of the data analysis with adjustment for variables that are considered to be the strongest confounders affecting the probability of receiving treatment and outcomes, as these variables are also more widely collected. Furthermore, sensitivity analyses suggested that including other potential confounders such as smoking status or BMI did not result in significant changes to the results.

In conclusion, in most settings, more than one of the available methods would be suitable for the types of analysis considered in this paper. In many cases, therefore, it may be beneficial to consider using more than one available method, to see if the results are consistent. Of course, in cases where 2 separate methods give the same effect estimate, this does not mean it is correct, but does add some reliability to the results. In cases where the methods gave very different effect estimates, this would act as a flag to re-examine the data, the assumptions of the methods, and the suitability of the analyses performed.

ACKNOWLEDGEMENTS

We thank people with CF and their families for consenting to their data being held in the UK CF Registry, and NHS teams in CF centres and clinics for the input of data into the Registry. We also thank the UK Cystic Fibrosis Trust and the Registry Steering Committee for access to anonymised UK CF Registry data.

Simon J. Newsome is supported through the CF-EpiNet Strategic Research Centre funded by the Cystic Fibrosis Trust. Ruth H. Keogh is supported by a Medical Research Council Methodology Fellowship (MR/M014827/1). Rhian M. Daniel is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (107617/Z/15/Z).

ORCID

Simon J. Newsome http://orcid.org/0000-0002-1067-1217
Ruth H. Keogh http://orcid.org/0000-0001-6504-3253
Rhian M. Daniel http://orcid.org/0000-0001-5649-9320

REFERENCES

1. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JAC. Methods for dealing with time-dependent confounding. *Stat Med*. 2013;32(9):1584-1618.
2. Ratjen F, Döring G. Cystic fibrosis. *Lancet*. 2003;361(9358):681-689.
3. Jeffery A, Charman S, Cosgriff R, Carr S. UK Cystic Fibrosis Registry Annual Data Report. 2016, Cystic Fibrosis Trust; 2017.
4. Bush A, Bilton D, Hodson M, eds.. *Hodson and Geddes' Cystic Fibrosis*. Boca Raton, Florida: Taylor & Francis; 2015.
5. Taylor-Robinson D, Archangelidi O, Carr S, et al. Data resource profile: the UK cystic fibrosis registry. *Int J Epidemiol*. 2017;dyx196. https://doi.org/10.1093/ije/dyx196
6. Fuchs HJ, Borowitz DS, Christiansen DH, et al. Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. *N Engl J Med*. 1994;331(10):637-642.
7. Yang C, Chilvers M, Montgomery M, Nolan SJ. Dornase alfa for cystic fibrosis. *Cochrane Database Syst Rev*. 2016;4. https://doi.org/10.1002/14651858.CD001127.pub3
8. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Stat*. 1978;6(1):34-58.
9. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part II: proofs of results. *Int J Biostat*. 2010;6(2). https://doi.org/10.2202/ 1557-4679.1242
10. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20(6):880-883.
11. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560.
12. Wang Y, Petersen ML, Bangsberg D, van der Laan MJ. Diagnosing bias in the inverse probability of treatment weighted estimator resulting from violation of experimental treatment assignment. *UC Berkeley Division of Biostatistics working paper series*. 2006;211. http://biostats.bepress.com/ucbbiostat/paper211
13. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656-664.
14. Petersen ML, Deeks SG, Martin JN, van der Laan MJ. History-adjusted marginal structural models for estimating time-varying effect modification. *Am J Epidemiol*. 2007;166(9):985-993.
15. Keogh RH, Daniel RM, Vanderweele TJ, Vansteelandt S. Analysis of longitudinal studies with repeated outcome measures: adjusting for time-dependent confounding using conventional methods. *Am J Epidemiol*. 2017;kwx311. https://doi.org/10.1093/aje/kwx311
16. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Modell*. 1986;7(9–12):1393-1512.

17. Robins JM, Wasserman L. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In: Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence Geiger D, ShenoyP, eds.; 1997; Morgan Kaufmann, San Francisco:409-420.
18. Robins JM, Blevins D, Ritter G, Wulfsohn M. G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*. 1992;3(4):319-336.
19. Vansteelandt S, Joffe M. Structural nested models and g-estimation: the partially realized promise. *Stat Sci*. 2014;29(4):707-731.
20. Vansteelandt S, Sjolander A. Revisiting g-estimation of the effect of a time-varying exposure subject to time-varying confounding. *Epidemiol Methods*. 2016;5(1):37-56.
21. Dukes O, Vansteelandt S. A note on g-estimation of causal risk ratios. *Am J Epidemiol*. 2018;kwx347. https://doi.org/10.1093/aje/kwx347
22. Snowden JM, Rose S, Mortimer KM. Implementation of g-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol*. 2011;173(7):731-738.
23. Naimi AI, Cole SR, Hudgens MG, Richardson DB. Estimating the effect of cumulative occupational asbestos exposure on time to lung cancer mortality: using structural nested failure-time models to account for healthy-worker survivor bias. *Epidemiology*. 2014;25(2):246-54.
24. Laber EB, Lizotte DJ, Qian M, Pelham WE, Murphy SA. Dynamic treatment regimes: technical challenges and applications. *Electron J Stat*. 2014;8(1):1225-1272.
25. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med*. 2006;25(24):4279-4292.
26. Williamson EJ, Forbes A, White IR. Variance reduction in randomised trials by inverse probability weighting using the propensity score. *Stat Med*. 2014;33(5):721-737.
27. Young JG, Tchetgen Tchetgen EJ. Simulation from a known Cox MSM using standard parametric models for the g-formula. *Stat Med*. 2014;33(6):1001-1014.
28. Van Der Laan MJ. Targeted maximum likelihood based causal inference: Part I. *Int J Biostat*. 2010;6(2). https://doi.org/10.2202/1557-4679.1211
29. Pang M, Schuster T, Filion KB, Schnitzer ME, Eberg M, Platt RW. Effect estimation in point-exposure studies with binary outcomes and high-dimensional covariate data—a comparison of targeted maximum likelihood estimation and inverse probability of treatment weighting. *Int J Biostat*. 2016;12(2). https://doi.org/10.1515/ijb-2015-0034
30. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol*. 2017;185(1):65-73.
31. Neugebauer R, Schmittdiel JA, van der Laan MJ. Targeted learning in real-world comparative effectiveness research with time-varying interventions. *Stat Med*. 2014;33(14):2480-2520.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Newsome SJ, Keogh RH, Daniel RM. Estimating long-term treatment effects in observational data: A comparison of the performance of different methods under real-world uncertainty. *Statistics in Medicine*. 2018;37:2367–2390. https://doi.org/10.1002/sim.7664

Appendix C

Supplementary Tables & Figures to Chapter 5

In Chapter 5 we only presented the results of the shortest- and longest-term effects. In all cases, the bias, empirical standard error, model-based standard error and MSE was smallest for the one-year effect and sequentially grew larger with longer-term effects. Here we present the kernel density plots and the tables of results for the two-, three-, and four- year results for the continuous outcome and the two- and three-year results for the count outcome. The SCMM method is not included in any of the following figures and tables, as this method only estimates the one-year effect, which was presented in the Chapter 5.

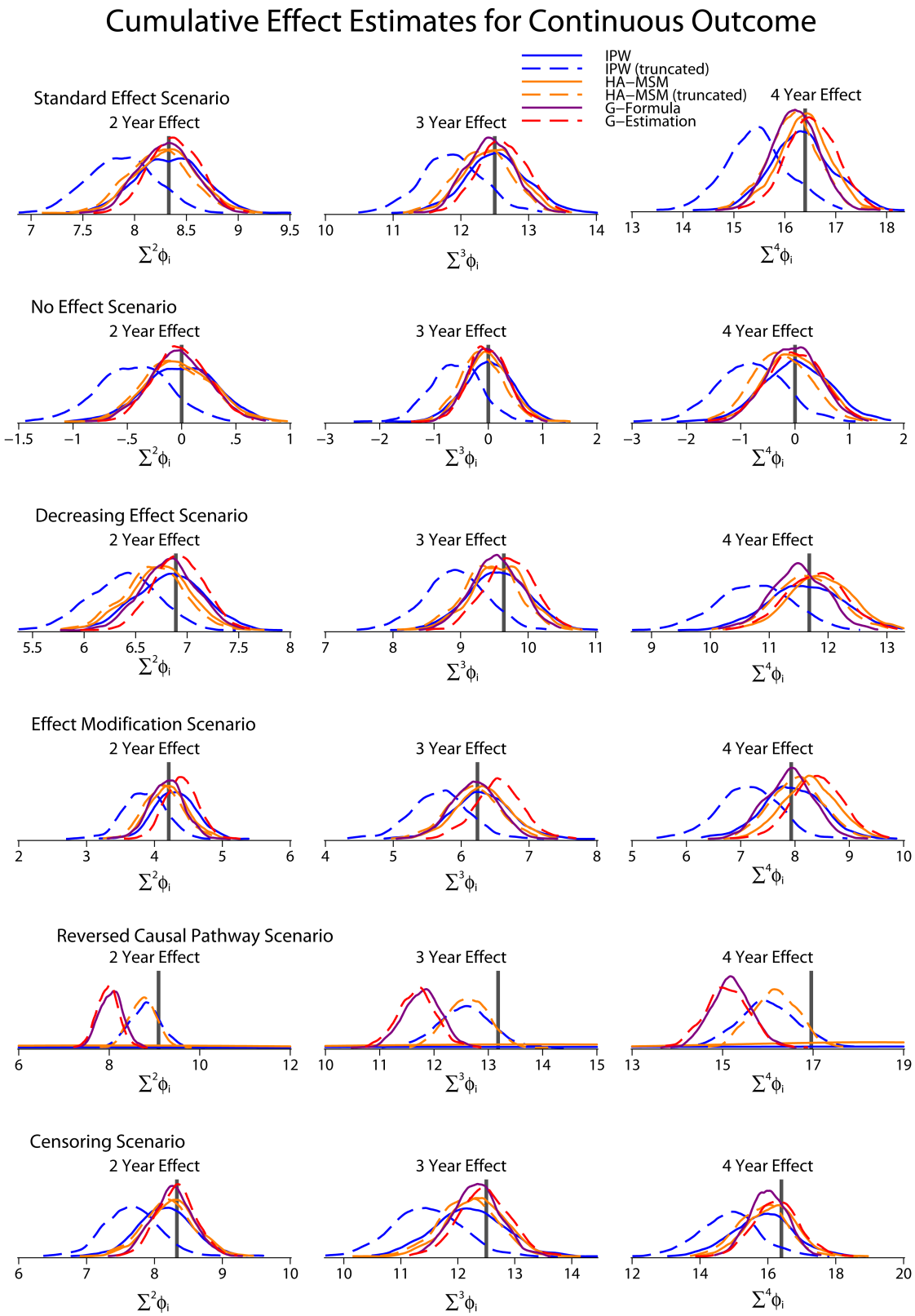


FIGURE C.1: Kernel density plots of two-, three- and four-year population-average effect estimates for a continuous outcome (The vertical line shows the correct effect)

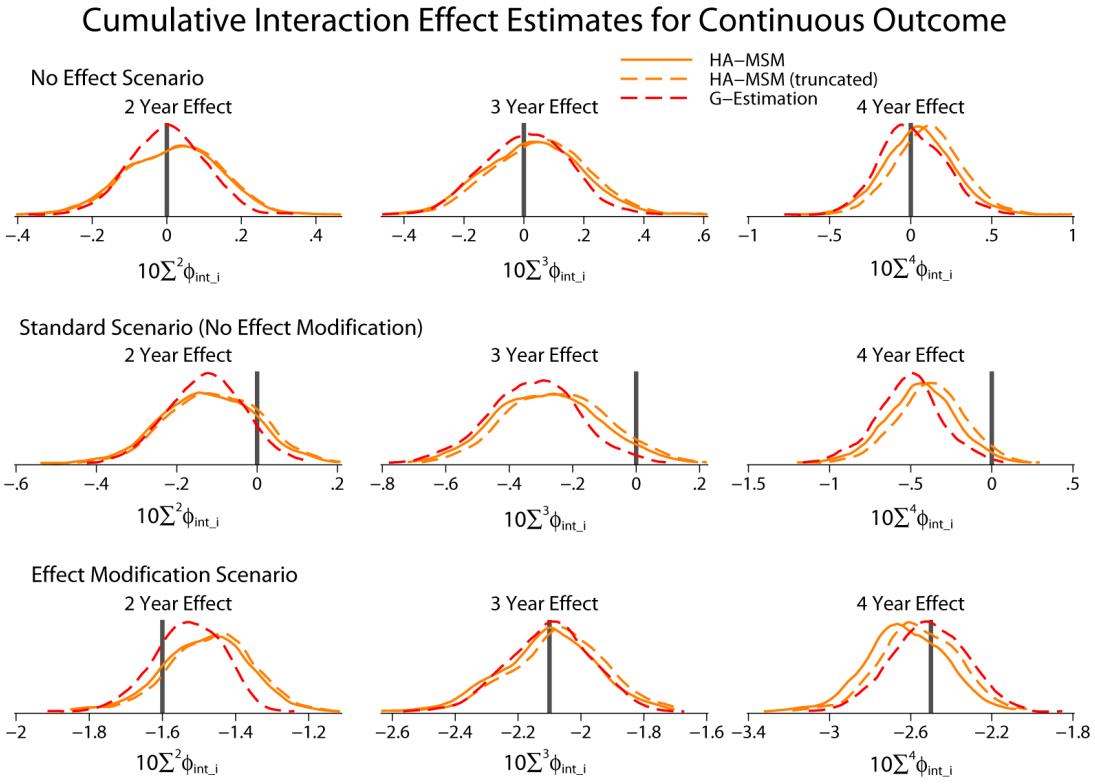


FIGURE C.2: Kernel density plots of two-, three- and four-year interaction effect estimates for a continuous outcome (The vertical line shows the correct effect)

Scenario	Method	n	Cumulative 2 Year Treatment Effect ($\sum_{i=1}^{2} \phi_i$)					Cumulative 3 Year Treatment Effect ($\sum_{i=1}^{3} \phi_i$)				
			Mean	Bias	Empirical SE	Model SE	MSE	Mean	Bias	Empirical SE	Model SE	MSE
Standard $\sum_1^2 \phi_i = 8.33$ $\sum_1^3 \phi_i = 12.50$	IPW of MSM	1000	8.34	0.010	0.32	0.52	0.10	12.53	0.026	0.46	0.65	0.21
	IPW of MSM (truncated)	1000	7.87	-0.46	0.32	0.49	0.31	11.86	-0.64	0.45	0.62	0.61
	HA-MSM	1000	8.31	-0.019	0.29	0.30	0.087	12.46	-0.039	0.40	0.40	0.16
	HA-MSM (truncated)	1000	8.23	-0.096	0.29	0.30	0.095	12.32	-0.18	0.39	0.39	0.19
	G-Formula	1000	8.30	-0.026	0.26	0.17	0.066	12.44	-0.065	0.35	0.20	0.12
	G-Estimation	1000	8.38	0.052	0.24	0.32	0.060	12.61	0.11	0.35	0.54	0.14
No Effect $\sum_1^2 \phi_i = 0.00$ $\sum_1^3 \phi_i = 0.00$	IPW of MSM	1000	-0.008	-0.008	0.32	0.52	0.10	0.005	0.005	0.45	0.65	0.20
	IPW of MSM (truncated)	1000	-0.44	-0.44	0.32	0.49	0.30	-0.61	-0.61	0.45	0.62	0.57
	HA-MSM	1000	-0.007	-0.007	0.31	0.30	0.093	-0.002	-0.002	0.40	0.40	0.16
	HA-MSM (truncated)	1000	-0.052	-0.052	0.30	0.30	0.095	-0.13	-0.13	0.40	0.40	0.17
	G-Formula	1000	-0.019	-0.019	0.27	0.17	0.072	-0.021	-0.021	0.36	0.20	0.13
	G-Estimation	1000	-0.004	-0.004	0.25	0.32	0.061	-0.003	-0.003	0.35	0.55	0.12
Decreasing Effect $\sum_1^2 \phi_i = 6.89$ $\sum_1^3 \phi_i = 9.64$	IPW of MSM	1000	6.83	-0.062	0.33	0.51	0.12	9.56	-0.082	0.46	0.65	0.22
	IPW of MSM (truncated)	1000	6.38	-0.51	0.33	0.49	0.37	8.92	-0.72	0.46	0.62	0.73
	HA-MSM	1000	6.74	-0.15	0.30	0.30	0.11	9.57	-0.073	0.41	0.40	0.17
	HA-MSM (truncated)	1000	6.67	-0.22	0.30	0.30	0.14	9.43	-0.21	0.41	0.40	0.21
	G-Formula	1000	6.82	-0.069	0.27	0.17	0.077	9.52	-0.12	0.36	0.20	0.14
	G-Estimation	1000	6.92	0.027	0.24	0.32	0.061	9.72	0.079	0.37	0.54	0.14
Effect Modification $\sum_1^2 \phi_i = 4.21$ $\sum_1^3 \phi_i = 6.24$	IPW of MSM	1000	4.30	0.09	0.31	0.49	0.092	6.30	0.065	0.44	0.60	0.19
	IPW of MSM (truncated)	1000	3.85	-0.36	0.31	0.47	0.22	5.69	-0.55	0.43	0.57	0.49
	HA-MSM	1000	4.21	0.0005	0.28	0.30	0.079	6.32	0.078	0.39	0.38	0.16
	HA-MSM (truncated)	1000	4.13	-0.768	0.28	0.29	0.084	6.18	-0.063	0.38	0.38	0.15
	G-Formula	1000	4.18	-0.030	0.25	0.16	0.064	6.20	-0.041	0.34	0.18	0.12
	G-Estimation	1000	4.39	0.18	0.24	0.32	0.091	6.55	0.31	0.35	0.53	0.22
Reversed Causal Pathway $\sum_1^2 \phi_i = 9.09$ $\sum_1^3 \phi_i = 13.18$	IPW of MSM	942	-27.87	-36.96	21.28	3.69	1818.41	-13.24	-26.42	29.12	4.75	1545.45
	IPW of MSM (truncated)	1000	8.80	-0.29	0.35	0.54	0.21	12.57	-0.61	0.49	0.67	0.61
	HA-MSM	926	1.22	-7.87	11.61	2.52	196.59	11.48	-1.70	12.48	2.93	158.51
	HA-MSM (truncated)	1000	8.75	-0.34	0.29	0.20	0.30	12.62	-0.56	0.38	0.40	0.46
	G-Formula	1000	8.04	-1.05	0.26	0.17	1.16	11.79	-1.39	0.34	0.19	2.05
	G-Estimation	1000	7.97	-1.12	0.24	0.33	1.31	11.69	-1.49	0.35	0.56	2.35

Scenario	Method	n	Cumulative 2 Year Treatment Effect ($\sum_{i=1}^{2} \phi_i$)					Cumulative 3 Year Treatment Effect ($\sum_{i=1}^{3} \phi_i$)				
			Mean	Bias	Empirical SE	Model SE	MSE	Mean	Bias	Empirical SE	Model SE	MSE
Censoring $\sum_{1}^{2} \phi_i = 8.33$ $\sum_{1}^{3} \phi_i = 12.50$	IPW of MSM	1000	8.16	-0.17	0.41	0.61	0.20	12.20	-0.30	0.61	0.81	0.46
	IPW of MSM (truncated)	1000	7.66	-0.67	0.40	0.59	0.61	11.45	-1.05	0.60	0.77	1.46
	HA-MSM	1000	8.26	-0.066	0.37	0.37	0.14	12.33	-0.17	0.50	0.53	0.28
	HA-MSM (truncated)	1000	8.19	-0.14	0.36	0.37	0.15	12.18	-0.32	0.50	0.52	0.36
	G-Formula	1000	8.27	-0.063	0.29	0.22	0.090	12.33	-0.17	0.41	0.27	0.19
	G-Estimation	1000	8.32	-0.011	0.28	0.39	0.081	12.47	-0.034	0.44	0.68	0.20

TABLE C.1: Simulation study results of two- and three-year population-average effect for continuous outcome ($\overline{D_t} \rightarrow L_t$) (NA signifies that the method does not estimate that effect)

Scenario	Method	n	Cumulative 4 Year Treatment Effect ($\sum_{i=1}^{4} \phi_i$)				
			Mean	Bias	Empirical SE	Model SE	MSE
Standard $\sum_{1}^{4} \phi_i = 16.40$	IPW of MSM	1000	16.34	-0.055	0.61	0.80	0.38
	IPW of MSM (truncated)	1000	15.45	-0.95	0.60	0.76	1.27
	HA-MSM	1000	16.38	-0.018	0.50	0.52	0.25
	HA-MSM (truncated)	1000	16.15	-0.25	0.50	0.52	0.31
	G-Formula	1000	16.21	-0.19	0.44	0.24	0.23
	G-Estimation	1000	16.53	0.13	0.49	0.81	0.25
No Effect $\sum_{1}^{4} \phi_i = 0.00$	IPW of MSM	1000	0.028	0.028	0.62	0.81	0.38
	IPW of MSM (truncated)	1000	-0.82	-0.82	0.62	0.78	1.05
	HA-MSM	1000	-0.011	-0.011	0.52	0.53	0.27
	HA-MSM (truncated)	1000	-0.25	-0.25	0.51	0.53	0.33
	G-Formula	1000	-0.023	-0.023	0.47	0.24	0.22
	G-Estimation	1000	-0.001	-0.001	0.51	0.84	0.26
Decreasing Effect $\sum_{1}^{4} \phi_i = 11.68$	IPW of MSM	1000	11.62	-0.062	0.63	0.80	0.40
	IPW of MSM (truncated)	1000	10.74	-0.94	0.63	0.77	1.27
	HA-MSM	1000	11.87	0.19	0.53	0.53	0.32
	HA-MSM (truncated)	1000	11.63	-0.049	0.53	0.52	0.28
	G-Formula	1000	11.50	-0.18	0.46	0.24	0.25
	G-Estimation	1000	11.78	0.10	0.51	0.82	0.27

Scenario	Method	n	Cumulative 4 Year Treatment Effect ($\sum_{i=1}^{4} \phi_i$)				
			Mean	Bias	Empirical SE	Model SE	MSE
Effect Modification $\sum_1^4 \phi_i = 7.93$	IPW of MSM	1000	7.95	0.019	0.57	0.71	0.32
	IPW of MSM (truncated)	1000	7.14	-0.79	0.56	0.68	0.94
	HA-MSM	1000	8.28	0.35	0.49	0.50	0.36
	HA-MSM (truncated)	1000	8.03	0.097	0.49	0.49	0.25
	G-Formula	1000	7.89	-0.037	0.44	0.22	0.19
	G-Estimation	1000	8.41	0.48	0.47	0.79	0.46
Reversed Causal Pathway $\sum_1^4 \phi_i = 16.96$	IPW of MSM	942	11.22	-5.74	21.78	5.61	1041.81
	IPW of MSM (truncated)	1000	15.99	-0.97	0.66	0.85	1.37
	HA-MSM	926	21.02	4.06	10.72	4.26	131.28
	HA-MSM (truncated)	1000	16.16	-0.80	0.51	0.53	0.90
	G-Formula	1000	15.19	-1.77	0.44	0.23	3.33
	G-Estimation	1000	15.10	-1.86	0.49	0.85	3.69
Censoring $\sum_1^4 \phi_i = 16.40$	IPW of MSM	1000	15.86	-0.54	0.90	1.08	1.10
	IPW of MSM (truncated)	1000	14.85	-1.55	0.88	1.03	3.16
	HA-MSM	1000	16.12	-0.28	0.73	0.75	0.61
	HA-MSM (truncated)	1000	15.84	-0.56	0.72	0.74	0.83
	G-Formula	1000	16.01	-0.39	0.54	0.35	0.44
	G-Estimation	1000	16.26	-0.14	0.67	1.08	0.47

TABLE C.2: Simulation study results of four-year population-average effect for continuous outcome ($\overline{D_t} \rightarrow L_t$) (NA signifies that the method does not estimate that effect)

Scenario	Method	n	Cumulative 2 Year Interaction Effect ($10\sum_{i=1}^{2}\phi_{\text{int}_i}$)					Cumulative 3 Year Interaction Effect ($10\sum_{i=1}^{3}\phi_{\text{int}_i}$)				
			Mean	Bias	Empirical SE	Model SE	MSE	Mean	Bias	Empirical SE	Model SE	MSE
No Effect $10\sum_{1}^{2}\phi_{\text{int}_i} = 0.00$ $10\sum_{1}^{3}\phi_{\text{int}_i} = 0.00$	HA-MSM	1000	0.015	0.015	0.12	0.12	0.016	0.030	0.030	0.16	0.16	0.026
	HA-MSM (truncated)	1000	0.023	0.022	0.12	0.12	0.016	0.055	0.055	0.16	0.16	0.028
	G-Estimation	1000	0.002	0.002	0.10	0.13	0.010	0.009	0.009	0.14	0.21	0.020
Standard $10\sum_{1}^{2}\phi_{\text{int}_i} = 0.00$, $10\sum_{1}^{3}\phi_{\text{int}_i} = 0.00$	HA-MSM	1000	-0.12	-0.12	0.11	0.12	0.028	-0.28	-0.28	0.16	0.16	0.10
	HA-MSM (truncated)	1000	-0.11	-0.11	0.11	0.12	0.024	-0.25	-0.25	0.16	0.16	0.085
	G-Estimation	1000	-0.13	-0.13	0.093	0.12	0.025	-0.32	-0.32	0.14	0.21	0.12
Effect Modification $10\sum_{1}^{2}\phi_{\text{int}_i} = -1.60$ $10\sum_{1}^{3}\phi_{\text{int}_i} = -2.10$	HA-MSM	1000	-1.48	0.12	0.12	0.12	0.027	-2.10	-0.003	0.15	0.16	0.023
	HA-MSM (truncated)	1000	-1.47	0.13	0.12	0.12	0.031	-2.07	0.030	0.15	0.15	0.023
	G-Estimation	1000	-1.53	0.071	0.094	0.12	0.014	-2.10	0.0007	0.13	0.19	0.018

TABLE C.3: Simulation study results of interaction effect for continuous outcome at two- and three-years (Results show change in effect of $\overline{D_t}$ on L_t per 10 change in L_{t-1}) (NA signifies that the method does not estimate that effect)

Scenario	Method	n	Cumulative 4 Year Interaction Effect ($10\sum_{i=1}^{4}\phi_{\text{int}_i}$)				
			Mean	Bias	Empirical SE	Model SE	MSE
No Effect $10\sum_{1}^{4}\phi_{\text{int}_i} = 0.00$	HA-MSM	1000	0.053	0.053	0.21	0.22	0.049
	HA-MSM (truncated)	1000	0.11	0.11	0.21	0.21	0.057
	G-Estimation	1000	0.010	0.010	0.20	0.32	0.039
Standard $10\sum_{1}^{4}\phi_{\text{int}_i} = 0.00$	HA-MSM	1000	-0.44	-0.44	0.22	0.22	0.24
	HA-MSM (truncated)	1000	-0.37	-0.37	0.21	0.21	0.19
	G-Estimation	1000	-0.51	-0.51	0.20	0.31	0.30
Effect Modification $10\sum_{1}^{4}\phi_{\text{int}_i} = -2.50$	HA-MSM	1000	-2.62	-0.12	0.19	0.20	0.053
	HA-MSM (truncated)	1000	-2.56	-0.058	0.19	0.20	0.040
	G-Estimation	1000	-2.51	-0.010	0.18	0.27	0.034

TABLE C.4: Simulation study results of interaction effect for continuous outcome at four-years (Results show change in effect of $\overline{D_t}$ on L_t per 10 change in L_{t-1}) (NA signifies that the method does not estimate that effect)

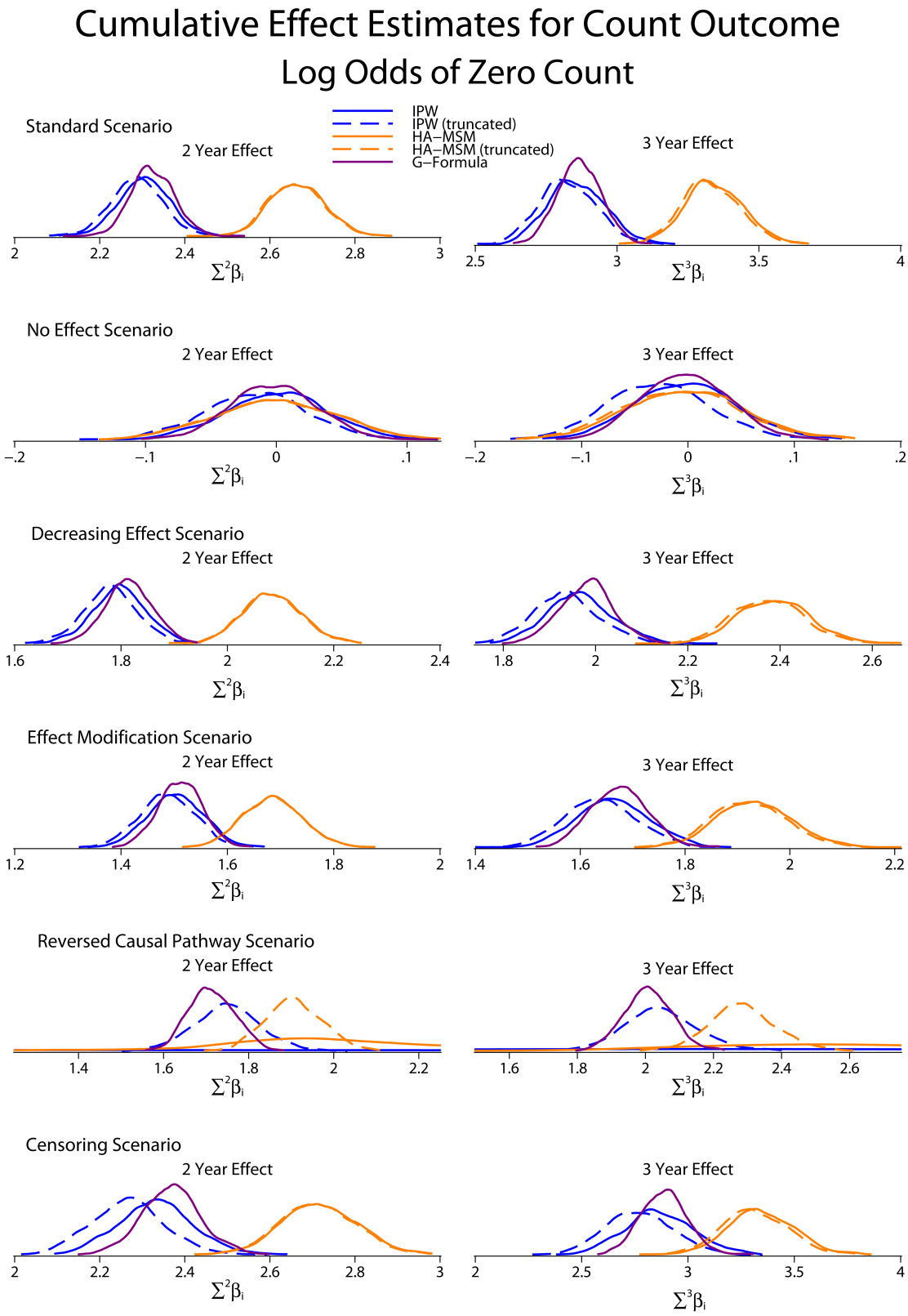


FIGURE C.3: Kernel density plots of two- and three-year population-average effect estimates for the odds of a zero count

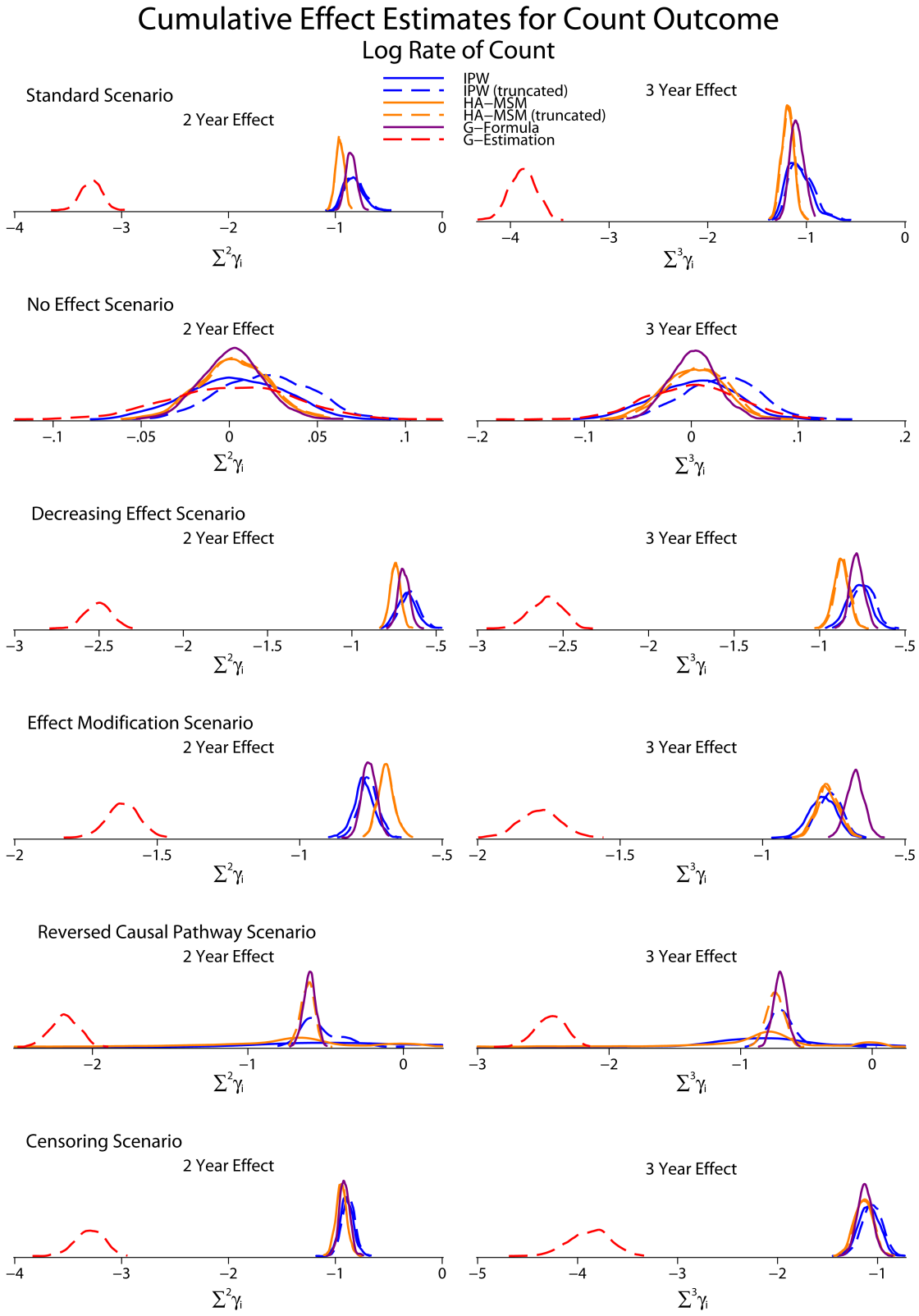


FIGURE C.4: Kernel density plots of two- and three-year population-average effect estimates for the rate of a count

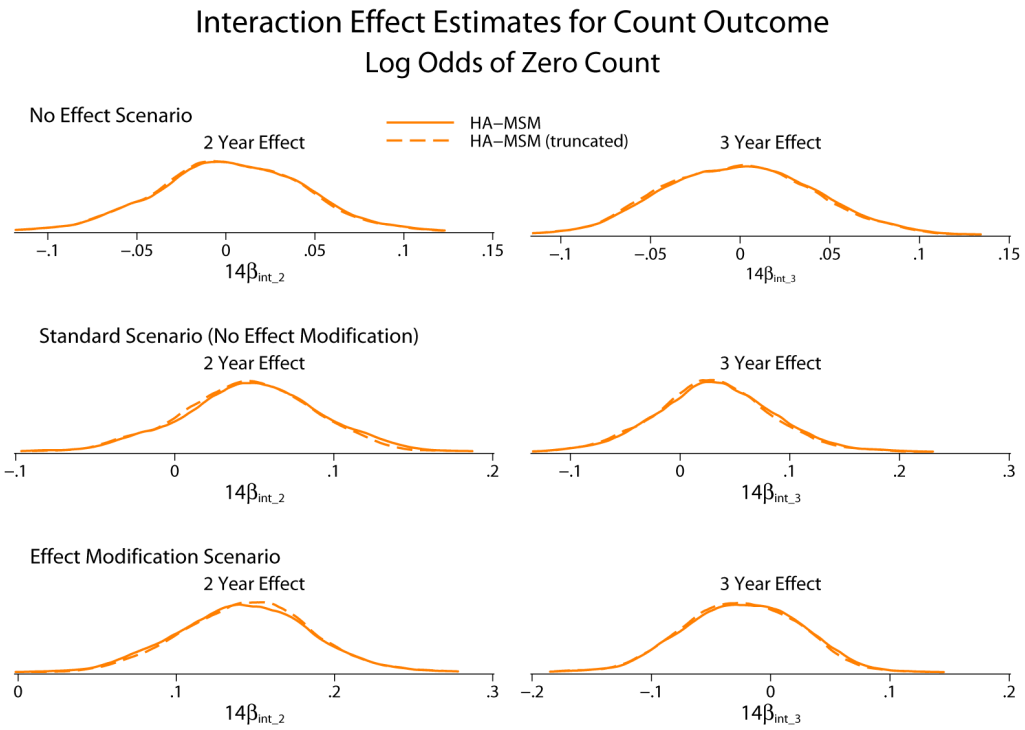


FIGURE C.5: Kernel density plots of two- and three-year interaction effect estimates for the odds of a zero count

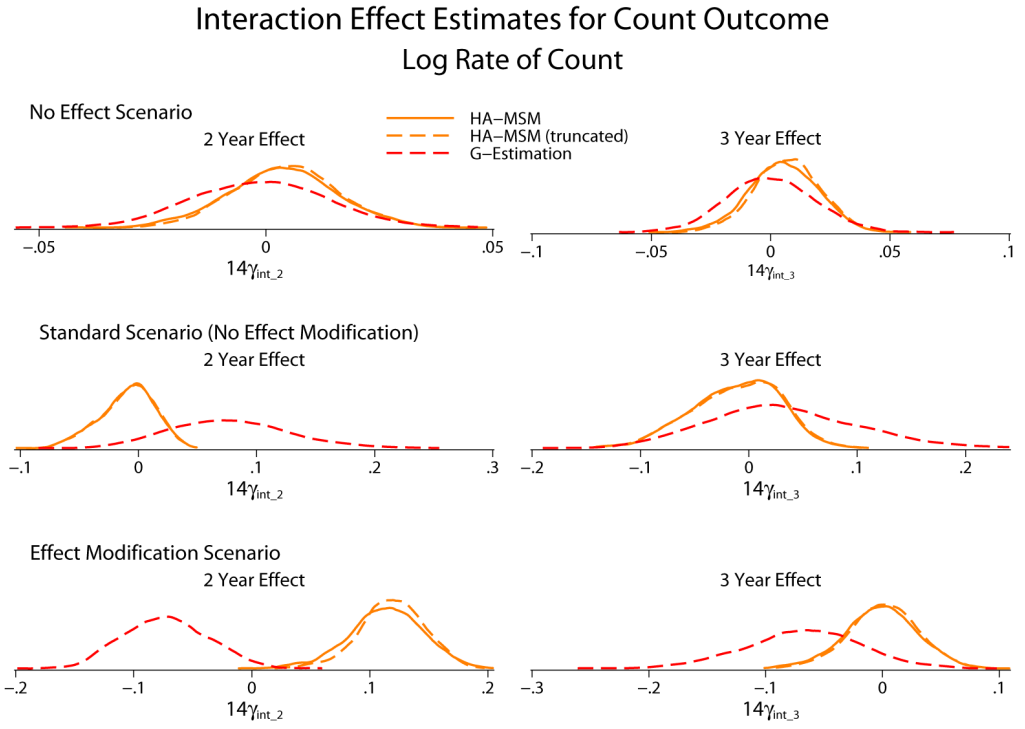


FIGURE C.6: Kernel density plots of two- and three-year interaction effect estimates for the rate of a count

Scenario	Estimate	Method	n	Cumulative 2 Year Treatment Effect					
				Log Odds Ratio of Zero Count ($\sum_{i=1}^{2} \beta_i$)			Log Rate Ratio of Count ($\sum_{i=1}^{2} \gamma_i$)		
				Mean	Empirical SE	Model SE	Mean	Empirical SE	Model SE
Standard	Marginal	IPW of MSM	1000	2.31	0.061	0.066	-0.84	0.083	0.082
		IPW of MSM (truncated)	1000	2.29	0.060	0.066	-0.81	0.083	0.081
		G-Formula	1000	2.32	0.050	0.028	-0.85	0.048	0.040
	Conditional	HA-MSM	1000	2.67	0.064	0.067	-0.96	0.040	0.039
		HA-MSM (truncated)	1000	2.66	0.064	0.067	-0.96	0.039	0.038
		G-Estimation	1000	NA	NA	NA	-3.28	0.096	0.024
No Effect	Marginal	IPW of MSM	1000	0.0007	0.040	0.044	0.004	0.027	0.031
		IPW of MSM (truncated)	1000	-0.016	0.040	0.043	0.021	0.025	0.028
		G-Formula	1000	-0.002	0.033	0.015	0.002	0.017	0.010
	Conditional	HA-MSM	1000	-0.00004	0.046	0.047	0.002	0.020	0.021
		HA-MSM (truncated)	1000	-0.002	0.046	0.047	0.004	0.019	0.021
		G-Estimation	1000	NA	NA	NA	0.001	0.035	0.024
Decreasing Effect	Marginal	IPW of MSM	1000	1.80	0.050	0.057	-0.67	0.057	0.057
		IPW of MSM (truncated)	1000	1.78	0.050	0.056	-0.65	0.056	0.055
		G-Formula	1000	1.81	0.044	0.023	-0.69	0.032	0.025
	Conditional	HA-MSM	1000	2.09	0.056	0.057	-0.74	0.031	0.031
		HA-MSM (truncated)	1000	2.08	0.056	0.057	-0.74	0.030	0.030
		G-Estimation	1000	NA	NA	NA	-2.50	0.076	0.024
Effect Modification	Marginal	IPW of MSM	1000	1.50	0.052	0.055	-0.77	0.036	0.038
		IPW of MSM (truncated)	1000	1.49	0.052	0.054	-0.76	0.034	0.035
		G-Formula	1000	1.51	0.041	0.021	-0.76	0.024	0.014
	Conditional	HA-MSM	1000	1.69	0.055	0.056	-0.70	0.029	0.028
		HA-MSM (truncated)	1000	1.68	0.055	0.056	-0.70	0.028	0.028
		G-Estimation	1000	NA	NA	NA	-1.62	0.056	0.023
Reversed Causal Pathway	Marginal	IPW of MSM	758	-1.81	5.47	0.81	0.12	0.99	0.11
		IPW of MSM (truncated)	1000	1.75	0.078	0.082	-0.53	0.13	0.11
		G-Formula	1000	1.71	0.054	0.027	-0.60	0.038	0.022
	Conditional	HA-MSM	727	2.15	1.13	0.39	-0.78	0.73	0.38
		HA-MSM (truncated)	1000	1.91	0.067	0.067	-0.62	0.044	0.043
		G-Estimation	1000	NA	NA	NA	-2.18	0.090	0.024

Scenario	Estimate	Method	n	Cumulative 2 Year Treatment Effect					
				Log Odds Ratio of Zero Count ($\sum_{i=1}^{2}\beta_i$)			Log Rate Ratio of Count ($\sum_{i=1}^{2}\gamma_i$)		
				Mean	Empirical SE	Model SE	Mean	Empirical SE	Model SE
Censoring	Marginal	IPW of MSM	1000	2.33	0.087	0.091	-0.90	0.065	0.063
		IPW of MSM (truncated)	1000	2.27	0.086	0.090	-0.88	0.062	0.061
		G-Formula	1000	2.37	0.069	0.039	-0.91	0.048	0.034
	Conditional	HA-MSM	1000	2.72	0.090	0.094	-0.94	0.053	0.052
		HA-MSM (truncated)	1000	2.71	0.089	0.093	-0.94	0.051	0.051
		G-Estimation	1000	NA	NA	NA	-3.30	0.14	0.15

TABLE C.5: Simulation study results of two-year population-average effects for count outcome ($D_t \rightarrow V_{t+1}$) (NA signifies that the method does not estimate that effect)

Scenario	Estimate	Method	n	Cumulative 3 Year Treatment Effect					
				Log Odds Ratio of Zero Count ($\sum_{i=1}^{3}\beta_i$)			Log Rate Ratio of Count ($\sum_{i=1}^{3}\gamma_i$)		
				Mean	Empirical SE	Model SE	Mean	Empirical SE	Model SE
Standard	Marginal	IPW of MSM	1000	2.86	0.099	0.10	-1.08	0.12	0.11
		IPW of MSM (truncated)	1000	2.83	0.099	0.10	-1.04	0.13	0.11
		G-Formula	1000	2.87	0.075	0.042	-1.09	0.069	0.056
	Conditional	HA-MSM	1000	3.34	0.10	0.10	-1.19	0.062	0.062
		HA-MSM (truncated)	1000	3.33	0.10	0.10	-1.18	0.061	0.061
		G-Estimation	1000	NA	NA	NA	-3.88	0.13	0.031
No Effect	Marginal	IPW of MSM	1000	-0.0009	0.048	0.055	0.003	0.037	0.040
		IPW of MSM (truncated)	1000	-0.027	0.048	0.054	0.028	0.034	0.036
		G-Formula	1000	-0.002	0.041	0.020	0.002	0.022	0.013
	Conditional	HA-MSM	1000	-0.001	0.055	0.057	0.001	0.028	0.028
		HA-MSM (truncated)	1000	-0.007	0.055	0.057	0.006	0.027	0.028
		G-Estimation	1000	NA	NA	NA	0.001	0.043	0.031
Decreasing Effect	Marginal	IPW of MSM	1000	1.97	0.071	0.074	-0.77	0.065	0.066
		IPW of MSM (truncated)	1000	1.94	0.071	0.074	-0.74	0.064	0.064
		G-Formula	1000	1.98	0.056	0.030	-0.78	0.040	0.028
	Conditional	HA-MSM	1000	2.39	0.078	0.079	-0.87	0.044	0.043
		HA-MSM (truncated)	1000	2.37	0.078	0.079	-0.87	0.043	0.042
		G-Estimation	1000	NA	NA	NA	-2.60	0.093	0.031

Scenario	Estimate	Method	n	Cumulative 3 Year Treatment Effect					
				Log Odds Ratio of Zero Count ($\sum_{i=1}^{3} \beta_i$)			Log Rate Ratio of Count ($\sum_{i=1}^{3} \gamma_i$)		
				Mean	Empirical SE	Model SE	Mean	Empirical SE	Model SE
Effect Modification	Marginal	IPW of MSM	1000	1.66	0.067	0.069	-0.79	0.047	0.048
		IPW of MSM (truncated)	1000	1.64	0.067	0.069	-0.76	0.043	0.043
		G-Formula	1000	1.68	0.054	0.028	-0.67	0.030	0.018
	Conditional	HA-MSM	1000	1.93	0.071	0.073	-0.77	0.039	0.039
		HA-MSM (truncated)	1000	1.92	0.071	0.073	-0.77	0.038	0.038
		G-Estimation	1000	NA	NA	NA	-1.78	0.071	0.031
Reversed Causal Pathway	Marginal	IPW of MSM	758	1.28	20.79	132.06	-0.60	0.90	0.13
		IPW of MSM (truncated)	1000	2.05	0.11	0.11	-0.67	0.14	0.10
		G-Formula	1000	2.01	0.071	0.038	-0.70	0.049	0.029
	Conditional	HA-MSM	727	3.51	2.76	1.54	-0.95	0.90	0.38
		HA-MSM (truncated)	1000	2.29	0.099	0.097	-0.74	0.065	0.062
		G-Estimation	1000	NA	NA	NA	-2.44	0.11	0.032
Censoring	Marginal	IPW of MSM	1000	2.87	0.16	0.15	-1.09	0.10	0.10
		IPW of MSM (truncated)	1000	2.77	0.16	0.15	-1.05	0.10	0.099
		G-Formula	1000	2.89	0.11	0.065	-1.13	0.076	0.050
	Conditional	HA-MSM	1000	3.35	0.16	0.16	-1.14	0.092	0.088
		HA-MSM (truncated)	1000	3.32	0.16	0.16	-1.13	0.090	0.087
		G-Estimation	1000	NA	NA	NA	-3.88	0.20	0.22

TABLE C.6: Simulation study results of three-year population-average effects for count outcome ($\overline{D}_t \rightarrow V_{t+1}$) (NA signifies that the method does not estimate that effect)

Scenario	Estimate	Method	n	2 Year Interaction Effect					
				Log Odds Ratio of Zero Count ($14\beta_{\text{int}_2}$)			Log Rate Ratio of Count ($14\gamma_{\text{int}_2}$)		
				$\overline{\beta}$	Empirical SE	Model SE	$\overline{\beta}$	Empirical SE	Model SE
No Effect	Conditional	HA-MSM	1000	0.003	0.040	0.039	0.004	0.012	0.011
		HA-MSM (truncated)	1000	0.002	0.039	0.039	0.005	0.012	0.011
		G-Estimation	1000	NA	NA	NA	-0.001	0.015	0.016
Standard (no effect modification)	Conditional	HA-MSM	1000	0.049	0.043	0.043	-0.010	0.023	0.019
		HA-MSM (truncated)	1000	0.046	0.041	0.041	-0.009	0.023	0.018
		G-Estimation	1000	NA	NA	NA	0.077	0.050	0.016
Effect Modification	Conditional	HA-MSM	1000	0.14	0.042	0.042	0.11	0.031	0.025
		HA-MSM (truncated)	1000	0.15	0.041	0.040	0.12	0.027	0.023
		G-Estimation	1000	NA	NA	NA	-0.072	0.036	0.016

TABLE C.7: Simulation study results of two-year interaction effects for count outcome (Results show change in effect of D_t on V_{t+1} per 14 change in V_t) (NA signifies that the method does not estimate that effect)

Scenario	Estimate	Method	n	3 Year Interaction Effect					
				Log Odds Ratio of Zero Count ($14\beta_{\text{int}_3}$)			Log Rate Ratio of Count ($14\gamma_{\text{int}_3}$)		
				$\overline{\beta}$	Empirical SE	Model SE	$\overline{\beta}$	Empirical SE	Model SE
No Effect	Conditional	HA-MSM	1000	0.00002	0.041	0.040	0.005	0.015	0.015
		HA-MSM (truncated)	1000	-0.002	0.041	0.040	0.007	0.014	0.014
		G-Estimation	1000	NA	NA	NA	0.00005	0.018	0.018
Standard (no effect modification)	Conditional	HA-MSM	1000	0.033	0.053	0.051	-0.012	0.041	0.028
		HA-MSM (truncated)	1000	0.030	0.051	0.050	-0.011	0.041	0.027
		G-Estimation	1000	NA	NA	NA	0.031	0.068	0.019
Effect Modification	Conditional	HA-MSM	1000	-0.023	0.049	0.048	-0.001	0.031	0.028
		HA-MSM (truncated)	1000	-0.025	0.048	0.047	0.001	0.030	0.027
		G-Estimation	1000	NA	NA	NA	-0.068	0.050	0.019

TABLE C.8: Simulation study results of three-year interaction effects for count outcome (Results show change in effect of $\overline{D}_t$ on V_{t+1} per 14 change in V_t) (NA signifies that the method does not estimate that effect)

Appendix D

Paper 3 - August 2018

This appendix contains the published version of the paper entitled "Investigating the effects of long-term dornase alfa use on lung function using registry data". It was accepted in the Journal of Cystic Fibrosis in August 2018. This paper is based on the work found in Chapters 6.

As the first author of this paper, I carried out the performed the analysis of the UK CF registry data, drafted all sections of the paper, produced all figures and tables, and prepared and approved the final version of the paper.

Student's Signature: 

Date: 29th January, 2019

Supervisor's Signature: 

Date: 29th January, 2019

ARTICLE IN PRESS

JCF-01693; No of Pages 8

ELSEVIER

Journal of Cystic Fibrosis xx (2018) xxx–xxx

Journal of Cystic Fibrosis

www.elsevier.com/locate/jcf

Original Article

Investigating the effects of long-term dornase alfa use on lung function using registry data

Newsome S.J. [a,*], Daniel R.M. [b], Carr S.B. [c], Bilton D. [d], Keogh R.H. [a]

[a] *Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK*
[b] *Division of Population Medicine, Cardiff University, Cardiff, UK*
[c] *Department of Paediatric Respiratory Medicine, Royal Brompton & Harefield NHS Foundation Trust, London, UK*
[d] *Faculty of Medicine, National Heart & Lung Institute, Imperial College London, London, UK*

Received 23 March 2018; revised 29 July 2018; accepted 2 August 2018
Available online xxxx

Abstract

Background: Dornase alfa (DNase) is one of the commonest cystic fibrosis (CF) treatments and is often used for many years. However, studies have not evaluated the effectiveness of its long-term use. We aimed to use UK CF Registry data to investigate the effects of one-, two-, three-, four- and five-years of DNase use on lung function to see if the benefits of short-term treatment use are sustained long term.
Methods: We analysed data from 4,198 people in the UK CF Registry from 2007 to 2015 using g-estimation. By controlling for time-dependent confounding we estimated the effects of long-term DNase use on percent predicted FEV_1 ($ppFEV_1$) and investigated whether the effect differed by $ppFEV_1$ at treatment initiation or by age.
Results: Considering the population as a whole, there was no significant effect of one-year's use of DNase; change in $ppFEV_1$ over one year was −0.1% in the treated compared to the untreated ($p = 0.51$) and this did not change with long-term use. However, treatment was estimated to be more beneficial in people with lower lung function ($p < 0.001$); those with $ppFEV_1 < 70\%$ at treatment initiation, showed an increase in lung function over one year that was sustained out to five years. The estimated effect of DNase did not depend on age ($p = 0.35$).
Conclusions: DNase improved lung function in individuals with reduced lung function, bringing a step-change in lung function, but no change in the slope of decline. There was no evidence for a benefit in lung function in those initiating treatment with $ppFEV_1 > 70\%$.

Keywords: DNase; Long-term treatment effect; Patient registry; UK Cystic Fibrosis Registry

1. Background

First licensed for use in the EU and the US in 1994, dornase alfa (DNase) is now one of the commonest cystic fibrosis (CF) treatments, used by almost 60% of people with CF in the UK in 2016 [1,2]. DNase is a mucolytic treatment administered via a nebulizer, decreasing the viscosity of sputum in the airways, aiming to aid in airway clearance, to improve lung function and decrease pulmonary exacerbations [3].

In a phase III clinical trial, using DNase once or twice daily over twenty-four weeks was shown to improve percent-predicted FEV_1 ($ppFEV_1$) [4]. Subsequently, a number of studies have examined the effect of longer-term use of DNase. For example, after 96 weeks of follow-up, treatment was shown to significantly improve $ppFEV_1$ in children aged 6 to 10 [5,6].

Most studies have focused on the absolute effect of DNase on lung function over a specified time period, i.e. on a step-change effect. Its impact on the rate of lung function decline is also important and this has been investigated in two studies. These studies showed that during DNase use the rate of lung function decline was less than the rate of decline in the same

* Corresponding author.
E-mail address: simon.newsome@lshtm.ac.uk (S.J. Newsome).

https://doi.org/10.1016/j.jcf.2018.08.004

patients prior to starting treatment and also less than the rate of decline in a comparator group of patients who never received treatment [7,8]. Only one study has attempted to evaluate the impact of using DNase for more than two years. This was a matched study with 76 patients where it was found that those receiving treatment over four years had a more gradual slope of decline in $ppFEV_1$ [9].

Until recently in the UK regional guidelines for CF use of DNase varied, but DNase tended to be recommended when a person's $ppFEV_1$ fell below 80%. However, in 2014 a national policy was approved allowing the use of DNase in anyone over six years of age. Thus, more recently, some centres have begun to routinely initiate DNase when a patient reaches six years of age. These differences in treatment practices provide an opportunity to use the UK CF Registry data to investigate the long-term effects of DNase in a diverse population.

There are, however, some challenges that must be addressed when attempting to use observational data for this purpose. The main issue when estimating treatment effects is confounding by indication, whereby more healthy individuals are less likely to receive treatment. A simple comparison of treated and untreated would therefore typically suggest that even an effective treatment is associated with worse outcomes. A further complexity of Registry data is that as it is longitudinal, not only do confounding variables affect both the outcome of interest and the probability of receiving treatment, but they are also themselves affected by whether the patient was receiving treatment or not in previous years. This issue is known as time-dependent confounding and traditional statistical methods will generally lead to biased results. There are several methods available that can deal with time-dependent confounding [10,11], including inverse probability weighted estimation of marginal structural models, g-computation formula and g-estimation. A recent investigation of the application of these methods using registry data showed that for a continuous outcome, such as lung function, g-estimation appeared to be the most reliable and flexible, in particular by accommodating estimation of treatment effect modification by a time-varying covariate [12].

In this paper, we aimed to use the UK CF Registry to investigate the effects of one-, two-, three-, four- and five-years of DNase use on lung function. Furthermore, we aimed to investigate whether there is evidence that treatment is more effective in younger people, as has previously been reported [7]. We also hypothesised that the effect of DNase may differ depending on lung function, and as such we examined whether there is evidence that the treatment effect is modified by previous measures of lung function.

2. Methods

2.1. UK cystic fibrosis registry

The UK Cystic Fibrosis Registry is a national database managed by the Cystic Fibrosis Trust. Each year people with CF attend an annual assessment at which data are collected on their current health as well as on the treatments they have received in the past year. More details about the registry can be found in the data resource profile [13].

People were eligible for inclusion in the study if they had at least two consecutive years of data in the Registry between 2007 and 2015, had not received DNase prior to 2007, had at least one year of treatment-free data, were at least six years old, had lung function data for at least two consecutive years and had not received a lung transplant. The first visit for everyone was therefore at a time when they were not and had never before received DNase treatment. Follow-up data were collected up to 2015, or were censored at death, transplant or loss to follow-up. If people who started receiving DNase stopped during their follow-up, they were censored at the time they stopped treatment. Fig. 1 shows a flow chart of the number of people included and excluded from the study population.

For this study, the primary outcome was change in lung function expressed as absolute change over time in $ppFEV_1$ calculated using the Global Lung Initiative (GLI) calculations [14]. Secondary analyses with ppFVC and $ppFEF_{25–75}$ as outcomes were also conducted. All outcomes are observed annually.

As well as previous measures of the outcome and DNase, we also adjusted for the following time-varying variables: annual number of hospital and home IV days, CF centre, other muco-

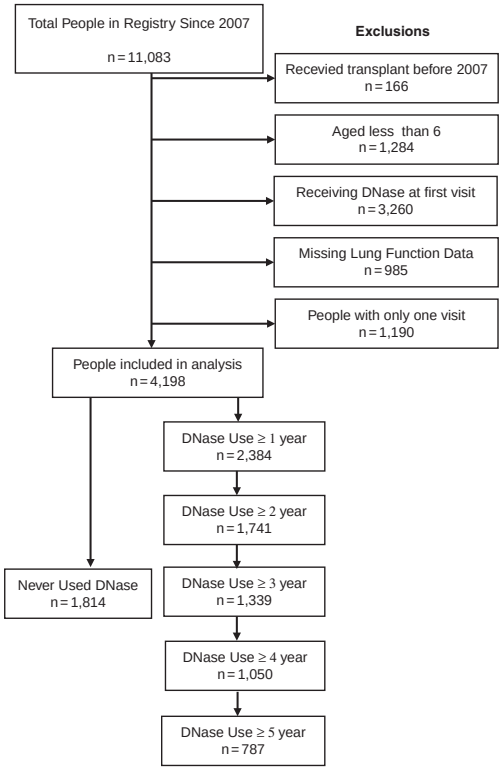


Fig. 1. Flowchart of people included in analysis.

Please cite this article as: Newsome SJ, et al, Investigating the effects of long-term dornase alfa use on lung function using registry data, J Cyst Fibros (2018), https://doi.org/10.1016/j.jcf.2018.08.004

active treatments, smoking status, CF related diabetes (CFRD), body mass index (BMI), allergic bronchopulmonary aspergillosis (ABPA) and infections (*P. aeruginosa*, *S. aureus* and *B. cepacia* complex); and the following non-time-varying variables: age, gender, ethnicity and genotype class (as defined by McKone et al) [15]. These were selected from the data collected in the Registry as variables that could affect both lung function and the decision to initiate DNase.

2.2. Statistical methods

In a preliminary exploration of the data, baseline covariates, i.e. those measured at the first visit, were summarised in all patients and separately in the following four subgroups: those who never received DNase, those who received DNase for at least one year, those who received DNase for at least five years, and those who stopped DNase during follow-up. Time-varying covariates were summarised across visits in the following five groups: at a patient's first visit, at subsequent visits among people not receiving DNase, at the first visit when people received DNase, at subsequent visits among people receiving DNase, and at the visit when people stopped receiving DNase. We also performed a univariable linear regression to estimate the crude difference in lung function between those receiving and not receiving treatment.

Data were then analysed using g-estimation to investigate the causal effect of up to five years of DNase use on lung function [16]. A recent paper by Vansteelandt and Sjolander has shown how to implement this method using standard statistical software [17]. Full details of this method can be found in the supplementary material to this paper. Briefly, the method works iteratively by first estimating a one-year treatment effect adjusting for the confounders of this effect, then estimating any additional two-year treatment effect adjusting for the relevant confounders of this effect and so on until the maximum follow-up time of interest (five years in our case). Thus, we obtain five different treatment effect estimates; these are the estimated differences in lung function between those using DNase for k years ($k = 1, \ldots, 5$) and those never receiving treatment, with the time-dependent confounding having been adjusted for at each stage.

Evidence that the treatment effect is modified by age or by previous lung function was investigated by including an interaction term between the relevant variable and DNase use at each visit.

For g-estimation, standard errors used to calculate p-values and 95% confidence intervals were estimated using the non-parametric bootstrap approach [18]. All analyses were performed using Stata (version 15.0, Stata Corp, College Station, Texas, USA).

3. Results

Overall, 4,198 people were included in the analysis, with a combined total of 20,923 annual assessments. The median number of follow-up visits per person was 5 (IQR 3–7). During follow-up, 2,384 (56.8%) people received DNase for at least one year, and most people who started using DNase continued to receive it indefinitely, with only 441 (18.5%) people stopping during follow-up. In total, 787 people had five or more consecutive years of DNase use.

Table 1 summarises the variables that were considered as confounders for the main analysis. When looking at this summary of the data, we found that those taking DNase tend to have worse symptoms. This is expected, because those needing treatment are expected to be in worse health. This finding is part of the phenomenon of confounding by indication. Particularly, the mean ppFEV$_1$ measured prior to visits when people were not receiving DNase was 80% compared to 72% prior to the visit when people started to receive DNase. Furthermore, the group who ever received DNase had a higher proportion of people with a high genotype class (76% vs 57%), a higher proportion with CFRD (23% vs 18%) and had more annual IV days (mean annual hospital IV days 10 vs 5, and mean annual home IV days 9 vs 5). Table 1 also summarises the group of people who stopped taking DNase during follow-up, but there were no noticeable differences between this group of people and those who continued to take DNase throughout follow-up.

Results from the univariable analysis (row 1 of Table 2) show that people receiving DNase had lower lung function compared to those not receiving treatment throughout follow-up. For example, at one year those receiving DNase had a lung function on average 7.1% lower than those not receiving treatment (95% CI -8.1% to −6.1%, $p < 0.0001$). Furthermore, the average annual decline in ppFEV$_1$ was 1.0% (95% CI 0.9% to 1.2%) in those not taking DNase compared to 1.3% (95% CI 1.1% to 1.5%) in those receiving DNase.

The results from the causal analysis using g-estimation are shown in Table 2 and Fig. 2. On a population level, we found no significant effect of DNase on ppFEV$_1$, with those on treatment estimated to have an absolute change in lung function of −0.1% over one year compared to someone not receiving treatment (95% CI -0.6% to 0.4%, $p = 0.65$). However, by year two this effect became more pronounced with an estimated difference of −0.7% (95%CI–1.4% to 0.05%, $p = 0.069$), and this trend continued out to year five, when it was estimated that on average receiving treatment for 5 years, compared to never receiving DNase, would result in an absolute change in lung function of −3.3% (95% CI -4.9% to −1.7%, $p < 0.0001$).

We found strong evidence that the effect of treatment differs depending on previous lung function ($p < 0.001$), with beneficial effects seen in those with low lung function. This is illustrated in Fig. 2, where beneficial effects are shown over the 5-year duration for people with baseline ppFEV$_1$ <70%, whereas for those with baseline ppFEV$_1$ >70% the rate of decline was steeper in those receiving treatment. For example, for an individual with baseline ppFEV$_1$ of 40%, initiating DNase was estimated to result in a lung function 1.6% higher after one year (95% CI 0.6% to 2.7%, $p = 0.002$) compared to not initiating DNase. Conversely, for an individual with a baseline ppFEV$_1$ of 80%, initiating DNase was estimated to result in a lung function 0.4% lower after one year (95% CI -0.1% to 0.9%, $p = 0.13$) compared to not initiating DNase.

Please cite this article as: Newsome SJ, et al, Investigating the effects of long-term dornase alfa use on lung function using registry data, J Cyst Fibros (2018), https://doi.org/10.1016/j.jcf.2018.08.004

Table 1
Descriptive Statistics of Key Variables. Data presented as mean (SD) for continuous variables and n (%) for binary and categorical variables.

Demographics (4,198 people)					
	All (n = 4,198)	Never used DNase (n = 1,814)	DNase use ≥ 1 year (n = 2,384)	DNase use ≥ 5 years (n = 787)	Stopped DNase during follow-up (n = 441)
Baseline age (Years)	19.8 (12.7)	21.4 (13.5)	18.7 (11.9)	19.1 (11.2)	21.0 (11.1)
Female	1923 (45.8)	806 (44.4)	1117 (46.9)	361 (45.9)	210 (47.6)
Caucasian	4061 (96.7)	1761 (97.1)	2300 (96.5)	758 (96.3)	432 (98.0)
Genotype class[a]					
High	2828 (67.4)	1031 (56.8)	1797 (76.1)	629 (79.9)	333 (75.5)
Low	563 (13.4)	353 (19.5)	210 (8.8)	51 (6.5)	35 (7.9)
None assigned	153 (3.6)	353 (19.5)	301 (12.6)	80 (10.2)	56 (12.7)
Missing	654 (15.6)	77 (4.2)	76 (3.2)	27 (3.4)	17 (3.9)
Longitudinal Data (20,923 observations)					
	First Visit [nobody using DNase] (n = 4,198)	Visits where not using DNase (n = 12,194)	First year using DNase (n = 2,384)	Subsequent years using DNase (n = 5,904)	Year after stopping DNase (n = 441)
Previous $ppFEV_1$	[d]	80.1 (20.9)	72.0 (21.8)	67.1 (21.9)	66.8 (22.6)
Annual change in $ppFEV_1$	[d]	−1.0 (9.9)	−0.4 (12.5)	−1.6 (9.5)	−1.7 (9.5)
Previous ppFVC[b]	[d]	90.1 (17.1)	84.5 (18.4)	81.4 (18.6)	81.3 (19.0)
Annual change in ppFVC[b]	[d]	−0.7 (10.1)	−0.3 (12.1)	−1.2 (10.1)	−1.3 (9.8)
Previous $ppFEF_{25-75}$[c]	[d]	71.4 (31.2)	63.9 (30.1)	58.1 (27.4)	59.3 (32.8)
Annual change in $ppFEF_{25-75}$[c]	[d]	−1.3 (21.7)	0.1 (26.7)	−1.0 (19.4)	−5.4 (22.9)
BMI (z-score)	0.13 (1.14)	0.29 (1.18)	0.00 (1.13)	−0.02 (1.14)	−0.02 (1.20)
Smoker					
No	3058 (72.8)	10,567 (86.7)	2138 (89.7)	5310 (89.9)	387 (87.8)
Occasionally	33 (0.8)	126 (1.0)	19 (0.8)	45 (0.8)	6 (1.4)
<1 packet per day	49 (1.2)	256 (2.1)	24 (1.0)	47 (0.8)	12 (2.7)
≥1 packet per day	20 (0.5)	69 (0.6)	6 (0.3)	14 (0.2)	3 (0.7)
Missing	1038 (24.7)	1176 (9.6)	197 (8.3)	488 (8.3)	33 (7.5)
Infections					
P. aeruginosa	2453 (58.4)	6460 (53.0)	947 (39.7)	2136 (36.2)	164 (37.2)
S. aureus	2643 (63.0)	7170 (58.8)	1324 (55.6)	3504 (59.3)	250 (56.7)
B. cepacia complex	80 (1.9)	334 (2.7)	86 (3.6)	271 (4.6)	17 (3.9)
ABPA	215 (5.1)	857 (7.0)	274 (11.5)	909 (15.4)	53 (12.0)
CFRD					
No	2226 (53.0)	7456 (61.1)	1360 (57.0)	3083 (52.2)	224 (50.8)
Yes	504 (12.0)	2150 (17.6)	550 (23.1)	1812 (30.7)	125 (28.3)
Missing	1468 (35.0)	2588 (21.2)	474 (19.9)	1009 (17.1)	92 (20.9)
Annual hospital IV days	4.1 (11.0)	4.5 (12.2)	10.4 (18.9)	12.4 (23.7)	11.6 (22.8)
Annual home IV days	4.1 (12.5)	4.7 (13.7)	9.0 (18.9)	10.5 (19.8)	10.7 (23.8)
Other muco-active treatments	200 (4.8)	1578 (12.9)	502 (21.1)	1940 (32.9)	136 (30.8)
Acetylcysteine	15 (0.4)	133 (1.1)	41 (1.7)	93 (1.6)	4 (0.9)
Hypertonic saline	184 (4.4)	1440 (11.8)	469 (19.7)	1831 (31.0)	130 (29.5)
Mannitol	2 (0.05)	39 (0.3)	13 (0.5)	124 (2.1)	7 (1.6)

[a] Genotype class as defined by McKone et al. [15].
[b] FVC data based on 20,714 observations,
[c] FEF_{25-75} data based on 3320 observations,
[d] First visit, so no previous measures available.

Table 2
Estimated Effect of DNase Use on percent predicted FEV_1 compared to never taking DNase.

	Years of DNase use									
	1		2		3		4		5	
	Est	95% CI	Est	95% CI	Est	95% CI	Est	95% CI	Est	95% CI
Results from univariable analysis										
Estimated population average effect										
	−7.1	−8.1, −6.1	−8.6	−9.7, −7.4	−10.6	−11.9, −9.3	−12.6	−14.0, −11.1	−14.2	−15.9, −12.6
Results from G-Estimation Analysis										
Estimated population average effect										
	−0.1	−0.6, 0.4	−0.7	−1.4, 0.05	−1.4	−2.3, −0.6	−2.5	−3.8, −1.3	−3.3	−4.9, −1.7
Estimated effect by baseline $ppFEV_1$										
20%	2.6	1.1, 4.2	2.0	−0.1, 4.2	0.9	−1.7, 3.6	3.0	−0.5, 6.5	4.1	−0.5, 8.7
40%	1.6	0.6, 2.7	1.0	−0.4, 2.5	0.1	−1.8, 1.9	0.9	−1.5, 3.3	1.2	−1.9, 4.4
60%	0.6	−0.01, 1.2	0.1	−0.9, 1.0	−0.8	−1.9, 0.4	−1.2	−2.7, 0.3	−1.6	−3.5, 0.3
80%	−0.4	−0.9, 0.1	−0.9	−1.6, −0.2	−1.6	−2.6, −0.7	−3.3	−4.5, −2.0	−4.5	−6.2, −2.8
100%	−1.4	−2.2, −0.6	−1.9	−3.0, −0.8	−2.5	−3.9, −1.0	−5.4	−7.4, −3.4	−7.3	−10.0, −4.6
Estimated change in effect per 10% change in baseline $ppFEV_1$										
	−0.5	−0.8, −0.2	−0.5	−0.9, −0.1	−0.4	−0.9, 0.003	−1.0	−1.7, −0.4	−1.4	−2.2, −0.6

The results from the g-estimation analysis provide information on whether any impact of DNase on lung function is in the form of a step change or whether the treatment modifies the slope of decline. In the groups for which we found a beneficial effect of treatment, the results were consistent with DNase resulting in a one-off step change in lung function, rather than a change in overall trajectory. Full results can be seen in Table 2, but taking individuals who start treatment with a baseline $ppFEV_1$ of 40% as an example, the estimated difference in absolute change in lung function at 5 years was of 1.2% (95% CI -1.9% to 4.4%, $p = 0.44$), very similar to the 1.6% benefit seen at one year.

We found no evidence that the effect of treatment on lung function differed depending on age at treatment initiation, $p = 0.61$. (Results of this analysis can be found in the supplementary material).

We also performed analyses to investigate the effect of DNase on FVC and $FEF_{25–75}$ and these results can be seen in supplementary material. The results from these analyses were broadly similar to the findings from the FEV_1 analysis.

4. Discussion

We used UK CF Registry data to estimate long-term effects of DNase use, controlling for confounding by indication by using state-of-the-art statistical methodology not previously applied to CF registry data. In our study, for individuals with a reduced lung function and not using DNase, we have shown that initiating DNase treatment and using it for one year brings a benefit such that $ppFEV_1$ is higher after one year than it would have been had those individuals not initiated DNase treatment. This beneficial effect appeared to remain with continued use of DNase out to five years, but with no overall modification of the lung function trajectory, as the estimated effect remained stable between years one and five.

One reason why it is important to estimate long-term effects of treatment is to see if a treatment modifies overall lung-function trajectory or just provides a one-off increase [19,20]. Fig. 3 shows two hypothetical lung-function trajectories for an individual treated with DNase compared to the lung-function trajectory of someone not receiving treatment. These trajectories show how treatment could either be disease modifying, where the lung function of those receiving treatment continues to grow wider apart from the lung function of those not receiving treatment, or alternatively how the overall lung-function trajectory could remain unchanged after an initial increase.

Crude comparisons between those who received and did not receive DNase clearly indicate that there is confounding by indication, such that individuals taking DNase tend to have worse health status than those not taking DNase. If this is not appropriately handled in the analysis, any estimates of the treatment effect would not have a causal interpretation. In this study, we made use of appropriate statistical methods to address the confounding by indication, accounting for the longitudinal setting, thereby showing how registries can be used to evaluate the long-term effects of treatment. Randomised controlled trials (RCT) are the gold standard for establishing treatment efficacy, but as previously discussed, it is preferable for a CF treatment to alter lung function trajectory rather than to provide a one-off improvement, and assessing change in trajectory requires longer follow-up than would typically be feasible in trials [21]. The analyses used in this paper take advantage of clear heterogeneity in treatment practices as the proportion of patients receiving DNase at individual CF centres ranges from <20% to >80% [2]. As the groups who receive and do not receive DNase include individuals with wide-ranging clinical characteristics, the statistical methods used in this paper can correct for confounding by indication as long as data on all confounders have been collected. We were able to adjust for a large number of variables using the data available in the UK CF

Please cite this article as: Newsome SJ, et al, Investigating the effects of long-term dornase alfa use on lung function using registry data, J Cyst Fibros (2018), https://doi.org/10.1016/j.jcf.2018.08.004

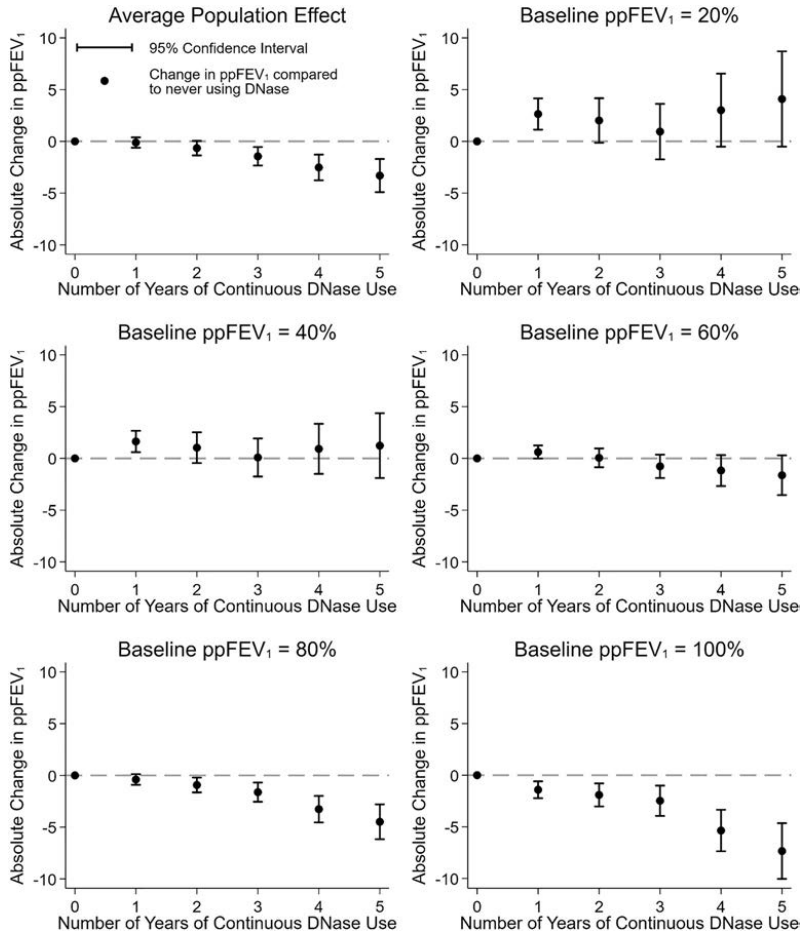


Fig. 2. Estimated mean difference in the change in $ppFEV_1$ between dornase alfa users compared to non-users (dashed grey line), by years of treatment. The top-left figure shows the estimated overall average population effect, with the remaining figures highlighting the differential effects based on $ppFEV_1$ at treatment initiation. The horizontal dashed grey line is the line indicating no treatment effect.

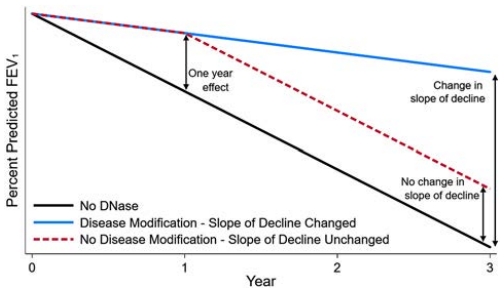


Fig. 3. Examples of possible lung function trajectories depending on treatment effect through time.

Registry, but it is not possible to verify whether confounding by indication has ever been completely dealt with and the causal interpretation of our results therefore depends on the assumption of no unmeasured confounding [22].

A previous registry study by Hodson et al. of the European Epidemiologic Registry of Cystic Fibrosis estimated a one-year treatment effect of DNase on ppFEV1 of 3.6% (95% CI 1.8% to 5.3%) and a two-year treatment effect of 2.5% (0.7% to 4.4%) [7]. These are larger population-average treatment effect estimates than obtained in our study, but the population for those studies had lower average baseline $ppFEV_1$, who were the patients we found benefitted most from treatment.

Only two previous studies have investigated the effects of DNase in people with $ppFEV_1 > 80\%$ and only one of these

Please cite this article as: Newsome SJ, et al, Investigating the effects of long-term dornase alfa use on lung function using registry data, J Cyst Fibros (2018), https://doi.org/10.1016/j.jcf.2018.08.004

included lung function as an outcome [23,24]. That study only administered DNase for four weeks and found no effect of treatment on ppFEV$_1$: treatment effect 0.1% increase in ppFEV$_1$ [24]. In our study, for individuals with higher lung function, those on DNase treatment had steeper trajectories of lung function decline than comparable individuals not receiving treatment. This may suggest that it would be more beneficial, in terms of lung function outcomes, to wait until lung function starts to decline before initiating DNase. However, as with all observational studies, it is possible that unobserved confounders affect these results. With the rich Registry data, we believe we have accounted for the covariates that could affect both lung function and the probability of receiving DNase treatment to account for confounding by indication, but it is possible that there are some unmeasured health-related variables that affected the decisions to initiate treatment in these individuals. For example, although an individual may have had a high lung function measurement at the previous annual assessment, they may have been beginning to show signs of lung function deterioration that was not picked up by having only one lung function measure per year.

Upon initiation of DNase, most people continue to receive the treatment indefinitely, with only very small numbers of people stopping treatment. Due to this, with the sample size available, it was not possible to estimate the effect of stopping treatment. However, as we observed that treatment only appeared to be beneficial in individuals with reduced lung function, future studies could investigate whether treatment needs to be continued in people who recover to higher levels of lung function.

A major strength of this study is the use of the UK CF Registry data, which are collected at regular intervals according to a standardised protocol. The data include a large number of variables that we could account for as potential confounders. One of the main limitations of this study is that there are no data available on levels of adherence to treatment. It is known to be particularly hard to measure adherence levels, but previous studies have estimated that average adherence levels to nebulised therapies, such as DNase, may range between 60% and 70% [25,26]. Specifically, for a longitudinal study, we may expect adherence levels to be higher at treatment initiation and decrease through time, which may partly explain why the observed effects are not as pronounced as in RCT, where adherence would typically be higher [27].

The UK CF Registry contains data on over 99% of people with CF in the UK, but a large number of these people were excluded from this analysis. The majority of these exclusions were due to people who were already receiving DNase prior to 2007 or people aged under six, and this is not considered to be a source of bias. Our results are applicable to people aged over six and estimate the effects of the first five years of DNase use. The method of g-estimation relies on having equally spaced visits, and therefore a number of people were censored due to missing lung function measurements during follow-up. We used so-called 'censoring weights' within g-estimation, which reweights individuals who remain in the study to account for those lost to follow-up.

The analyses presented in this paper use the FEV$_1$ measure obtained on the date of a patient's annual review. An alternative approach would be to use the best FEV$_1$ measure obtained since the last annual review. However, best FEV$_1$ has only been collected in the UK CF Registry since 2012 and the aim of this paper, to estimate the effects of long-term DNase use, would not be possible with the number of best FEV$_1$ observations available. The unknown timing of the best FEV1 measure relative to the other measures would also present additional challenges for the analysis. According the Registry protocol, annual reviews take place at a time when the patient is stable, and ongoing validation procedures indicate good adherence to this protocol [28]. Furthermore, there is no evidence to suggest that those receiving DNase are more likely to have their annual review during an unstable period compared to those not receiving DNase, meaning that using the FEV$_1$ measure obtained during the annual assessment should not result in any bias.

It is also acknowledged that spirometry measures, such as FEV$_1$, may not be sufficiently sensitive to detect the early stages of lung function decline, and it has been suggested that other measures such as lung clearance index (LCI) may give a better indication of early lung function deterioration [29,30]. Unfortunately, LCI is not collected in the UK CF Registry, so it was not possible to investigate this. However, FEF$_{25–75}$ is collected in the registry, albeit less reliably than FEV$_1$ (3320 FEF$_{25–75}$ measurements compared to 20,923 FEV$_1$ measurements in this analysis), and the results showed similar findings to the findings with FEV$_1$, but with much larger confidence intervals, reflecting the lack of measurements and increased variability of this measure (see Supplement).

In conclusion, we have shown a beneficial long-term effect of DNase in people with reduced lung function, but with no overall change in lung-function trajectory. There is a differential effect of treatment based on lung function at treatment initiation with no improvements in lung function seen in individuals initiating treatment with ppFEV$_1$ higher than 70%. Finally, this study highlights the potential of registries in investigating the effects of long-term treatment use and that issues of confounding-by-indication can be addressed with appropriate statistical methods.

Funding & competing interests

SJN, SBC, DB and RHK are part of the CF-EpiNet Strategic Research Centre supported by a grant from the Cystic Fibrosis Trust. RMD is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (107,617/Z/15/Z). SBC is a principal investigator (PI) for a CF Trust Registry-based pharmacovigilance (PASS) study for Pharmaxis, and is a PI for a CF Trust Vertex-funded Registry-based observational study. DB is a principal investigator for two CF Trust Registry-based pharmacovigilance studies for Vertex and Teva pharmaceutical companies. RHK is supported by a Medical Research Council Methodology Fellowship (MR/M014827/1).

Please cite this article as: Newsome SJ, et al, Investigating the effects of long-term dornase alfa use on lung function using registry data, J Cyst Fibros (2018), https://doi.org/10.1016/j.jcf.2018.08.004

Acknowledgements

We thank people with cystic fibrosis and their families for consenting to their data being held in the UK CF Registry, and NHS teams in CF centres and clinics for the input of data into the Registry. We also thank the UK Cystic Fibrosis Trust and the Registry Steering Committee for access to anonymized UK CF Registry data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcf.2018.08.004.

References

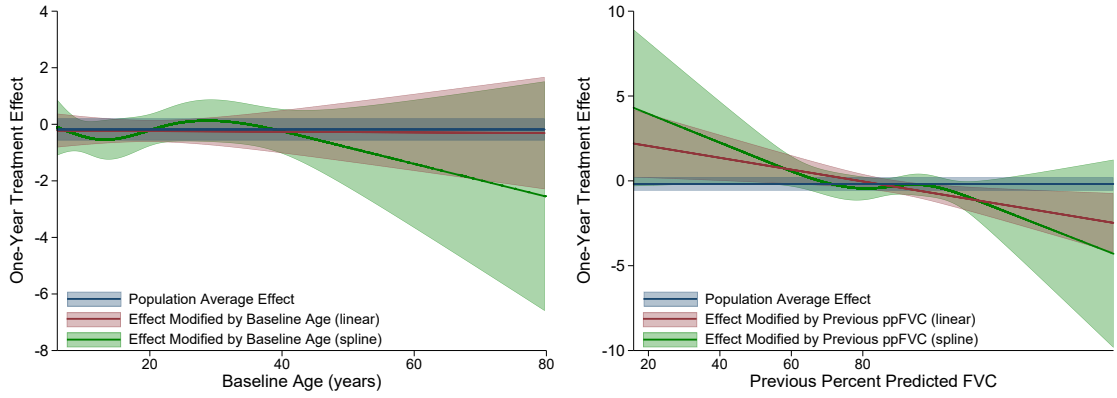
[1] Pulmozyme. 2500 U/2.5ml, nebuliser solution. Available from: Internet https://www.medicines.org.uk/emc/medicine/1723.
[2] UK Cystic Fibrosis Registry Annual Data Report 2016 Internet. 2016. Available from:file:///C:/Users/User/Downloads/2016 Registry Annual Data Report (1).pdf.
[3] Bush A, Bilton D, Hodson M. Hodson and Geddes' Cystic Fibrosis. Taylor & Francis; 2015; 694 Available from: Internet https://books.google.com/books?id=x-3UMgEACAAJ&pgis=1.
[4] Fuchs H, Borowitz D, Christiansen D, Morris E, Nash M, Ramsey B, et al. Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. N Engl J Med 1994;331(10):637–42.
[5] Quan JM, Tiddens HA, Sy JP, SG McKenzie, Montgomery MD, Robinson PJ, et al. A two-year randomized, placebo-controlled trial of dornase alfa in young patients with cystic fibrosis with mild lung function abnormalities. J Pediatr 2001;139(6):813–20 Internet Available from: http://www.ncbi.nlm.nih.gov/pubmed/11743506.
[6] Robinson PJ. Dornase alfa in early cystic fibrosis lung disease. Pediatr Pulmonol 2002;34(3):237–41 Internet Available from: http://www.ncbi.nlm.nih.gov/pubmed/12203856.
[7] Hodson ME, McKenzie S, Harms HK, Koch C, Mastella G, Navarro J, et al. Dornase alfa in the treatment of cystic fibrosis in Europe: a report from the epidemiologic registry of cystic fibrosis. Pediatr Pulmonol 2003; 36(5):427–32 Internet Available from: http://www.ncbi.nlm.nih.gov/pubmed/14520726.
[8] Konstan MW, Wagener JS, Pasta DJ, Millar SJ, Jacobs JR, Yegin A, et al. Clinical use of dornase alpha is associated with a slower rate of FEV1 decline in cystic fibrosis. Pediatr Pulmonol 2011;46(6):545–53 Available from:Internet http://www.ncbi.nlm.nih.gov/pubmed/21438174.
[9] Shah PL, Conway S, Scott SF, Rainisio M, Wildman M, Stableforth D, et al. A case-controlled study with dornase alfa to evaluate impact on disease progression over a 4-year period. Respiration 2001;68(2):160–4 Available from: Internet http://www.ncbi.nlm.nih.gov/pubmed/11287830.
[10] Daniel R, Cousens S, De Stavola B, Kenward M, Sterne J. Methods for dealing with time-dependent confounding. Stat Med 2013;32(9): 1584–618 Available from: Internet https://doi.org/10.1002/sim.5686.
[11] Hernán MA, Robins JM. Causal Inference. Boca Raton chapman & hall/CRC; 2018 Available from: Internet https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.
[12] Newsome S, Keogh R, Daniel R. Estimating long-term treatment effects in observational data: a comparison of the performance of different methods under real-world uncertainty. Stat Med 2018;37:2367–90.
[13] Taylor-Robinson D, Archangelidi O, Carr S, Cosgriff R, Gunn E, Keogh R, et al. Data Resource Profile: the UK cystic fibrosis registry. Int J Epidemiol 1 Feb 2018;47(1):9e–10e. https://doi.org/10.1093/ije/dyx196.
[14] Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J 2012;40(6):1324–43.
[15] McKone E, Goss C, Aitken M. CFTR genotype as a predictor of prognosis in cystic fibrosis. Chest 2006;130(5):1441–7.
[16] Robins J, Blevins D, Ritter G, Wulfsohn M. G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology 1992;3:319–36.
[17] Vansteelandt S, Sjolander A. Revisiting g-estimation of the effect of a time-varying exposure subject to time-varying confounding. Epidemiol Methods 2016;5(1):37–56 Available from:Internet http://www.degruyter.com/view/j/em.ahead-of-print/em-2015-0005/em-2015-0005.xml.
[18] Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1986;1(1): 54–77.
[19] Konstan MW, Ratjen F. Effect of dornase alfa on inflammation and lung function: potential role in the early treatment of cystic fibrosis. J Cyst Fibros 2012;11(2):78–83 Available from: Internet https://doi.org/10.1016/j.jcf.2011.10.003.
[20] Konstan MW. Dornase alfa and progression of lung disease in cystic fibrosis. Pediatr Pulmonol 2008;43(S9):S24–8 Available from: Internet https://doi.org/10.1002/ppul.20858.
[21] Szczesniak R, Heltshe S, Stanojevic S, Mayer-Hamblett N. Use of FEV1 in cystic fibrosis epidemiologic studies and clinical trials: a statistical perspective for the clinical researcher. J Cyst Fibros 2017;16(3):318–26.
[22] Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data. Am J Public Health 2018;108(5): 616–9.
[23] Paul K, Rietschel E, Ballmann M, Griese M, Worlitzsch D, Shute J, et al. Effect of treatment with dornase alpha on airway inflammation in patients with cystic fibrosis. Am J Respir Crit Care Med 2004;169:719–25.
[24] Amin R, Subbarao P, Lou W, Jabar a, Balkovec S, Jensen R, et al. The effect of dornase alfa on ventilation inhomogeneity in patients with cystic fibrosis. Eur Respir J 2011;37(4):806–12.
[25] Shakkottai A, Kidwell K, Townsend M, Nasr S. A five-year retrospective analysis of adherence in cystic fibrosis. Pediatr Pulmonol 2015;50(12): 1224–9.
[26] Quittner A, Zhang J, Marynchenko M, Chopra P, Signorovitch J, Yushkina Y, et al. Pulmonary medication adherence and health-care use in cystic fibrosis. Chest 2014;146(1):142–51 Available from: Internet https://doi.org/10.1378/chest.13-1926.
[27] Kettler L, Sawyer S, Winefield H, Greville H. Determinants of adherence in adults with cystic fibrosis. Thorax 2002;57(5):459–64 Available from: Internet http://thorax.bmj.com/cgi/doi/10.1136/thorax.57.5.459.
[28] Hoo Z, Curley R, Campbell M, Walters S, Wildman M. The importance of data issues when comparing cystic fibrosis registry outcomes between countries: are annual review FEV 1 in the UK only collected when subjects are well? J Eval Clin Pract 2018:1–7 Available from: Internet https://doi.org/10.1111/jep.12967.
[29] Davies J, Cunningham S, Alton E, Innes J. Lung clearance index in CF: a sensitive marker of lung disease severity. Thorax 2008;63(2):96–7 Available from: Internet http://thorax.bmj.com/cgi/doi/10.1136/thx.2007.088575.
[30] Pittman J, Rosenfeld M. Elementary, my dear Watson! The accumulating evidence for the lung clearance index in monitoring early cystic fibrosis lung disease. Am J Respir Crit Care Med 2017;195(9):1131–2.

Please cite this article as: Newsome SJ, et al, Investigating the effects of long-term dornase alfa use on lung function using registry data, J Cyst Fibros (2018), https://doi.org/10.1016/j.jcf.2018.08.004

Appendix E

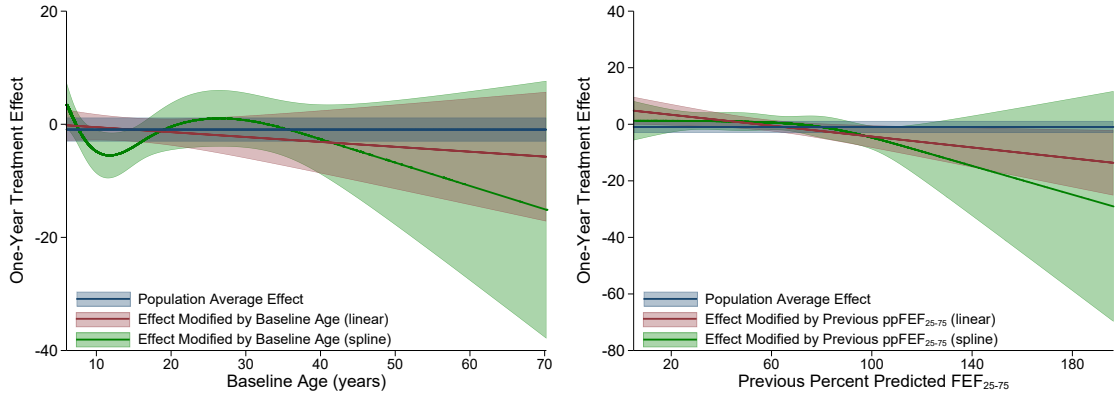
Supplementary Tables & Figures to Chapter 6

Here we present the results of the analyses with percent predicted FVC and percent predicted FEF_{25-75} as outcomes. These are both measures of lung function similar to $ppFEV_1$. The analyses presented are identical to the analysis of $ppFEV_1$, starting with an analysis of the one-year effect using SCMM, including estimating effect modification by baseline age or previous lung function, and then estimating the one-, two-, three-, four- and five-year effects using all available methods.



(A) (B)

FIGURE E.1: Estimated one-year effect of DNase on ppFVC depending on baseline age or previous ppFVC



(A) (B)

FIGURE E.2: Estimated one-year effect of DNase on $ppFEF_{25-75}$ depending on baseline age or previous $ppFEF_{25-75}$

Method		1 Year Treatment Effect		
		Est.	95% CI	p
Population Average		-0.18	-0.57, 0.21	0.36
Modified by Age	Intercept	-0.21	-0.97, 0.54	0.58
	Interaction	-0.012	-0.34, 0.31	0.94
Modified by Previous ppFVC	Intercept	2.75	0.33, 5.17	0.026
	Interaction	-0.35	-0.62, -0.075	0.012

TABLE E.1: Estimated one-year effects of DNase on ppFVC (The intercept term is the estimated effect if the age or previous ppFVC is 0, and the interaction term is the change in effect per 10 increase in the modifier)

Method		1 Year Treatment Effect		
		Est.	95% CI	p
Population Average		-0.93	-3.04, 1.17	0.38
Modified by Age	Intercept	0.34	-3.29, 3.97	0.85
	Interaction	-0.87	-2.89, 1.16	0.40
Modified by Previous $ppFEF_{25-75}$	Intercept	5.29	0.044, 10.53	0.048
	Interaction	-0.96	-1.79, -0.14	0.022

TABLE E.2: Estimated one-year effects of DNase on $ppFEF_{25-75}$ (The intercept term is the estimated effect if the age or previous $ppFEF_{25-75}$ is 0, and the interaction term is the change in effect per 10 increase in the modifier)

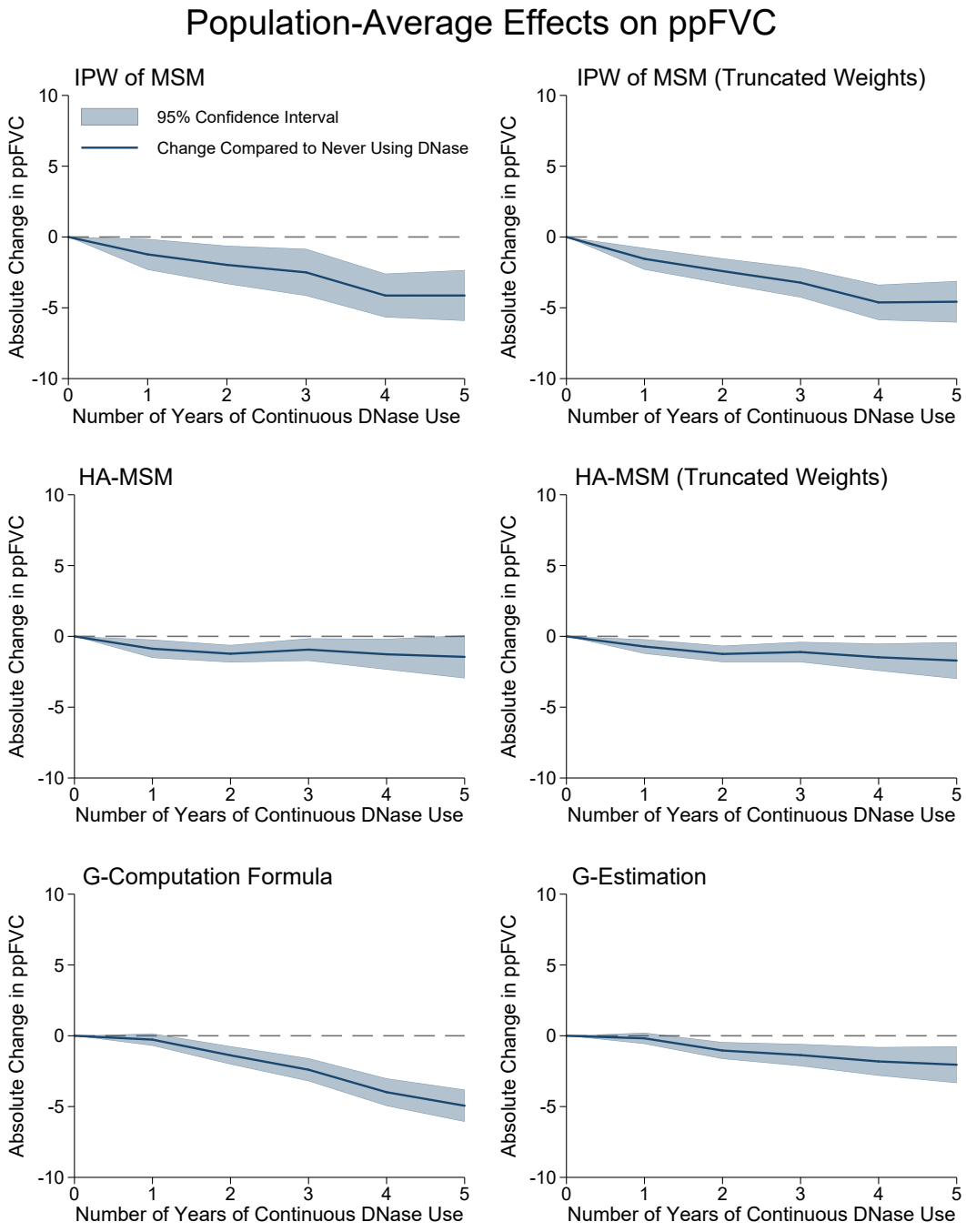


FIGURE E.3: Estimated population-average effects of DNase on ppFVC

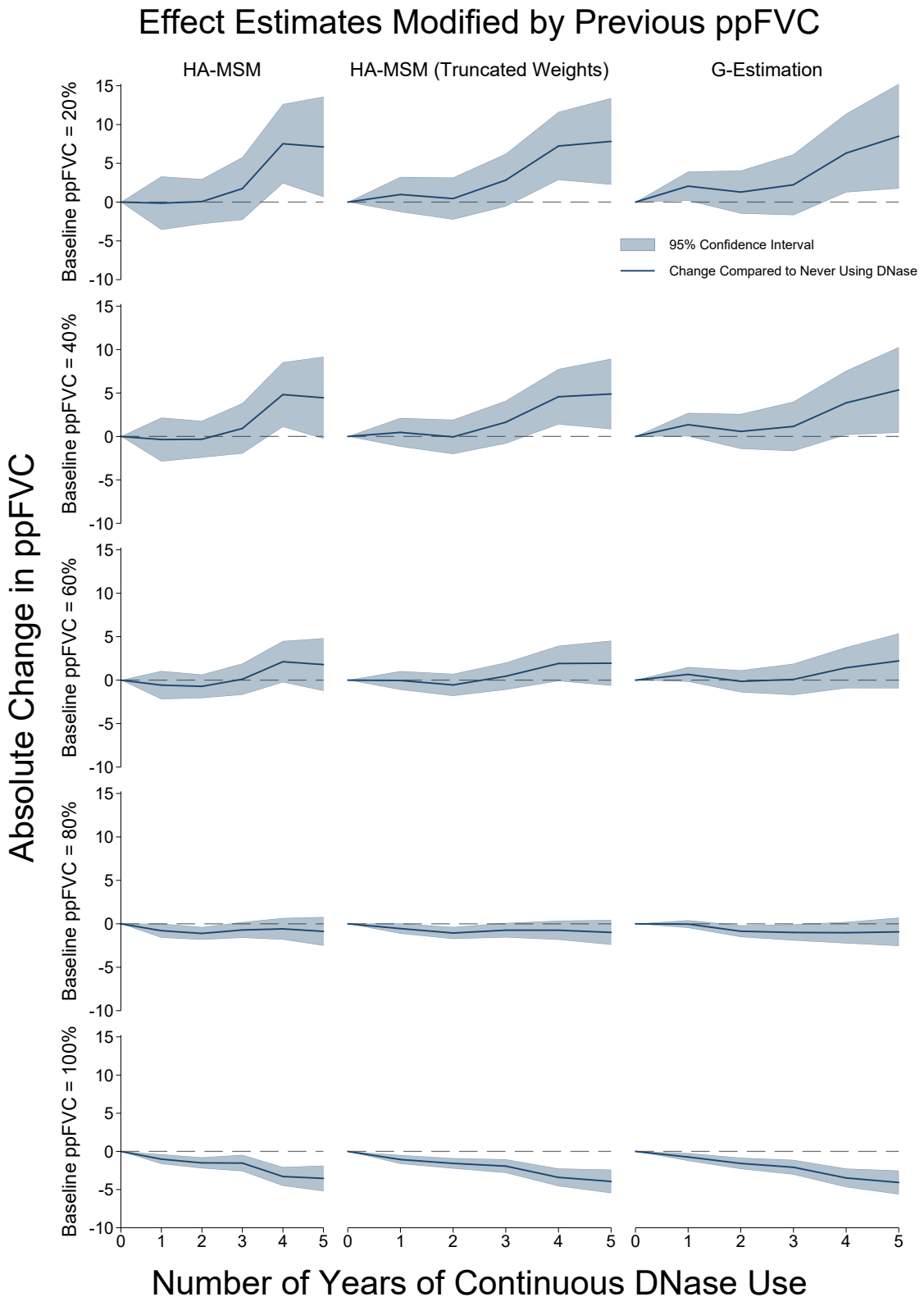


FIGURE E.4: Estimated effects of DNase on ppFVC modified by previous ppFVC

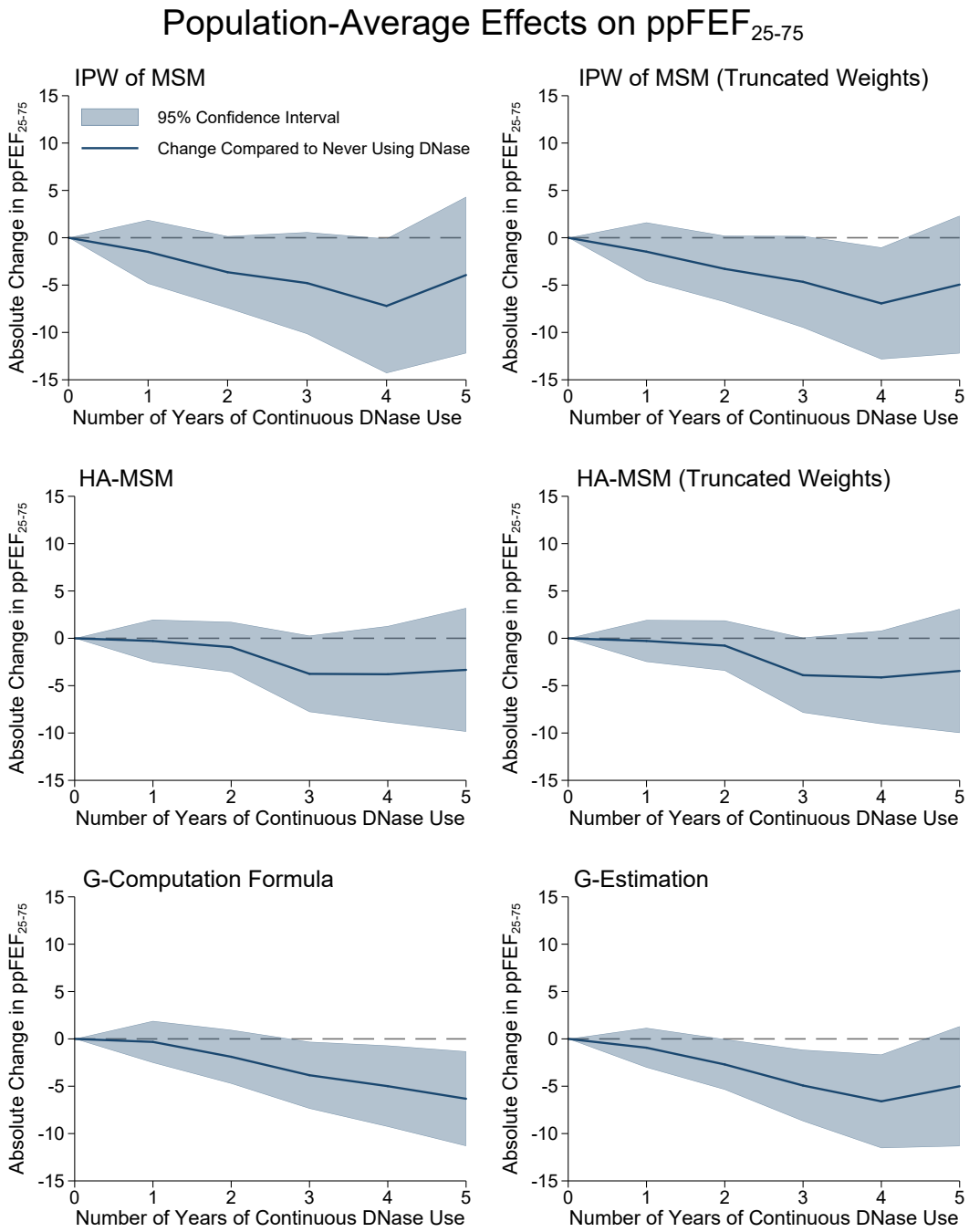


FIGURE E.5: Estimated population-average effects of DNase on ppFEF_{25-75}

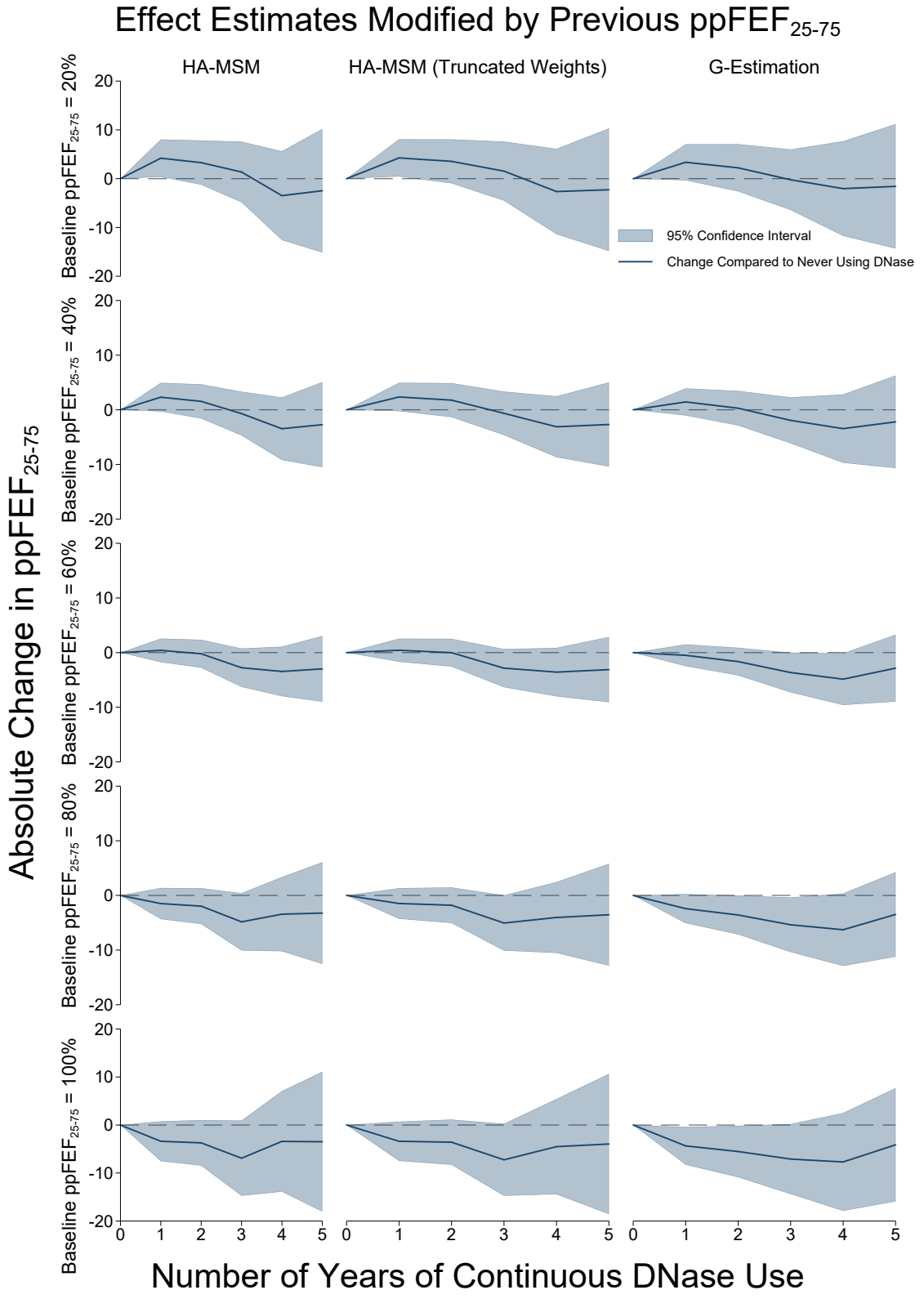


FIGURE E.6: Estimated effects of DNase on $ppFEF_{25-75}$ modified by previous $ppFEF_{25-75}$

Method	1 Year Treatment Effect			Cumulative 2 Year Treatment Effect			Cumulative 3 Year Treatment Effect		
	Est.	95% CI	p	Est.	95% CI	p	Est.	95% CI	p
IPW of MSM	-1.23	-2.32, -0.14	0.027	-1.97	-3.32, -0.62	0.004	-2.50	-4.16, -0.84	0.003
IPW of MSM (truncated)	-1.54	-2.31, -0.77	<0.0001	-2.41	-3.30, -1.51	<0.0001	-3.22	-4.27, -2.17	<0.0001
HA-MSM	-0.87	-1.52, -0.23	0.008	-1.22	-1.84, -0.61	<0.0001	-0.93	-1.73, -0.14	0.021
HA-MSM (truncated)	-0.72	-1.23, -0.21	0.006	-1.24	-1.82, -0.66	<0.0001	-1.10	-1.82, -0.39	0.003
G-Formula	-0.27	-0.71, 0.16	0.22	-1.38	-2.02, -0.74	<0.0001	-2.40	-3.21, -1.58	<0.0001
G-Estimation	-0.18	-0.58, 0.22	0.37	-1.04	-1.64, -0.45	<0.001	-1.36	-2.14, -0.58	<0.001
Method	Cumulative 4 Year Treatment Effect			Cumulative 5 Year Treatment Effect					
	Est.	95% CI	p	Est.	95% CI	p			
IPW of MSM	-4.13	-5.67, -2.59	<0.0001	-4.13	-5.92, -2.33	<0.0001			
IPW of MSM (truncated)	-4.62	-5.87, -3.37	<0.0001	-4.57	-6.03, -3.11	<0.0001			
HA-MSM	-1.26	-2.35, -0.18	0.023	-1.45	-2.96, 0.071	0.062			
HA-MSM (truncated)	-1.47	-2.44, -0.51	0.003	-1.71	-3.00, -0.41	0.010			
G-Formula	-3.98	-4.96, -3.01	<0.0001	-4.94	-6.08, -3.80	<0.0001			
G-Estimation	-1.81	-2.82, -0.80	<0.001	-2.04	-3.34, -0.75	0.002			

TABLE E.3: Estimated cumulative population average effects of DNase on ppFVC

Method		1 Year Treatment Effect			Cumulative 2 Year Treatment Effect			Cumulative 3 Year Treatment Effect		
		Est.	95% CI	p	Est.	95% CI	p	Est.	95% CI	p
HA-MSM	Intercept	0.083	-4.28, 4.44	0.97	0.46	-3.23, 4.14	0.81	2.56	-2.63, 7.75	0.33
	Interaction	-0.11	-0.57, 0.36	0.65	-0.20	-0.60, 0.21	0.34	-0.41	-0.99, 0.18	0.17
HA-MSM (truncated)	Intercept	1.49	-1.39, 4.37	0.31	0.96	-2.49, 4.41	0.59	4.02	-0.32, 8.36	0.070
	Interaction	-0.25	-0.57, 0.059	0.11	-0.25	-0.63, 0.13	0.19	-0.59	-1.08, -0.11	0.016
G-estimation	Intercept	2.75	0.35, 5.16	0.025	2.01	-1.54, 5.56	0.27	3.31	-1.70, 8.32	0.20
	Interaction	-0.35	-0.62, -0.078	0.012	-0.36	-0.75, 0.038	0.076	-0.54	-1.09, 0.015	0.057
Method		Cumulative 4 Year Treatment Effect			Cumulative 5 Year Treatment Effect					
		Est.	95% CI	p	Est.	95% CI	p			
HA-MSM	Intercept	10.23	3.70, 16.75	0.002	9.78	1.53, 18.03	0.020			
	Interaction	-1.35	-2.07, -0.64	<0.001	-1.33	-2.24, -0.43	0.004			
HA-MSM (truncated)	Intercept	9.89	4.29, 15.49	<0.001	10.77	3.64, 17.90	0.003			
	Interaction	-1.33	-1.95, -0.71	<0.0001	-1.47	-2.26, -0.68	<0.001			
G-estimation	Intercept	8.77	2.30, 15.23	0.008	11.64	3.03, 20.24	0.007			
	Interaction	-1.22	-1.93, -0.51	<0.001	-1.57	-2.51, -0.63	0.001			

TABLE E.4: Estimated cumulative effects of DNase on ppFVC modified by previous ppFVC (The intercept term is the estimated effect if previous ppFVC was 0%, and the interaction term is the change in effect per 10% increase in previous ppFVC)

Method	1 Year Treatment Effect			Cumulative 2 Year Treatment Effect			Cumulative 3 Year Treatment Effect		
	Est.	95% CI	p	Est.	95% CI	p	Est.	95% CI	p
IPW of MSM	-1.49	-4.85, 1.88	0.39	-3.64	, -7.44, 0.16	0.061	-4.79	-10.16, 0.59	0.081
IPW of MSM (truncated)	-1.47	-4.56, 1.62	0.35	-3.29	-6.79, 0.21	0.065	-4.66	-9.49, 0.18	0.059
HA-MSM	-0.28	-2.54, 1.97	0.81	-0.93	-3.58, 1.73	0.49	-3.75	-7.78, 0.28	0.068
HA-MSM (truncated)	-0.28	-2.49, 1.93	0.81	-0.77	-3.42, 1.88	0.57	-3.89	-7.86, 0.069	0.054
G-Formula	-0.32	-2.53, 1.89	0.78	-1.89	-4.73, 0.95	0.19	-3.84	-7.37, -0.31	0.033
G-Estimation	-0.93	-3.03, 1.16	0.38	-2.70	-5.37, -0.037	0.047	-4.92	-8.70, -1.15	0.010
Method	Cumulative 4 Year Treatment Effect			Cumulative 5 Year Treatment Effect					
	Est.	95% CI	p	Est.	95% CI	p			
IPW of MSM	-7.20	-14.31, -0.097	0.047	-3.94	-12.20, 4.33	0.35			
IPW of MSM (truncated)	-6.93	-12.84, -1.03	0.021	-4.95	-12.22, 2.33	0.18			
HA-MSM	-3.79	-8.87, 1.29	0.14	-3.33	-9.86, 3.20	0.32			
HA-MSM (truncated)	-4.13	-9.06, 0.81	0.10	-3.45	-10.00, 3.11	0.30			
G-Formula	-4.99	-9.28, -0.70	0.023	-6.32	-11.32, -1.32	0.013			
G-Estimation	-6.60	-11.54, -1.66	0.009	-5.00	-11.33, 1.33	0.12			

TABLE E.5: Estimated population average effects of DNase on ppFEF_{25-75}

Method		1 Year Treatment Effect			Cumulative 2 Year Treatment Effect			Cumulative 3 Year Treatment Effect		
		Est.	95% CI	p	Est.	95% CI	p	Est.	95% CI	p
HA-MSM	Intercept	6.09	0.77, 11.42	0.025	5.04	-1.19, 11.28	0.11	3.44	-5.50, 12.39	0.45
	Interaction	-0.95	-1.79, -0.11	0.027	-0.88	-1.84, 0.093	0.076	-1.03	-2.57, 0.50	0.19
HA-MSM (truncated)	Intercept	6.16	0.87, 11.46	0.022	5.33	-0.87, 11.53	0.092	3.74	-4.89, 12.38	0.40
	Interaction	-0.95	-1.79, -0.12	0.025	-0.89	-1.85, 0.071	0.070	-1.10	-2.56, 0.36	0.14
G-estimation	Intercept	5.29	0.13, 10.45	0.045	4.16	-2.68, 11.01	0.23	1.50	-7.22, 10.22	0.74
	Interaction	-0.96	-1.78, -0.15	0.021	-0.97	-2.08, 0.14	0.087	-0.86	-2.29, 0.57	0.24
Method		Cumulative 4 Year Treatment Effect			Cumulative 5 Year Treatment Effect					
		Est.	95% CI	p	Est.	95% CI	p			
HA-MSM	Intercept	-3.49	-16.57, 9.58	0.60	-2.22	-20.52, 16.08	0.81			
	Interaction	0.008	-2.17, 2.18	0.99	-0.13	-3.19, 2.94	0.94			
HA-MSM (truncated)	Intercept	-2.17	-14.66,10.31	0.73	-1.85	-20.07,16.38	0.84			
	Interaction	-0.23	-2.29, 1.83	0.82	-0.21	-3.27, 2.85	0.89			
G-estimation	Intercept	-0.62	-14.32, 13.08	0.93	-0.94	-18.56, 16.68	0.92			
	Interaction	-0.71	-2.90, 1.48	0.53	-0.32	-2.98, 2.35	0.82			

TABLE E.6: Estimated cumulative effects of DNase on ppFEF_{25-75} modified by previous ppFEF_{25-75} (The intercept term is the estimated effect if previous ppFEF_{25-75} was 0%, and the interaction term is the change in effect per 10% increase in previous ppFEF_{25-75})

Appendix F

Supplementary Tables & Figures to Chapter 9

Here we present the results of the analyses with percent predicted FVC and percent predicted FEF_{25-75} as outcomes. These are both measures of lung function similar to $ppFEV_1$. The analyses presented are identical to the analysis of $ppFEV_1$, starting with an analysis of the one-year effect using SCMM, including estimating effect modification by baseline age or previous lung function, and then estimating the one-, two-, three-, four- and five-year effects using all available methods.

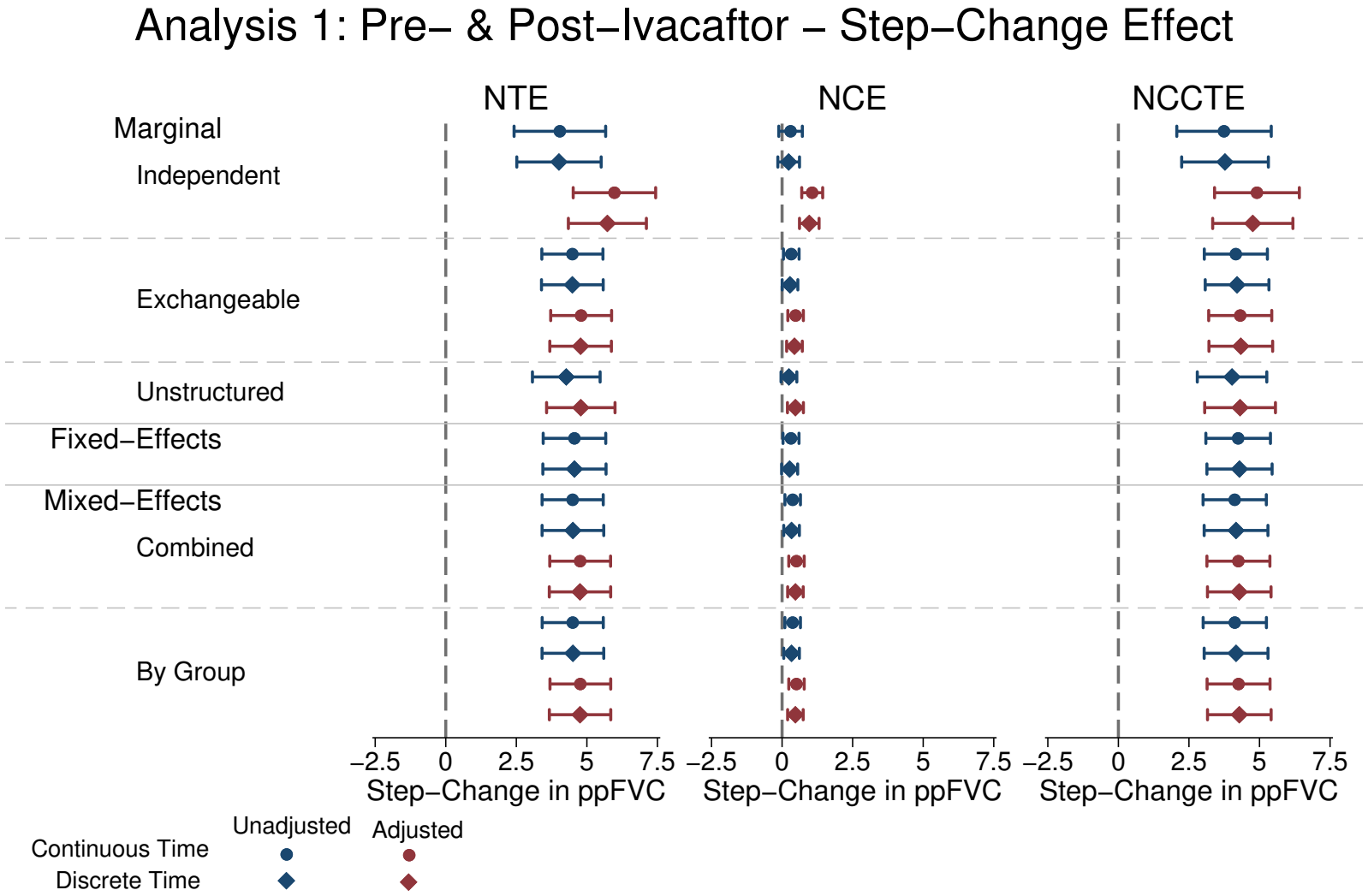


FIGURE F.1: Estimated step-change effect on ppFVC when comparing post-ivacaftor period to pre-ivacaftor period

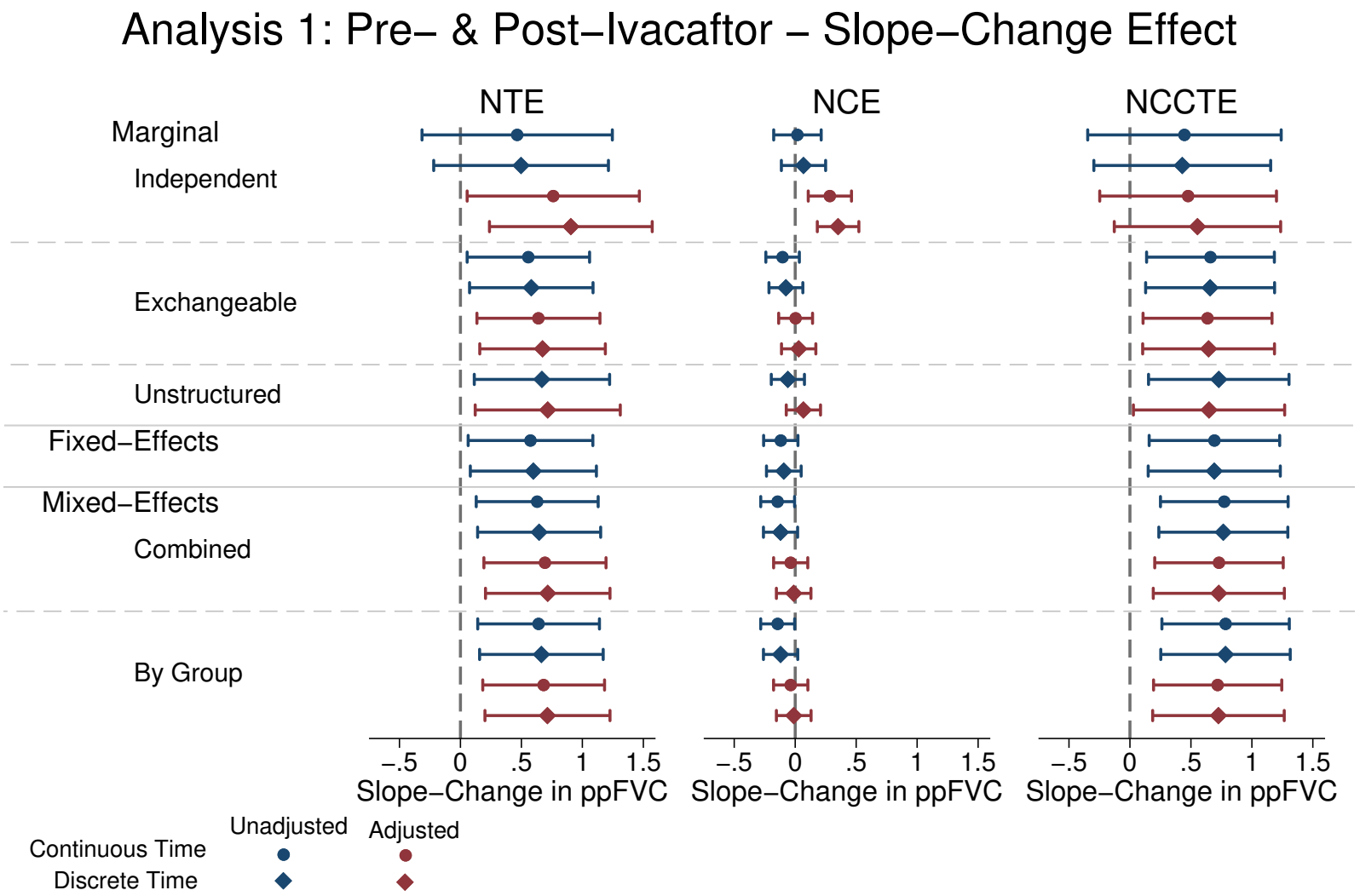


FIGURE F.2: Estimated slope-change effect on ppFVC when comparing post-ivacaftor period to pre-ivacaftor period

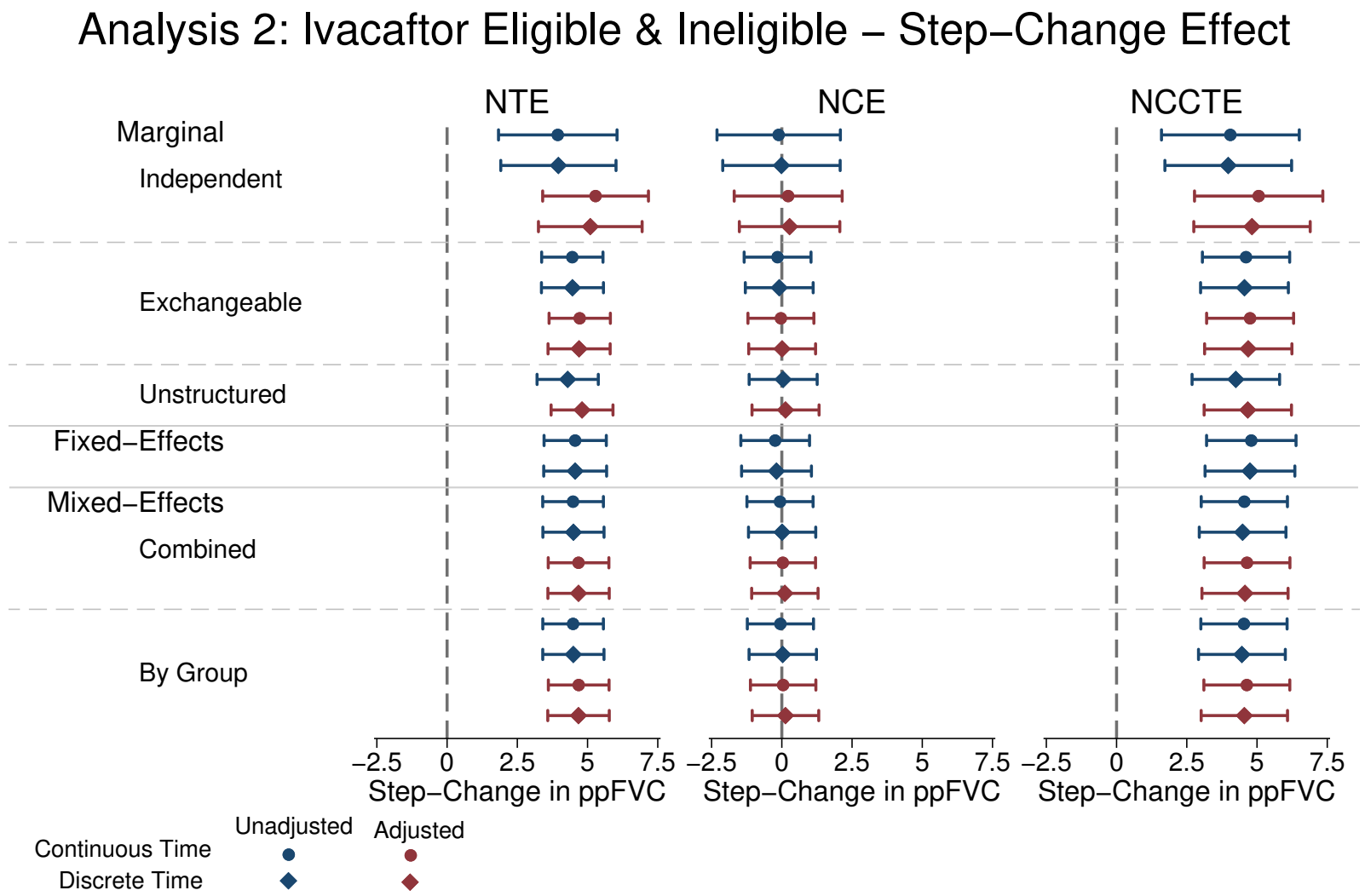


FIGURE F.3: Estimated step-change effect on ppFVC when comparing those eligible for ivacaftor to those ineligible for ivacaftor

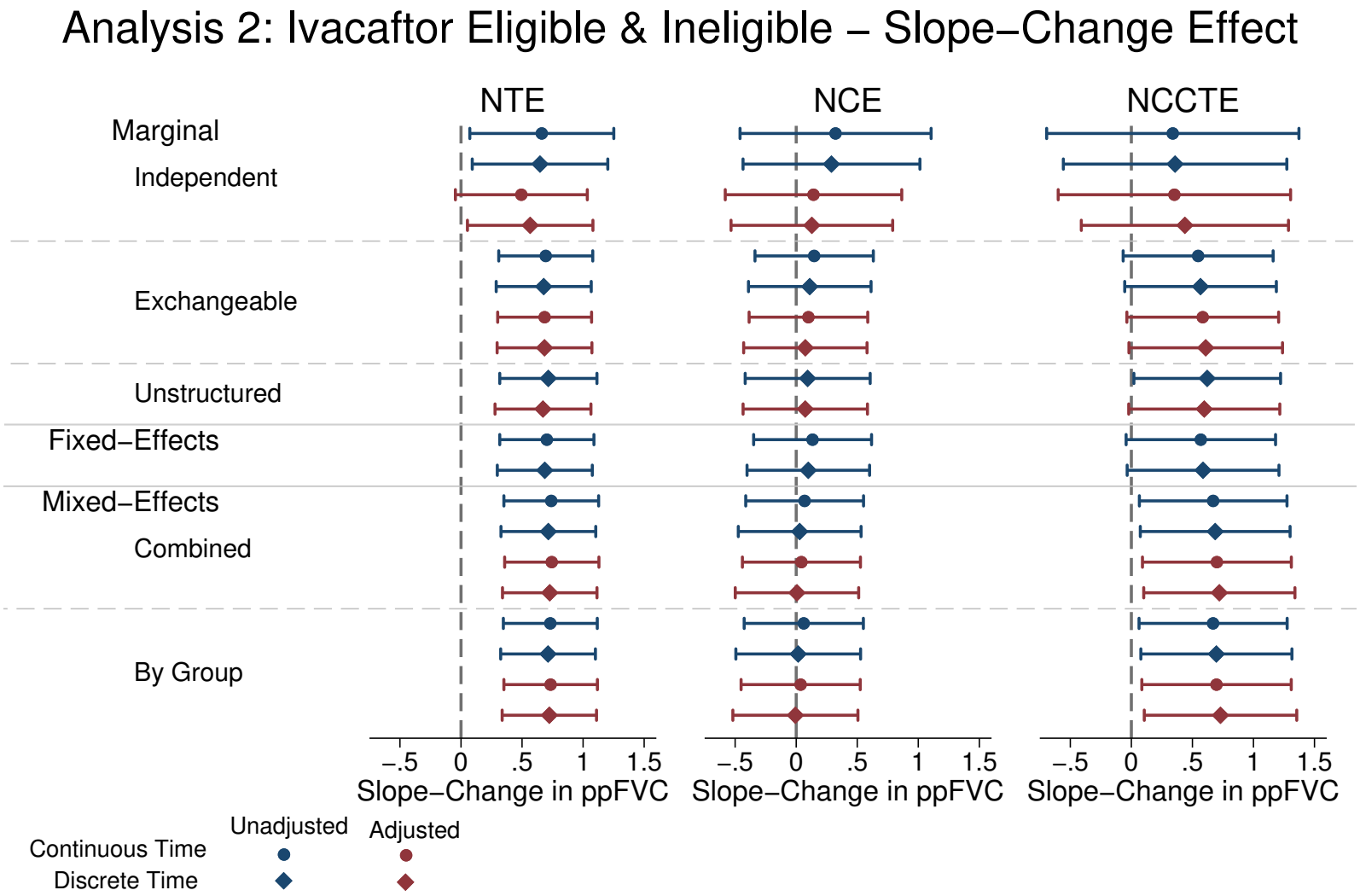


FIGURE F.4: Estimated slope-change effect on ppFVC when comparing those eligible for ivacaftor to those ineligible for ivacaftor

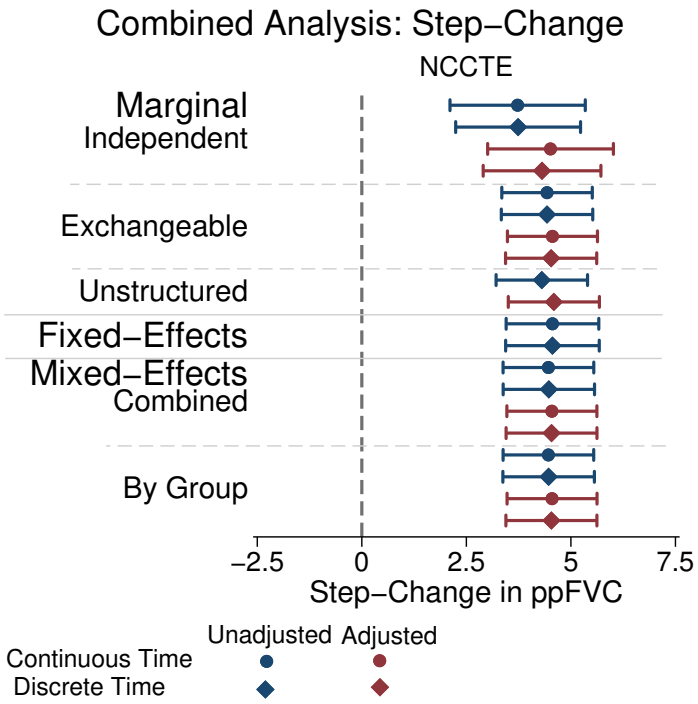


FIGURE F.5: Estimated step-change effect on ppFVC from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time-period prior to the availability of ivacaftor

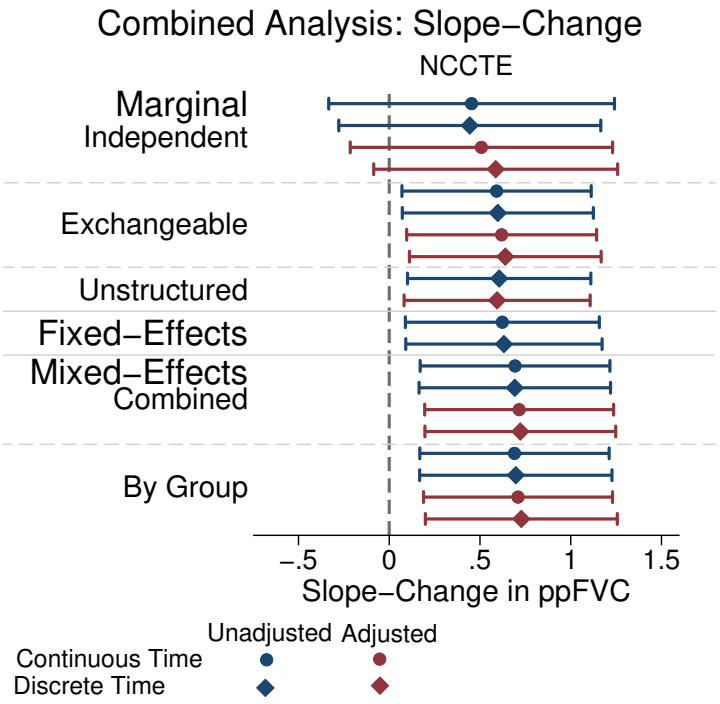


FIGURE F.6: Estimated slope-change effect on ppFVC from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time-period prior to the availability of ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 1: Pre- & Post Comparison - Step-Change Effect											
				NTE (B vs A)				NCE (D vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	4.04	0.83	2.42, 5.66	<0.001	0.30	0.21	-0.12, 0.72	0.16	3.74	0.85	2.07, 5.42	<0.001
			Discrete	4.01	0.76	2.52, 5.51	<0.001	0.23	0.20	-0.15, 0.62	0.23	3.78	0.78	2.24, 5.31	<0.001
		Adjusted	Continuous	5.98	0.75	4.52, 7.44	<0.001	1.07	0.19	0.70, 1.44	<0.001	4.91	0.77	3.41, 6.41	<0.001
			Discrete	5.73	0.71	4.34, 7.11	<0.001	0.96	0.18	0.62, 1.31	<0.001	4.76	0.73	3.34, 6.18	<0.001
	Exchangeable	Unadjusted	Continuous	4.49	0.55	3.40, 5.57	<0.001	0.33	0.14	0.05, 0.60	0.021	4.16	0.57	3.04, 5.28	<0.001
			Discrete	4.49	0.56	3.39, 5.58	<0.001	0.28	0.14	0.00, 0.56	0.048	4.20	0.58	3.08, 5.33	<0.001
		Adjusted	Continuous	4.80	0.55	3.72, 5.87	<0.001	0.48	0.14	0.20, 0.75	<0.001	4.32	0.57	3.20, 5.43	<0.001
			Discrete	4.78	0.56	3.69, 5.87	<0.001	0.44	0.14	0.16, 0.72	0.002	4.34	0.58	3.21, 5.47	<0.001
	Unstructured	Unadjusted	Discrete	4.27	0.61	3.07, 5.47	<0.001	0.24	0.14	-0.04, 0.52	0.090	4.03	0.63	2.80, 5.26	<0.001
		Adjusted		4.78	0.62	3.57, 6.00	<0.001	0.47	0.14	0.19, 0.75	0.001	4.31	0.64	3.06, 5.56	<0.001
Fixed-Effects		Unadjusted	Continuous	4.56	0.56	3.45, 5.67	<0.001	0.32	0.14	0.03, 0.60	0.029	4.24	0.58	3.10, 5.39	<0.001
			Discrete	4.56	0.57	3.44, 5.68	<0.001	0.27	0.15	-0.02, 0.55	0.065	4.29	0.59	3.14, 5.45	<0.001
Mixed-Effects	Combined	Unadjusted	Continuous	4.50	0.55	3.41, 5.58	<0.001	0.38	0.14	0.10, 0.65	0.008	4.12	0.57	3.00, 5.24	<0.001
			Discrete	4.51	0.56	3.41, 5.60	<0.001	0.34	0.14	0.06, 0.61	0.018	4.17	0.58	3.04, 5.30	<0.001
		Adjusted	Continuous	4.76	0.55	3.68, 5.84	<0.001	0.51	0.14	0.24, 0.78	<0.001	4.25	0.57	3.14, 5.37	<0.001
			Discrete	4.76	0.55	3.67, 5.84	<0.001	0.47	0.14	0.20, 0.75	<0.001	4.28	0.57	3.16, 5.41	<0.001
	By Group	Unadjusted	Continuous	4.50	0.55	3.42, 5.58	<0.001	0.37	0.14	0.10, 0.65	0.008	4.12	0.57	3.00, 5.24	<0.001
			Discrete	4.50	0.56	3.41, 5.60	<0.001	0.34	0.14	0.06, 0.61	0.018	4.17	0.58	3.04, 5.30	<0.001
		Adjusted	Continuous	4.77	0.55	3.69, 5.84	<0.001	0.51	0.14	0.24, 0.78	<0.001	4.26	0.57	3.14, 5.37	<0.001
			Discrete	4.76	0.56	3.67, 5.85	<0.001	0.47	0.14	0.20, 0.75	<0.001	4.28	0.57	3.16, 5.41	<0.001

TABLE F.1: Estimated step-change effect on ppFVC when comparing post-ivacaftor period to pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 1: Pre- & Post Comparison - Slope-Change Effect											
				NTE (B vs A)				NCE (D vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	0.47	0.40	-0.31, 1.25	0.24	0.02	0.10	-0.18, 0.21	0.85	0.45	0.40	-0.35, 1.24	0.27
			Discrete	0.50	0.37	-0.22, 1.21	0.17	0.07	0.09	-0.11, 0.25	0.46	0.43	0.37	-0.30, 1.15	0.25
		Adjusted	Continuous	0.76	0.36	0.06, 1.47	0.035	0.28	0.09	0.11, 0.46	0.002	0.48	0.37	-0.25, 1.20	0.20
			Discrete	0.91	0.34	0.24, 1.57	0.008	0.35	0.087	0.18, 0.52	<0.001	0.55	0.35	-0.13, 1.24	0.11
	Exchangeable	Unadjusted	Continuous	0.56	0.26	0.06, 1.06	0.030	-0.10	0.07	-0.24, 0.03	0.14	0.66	0.27	0.14, 1.18	0.013
			Discrete	0.58	0.26	0.07, 1.09	0.024	-0.08	0.07	-0.22, 0.06	0.28	0.66	0.27	0.13, 1.19	0.015
		Adjusted	Continuous	0.64	0.26	0.14, 1.14	0.013	0.00	0.07	-0.14, 0.14	0.96	0.64	0.27	0.11, 1.16	0.018
			Discrete	0.67	0.26	0.16, 1.19	0.010	0.03	0.07	-0.11, 0.17	0.69	0.64	0.28	0.10, 1.19	0.019
	Unstructured	Unadjusted	Discrete	0.67	0.28	0.11, 1.22	0.018	-0.06	0.069	-0.20, 0.08	0.39	0.73	0.29	0.15, 1.30	0.013
		Adjusted		0.72	0.30	0.12, 1.31	0.018	0.07	0.07	-0.07, 0.21	0.35	0.65	0.32	0.03, 1.27	0.040
Fixed-Effects		Unadjusted	Continuous	0.57	0.26	0.06, 1.09	0.028	-0.12	0.07	-0.26, 0.02	0.10	0.69	0.27	0.16, 1.23	0.011
			Discrete	0.60	0.26	0.08, 1.12	0.023	-0.09	0.07	-0.24, 0.05	0.20	0.69	0.28	0.15, 1.23	0.012
Mixed-Effects	Combined	Unadjusted	Continuous	0.63	0.26	0.13, 1.13	0.014	-0.14	0.07	-0.29, -0.01	0.042	0.77	0.27	0.25, 1.30	0.004
			Discrete	0.65	0.26	0.14, 1.15	0.012	-0.12	0.07	-0.26, 0.02	0.092	0.77	0.27	0.24, 1.29	0.005
		Adjusted	Continuous	0.69	0.26	0.19, 1.19	0.007	-0.04	0.07	-0.18, 0.10	0.60	0.73	0.27	0.20, 1.26	0.007
			Discrete	0.72	0.26	0.21, 1.23	0.006	-0.01	0.07	-0.15, 0.13	0.86	0.73	0.27	0.19, 1.27	0.008
	By Group	Unadjusted	Continuous	0.64	0.25	0.14, 1.14	0.012	-0.14	0.071	-0.28, -0.00	0.043	0.78	0.27	0.26, 1.31	0.003
			Discrete	0.66	0.26	0.16, 1.17	0.010	-0.12	0.07	-0.26, 0.02	0.095	0.78	0.27	0.25, 1.31	0.004
		Adjusted	Continuous	0.68	0.25	0.18, 1.18	0.007	-0.04	0.07	-0.18, 0.10	0.61	0.72	0.27	0.19, 1.24	0.007
			Discrete	0.71	0.26	0.20, 1.23	0.006	-0.01	0.07	-0.15, 0.13	0.87	0.73	0.28	0.19, 1.27	0.008

TABLE F.2: Estimated slope-change effect on ppFVC when comparing post-ivacaftor period to pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 2: Eligible & Ineligible Comparison - Step-Change Effect											
				NTE (B vs A)				NCE (D vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	3.94	1.08	1.83, 6.05	<0.001	-0.11	1.12	-2.31, 2.08	0.92	4.05	1.25	1.60, 6.50	0.001
			Discrete	3.96	1.05	1.91, 6.01	<0.001	-0.01	1.07	-2.10, 2.08	0.99	3.98	1.15	1.72, 6.23	<0.001
		Adjusted	Continuous	5.29	0.96	3.40, 7.17	<0.001	0.23	0.98	-1.70, 2.15	0.82	5.06	1.17	2.78, 7.34	<0.001
			Discrete	5.10	0.94	3.25, 6.94	<0.001	0.28	0.91	-1.51, 2.06	0.76	4.82	1.06	2.75, 6.89	<0.001
	Exchangeable	Unadjusted	Continuous	4.46	0.56	3.37, 5.55	<0.001	-0.15	0.61	-1.34, 1.04	0.80	4.61	0.79	3.06, 6.16	<0.001
			Discrete	4.46	0.56	3.36, 5.57	<0.001	-0.09	0.62	-1.30, 1.12	0.88	4.55	0.80	2.99, 6.12	<0.001
		Adjusted	Continuous	4.72	0.56	3.63, 5.81	<0.001	-0.03	0.60	-1.21, 1.14	0.96	4.75	0.79	3.21, 6.30	<0.001
			Discrete	4.70	0.56	3.59, 5.80	<0.001	0.01	0.61	-1.18, 1.20	0.99	4.69	0.79	3.13, 6.24	<0.001
	Unstructured	Unadjusted	Discrete	4.29	0.56	3.20, 5.38	<0.001	0.05	0.62	-1.16, 1.26	0.94	4.25	0.80	2.69, 5.80	<0.001
		Adjusted		4.80	0.56	3.70, 5.91	<0.001	0.13	0.61	-1.06, 1.32	0.83	4.67	0.79	3.12, 6.23	<0.001
Fixed-Effects		Unadjusted	Continuous	4.56	0.57	3.45, 5.67	<0.001	-0.24	0.62	-1.46, 0.99	0.70	4.80	0.81	3.20, 6.39	<0.001
			Discrete	4.56	0.57	3.44, 5.68	<0.001	-0.19	0.63	-1.43, 1.06	0.76	4.75	0.82	3.15, 6.35	<0.001
Mixed-Effects	Combined	Unadjusted	Continuous	4.48	0.55	3.40, 5.56	<0.001	-0.06	0.60	-1.24, 1.11	0.91	4.55	0.78	3.01, 6.08	<0.001
			Discrete	5.50	0.56	3.41, 5.59	<0.001	0.01	0.61	-1.19, 1.21	0.99	4.49	0.79	2.95, 6.03	<0.001
		Adjusted	Continuous	4.68	0.55	3.60, 5.76	<0.001	0.04	0.59	-1.13, 1.20	0.95	4.64	0.78	3.12, 6.17	<0.001
			Discrete	4.68	0.56	3.59, 5.77	<0.001	0.11	0.60	-1.07, 1.29	0.86	4.57	0.78	3.04, 6.11	<0.001
	By Group	Unadjusted	Continuous	4.49	0.55	3.41, 5.56	<0.001	-0.05	0.60	-1.23, 1.13	0.93	4.54	0.78	3.00, 6.07	<0.001
			Discrete	4.50	0.56	3.41, 5.59	<0.001	0.03	0.61	-1.17, 1.23	0.96	4.46	0.79	2.92, 6.01	<0.001
		Adjusted	Continuous	4.68	0.55	3.60, 5.77	<0.001	0.05	0.59	-1.12, 1.21	0.94	4.64	0.78	3.11, 6.17	<0.001
			Discrete	4.68	0.56	3.58, 5.77	<0.001	0.13	0.60	-1.05, 1.31	0.83	4.55	0.78	3.01, 6.09	<0.001

TABLE F.3: Estimated step-change effect on ppFVC when comparing those eligible for ivacaftor to those ineligible for ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 2: Eligible & Ineligible Comparison - Slope-Change Effect											
				NTE (B vs A)				NCE (D vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	0.66	0.30	0.07, 1.25	0.028	0.32	0.40	-0.46, 1.10	0.42	0.34	0.53	-0.69, 1.37	0.52
			Discrete	0.65	0.28	0.09, 1.20	0.022	0.29	0.37	-0.44, 1.01	0.43	0.36	0.47	-0.56, 1.27	0.44
		Adjusted	Continuous	0.49	0.28	-0.05, 1.03	0.073	0.14	0.37	-0.58, 0.86	0.70	0.35	0.49	-0.60, 1.30	0.47
			Discrete	0.57	0.26	0.05, 1.08	0.031	0.13	0.34	-0.53, 0.79	0.70	0.44	0.43	-0.41, 1.29	0.31
	Exchangeable	Unadjusted	Continuous	0.69	0.20	0.31, 1.08	<0.001	0.15	0.25	-0.34, 0.63	0.55	0.55	0.31	-0.07, 1.16	0.081
			Discrete	0.68	0.20	0.29, 1.07	<0.001	0.11	0.26	-0.39, 0.61	0.66	0.57	0.32	-0.05, 1.19	0.074
		Adjusted	Continuous	0.68	0.20	0.30, 1.07	<0.001	0.10	0.25	-0.39, 0.59	0.69	0.58	0.32	-0.04, 1.21	0.065
			Discrete	0.68	0.20	0.30, 1.07	<0.001	0.08	0.26	-0.43, 0.58	0.77	0.61	0.32	-0.02, 1.24	0.058
	Unstructured	Unadjusted	Discrete	0.72	0.20	0.32, 1.11	<0.001	0.09	0.26	-0.42, 0.61	0.72	0.62	0.31	0.02, 1.22	0.043
		Adjusted		0.67	0.20	0.28, 1.06	<0.001	0.07	0.26	-0.44, 0.58	0.78	0.60	0.32	-0.02, 1.22	0.059
Fixed-Effects		Unadjusted	Continuous	0.70	0.20	0.32, 1.09	<0.001	0.13	0.25	-0.35, 0.62	0.59	0.57	0.31	-0.04, 1.18	0.068
			Discrete	0.69	0.20	0.30, 1.08	<0.001	0.10	0.26	-0.40, 0.60	0.70	0.59	0.32	-0.03, 1.21	0.064
Mixed-Effects	Combined	Unadjusted	Continuous	0.74	0.20	0.35, 1.13	<0.001	0.07	0.25	-0.41, 0.55	0.78	0.67	0.31	0.07, 1.27	0.030
			Discrete	0.72	0.20	0.33, 1.10	<0.001	0.03	0.26	-0.47, 0.53	0.91	0.69	0.31	0.07, 1.30	0.028
		Adjusted	Continuous	0.74	0.20	0.36, 1.13	<0.001	0.04	0.25	-0.44, 0.53	0.86	0.70	0.31	0.09, 1.31	0.024
			Discrete	0.73	0.20	0.34, 1.11	<0.001	0.01	0.26	-0.50, 0.51	0.98	0.72	0.32	0.10, 1.34	0.023
	By Group	Unadjusted	Continuous	0.73	0.20	0.35, 1.12	<0.001	0.06	0.25	-0.43, 0.55	0.80	0.67	0.31	0.06, 1.28	0.031
			Discrete	0.71	0.20	0.32, 1.10	<0.001	0.02	0.26	-0.49, 0.53	0.95	0.70	0.32	0.08, 1.31	0.027
		Adjusted	Continuous	0.73	0.20	0.35, 1.12	<0.001	0.04	0.25	-0.45, 0.52	0.88	0.70	0.31	0.09, 1.31	0.025
			Discrete	0.72	0.20	0.34, 1.11	<0.001	-0.01	0.26	-0.52, 0.51	0.98	0.73	0.32	0.11, 1.35	0.022

TABLE F.4: Estimated slope-change effect on ppFVC when comparing those eligible for ivacaftor to those ineligible for ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Combined Analysis Step-Change Effect			
				est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	3.73	0.83	2.11, 5.34	<0.001
			Discrete	3.74	0.76	2.24, 5.23	<0.001
		Adjusted	Continuous	4.51	0.77	3.01, 6.02	<0.001
			Discrete	4.31	0.72	2.90, 5.72	<0.001
	Exchangeable	Unadjusted	Continuous	4.43	0.55	3.35, 5.51	<0.001
			Discrete	4.43	0.56	3.34, 5.52	<0.001
		Adjusted	Continuous	4.56	0.55	3.48, 5.64	<0.001
			Discrete	4.53	0.56	3.44, 5.62	<0.001
	Unstructured	Unadjusted	Discrete	4.30	0.56	3.21, 5.40	<0.001
		Adjusted		4.59	0.56	3.5, 5.68	<0.001
Fixed-Effects		Unadjusted	Continuous	4.56	0.57	3.45, 5.67	<0.001
			Discrete	4.56	0.57	3.44, 5.68	<0.001
Mixed-Effects	Combined	Unadjusted	Continuous	4.46	0.55	3.38, 5.55	<0.001
			Discrete	4.47	0.56	3.38, 5.57	<0.001
		Adjusted	Continuous	4.55	0.55	3.47, 5.62	<0.001
			Discrete	4.54	0.55	3.45, 5.62	<0.001
	By Group	Unadjusted	Continuous	4.46	0.55	3.38, 5.55	<0.001
			Discrete	4.47	0.56	3.37, 5.56	<0.001
		Adjusted	Continuous	4.55	0.55	3.47, 5.62	<0.001
			Discrete	4.53	0.56	3.45, 5.62	<0.001

TABLE F.5: Estimated step-change effect on ppFVC from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time-period prior to the availability of ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Combined Analysis Slope-Change Effect			
				est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	0.45	0.40	-0.33, 1.24	0.26
			Discrete	0.44	0.37	-0.28, 1.17	0.23
		Adjusted	Continuous	0.51	0.37	-0.21, 1.23	0.17
			Discrete	0.59	0.34	-0.09, 1.26	0.087
	Exchangeable	Unadjusted	Continuous	0.59	0.27	0.07, 1.11	0.026
			Discrete	0.60	0.27	0.07, 1.12	0.026
		Adjusted	Continuous	0.62	0.27	0.10, 1.14	0.020
			Discrete	0.64	0.27	0.11, 1.17	0.018
	Unstructured	Unadjusted	Discrete	0.61	0.26	0.10, 1.11	0.019
		Adjusted		0.59	0.26	0.08, 1.11	0.023
Fixed-Effects		Unadjusted	Continuous	0.62	0.27	0.09, 1.16	0.022
			Discrete	0.63	0.28	0.09, 1.17	0.022
Mixed-Effects	Combined	Unadjusted	Continuous	0.69	0.27	0.17, 1.22	0.009
			Discrete	0.69	0.27	0.16, 1.22	0.010
		Adjusted	Continuous	0.72	0.27	0.20, 1.24	0.007
			Discrete	0.72	0.27	0.20, 1.25	0.007
	By Group	Unadjusted	Continuous	0.69	0.27	0.17, 1.21	0.010
			Discrete	0.70	0.27	0.17, 1.23	0.010
		Adjusted	Continuous	0.71	0.27	0.19, 1.23	0.008
			Discrete	0.73	0.27	0.20, 1.26	0.007

TABLE F.6: Estimated slope-change effect on ppFVC from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time-period prior to the availability of ivacaftor

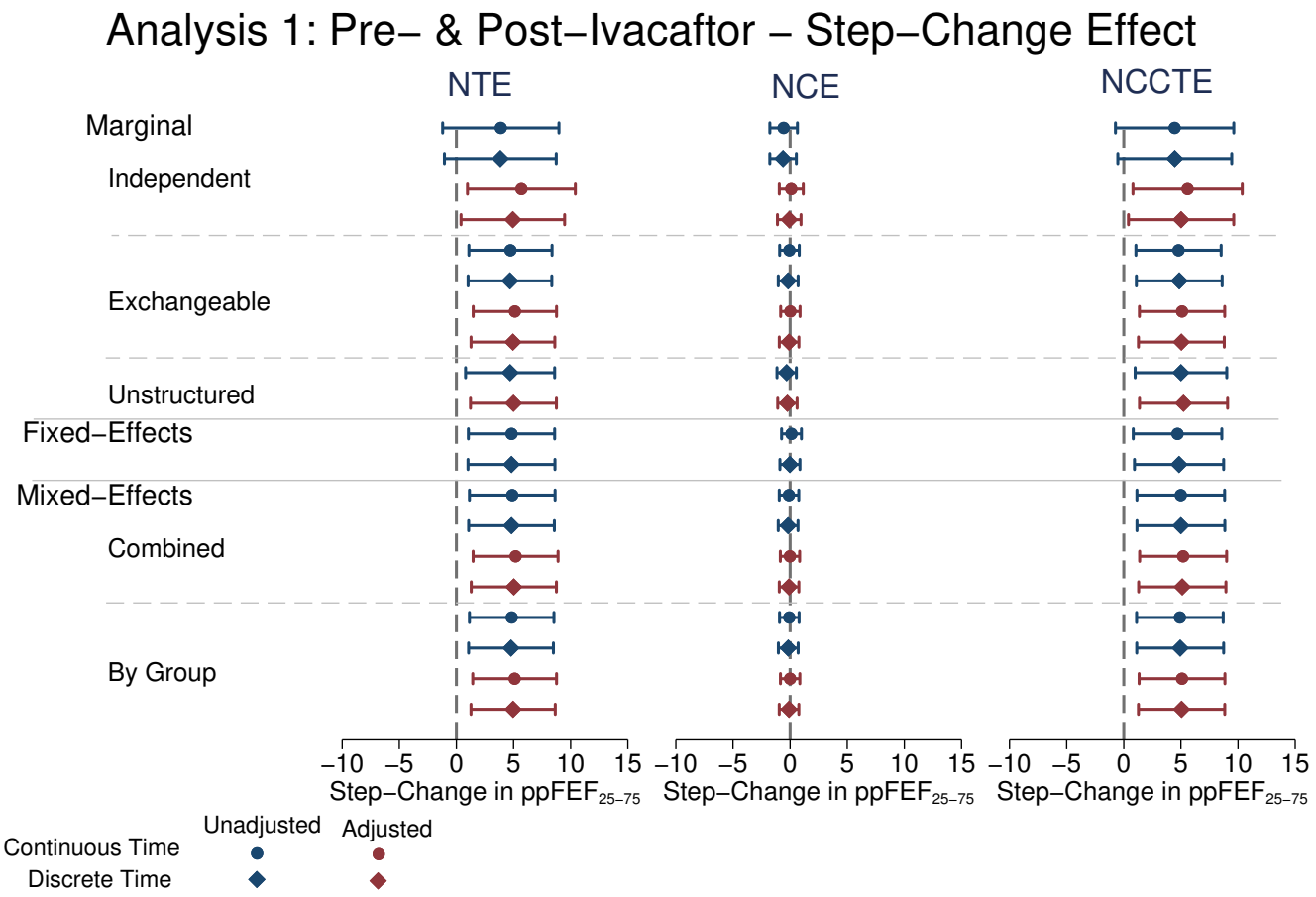


FIGURE F.7: Estimated step-change effect on $ppFEF_{25-75}$ when comparing post-ivacaftor period to pre-ivacaftor period

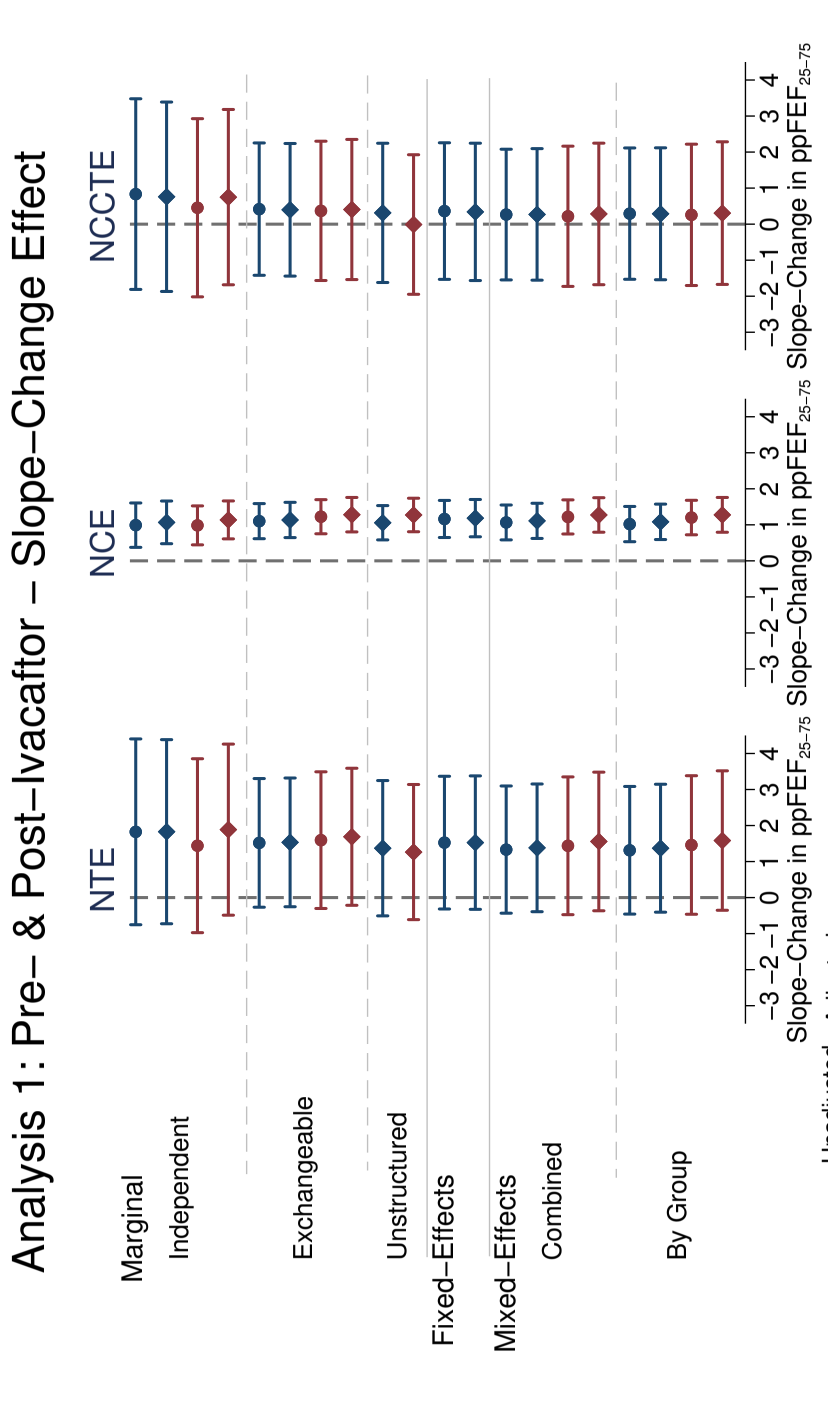


FIGURE F.8: Estimated slope-change effect on $ppFEF_{25-75}$ when comparing post-ivacaftor period to pre-ivacaftor period

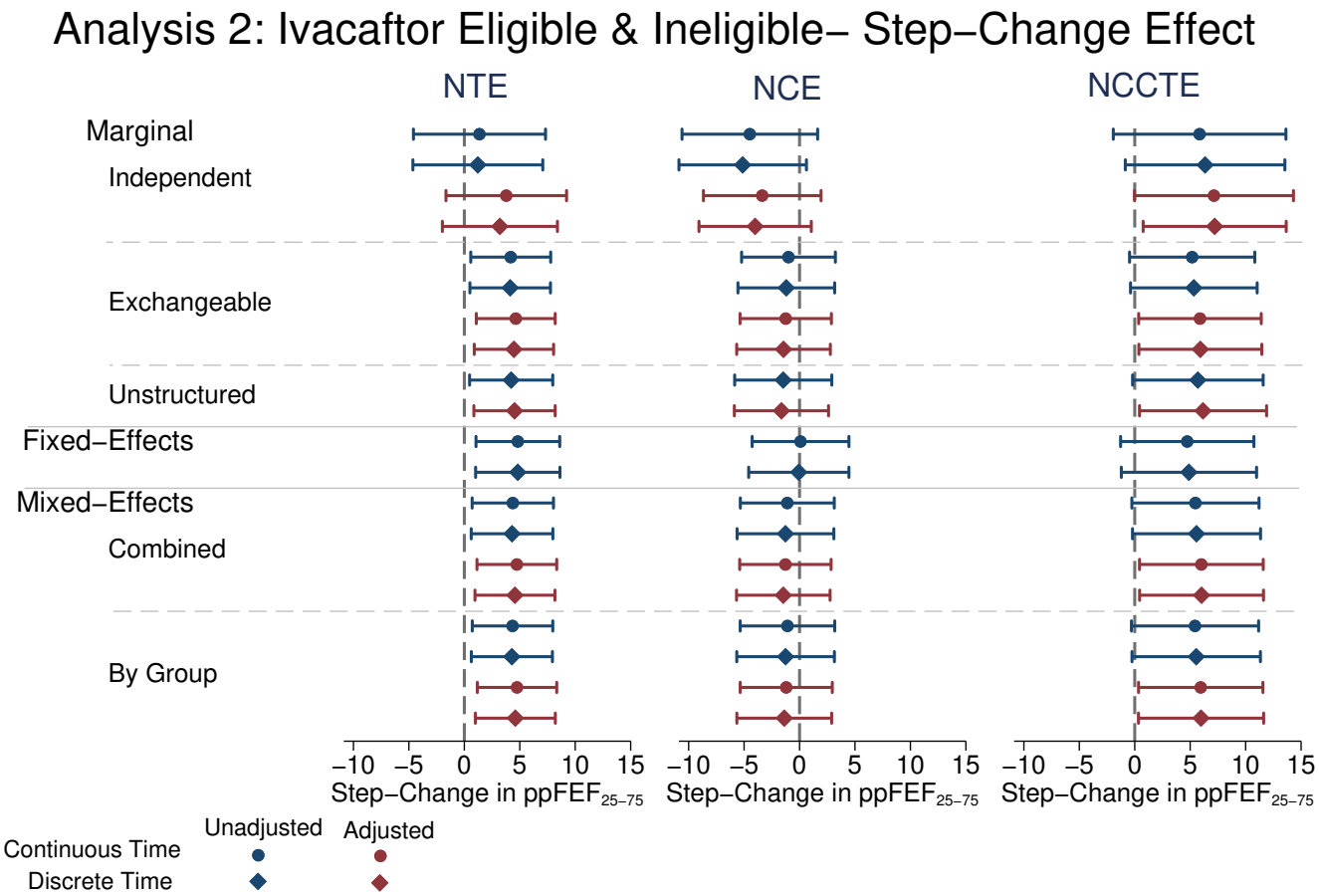


FIGURE F.9: Estimated step-change effect on $ppFEF_{25-75}$ when comparing those eligible for ivacaftor to those ineligible for ivacaftor

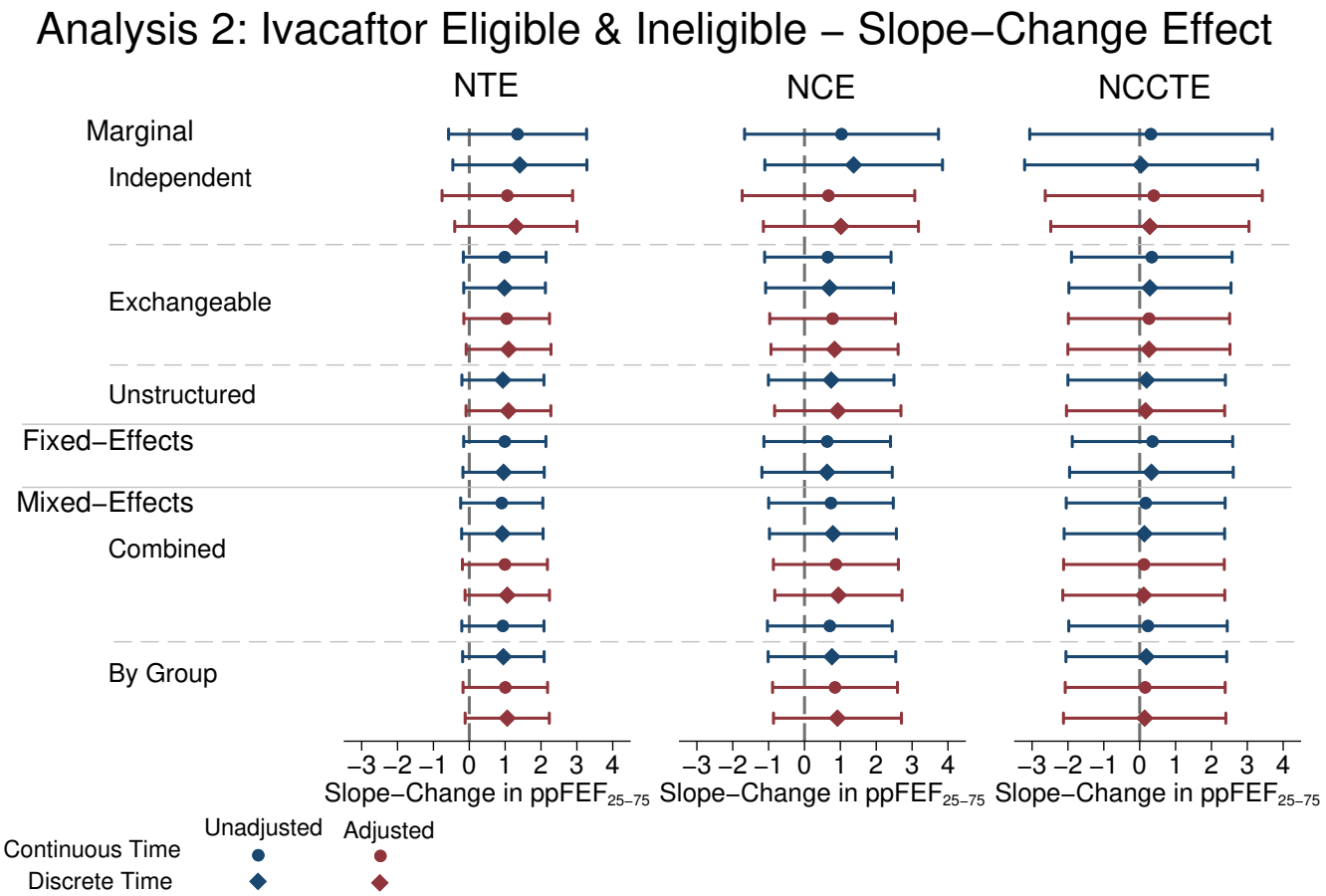


FIGURE F.10: Estimated slope-change effect on $ppFEF_{25-75}$ when comparing those eligible for ivacaftor to those ineligible for ivacaftor

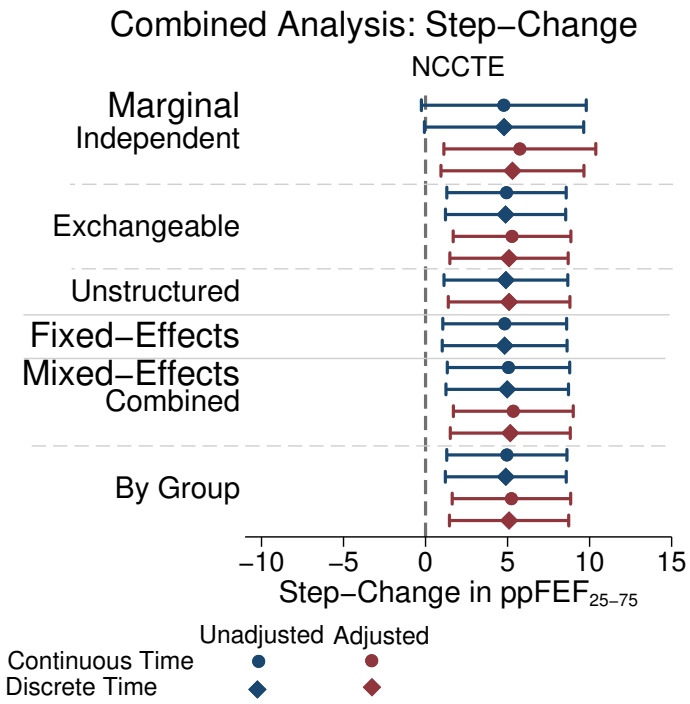


FIGURE F.11: Estimated step-change effect on ppFEF_{25-75} from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time-period prior to the availability of ivacaftor

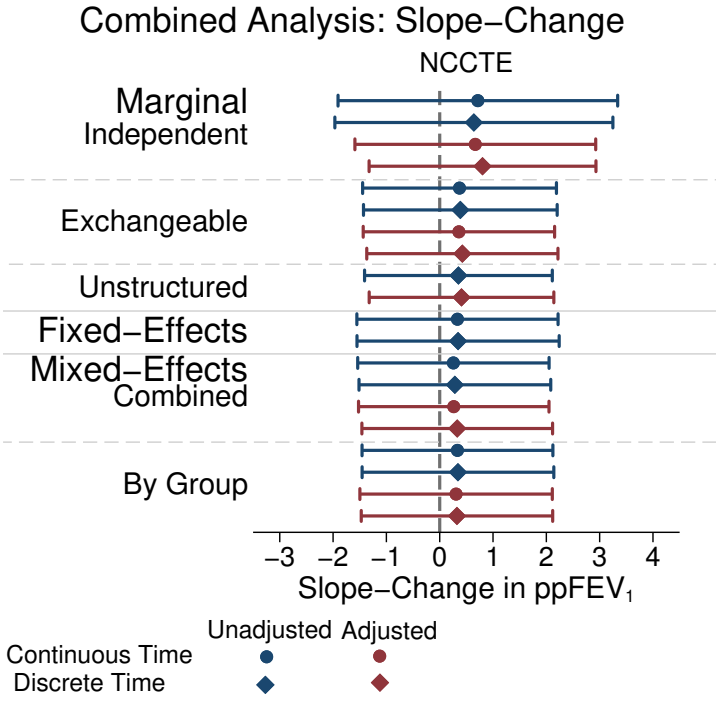


FIGURE F.12: Estimated slope-change effect on ppFEF_{25-75} from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time-period prior to the availability of ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 1: Pre- & Post Comparison - Step-Change Effect											
				NTE (B vs A)				NCE (D vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	3.89	2.60	-1.20, 8.98	0.13	-0.57	0.62	-1.77, 0.64	0.36	4.45	2.64	-0.73, 9.63	0.092
			Discrete	3.85	2.50	-1.05, 8.74	0.12	-0.62	0.59	01.78, 0.54	0.30	4.46	2.54	-0.52, 9.45	0.079
		Adjusted	Continuous	5.69	2.41	0.97, 10.41	0.018	0.10	0.54	-0.95, 1.15	0.85	5.59	2.44	0.80, 10.37	0.022
			Discrete	4.94	2.31	0.41, 9.47	0.033	-0.08	0.53	-1.11, 0.95	0.88	5.02	2.35	0.41, 9.63	0.033
	Exchangeable	Unadjusted	Continuous	4.73	1.86	1.09, 8.37	0.011	-0.06	0.44	-0.92, 0.80	0.89	4.79	1.90	1.06, 8.52	0.012
			Discrete	4.69	1.87	1.02, 8.35	0.012	-0.17	0.44	-1.04, 0.70	0.70	4.86	1.92	1.10, 8.61	0.011
		Adjusted	Continuous	5.12	1.86	1.47, 8.77	0.006	0.02	0.43	-0.83, 0.87	0.96	5.10	1.91	1.37, 8.84	0.007
			Discrete	4.95	1.87	1.28, 8.61	0.008	-0.09	0.44	-0.95, 0.76	0.83	5.04	1.92	1.28, 8.80	0.009
	Unstructured	Unadjusted	Discrete	4.70	1.99	0.80, 8.60	0.018	-0.30	0.43	-1.15, 0.55	0.49	5.00	2.05	0.99, 9.01	0.014
		Adjusted		5.00	1.92	1.23, 8.76	0.009	-0.24	0.43	-1.08, 0.61	0.59	5.23	1.97	1.37, 9.09	0.008
Fixed-Effects		Unadjusted	Continuous	4.83	1.93	1.05, 8.60	0.012	0.12	0.44	-0.75, 0.99	0.78	4.70	1.98	0.83, 8.58	0.017
			Discrete	4.82	1.94	1.02, 8.63	0.013	-0.02	0.45	-0.90, 0.86	0.97	4.84	1.99	0.94, 8.74	0.015
Mixed-Effects	Combined	Unadjusted	Continuous	4.89	1.91	1.14, 8.64	0.011	-0.10	0.44	-0.95, 0.75	0.82	4.99	1.96	1.15, 8.83	0.011
			Discrete	4.82	1.91	1.07, 8.58	0.012	-0.18	0.44	-1.04, 0.69	0.69	5.00	1.96	1.15, 8.84	0.011
		Adjusted	Continuous	5.18	1.90	1.46, 8.90	0.006	-0.01	0.43	-0.86, 0.83	0.98	5.19	1.94	1.39, 9.00	0.007
			Discrete	5.03	1.90	1.30, 8.76	0.008	-0.09	0.44	-0.95, 0.76	0.83	5.12	1.95	1.30, 8.94	0.009
	By Group	Unadjusted	Continuous	4.84	1.89	1.14, 8.54	0.010	-0.07	0.44	-0.93, 0.79	0.87	4.91	1.93	1.12, 8.70	0.011
			Discrete	4.78	1.89	1.07, 8.49	0.012	-0.16	0.44	-1.03, 0.71	0.72	4.94	1.94	1.14, 8.74	0.011
		Adjusted	Continuous	5.10	1.87	1.43, 8.77	0.006	0.00	0.43	-0.84, 0.85	0.99	5.10	1.92	1.34, 8.85	0.008
			Discrete	4.96	1.88	1.28, 8.65	0.008	-0.09	0.44	-0.95, 0.76	0.83	5.06	1.93	1.28, 8.84	0.009

TABLE F.7: Estimated step-change effect on ppFEF_{25-75} when comparing post-ivacaftor period to pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 1: Pre- & Post Comparison - Slope-Change Effect											
				NTE (B vs A)				NCE (D vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	1.83	1.32	-0.75, 4.40	0.17	0.99	0.31	0.38, 1.61	0.002	0.83	1.35	-1.81, 3.48	0.54
			Discrete	1.83	1.30	-0.73, 4.38	0.16	1.07	0.30	0.48, 1.66	<0.001	0.76	1.34	-1.87, 3.39	0.57
		Adjusted	Continuous	1.44	1.23	-0.97, 3.85	0.24	0.99	0.28	0.44, 1.53	<0.001	0.45	1.26	-2.02, 2.93	0.72
			Discrete	1.89	1.21	-0.49, 4.26	0.12	1.14	0.27	0.61, 1.66	<0.001	0.75	1.24	-1.68, 3.18	0.55
	Exchangeable	Unadjusted	Continuous	1.52	0.91	-0.27, 3.30	0.096	1.10	0.25	0.61, 1.59	<0.001	0.42	0.94	-1.42, 2.25	0.66
			Discrete	1.53	0.91	-0.25, 3.32	0.093	1.14	0.25	0.65, 1.62	<0.001	0.40	0.94	-1.44, 2.24	0.67
		Adjusted	Continuous	1.59	0.97	-0.30, 3.49	0.099	1.23	0.24	0.75, 1.70	<0.001	0.37	0.99	-1.56, 2.30	0.71
			Discrete	1.69	0.97	-0.21, 3.59	0.082	1.28	0.24	0.80, 1.76	<0.001	0.41	0.99	-1.54, 2.35	0.68
	Unstructured	Unadjusted	Discrete	1.37	0.96	-0.51, 3.25	0.15	1.06	0.24	0.58, 1.53	<0.001	0.31	0.99	-1.62, 2.25	0.75
		Adjusted		1.26	0.96	-0.61, 3.14	0.19	1.27	0.24	0.81, 1.74	<0.001	-0.01	0.99	-1.95, 1.93	0.99
Fixed-Effects		Unadjusted	Continuous	1.53	0.94	-0.32, 3.37	0.10	1.16	0.26	0.65, 1.68	<0.001	0.36	0.97	-1.53, 2.26	0.71
			Discrete	1.53	0.94	-0.33, 3.38	0.11	1.19	0.26	0.67, 1.70	<0.001	0.34	0.97	-1.57, 2.25	0.73
Mixed-Effects	Combined	Unadjusted	Continuous	1.33	0.90	-0.43, 3.10	0.14	1.07	0.25	0.58, 1.55	<0.001	0.27	0.93	-1.55, 2.08	0.77
			Discrete	1.38	0.90	-0.39, 3.15	0.13	1.11	0.25	0.62, 1.60	<0.001	0.27	0.93	-1.56, 2.10	0.77
		Adjusted	Continuous	1.44	0.98	-0.47, 3.35	0.14	1.22	0.24	0.75, 1.69	<0.001	0.22	0.99	-1.73, 2.16	0.83
			Discrete	1.56	0.98	-0.37, 3.48	0.11	1.27	0.24	0.80, 1.75	<0.001	0.28	1.00	-1.68, 2.25	0.78
	By Group	Unadjusted	Continuous	1.31	0.90	-0.46, 3.08	0.15	1.02	0.25	0.53, 1.51	<0.001	0.29	0.93	-1.53, 2.11	0.75
			Discrete	1.37	0.91	-0.40, 3.15	0.13	1.08	0.25	0.59, 1.58	<0.001	0.29	0.93	-1.54, 2.12	0.76
		Adjusted	Continuous	1.46	0.98	-0.46, 3.38	0.14	1.20	0.24	0.72, 1.68	<0.001	0.26	1.00	-1.70, 2.22	0.80
			Discrete	1.58	0.99	-0.35, 3.52	0.11	1.28	0.25	0.79, 1.76	<0.001	0.31	1.01	-1.67, 2.28	0.76

TABLE F.8: Estimated slope-change effect on ppFEF_{25-75} when comparing post-ivacaftor period to pre-ivacaftor period

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 2: Eligible & Ineligible Comparison - Step-Change Effect											
				NTE (B vs A)				NCE (D vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	1.37	3.04	-4.59, 7.32	0.65	-4.48	3.12	-10.60, 1.63	0.15	5.85	3.97	-1.94, 13.64	0.14
			Discrete	1.22	2.99	-4.64, 7.07	0.68	-5.13	2.93	-10.88, 0.62	0.080	6.34	3.67	-0.85, 13.54	0.084
		Adjusted	Continuous	3.78	2.77	-1.66, 9.22	0.17	-3.36	2.71	-8.67, 1.94	0.21	7.15	3.66	-0.02, 14.31	0.051
			Discrete	3.21	2.65	-1.98, 8.40	0.23	-4.00	2.58	-9.07, 1.06	0.12	7.21	3.29	0.76, 13.66	0.028
	Exchangeable	Unadjusted	Continuous	4.18	1.84	0.58, 7.79	0.023	-1.00	2.16	0.58, 7.79	0.023	5.18	2.88	-0.47, 10.83	0.072
			Discrete	4.13	1.85	0.50, 7.77	0.026	-1.19	2.22	-5.55, 3.16	0.59	5.33	2.91	-0.38, 11.04	0.067
		Adjusted	Continuous	4.64	1.82	1.08, 8.20	0.011	-1.25	2.10	-5.37, 2.87	0.55	5.88	2.82	0.36, 11.41	0.037
			Discrete	4.47	1.83	0.89, 8.05	0.014	-1.45	2.15	-5.67, 2.78	0.50	5.92	2.83	0.38, 11.46	0.036
	Unstructured	Unadjusted	Discrete	4.22	1.91	0.47, 7.97	0.027	-1.47	2.23	-5.85, 2.91	0.51	5.69	3.00	-0.19, 11.58	0.058
		Adjusted		4.53	1.87	0.86, 8.20	0.016	-1.63	2.17	-5.88, 2.61	0.45	6.16	2.92	0.43, 11.89	0.035
Fixed-Effects		Unadjusted	Continuous	4.83	1.93	1.05, 8.60	0.012	0.09	2.23	-4.27, 4.45	0.97	4.73	3.07	-1.28, 10.75	0.12
			Discrete	4.82	1.94	1.02, 8.63	0.013	-0.07	2.31	-4.59, 4.45	0.98	4.89	3.11	-1.21, 10.99	0.12
Mixed-Effects	Combined	Unadjusted	Continuous	4.37	1.87	0.71, 8.04	0.019	-1.10	2.16	-5.34, 3.13	0.61	5.47	2.93	-0.26, 11.21	0.061
			Discrete	4.30	1.88	0.62, 7.98	0.022	-1.27	2.23	-5.63, 3.10	0.57	5.57	2.95	-0.20, 11.35	0.059
		Adjusted	Continuous	4.74	1.83	1.14, 8.33	0.010	-1.28	2.10	-5.40, 2.85	0.54	6.01	2.85	0.43, 11.59	0.035
			Discrete	4.56	1.84	0.96, 8.17	0.013	-1.46	2.15	-5.68, 2.76	0.50	6.03	2.85	0.44, 11.61	0.034
	By Group	Unadjusted	Continuous	4.35	1.85	0.72, 7.98	0.019	-1.09	2.17	-5.35, 3.17	0.62	5.44	2.93	-0.30, 11.17	0.063
			Discrete	4.29	1.87	0.63, 7.94	0.022	-1.26	2.25	-5.66, 3.14	0.58	5.55	2.95	-0.24, 11.33	0.061
		Adjusted	Continuous	4.75	1.83	1.17, 8.33	0.009	-1.20	2.12	-5.35, 2.95	0.57	5.95	2.86	0.34, 11.56	0.037
			Discrete	4.60	1.84	1.00, 8.20	0.012	-1.38	2.18	-5.65, 2.90	0.53	5.98	2.88	0.34, 11.62	0.038

TABLE F.9: Estimated step-change effect on ppFEF_{25-75} when comparing those eligible for ivacaftor to those ineligible for ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Analysis 2: Eligible & Ineligible Comparison - Slope-Change Effect											
				NTE (B vs A)				NCE (D vs C)				NCCTE			
				est.	SE	95% CI	p	est.	SE	95% CI	p	est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	1.35	0.98	-0.58, 3.27	0.17	1.03	1.38	-1.67, 3.74	0.45	0.31	1.72	-3.07, 3.69	0.86
			Discrete	1.41	0.95	-0.46,3.28	0.14	1.37	1.26	-1.11, 3.85	0.28	0.04	1.66	-3.21, 3.29	0.98
		Adjusted	Continuous	1.06	0.93	-0.76, 2.88	0.25	0.67	1.23	-1.74, 3.07	0.59	0.39	1.54	-2.63, 3.42	0.80
			Discrete	1.30	0.87	-0.41, 3.00	0.14	1.02	1.10	-1.15, 3.18	0.36	0.28	1.41	-2.48, 3.05	0.84
	Exchangeable	Unadjusted	Continuous	0.99	0.59	-0.17, 2.14	0.093	0.65	0.90	-1.11, 2.41	0.47	0.34	1.14	-1.90, 2.58	0.77
			Discrete	0.98	0.58	-0.16, 2.12	0.091	0.70	0.91	-1.08, 2.48	0.44	0.29	1.15	-1.97, 2.55	0.80
		Adjusted	Continuous	1.04	0.61	-0.15, 2.24	0.087	0.78	0.89	-0.96, 2.54	0.38	0.26	1.15	-1.99, 2.51	0.82
			Discrete	1.10	0.60	-0.09, 2.28	0.069	0.84	0.90	-0.93, 2.61	0.35	0.26	1.15	-2.00, 2.52	0.82
	Unstructured	Unadjusted	Discrete	0.94	0.58	-0.21, 2.08	0.11	0.74	0.89	-1.01, 2.50	0.40	0.19	1.12	-2.00, 2.39	0.86
		Adjusted		1.09	0.60	-0.09, 2.27	0.070	0.93	0.90	-0.83, 2.69	0.30	0.16	1.13	-2.04, 2.37	0.88
Fixed-Effects		Unadjusted	Continuous	0.99	0.58	-0.16, 2.14	0.090	0.63	0.90	-1.13, 2.40	0.48	0.36	1.14	-1.88, 2.59	0.75
			Discrete	0.96	0.58	-0.18, 2.09	0.098	0.63	0.93	-1.19, 2.45	0.50	0.33	1.16	-1.96, 2.61	0.78
Mixed-Effects	Combined	Unadjusted	Continuous	0.91	0.58	-0.24, 2.05	0.12	0.74	0.89	-1.00, 2.48	0.41	0.17	1.13	-2.05, 2.38	0.88
			Discrete	0.92	0.58	-0.21, 2.06	0.11	0.79	0.90	-0.98, 2.56	0.38	0.13	1.14	-2.11, 2.37	0.91
		Adjusted	Continuous	0.99	0.60	-0.19, 2.18	0.10	0.88	0.89	-0.87, 2.62	0.32	0.12	1.14	-2.12, 2.36	0.92
			Discrete	1.06	0.60	-0.12, 2.24	0.078	0.95	0.91	-0.83, 2.72	0.30	0.11	1.15	-2.14, 2.37	0.92
	By Group	Unadjusted	Continuous	0.94	0.58	-0.21, 2.08	0.11	0.71	0.89	-1.03, 2.44	0.43	0.23	1.13	-1.98, 2.44	0.84
			Discrete	0.95	0.58	-0.18, 2.09	0.10	0.77	0.91	-1.01, 2.54	0.40	0.18	1.14	-2.06, 2.43	0.87
		Adjusted	Continuous	1.00	0.60	-0.17, 2.18	0.095	0.85	0.89	-0.89, 2.59	0.34	0.15	1.14	-2.08, 2.38	0.89
			Discrete	1.06	0.60	-0.11, 2.23	0.077	0.92	0.91	-0.86, 2.70	0.31	0.14	1.15	-2.13, 2.40	0.90

TABLE F.10: Estimated slope-change effect on ppFEF_{25-75} when comparing those eligible for ivacaftor to those ineligible for ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Combined Analysis Step-Change Effect			
				est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	4.77	2.56	-0.25, 9.80	0.063
			Discrete	4.79	2.48	-0.07, 9.64	0.053
		Adjusted	Continuous	5.75	2.36	1.12, 10.38	0.015
			Discrete	5.30	2.22	0.94, 9.65	0.017
	Exchangeable	Unadjusted	Continuous	4.93	1.86	1.30, 8.57	0.008
			Discrete	4.88	1.87	1.21, 8.54	0.009
		Adjusted	Continuous	5.27	1.83	1.68, 8.86	0.004
			Discrete	5.09	1.84	1.48, 8.69	0.006
	Unstructured	Unadjusted	Discrete	4.90	1.93	1.12, 8.67	0.011
		Adjusted		5.09	1.89	1.39, 8.80	0.007
Fixed-Effects		Unadjusted	Continuous	4.83	1.93	1.05, 8.60	0.012
			Discrete	4.82	1.94	1.02, 8.63	0.013
Mixed-Effects	Combined	Unadjusted	Continuous	5.06	1.90	1.33, 8.79	0.008
			Discrete	4.98	1.91	1.24, 8.71	0.009
		Adjusted	Continuous	5.35	1.86	1.70, 9.00	0.006
			Discrete	5.16	1.87	1.51, 8.82	0.006
	By Group	Unadjusted	Continuous	4.96	1.87	1.29, 8.62	0.008
			Discrete	4.89	1.88	1.21, 8.58	0.009
		Adjusted	Continuous	5.23	1.84	1.62, 8.85	0.004
			Discrete	5.09	1.85	1.46, 8.73	0.006

TABLE F.11: Estimated step-change effect on ppFEF_{25-75} from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time-period prior to the availability of ivacaftor

Model	Matrix Structure	$\boldsymbol{H}$	Time	Combined Analysis Slope-Change Effect			
				est.	SE	95% CI	p
Marginal	Independent	Unadjusted	Continuous	0.72	1.34	-1.91, 3.34	0.59
			Discrete	0.64	1.33	-1.97, 3.25	0.63
		Adjusted	Continuous	0.67	1.15	-1.59, 2.93	0.56
			Discrete	0.80	1.09	-1.32, 2.93	0.46
	Exchangeable	Unadjusted	Continuous	0.37	0.93	-1.45, 2.19	0.69
			Discrete	0.39	0.93	-1.43, 2.20	0.68
		Adjusted	Continuous	0.36	0.92	-1.44, 2.15	0.69
			Discrete	0.42	0.91	-1.37, 2.22	0.64
	Unstructured	Unadjusted	Discrete	0.35	0.90	-1.41, 2.11	0.70
		Adjusted		0.41	0.88	-1.32, 2.14	0.64
Fixed-Effects		Unadjusted	Continuous	0.33	0.96	-1.55, 2.22	0.73
			Discrete	0.34	0.97	-1.55, 2.24	0.72
Mixed-Effects	Combined	Unadjusted	Continuous	0.26	0.92	-1.54, 2.05	0.78
			Discrete	0.28	0.92	-1.51, 2.08	0.76
		Adjusted	Continuous	0.26	0.91	-1.52, 2.05	0.77
			Discrete	0.33	0.91	-1.46, 2.12	0.72
	By Group	Unadjusted	Continuous	0.33	0.91	-1.46, 2.12	0.72
			Discrete	0.34	0.92	-1.46, 2.14	0.71
		Adjusted	Continuous	0.31	0.92	-1.50, 2.11	0.74
			Discrete	0.32	0.92	-1.47, 2.12	0.72

TABLE F.12: Estimated slope-change effect on ppFEF_{25-75} from combined analysis comparing those currently receiving ivacaftor both to those currently not receiving ivacaftor and those in the time-period prior to the availability of ivacaftor