Significant factors and indicators of NHS providers' performance in meeting referral to treatment waiting time standards

MATH5872M: Dissertation in Data Science and Analytics

Supervised by:
[redacted] (University of Leeds) and [redacted] (Department of Health)

Submitted in accordance with the requirements for the degree of

Master of Science in Data Science and Analytics

The University of Leeds, School of Mathematics

August 2017

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Abstract

NHS providers are subject to waiting time standards that stipulate that a patient on an elective care pathway should wait no more than 18 weeks for treatment. However, performance is declining, and the factors influencing national and provider-level performance need to be better understood. The monthly data publicly published by the Department of Health is an historical record of provider performance; there is also a need for insight to aid future performance. By understanding the factors that influence performance and identifying characteristics of failing providers, the future performance of provider waiting times can be improved. This thesis uses statistical methods such as time series, classification trees, logistic regression and distribution fitting to offer a data-driven approach to identify key factors and early indicators for 190 NHS providers. This thesis finds that the significant factors of provider performance are providers who are previously failing, size of waiting lists and the type of treatment. This research and analysis provides a platform for a series of further studies that can best address how provider and national waiting times are likely to change in the future and how this might be improved.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The National Health Service (NHS) in England uses data to improve performance and provide public accountability to its patients. A key performance measure to assess the quality of elective (planned, non-urgent) care is the amount of time a patient waits from referral by a GP or healthcare professional to further treatment. This consultant-led referral to treatment (RTT) is subject to the NHS pledge to be treated within maximum RTT waiting times. Maximum RTT waiting times are subject to revision; the current 18 week target was introduced in April 2012 (NHS England, 2017a). It was brought in to give patients an informed choice of care, improve patients' experiences and ensure an operational standard for commissioning groups and providers to improve performance. Consequently, driven by the introduction of standards and the publication of waiting time data, the NHS England *Five Year Forward View* (2014a) championed significantly shorter waiting times since 1990 — but has since noted concerns about an increase in recent years RTT waiting time performance (NHS England, 2017b). This study of the consultant-led RTT waiting time data will investigate the significant factors and early indicators for individual providers and national performance.

This introductory chapter outlines the key terminology relating to the NHS and referral-to-treatment pathways. Chapter 2 outlines the datasets used, along with the methods and approaches taken to prepare the data for analysis. The methodology of models is outlined in Chapter 3, introducing summary statistics, time series analysis, classification trees, logistic regression and distribution fitting models. This includes a technical overview of each method and application using computer software. Chapter 4 explores the results and interpretations in detail, as well as the limitations of each method. Finally, Chapter 5 draws together the conclusions of this report for national and provider-level performance. A final summary of limitations and potential for future work is also presented here.

1.1 Key terminology

To provide clarity, the following definitions will be used throughout this report. These definitions are annotated from and further explained in the *RTT Annual Report 2016/17* (NHS

England, 2017a) and *RTT Rules Suite* (Department of Health, 2015a).

1.1.1 NHS England

Performance is measured for various levels within the NHS architecture. National performance encompasses all NHS treatment, further split by four geographical regions: North, Midlands & East, London, and South. Local services are currently allocated by 207 Clinically Commissioning Groups (CCGs) (NHS Clinical Commisioners, nd), who are responsible for commissioning services from a range of providers (including NHS trusts and private sector organisations). A provider is an organisation that provides care to patients across various type of service/treatment function, such as Dermatology, Neurology or General Surgery.

1.1.2 Referral to treatment (RTT) pathway

Patients who are referred for non-emergency (elective) care by an NHS consultant start a referral to treatment (RTT) pathway. The pathway measures (in weeks) how long it takes for the patient to be treated, or how long they are still waiting, recorded at the end of each month. A patient may have more than one RTT referral, which are all included individually in the pathway waiting times data.

Clock start: The date that the RTT pathway starts, as soon as the consultant-led referral letter, or online e-referral, is received by the subsequent NHS provider.

Clock stop: The date that the RTT pathway ends, usually if first treatment begins, patient declines treatment or a clinical decision is made not to treat the patient. There are several scenarios that may end the RTT pathway (or, 'stop the clock'):

- Patient is admitted to hospital for treatment, such as an operation.
- Treatment begins without hospital admission, such as self-monitoring or monitoring by a healthcare professional.
- Initial contact is made for fitting medical devices, or advice from a clinician is given to manage the patient's condition.
- Healthcare professional makes a decision to not treat, reassessing the need for further treatment.
- Patient declines or does not attend first treatment, is medically unfit due to multiple conditions or dies prior to receiving treatment.

1.1.3 Incomplete pathway

A patient still waiting for treatment at the end of each month is recorded as waiting x weeks and an incomplete pathway — often referred to colloquially as the NHS waiting list. The size of the RTT waiting list is simply the total number of incomplete RTT pathways.

Incomplete pathway data should be considered as a snapshot of patients still waiting for treatment at the end of the month. This is because any incomplete pathway waiting longer than four to five weeks will have been in last month's data — this applies for a large proportion of incomplete pathways.

1.1.4 Completed pathway

A patient who has received their first definitive treatment is recorded as a completed (**admitted** or **non-admitted**) pathway. Though completed pathways are no longer subject to waiting time standards, it is still mandatory for this data to be submitted by NHS providers.

Admitted pathway: A pathway that ends during the month due to an admission to hospital. This can include an overnight stay or day case admission, also known as inpatient treatments.

Non-admitted pathway: A pathway that ends during the month not requiring an admission to a hospital. This can include community care, monitoring or non-treatment, and are also known as outpatient treatments.

1.1.5 Waiting time standards

To improve patient satisfaction, patients have the right to start treatment without unnecessary delay — that is, to minimise the length of an incomplete pathway. Various operational waiting time standards have been brought in since the introduction of the standard by the government in 2005 (Department of Health, 2004). The current NHS Constitution pledge (Department of Health, 2015b), introduced in April 2012, states:

- Patients have the right to consultant-led treatment within 18 weeks from their first point of referral.
- An operational target stating that 92% of incomplete pathways should be no longer than 18 weeks.
- A zero tolerance of RTT pathway waits longer than 52 weeks (introduced April 2013).

1.2 Example scenario: Mrs Smith's RTT pathway

Mrs Smith has a minor fall at home and books an appointment to see her GP.

1. The GP assesses there is no urgent treatment needed, but refers Mrs Smith for an X-ray as a precautionary measure. This **starts the clock** for Mrs Smith's consultant-led RTT pathway.

2. Mrs Smith waits several weeks for an appointment to become available. As such, the number of weeks she has been waiting at the end of the month is included in the relevant month's — and subsequent months' — **incomplete pathway** waiting times.

3. Mrs Smith is able to complete her X-ray appointment in week 10. This **stops the clock** for her RTT pathway.

4. Mrs Smith's pathway is no longer recorded as an incomplete pathway. It is recorded in the *9-10 weeks* **non-admitted pathway** (since her appointment did not require an admission to hospital) waiting time data.

Further treatment as a result of Mrs Smith's x-ray will start a new pathway.

Chapter 2

Consultant-led referral to treatment (RTT) waiting time data

This chapter briefly introduces the datasets used throughout the project, with a summary of the steps taken to prepare, pre-process and validate the data ready for analysis.

2.1 Explaining the data

This study of consultant-led elective treatment waiting times uses monthly and time series public datasets from the NHS England Statistics website, available at `https://www.england.nhs.uk/statistics/statistical-work-areas/rtt-waiting-times/`. The datasets used include aggregated national time series, provider-based time series and provider-level data published monthly between August 2007 and March 2017. The aim of this study is to provide an analysis of the incomplete waiting list nationally and for individual providers.

There are two types of raw data files provided by NHS England: national or summary time series (one dataset) or monthly data (116 datasets) between August 2007 and March 2017. Preparation of the data and the analysis that follows is split into two core sections: national performance and provider-level performance.

2.1.1 National data

The national performance data used comprises of two datasets: a time series of published incomplete RTT pathways and a time series estimation of missing total incomplete waiting data. Table 2.1 summarises the variables included in the analysis.

The missing estimates dataset is used to adjust the total number of incomplete pathways only, but ignored for outcome measures such as the percentage of pathways waiting within 18 weeks as this would be unsuitable to adjust accurately. This is further explained in Section 2.2.2.

Variable	Type	Description
Published national data		
Date	date	Year and month of the incomplete pathway waiting list snapshots. Annual data collection period is April-March.
No. $\leq$ 18 weeks	discrete	Number of pathways waiting within 18 weeks of referral.
% $\leq$ 18 weeks	percentage	Percentage of total pathways waiting within 18 weeks.
No. $>$ 18 weeks	discrete	Number of pathways waiting longer than 18 weeks.
No. $>$ 52 weeks	discrete	Number of pathways waiting longer than 52 weeks.
% $>$ 52 weeks	percentage	Percentage of total pathways waiting longer than 52 weeks.
Total waiting (mill)	discrete	Total number of published incomplete pathways.
Missing data estimates		
Missing estimate (mill)	discrete	Estimates for the number of incomplete pathways not recorded by providers (*see Section 2.2.3*).
Adjusted total waiting (mill)	discrete	Summed published and missing data estimates of the total number of incomplete pathways.

Table 2.1: *Summary of national data variables.*

2.1.2 Provider-level data

The provider-level data consists of two types of dataset: monthly provider data (August 2007-March 2017; 116 datasets) and a time series summary of the total incomplete pathways and percentage of pathways within 18 weeks (2 datasets) for each provider. Table 2.2 gives a summary and explanation of the variables included in later analysis.

The main aim of this study is to assess the performance of providers to meet the operational incomplete waiting standard introduced in April 2012, so the final number of monthly datasets used is 60 (April 2012 - March 2017). Similarly, the time series data is cleaned to exclude data from pre-April 2012. Further explanation of data cleaning steps and methods are discussed in Section 2.2.

Variable	Type	Description
Monthly provider-level data		
Provider name/code	categorical	Name and three-digit code of each organisation providing elective care *(e.g. RAL = Royal Free London Trust).*
Region code	categorical	Three-digit code corresponding to four regions: *Y54 = North* *Y55 = Midlands & East* *Y56 = London* *Y57 = South*
Treatment function	categorical	Each RTT is assigned one of 19 specialities to best capture the main treatment area *(e.g. Dermatology, Cardiology, and Other)*.
No. of pathways by week since referral	discrete	Number of incomplete pathways in weekly time bands (*Further explained in Table A.1 in Appendix A*).
Total waiting	discrete	Total number of incomplete pathways by provider.
No. $\leq$ 18 weeks	discrete	Number of pathways waiting within 18 weeks of referral.
% $\leq$ 18 weeks	percentage	Percentage of total pathways waiting within 18 weeks.
Time series provider-level data		
Provider code	categorical	As previous.
Date	date	Year and month of the incomplete pathway waiting list snapshots.
Total waiting	discrete	As previous.
% $\leq$ 18 weeks	percentage	As previous.

Table 2.2: *Summary of provider-level data variables.*

2.2 Data preparation

The data preparation process is used to identify incomplete and inaccurate records in each dataset. It has also been used to assess irrelevant data that may be omitted for this study. Since the publication of data spans five years, with significant improvements and rigour now placed on data submission standards, there are inconsistent styles of reporting data. The types of cleaning required for the datasets include:

- Redundant data
- Missing data
- Data entry design
- Coding inconsistencies

The following section discusses the process of identifying and dealing with incomplete, inaccurate and irrelevant data — an iterative process outlined in Figure 2.1, as part of the cyclic Cross Industry Standard Process (CRISP) for data mining framework (Chapman et al., 2000).

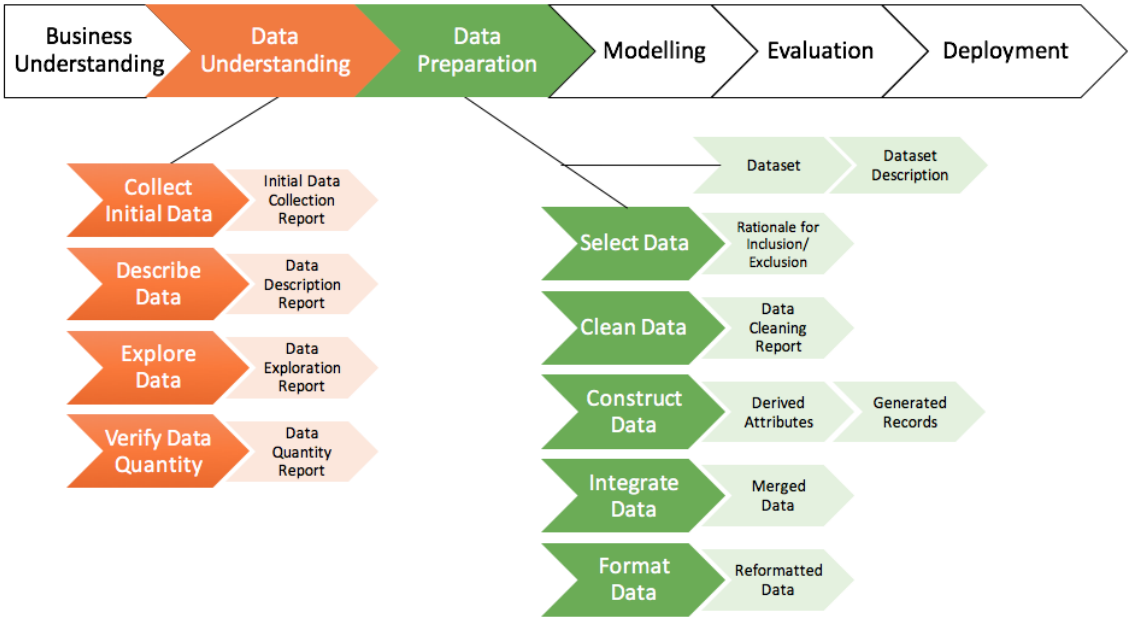

Figure 2.1: *The data understanding and preparation stages of data analysis; annotated from the CRISP Data Mining framework guide (Chapman et al., 2000). Source:* `https://www.the-modeling-agency.com/crisp-dm.pdf`.

2.2.1 Redundant data

One of the first steps before data cleaning, is to take initial steps to remove redundant data. The aim of this study is to assess the performance of RTT pathways by providers and nationally. For exploratory national analysis, all time data is considered. However, for further national analysis and all provider-level analysis the data prior to April 2012 is omitted. This corresponds to the introduction of the operational waiting time target — 92% of incomplete pathways should be within 18 weeks — which is the main outcome measure of RTT performance. The reason to exclude all provider-level analysis pre-April 2012 is due to the high prevalence of incomplete and inaccurate data.

Redundant data is prevalent in the monthly submissions, where providers that are no longer active have been temporarily recorded as zero pathway observations. The number of redundant submissions is outlined in Table 2.3. By removing these false observations, there is no loss of information and the accuracy of summary statistics, such as the mean, is improved.

Another example of redundant data is a provider that has accurately submitted zero pathways for some months. This is usually noticed and removed after a few iterations of the data cleaning

process (Figure 2.1). Although the observations are valid, they have no added value to models or analysis. Furthermore, a zero pathway month-provider observation can cause complications for variables such as meeting the pathway target, altering the expected mean of the overall data despite having no pathways. Consequently, an additional total of 13 observations have been removed:

- 8 zero-pathway months for provider *RYX* (November 2013 to June 2014).
- 5 zero-pathway months for provider *RT5* (November 2014 to March 2015).

2.2.2 Missing data

The first stage of cleaning the data was to identify empty and non-reported data. These are classified as two types of missing data:

- Empty cells
- DNS — 'did not submit'

Empty or **blank cells** must be considered with the context of the dataset. The lack of data should explain that there is no data available for this variable. This is particularly prevalent in the time series dataset, as a sequence of empty cells suggests there was no available data to be recorded in that time period. An example in the provider-level datasets is a provider that changes its provider code, closes, or merges with another provider, so its original code is no longer active within the April 2012 - March 2017 period.

The approach to resolving closed or merging providers was to remove them from the dataset. This is because closed providers are often planned or due to extreme circumstances, resulting in external factors such as reductions in funding, major workflow changes or less pressure on performance standards. Similarly, there were three providers opening after April 2012 which were removed from the dataset. Table 2.3 gives a summary of the number of merged or closed providers for each month.

The effect of merged providers is a key consideration for providers remaining in the data, where the total number of pathways may suddenly increase or decrease as pathways are transferred within the system. It was not possible to track where all pathways were transferred in the event of closed or merging providers, since pathways would often split across several other providers. This might explain some dramatic changes in performance.

However, empty cells might also be due to an error, such that there exists or could exist data for that record but which has not been included in the dataset. This is noticeable as a provider having many months of pathways with no clear sign of tending to zero, and randomly missing pathways for one or two months followed by a return to the norm in subsequent months.

For these rarer instances, an effort is made to validate and input the missing data. If the data cannot be located in other datasets or from request to the data providers, a decision is made to

classify it as truly missing data (i.e. no data available) — and dealt with as previously mentioned — or as DNS data.

For the **DNS** data, there are two main approaches to dealing with the missing data: estimating the datapoint, or deleting the record. The latter option results in missing out on other valuable data within this that has been recorded successfully, and is hence avoided. Instead, an estimation of the data was created.

To estimate the data, the last recorded data point is assumed for the missing data period. This is shown in Figure 2.2, where provider *RHW* did not submit the total number of incomplete pathways data between July and December 2014, with estimated data from June 2014. Estimating the data based on the last reported submission is the same method used for the missing data estimates in the national time series dataset as published by the Department of Health (see Figure A.1 in Appendix A). Consequently, there is a clear and logical continuity of how missing data has been dealt with across different datasets.

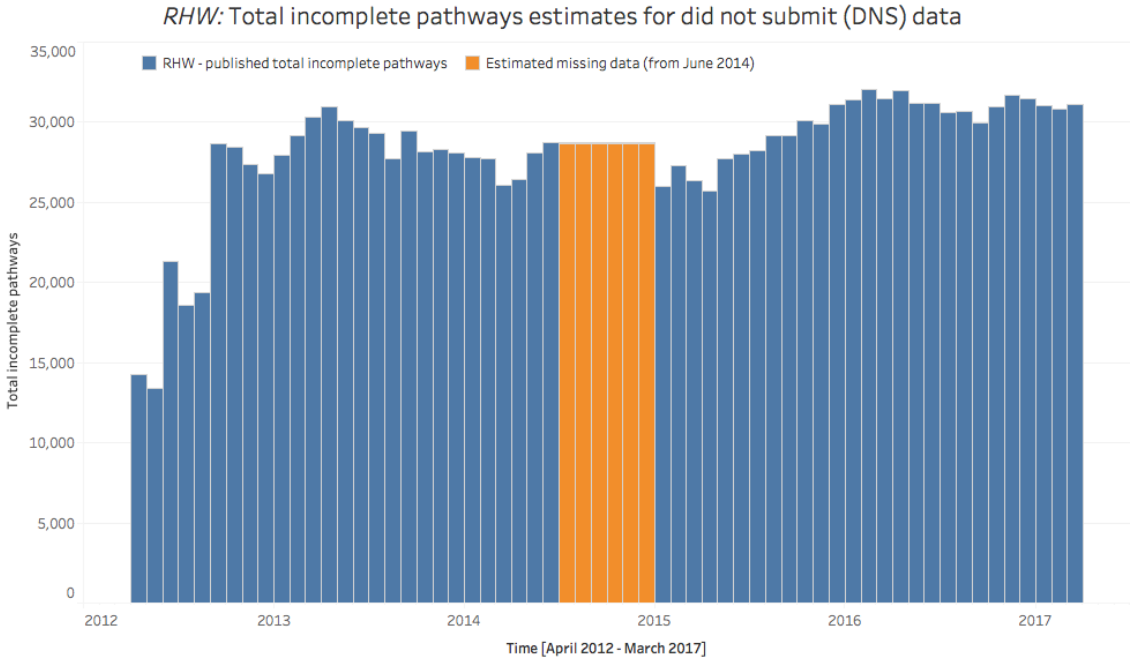

Figure 2.2: *Estimated data (orange) for total number of incomplete pathways for provider RHW between April 2012 and March 2017. Estimated from last submitted data — June 2014.*

Some considerations require further analysis based on these estimations. Though in many cases this estimation is only taken for a few months at a time, some providers have estimated data for up to 34 consecutive months (out of 60). A total of 289 months of estimated data have been included in the final dataset, as further detailed in Table 2.4.

A summary of how many providers are included in the final dataset:

- **215** providers submitted data in April 2012.
- **25** providers were removed, as they closed before March 2017.
- **3** new providers opened post-April 2012 were not included.
- **190** providers with complete data for April 2012 - March 2017.

Month	Active providers	(Change from previous month) Opened/new code	Closed/merged	(Active provider data) Submitted data	Did not submit	Redundant*
Apr-12	215	0	0	214	1	2
May-12	213	0	2	212	1	2
Jun-12	212	0	1	211	1	2
Jul-12	211	0	1	211	0	0
Aug-12	210	0	1	209	1	0
Sep-12	210	0	0	209	1	0
Oct-12	207	0	3	205	2	0
Nov-12	207	0	0	205	2	0
Dec-12	207	0	0	205	2	0
Jan-13	207	0	0	205	2	0
Feb-13	207	0	0	205	2	0
Mar-13	207	0	0	206	1	0
Apr-13	203	0	4	203	0	1
May-13	204	0	0	204	0	1
Jun-13	203	0	0	203	0	0
Jul-13	203	0	0	203	0	0
Aug-13	203	0	0	203	0	0
Sep-13	203	0	0	202	1	0
Oct-13	202	0	1	200	2	0
Nov-13	202	0	0	200	2	1
Dec-13	202	0	0	199	3	1
Jan-14	201	0	1	198	3	1
Feb-14	201	0	0	196	5	1
Mar-14	199	0	0	192	7	2
Apr-14	201	0	0	196	5	2
May-14	201	0	0	196	5	2
Jun-14	201	0	0	197	4	1
Jul-14	201	1	1	195	6	0
Aug-14	201	0	0	195	6	0
Sep-14	201	0	0	194	7	0
Oct-14	199	1	3	192	7	0
Nov-14	198	0	1	191	7	1
Dec-14	198	0	0	190	8	1
Jan-15	197	0	0	190	7	0
Feb-15	195	0	2	188	7	0
Mar-15	195	0	0	188	7	1
Apr-15	195	0	0	187	8	0
May-15	195	0	0	187	8	0
Jun-15	194	0	1	187	7	0
Jul-15	194	0	0	185	9	0
Aug-15	194	0	0	185	9	0
Sep-15	193	0	1	184	9	0
Oct-15	193	0	0	183	10	0
Nov-15	193	0	0	183	10	0
Dec-15	193	0	0	183	10	0
Jan-16	193	0	0	184	9	0
Feb-16	193	0	0	184	9	0
Mar-16	193	0	0	186	7	0
Apr-16	194	1	1	186	8	1
May-16	193	0	0	185	8	0
Jun-16	194	0	0	185	9	0
Jul-16	194	0	0	185	9	0
Aug-16	194	0	0	185	9	0
Sep-16	194	0	0	185	9	0
Oct-16	194	0	0	188	6	0
Nov-16	194	0	0	189	5	0
Dec-16	194	0	0	189	5	0
Jan-17	194	0	0	190	4	0
Feb-17	193	0	1	190	3	0
Mar-17	193	0	0	190	3	0
Total		3	25			

Table 2.3: *Summary of the number of providers submitting data each month.*
**Redundant data submissions are previously closed providers that incorrectly remain in the data as zero pathways.*

Org code	DNS start	DNS end	Months estimated post-April 12	Estimated from
RYJ	Dec-11	May-12	2	Nov-11
RFR	Jun-12	Jun-12	1	May-12
RVJ	Aug-12	Feb-13	7	Jul-12
RAE	Oct-12	Mar-13	6	Sep-12
RKE	Sep-13	May-14	9	Aug-13
RF4	Dec-13	Sep-16	34	Nov-13
RGQ	Feb-14	Mar-14	2	Jan-14
RTG	Feb-14	Mar-14	2	Jan-14
RBK	Mar-14	Sep-16	31	Feb-14
RMP	Mar-14	Nov-14	9	Feb-14
RHW	Jul-14	Dec-14	6	Jun-14
RQW	Jul-14	Sep-15	15	Jun-14
RR1	Jul-14	Feb-15	8	Jun-14
R1H	Sep-14	Mar-17	31	Aug-14
RDE	Dec-14	Apr-15	5	Nov-14
RPA	Dec-14	May-15	6	Nov-14
RM2	Mar-15	Jul-15	5	Feb-15
RJZ	Apr-15	Feb-16	11	Mar-15
RLQ	May-15	Nov-16	19	Apr-15
RT5	Jul-15	Nov-15	5	Jun-15
RV9	Jul-15	Dec-15	6	Jun-15
RP4	Aug-15	Dec-16	17	Jul-15
RPA	Oct-15	Sep-16	12	Sep-15
RQX	Oct-15	Feb-16	5	Sep-15
RNQ	Dec-15	Jan-17	14	Nov-15
RJF	Apr-16	Oct-16	7	Mar-16
RJ7	Jun-16	Mar-17	10	May-16
RTE	Dec-16	Mar-17	4	Nov-16
Total			289	

Table 2.4: *Summary of estimated data for providers. Notably,* RYJ *did not submit (DNS) data between December 2011 and May 2012, however only two months have been estimated post-April 2012; also,* RPA *has two periods of DNS.*

2.2.3 Data errors

The final stage in data preparation is to check for idiosyncrasies of the data. Having omitted redundant data and estimated missing data, the data is checked for final additional problems that will affect analysis. There is no general rule for identifying these sort of errors, and are usually as a result of data recording practice, human error or software coding.

The first necessary cleaning corresponds to an idiosyncrasy in how the data has been recorded. In the national time series dataset, the median wait for incomplete RTT pathways is recorded as an estimation between 0 and 52 weeks. However, for months where the median wait is greater than 52 weeks, it has been recorded as 52+ and hence is no longer a quantitative variable. To prepare the data for visual analysis, these are adjusted to 100 to demonstrate the effect of long waiting times, and clearly marked as adjusted figures.

Another idiosyncrasy of the data is due to inconsistencies in software coding. This particular error is due to hidden formulae within Microsoft Excel. Instances of losing the number of significant figures had to be adjusted back to raw numbers, and variables input as a string would sometimes include an apostrophe preceding the data. For example, a provider code *R1A* had hidden formula *'R1A*. This causes problems for data pre-processing and data analysis, as matching variables across datasets is not possible (since coding $\textit{R1A} \neq \textit{'R1A}$). Several methods were used to amend this, such as Excel's copy and paste as values feature, and a cleaning formula such as `=RIGHT(A4,3)`, which takes a string of length 3 from cell A4, and hence removes the unwanted apostrophe.

2.3 Data pre-processing

The data pre-processing step adapts the clean data to create new variables, interpretations and formats. Combining old and new variables, along with a validation of the data, results in the final datasets that are used for analysis of RTT waiting time performance.

2.3.1 New variables

In addition to the variables provided in the original datasets — outlined in Section 2.1 — further variables were manipulated to provide insight for the analysis. These offer alternative outcome measures to be compared, as well as further predictor variables to explain the data.

The operational waiting time target for incomplete RTT pathways is that 92% should have been waiting less than 18 weeks. As an alternative measure of this percentage, a binary variable *met target* was created describing if the provider passed: **1**, or failed: **0**, the target each month. A time series dataset of this binary variable is easy to manipulate for summary results such as how long a provider meets the target, how long it fails for, and the time taken to recover from missing the target.

This variable can be further used to derive lagged variables such as provider performance in previous months. The lagged x months variable, where each value of x — an integer between 1

and 60 — corresponds to a new variable, provides extra insight for predicting future data.

With 60 separate datasets containing the most detailed information about providers, some further data pre-processing was undertaken to include these in the final dataset. The treatment function (as described in Table 2.2), indicates the number of pathways for different specialities within each provider. These variables were transformed into new variables, case mix percentage, and found for all 19 treatment types as the percentage of speciality out of all treatments by a provider. An example is given below for provider *R1A* in March 2017:

Treatment function	% of total pathways (Mar-17)
GeneralSurgery	8.9%
Urology	12.3%
TraumaOrthopaedics	12.7%
ENT	21.3%
Ophthalmology	0.9%
OralSurgery	8.6%
Neurosurgery	0.0%
PlasticSurgery	0.0%
CardiothoracicSurgery	0.0%
GeneralMedicine	6.6%
Gastroenterology	0.0%
Cardiology	0.0%
Dermatology	1.3%
ThoracicMedicine	0.0%
Neurology	0.0%
Rheumatology	0.0%
GeriatricMedicine	0.0%
Gynaecology	13.4%
Other	14.0%
Total	100.0%

As previously described, providers which did not submit data in some months (see Table 2.4) were estimated from the last available month instead.

With a number of new variables and datasets available for further summary and explanatory analysis, two final datasets were created in a format for provider-level analysis:

Provider-level by month: 11387 observations, including estimated months (190 providers $\times$ 60 months $-$ 13 zero pathway months).

Provider-level by month and case-mix: 221960 observations, excluding estimated months ((190

providers $\times$ 60 months $\times$ 20 case mixes) $-$ (289 missing months + 13 zero pathway months) $\times$ 20 case mixes)). *Note, 20 case mixes = 19 specialities + total*.

2.3.2 Data validation

Data validation provides a final check that data cleaning and data pre-processing stages are completed with minimal error. This is done by comparing summary results and randomly assigned records with original and external datasets. Some examples of data validation are shown below:

- The number of providers submitting data each month, did not submit and met the target was manually collected from the monthly provider datasets and checked against provider time series and new variables.
- The region codes for each provider were checked against an external dataset listing all NHS organisation codes and addresses as found on the NHS website.
- A random check was completed by assigning providers a number in the final dataset, and randomly inspecting five so that each variable is consistent with the original datasets.

Chapter 3

Methodology

This chapter gives an overview of the methods used to analyse the RTT waiting time data. The theory of statistical methods aided by computer software such as Microsoft Excel, Tableau and R are explained here. First, methods for acquiring and understanding summary statistics are introduced. This is followed by a theoretical overview of four models which have been selected appropriate to the dataset and aims of the analysis: time series, classification tree, logistic regression and distribution fitting. Finally, principles and methods for assessing model fit are discussed.

3.1 Summary statistics

Summary statistics are explored as an initial description of the data, using R for quantitative outputs and Tableau for visual interpretation. This exploratory analysis provides a basis for the choice of later analysis methods.

Descriptive statistics give a greater understanding of how each variable relates to the overall dataset, and in particular to the outcome variables — in this case, percentage of pathways within 18 weeks or met operational target. An example of these statistics is summarised in Table 3.1 for the provider-level data, where N is the number of observations and a range of statistics for the variables: total incomplete pathways, % within 18 weeks, met target and 19 case mix variables (*Cardiology, . . . , Urology*).

The % within 18 weeks has a mean of 94%, which suggests that the average provider-month performs above the operational waiting time target (92%). The standard deviation, 4.8%, which explains how the data varies from the mean, suggests that this percentage is fairly consistent — that is, those provider-months that vary greatly from the mean are likely extreme and due to external factors rather than random noise. Notably, the case mix percentage variables are not normally distributed as demonstrated by standard deviation exceeding the mean. It is necessary to consider the skewed nature of these variables in future analyses. As a different outcome measure, met target is a binary variable explaining a provider meeting the target. The summary shows that approximately 83% of provider-months in our data have at least 92% of their path-

Variable	N	Mean	St. Dev.	Min	Max
% within 18 weeks	**11,387**	**94%**	**4.8%**	**0%**	**100%**
Total incomplete pathways*	11,387	15,831	13,461	2	86,686
Cardiology	11,387	4.4%	6.1%	0%	69%
CardiothoracicSurgery	11,387	0.4%	2.7%	0%	33.1%
Dermatology	11,387	4.6%	5.5%	0%	80.2%
ENT	11,387	5.9%	4.8%	0%	100%
Gastroenterology	11,387	3.4%	2.9%	0%	13.3%
GeneralMedicine	11,387	1.7%	3.6%	0%	100%
GeneralSurgery	11,387	7.4%	5.8%	0%	27%
GeriatricMedicine	11,387	0.5%	0.8%	0%	15.1%
Gynaecology	11,387	5.4%	7%	0%	100%
Neurology	11,387	2.5%	5.8%	0%	71.9%
Neurosurgery	11,387	0.6%	2.2%	0%	30.5%
Ophthalmology	11,387	8.3%	9.1%	0%	100%
OralSurgery	11,387	4.5%	8.6%	0%	100%
Other	11,387	27.2%	29.2%	0%	100%
PlasticSurgery	11,387	1.3%	3.5%	0%	42.8%
Rheumatology	11,387	2.2%	6.9%	0%	100%
ThoracicMedicine	11,387	1.9%	3.5%	0%	40.9%
TraumaOrthopaedics	11,387	13.5%	15.4%	0%	100%
Urology	11,387	4.1%	2.9%	0%	21%

		Number and % for levels of categorical data			
Met target	N		*Met*	*Failed*	
	11,387		**9441 (82.9%)**	**1946 (17.1%)**	
Region		*North*	*Midlands & East*	*London*	*South*
	11,387	3480 (30.6%)	3655 (32.1%)	1672 (14.7%)	2580 (22.7%)

Table 3.1: *Summary statistics of provider-level variables. Outcome variables highlighted in bold. *Mean and standard deviation rounded to nearest whole number.*

ways waiting less than 18 weeks. The categorical variable, region, is not an even split across the full data, and separating analysis by region could explain a varying provider performance depending on a provider's location.

Visualisations created in Tableau offer another method to assess key insights. This method produces insights that may be more noticeable in a graph rather than tabular form. Two examples are a monthly graph of total incomplete pathways (Figure 3.1), and a scatterplot of the outcome variable, % of pathways within 18 weeks, and an explanatory variable, total incomplete pathways (Figure 3.2).

The monthly graph (Figure 3.1) of total incomplete pathways indicates an annual increasing

trend in waiting list size, but also introduces the possibility of seasonality — each year the total number of incomplete pathways increase dramatically with strong seasonal increases from January to May, levelling off or decreasing toward December. The next step is to explore this further in time series analysis.

The scatterplot (Figure 3.2) supports an intuitive hypothesis that there is a relationship between the number of pathways on the national waiting list and the operational waiting time target performance. To further understand this relationship, such as exploring other variables that may influence the national and provider performance, classification and regression models are chosen as the next step.

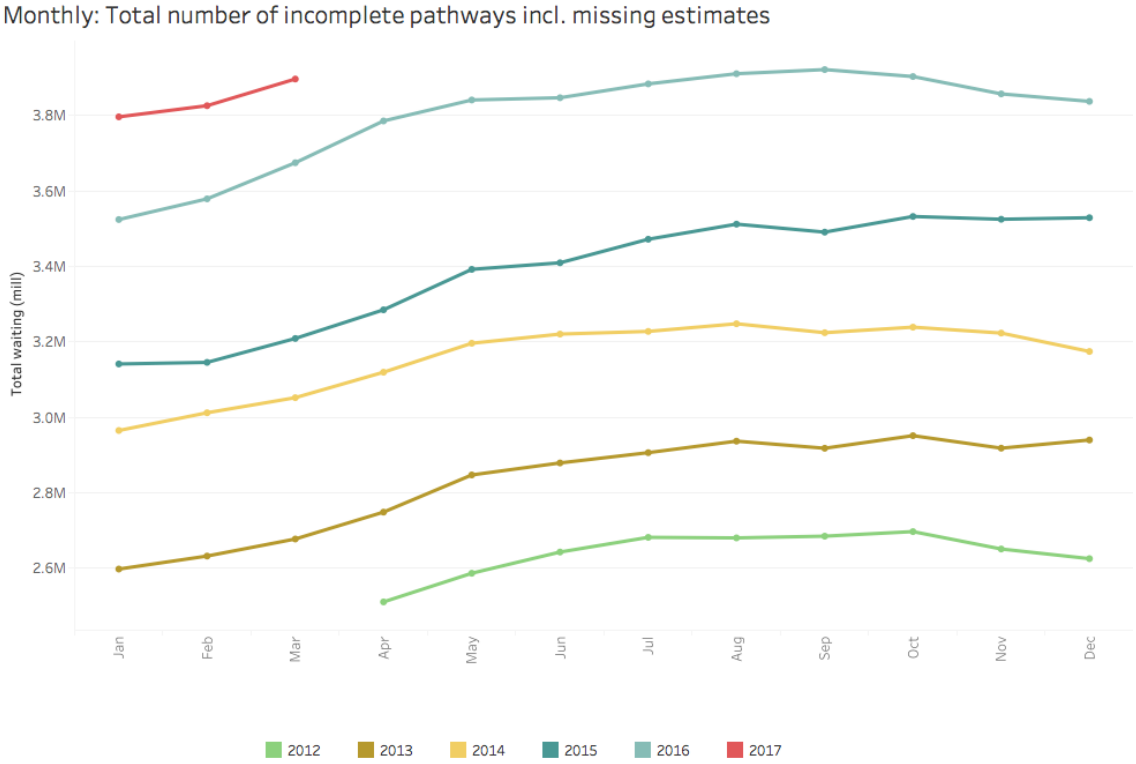

Figure 3.1: *National total number of incomplete pathways each month, split by year (colour key).*

The summary statistics explained here are a brief summary of the exploratory analysis undertaken. The next steps are to use the summary statistics to explain and justify initial findings such as seasonality; or focal points to further explore, such as which variables are driving national and provider performance.

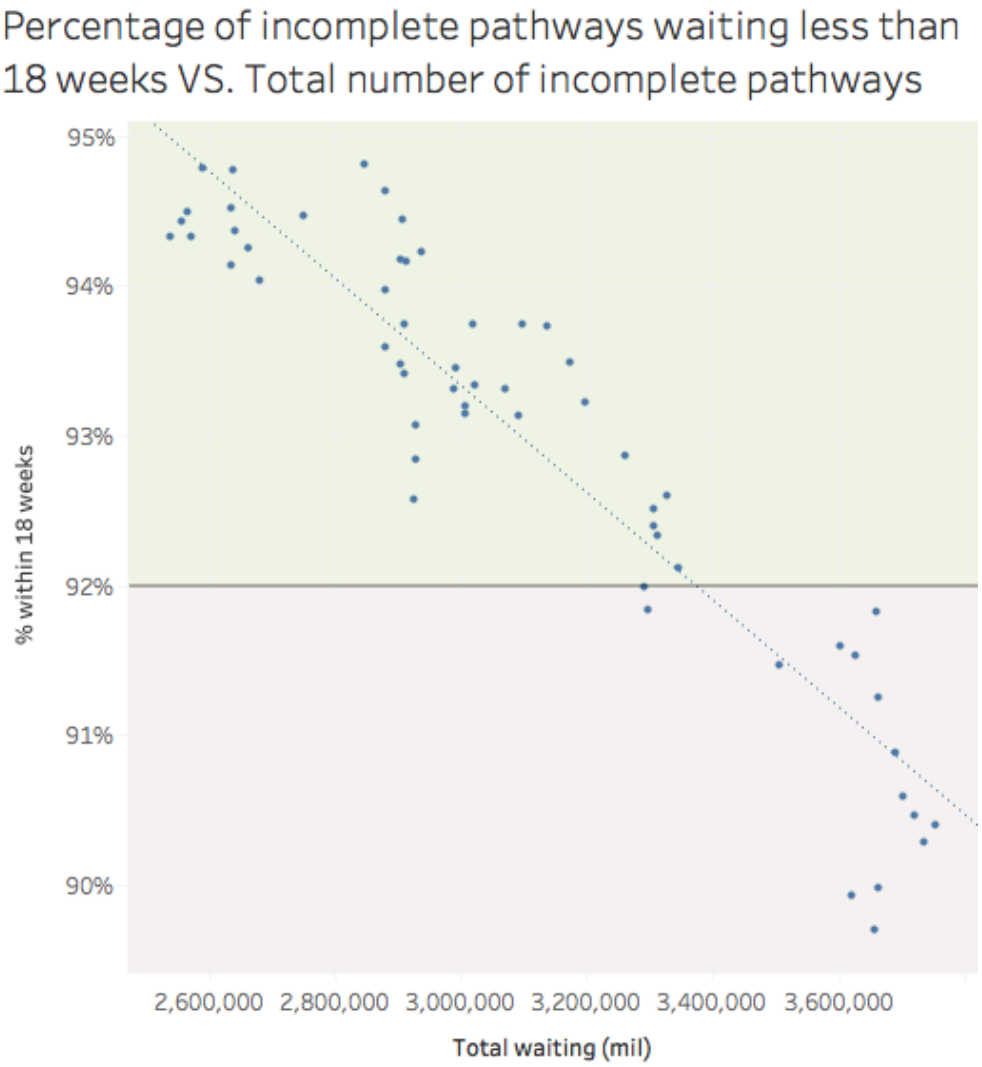

Figure 3.2: *National relationship between total incomplete pathways and % of pathways within 18 weeks. Dotted line indicates possible general trend.*

3.1.1 R syntax: *summary, stargazer*

The `summary` and `stargazer` commands in R gives a range of descriptions of each variable in the dataset — see previously in Table 3.1.

3.2 Time series analysis

Time series analysis is the first statistical tool to explore since the data describes a stochastic process in which observations are recorded at discrete weekly or monthly intervals. A time series approach is useful for understanding and extracting statistics and characteristics of the data. Unlike other statistical methods, time series data consists of successive observations that are not independent, due to an ordered time variable. This suggests that a correlation between

subsequent time points is best explained by a dependence on past values (Shumway and Stoffer, 2000).

The notation used throughout this section is adapted from Brockwell and Davis (2002). The data is modelled as a random variable X_t, at time t = 0, $\pm 1, \pm 2, \ldots$ representing equally spaced intervals. Further, the dependence on past points can be described by notation that x_t (representing an individual observation, rather than the random variable X_t), is dependent in some way on past values $x_{t-1}, x_{t-2}, \ldots$.

One of the key outcomes is to explain the overall data as a decomposition of components, such as the additive model

$$X_t = m_t + s_t + Z_t \tag{3.1}$$

where at time t: m_t is the trend (or mean level) component; s_t is the seasonal component; and Z_t is the remainder component.

The trend is a slowly changing function representing the overall tendency of the time series, while the seasonal component explains cyclic fluctuations in a known calendar period, d. For example, a time series of daily observations with a weekly seasonal pattern would have cycle $d = 7$, such that $s_t = s_{t \pm d}$. The remainder component is random noise that is stationary, such that there is no dependence on time or between remainder observations.

Depending on how the seasonal and remainder vary over time, such as an increasing or decreasing seasonal effect, the decomposition model may be more appropriate as the multiplicative model

$$X_t = m_t \times s_t \times Z_t. \tag{3.2}$$

This can be written as the additive model (as in equation 3.1) by taking a log transformation of the components;

$$\begin{aligned} X_t &= m_t \times s_t \times Z_t \\ \log(X_t) &= \log(m_t \times s_t \times Z_t) \\ &= \log(m_t) + \log(s_t) + \log(Z_t). \end{aligned}$$

As suggested in Figure 3.1 and confirmed by a de-trended plot of total incomplete pathways in Figure 3.3, the incomplete pathways time series data involves a significant seasonal effect that does not vary over time. Therefore, the additive model is appropriate for this time series analysis.

The procedure for decomposing the additive model data is:

1. Calculate the trend component; m_t
2. De-trend the time series; $X_t - m_t$
3. Calculate the seasonal component from de-trended series; s_t

4. Calculate the random component; $Z_t = X_t - m_t - s_t$

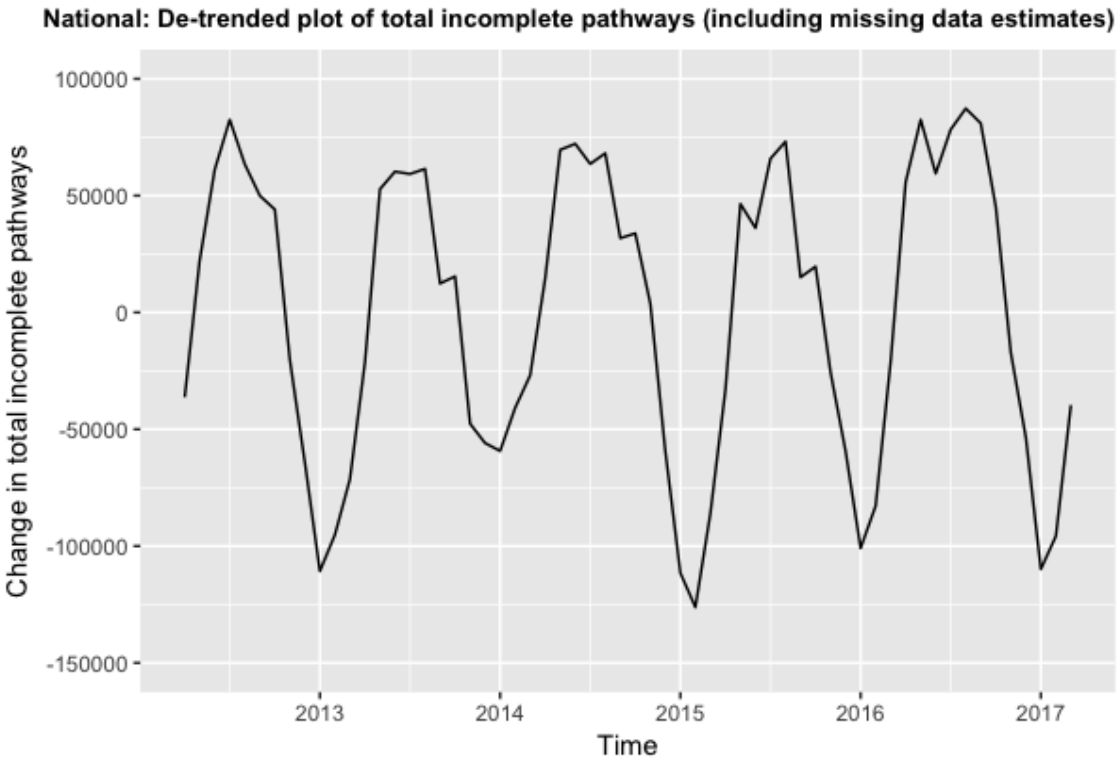

Figure 3.3: *National de-trended total number of incomplete pathways.*

The decomposed time series can be used to explain an overall trend and the effect of seasonality, such as the incomplete waiting list decreasing in December and rapidly increasing from January to May (Figure 3.1). Additionally, the de-seasonalised data can be extracted and used for further analysis. The total incomplete pathways for providers with seasonal effect removed can be interpreted as the usual capacity of a provider. This provides extra insight that may influence performance of providers meeting the operational waiting time target.

3.2.1 Loess decomposition

There are many methods for estimating the trend and seasonality as discussed by Chatfield (2004), Brockwell and Davis (2002) and Shumway and Stoffer (2000). On balance of recommendation and ease of programming, the method used to decompose the time series data is an iterative filtering procedure called STL (Seasonal and Trend decomposition using Loess), developed by Cleveland et al. (1990).

The Loess decomposition method uses locally weighted scatterplot smoothing to estimate the trend component of the data. By taking subsets of the data, least squares regression estimates are fitted. This is done by iteratively fitting local regression lines to each subset of the data, with points close to the centre of each subset weighted (also known as nearest neighbour-

hood weighting (Cleveland et al., 1990)), such that they have a greater influence on the local regression line. Equally, points which are furthest from the centre will be weighted least, and have less effect on the local regression line. The centre point is then taken as the Loess curve estimate. By repeating this procedure, moving the subset window across the dataset, a series of Loess estimates are fitted to create a smooth trend estimate of the overall time series.

With the trend estimated through Loess smoothing, the seasonal component can then be calculated from the de-trended series. A Loess smoothing procedure is used again for the seasonal decomposition across the time series for each month. However, with the assumption that the seasonal variation is constant across years, the method becomes a least squares regression across the whole data — for example, January has five data points representing each year — and is simply the mean value of all observations for each month.

Finally, the remainder is calculated as the difference between the de-trended series and the seasonal estimation.

There are advantages and disadvantages to using the STL procedure (Hyndman and Athanasopoulos, 2013). The main advantage is the iterative calculation of components, robustness to outliers and user-control in calculation. However, the disadvantages include the lack of model to extrapolate (useful as a basis for forecasting) and limitations for calendar variation such as the number of days in each month.

3.2.2 R syntax: *stl*

An explanation of the R package `stl` is explained here, with particular focus on criteria that optimise the process and its output.

`tsdata = ts(dataset, frequency = 12, start = c(2012, 4))` # The dataset is saved as the required time series format. Since we have monthly data, a 12-month frequency starting in April 2012 is specified.

`stl_data = stl(tsdata, ''periodic'')` # The time series data is decomposed using the STL procedure. The "periodic" criteria specifies the window for the function to calculate the seasonality, in this case across the whole series since the variation is assumed constant.

The output of the STL decomposition can then be used for further analysis by extracting each component:

`seasonal <- stl_data$time.series[,1]` # Extracts the seasonal component.

`trend <- stl_data$time.series[,2]` # Extracts the trend component.

`random <- stl_data$time.series[,3]` # Extracts the remainder component.

`write.csv(stl_data - seasonal,"De-seasonalCSV.csv")` # Creates a CSV file of the de-seasonalised data.

3.3 Classification tree model

The classification tree model is a supervised segmentation aiming to predict the classification of a target variable. The advantage of a tree structure is the elegant procedure and the clear, effective output making it easy to interpret. The theory and R implementation is explained here, with particular focus on entropy and information gain used to decide each split.

As outlined in Provost and Fawcett (2013), the model splits data observations by taking into account attributes (also known as variables) that aim to form 'pure' subgroups of the same target variable classification. For example, the subgroups are best split to contain the same classification of the provider operational target performance, *met target* or *fail target*. This process involves using predictor variables to form a series of subsets, where a unique path exists for every data point, ending with a final classification of predominantly one target variable classification.

For each predictor variable there is a binary choice according to some rule. A categorical variable, such as month, will have a binary choice *yes/no* to answer an equality rule, 'Month is April?'; a continuous or count variable uses a numerical inequality with the same binary *yes/no* choice to answer a rule such as 'The number of total incomplete pathways is greater than 100?'. The choice of each predictor variable split is chosen according to which variable is most informative of the target variable, and calculated using entropy and information gain.

3.3.1 Entropy and information gain

Entropy measures how impure a set of data with various properties, with respect to the target variable (Provost and Fawcett, 2013). The formula for entropy is

$$entropy = \sum_j -p_j \log_2(p_j), \tag{3.3}$$

where each p_j is the probability of property j within the set being evaluated. A set of data with 0 entropy has minimal impurity (and therefore is pure), meaning all the data are of the same target variable class. In contrast, an entropy of 1 corresponds to a maximum impurity where the target classes are perfectly split.

Entropy is the core measure used to evaluate the informative attributes with respect to the target variable. Starting with an impure dataset, informative attributes are introduced to split the data and make purer subsets of data. The information gain is the measure of how much an attribute changes the entropy as a result of the split. A decrease in entropy summed across the created subsets is a positive information gain. That is, an attribute has added information to the parent state (pre-split data state). This is weighted by the proportion of instances for each

subset, and defined by,

$$IG(parent, subset) = entropy(parent) - \sum_i p(subset_i) \times entropy(subset_i). \quad (3.4)$$

For example, a set of 10 people is classified with respect to the target variable, enjoy watching football with a binary response *yes/no*: 6/10 *yes* and 4/10 *no*. The purity measure, or entropy, for the entire data set is:

$$\begin{aligned} entropy(parent) &= -\frac{6}{10} \times \log\left(\frac{6}{10}\right) \quad - \quad \frac{4}{10} \times \log\left(\frac{4}{10}\right) \\ &= 0.971 \text{ (3dp)}. \end{aligned}$$

The parent entropy is very impure (≈ 1), since there is close to an even split of people with each property (enjoy watching football or not). The inclusion of an attribute such as gender is considered to better inform the classification of the target variable. The 10 people are described as 4 female and 6 male; 4/4 females responded *yes* (0/4 *no*), and 2/6 males responded *yes* (4/6 *no*). By first calculating the entropy for each subset using equation 3.3,

$$\begin{aligned} entropy(female) &= -\frac{4}{4} \times \log\left(\frac{4}{4}\right) \quad - \quad \frac{0}{4} \times \log\left(\frac{0}{4}\right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} entropy(male) &= -\frac{2}{6} \times \log\left(\frac{2}{6}\right) \quad - \quad \frac{4}{6} \times \log\left(\frac{4}{6}\right) \\ &= 0.918 \text{ (3dp)}, \end{aligned}$$

the information gain can be calculated using equation 3.4, where $p(female) = 4/10$ and $p(male) = 6/10$):

$$\begin{aligned} IG(parent, gender) &= 0.971 - \left(\left(\frac{4}{10} \times 0\right) + \left(\frac{6}{10} \times 0.918\right)\right) \\ &= 0.971 - 0.551 \\ &= 0.420. \end{aligned}$$

So gender substantially reduces entropy and is considered an informative attribute. The classification tree would split for gender, enabling more accurate predictions to classify the 10 people who enjoy watching football or not.

Though only a simple example above, this calculation is simply repeated for each included

attribute to better inform the target variable classification. The attribute with the highest information gain is therefore the most informative and is used to split the model. As a recursive procedure, this process can continue for further splits until all possible pure subsets have been created — although this is a strong case of overfitting. A criterion parameter, which specifies a cut-off threshold for each subset's information gain, is applied to prevent overfitting and optimise classification trees.

3.3.2 Variable selection

The target variable used for the classification tree in data analysis is provider performance, met target, classified as either *met* or *failed*. The choice of predictor variables included for the model to select based on information gain are total incomplete pathways, case mix percentages, deseasonalised total incomplete pathways (as discussed in Section 3.2) and the lagged variable met target 'x' months ago. Notably, year and month are excluded from this model (hence it is appropriate to include the de-seasonalised variable). This is because the classification tree is useful to predict future performance, so the time variables have been omitted as they are irrelevant in this situation. For example, a binary choice of year may be 'Year is 2015?', which will always be *no* for future performance.

A limitation of including the lagged variable, met target 'x' months ago, is the first 18 months, from April 2012 to September 2013, have an *NA* value since the previous performance is no longer applicable. Since the operational target was introduced in April 2012, it would be unlikely for providers to meet the future performance measure, and so it is inappropriate to calculate this. This results in information gain values being calculated using 190 less observations for each x in the met target 'x' months ago ($0 < x < 18$) — so met target *1 month ago* will be calculated from $11,387 - 190 = 11,197$; ... *2 months ago* from $11,387 - (190 \times 2) = 11,007$; and so on.

3.3.3 R syntax: *rpart*

The classification tree calculation and plots are undertaken using `rpart` in R. To optimise the models, some important additional parameters were specified as explained here.

`g <- runif(nrow(data)) shuffled <- data[order(g),]` # The classification tree is optimally created using a randomly shuffled dataset.

`(parms=list(split="information"))` # "Information" specifies the splits to be calculated using information gain (and entropy).

`(control = rpart.control(minsplit = 1, cp=0.0001)` # Criterion parameter 'cp' allows the model to split if information gain of the predictor variable is at least 0.0001 — this will result in a very large, overfitted tree.

`prune(overfitted_model, cp= overfitted_model $cptable [which.min (overfitted_model$cptable[,"xerror"]),"CP"])` # Optimally prunes the overfitted model based on the criterion parameter. This can be specified as a specific value, or to minimise the standard error of the fit as shown here.

The weighted model allows for a loss matrix to be specified. This can be used to minimise types of mis-classification, such as false negatives (Type II error). This can be specified in R by,

`(parms=list(loss=matrix(c(0,2,1,0), byrow=TRUE, nrow=2)))` # Specify loss matrix. False negatives weighted twice the loss of false positives, and will be minimised as a result.

3.4 Logistic regression model

A regression model is used to assess which variables are significantly influential to provider performance measures — met target, and % pathways within 18 weeks. An optimum model can be further used to make predictions for new data. A reminder of generalised linear model theory is explained here, with focus on the logistic response variable and deviance model fit tests.

The response variables are discrete counts of successes and failures, with probability p_i of success. For met target, each provider-month is modelled as a Bernoulli distribution $Y_{ij} \sim Bernoulli(p_i)$ — a single trial with either success, *met target* or failure, *failed target*. The % pathways within 18 weeks is modelled as a binomial distribution where each pathway is described as an individual trial with a count of successes for pathways within 18 weeks and a count of failures otherwise, $Y_i \sim Bin(n_i, p_i)$. In both cases, the response variable Y is modelled with binomial errors, since the model is fitting probabilities ranging between 0 and 1 ($0 \leq p_i \leq 1$). The model with logit-link function is

$$\theta = logit(P) = X\beta \tag{3.5}$$

where θ is the linear predictor written as the canonical link function, X is the design matrix of explanatory variables and β is the estimated coefficients of each explanatory variable. Notably for categorical explanatory variables, such as region, a dummy matrix is formulated corresponding to its number of levels. The statistical significance of coefficient estimates is evaluated using a p-value hypothesis test that the estimate is significantly different from zero at the 95% level, and therefore has a relationship with the baseline categorical levels of the response variable. Using the inverse logit-link function, the probability of each model's success variable can be calculated as

$$p_i = \exp\left(\frac{\theta}{1+\theta}\right). \tag{3.6}$$

To assess the overall model fit, two deviance tests are undertaken: log-likelihood ratio statistic and goodness-of-fit test. Given an initial Model 1 with deviance, D_1 and the number of

parameters, p_1 ($= p_1'$ estimated explanatory variable parameters + 1 intercept term), the log-likelihood tests the inclusion of extra parameters in a larger Model 2 (deviance, D_2 and parameters, p_2). Then,

$$D_1 - D_2 \sim \chi^2_{p_2 - p_1}. \tag{3.7}$$

And the goodness-of-fit test for a model, say Model 2 (where degrees of freedom, $df_2 = n - p_2$ where n is the number of observations), is

$$D_2 \sim \chi^2_{df_2}. \tag{3.8}$$

3.4.1 Variable selection

As with classification tree analysis, the explanatory variables used in the model have to be chosen carefully. A notable inclusion of this model is Org Code, so that each provider is included in the model with 190 levels. Notably, the variables region, case mix percentages and lagged variables met target 'x' months ago are no longer appropriate to be included. First, since region is describing each provider, it is already explained in the model. The case mix percentages is a compositional variable — since the percentages are not independent of the other case mix types — and instead, the original counts of case mix incomplete pathways are included instead. One consequently excluded variable is total incomplete pathways as this is explained by the counts across all case mix counts (since $\sum_i count\ of\ case\ mix\ incomplete\ pathways_i = total\ incomplete\ pathways$). Finally, the lagged variables of past performance are excluded due to perfect separation. This is due to these variables being a singular and perfect measure of prediction to future performance measures (see later classification tree analysis, Section 4.3).

An alternative technique to dealing with a compositional variable such as case mix percentage, is to use partial least squares. This method reduces the dimensions of the explanatory variables by creating linear combinations using ordinary least squares. However, since the model aim was to be easily interpretable and extended for future use by a range of stakeholders, the original variables are most useful to the model. Future methods to include some of the excluded variables in the logistic regression are to split the original data into subsets, such as four different models corresponding to each region.

3.4.2 R syntax: *glm*

The logistic regression model is performed using `glm` in R. Further to model fitting, additional commands were necessary to perform analysis as detailed here.

```
OrgCode<-relevel(OrgCode, ref="RA2") # The baseline provider (OrgCode) is
selected to best represent the majority of the data (i.e. unlikely to be an extreme/outlier
provider).
```

```
upper <- (summary(model)$coefficients[,1] +
1.96*summary(model)$coefficients[,2]) # upper CI for coefficient estimates.

lower <- (summary(model)$coefficients[,1] -
1.96*summary(model)$coefficients[,2]) # lower CI for coefficient estimates.
```

3.5 Distribution fitting

The weekly waiting times are key to understanding the overall operational waiting time standards. The weekly waiting time counts are discrete counts which can be plotted as a piecewise linear graph. The decreasing cumulative proportion of weekly waiting times is an indication of how the volume of incomplete pathways changes as the weeks waiting approached the operational standard. The goal of distribution fitting analysis is to fit the observed empirical data to a theoretical distribution, with a general model that can be applied to new data (Collett, 2015). A suggested distribution that the observed data values follow is an exponential decreasing cumulative distribution function (equivalent to an exponential decay function).

The random variable, time, is exponentially distributed:

$$T \sim \exp(\lambda) \quad \text{(where } \lambda > 0 \text{ is the rate parameter)},$$

with probability distribution function (pdf) given by:

$$f(t;\lambda) = \begin{cases} \lambda e^{-t\lambda} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{3.9a}$$

and cumulative distribution function (cdf) given by:

$$F(t;\lambda) = \begin{cases} 1 - e^{-t\lambda} & t \geq 0 \\ 0 & t < 0. \end{cases} \tag{3.9b}$$

Finally, since the observed data is describing a decreasing cumulative distribution, equation 3.10 is transformed as:

$$1 - F(t;\lambda) = \begin{cases} e^{-t\lambda} & t \geq 0 \\ 1 & t < 0. \end{cases} \tag{3.10}$$

The rate parameter λ is equal to $1/\mu$, where μ is the mean number of weeks waiting.

The two hypotheses of the fitted distribution are:

H_o**:** The empirical data are consistent with an exponential distribution, $T \sim \exp(\lambda)$.

H_a**:** The empirical data *differ* from an exponential distribution.

A successful model fit offers an alternative measure to be used in further analysis. For example, the mean number of weeks waiting as a response variable would explain the underlying outcome variable necessary to meet the operational target. Further uses include a survival analysis by including complete pathway data, *admitted* and *non-admitted* pathways. The weekly waiting time data for providers can be interpreted as a survival function for each month, where the proportion of subjects surviving over time is equivalent to the proportion of pathways having waited 'x' number of weeks.

3.5.1 Goodness-of-fit test

To model the fit of the empirical observation data to the theoretical distribution, a chi-square goodness of fit test is used as outlined by McDonald (2014). With observed data O_i and the expected count E_i (estimated from the fitted distribution assuming the hypothesis is correct) for each count in category i in the k discrete time bands ($i = 1, 2, \ldots, k$), the hypotheses are:

$$H_o\text{:}\ O_i = E_i.$$

$$H_a\text{:}\ O_i \neq E_i.$$

The chi-square test statistic,

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{3.11}$$

with $k - 1$ degrees of freedom. The critical value of the chi-square test is therefore χ^2_{k-1}, where large (greater than χ^2_{k-1}) test statistics reject the null hypothesis and small values are insufficient to reject the null hypothesis.

3.5.2 R syntax: *chisq.test*

The plots and goodness-of-fit test are undertaken in R, with the function listed in Appendix A.

The Monte Carlo simulation method offers a robust alternative to creating p-values based on many repetitions of the expected data and testing the likelihood of the observed values occurring with random error (Hope, 1968).

Chapter 4

Data analysis

4.1 Summaries

Summary statistics can reveal key insights of the data, or stimulate further analysis. An example of insight derived from summary statistics is an assessment of the cleaned dataset in comparison to the original provider data, since not all providers are included in the cleaned datasets for analysis. Table 4.1 compares the unadjusted provider dataset with the final, cleaned provider dataset across variables at the end of each year.

The insight across variables demonstrates that although the cleaned data contain fewer observations than the original national overview, they are a fair representation across year end snapshots. In general, the final dataset follows the trend across years for all variables, although the outcome variables, % within 18 weeks, and % providers meeting target, tend to predict worse performance than the original dataset — demonstrated by a negative percentage change from the original dataset. This is a consideration for analysis based on the final dataset. As expected, there is a notable difference for the total and average number of incomplete pathways, since the final dataset has been cleaned to remove some provider data and includes missing data estimates. The relatively small difference for performance measures across each provider suggests that this difference is split equally across providers, and has not removed any significant features of the original data.

Figures 4.1 and 4.2 provide insight into the national and provider performance. These are based on the national datasets, including missing data estimates for the total incomplete pathways. There is a strong negative trend of national performance over time, supported by a similar negative trend associated with the number of providers meeting the target. One notable finding, as shown in Figure 4.2, is that the national performance fails the operational target when less than 74% of providers fail their individual target.

As shown in Tables 4.2 and 4.3, the performance of providers varies between different case mix types and between regions in March 2017. Notably *Neurosurgery* and *Trauma & Orthopaedics* case mix variables have the worst performance, with the latter contributing significantly to the overall national percentage, 90%. A possible explanation is that departments

Variable	Year end	Original*	Final dataset	Difference (%)
Total incomplete pathways	Mar-13	2,564,233	2,464,761	-3.88%
	Mar-14	2,788,169	2,829,883	1.50%
	Mar-15	2,866,657	2,994,020	4.44%
	Mar-16	3,340,613	3,460,503	3.59%
	Mar-17	3,547,699	3,667,836	3.39%
Avg. total pathways (mean)**	Mar-13	12,448	12,972	4.22%
	Mar-14	14,522	14,894	2.56%
	Mar-15	15,248	15,758	3.34%
	Mar-16	17,960	18,213	1.41%
	Mar-17	18,672	19,304	3.39%
% within 18 weeks	Mar-13	94.12%	94.15%	0.03%
	Mar-14	93.54%	93.45%	-0.10%
	Mar-15	92.96%	92.69%	-0.27%
	Mar-16	91.22%	90.82%	-0.40%
	Mar-17	90.02%	89.85%	-0.17%
% providers meeting target	Mar-13	94.66%	95.26%	0.60%
	Mar-14	91.15%	89.47%	-1.67%
	Mar-15	82.35%	80.00%	-2.35%
	Mar-16	70.97%	69.47%	-1.49%
	Mar-17	64.74%	63.16%	-1.58%

Table 4.1: *Comparison of summary statistics between original data and cleaned provider-level datasets. Since the incomplete pathways are snapshots at the end of each month, the year end is more appropriate than a summation across each year.* ***Difference (%) = final - original*** *(as % of* ***original****).*
**Original provider dataset does not include missing data estimates.*
***Avg. total pathways rounded to nearest whole number.*

lack resources, such as fewer doctors or beds. The contrasting waiting list size of the two worst performing case mixes (*Neurosurgery* and *Trauma & Orthopaedics*) suggest two different underlying causes. Neurosurgery is a more significant surgery type, and taking into account that this is elective (non-urgent) suggests that a patient is more likely to take their time before deciding to have surgery. In contrast, Trauma & Orthopaedics is usually a simple procedure but much more common and so there is a large demand on the service. Further, whilst *Geriatric Medicine* has 97% pathways within 18 weeks, the total number of pathways is relatively low and therefore the high performance has a low relative impact on national performance. The performance of pathways within 18 weeks across regions is similar, but the percentage of providers meeting the target in *South*, and less-so in *Midlands & East*, is significantly lower. The overall percentage of providers meeting the target is less than 65%, with the percentage within 18 weeks at 90%. This suggests that there might be a significant set of providers driving the national performance.

This is a starting point to better understand the significance and extent of an individual provider failing a target.

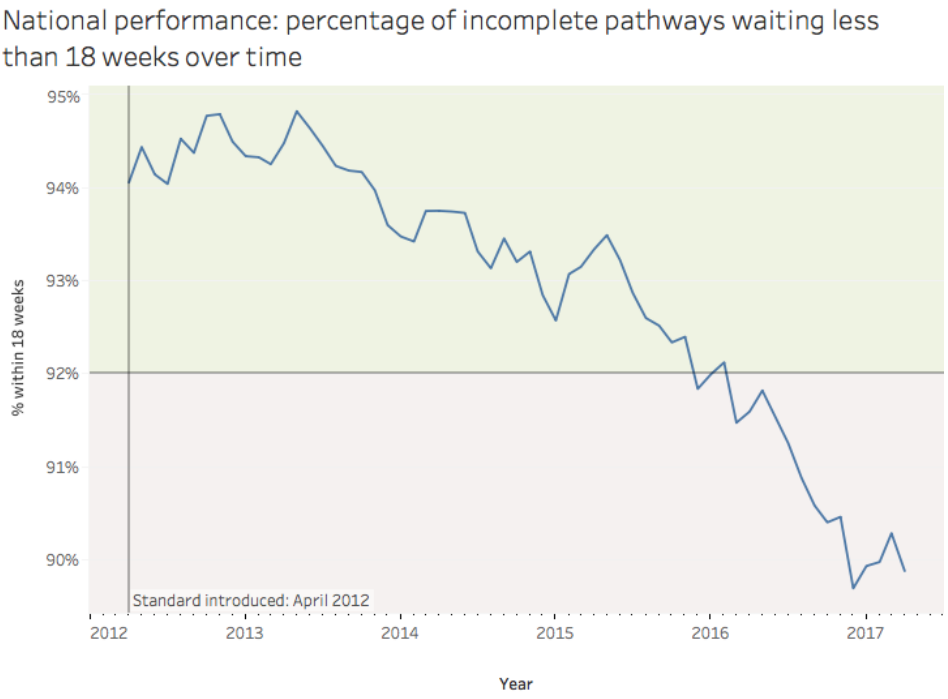

Figure 4.1: *National performance of the % of incomplete pathways waiting less than 18 weeks between April 2012 and March 2017. Shaded regions represent performance meeting operational target (green) and failing (red).*

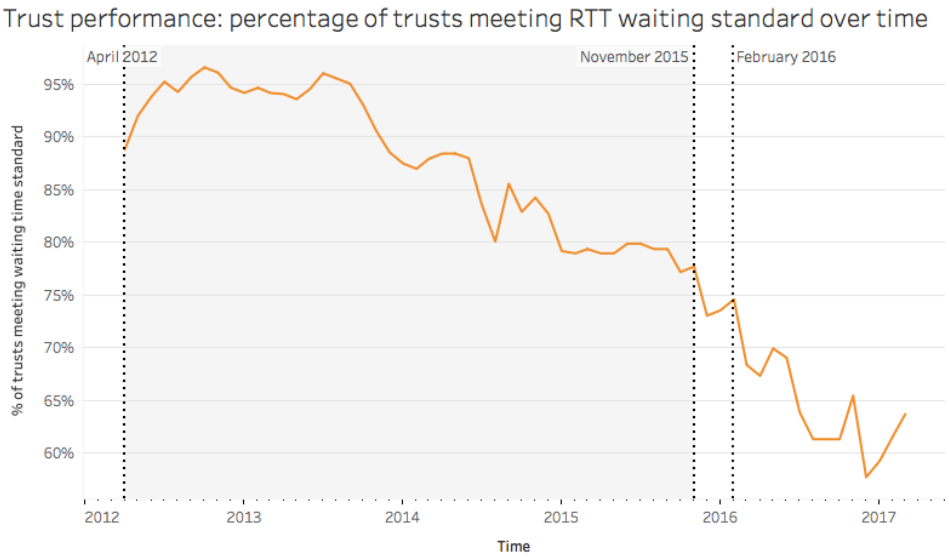

Figure 4.2: *Percentage of providers meeting operational target over time between April 2012 and March 2017. Reference lines correspond to national performance meeting operational target (see Figure 4.1).*

Treatment function	Total incomplete pathways	% within 18 weeks	Active providers for treatment type	% providers met target
Cardiothoracic Surgery	8,016	89.00%	61	75.41%
Geriatric Medicine	20,738	**97.16%**	134	95.52%
Neurosurgery	29,430	83.32%	38	47.37%
Plastic Surgery	48,812	86.43%	89	55.06%
General Medicine	51,306	**94.75%**	128	85.16%
Rheumatology	66,691	**95.43%**	141	87.23%
Thoracic Medicine	88,512	**93.08%**	133	80.45%
Neurology	107,393	89.69%	108	65.74%
Oral Surgery	152,303	89.01%	123	47.15%
Gastroenterology	171,892	91.85%	134	69.40%
Dermatology	173,276	**92.17%**	124	74.19%
Urology	184,054	88.46%	142	46.48%
Cardiology	189,627	**92.13%**	143	70.63%
Gynaecology	212,327	90.80%	143	67.83%
ENT	256,937	89.47%	138	56.52%
General Surgery	286,395	86.76%	140	36.43%
Ophthalmology	370,144	91.23%	132	68.18%
Trauma & Orthopaedics	389,338	84.01%	151	25.83%
Other	740,508	**92.09%**	184	77.17%
All treatment	3,547,699	90.02%	190	64.74%

Table 4.2: *National and provider performance for each treatment type in March 2017. The data is taken from the original provider dataset, and does not include missing data estimates.*

Region	Total incomplete pathways	% within 18 weeks	Number of providers	% of providers met target
North	1,046,306	90.87%	59	79.66%
Midlands & East	1,016,356	89.81%	61	62.30%
London	635,191	89.05%	27	70.37%
South	849,846	89.94%	43	44.19%
National	3,547,699	90.02%	190	64.74%

Table 4.3: *National and provider performance for each region in March 2017. The data is taken from the original provider dataset, and does not include missing data estimates.*

4.2 Time series analysis

Time series decomposition is useful to explain the underlying patterns of the overall time series, into trend, seasonal and random components. Section 3.2 introduced the method and R package `stl` used to decompose the data. The decomposition analysis has been applied to the total incomplete pathways for three perspectives: nationally, by treatment (or case mix) type, and for each provider.

The main plot outputted for the model (Figure 4.3) shows the original dataset, the trend component, seasonal component and remainder component. The bars on the right hand side of each graph give a visualisation of the scales of variance explained by each component, equally interpreted by the magnitude of the units on the y-axis. As a general interpretation, the bars demonstrate the magnitude of each component's variation contained in the original data, if they were all plotted on the same axis. For example, Figure 4.3 shows the national incomplete pathways time series decomposition. The bar on the original data plot can be considered as a single unit of variation, where national incomplete pathways range from 2.5 million to nearly 4 million. The trend component has a similar scale, as demonstrated by the bar and y-axis scale, suggesting that a large amount of variance is explained by the trend (relative to the overall data variation). The large bars and small scales for both seasonal and random components suggest that the variation attributed to these components is much smaller than the trend. Although the seasonal component is of little importance relative to the trend, the seasonal variance is more influential than the random variance as suggested by the smaller bar on the seasonal plot. The models that account for most of the variation as trend and/or seasonality suggest a good decomposition model fit.

The decomposition of national data is shown in Figure 4.3. The trend explains most of the variation in the data, as a steady increase in the size of the number of incomplete pathways. The trend seems to level off around mid-2014 before returning to the previous trend by the start of 2016. This suggests that in general the national waiting list is increasing each year. The seasonal component only explains a small amount of the data variation, remaining constant over time (as expected in our additive model). The size of the waiting list peaks in summer months and dramatically decreases toward year ends. This is further shown in Figure 4.4, assessing the seasonal pattern of the de-trended data. The variation of the five observations for each month (as plotted by the black line) is explained by the random component, suggesting minimal variation is not explained by the seasonal component. Notably, the December seasonal effect explains nearly all data variation in the de-trended series. Finally, a test of the random component is shown in Figure 4.5. This is comparable to the decomposed plot, where the variation is random over time with very few outliers. However, there is still some seasonality present in the remainder that is not captured by the model. This might be explained by a varying seasonal effect from year to year. The histogram confirms that the random noise can be assumed as normally distributed — important as a modelling principle of stationarity — but also to highlight possible outliers. The

two peaks either side of ± 50,000 correspond to the high random components in Feb/March 2015 and April 2016 (identifiable in Figure 4.3), that may be further explained by external factors. One possible explanation for the Feb/March 2015 is the high staff sickness levels in the NHS in the previous months Nov 2014 - January 2015 (Health and Social Care Information Centre, 2015).

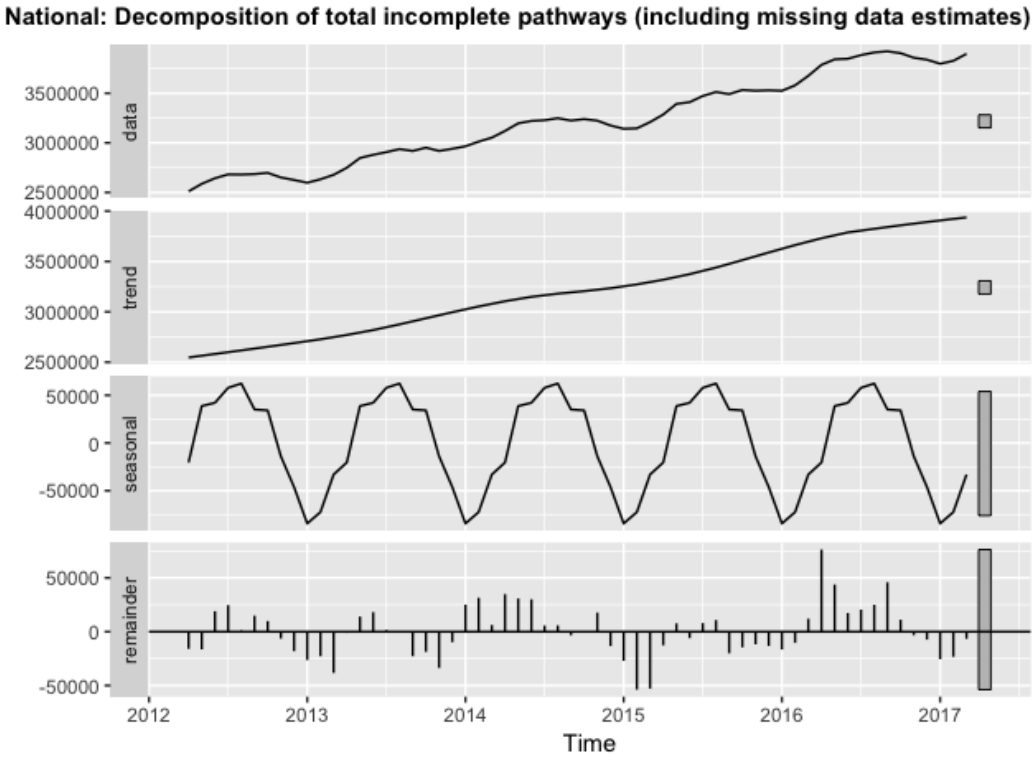

Figure 4.3: *Decomposition of national total incomplete pathways.*

The decomposed treatment type of pathway waiting lists is similar to the national decomposition. Figure 4.6 is an example decomposition of one of the 19 treatment/case mix time series data: *Cardiology*. Although the size of the case mix waiting list is only 4% of the national waiting list, the trend explains most of the variance in the data and is very similar in shape. The next most important is variation explained by the random component as shown by the relative bar sizes. Hence, the seasonal component now explains the least amount of variation in the data and suggests a weak seasonality is present for the volume of *Cardiology* waiting lists. It can thus be interpreted that Cardiology is a treatment type that is equally prevalent throughout the year, and is less susceptible to seasonal effects such as cold weather. A further analysis of each treatment type may further explain the national seasonal effect influenced by certain types of treatment.

The final decomposition of waiting list size was applied to each provider, as shown for *R1A* in Figure 4.7. This is a comparatively much smaller magnitude of waiting list, and varies much more for each provider. The size of the waiting list has less variance explained by trend, as providers who merge or expand to take on more pathways will have a greater effect on the

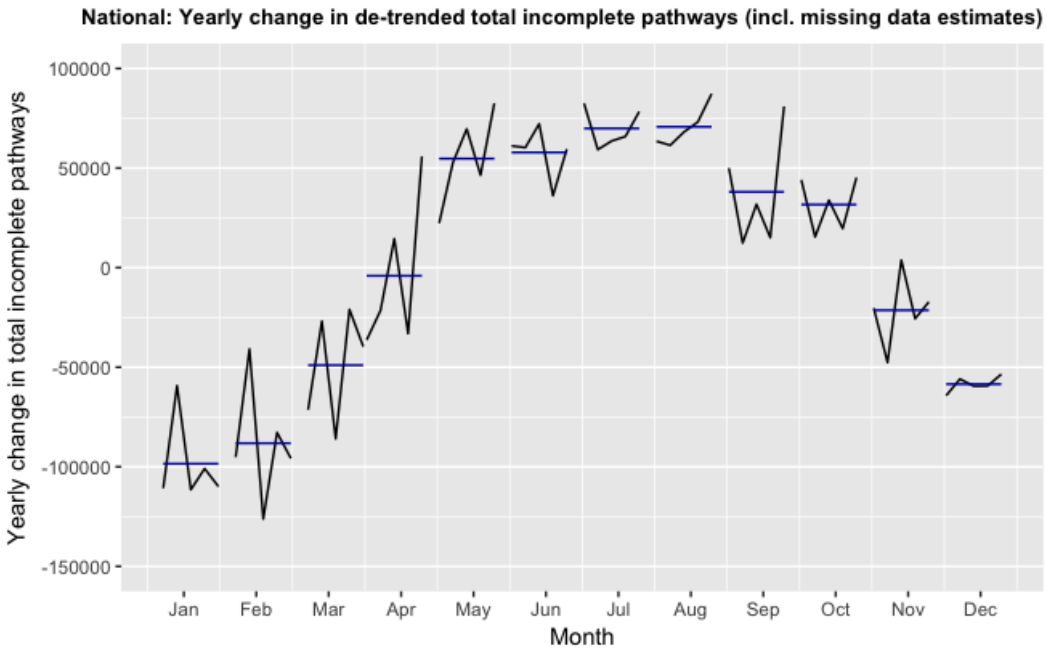

Figure 4.4: *Month plot showing the variance in seasonal component of national total incomplete pathways. Blue lines are the mean seasonal effects.*

overall trend. However, the trend component for *R1A* is still the most influential, although very similarly explained by the random component. The seasonal effect is very weak for this example, and suggests seasonal decomposition is not necessarily appropriate for this provider since a lot of variance is still accounted as random noise. Although this provider demonstrates a weak set of trend and seasonal components, the performance differs for each provider. A comparative measure across all 190 providers would demonstrate the effect each provider has on the aggregated treatment type and national waiting lists.

Overall, the decomposition of time series data into seasonal, trend and random components is useful to explain the underlying features of the data. In particular, it is evident that a trend is the most important component with some seasonal components shown too. These components can be further used to create variables such as a de-seasonalised waiting list and de-trended waiting list. The exclusion of each component can be used for models where the relative component variable is unavailable or inappropriate to use. For example, the de-seasonalised waiting list is useful where the month is excluded, and de-trended when the year is excluded. The de-seasonalised waiting list has been calculated for each provider, for use in later analysis where the month variable is excluded. This can be interpreted as the provider's capacity level to deal with waiting lists, where the observations with unadjusted waiting list size exceeding the 'capacity level' results in a stretched use of finite resources — and potential negative effect on provider performance.

The extracted de-seasonalised total incomplete pathways, as outlined in Section 3.2, does

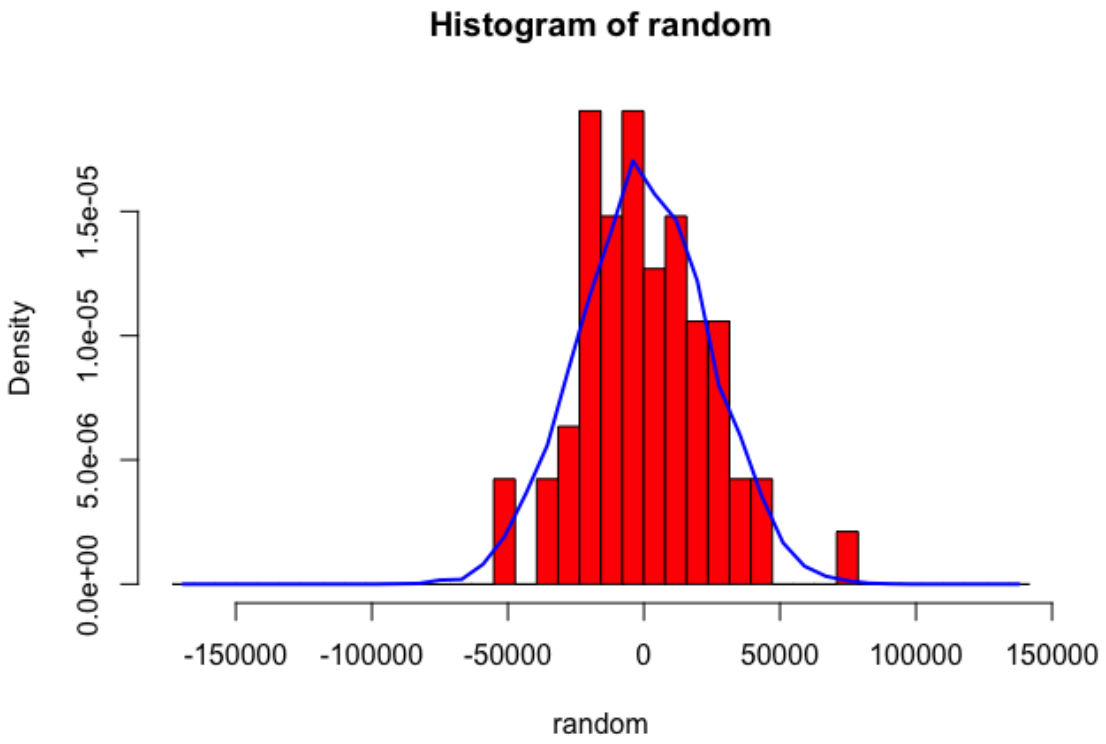

Figure 4.5: *Histogram to test the normality of the random component of national total incomplete pathways decomposition.*

not take into account that the number of pathways is a discrete count variable. Consequently, providers with a small number of pathways in some months will be de-seasonalised, resulting in a negative number. Since this is not logically possible for count data, the de-seasonalised waiting list size is validated, with all negative counts being adjusted to 0. There were seven observations that were adjusted for provider *RYX* and are explained due to a very low number of pathways in one year, with the model assuming constant seasonality levels throughout. A more flexible model allowing for changing seasonality should be explored as a more effective method for decomposing provider-level total incomplete pathways.

A limitation of the current time series model is that the monthly seasonal effect does not take into account the number of days in each month. For example, the seasonal effect of February only applies to 28 days worth of data, compared with March having 31 days. Furthermore, the number of working days in each month and year will vary by excluding weekends and bank holidays. An approach to address this would be to look at the total incomplete pathways per working day. Future models could be further applied to other underlying predictor variables of provider performance.

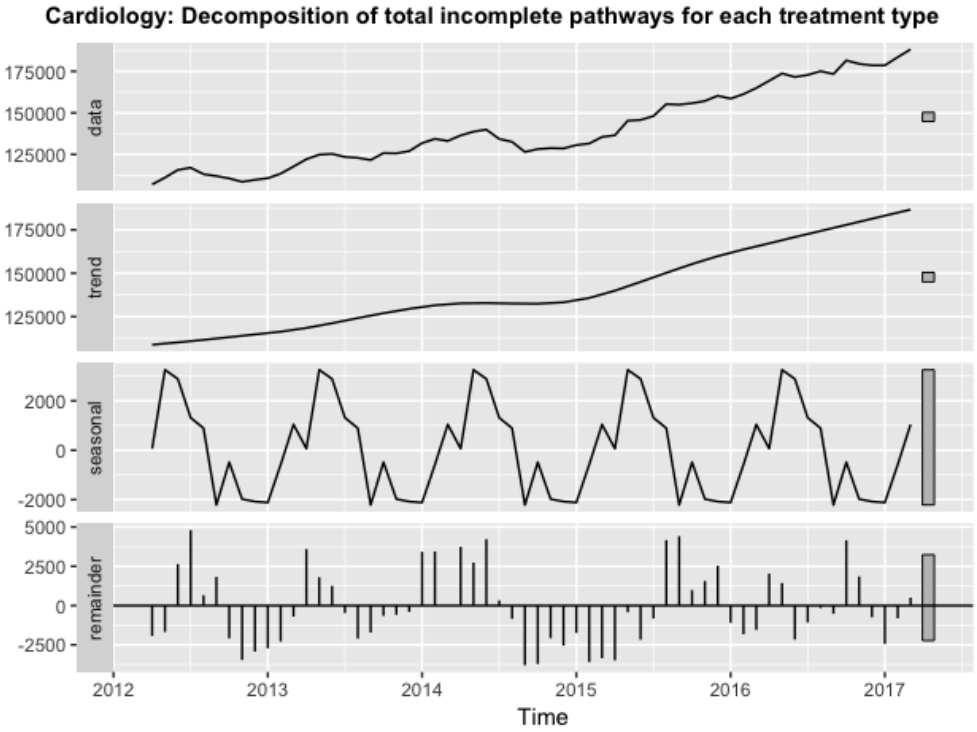

Figure 4.6: *Example decomposition of total incomplete pathways for treatment type: Cardiology.*

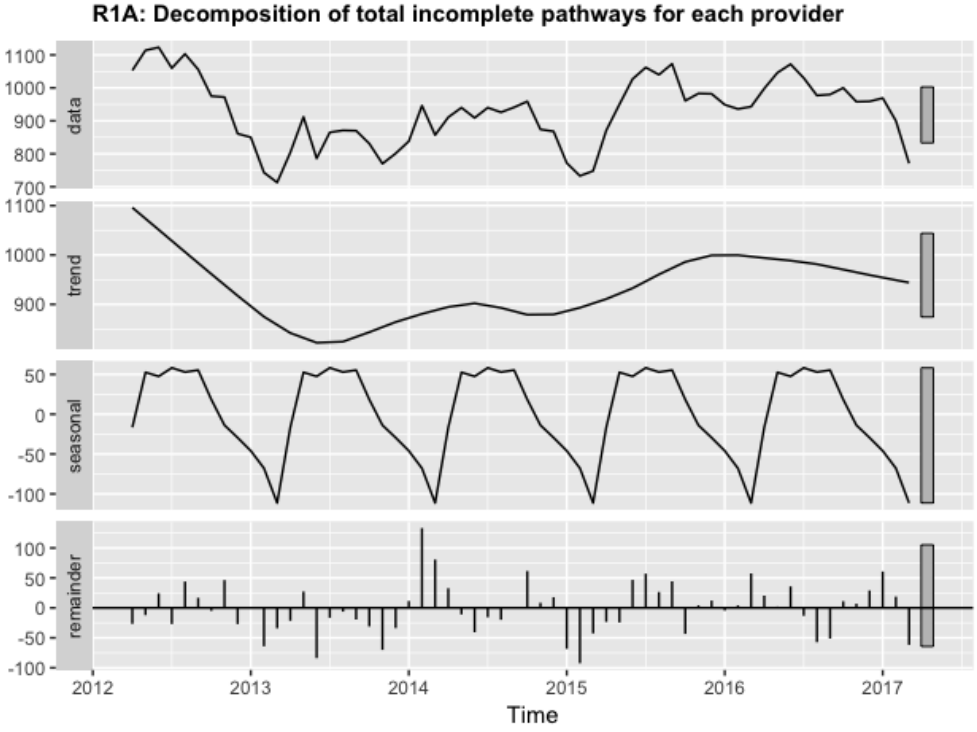

Figure 4.7: *Example decomposition of total incomplete pathways for provider: R1A.*

4.3 Classification tree analysis

The classification tree analysis has been performed in R using package `rpart` (as outlined in Section 3.3). A number of models were fitted to assess which best fit the data and give the best insight into provider-level performance. The main output from a classification tree model is the final plot containing four key components:

- The *Pass*/*Fail* classification at each node.
- The probability of a provider meeting/passing the operational target. This is also indicated by the shaded background of each node: *solid blue = very likely to fail target*, ranging to *solid green = very likely to meet target*.
- The percentage of data being explained at each node split.
- The variable used for splitting, described by left for *yes* and right for *no*.

A description of the models and their performance is given in Table 4.4. The exclusion of lagged variables is listed stepwise, such that the first model includes all variables, the second model includes all variables but excludes *last month*, the third model excludes *last month* and also *2 months previous* (3 months in total), and so on. The measure of a model's performance has been computed using training-test set principles, with the final model computed using the full dataset. The most useful measures of performance for the classification tree are accuracy and specificity, since the model should maximise its predictive performance for providers passing or failing, but also to help recognise a provider that might fail. The specificity can be further optimised in models that weight the importance of minimising false positive predictions (providers that are predicted to pass the waiting time target, but actually fail) greater than false negative predictions (providers predicted to fail, but actually pass). These models are described and compared in Table 4.5.

The first model to be plotted is shown in Figure 4.8, which includes all predictor variables including every lagged performance predictor: *last month, 2 months previous, . . . , 18 months previous*. This model shows that the lagged variable *last month* describes enough information from all possible predictor variables to predict the performance of future providers. A provider that is failing in the previous month has only a 10% chance of passing in the subsequent month, whilst a provider is 97% likely to continue to meet the operational waiting time target having done so in the previous month.

The second plotted model, chosen due to the inclusion of a lagged variable greater than *5 months previous* with the highest accuracy and specificity score, is shown in Figure 4.9. The most significant predictor of performance is the provider's historical performance six months ago; 91% probability of passing the target if the provider was doing so six months previously. However for those providers failing historically, there is a varying probability of failure depending on:

1. First, the **total number of incomplete pathways**:

   [a:] If the total is greater than or equal to 21,000, then there is only a 17% probability of a provider passing.

2. If the total is less than 21,000 then the proportion of **Oral Surgery** incomplete pathways is taken into consideration:

   [a:] If a provider's total pathways consist of greater than 6.4% *Oral Surgery*, there is a 24% probability of passing.

   [b:] If the provider's *Oral Surgery* pathways is less than 6.4% of the total pathways, then the proportion of **Rheumatology** is considered:

   [b(i):] If the percentage of a provider's total pathways consist of greater than 2.4% *Rheumatology* then the provider only has an 18% probability of passing

   [b(ii):] However, a provider with less than 2.4% *Rheumatology*, indicates a 60% probability of passing the target.

		Predicting to meet target		*Predicting to fail target*	
Stepwise excluded variables	Accuracy	PPV	Sensitivity	NPV	Specificity
- (all variables included)	0.9574	0.9754	0.9730	0.8728	0.8830
Last month	0.9402	0.9644	0.9632	0.8256	0.8304
2 month previous	0.9271	0.9471	0.9657	0.8194	0.7427
3 month previous	0.9088	0.9291	0.9632	0.7872	0.6491
4 month previous	0.8886	0.9163	0.9522	0.7194	0.5848
5 month previous	0.8865	0.9027	0.9669	0.7611	0.5029
6 month previous	0.8815	0.8994	0.9645	0.7411	0.4854
7 month previous	0.8825	0.8968	0.9694	0.7619	0.4678
8 month previous	0.8784	0.8928	0.9694	0.7525	0.4444
9 month previous	0.8744	0.8879	0.9706	0.7474	0.4152
10 month previous	0.8632	0.8804	0.9657	0.6957	0.3743
11 month previous	0.8561	0.8744	0.9645	0.6667	0.3392
12 month previous	0.8571	0.8738	0.9669	0.6786	0.3333
13 month previous	0.8521	0.8633	0.9755	0.6923	0.2632
14 month previous	0.8652	0.8684	0.9865	0.8167	0.2865
15 month previous	0.8713	0.8773	0.9816	0.7973	0.345
16 month previous	0.8713	0.8858	0.9694	0.734	0.4035
17 month previous	0.8683	0.8803	0.973	0.7412	0.3684
18 month previous	0.8734	0.8784	0.9828	0.8108	0.3509

*Table 4.4: Description of classification tree models and performance measures. The stepwise excluded variables are removed sequentially in each model such that the **18 month previous** does not include any of the previous excluded variables (**17 month previous**, ... **Last month**). *Highlighted models have been optimally selected for plotting.*

Stepwise excluded variables	Accuracy	*Predicting to meet target* PPV	Sensitivity	*Predicting to fail target* NPV	Specificity
- (all variables included)	0.9574	0.9754	0.9730	0.8728	0.8830
Last month	0.9402	0.9644	0.9632	0.8256	0.8304
2 month previous	0.9271	0.9471	0.9657	0.8194	0.7427
3 month previous	0.9088	0.9291	0.9632	0.7872	0.6491
4 month previous	0.8886	0.9163	0.9522	0.7194	0.5848
5 month previous	0.8815	0.9059	0.9559	0.7143	0.5263
6 month previous	0.8815	0.9059	0.9559	0.7143	0.5263
7 month previous	0.8734	0.8985	0.9547	0.6917	0.4854
8 month previous	0.8683	0.8933	0.9547	0.6783	0.4561
9 month previous	0.8673	0.8969	0.9485	0.6613	0.4795
10 month previous	0.8592	0.8808	0.9596	0.6633	0.3801
11 month previous	0.8582	0.9024	0.9289	0.6054	0.5205
12 month previous	0.8582	0.9033	0.9277	0.604	0.5263
13 month previous	0.8592	0.9044	0.9277	0.6067	0.5322
14 month previous	0.8582	0.9043	0.9265	0.6026	0.5322
15 month previous	0.8632	0.903	0.935	0.6268	0.5205
16 month previous	0.8582	0.8953	0.9375	0.6165	0.4795
17 month previous	0.8531	0.9028	0.9216	0.5844	0.5263
18 month previous	0.8663	0.8968	0.9473	0.656	0.4795

Table 4.5: *Description of weighted classification tree models and performance measures. The stepwise excluded variables are removed sequentially in each model such that the **18 month previous** does not include any of the previous excluded variables (**17 month previous**, ... **Last month**). *Highlighted models have been optimally selected for plotting.*

The order of the predictor variables indicates which contain the most information for a provider passing or failing the incomplete waiting time target.

The third plotted model excludes all lagged variables, so does not directly take into account the provider's previous performance. This results in a much larger tree, where the optimum predictor variables included are the number of total incomplete pathways and eight case mix percentages. Notably, the total number of incomplete pathways is the most dominant predictor and also further split as the second best predictor. This can thus be interpreted that if the size of the provider is small (small total number of pathways) then the provider is very likely (91%) to meet the operational target. Further, if the provider is very large with more than 62,000 pathways then the provider is 97% likely to fail (although only based on $\sim 1\%$ of the data). For providers within the threshold $21,000 <$ size of provider $< 62,000$, then a range of case mix percentage variables are influential according to the remaining 30-31% of the data.

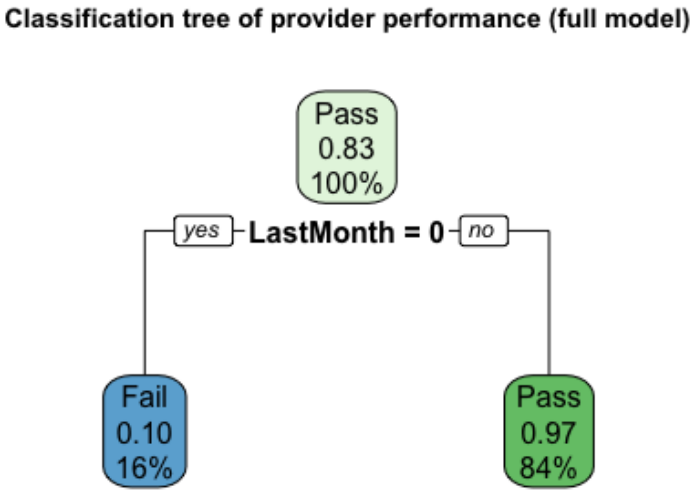

Figure 4.8: *Optimum classification tree of provider performance, where all variables have been included for consideration as predictor variables. ****Last month = 0*** *is equivalent to a provider failing in the last month.*

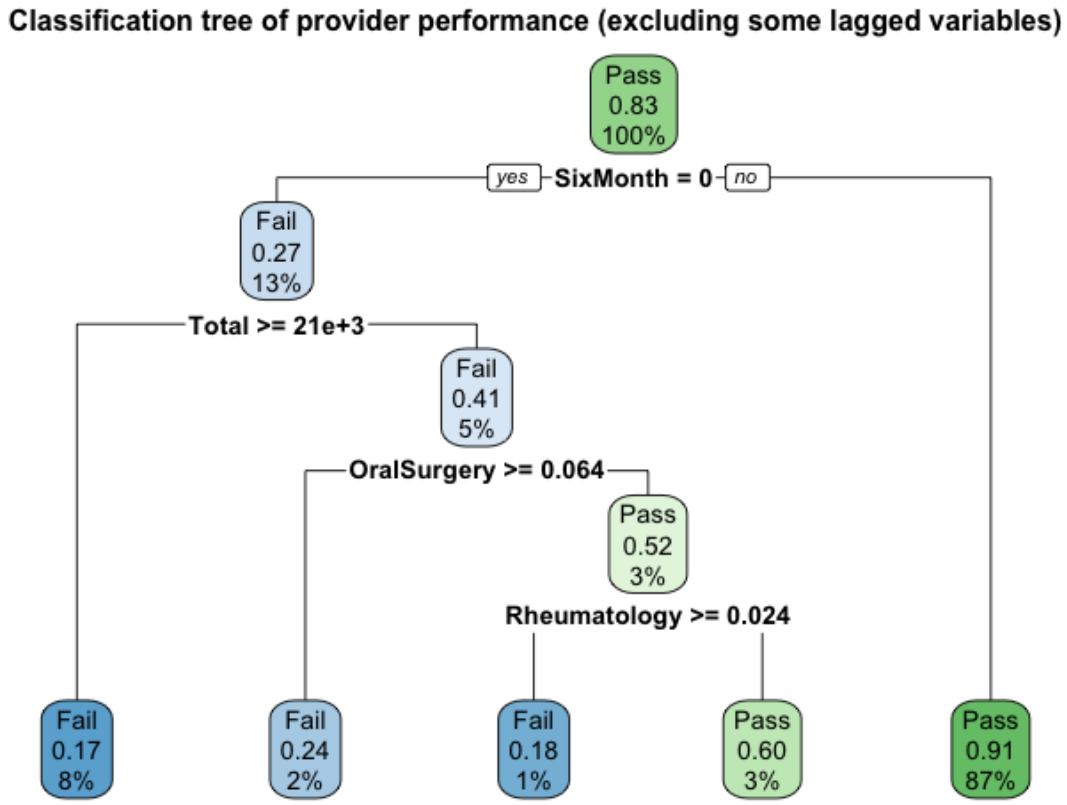

Figure 4.9: *Optimum classification tree of provider performance, where lagged variables with less than 6 months predictive information have been excluded as predictor variables. ****Six-Month = 0*** *is equivalent to a provider failing 6 months ago.*

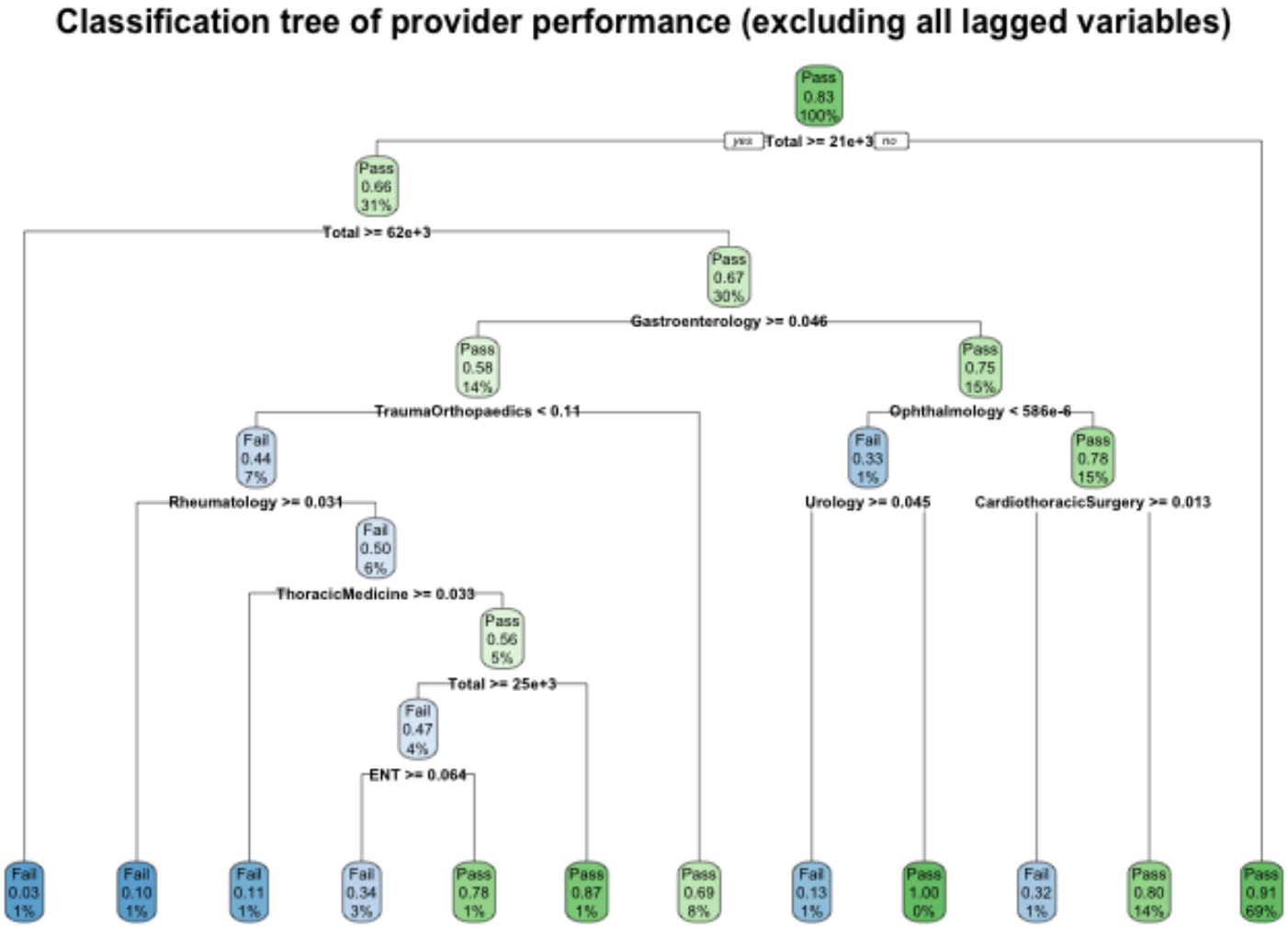

Figure 4.10: *Optimum classification tree of provider performance, where all lagged variables have been excluded as predictor variables.*

The final plotted model is a weighted model to minimise the incorrect classification of false positives — providers that are predicted to pass, but actually fail. This is shown in Figure 4.11. This model is useful if looking to minimise the number of failing providers, with the model significantly maximising the specificity measure in comparison to the unweighted models. However, the model's improved specificity performance is only realised past the stepwise exclusion of *4 months previous*. Assuming a predictive time period greater than four months is desirable, this model is best suited to spot providers that are failing. On the other hand, the model compromises the correct classification of providers meeting the target, which would result in wasted (scarce) resources if allocated solely on this model's findings.

The classification analysis has found that the most informative predictor variables are the lagged variables of past performance. For the non-weighted model, the inclusion of a lagged variable up to and including *5 months previous* is a single predictor for a provider's performance in the future month. A provider that is failing is predicted as likely to continue to fail, and vice versa. The advantage of removing some lagged variables allows for an earlier forecast of performance and potential early indicators for providers. However, one notable limitation of the analysis is the use of current month counts and percentages of incomplete pathways needed to predict a provider's next month classification. To better predict future performance, a suggestion for future analysis is to include a full range of lagged variables (lagged incomplete pathways, case mixes, etc.) relevant to the length of forecasting time period.

The total incomplete pathways can be interpreted as the size of a provider, and is either the second or most dominant predictor in many other models. This follows the basic linear relationship demonstrated nationally in Figure 3.2, the greater the size of the waiting list the less likely it is that providers meet the operational target. The model did not include de-seasonalised total incomplete pathways (an alternative measure to month) as an informative predictor. A suggested next step is to consider de-trended or random component (de-trended and de-seasonalised) of total incomplete pathways to explain the effect of yearly trends. Finally, 12 of the 19 case mix percentage variables are dominant predictors at some point in the models. The seven case mixes not included are *General Medicine, Neurosurgery, Plastic Surgery, Cardiology, Neurology, Geriatric Medicine* and *Gynaecology*.

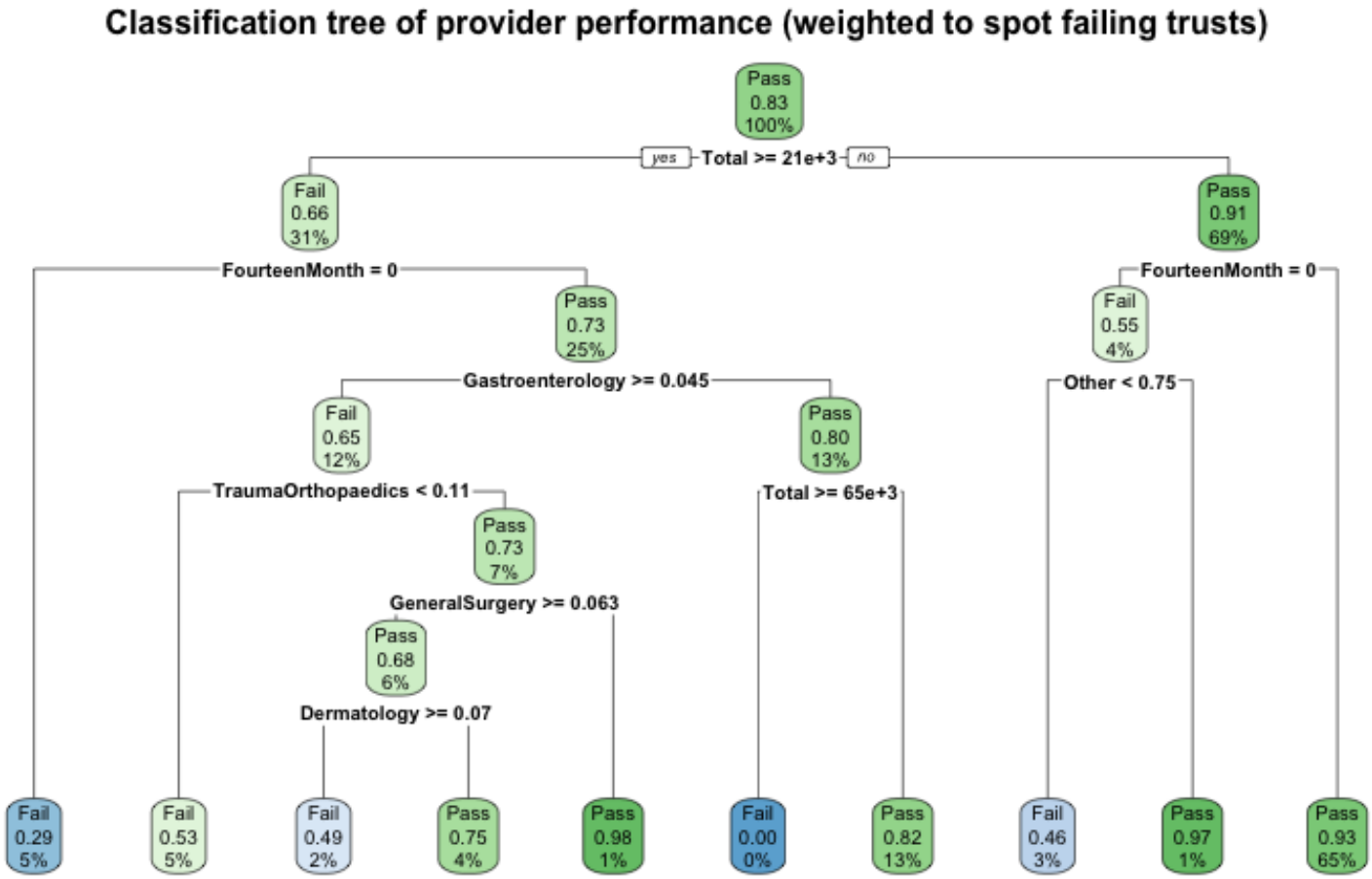

Figure 4.11: *Optimum classification tree of provider performance weighted to correctly predict failing providers, where some lagged variables have been excluded as predictor variables. ****FourteenMonth = 0*** *is equivalent to a provider failing 14 months ago.*

4.4 Logistic regression analysis

A logistic regression model was fitted to the provider-monthly level data to determine which factors are significantly affecting a provider's performance. First, a model assessing the provider's probability to meet the overall operational waiting target that 92% of pathways meet the 18 week deadline; secondly, the underlying probability that an incomplete pathway has been waiting less than 18 weeks. The inclusion of variables and choice of predictor variables was explained in Section 3.4. Recall that the major change from previous analysis is the case mix percentage variable which is transformed to case mix counts (i.e. case mix percentage multiplied by the sum of total incomplete pathways in the provider-month), to avoid the use of a compositional variable.

For both models, the coefficients of categorical predictor variables correspond to a $logit(p)$-unit difference from the reference group:

- Provider (OrgCode) *RA2* in *April*.

The significance levels of categorical predictor variables imply that there is sufficient evidence that the coefficient is non-zero, such that the corresponding level is significantly different from the base/reference level (as given above).

4.4.1 Met operational target

The first model assesses the operational waiting target performance, as in previous classification tree analysis. The response variable is the binary variable, met target — *met* or *failed*. As an initial descriptive analysis, all possible variables were included in the regression model.

The variables that are significant to the model are year, month, provider code, *General Surgery, Trauma & Orthapaedics, ENT, Gastronenterology, Cardiology, Thoracic Medicine, Neurology, Rheumatology, Geriatric Medicine, Gynaecology* and *Other*. This can be interpreted that these variables have an impact on the success of a provider to meet the operational target, with the resources of each treatment type needing to be adequately allocated to achieve the best performance. For levels of categorical variables that are not significant, such as *August* or *R1A*, the levels are similar to the reference level (month: *April*, or provider: *RA2*). An annotated example of the R output is given in Appendix B.

Of particular note are the 8 case mix variables which are not significant at the 5% level. This suggests their coefficient estimates are not significantly different from 0 and may not influence meeting the operational target. Consequently, a marginally improved final model is calculated using the deviance goodness-of-fit tests (`Residual deviance:4779.6 on 11172 degrees of freedom`), in which fewer degrees of freedom are used to explain a significantly similar amount of deviance. In this model, 6 of the case mixes are excluded from the model, *Urology, Ophthalmology, Oral Surgery, Neurosurgery, General Medicine* and *Dermatology* (the 2 remaining in the model are *Plastic Surgery* and *Cardiothoracic Surgery*). Assuming

that interaction terms would not be significant, the predictive power of this final logistic regression model is assessed against new data from April and May 2017. The results are shown in Table 4.6.

The predictive model, where new provider data was available, accurately predicts 3/3 of the top providers but only 1/5 of the bottom providers in April; and 2/3 of top providers and 3/5 of bottom providers in May. This gives an overall accuracy of 56%. Notably, the poor performance for predicting the bottom five providers is affected by providers with a very small number of incomplete pathways, where the change in performance is much more variable due to one or two pathways meeting the 18 weeks deadline. Further tests across a larger number of providers would better represent the accuracy of the logistic model.

Notably the top and bottom five providers, as ranked by their coefficients, have been used to assess the predictive performance. In general, a higher coefficient results in a higher probability of the relative provider meeting the target, However, these ranked coefficients should be treated with caution as suggestive of provider 'league tables'. First, the coefficient is relevant to the reference levels including month — this means that they are ranked according to performance in April. Also, caution should be applied due to overlapping confidence intervals (see Figure A.2 in Appendix B), implying the difference between providers is likely to be significant random noise and susceptible to change. The ranking can differentiate between good and bad providers (high and low coefficients respectively) in general, but is less conclusive as a precisely ordered list. The ranking is only used for model testing purposes here.

4.4.2 Pathway within 18 weeks standard

The second model assesses the provider's performance of pathways waiting less than 18 weeks. In this case, the response variable is a count of successes and failures, within 18 weeks — *success*, or longer than 18 weeks — *failure*. In contrast to the Bernoulli trials above, the model uses pathways within 18 weeks as a count of successes. This is used to model the probability that a provider's pathways are waiting less than 18 weeks. This is the underlying performance that gives the operational target, where the probability of an incomplete pathway within 18 weeks is 0.92.

All variables in this model are significant such that the coefficient accurately describes the difference from the reference levels. Notably, only three providers (*RQM, RXM* and *RYW*) are considered not significantly different to the baseline provider. This is contrasting to the met target model where providers are more likely not similar to each other. One possible explanation for all the significant variables in this model is a greater predictive power in predicting the underlying success of each pathway, rather than the binary operational waiting target being met. However, as noted by the deviance measures, this model has a much larger deviance in both the null model and fitted model. This suggests that the model would benefit from additional variables and interaction effects to predict performance with greater confidence. (An annotated

example of the R output of this model is given in Appendix B).

The performance of this model as a predictor for future months, April and May 2017, is assessed for top and bottom five providers (shown in Table 4.7), as previously discussed in the met target model analysis. For this model, the estimated coefficient confidence intervals overlap less frequently (see Figure A.3 in Appendix B) suggesting a more accurate ranked performance position for providers. The predictive performance of this model is measured in relation to the operational target. If the probability of a pathway within 18 weeks is greater than 0.92, then the provider would expect to meet the operational target. The model accurately predicts 4/4 top providers and 2/5 bottom providers in both April and May, with an overall 67% accuracy rate. Four of the bottom providers are the same as in the previous model, and are affected by a low number of pathways in the modelling process. Though only based on a small sample of providers, the model tends to predict top provider performance more accurately than for bottom providers.

Overall, the predictive performance of both models is more suited to top providers, although the sample of providers is very small in each test ($\sim 5\%$ of the providers included in the model). An extension of the model's performance for all providers would refine the accuracy. The inclusion of standard error predictions would also allow for the model's initial coefficient estimate accuracy.

A limitation of the logistic regression analysis is the absence of interaction terms. It is a good principle to consider interaction terms between variables; however, this was not possible due to R computational difficulties handling the large number of levels in organisation code. A potential next step would be to look at interaction variables, particularly between providers (organisation codes) and case mix specialities that may be significant.

Furthermore, with an optimum model confirmed by this analysis, a comparative look at the size of each variable's effect on the two outcomes could be explored by future studies. In general, the larger the predictor variable coefficient the greater the impact a variable has on the response variable. However, a notable difference in the units for quantitative and qualitative variables must also be considered. A unit change in year, say from 2012 to 2013, is not equally weighted to a unit change, say from *April* to *May*, in the month levels. A suggested approach is to standardise the parameter coefficients, giving an optimal presentation and interpretable set of coefficients.

	April (2017)			
	Predicted		**Actual**	
Provider	Prob (meeting target)	Met/Fail (*if Prob $>$0.5)	Met/Fail	(% within 18 weeks)
R1H			*Did not submit data*	
RDU	0.9091	Pass	Pass	(92.2%)
RTE			*Did not submit data*	
RAL	0.5617	Pass	Pass	(92.2%)
RWE	0.2756	Fail	Fail	(91.3%)
RAT	0.2833	Fail	Pass	(99.5%)
RNN	0.3011	Fail	Pass	(97.0%)
RT1	0.4591	Fail	Pass	(98.9%)
RLQ	0.0192	Fail	Fail	(75.8%)
RT5	0.4864	Fail	Pass	(100.0%)
	May (2017)			
R1H			*Did not submit data*	
RDU	0.9068	Pass	Pass	(93.2%)
RTE			*Did not submit data*	
RAL	0.5877	Pass	Pass	(92.6%)
RWE	0.2714	Fail	Pass	(92.3%)
RAT	0.3558	Fail	Pass	(100.0%)
RNN	0.3576	Fail	Pass	(98.5%)
RT1	0.5259	Pass	Pass	(100.0%)
RLQ	0.0284	Fail	Fail	(76.5%)
RT5	0.5544	Pass	Pass	(100.0%)

Table 4.6: *Performance of logistic regression predictions of top and bottom five providers to meet operational target.*

	April (2017)			
	Predicted		**Actual**	
Provider	Prob (pathway <18 weeks)	(Met target) * if Prob >0.92	% within 18 weeks	(Met/Fail target)
R1C	0.9948	(Pass)	99.3%	(Pass)
RX4	0.9964	(Pass)	100.0%	(Pass)
RYG	0.9972	(Pass)	99.8%	(Pass)
RWN			*Closed provider/inactive code*	
RY6	0.9995	(Pass)	99.9%	(Pass)
RT1	0.8560	(Fail)	98.9%	(Pass)
RLQ	0.7946	(Fail)	75.8%	(Fail)
RT5	0.8659	(Fail)	100.0%	(Pass)
RL1	0.8425	(Fail)	91.2%	(Fail)
RAT	0.8956	(Fail)	99.5%	(Pass)
	May (2017)			
R1C	0.9951	(Pass)	99.2%	(Pass)
RX4	0.9966	(Pass)	100.0%	(Pass)
RYG	0.9974	(Pass)	100.0%	(Pass)
RWN			*Closed provider/inactive code*	
RY6	0.9995	(Pass)	99.7%	(Pass)
RT1	0.8626	(Fail)	100.0%	(Pass)
RLQ	0.8046	(Fail)	76.5%	(Fail)
RT5	0.8721	(Fail)	100.0%	(Pass)
RL1	0.8495	(Fail)	91.4%	(Fail)
RAT	0.9013	(Fail)	100.0%	(Pass)

Table 4.7: *Performance of logistic regression predictions of top and bottom five providers' incomplete pathways waiting less than 18 weeks.*

4.5 Distribution fitting analysis

The distribution fitting method was applied to monthly provider-level data variable, number of pathways by week since referral. Since a provider must aim for 92% of their incomplete pathways as having been waiting less than 18 weeks, it is expected that in each month the number of pathways waiting *0-1 weeks, 1-2, 2-3 . . .* is far greater than those close to the *17-18 weeks* deadline and beyond. To model this, the weekly data was transformed to a decreasing cumulative distribution of the incomplete pathways in each month. Further, by taking the proportion of total pathways in each weekly time band, an observed probability distribution is calculated. Figures 4.12 and 4.13 show the original weekly waiting time for total incomplete pathways and the decreasing cumulative probability distribution for provider *R1A*.

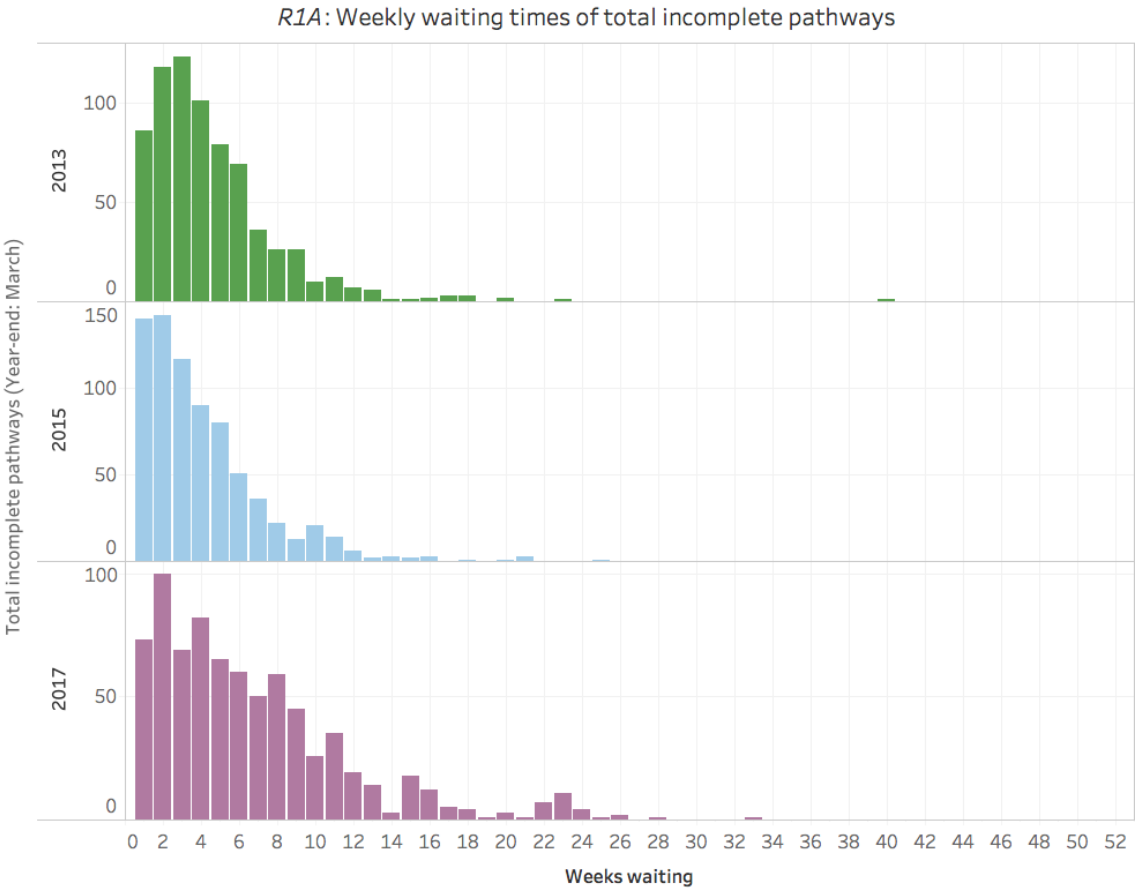

Figure 4.12: *Weekly waiting times of incomplete pathways for provider* R1A. *Year-end (March) snapshots for each year.*

The decreasing cumulative proportion of weekly waiting times is an indication of how the volume of incomplete pathways changes as the weeks waiting approached the operational target. A fitted distribution aims to model if a provider is performing at a rate such that 8% of their incomplete pathways are likely to be waiting less than 18 weeks (cross-hair point on Figure 4.13).

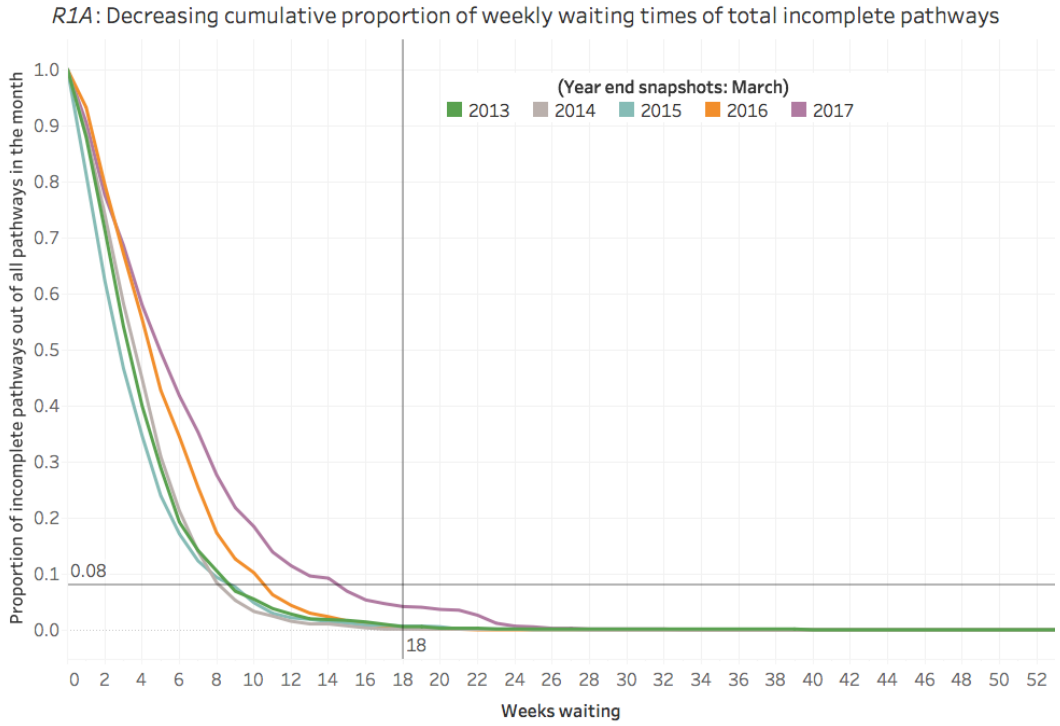

Figure 4.13: *Decreasing cumulative proportion of weekly waiting times of total incomplete pathways for provider* R1A. *All incomplete pathways have waited 0 weeks; week 52 corresponds to pathways waiting 52+ weeks.*

A suggested distribution that the observed weekly waiting time data values follow is an exponential decreasing cumulative distribution function (equivalent to an exponential decay function):

$$1 - F(t;\lambda) = \begin{cases} e^{-t\lambda} & t \geq 0 \\ 1 & t < 0. \end{cases}$$

Recall the set of hypotheses, as introduced in Section 3.5, are:

H_o**:** The empirical data are consistent with an exponential distribution, $T \sim \exp(\lambda)$.

H_a**:** The empirical data *differ* from an exponential distribution.

As outlined in Section 3.5, the rate parameter λ is equal to $1/\mu$, where μ is the mean number of weeks waiting. The distribution is fitted to two providers, *R1A* and *RBA*, for each of the 60 months of weekly waiting times for the total number of incomplete pathways. The chi-square goodness-of-fit test was used to assess the null hypothesis with a 5% significance level. The results of the test are shown in Table 4.8.

For provider *R1A*, the results indicate that most of the fitted exponential models have statistically significant p-values, such that we reject the null hypothesis and accept the alternate

hypothesis — the exponential model is not a good fit for the empirical data. Although the fitted exponential model is, at first look, likely to be a good fit (see Figure 4.14), the variance in absolute numbers suggests the differences between observed and expected are significant. However, for some months in the *R1A* models with high p-values (> 0.05), suggest that there is not enough evidence to reject H_0. For these instances, the observed data are consistent with an exponential distribution (an example is shown in Figure 4.15). Of these, six instances correspond to January and February models in 2013, 2014 and 2015. This supports the previous time series analysis that seasonality is present and with future study the de-seasonalised weekly pathways data could be investigated.

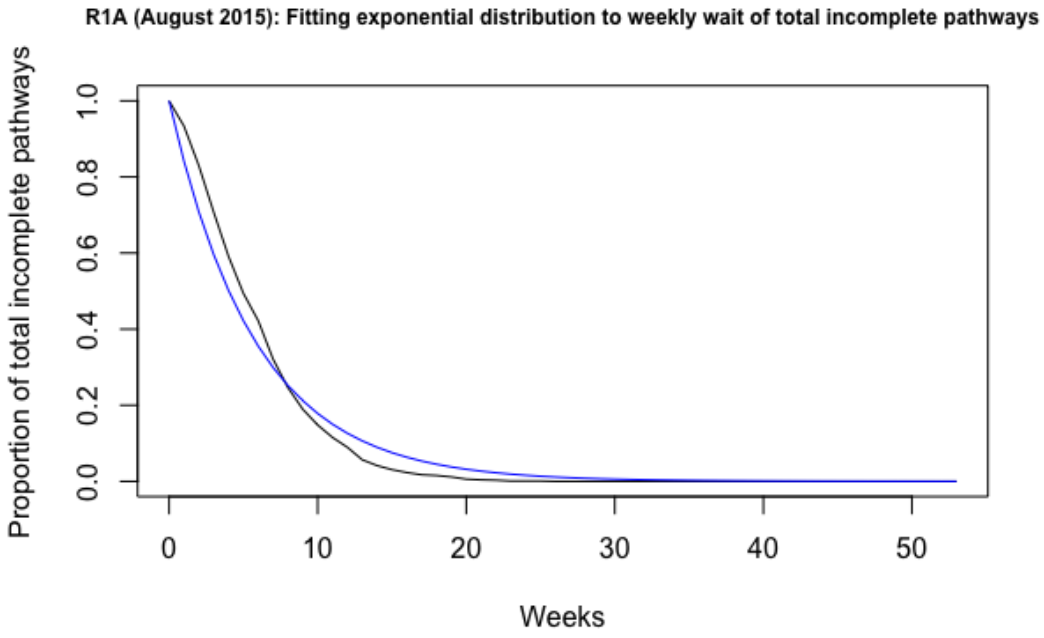

Figure 4.14: *Fitting decreasing exponential cumulative distribution to empirical weekly waiting times for provider R1A (August 2015). Blue line = fitted exponential; Black line = empirical data.*

With 60 monthly models for each provider and 5% significance level, we would expect 3 (5% of 60) to randomly pass the goodness-of-fit test; however, there are 9 monthly models suggesting that although in general the null hypothesis is rejected, there is some evidence to suggest that the exponential distribution is a good fit. Comparing the suggested exponential model for other providers would be the next step to more conclusively assess overall distribution fitting.

Provider *RBA* is a less stable provider than *R1A*, due to an increased variance in terms of total incomplete pathways and operational waiting time target performance. The goodness-of-fit monthly tests all have significant p-values to reject the null hypothesis, and reflect that the exponential distribution is much more significantly concluded as not a suitable fit for the weekly

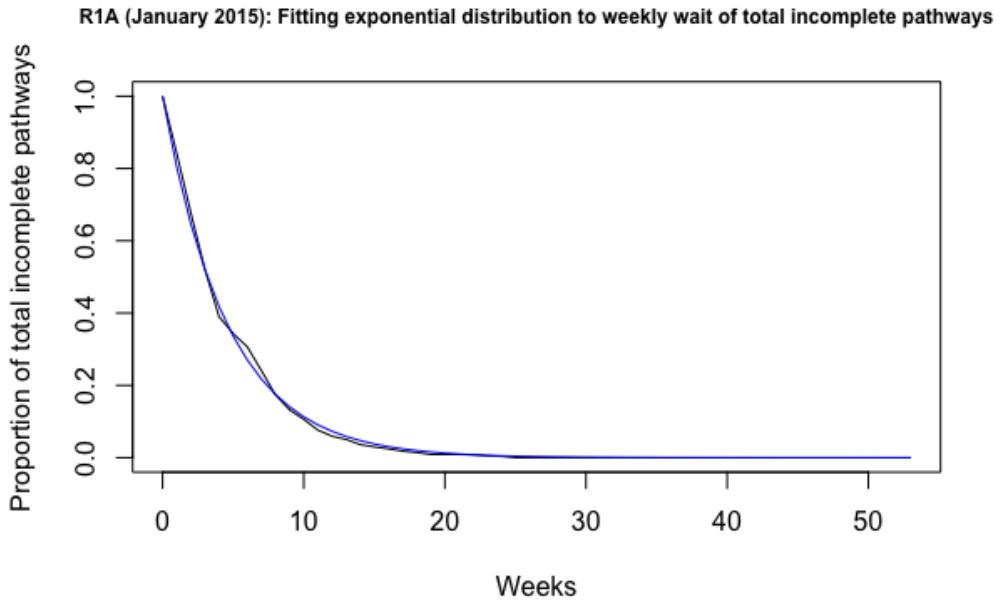

Figure 4.15: *Fitting decreasing exponential cumulative distribution to empirical weekly waiting times for provider R1A (January 2015). Blue line = fitted exponential; Black line = empirical data.*

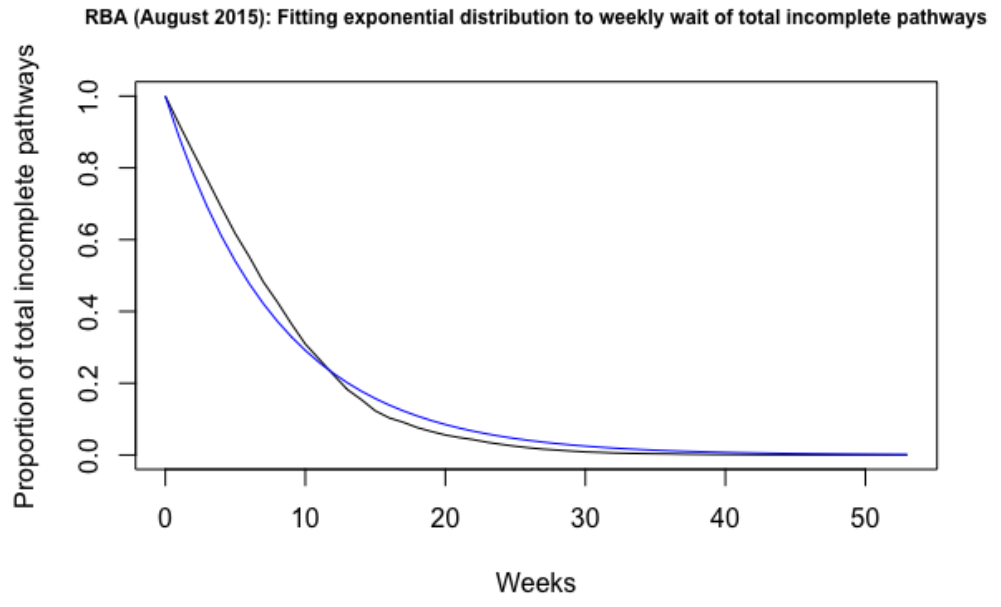

Figure 4.16: *Fitting decreasing exponential cumulative distribution to empirical weekly waiting times for provider RBA (August 2015). Blue line = fitted exponential; Black line = empirical data.*

total incomplete pathways. Finally, this is demonstrated in Figure 4.16, where the exponential distribution fails to explain the empirical data indicating a bad fit for provider *RBA* in August 2015.

The distribution fitting helps to better understand the underlying weekly waiting times for incomplete pathways. For some provider months, the exponential distribution is sufficient to predict the count of weekly waiting incomplete pathways. Since the exponential distribution only depends on the scale parameter, λ ($= 1/\mu$), this can be easily applied for future use. A suggested starting point to consider other models is the Weibull distribution, which considers an extra shape parameter, k.

	R1A		*RBA*	
Date	p-value	Reject H_0	p-value	Reject H_0
Apr 2012	0.024	Yes	0.001	Yes
May 2012	0.015	Yes	0.001	Yes
Jun 2012	0.012	Yes	0.001	Yes
Jul 2012	0.012	Yes	0.001	Yes
Aug 2012	0.0005	Yes	0.001	Yes
Sep 2012	0.014	Yes	0.001	Yes
Oct 2012	0.013	Yes	0.001	Yes
Nov 2012	0.026	Yes	0.001	Yes
Dec 2012	0.0005	Yes	0.001	Yes
Jan 2013	0.073	No	0.001	Yes
Feb 2013	0.133	No	0.001	Yes
Mar 2013	0.022	Yes	0.001	Yes
Apr 2013	0.048	Yes	0.001	Yes
May 2013	0.001	Yes	0.001	Yes
Jun 2013	0.311	No	0.001	Yes
Jul 2013	0.008	Yes	0.001	Yes
Aug 2013	0.001	Yes	0.001	Yes
Sep 2013	0.175	No	0.001	Yes
Oct 2013	0.025	Yes	0.001	Yes
Nov 2013	0.019	Yes	0.001	Yes
Dec 2013	0.0005	Yes	0.001	Yes
Jan 2014	0.083	No	0.001	Yes
Feb 2014	0.059	No	0.001	Yes
Mar 2014	0.009	Yes	0.001	Yes
Apr 2014	0.001	Yes	0.001	Yes
May 2014	0.0005	Yes	0.001	Yes
Jun 2014	0.001	Yes	0.001	Yes
Jul 2014	0.0005	Yes	0.001	Yes
Aug 2014	0.0005	Yes	0.001	Yes
Sep 2014	0.001	Yes	0.001	Yes
Oct 2014	0.001	Yes	0.001	Yes
Nov 2014	0.0005	Yes	0.001	Yes
Dec 2014	0.0005	Yes	0.001	Yes
Jan 2015	0.538	No	0.001	Yes
Feb 2015	0.109	No	0.001	Yes
Mar 2015	0.388	No	0.001	Yes
Apr 2015	0.043	Yes	0.001	Yes
May 2015	0.0005	Yes	0.001	Yes
Jun 2015	0.0005	Yes	0.001	Yes
Jul 2015	0.0005	Yes	0.001	Yes
Aug 2015	0.0005	Yes	0.001	Yes
Sep 2015	0.0005	Yes	0.001	Yes
Oct 2015	0.001	Yes	0.001	Yes
Nov 2015	0.0005	Yes	0.001	Yes
Dec 2015	0.0005	Yes	0.001	Yes
Jan 2016	0.0005	Yes	0.001	Yes
Feb 2016	0.020	Yes	0.001	Yes
Mar 2016	0.0005	Yes	0.001	Yes
Apr 2016	0.0005	Yes	0.001	Yes
May 2016	0.0005	Yes	0.001	Yes
Jun 2016	0.0005	Yes	0.001	Yes
Jul 2016	0.0005	Yes	0.001	Yes
Aug 2016	0.0005	Yes	0.001	Yes
Sep 2016	0.0005	Yes	0.001	Yes
Oct 2016	0.0005	Yes	0.001	Yes
Nov 2016	0.0005	Yes	0.001	Yes
Dec 2016	0.0005	Yes	0.001	Yes
Jan 2017	0.0005	Yes	0.001	Yes
Feb 2017	0.0005	Yes	0.001	Yes
Mar 2017	0.0005	Yes	0.001	Yes

Table 4.8: *Summary of goodness-of-fit tests for exponential distribution fitting to empirical weekly total incomplete pathways data. Highlighted cells: blue = insufficient evidence to reject null hypothesis; green = correspond to months of plotted Figures 4.14, 4.15 and 4.16.*

Chapter 5

Conclusions

5.1 National performance

The national performance of incomplete pathways to be waiting less than 18 weeks has been declining since 2013, failing the operational target for the first time in November 2015 and not recovering after failing for a second time in February 2016. The performance varies with respect to different treatment types. In the most recent March 2017 data, *Neurosurgery* and *Trauma & Orthopaedics* pathways had worst performance, with *Geriatric Medicine* and *Rheumatology* performing best. The national performance between regions is similar, with the *North* slightly outperforming the national average.

One of the major underlying factors of incomplete pathway performance is the total number of incomplete pathways, also known as the waiting list size. This has been steadily increasing in the same period since 2013, and is affected by seasonal changes. The higher waiting list in summer months and levelling-off or decreases in December are of particular note. One possible explanation for extreme waiting list sizes is higher staff sickness levels in the preceding months. The upward trend is evident across waiting list sizes for different case mixes, but with a less significant seasonal pattern.

5.2 Provider-level performance

The provider-level performance of incomplete pathways to be waiting less than 18 weeks has decreased in line with the national performance. A notable relationship between national and provider-level performance is the 74% critical point of providers meeting the operational target necessary for the national performance target to be met. The percentage of providers meeting the operational target across treatment types replicates the national performance in general. However, *Neurosurgery* (as for national performance) and *General Surgery* are the worst performers. For regional provider-level performance, the percentage of providers meeting the operational target varies significantly. *South* is the worst region, despite having the second highest overall

national performance. In contrast, pathways from providers in the *North* have the best performance and patients are more likely to be waiting less than 18 weeks.

The underlying performance of the waiting list (total incomplete pathways) for providers varies between providers. The trend increases nationally, but this is not always the case for each individual provider. This is due to the varying sizes in waiting lists between providers, subject to sudden waiting list changes from merging and closing providers. There is weak evidence of a seasonal effect which mimics the national performance. However, if present, a de-seasonalised provider waiting list can be interpreted and further utilised as a provider's maximum capacity.

Classification tree models demonstrate that the most informative predictor variable of performance is the previous performance in the last month. This is extended to the most informative predictor being lagged performance up to *5 months previously*, until other variables are significant for predictions. The waiting list size is the most important non-lagged variable, with 12 of the 19 case mix percentages also prevalent. For a range of future performance predictions, the classification models can be adapted to include or exclude previous performance or to ensure failing providers are correctly spotted.

The logistic regression model found year, month, provider and 11 of the 19 case mix counts to be significant in predicting the operational target performance for providers. Notably, the case mix variables that are insignificant in either classification tree analysis or logistic regression analysis of meeting the operational target are *General Medicine, Neurosurgery* and *Plastic Surgery*. Providers tend to either be good or bad at meeting the operational target, but a ranking within these categories is less clearly apparent. The refined logistic model is effective for predicting successful provider performance, but fails for providers with small waiting lists. The underlying performance of an incomplete pathway to be waiting less than 18 weeks was found to be significantly influenced by year, month, provider and all case mix counts. The providers are more accurately ranked by how likely a pathway will be waiting within 18 weeks. Model predictions on future months are accurate for top performing providers, but again fail to accurately predict bottom providers with a small waiting list.

Weekly waiting times give a deeper underlying insight into a provider meeting the operational target. The exponential distribution with a mean parameter can be used to model weekly waiting times for some providers in certain months, but is not suitable as a generalisation. The inclusion of other parameters is likely to improve a more generalised fitted distribution. The parameters measured hold key underlying information about performance and can be used as effective target measures for future models.

5.3 Observations

This thesis has explored some of the models and analysis that identify key factors and early indicators for NHS providers' performance in meeting referral to treatment waiting time standards. The discussion has demonstrated some models are better suited to this; logistic regression has found the significant factors influencing performance, and classification trees demonstrate a useful tool as an early indicator for performance. To further support these methods, time series describes the performance over time, with distribution fitting exploring the underlying 18 week standard. These techniques for better understanding NHS providers' performance are a significant starting point to help achieve high standards of RTT elective care pathways.

5.4 Considerations

The models and analysis are subject to some considerations and limitations. The study explores the RTT incomplete waiting time data with just a few possible explanatory variables considered. The models are powerful tools to explain the data of the variables used, but will be further aided by including more variables. Further understanding and significant factors should be sought by communicating and working with healthcare providers and Clinical Commissioning Groups. A suggested set of external variables to consider are the finances and resources of providers (such as the number of staff or capacity size) and the effect of random events such as staff illness or IT system issues.

Several assumptions and limitations have been applied to the models. A statistical assumption is that the time dependent data are considered random between months, which should be further explored and evaluated. In time series analysis, the seasonal effect does not take into account the number of working days in each month which might offer a more accurate decomposition of waiting lists. Classification tree and logistic models are an explanation of the variables included in the model (rather than a definitive measure of waiting time performance), such that the model fit would benefit from the consideration of interaction terms (computationally difficult using current software) as well as further explanatory variables.

Finally, the application of models in the future should be considered. The RTT waiting list target was notably missing in the latest NHS Five Year Forward framework (NHS England, 2017b). This implies a relaxation on providers to meet the operational target, as reported for non-urgent operations such as hernia or knee surgery (Campbell, 2017). Also, the providers included in this model are likely to be affected by the merging or closing of other providers — with one provider closing in April 2017 (one month after the studied time period). This is likely to have a substantial effect on other providers that take on the closed provider's pathways.

5.5 Future work

There is significant potential for future study building on the models and analysis previously discussed. Some suggested starting points building on the current models are the alternative total incomplete pathways per (working) day variable and identifying further parameters such as the shape parameter (Weibull distribution) for distribution fitting. Future study of analysis using working days would be particularly relevant to the seven-day NHS proposals (Department of Health, 2015c) — for example, what effect would seven day elective care have on meeting referral to treatment standards?

The changing state of failing providers is of added importance to ensure providers are held accountable for the best quality of care. A starting point for this is the maximum number of months providers consecutively fail, and how likely a provider is to recover having previously failed the operational target (see Appendix A, Figure A.4 and A.5, respectively).

Though not considered here, the completed pathway data explains the pathways that are no longer waiting for treatment. The inclusion of these datasets could add further insight. One such opportunity is to perform a survival analysis (alongside a successful distribution fitting model) where the proportion of pathways having waited 'x' number of weeks is considered as having survived 'x' weeks.

Finally, the assessed data only considers public sector providers. However, there is a notable difference in performance between public and independent sectors as shown in Figure A.6 (see Appendix A). The contrasting performance, and changing healthcare architecture, might further explain patterns in provider and national performance.

The importance of achieving referral to treatment standards and the worsening performance nationally signifies that such further work would be worthwhile to develop the analyses undertaken in this study.

Appendix A

Additional tables and figures

Time bands	Corresponding number of days since referral	Notes
0-1 weeks	0, 1, 2, 3, 4, 5, 6 and 7 days	*Includes eight rather than seven days*
>1-2 weeks	8, 9, 10, 11, 12, 13 and 14 days	
⋮	⋮	
>17-18 weeks	120, 121, 122, 123, 124, 125 and **126** days	*126 days is equivalent to 18 weeks*
>18-19 weeks	127, 128, 129, 130, 131, 132 and 133 days	*Longer than operational target*
⋮	⋮	
>51-52 weeks	358, 359, 360, 361, 362, 363 and 364 days	
52+ weeks	365 days and more	

Table A.1: *Summary of RTT pathways assigned to weekly time bands.*

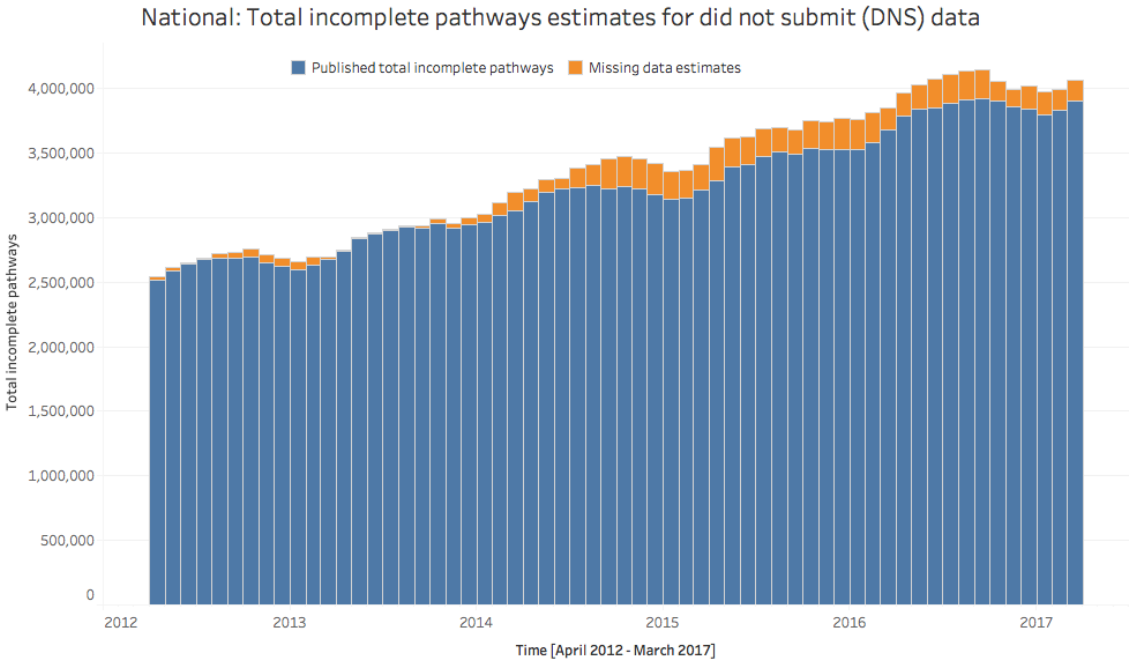

Figure A.1: *Estimated data (orange) for total number of incomplete pathways nationally between April 2012 and March 2017. Estimated from last submitted data for each month a provider did not submit data.*

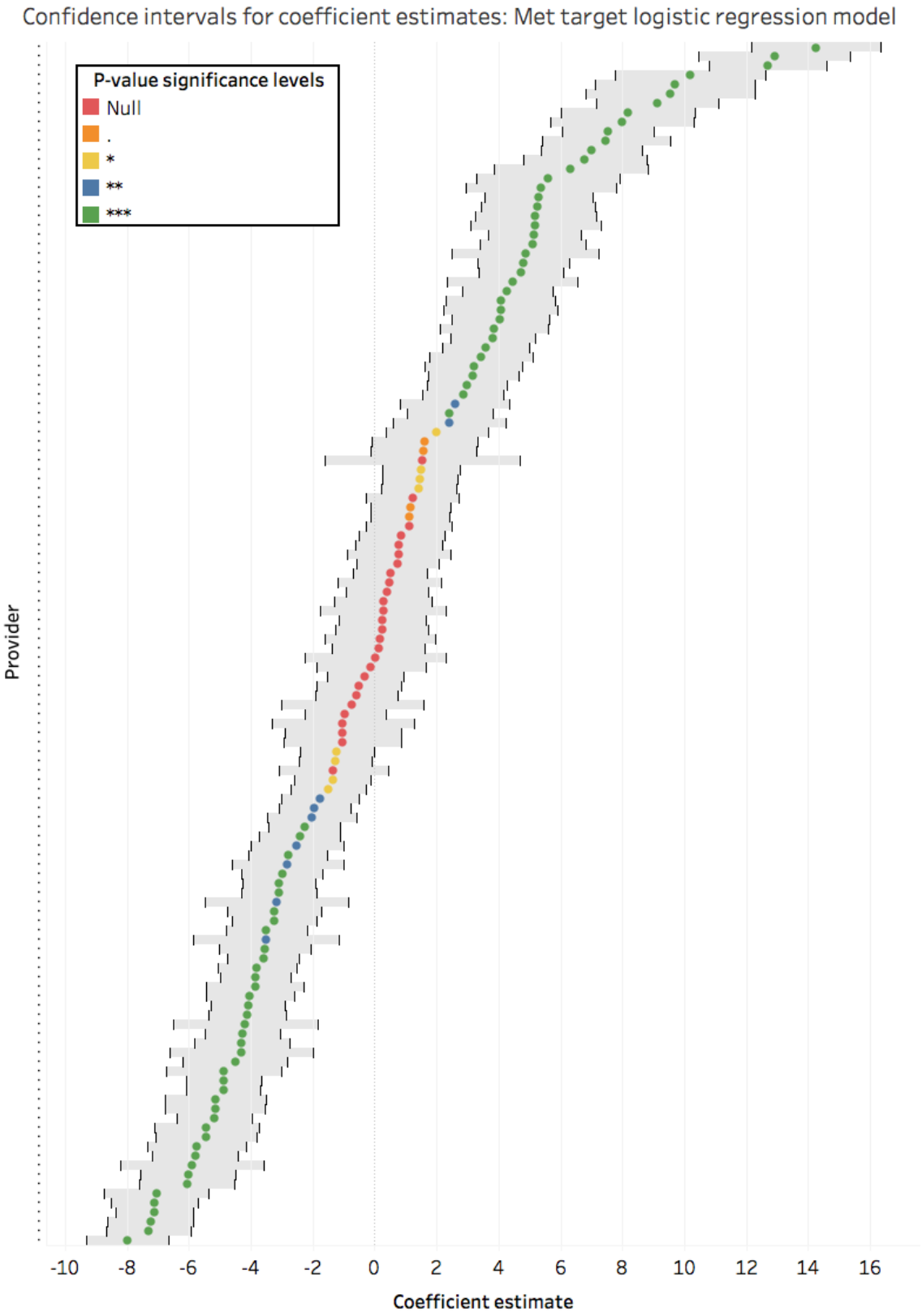

Figure A.2: Confidence interval for coefficient estimates: met target logistic regression model. Notably, 61 providers have been excluded as the confidence intervals were extreme. This corresponds to poorly fitted estimates with p-values very close to 1.

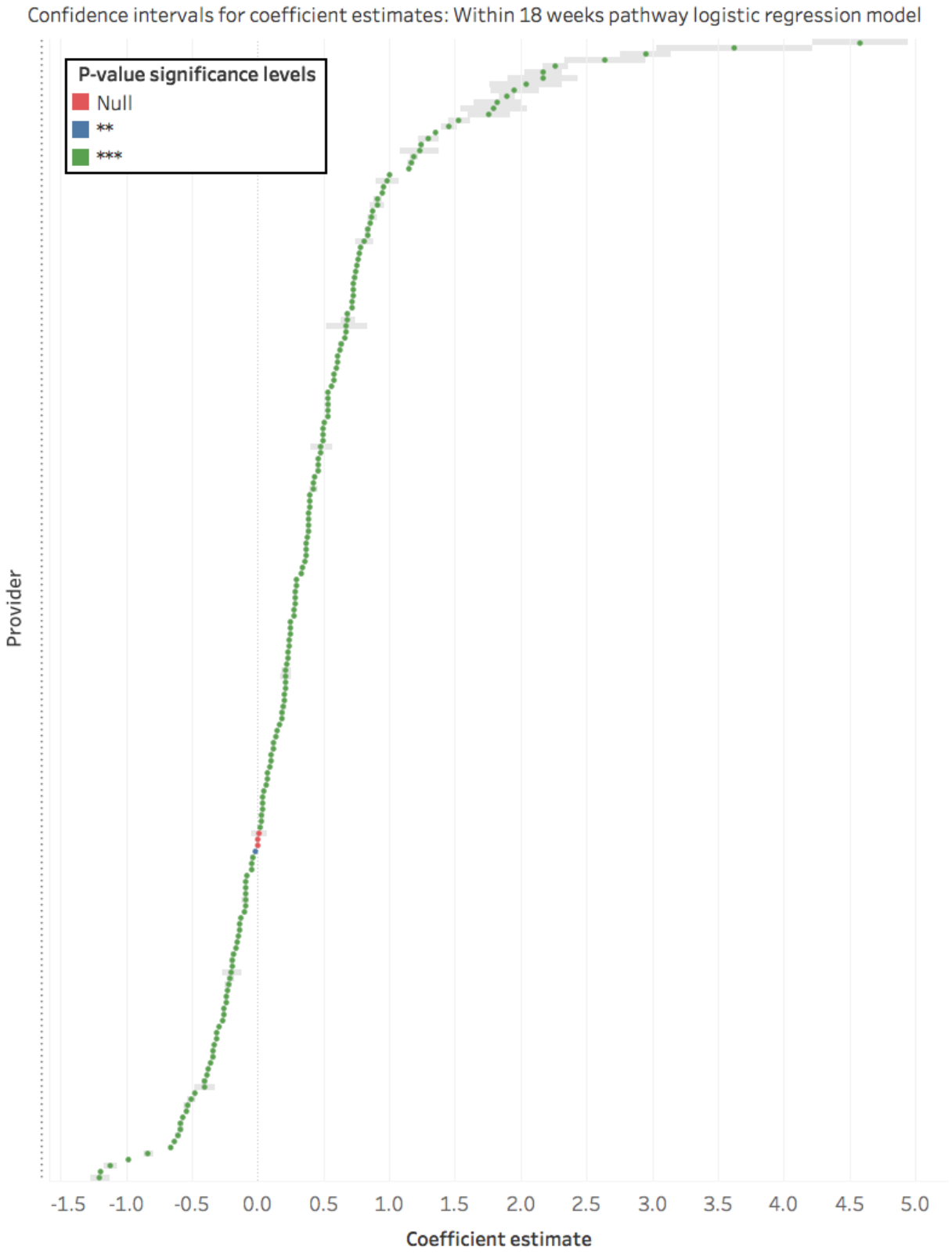

Figure A.3: *Confidence interval for coefficient estimates: within 18 weeks pathway logistic regression model*

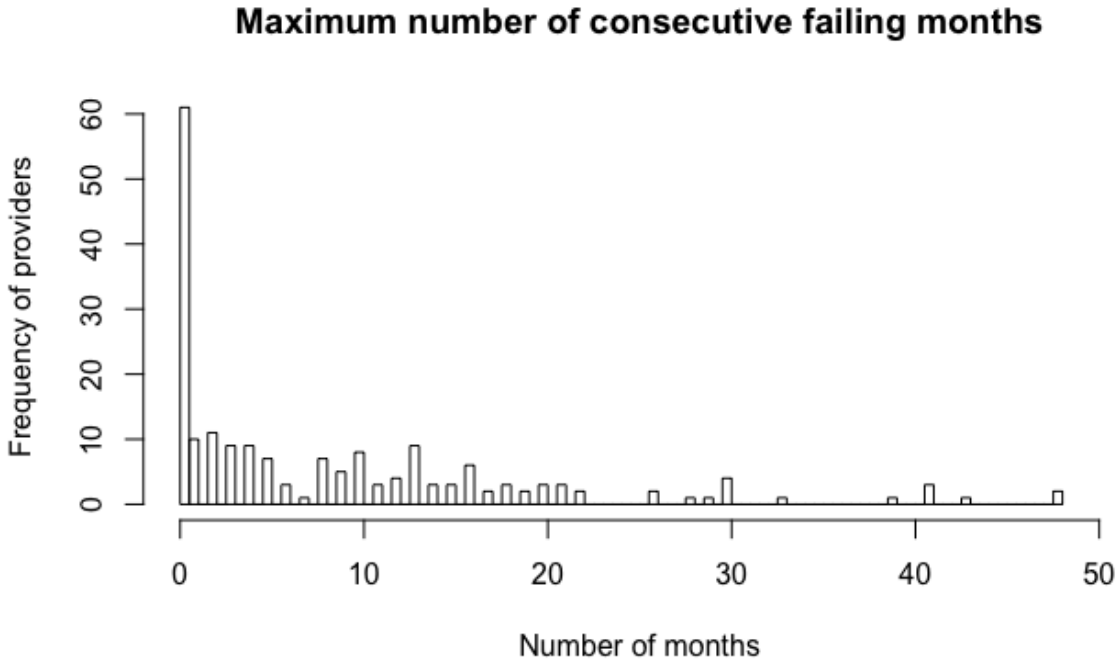

Figure A.4: *Maximum number of months providers consecutively fail.*

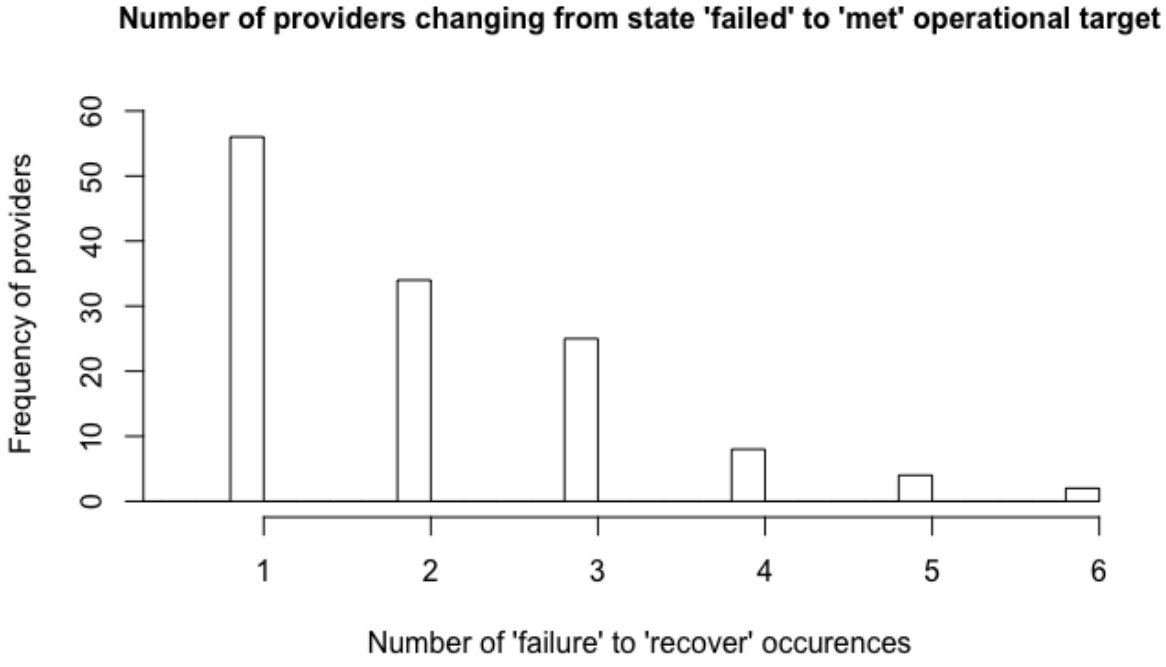

Figure A.5: *Ability of providers to recover from failing operational target. Notably only includes providers who have failed at least one month (61 providers have always passed the operational target between April 2012 - March 2017.*

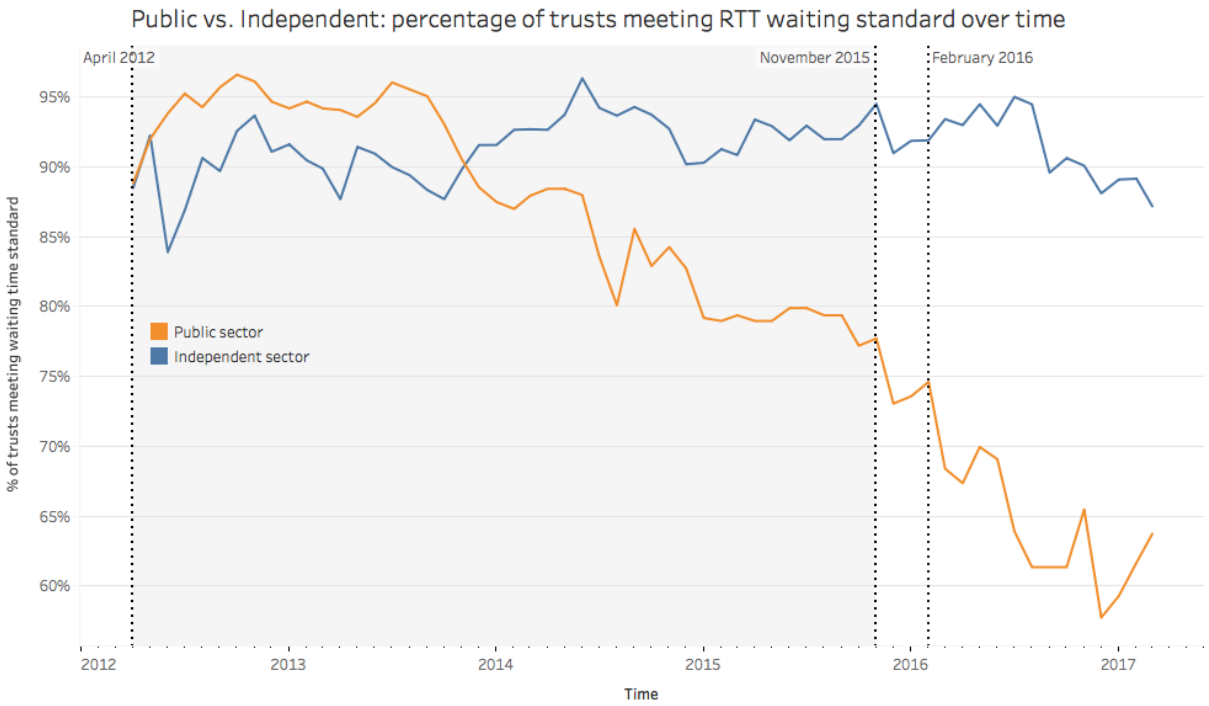

Figure A.6: Public — Independent sector comparison of the percentage of providers meeting operational target over time between April 2012 and March 2017. Reference lines correspond to national performance meeting operational target (see Figure 4.1).

Appendix B

R formulae & output

B.1 R Code

B.1.1 Section 3.5.2 — Chi-square test

```
f <- function(i)

{rate <- 1/Mean[i,]
```

#Calculated rate parameter: 1 / mean number of weeks incomplete pathways have been waiting for.

```
plot(x = Weeks, y = R1A[,i+2], type = 'l')
lines(Weeks, exp(-Weeks*rate), col = "blue")
```

Plot of observed decreasing cdf values ("i+2" avoids unwanted columns) and expected decreasing exponential cdf (in blue).

```
observed <- (R1A[2:54,i+2])*Total[i,1]
expected <- (exp(-Weeks*rate))*Total[i,1]
```

Extracts numerical vector of absolute observed values and absolute expected values.

```
return(chisq.test(observed,p=expected, rescale.p = TRUE,
simulate.p.value = TRUE))}
```

Returns p-value of null hyptohesis (H0). 'rescale.p' ensures expected probabilities add up to 1. 'simulate.p.value' calculates the H0 p-values using Monte-carlo simulation (R default = 2000 iterations).

```
R1A_p_values <- sapply(1:60, f)
```

Apply function for each of the 60 provider months; saved as a vector.

B.2 R output

B.2.1 Section 4.4.1 — Met operational target

```
Call:
glm(formula = Met ~ Year + Month + OrgCode + GeneralSurgery +
Urology + TraumaOrthopaedics + ENT + Ophthalmology + OralSurgery
+ Neurosurgery + PlasticSurgery + CardiothoracicSurgery +
GeneralMedicine + Gastroenterology + Cardiology + Dermatology +
ThoracicMedicine + Neurology + Rheumatology + GeriatricMedicine +
Gynaecology + Other, family = binomial(logit), data = all_in_one)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.7687   0.0000   0.0065   0.3341   2.5851

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)           1.112e+03  7.738e+01  14.372  < 2e-16 ***
Year                 -5.478e-01  3.846e-02 -14.242  < 2e-16 ***
MonthAugust           6.900e-03  1.848e-01   0.037 0.970208
MonthDecember        -3.848e-01  1.800e-01  -2.138 0.032521 *
MonthFebruary         1.709e-01  1.805e-01   0.947 0.343866
MonthJanuary          2.394e-02  1.805e-01   0.133 0.894453
MonthJuly             2.475e-01  1.870e-01   1.324 0.185518
MonthJune             3.655e-01  1.876e-01   1.948 0.051392 .
MonthMarch            1.908e-01  1.791e-01   1.065 0.286663
MonthMay              2.735e-01  1.865e-01   1.467 0.142472
MonthNovember         1.190e-01  1.848e-01   0.644 0.519606
MonthOctober          6.343e-02  1.840e-01   0.345 0.730272
MonthSeptember        2.531e-01  1.876e-01   1.349 0.177337
OrgCodeR1A            1.134e+01  1.329e+03   0.009 0.993189
OrgCodeR1C            1.194e+01  1.329e+03   0.009 0.992833
OrgCodeR1D           -4.127e+00  8.103e-01  -5.092 3.54e-07 ***
OrgCodeR1E            9.863e+00  1.301e+03   0.008 0.993951
OrgCodeR1F           -5.204e+00  6.633e-01  -7.845 4.32e-15 ***
     ⋮                     ⋮          ⋮         ⋮       ⋮
OrgCodeRYX           -4.310e+00  1.234e+00  -3.493 0.000477 ***
OrgCodeRYY           -3.259e+00  1.223e+00  -2.665 0.007689 **
```

```
OrgCodeTAD             1.098e+01  1.333e+03   0.008 0.993425
OrgCodeTAJ             1.098e+01  1.332e+03   0.008 0.993423
GeneralSurgery        -1.036e-03  1.234e-04  -8.394  < 2e-16 ***
Urology               -1.895e-06  2.629e-04  -0.007 0.994248
TraumaOrthopaedics    -1.247e-03  1.240e-04 -10.057  < 2e-16 ***
ENT                   -1.147e-03  1.927e-04  -5.952 2.64e-09 ***
Ophthalmology          3.549e-05  1.053e-04   0.337 0.736170
OralSurgery           -1.477e-04  1.597e-04  -0.925 0.355019
Neurosurgery          -8.834e-04  5.842e-04  -1.512 0.130499
PlasticSurgery        -5.573e-04  3.882e-04  -1.435 0.151153
CardiothoracicSurgery -1.294e-03  7.843e-04  -1.650 0.098932 .
GeneralMedicine        2.908e-04  2.543e-04   1.144 0.252785
Gastroenterology      -8.914e-04  1.974e-04  -4.515 6.35e-06 ***
Cardiology            -5.047e-04  1.983e-04  -2.545 0.010940 *
Dermatology            8.330e-05  1.505e-04   0.554 0.579853
ThoracicMedicine       1.610e-03  3.744e-04   4.302 1.69e-05 ***
Neurology             -7.156e-04  2.980e-04  -2.401 0.016340 *
Rheumatology           1.091e-03  3.161e-04   3.452 0.000556 ***
GeriatricMedicine      2.826e-03  9.355e-04   3.021 0.002517 **
Gynaecology            4.703e-04  2.376e-04   1.979 0.047800 *
Other                 -2.683e-04  6.129e-05  -4.378 1.20e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 10414.7  on 11386  degrees of freedom
Residual deviance:  4775.1  on 11166  degrees of freedom
```

B.2.2 Section 4.4.2 — Pathway within 18 weeks standard

```
Call:
glm(formula = 18WeekSuccess ~ Year + Month + OrgCode + GeneralSu
rgery + Urology + TraumaOrthopaedics + ENT + Ophthalmology + Oral
Surgery + Neurosurgery + PlasticSurgery + CardiothoracicSurgery +
GeneralMedicine + Gastroenterology + Cardiology + Dermatology +
ThoracicMedicine + Neurology + Rheumatology + GeriatricMedicine +
Gynaecology + Other, family = binomial(logit), data = all_in_one)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
```

```
-4.7687   0.0000   0.0065   0.3341   2.5851

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)           1.546e+02  5.941e-01  260.207  < 2e-16 ***
Year                 -7.517e-02  2.952e-04 -254.659  < 2e-16 ***
MonthAugust          -1.843e-02  1.444e-03  -12.761  < 2e-16 ***
MonthDecember        -1.316e-01  1.412e-03  -93.209  < 2e-16 ***
MonthFebruary        -3.648e-02  1.428e-03  -25.545  < 2e-16 ***
MonthJanuary         -6.097e-02  1.432e-03  -42.576  < 2e-16 ***
MonthJuly             1.200e-02  1.449e-03    8.280  < 2e-16 ***
MonthJune             3.294e-02  1.458e-03   22.597  < 2e-16 ***
MonthMarch           -1.353e-02  1.419e-03   -9.536  < 2e-16 ***
MonthMay              5.476e-02  1.466e-03   37.344  < 2e-16 ***
MonthNovember        -5.193e-02  1.432e-03  -36.251  < 2e-16 ***
MonthOctober         -4.740e-02  1.433e-03  -33.088  < 2e-16 ***
MonthSeptember       -2.969e-02  1.444e-03  -20.559  < 2e-16 ***
OrgCodeR1A            1.301e+00  3.921e-02   33.183  < 2e-16 ***
OrgCodeR1C            2.262e+00  4.774e-02   47.389  < 2e-16 ***
OrgCodeR1D           -9.185e-02  1.386e-02   -6.628 3.41e-11 ***
OrgCodeR1E            4.215e-01  1.552e-02   27.154  < 2e-16 ***
     ⋮                     ⋮          ⋮         ⋮       ⋮
OrgCodeRQM           -1.700e-03  6.341e-03   -0.268 0.788631
     ⋮                     ⋮          ⋮         ⋮       ⋮
OrgCodeRXM            1.084e-02  3.099e-02    0.350 0.726452
     ⋮                     ⋮          ⋮         ⋮       ⋮
OrgCodeRYW            4.963e-04  8.773e-03    0.057 0.954889
OrgCodeRYX            2.172e+00  7.468e-02   29.090  < 2e-16 ***
OrgCodeRYY            1.453e+00  3.099e-02   46.881  < 2e-16 ***
OrgCodeTAD            1.953e+00  9.365e-02   20.850  < 2e-16 ***
OrgCodeTAJ            1.796e+00  1.297e-01   13.855  < 2e-16 ***
GeneralSurgery       -5.195e-05  6.383e-07  -81.391  < 2e-16 ***
Urology              -8.005e-05  1.361e-06  -58.803  < 2e-16 ***
TraumaOrthopaedics   -8.010e-05  6.776e-07 -118.216  < 2e-16 ***
ENT                  -1.282e-04  8.833e-07 -145.100  < 2e-16 ***
Ophthalmology        -4.754e-06  5.438e-07   -8.742  < 2e-16 ***
```

```
OralSurgery           -1.878e-05  5.497e-07  -34.172  < 2e-16 ***
Neurosurgery          -1.301e-04  2.203e-06  -59.055  < 2e-16 ***
PlasticSurgery        -1.090e-04  1.911e-06  -57.045  < 2e-16 ***
CardiothoracicSurgery  1.281e-04  6.192e-06   20.683  < 2e-16 ***
GeneralMedicine       -5.351e-05  1.262e-06  -42.395  < 2e-16 ***
Gastroenterology      -1.061e-05  1.026e-06  -10.335  < 2e-16 ***
Cardiology            -2.636e-05  1.003e-06  -26.273  < 2e-16 ***
Dermatology           -9.686e-05  7.816e-07 -123.914  < 2e-16 ***
ThoracicMedicine       1.874e-05  2.212e-06    8.470  < 2e-16 ***
Neurology             -6.519e-05  1.530e-06  -42.606  < 2e-16 ***
Rheumatology           1.727e-04  1.970e-06   87.693  < 2e-16 ***
GeriatricMedicine      1.992e-04  3.499e-06   56.931  < 2e-16 ***
Gynaecology            1.461e-05  1.265e-06   11.548  < 2e-16 ***
Other                  3.009e-06  2.616e-07   11.504  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 4034232  on 11386  degrees of freedom
Residual deviance: 1205986  on 11166  degrees of freedom
```

Bibliography

Brockwell, P. and Davis, R. (2002). *Intoduction to Time Series and Forecasting*. Springer, New York, 2nd edition.

Campbell, D. (2017). NHS 'waving white flag' as it axes 18-week waiting time operation target. *The Guardian (online)*. [Accessed 29 June 2017]. Available from: https://www.theguardian.com/society/2017/mar/31/nhs-surgery-target-operations-cancelled-simon-stevens.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). CRISP-DM 1.0. *SPSS*. [Accessed 29 July 2017]. Available from ttps://www.te-modeling-agency.com/crisp-dm.pdf.

Chatfield, C. (2004). *The Analysis of Time Series: An Introduction*. Chapman & Hall, New York, 6th edition.

Cleveland, R., Cleveland, W., McRae, J., and Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6(1):3–73.

Collett, D. (2015). *Modelling Survival Data in Medical Research*. Chapman & Hall, Boca Raton, FL., 3rd edition.

Department of Health (2004). The NHS Improvement Plan: Putting People at the Heart of Public Services. [Accessed 2 July 2017]. Available from http://webarchive.nationalarchives.gov.uk/20130104225700/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4084476.

Department of Health (2015a). Referral to treatment consultant-led waiting times rules suite: October 2015. [Accessed 8 July 2017]. Available from https://www.gov.uk/government/publications/right-to-start-consultant-led-treatment-within-18-weeks.

Department of Health (2015b). The Handbook to The NHS Constitution. [Accessed 2 July 2017]. Available from https://www.gov.uk/government/uploads/system/uploads/attachment{_}data/file/474450/NHS{_}Constitution{_}Handbook{_}v2.pdf.

Department of Health (2015c). 7-day NHS services: a factsheet. [Accessed 24 August 2017]. Available from https://www.gov.uk/government/publications/7-day-nhs-services-a-factsheet.

Dobson, A. and Barnett, A. (2008). *An Introduction to Generalized Linear Models*. Chapman & Hall, Boca Raton, FL., 3rd edition.

Health and Social Care Information Centre (2015). NHS Sickness Absence Rates. [Accessed 23 August 2017]. Available from https://digital.nhs.uk/catalogue/PUB17903.

Hlavac, M. (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2. Available at: http://CRAN.R-project.org/package=stargazer.

Hope, A. (1968). A Simplified Monte Carlo Significance Test Procedure. *Journal of the Royal Statistics Society. Series B*, 30(3):582–598.

Hyndman, R. and Athanasopoulos, G. (2013). Forecasting: principles and practice. *OTexts (online), Melbourne*. Section 6/7. [Accessed 01 August 2017]. Available from https://www.otexts.org/fpp.

McDonald, J. (2014). *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, Maryland, 3rd edition.

NHS Clinical Commisioners (n.d). About CCGs. *NHS Confederation (online)*. [Accessed 21 August 2017]. Available from: https://www.nhscc.org/ccgs/.

NHS England (2014a). Five Year Forward View. [Accessed 2 July 2017]. Available from https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf.

NHS England (2014b). Understanding the New NHS. [Accessed 9 July 2017]. Available from https://www.england.nhs.uk/wp-content/uploads/2014/06/simple-nhs-guide.pdf.

NHS England (2017a). Referral to treatment (RTT) waiting times statistics for consultant-led elective care - 2016/17. [Accessed 8 July 2017]. Available from https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2017/06/RTT-Annual-Report-2016-17-v0.9-final.pdf.

NHS England (2017b). Next Steps on the NHS Five Year Forward View. [Accessed 28 June 2017]. Available from: https://www.england.nhs.uk/wp-content/uploads/2017/03/NEXT-STEPS-ON-THE-NHS-FIVE-YEAR-FORWARD-VIEW.pdf on 28 June 2017.

Provost, F. and Fawcett, T. (2013). *Data Science for Business*. O'Reilly, California, 1st edition.

Shumway, R. and Stoffer, D. (2000). *Time Series Analysis and Its Applications*. Springer, New York.